

## THÈSE

Pour obtenir le grade de

# DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Sciences pour l'Ingénieur**

Arrêtée ministériel : 25 mai 2016

Présentée par

**Pierre Nagorny**

Thèse dirigée par **Maurice Pillet**  
et codirigée par **Eric Pairel**

préparée au sein du **Laboratoire SYMME**  
dans **l'École Doctorale SISEO**

## Contrôle automatique non-invasif de la Qualité des produits

Application au procédé d'injection-moulage des  
thermoplastiques

Thèse soutenue publiquement le "**Date de soutenance**",  
devant le jury composé de :

**Président**

Laboratoire, Président

**Rapporteur**

Laboratoire, Rapporteur

**Examinateur**

Laboratoire, Examinateur

**Maurice Pillet**

Laboratoire SYMME, Université Savoie Mont Blanc, Directrice de thèse

**Eric Pairel**

Laboratoire SYMME, Université Savoie Mont Blanc, Co-Directeur de thèse

**Invité**

Laboratoire, Invité





## Remerciements

Ces quelques lignes seront nettement insuffisantes pour remercier ceux qui doivent l'être. Je m'excuse par avance de ne pas avoir nominativement cité les nombreuses personnes concernées. Le lecteur se devra d'effectuer une recherche bibliographique pour réunir la liste complète de ces personnes : co-auteurs, membres de groupements de travaux, doctorants de l'École Doctorale SISEO en 2016-2019 et les étudiants de l'école d'ingénieur Polytech Annecy qui ont été d'excellents cobayes d'enseignement.

Je souhaite remercier les membres du jury M. XXX et M. XXX pour l'intérêt qu'ils ont porté à ce travail et à M. XXX pour avoir été le rapporteur. Leurs remarques et suggestions furent précieuses.

Je remercie particulièrement Éric Pairel et Maurice Pillet, mes directeurs de thèse pour :

- leur soutien humain infaillible,
- la direction idéale de ces travaux de recherche,
- le partage de leurs expériences académique, professorale, de recherche.

Le financement de ce travail et son accomplissement, n'aurait pu être obtenu sans leur persévérance. Je remercie BPI France d'avoir financé ce travail, et de manière générale l'effort de la France pour le financement de la Recherche. Merci aux Bibliothèques Universitaires pour leur travail de préservation des ouvrages de références, et à tous les sites de l'Internet qui rendent accessibles une si vaste connaissance.

Cette ligne me permet de remercier l'ensemble des membres du laboratoire SYMME d'Annecy-Le Bourget-du-Lac. Je garde un excellent souvenir de tous les fructueux échanges qui m'ont permis d'approfondir et de faire germer les idées. Merci aux chercheurs, ingénieurs et aux doctorants pour avoir égayés tous ces moments.

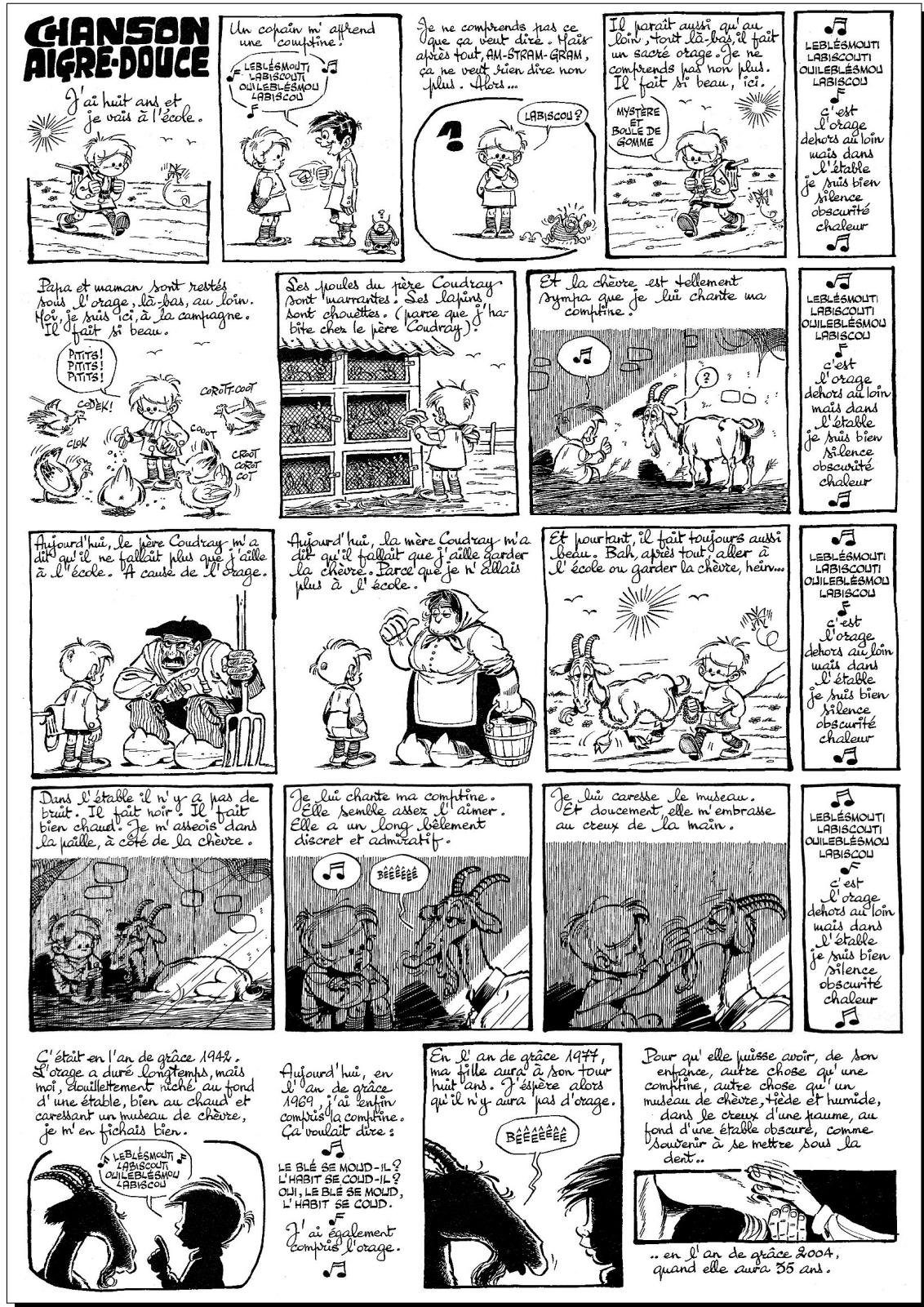
Coté famille, je dois d'énormes mercis à mes grands-parents —Mamé—, mes parents modèles —Maman Monique et Papa Fred—, ma merveilleuse femme Margot et nos joyeux enfants qui ont égayés la rédaction de ce manuscrit. Merci Margot pour tes relectures attentives qui ont permis de corriger ce manuscrit.

Enfin, merci à vous, lecteurs, qui faites vivre ce manuscrit. J'espère qu'il vous permettra d'obtenir des réponses et qu'il saura créer de nouvelles interrogations, tout en vous divertissant avant de vous endormir.

## Dédicace

*À ma généreuse famille, pour tout le soutien reçu, qui m'a permis de mener à bien ce travail.*





© Gotlib / Dargaud <sup>1</sup>

Mes grands-parents ont connu l'orage.

En 2000, j'avais huit ans, il n'y avait pas d'orage chez moi. Je suis allé à l'école.

En 2027, ma fille aura à son tour huit ans. J'espère qu'il n'y aura pas d'orage.

1. Extraits de deux planches de Marcel Gotlib, publiées en Novembre 1969 dans le N°525 de l'hebdomadaire Pilote, puis dans la Rubrique-à-Brac Taume 2 [Got71], publiée aux éditions Dargaud.



# Table des matières

Remerciements & Dédicace . . . . .	i
Épigraphie . . . . .	iii
Table des matières . . . . .	v
Résumé . . . . .	ix
Notations . . . . .	xi
<b>Introduction</b>	<b>1</b>
<b>1 Contexte industriel et objectifs de recherche</b>	<b>7</b>
1.1 Maîtrise du procédé d'injection-moulage des thermoplastiques . . . . .	8
1.1.1 Cartographie bibliographique . . . . .	8
1.1.2 Modélisation du procédé d'injection-moulage . . . . .	9
1.1.3 Vers le pilotage du point de fonctionnement : Étude bibliographique . . . . .	14
1.1.4 La mesure de la qualité du produit : un verrou technologique au pilotage . . . . .	23
1.2 Objectifs de recherche : enjeux du déploiement du contrôle qualité en ligne de production . . . . .	25
1.2.1 Viabilité économique du déploiement de la mesure de la qualité en ligne de production . . . . .	25
1.2.2 Faisabilité technologique de la mesure de la qualité en ligne de production . . . . .	27
1.2.3 Formalisation de la notion de qualité et intégration de l'expert humain . . . . .	28
<b>2 Mesure industrielle non-invasive de la qualité</b>	<b>31</b>
2.1 Qualité d'un produit en injection-moulage des thermoplastiques . . . . .	32
2.2 Mesures en ligne de production . . . . .	33
2.2.1 Mesures indirectes des caractéristiques du produit . . . . .	36
2.2.2 Mesures directes des caractéristiques du produit . . . . .	38
2.2.3 Définition du degré d'invasivité d'un système de mesure pour un procédé . . . . .	39
2.2.4 Corrélation entre les caractéristiques des pièces chaudes et des pièces finales . . . . .	41
2.3 Apport de l'imagerie non-conventionnelle pour la mesure de la qualité . . . . .	43
2.3.1 Mesure géométrique par triangulation laser . . . . .	44
2.3.2 Apport de la thermographie pour le contrôle de la géométrie des pièces . . . . .	46
2.3.3 Apport de la polarimétrie pour le contrôle des défauts d'aspect . . . . .	52
2.4 Traitement et fusion de l'information issue de mesures multimodales . . . . .	59
2.4.1 Nécessité de l'extraction de l'information pertinente . . . . .	60
2.4.2 Traitement d'images amont . . . . .	60
2.4.3 Descripteurs statistiques . . . . .	61
2.4.4 Sélection automatique de descripteurs pertinents . . . . .	65
2.4.5 Description d'images par apprentissage statistique . . . . .	65
2.5 Synthèse et perspectives . . . . .	67
2.5.1 Synthèse : mesure non-invasive en cycle industriel . . . . .	67
2.5.2 Perspectives de recherche . . . . .	67

<b>3 Modélisation de la notion de qualité par apprentissage</b>	<b>72</b>
3.1 Construction du jeu de données d'apprentissage à partir de l'expertise humaine . . . . .	75
3.1.1 Exploration du procédé par plan d'expériences . . . . .	75
3.1.2 Apprentissage en ligne . . . . .	76
3.2 Annotation humaine du jeu de données en contexte industriel . . . . .	78
3.2.1 Annotation hors ligne . . . . .	79
3.2.2 Annotation en ligne de production . . . . .	81
3.2.3 Intégration de l'humain dans le système d'apprentissage . . . . .	81
3.2.4 Assistance à l'annotation humaine . . . . .	82
3.3 Construction d'une métrique de la qualité par apprentissage . . . . .	84
3.3.1 Apprentissage supervisé . . . . .	84
3.3.2 Apprentissage par transfert de domaine . . . . .	103
3.3.3 Apprentissage semi-supervisé . . . . .	104
3.3.4 Apprentissage non-supervisé . . . . .	112
3.4 Optimisation automatique des hyper-paramètres d'un modèle . . . . .	129
3.4.1 Optimisation par recherche aléatoire . . . . .	130
3.4.2 Optimisation par plan d'expériences . . . . .	131
3.4.3 Optimisation par métamodèle bayésien . . . . .	132
3.4.4 Méthodes de réduction du coût d'évaluation du métamodèle . . . . .	134
3.4.5 Optimisation par algorithmes bandits . . . . .	136
3.4.6 Optimisation par algorithmes évolutionnistes . . . . .	136
3.4.7 Optimisation de l'architecture de réseaux de neurones profonds . . . . .	137
3.4.8 Conclusion : AutoML pour l'industrie . . . . .	137
3.5 Conclusion : Modélisation par apprentissage pour l'industrie . . . . .	138
<b>4 Conception d'un système de mesure de la Qualité</b>	<b>142</b>
4.1 Conception d'un dispositif de mesure adapté aux contraintes industrielles . . . . .	143
4.1.1 Imagerie non-conventionnelle pour la mesure non-invasive de la Qualité .	143
4.1.2 Détection de pièce robuste . . . . .	143
4.2 Solution logicielle . . . . .	143
4.2.1 Architecture logicielle . . . . .	143
4.2.2 Interface Homme-Machine . . . . .	143
4.2.3 Apprentissage de la notion de Qualité . . . . .	144
4.3 Application industrielle . . . . .	144
4.3.1 Contrôle de la qualité par apprentissage en industrie . . . . .	144
4.3.2 Résultats obtenus par le dispositif de mesure TheEye . . . . .	144
4.4 Évaluation de la performance du système de mesure de la Qualité . . . . .	144
4.4.1 Comparaison avec un expert humain . . . . .	144
4.4.2 Métrique de confiance industrielle . . . . .	144
<b>Conclusion générale</b>	<b>146</b>
<b>Bibliographie</b>	<b>I</b>
<b>Table des figures</b>	<b>XXVII</b>
<b>Liste des tableaux</b>	<b>XXXI</b>

<b>A Annexe 1 : Titre</b>	<b>XXXIII</b>
A.1 Section 1 . . . . .	XXXIV
A.1.1 Sous section 1 . . . . .	XXXIV
A.1.2 Sous section 2 . . . . .	XXXIV
<b>B Annexe 2 : Titre</b>	<b>XXXV</b>
B.1 Section 1 . . . . .	XXXVI
B.1.1 Sous section 1 . . . . .	XXXVI
B.1.2 Sous section 2 . . . . .	XXXVI
<b>C Annexe 3 : Titre</b>	<b>XXXVII</b>
C.1 Section 1 . . . . .	XXXVIII
C.1.1 Sous section 1 . . . . .	XXXVIII
C.1.2 Sous section 2 . . . . .	XXXVIII



---

## Résumé

Le procédé d'injection moulage des thermoplastiques permet la production industrielle de produits, dont les exigences de qualités géométriques et d'aspects sont croissantes. Maîtriser la qualité du produit nécessite de maîtriser le procédé d'injection. Les contraintes posées par le respect du cycle industriel sont nombreuses. Ces travaux de doctorats s'inscrivent dans la thématique du contrôle des produits, spécifiquement intégré sur la ligne de production. Les technologies de mesures non invasives permettent le respect des contraintes industrielles, ainsi qu'une simplification de la mise en place et une mobilité dans l'atelier. L'apport de l'imagerie non-conventionnelle, avec la thermographie et la polarimétrie, est étudiée. La classification automatique des pièces selon leur qualité est effectuée à l'aide d'un modèle de la qualité de la pièce construit par apprentissage supervisé. Un dispositif de contrôle de la qualité des produits, dès la sortie du moule, est présenté. Ce dernier utilise l'information issue de multiples capteurs et l'apprentissage supervisé itératif du modèle, au fil de la production. Plusieurs cas d'études industriels permettent d'évaluer les performances et la viabilité de la solution retenue.

## Abstract

The injection molding process of thermoplastics allows the industrial production of products, whose requirements of geometric qualities and aspects are increasing. Mastering the quality of the product requires mastering the injection process. The constraints posed by the respect of the industrial cycle are numerous. This doctoral work is part of the product control theme, specifically integrated into the production line. Non-invasive measurement technologies allow compliance with industrial constraints, simplification of implementation and mobility in the workshop. The contribution of non-conventional imaging, with thermography and polarimetry, is studied. The automatic classification of parts according to their quality is carried out using a model of the quality of the part built by supervised learning. A device for checking the quality of the products, right out of the mold, is presented. The latter uses information from multiple sensors and the iterative supervised learning of the model, during production, during the process adjustment phase. Several cases of industrial studies make it possible to evaluate the performance and the viability of the chosen solution.



---

# Notations

## Statistique

$p(x)$	Distribution de probabilité des valeurs $x$
$\hat{x}$	Approximation/prédiction des valeurs de $x$
$\bar{x}$	Moyenne des valeurs $x$
$\sigma$	Ecart-type
$\sigma^2$	Variance
$\mathcal{N}$	Distribution Gaussienne

## Algorithmique et informatique

$O(n)$	Complexité algorithmique Grand O (O( $n$ ) note ici une complexité linéaire)
GPU	<i>Graphical Process Unit</i> : Processeur graphique qui accélère les calculs massivement parallèles
RAM	<i>Random Access Memory</i> : Mémoire opérationnelle nécessaire au support des calculs

## Apprentissage automatique

Modèle	Application qui réalise une tâche (classification, régression, etc.)
Apprentissage	Ajustement des poids du modèle pour réaliser l'objectif
Hyper-paramètres	Variables qui modifient les performances d'un algorithme d'apprentissage
$\mathcal{L}$	Fonction de coût que l'on cherche à minimiser lors de l'apprentissage d'un modèle
<b>W</b>	Matrice des poids du modèle
<b>x</b>	Variables d'entrée du modèle
$y$	Variable de sortie du modèle
$\ \mathbf{x}\ _1$	Norme $\ell_1$ , absolue de $\mathbf{x}$
$\ \mathbf{x}\ _2^2 = \ \mathbf{x}\ ^2$	Norme $\ell_2$ , euclidienne de $\mathbf{x}$
$\eta$	Facteur d'apprentissage, coefficient d'ajustement des poids à chaque itération
$\theta$	Paramètres du modèle, tel que $f_\theta(x)$ ait pour variable $x$ et pour paramètres $\theta$
ACP	Analyse en Composante Principale

Dans ce travail, nous utiliserons les conventions de notations suivantes : les variables scalaires sont en minuscule ( $x$ ), les vecteurs sont en minuscule grasse ( $\mathbf{x}$ ) et les matrices sont en majuscule grasse ( $\mathbf{X}$ ). Lorsque cela est nécessaire, ces notations sont augmentées en indice d' identifiants de références ( $x_a$ ) ou d'indexation des valeurs des vecteurs ( $\mathbf{x}_i$  pour  $i \in [0; k]$ ).



## Introduction générale

*L*'industrie manufacturière est en quête croissante d'une production à la qualité optimale. Le besoin du client final est la fonction du produit. Des besoins fonctionnels techniques, comme des besoins fonctionnels d'aspects, apparaissent aujourd'hui sur les mêmes pièces, avec des niveaux d'exigences croissants. La notion de qualité d'un produit est compliquée à spécifier. De nombreuses normes permettent de définir des tolérances sur les produits. Cependant, les normes ne permettent pas de spécifier toutes les caractéristiques d'un produit. C'est le cas de la spécification de la qualité d'aspect. Chaque profession a construit sa définition et chaque entreprise du secteur l'a enrichie. L'exploitation de ces définitions, afin de construire une métrique de la qualité, est un véritable défi industriel.

Pour atteindre la qualité optimale, il est nécessaire de maîtriser le procédé de production, voir dans le meilleur des cas, de le réguler par optimisation en boucle fermée. La maîtrise de la qualité s'obtient après une démarche itérative. Il nécessite de mesurer la qualité, ce qui demande la mise en place de système de métrologie couplée à une métrique de la qualité adaptée. La mesure robuste de la qualité permet de mettre sous contrôle le procédé par des moyens statistiques afin de détecter les dérives dans la qualité produite. Enfin, il devient possible d'effectuer une rétro-action sur les réglages du procédé afin d'optimiser la qualité produite.

La non-qualité est l'ensemble des coûts qui sont engendrés par la fabrication de produits qui ne répondent pas au cahier des charges (ce sont des produits de "mauvaise" qualité). Les problématiques environnementales et économiques convergent autour de la nécessité de limiter la non-qualité. Il est coûteux tant économiquement qu'environnementalement de procéder aux étapes successives de finitions du produit alors que celui-ci était de mauvaise qualité dès la première étape. Écarter les produits ne répondant pas au cahier des charges, dès la fin de l'étape du procédé, permet un gain économique conséquent. La gestion du cycle de vie des produits de mauvaise qualité est également rendue possible. Un circuit spécifique de recyclage peut être mis en place à chaque étape de production. Plus encore, détecter une dérive de qualité produite sur le procédé permet d'alerter et de stopper le moyen de production. Sur des procédés de productions stables où les opérateurs humains interviennent peu, cela permet d'économiser la matière première de plusieurs centaines, voire de plusieurs milliers de pièces.

Plusieurs contraintes industrielles et technologiques entrent en jeu dans la conception d'un système de mesure de la qualité. En premier lieu, le temps de cycle du procédé industriel doit être respecté. Il ne doit pas être impacté par la technologie de mesure. La notion de qualité doit être spécifiée afin de réaliser une mesure répétable. Enfin, le coût économique du moyen de métrologie doit être inférieur au coût de la production de rebuts.

L'injection-moulage des thermoplastiques est un procédé à haute cadence, peu coûteux et répétable. Une fois la machine de production réglée, le procédé est souvent très stable. Il peut produire de manière continue plusieurs milliers de pièces sans intervention humaine. Le coût de la matière première pour le moulage est faible en comparaison du coût des étapes de finitions suivantes, telles que les étapes de peintures. Éviter aux pièces de mauvaise qualité d'être peintes est une priorité. En second temps, un système de rétroaction pourra être déployé sur le procédé, afin de maximiser la qualité produite.

Le procédé d'injection-moulage des thermoplastiques consiste à injecter un polymère visqueux dans un outillage massif refroidi ; à maintenir une pression de compactage pendant une durée spécifique ; puis à éjecter la pièce tout en préparant en parallèle la matière nécessaire au cycle suivant. Il est possible d'ajuster précisément les pressions et les durées d'ouvertures et de fermetures de multiples buses d'injection de manière séquentiel. Ainsi, l'espace des variables de pilotage est grand et les variables sont continues.

La production de pièce par moulage des thermoplastiques suit généralement la séquence suivante :

1. Injection-moulage de la pièce
2. Stockage tampon des pièces
3. Étapes de peintures
4. Stockage
5. Expédition

Afin de minimiser les coûts d'immobilisation du stockage, on cherche à réduire au maximum le stockage entre les étapes. Les défauts géométriques et d'aspect apparaissent majoritairement lors des étapes de déplacements des pièces entre les étapes successives de la production. Pour des pièces techniques au cahier des charges rigoureux, un contrôle qualité avant expédition est donc obligatoire. Les défauts géométriques et d'aspect sont également produits lors de l'étape initiale d'injection-moulage. De nombreux défauts dit "d'aspect" sont des défauts géométriques à l'échelle de la centaine de micromètres. Ceux-ci modifient les réflexions des lumières incidentes et cause des changements de luminosité brusques, qui ne sont pas acceptables pour le cahier des charges du produit fini. Les étapes de peintures peuvent mettre en valeurs des défauts en introduisant des surfaces spéculaires.

### **Le FUI SAPRISTI : un projet de recherche collaborative de 2016 à 2019**

L'obtention d'une chaîne de production compétitive passe par le développement, l'optimisation et l'automatisation de chacune de ces étapes. Dans cette optique, différents consortiums constitués d'industriels et d'académiques travaillent sur des projets de recherche collaboratives (FUI). Ces travaux s'inscrivent dans le projet collaboratif FUI<sup>2</sup> SAPRISTI<sup>3</sup> du 21<sup>ème</sup> appel à projet FUI Régions.<sup>4</sup> Le projet FUI SAPRISTI est labellisé par trois pôles de compétitivités :

- Plastipolis,
- iDforCAR,
- Mont-Blanc Industries.

Les partenaires du consortium FUI SAPRISTI sont :

- Acsystème, développeur et intégrateur de solutions en automatique et traitement de l'information,
- Le Centre Technique CT-IPC, acteur européen en développement et recherche pour le secteur de la plasturgie,
- Le laboratoire LaMCoS de l'INSA de Lyon, expert de la modélisation du procédé d'injection-moulage des thermoplastiques,

---

2. FUI : Fond Unique Interministériel

3. SAPRISTI : Système Autocorrectif pour la PRoduction zéro défaut de plèces pLaSTIques

4. [Résultats du 21<sup>ème</sup> appel à projets FUI sur competitivite.gouv.fr](http://www.competitivite.gouv.fr) : 60 projets bénéficieront d'une aide de l'État de 43 Millions d'Euros ainsi que d'une aide des collectivités territoriales et des fonds communautaires (FEDER) de 37,8 Millions d'Euros.

- [Plastic Omnium](#), équipementier automobile mondial
- [Renault](#), automobiliste concepteur et équipementier de ses équipements,
- [SISE](#), acteur européen de la régulation des procédés en plasturgie, dont notamment le contrôle de l'injection séquentiel,
- Le [laboratoire SYMME<sup>5</sup>](#) de l'[Université Savoie Mont Blanc](#), expert en réglage automatique et en optimisation de procédés industriels, dans le cadre de son axe de recherche sur la Maîtrise de la Qualité des Produits pour l'Industrie.

L'injection-moulage des thermoplastiques est un procédé de production bas coût haut-volume (*Low-Cost, High Volume*) particulièrement prisé des industriels. Le marché de la plasturgie représente en France 3500 entreprises pour 30,2 Milliards de Chiffre d'Affaires en 2018<sup>6</sup>. Dans ce contexte, l'enjeu du projet FUI SAPRISTI est triple. Le projet a pour triple objectifs de :

- Déterminer des indicateurs pertinents à mettre sous contrôle.
- Mieux comprendre les phénomènes physiques qui entrent en jeu dans le procédé, afin de modéliser le procédé en prenant en compte les indicateurs sélectionnés.
- Optimiser le procédé en boucle fermé à partir de la modélisation et des indicateurs, afin de produire une qualité optimale.

Dans ce cadre, les objectifs pour le laboratoire SYMME sont dans un premier temps de trouver quelles sont les variables pertinentes à étudier sur le procédé. Ces variables peuvent être des mesures directes sur le procédé ou bien des mesures réalisées sur les produits. Dans cette démarche, les contraintes industrielles de sont pris en compte pour choisir des moyens de mesures compatibles avec le procédé. Ensuite, il s'agit de modéliser les effets de ces variables avec la qualité finale des produits. Pour cela, nous utiliserons des modèles construits par apprentissage statistique. L'objectif final du FUI SAPRISTI est de mettre sous contrôle le procédé afin de maximiser la qualité produite. Cet objectif ne peut être atteint que si les précédents objectifs sont résolus. Cette nécessité de disposer d'un moyen de contrôle automatique de la qualité des produits a donné lieu à notre problématique de recherche.

## Problématique de recherche

Dans le cas de l'injection-moulage des thermoplastiques, aucun système de mesure de la qualité ne répond actuellement aux exigences des partenaires du FUI SAPRISTI ; que se soient en matière de robustesse de la mesure, ou en matière de compatibilité avec les contraintes posées par le procédé. Dans ce cadre, ce travail de doctorat cherche à développer un dispositif de mesures non-invasif de la qualité des produits, dès la sortie de la machine, qui répond aux problématiques industrielles.

Nous avons identifié trois verrous au déploiement massif du contrôle qualité à cent pour-cent dès la sortie de la machine :

- économique
- technologique
- humain

Nous évaluerons et chercherons à résoudre chacun de ces verrous afin de proposer un moyen de mesure industriellement viable. Nous présentons les méthodes et les moyens retenus pour lever ces verrous.

---

5. SYMME : SYstèmes et Matériaux pour la MÉcatronique

6. Synthèse 2018 du secteur réalisée par la Fédération Française de la Plasturgie.

Des échanges avec les partenaires du FUI SAPRISTI nous ont permis de prendre un compte les limites que posent le procédé industriel. La pratique industrielle actuelle propose d'utiliser des mesures intégrées dans l'outillage. L'intégration des capteurs nécessaire à ces mesures est invasive : l'ingénierie nécessaire à la conception de l'outillage avec capteur est couteuse ; l'instrumentation des capteurs est fragile et la maintenance nécessite un arrêt prolongé de la production puisque c'est l'outillage qui doit être expertisé.

La notion de qualité est subjective. Malgré les normes, la notion de qualité est souvent dépendante des connaissances holistiques du responsable expert qualité. Il est difficile de modéliser ce savoir dans un système de contrôle automatisé. Les récentes avancées en apprentissage statistique, dont les réseaux de neurones profonds (*Deep Learning*), permettent de concevoir un système qui intègre un modèle de la notion de qualité. Ce modèle est construit à partir de l'expertise humaine, par apprentissage supervisé. L'apprentissage du modèle est réalisé au fur et à mesure de la production, de manière à être intégré au procédé de production. Ce dispositif de mesure s'appuie également sur l'apport de l'imagerie non-conventionnelle qui permettent d'acquérir des informations que l'œil de l'expert humain ne peut percevoir. Des essais expérimentaux réalisés avec un prototype de dispositif de mesures montrent des résultats encourageants.

### Démarche de recherche

Afin d'identifier les variables du procédé d'injection-moulage des thermoplastiques qui sont susceptibles de contenir une information sur la qualité finale du produit, nous avons étudié la littérature. Notre état de l'art propose une revue des publications majeures sur le pilotage du procédé d'injection-moulage, depuis 1975 à ce jour.

Afin de répondre aux verrous qui limitent le déploiement du contrôle qualité, nous n'avons pas retenu la mesure invasive dans nos travaux. Aussi, nous avons choisi de définir le périmètre de notre recherche au contrôle qualité non-invasif pour le procédé industriel. Cela permet de limiter la complexité de la mise en place des dispositifs de mesure.

L'expérimentation occupe une part importante de ces travaux. Dans un premier temps, cela nous a permis de comprendre le procédé et la pratique industrielle. Nous avons étudié l'utilisation de moyens de mesures issus de l'imagerie non-conventionnel : scanner laser, thermographie, polarimétrie, ainsi que de mesures intégrées dans l'outillage que nous n'avons pas retenu. Par la suite, nous avons cherché à valider par itérations successives notre dispositif de mesure. Notre démarche de recherche est phénoménologique. Nous nous appuyons sur des expériences, plus que sur les théories chimico-thermo-dynamique des polymères. Dans le cas où l'on dispose d'un grand nombre de cas d'études (échantillons), cette démarche est particulièrement adaptée à l'utilisation de l'apprentissage statistique pour modéliser les phénomènes.

À partir des mesures, il s'agit de concevoir un système capable de discerner les pièces de bonne qualité, des pièces de mauvaise qualité. Nous orientons nos travaux dans l'utilisation de méthode d'apprentissage supervisé. L'expert qualité humain transmet son savoir holistique au système qui réalise le contrôle de manière autonome.

Le développement du système de mesures s'inscrit dans une démarche de validation de l'ensemble de nos propositions. Les modèles construits par apprentissage statistique nécessitent un jeu de données d'apprentissage. Posséder un dispositif de mesures nous permet d'enregistrer les données requises pour l'apprentissage du modèle. Le dispositif de mesures est conçu par une intégration verticale de l'imagerie au traitement des données. Cela permet de maîtriser l'ensemble de la chaîne d'acquisition et d'analyse des données, et de pouvoir l'optimiser dans son ensemble, pour répondre à notre problème. La validation du système est alors réalisée en condition industrielle.

## Organisation du manuscrit

Ce manuscrit est constitué de quatre chapitres qui peuvent être indépendamment parcourus. Le lecteur sera orienté au gré de sa lecture, vers les sections d'autres chapitres, lorsque des informations complémentaires pourront lui être utiles. Le premier chapitre 1 présente de manière détaillée (i) le contexte industriel du travail de doctorat; (ii) les objectifs de recherche pour le déploiement du contrôle qualité dès la sortie de l'injection-moulage. Le deuxième chapitre 2 a pour objectifs de (i) présenter les différentes technologies de métrologies existantes; (ii) le choix de l'imagerie non-conventionnelle. Le troisième chapitre 3 propose d'étudier les moyens de formalisation de la notion de Qualité par apprentissage et de les évaluer sur notre cas d'application industriel. Le quatrième chapitre 4 détaille (i) la conception d'un nouveau dispositif de mesures ; (ii) son évaluation industrielle. Enfin, en conclusion de ce manuscrit un bilan concernant les travaux de ces trois années de doctorat est dressé et une discussion est proposée concernant les nombreuses perspectives.

## Contributions scientifiques

Ces travaux de doctorat proposent une large étude des techniques disponibles, qui sont requises pour mettre en place le contrôle automatique des produits sur les chaînes de production. Le travail s'intéresse à l'apport de l'imagerie non-conventionnelle, aux méthodes d'apprentissages automatiques et à la conception d'un système complet de contrôle automatique de la qualité.

Ces travaux ont permis le développement d'un système de contrôle automatique de la qualité des pièces plastiques, dès la sortie du moule, de manière non-invasive, sans contact avec la pièce et avec un impact minimal sur le procédé. Ce système associe un banc de mesure spécialement développé pour répondre aux contraintes industrielles, ainsi qu'un logiciel afin d'intégrer et de réaliser le traitement des sources d'information. Le logiciel utilise les possibilités de calculs distribués offertes par les solutions *Cloud*. Une interface homme-machine, sous la forme d'*application web*, a été développé pour permettre à un expert humain de transmettre sa connaissance à la machine. Ce système a été évalué en conditions industrielles, sur le procédé d'injection-moulage des thermoplastiques.

Ces travaux de doctorat ont fait l'objet de communications en congrès scientifiques internationaux [NPP17b; NPP17a; Nag+17; Nag+18; Nag+19], ainsi que d'un colloque national [NPP17b]. Un Article en revue à Comité de Lecture a été soumis fin Septembre 2019.

## Valorisation scientifique

Notre système de mesures de la qualité a été lauréat du sixième concours *OutOfLab*<sup>7</sup>, organisé par la SATT<sup>8</sup> [Linksium Grenoble Alpes](#). Ce prix nous permet de poursuivre la recherche et les développements, vers une application commerciale.

Le dispositif de contrôle de la qualité conçu pendant ces travaux de doctorat a fait l'objet d'un dépôt de brevet : "United States Patent Application" [NPP19].

---

7. Résultats de la sixième édition du concours *OutOfLabs*

8. SATT : Société d'Accélération du Transfert de Technologies



# 1

## Contexte industriel et objectifs de recherche

---

*Dans ce premier chapitre, nous positionnons notre travail par rapport à l'état de l'art de l'existant, en injection-moulage des thermoplastiques. Nous identifions des démarches qui permettent d'améliorer la modélisation du procédé. Enfin, nous identifions les verrous au déploiement du contrôle de la qualité à cent pourcent, ainsi que nos objectifs de recherche pour les lever.*

---

### Sommaire

---

<b>1.1</b>	<b>Maîtrise du procédé d'injection-moulage des thermoplastiques . . . . .</b>	<b>8</b>
1.1.1	Cartographie bibliographique . . . . .	8
1.1.2	Modélisation du procédé d'injection-moulage . . . . .	9
1.1.3	Vers le pilotage du point de fonctionnement : Étude bibliographique .	14
1.1.4	La mesure de la qualité du produit : un verrou technologique au pilotage	23
<b>1.2</b>	<b>Objectifs de recherche : enjeux du déploiement du contrôle qualité en ligne de production . . . . .</b>	<b>25</b>
1.2.1	Viabilité économique du déploiement de la mesure de la qualité en ligne de production . . . . .	25
1.2.2	Faisabilité technologique de la mesure de la qualité en ligne de production	27
1.2.3	Formalisation de la notion de qualité et intégration de l'expert humain .	28

---

## 1.1 Maîtrise du procédé d'injection-moulage des thermoplastiques

Dans cette section, nous réalisons une synthèse des techniques employées dans l'industrie afin de maîtriser la variabilité du procédé d'injection-moulage. Il s'agit dans un premier temps d'effectuer des mesures pertinentes sur l'état du procédé, puis de les analyser afin de pouvoir proposer une modification des réglages adaptée.

### 1.1.1 Cartographie bibliographique

Notre étude bibliographique “Pilotage En Injection Plastique - État de l’Art” [NPP17a] nous permet de proposer un graphe relationnel à partir des mots-clés associés à 421 publications, de 1970 à 2015, Figure 1.1.

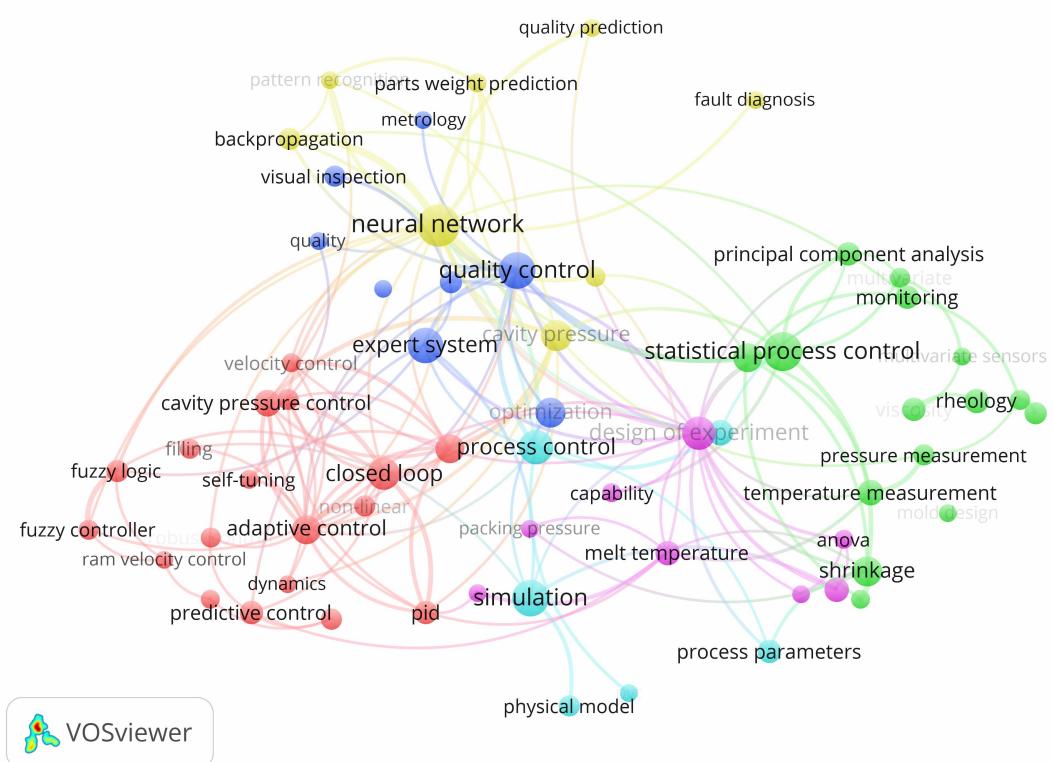


Figure 1.1 – Cartographie bibliographique de la maîtrise du procédé d'injection-moulage.

Les mots-clés sont retenus pour l'affichage s'ils apparaissent plus de quatre fois. Ils sont regroupés par liens communs. Nous utilisons le formalisme *VOS* (*Visualization Of Similarities*) proposé par VAN ECK et WALTMAN, pour modéliser notre base bibliographique [vW06 ; van+10]. Nous identifions des axes de recherches importants :

- la modélisation physique et la simulation du procédé (couleur cyan),
- la régulation du point de fonctionnement par la théorie du contrôle (couleur rouge),
- l'analyse statistique (couleur verte),
- l'apport des systèmes experts pour régler (couleur bleue),
- l'apport des systèmes à réseaux des neurones (couleur jaune).

L'analyse statistique est utilisée pour détecter les défaillances du procédé. Elle s'appuie sur la mesure *in-situ* de variables du procédé, parmi lesquelles les plus utilisées sont la température

et la pression. Les réseaux de neurones sont souvent utilisés pour prédire les caractéristiques des pièces. Les caractéristiques de qualités du produit, comme la masse ou les dimensions, sont calculées par les réseaux à partir de la mesure de la pression dans le moule. La littérature propose de réguler les paramètres du procédé d'injection à partir de la prédiction de caractéristiques des pièces. La prédiction à partir de variables du procédé est préférée à une mesure directe sur la pièce, car la métrologie n'est pas assez développée pour permettre une mesure pièce à pièce en cycle industriel. À contrario, l'enregistrement des mesures de pression pendant le cycle est très répandue, car elle est implantée par les fabricants de moules et les fabricants de systèmes de régulations du procédé. Enfin, nous distinguons (de couleur rouge) les efforts pour la régulation du procédé. De nombreuses méthodes ont été étudiées et nous remarquons l'intérêt porté aux méthodes adaptatives. Cependant, chacune de ces méthodes ne s'intéressent qu'à un nombre limité de variables du procédé et des caractéristiques du produit.

### 1.1.2 Modélisation du procédé d'injection-moulage

Le procédé d'injection-moulage des thermoplastiques est cyclique et séquentiel. Chaque phase possède des paramètres qui influent de manière non-linéaire sur les phases suivantes et, à terme, sur les caractéristiques du produit fini. Le moulage par injection est l'étape clé du processus de fabrication des pièces plastiques. Elle répercute ses défauts sur l'ensemble des étapes suivantes. Et ce sont les étapes de finitions qui ajoutent le plus de valeur au produit. Un défaut dans l'une des phases impactera l'ensemble de la chaîne de production et diminuera le rendement global du moyen de production.

Afin de mettre en évidence la séquence d'un cycle du procédé, nous proposons un schéma bloc d'un cycle d'injection, Figure 1.2, dans lequel le dosage est réalisé en parallèle du refroidissement.

#### 1.1.2.1 Proposition d'un modèle systémique composé de phases interdépendantes

Le moulage par injection plastique se décompose en phases interdépendantes, qui conditionnent le produit fini. KAZMER et HATCH proposent un découpage en cinq phases [KH99], Figure 1.3, qui est une évolution du découpage en trois phases proposé par MA [Ma74]. Cette représentation systémique du procédé fait apparaître les variables d'état intermédiaires. Elle fait également apparaître les relations de dépendance entre les phases, qui sont la cause des caractéristiques du produit. La dimension temporelle n'est pas prise en compte. Cela ne permet pas de faire apparaître les opérations qui sont réalisées en parallèles, comme la phase de dosage et la phase d'injection.

Nous proposons d'ajouter la dimension temporelle avec un chronogramme du cycle d'injection, Figure 1.4. Ce dernier prend en compte les relations temporelles contiguës et parallèles des différentes phases. On distingue les phases principales d'injection, de maintien et de refroidissement. La phase de dosage est réalisée en temps masqué. Elle peut débuter pendant le refroidissement, dès que le canal d'alimentation est solidifié. Elle se termine avant le début de la phase d'injection. THYREGOD définit le temps de refroidissement comme facteur principal de la qualité et de la durée du temps de cycle, pour une production multi-empreintes [Thy01]. Il propose ainsi d'optimiser le ratio entre durée de la phase de refroidissement et profit économique réalisé par heure. Ce critère permet d'optimiser qualité et productivité. L'objectif est de réduire la durée des phases du cycle. La durée incompressible de la phase de dosage, réalisée en temps masqué, est la limite à cette optimisation. Les systèmes de canaux chauds (*hot runner*) réduisent les durées d'injection et de refroidissement : les canaux de distribution de la matière sont maintenus à température par des éléments chauffants de sorte que la matière reste à l'état fondu dans les canaux pendant l'ensemble du cycle. Enfin, des moules à refroidissement actif, capable de dissiper rapidement la chaleur, réduisent la durée du refroidissement [KH99].

Nous proposons de découper le procédé en cinq phases successives. Nous proposons une représentation Zig Zag, Figure 1.5, qui est complémentaire du schéma-bloc précédent. La

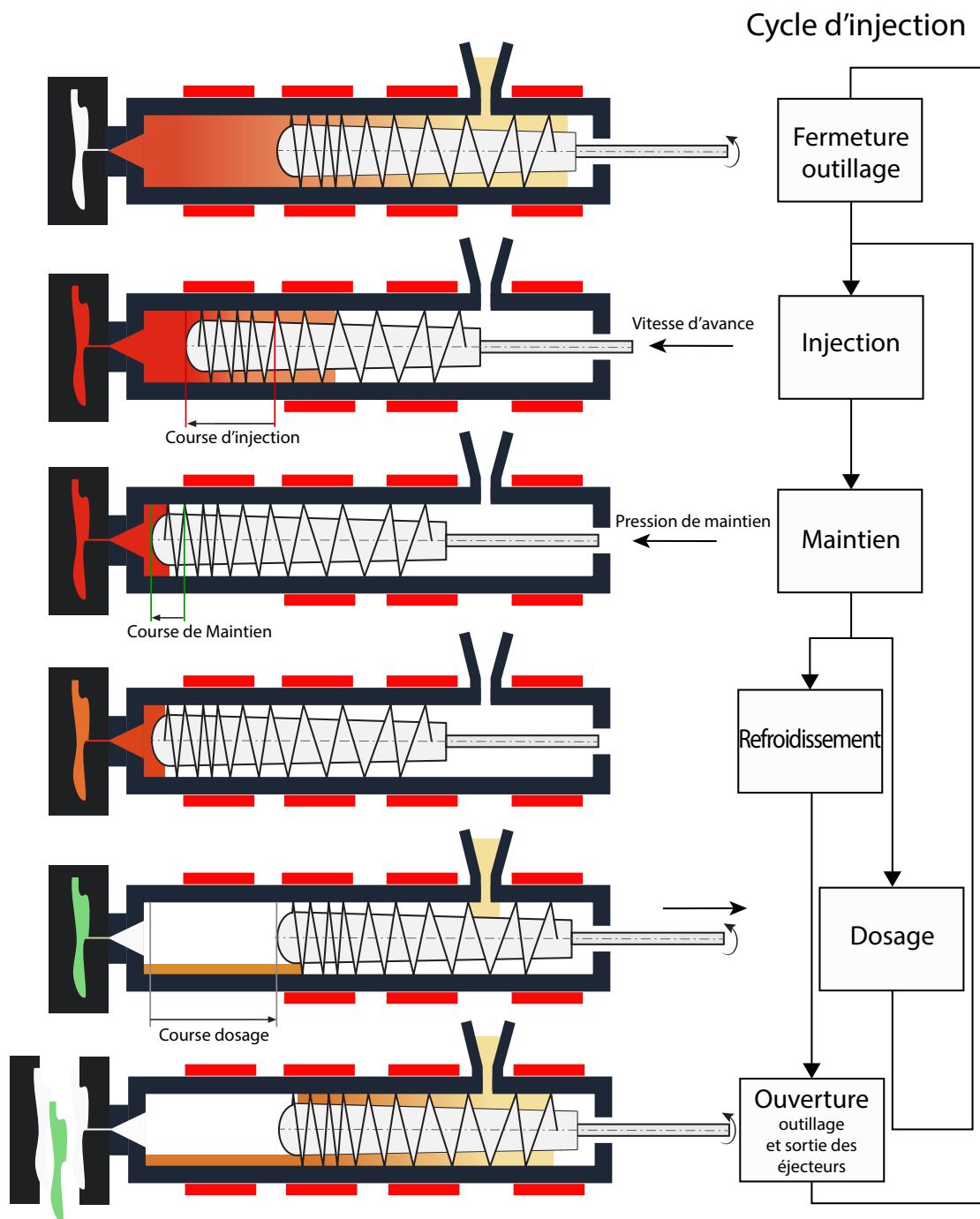


Figure 1.2 – Schéma-bloc du procédé d'injection-moulage.

## 1.1. Maîtrise du procédé d'injection-moulage des thermoplastiques

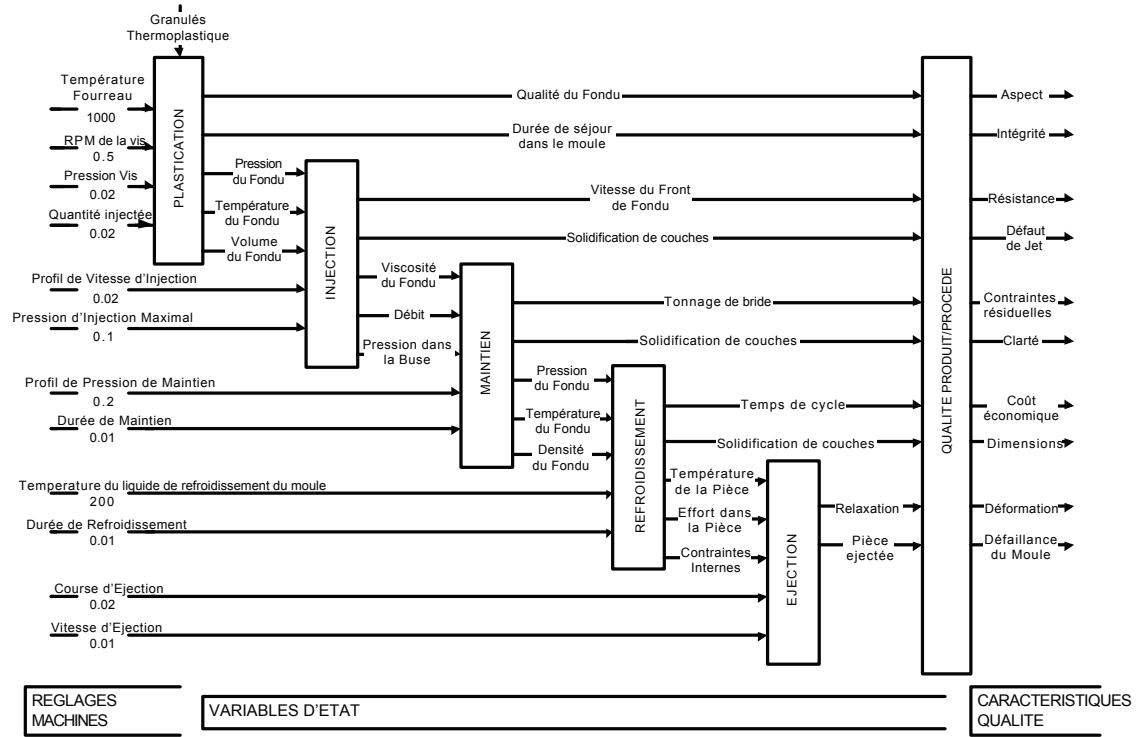


Figure 1.3 – Vue systémique du procédé d'injection-moulage, adaptée de KAZMER et HATCH [KH99].

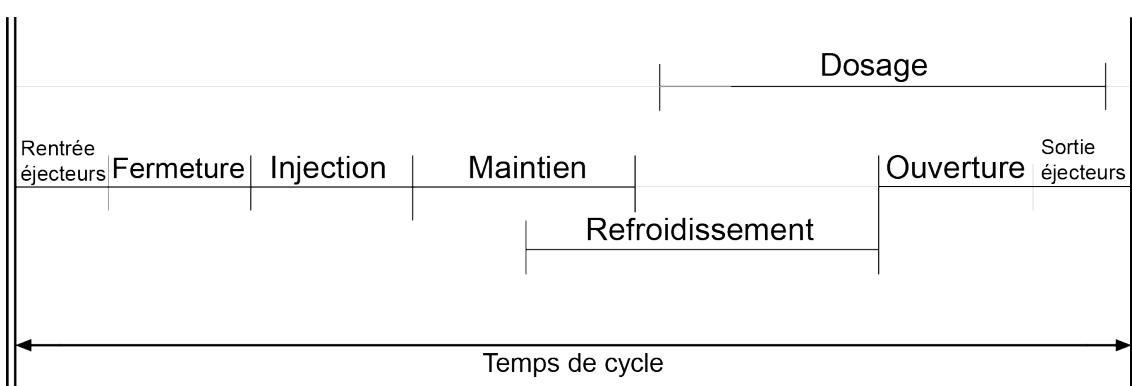


Figure 1.4 – Chronogramme du cycle d'injection-moulage.

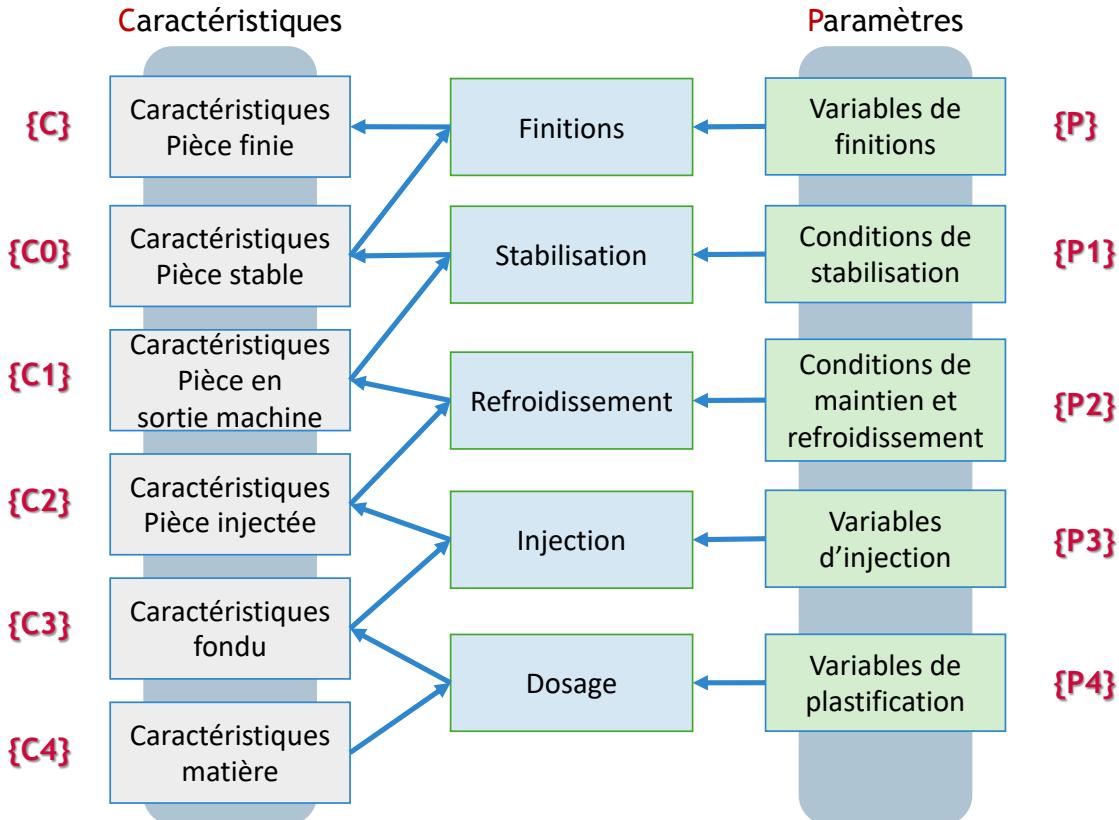


Figure 1.5 – Représentation Zig Zag du procédé d'injection-moulage.

représentation Zig Zag est proposée par SUH dans sa méthodologie *Axiomatic Design* [Suh90]. À la différence de la vue systémique de la Figure 1.3, la représentation Zig Zag fait apparaître les enjeux du procédé : les variables du procédé  $P_i$  et les caractéristiques du produit lors de chacune des phases  $C_i$ . Les caractéristiques du produit fini  $C_i$  sont fonctions de l'ensemble des phases. De plus, chaque phase est fonction des phases précédentes et des variables du procédé. Si un réglage est modifié au niveau de l'une des étapes, les caractéristiques du procédé pour toutes les étapes suivantes seront modifiées. Dans la suite de notre présentation des méthodes de pilotage, nous indiquerons à quelle étape de notre représentation correspondent les études citées.

### 1.1.2.2 Revue bibliographique de la modélisation du procédé d'injection-moulage

En 1987, dans un État de l'Art du pilotage des procédés d'injection-moulage, AGRAWAL, PANDELIDIS et PECHT estiment que de futurs travaux devront réaliser un modèle multivarié du procédé d'injection-moulage, afin de prendre en compte l'ensemble des variables [APP87]. Pourtant, aucun des modèles proposés dans la littérature ne prend en compte l'ensemble des variables de ce procédé. La nature séquentielle et cyclique du procédé rend difficile la réalisation de cette démarche.

#### Modélisations théoriques du procédé d'injection-moulage

Ces modèles doivent prendre en compte des interactions multi-physiques. La connaissance de la physique des polymères reste un champs de recherche ouvert. Le lecteur intéressé par les avancées de la modélisation cristallographique et micro-mécaniques des polymères, en injection-moulage, pourra se référer à la monographie de GALESKI et REGNIER [GR09]. Les auteurs discutent des contraintes thermo-mécaniques qui s'exercent sur la matière pendant un cycle de production. Ils analysent les conséquences de ces contraintes sur les propriétés cristallographiques et sur les

propriétés mécaniques des pièces.

En 1978, SHANKAR a le premier développé un modèle non linéaire en temps discret, pour optimiser la vitesse d'avance de la vis et le débit d'injection, dans le but d'obtenir une pièce aux caractéristiques définies [Sha78]. À la suite de ce travail, SHANKAR et PAUL ont proposé un modèle similaire de l'ensemble du procédé [SP82]. Le modèle est validé par une comparaison entre des simulations et des essais expérimentaux. CHIU, SHIH et WEI analysent les profils de pression dans le moule, ce qui leur permet de proposer un modèle dynamique du procédé [CSW91]. Huit variables d'état sont utilisées : délai avant le début du dosage, pression d'arrivée de la matière pour le dosage ( $C_3, P_4$  §1.5), pression d'injection, position de la vis, pression de la vis, volume du polymère dans l'empreinte et débit du polymère ( $C_2, P_3$ ). En 2000, DELAUNAY et al. montrent l'influence de l'évolution du contact entre la matière injectée et le moule pendant le cycle d'injection [Del+00]. Les précédents travaux assument que le contact entre le moule et la matière est parfait et que les températures à l'interface des deux phases sont identiques. Or une différence de plus de 20°C est cependant mesurée ( $C_2, P_3$ ). À partir d'une mesure invasive par un capteur de température dans l'outillage, ce travail identifie un modèle qui prend en compte l'évolution du contact d'un point de vue thermique. Les auteurs concluent sur la difficulté d'identification de leur modèle, lorsque d'autres paramètres du procédé sont modifiés. Les limites de la modélisation mathématique du procédé d'injection-moulage sont les hypothèses fortes qui sont faites en amont. BÉREAUX et al. proposent un modèle de l'étape de plastification ( $P_4$  §1.5) [Bér+04]. À partir des lois d'écoulement non-newtoniennes et de transformation de la matière, MOGUEDET, NAMY et BÉREAUX proposent un modèle par simulation en éléments finis [MNB09]. Ces simulations sont coûteuses en ressources de calculs et en temps. Leurs utilisations pour ajuster les réglages nécessite que le résultat soit obtenu pendant la durée du temps de cycle du procédé. Malgré les puissances de calculs aujourd'hui disponibles, cette durée est toujours incompatible. Les simulations éléments-finis sont en revanche indispensables pour la conception des pièces et des outillages, car elles permettent le dimensionnement des structures. Elles permettent également d'optimiser les conceptions [GTW08].

Dans la suite de nos travaux, nous ne nous appuierons pas sur la connaissance des modèles physiques. Nous utiliserons une démarche empirique pour modéliser le procédé. Nous nous intéresserons en particulier à la modélisation des caractéristiques des pièces produites, pour des paramètres du procédé données.

### Modélisation empirique par plan d'expérience

Les plans d'expériences permettent de construire un modèle polynomiale approximé d'un système à plusieurs paramètres. Ils permettent d'adapter la quantité d'essais expérimentaux à réaliser en fonction de l'ordre du modèle que l'on cherche à obtenir. Les essais sont réalisés suivant une table de variation des paramètres. Nous discutons des plans d'expériences dans le Chapitre 3 §3.4.2. Le modèle obtenu permet d'optimiser une sortie en fonction des paramètres, et ainsi de trouver un point de fonctionnement idéal. En 1994, BLYSKAL et MEHERAN utilisent les plans d'expériences afin de déterminer les paramètres optimaux pour produire des dimensions de pièces cibles [BM94]. De nombreux travaux utiliseront par la suite cette démarche. En 2013, FEI, MEHAT et KAMARUDDIN réalisent une étude rétrospective de l'utilisation des méthodes de Taguchi en injection-moulage des thermoplastiques [FMK13]. Ils distinguent deux utilisations faites dans la littérature : produire un jeu de données de mesures, pour ensuite construire un modèle par réseaux de neurones (nous réaliserons cette démarche dans le Chapitre 3 §§??); optimiser la conception d'une pièce et de son outillage en simulation numérique.

### Modélisation empirique par réseaux de neurones

L'utilisation des réseaux de neurones dans les systèmes industriels est proposée dès 1980. Le procédé d'injection plastique possède des paramètres interdépendants et non-linéaires. Par leur construction, les réseaux de neurones sont des fonctions non-linéaires adaptées à ces problèmes

aux multiples entrées. Nous détaillons les algorithmes à réseaux de neurones dans le Chapitre 3 §3.3.1.3. En 2000, SCHNERR-HÄSELBARTH et MICHAELI proposent le système *Intelligent Quality Control* [SM00]. Un algorithme à réseaux de neurones est utilisé pour prédire la qualité des pièces produites à partir des variables du procédé. Les données d'apprentissage proviennent d'un plan factoriel à trois niveaux, sur trois paramètres du procédé : température du fondu, pression de maintien et vitesse d'injection ( $P_4, P_3, P_2$  §1.5). 150 points de mesures de la pression dans le moule ( $C_2, C_0$  §1.5) sont enregistrés pendant les phases d'injection et de maintien. La masse de la pièce est mesurée avec une précision de 1 milligramme et la plage d'essai couvre une variation de masse de 1,4%. Le réseau est alors capable de prédire la masse des pièces avec une exactitude de 86 à 95,2%, soit une moyenne de 7 milligramme d'erreur sur la masse. Ces résultats mettent en valeur l'intérêt des modèles qui utilisent des réseaux de neurones. Enfin, dans sa thèse, CHEN et al. proposent une méthodologie pour établir le profil de vitesse d'injection ( $P_3$  §1.5) [Che+02]. Il modélise l'avancée du front du fondu à l'aide de réseaux de neurones. Il conclut sur la nécessité de garantir une vitesse de front constante lors de l'injection, afin d'obtenir une qualité produit optimale. Il modélise également la température du fondu par réseau de neurones, et obtient un profil de température à imposer pour la plastification ( $P_2$  §1.5).

### 1.1.3 Vers le pilotage du point de fonctionnement : Étude bibliographique

Nous réalisons une synthèse de notre étude bibliographique “Pilotage En Injection Plastique - État de l'Art” [NPP17a]. L'historique des avancées réalisées pour la maîtrise du procédé d'injection-moulage montre une évolution vers le pilotage cycle après cycle des paramètres du procédé. L'objectif du pilotage est d'optimiser le procédé afin d'obtenir la meilleure qualité possible pour le produit. Réaliser le pilotage du point de fonctionnement nécessite la maîtrise de l'ensemble

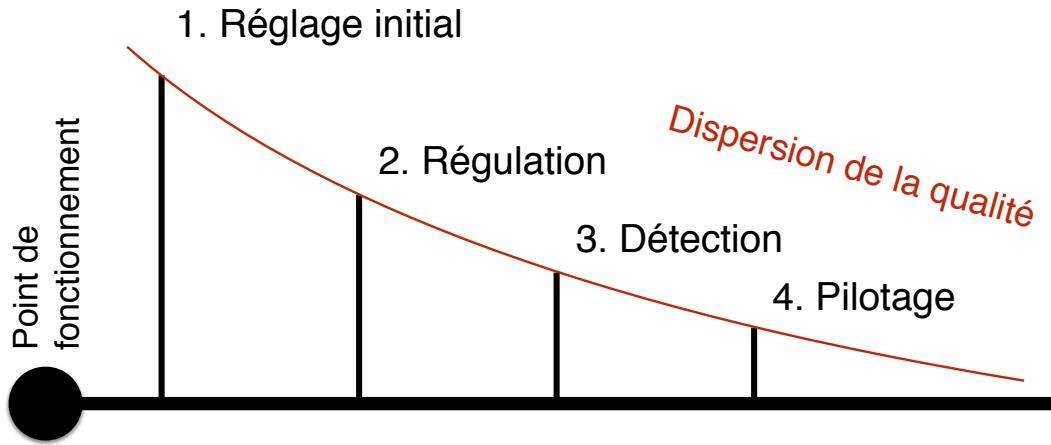


Figure 1.6 – Vers le pilotage du point de fonctionnement du procédé pour optimiser la qualité du produit.

des étapes précédentes : le réglage initial d'un point de fonctionnement, la régulation de ce point de fonctionnement et la détection de situations hors-contrôles. La Figure 1.6 récapitule cette démarche.

#### 1.1.3.1 Réglage initial du point de fonctionnement

La pratique industrielle du procédé d'injection-moulage fait appel au savoir-faire du technicien pour régler le point de fonctionnement initial. Un point de fonctionnement est associé aux caractéristiques du produit. De plus, le procédé d'injection-moulage n'est pas bijectif : il existe plusieurs points de fonctionnement pour des caractéristiques de produit identiques. L'approche classique de réglage employée par le technicien régulateur est l'essai-erreur. Aussi, lors du démarrage

du procédé, les premiers cycles de la production sont dédiés au réglage initial. Des échantillons sont pris quelques secondes après leurs sorties du moule et leurs qualités sont rapidement mesurées par observation visuelle. Dans un premier temps, il s'agit de contrôler l'absence de manque de matière (trous dans la pièce), puis les déformations géométriques importantes et enfin les traces de brûlures et givrages. L'opérateur utilise ensuite ses connaissances pour améliorer les réglages et respecter le cahier des charges. Sur les pièces les plus techniques, des mesures géométriques manuelles sont effectuées à l'aide d'un comparateur.

RICHARD et al. analysent les stratégies employées par des techniciens régulateurs [Ric+09]. Les observations ont été recueillies sur un intervalle de dix années par PASTRÉ et SAMURÇAY [Pas94 ; PS04] auprès de 13 régulateurs d'une entreprise de fabrication de produits, dont la qualité de finition est élevée [Pas94 ; PS04]. Leurs niveaux de compétences et leurs parcours professionnels sont variés. Certains ont une formation technique, mais beaucoup ont appris par la pratique sans formation initiale. Les résultats de cette étude montrent que la cause principale des différences de réaction entre régulateurs —la dispersion de réglages— est liée à la difficulté de lecture de la courbe de la pression hydraulique mesurée par la machine. La forme de cette courbe est utilisée par les techniciens régulateurs pour déduire l'état du procédé. Les causes de défauts sont connues de l'ensemble des régulateurs. Les systèmes d'assistance au réglage proposent d'ajouter des indicateurs de l'état du procédé, ce qui permet d'éviter les difficultés de lecture de la courbe de pression.

### Apport des systèmes experts

Les systèmes experts permettent de choisir des couples paramètres-procédés fonctionnels, afin de produire des pièces plus rapidement, lors d'une mise en place d'une production. Ces systèmes s'appuient sur une base de connaissances : un ensemble de relations causales pondérées, qui a été construit à partir d'une multitude d'essais et qui modélise également des règles empiriques. Ces systèmes sont complétés et sont validés par des simulations numériques [Jan92]. En 1993, KAMEOKA, HARAMOTO et SAKAI vérifient que le système expert est capable de proposer des réglages identiques à ceux d'un technicien-régulateur de niveau moyen [KHS93]. Des systèmes hybrides peuvent être constitués de règles empiriques et de cas d'applications spécifiques [SS97]. Ces systèmes simulent une démarche de réglage humaine basée sur la connaissance de relations linéaires simples entre les paramètres du procédé. BOZDANA et EYERCIOĞLU proposent une base de connaissances de 623 presses à injecter et 27 matériaux différents [BE02]. Ce système expert permet de sélectionner le couple machine-matière le plus approprié, pour une pièce donnée.

#### 1.1.3.2 Régulation du point de fonctionnement

Diminuer la dispersion d'un procédé de production permet de limiter les rebuts. Pour cela, il est nécessaire de maîtriser le procédé. Il est possible d'asservir les paramètres réglables pour garantir que le procédé reste sur le point de fonctionnement défini. Cependant, asservir un procédé non linéaire aux multiples commandes et sorties est compliqué. La littérature de la théorie du contrôle automatique propose de nombreux travaux appliqués aux productions industrielles. Les recherches aboutissent souvent à des solutions commerciales. Dans ce domaine d'étude, nous notons le dynamisme du secteur de la chimie de synthèse. La régulation du procédé s'appuie sur l'analyse de la transformation de la matière, qui change d'état au cours du cycle. Les contraintes qui lui sont appliquées pendant le cycle conditionnent son état final. Le produit final correspond à la pièce sortie de la machine, après une durée de stabilisation, durant laquelle des modifications géométriques peuvent se produire (par exemple des déformations liées aux contraintes internes qui se relâchent). Les premiers développements ont porté sur la régulation de paramètre unique.

### Utilisation de l'Automatique

Dès 1975, un asservissement est breveté par LACZKO, sous la forme d'une consigne sur la viscosité de la matière [Lac75]. La pression mesurée dans l'outillage permet de réguler la pression

hydraulique qui est appliquée pendant le cycle d'injection, afin de garantir une consigne sur la viscosité de la matière. Cette asservissement est hydraulique et ne nécessite pas de micro-contrôleur.

Dans la littérature, la caractéristique du produit la plus étudiée est la masse de la pièce ( $C_0$  §1.5). Le paramètre du procédé qui est le plus étudié pour la régulation est la pression dans l'outillage [AKP85; Kam+87] ( $P_3$  §1.5). Sur ce sujet, ABU FARA évalue l'utilisation des régulateurs PI et PID (*Proportionnel Intégrateur Dérivateur*) [Abu88]. Dans son travail de doctorat, le système hydraulique du vérin d'injection de la presse est modifié pour inclure deux servovalves, afin de piloter la pression d'injection et d'asservir la pression mesurée directement dans le moule. Les valeurs des correcteurs Proportionnel, Intégrateur et Dérivateur sont choisies à partir de simulations. Les expérimentations montrent que la valeur de la pression dans l'outillage a un comportement non linéaire. La régulation est effectuée à l'aide d'un gain échelonné sur la servovalve, selon la valeur de la pression. Ce procédé de régulation provoque néanmoins des oscillations de pression à la fin de la phase d'injection, car la commande de la valve est binaire.

En 1987, AGRAWAL, PANDELIDIS et PECHT réalisent un état de l'art du contrôle du procédé d'injection-moulage, dans lequel il indique que les régulateurs PI (*Proportionnel Intégrateur*) et PID (*Proportionnel Intégrateur Dérivateur*) sont difficilement utilisables, car il n'existe pas d'état d'équilibre pour l'ensemble de la plage d'états du procédé [APP87]. Il conclut sur l'intérêt des contrôleurs autorégulateurs, contrôleurs optimaux et contrôleurs prédictifs. Ces contrôleurs ne réagissent qu'à une unique variable, mais le procédé à des interactions multivariées. Les contrôleurs uni-variés entraînent un risque pour la robustesse de la régulation, car ils ne prennent pas en compte les interactions entre les paramètres. C'est pourquoi, ils peuvent avoir l'effet opposé de celui souhaité. Le risque est de dérégler un procédé qui est stabilisé. Le contrôleur uni-varié peut appliquer une correction idéale pour variable, mais cette correction sera erronée pour les autres variables, d'où l'effet opposé qui dérèglera le procédé. Enfin, AGRAWAL, PANDELIDIS et PECHT argumentent la proposition que les caractéristiques de la matière fondue doivent être mesurées en direct, plutôt que de mesurer des variables de la machine ; ceci permet de limiter les erreurs dû aux mesures indirectes. C'est le début de la mise en place de capteurs directement au contact de la matière, dans l'outillage. Nous qualifions ces méthodes de mesure "invasives" pour le procédé. Elles nécessitent une intégration compliquée au sein de l'outillage.

Par la suite, ABU FARA, KAMAL et PATTERSON développent une régulation de la pression dans le moule qui est capable de rejeter des perturbations externes, tel que la variation des propriétés du polymère et de la matière fondue ( $P_1$  et  $P_2$  §1.5) [AKP90]. Il conclut sur la nécessité d'implémenter un contrôle multivarié et adaptatif. En 1995, KAZMER et BARKAN proposent le système *Dynamic Feed Control* afin d'équilibrer le débit d'injection de la matière pour les moules multi-empreintes ( $P_3$  §1.5) [KB95]. Plusieurs valves permettent de réguler la pression dans chacune des empreintes, pendant la durée du cycle. Les résultats obtenus montrent une réduction de 75% de la variabilité de la masse des pièces produites. La capacité long terme (Cpp) du procédé augmente de 0,52 à 1,67. Ce système demande un développement spécifique de l'outillage du moule.

Des travaux s'intéressent à la régulation de la température de la matière fondue ( $P_4$  §1.5), par KAMAL, PATTERSON et GOMES [KPG86; GPK86], puis par GUSTAFSON et LEBOW [GL87]. La vitesse d'avance de la vis lors de l'injection ( $P_3$  §1.5) est également régulée par PANDELIDIS et AGRAWAL [PA88]. Une boucle de régulation Linéaire Quadratique (commande  $LQ$ ) est conçue et ses paramètres sont identifiés. Les résultats expérimentaux montrent de meilleures performances qu'avec une régulation PID. Le temps de réponse aux perturbations est notamment plus rapide.

### Apport de la modélisation par réseaux de neurones

Plusieurs travaux proposent de modéliser une partie du procédé d'injection afin de réguler certains paramètres du procédé. Nous détaillerons la théorie des réseaux de neurones dans le Chapitre 3 §3.3.1.3.

DEMIRCI, COULTER et GÜÇERI régule la vitesse d'avancée du front de la matière fondu à partir des mesures de la pression d'injection ( $P_3$  §1.5) [DCG97]. Dans le même temps, WOLL et COOPER [WC97] propose de réguler deux variables du procédé : la pression d'injection et la température du fourreau ( $P_3, P_2$  §1.5), afin d'obtenir un profil d'évolution de la pression dans le moule déterminé. Un réseau de 36 neurones d'entrées, 3 neurones intermédiaires et 2 neurones de sorties est retenu. Le modèle prend en entrée un échantillonnage de 36 valeurs discrètes du profil de pression et calcule en sortie les dimensions de la pièce qui sera produite. Le réseau de neurone modélise cette relation à partir d'un jeu de données de 81 profils de pressions. Ces profils ne sont pas obtenus par expérimentation, mais ils sont simulés par un modèle physique simplifié du procédé. Le réseau de neurones approxime la simulation qui couteuse en ressource de calcul. Le réseau propose une réponse sur deux paramètres à partir de 36 variables d'entrées. Les dimensions des échantillons obtenues expérimentalement avec cette régulation sont mesurées à l'aide un pied à coulisse digital (précision de 10 micromètres), plusieurs jours après la production. L'utilisation du réseau permet de régler et réguler le procédé afin d'obtenir une longueur d'échantillon voulue. Les auteurs concluent sur la nécessité d'utiliser un modèle multivarié non linéaire pour modéliser le procédé d'injection. Les réseaux de neurones sont particulièrement adaptés à cette tâche.

MICHAELI et SCHREIBER régulent la pression dans l'outillage à partir de la consigne de pression du vérin hydraulique d'injection [MS09]. Un modèle par réseau de neurones dispose d'une base d'apprentissage de 15 cycles d'injection-moulage avec différentes configurations de réglages. La température de la matière dans le moule est mesurée par capteur infrarouge ( $C_2$  Figure 6). La pression d'injection est régulée ( $P_3$  §1.5) afin de suivre le diagramme Pression-Volume-Température, de transformation du matériau pendant l'ensemble du cycle. Avec la mise en place de cette régulation, les résultats expérimentaux montrent qu'une augmentation de 20°C de la température de la matière fondu ne cause qu'une augmentation de 0,07% de la masse de la pièce ; ce qui est négligeable. En comparaison, la même augmentation, sans la régulation, entraîne une diminution de 1.27% qui n'était pas acceptable.

### Apport du contrôle adaptatif

Le contrôle adaptatif<sup>1</sup> est expérimenté en injection-moulage dès 1983 par SANSCHAGRIN [San83]. 10 paramètres sont régulés ( $P_4, P_3, P_2$  §1.5) et 3 caractéristiques sont mesurées : masse de la pièce ( $C_0$ ), pression maximale dans la cavité ( $C_2$ ), écartement maximal du moule ( $C_2$ ). L'algorithme d'identification utilise la méthode des moindres carrés généralisés<sup>2</sup>. La température du fourreau ( $P_4$  §1.5) n'a pas été prise en compte dans cette étude. L'auteur justifie cette absence car une modification de la consigne de la température du fourreau est effective après un retard de dix cycles d'injection, à cause de l'inertie thermique de l'outillage. Trois séries d'essais sont réalisées. Chaque série d'essais ne s'intéresse qu'à quatre paramètres sur les dix possibles. Cette démarche peut créer des corrélations statistiques biaisées. Aussi, il est préférable de réaliser l'analyse de l'ensemble des facteurs lors d'un même essai. L'essai est réalisé pour des pièces à paroi fine (1 millimètre) et des pièces à paroi épaisse (4 millimètres). Les variations obtenues montrent qu'il est plus difficile d'obtenir une répétabilité de la masse sur les pièces épaisses. Le paramètre le plus influent sur les pièces est la pression de maintien. À la suite des essais, les trois paramètres les plus influents ont été retenus pour implémenter une régulation en boucle fermée. Suite à la régulation de ces trois paramètres, la masse des pièces varie de 0,4% à 3,6% et de 0,1 à 0,3% sans la régulation ; soit une diminution de la variation proche d'un facteur dix.

Dans ces travaux de doctorat, DEVOS réalise une étude exhaustive des paramètres du procédé qui sont influent sur la qualité géométrique des pièces [Dev90]. Il utilise un correcteur PID pour réguler la pression du vérin d'injection, à partir de la pression mesurée à l'intérieur de l'outillage.

1. Le contrôle adaptatif est développé dès 1979 par LANDAU ; EGARDT [Lan79 ; Ega79]

2. L'algorithme des moindres carrés généralisés qui est utilisé dans le travail SANSCHAGRIN a été initialement implanté en 1976 par BETHOUX pour l'étude de la fabrication de papiers et de la distillation [Bet76]

Avec ce système, la variabilité géométrique et la variabilité massique sont diminuées de moitié. L'étude conclue que la mesure de pression maximale dans l'outillage (**C2 §1.5**) est une mesure qui peut être choisie comme variable de contrôle du passage de la phase d'injection, à la phase de maintien : c'est le point de "commutation". Cette démarche est aujourd'hui implémentée chez la plupart des fabricants de systèmes de régulation pour l'injection-moulage.

Dans ses travaux de doctorat, TSOI compare différents contrôleurs pour réguler la vitesse de l'avance de la vis, à partir de la mesure de la pression dans l'outillage [Tso97]. Il propose l'utilisation de contrôleurs à logique floue, afin de suivre un profil de vitesse. La logique floue permet de modéliser sous forme continue des relations empiriques, définies à la manière des systèmes experts. HUANG et LEE [HL00] contrôlent la vitesse d'avance de la vis à l'aide d'un contrôleur à logique floue, afin de réguler la pression qui est mesurée dans la buse, pendant les phases d'injection et de maintien (**C3** et **C2 §1.5**). Les expériences sont compilées dans des graphiques de réponses à une commande de pression définie. Il n'y a pas d'évaluation numérique. Les graphiques montrent la réponse satisfaisante du contrôleur à la suite d'un changement d'outillage et à la suite d'un changement de température du fourreau.

La régulation du procédé d'injection est une thématique d'étude riche. L'utilisation d'une grande variété de contrôleurs uni-varié ou bi-varié a été proposée. Cependant, la régulation de l'ensemble des variables du procédé n'a jamais été proposée. Nous posons l'hypothèse que la cause est dû à la complexité de la modélisation du système complet.

### 1.1.3.3 Détection de situations hors-contrôles

Les variations des caractéristiques produit observées sur des produits peuvent avoir différentes origines. Selon la terminologie de SHEWHART définit dans "Economic Quality Control of Manufactured Product" [She30], l'origine des variations se classent en deux catégories : les causes communes et les causes spéciales. Les causes communes proviennent de nombreuses sources de perturbations aléatoires. Elles sont inhérentes au procédé. Elles entraînent une dispersion selon une loi Normale des caractéristiques des pièces produites. En revanche, les causes spéciales entraînent des écarts de caractéristiques qui dépassent la dispersion aléatoire. Les causes spéciales doivent être détectées, puis leurs origines devront être identifiées. Dès 1997, SHERBELIS et al. proposent de définir une fenêtre acceptable dans laquelle les caractéristiques du procédé doivent se situer [She+97]. Cela permet de prendre en compte la variabilité naturelle du procédé.

En 1991, la *Society of the Plastics Industry* [Ber91] définit la tolérance dimensionnelle acceptable, en injection-moulage, dans un intervalle de 0,2% à 0,4% des dimensions attendues. En 2010, KAZMER et al. étudie différents procédés de régulation [Kaz+10]. Il observe, quel que soit la méthode utilisée, une dispersion des dimensions inférieure à 0,3%. Il conclut que les tolérances précédemment proposées peuvent être dépassée avec des machines modernes et des stratégies de régulations avancées. Les tolérances attendues en injection-moulage sont définies par l'Association Française de NORmalisation (AFNOR) [Nor87]. Les tolérances dimensionnelles sont spécifier en pourcentage d'écart. De plus, elles diffèrent en fonction du matériau plastique employé. La récente norme ISO20457 ([ISO18]) reprend en grande partie cette norme établie par l'AFNOR. Elle spécifie également la méthode de calcul à employer pour le défaut géométrique de diminution des dimensions après le refroidissement (rétractation, retrait ou *shrinkage*). La norme définit également les causes externes possibles à la variation des dimensions (météorologie, contraintes mécaniques extérieures), ainsi que les causes liées au procédé (mauvais calcul du retrait, déformation de l'outillage, usure de l'outillage). Les conditions sont également spécifiées : mesures à réaliser entre 16 et 72 heures après la production, une fois que les pièces ont été stockées à  $23^{\circ}\text{C} \pm 2^{\circ}\text{C}$  et  $50\% \pm 10\%$  d'hygrométrie. La normalisation des tolérances permet de connaître l'amplitude à surveiller. Dans le cadre du FUI SAPRISTI, nous travaillons en particulier sur le contrôle de défauts d'aspect, dont la cause est géométrique (défaut de retassure). L'ordre de grandeur de ces défauts est de l'ordre de la dizaine de micromètres, pour des pièces de plusieurs

mètres. La tolérance requise pour limiter ces défauts est trois ordres de grandeur plus grande que les tolérances requises dans les normes.

### Analyse en Composante Principale

Une presse instrumentée moderne fournit une quantité d'informations conséquente sur ses paramètres propres, de même que sur l'état de la matière. Ces données sont produites pour chaque cycle. L'Analyse en Composantes Principales §3.3.4.3 permet de réduire la dimension des données. Cette méthode est employée par plusieurs travaux en injection-moulage des thermoplastiques. LU et al. [Lu+04] réalisent une ACP sur des mesures réalisés pendant 60 cycles d'injection, pour 16 variables du procédés ( $C_0, C_1, C_2$  §1.5). Le regroupement des composantes fait apparaître quatre phases distinctes, dont deux phases transitoires. Cette division correspond bien aux phases du procédé définies dans l'industrie : injection, maintien, refroidissement, plastification. Une analyse détaillée, phases par phases, montre que les températures du fourreau ne sont pas corrélées avec les autres variables du procédé. De plus, les variables dominantes, pour chacune des phases, ne sont pas toujours les mêmes. Enfin, des variables qui ne sont pas importantes sont identifiées. Trois défaillances sont alors introduits en production : une variation de la composition de la matière, une défaillance du capteur de régulation de la température du fourreau et une défaillance de la valve anti-retour de la buse d'injection. Ces trois défaillances sont très communs en injection-moulage. Les trois défaillances sont détectées en temps réel, par un test  $T^2$  de HOTELLING [Hot33] sur les composantes de l'ACP.

### Maîtrise Statistique des Procédés

Les travaux fondateurs de la Maîtrise Statistique des Procédés (MSP), *Statistical Process Control* sont réalisés dès 1930 [She30 ; She31]. En 2003, PILLET, MAIRE et BRONET proposent des principes pour adapter la MSP au procédé d'injection-moulage [PMB03]. Le choix des variables à contrôler peut se faire sur les paramètres du procédé et sur les caractéristiques du produit. Il est recommandé de choisir peu de variables. Celles-ci doivent être les plus caractéristiques possibles des défauts déterminé par le cahier des charges. Enfin, il est recommandé de choisir des variables faciles à mesurer. Le procédé d'injection-moulage est capable de produire une pièce toutes les 30 secondes, c'est pourquoi le risque de faux positif est ajusté. Ce risque est généralement défini à plus ou moins trois fois l'écart-type de la distribution de la valeur surveillée. Dans le cas de l'injection, cela correspondrait à une fausse alarme toutes les 2,5 heures, soit 0,27% de pièces hors-contrôles. Cette fréquence est trop élevée, c'est pourquoi les auteurs recommandent d'ajuster le risque à 4,5 fois l'écart-type, soit 0,00068% de pièces hors-contrôles, soit une pièce toutes les 992 heures de production. Cette démarche diminue drastiquement l'intervalle de tolérance du procédé. En cas de détection de situation hors-contrôle, il est recommandé d'éjecter les pièces concernées. En cas de répétition de cette situation hors-contrôle sur plusieurs pièces, il est recommandé d'arrêter la machine et de faire intervenir le technicien-régulateur. En 2008, KAZMER, WESTERDALE et HAZEN étudie l'utilisation de la démarche MSP [KWH08]. Il évalue séparément la mise sous contrôle statistique des variables de la machine ( $P_1, P_2, P_3$  §1.5) et la mise sous contrôle statistique des variables mesurées dans l'outillage ( $C_1$  et  $C_2$ ). Les paramètres du procédé sont perturbés suivant un plan d'expériences §3.4.2. Ce travail montre la pertinence de cette méthode. Néanmoins, les recommandations de PILLET, MAIRE et BRONET [PMB03] ne sont pas prises en compte : la mise sous contrôle n'est pas réalisé sur les variables les plus pertinentes. En 2016, LIU, LIU et ZHANG évaluent un algorithme à fenêtre mobile, afin d'identifier les différentes phases du procédé avant d'appliquer le contrôle MSP [LLZ16]. ZHANG et al. [Zha+16] propose de limiter la surveillance aux seuls paramètres du procédé ( $P_2$  §1.5), afin d'éviter d'instrumenter l'outillage. Il se contraint à utiliser les mesures systèmes de régulation des machines : la pression hydraulique et la position de la vis dans le fourreau. L'objectif est de prédire les défauts sur la masse des pièces ( $C_3$  §1.5). La méthode proposée atteint 91,48% de détections de défauts de masse des pièces. Cette méthode obtient de meilleures performances, en comparaison avec une Analyse en Composante Principale.

### 1.1.3.4 Pilotage des caractéristiques de la pièce produite

Le pilotage d'un procédé demande dans un premier temps de détecter les situations hors-contrôles. Dans le cas où le procédé n'est pas dans sa configuration cible, on mesure l'écart entre la situation hors-contrôle et la situation cible. Puis on ajuste un ou plusieurs paramètres réglables du procédé, afin de se rapprocher de la situation cible. Plusieurs stratégies de pilotage ont été proposées dans la littérature. Certaines utilisent une modélisation physique simplifiée du procédé, d'autres s'appuient sur une base de connaissance préétablie ou utilisent une analyse statistique. Ces systèmes reposent sur l'instrumentation de la presse et de l'outillage pour mesurer les conditions appliquées au polymère. L'évaluation des performances de ces stratégies passe par la mesure de dispersions des caractéristiques des produits. Dans une démarche pragmatique, le pilotage des caractéristiques du produit peut également se réaliser en mesurant directement les caractéristiques des produits ( $C_1$  et  $C_0$  §1.5). Dans le cadre de nos travaux, nous mesurerons la qualité des pièces juste après la sortie de l'outillage ( $C_1$ ).

Les essais de modélisation informatique ont émergé avec le développement des valves de décompression à commande électrique [Ma74]. En 2001, KAZMER et DANAIS synthétisent les avancées réalisées dans le cadre du pilotage du procédé [KD01]. Ils définissent deux contraintes principales au développement du pilotage :

- la méconnaissance des relations entre les paramètres d'entrées et les caractéristiques finales du produit ( $P_i$  et  $C_i$  §1.5),
- le manque de possibilité d'ajustement des réglages du procédé.

Ils argumentent que l'amélioration d'une caractéristique du produit se fera souvent au détriment d'une autre caractéristique, ou bien par l'augmentation des coûts de produit. Cette contrainte de coût est courante pour les productions industrielles à haute valeur ajoutée.

L'ajustement statistique des procédés (*Statistical Process Adjustment*) est un domaine situé à l'intersection de la régulation automatique et de la Maîtrise Statistique des Procédés. Ce domaine d'étude étudie les relations entre les paramètres du procédé et les caractéristiques du produit, par une approche statistique. Un exemple de relations entre les paramètres du procédé d'injection-moulage, qui conditionnent les caractéristiques de la pièce produite, concerne la vitesse d'avance ( $P_3$  §1.5). Celle-ci est liée à la viscosité de la matière ( $C_3$ ) et elle définit le débit d'injection ( $P_3$ ).

Il y a également une influence des paramètres du procédé sur la structure interne de la pièce. Les travaux de doctorat de GIROUD étudie l'influence des paramètres du procédé ( $P_3$ ) sur les contraintes internes ( $C_1$ ) [Gir01]. Le profil de pression appliqué à la pièce pendant la phase d'injection conditionne les contraintes internes résiduelles. Ces contraintes seront figées dans la pièce par la modification de la forme et de l'orientation des chaînes de polymères. La relaxation de ces contraintes peut provoquer des déformations géométriques. C'est une des causes principales du retrait, qui apparaît plusieurs heures après la production.

DEL CASTILLO réalise une rétrospective des travaux réalisés dans le domaine de l'Ajustement Statistique des Procédés [Del06]. Il identifie les méthodes de modélisation bayésiennes comme des pistes privilégiées de recherche, par exemple dans leurs applications en contrôle. Nous présenterons ces méthodes d'optimisations dans le Chapitre 3 §3.4.3. Il est parfois difficile de mesurer les pièces en sortie de machine. Les causes sont une durée de la mesure trop longue qui est supérieure au temps de cycle, un coût trop élevé ou simplement une mesure non faisable. De plus, le manque de spécifications concernant les caractéristiques de la qualité des pièces n'a pas encouragé l'utilisation de la mesure directe. Excepté pour le dimensionnel, la notion de qualité dans les cahiers des charges sont rarement normalisés. Elles nécessitent l'interprétation humaine, ce qui est difficile à transmettre à programmer dans une machine (c'est un des objectifs de ce travail, présenté dans le Chapitre 3.3). Pour remédier à ces limites, de nombreux systèmes ont choisi de prédire les caractéristiques des pièces produites ( $C_1$  §1.5) à partir de variables du procédé ( $P_i$ ). La prédiction obtenue est ensuite utilisée à des fins d'ajustement des réglages pour tendre vers un procédé optimal.

### Ajustement du procédé à partir des caractéristiques prédites

En 1993, HAEUSSLER et WORTBERG [HW93] utilisent un réseau de neurones pour prédire la masse et la longueur des pièces ( $C_1$  §1.5), à partir des paramètres du procédé ( $P_3, P_2$ ). Le réseau prédit la masse des pièces avec 80% de justesse et la longueur avec 88,2% de justesse. Ces résultats sont comparés avec une régression non-linéaire, qui atteint une justesse de 75% pour la masse et 79% pour la longueur. Cette étude conclut sur la nécessité de l'utilisation d'un modèle adaptatif pour un réaliser le pilotage en cycle industriel. Le modèle prédictif doit être enrichi à partir des mesures, au fur et à mesure des cycles. Le modèle ne doit pas se limiter à modéliser un jeu de données statique. Nous reprenons cette idée dans nos travaux, en proposant un modèle de la notion de qualité qui est enrichi au fur et à mesure de la production §3.3.

En 1996, WOLL, COOPER et SOUDER proposent d'utiliser un réseau de neurones pour prédire la masse des pièces, à partir des profils de pressions mesurées dans l'outillage et dans la buse [WCS96]. Ce modèle permet d'accepter ou de rebouter la pièce produite par l'analyse de l'écart statistique MSP §1.1.3.3. Le réseau de neurones est appris comme un modèle inverse du procédé. Les profils de pressions sont les variables d'entrées ( $C_2, C_1$  §1.5) et les variables de sorties sont les valeurs des consignes de pression de maintien ou de température du fourreau ( $P_2, P_1$ ). Les paramètres sont ajustés cycles après cycles en comparant le profil de pression courant, avec le profil cible. L'analyse MSP est réalisée sur la valeur de pression maximale qui est mesurée dans l'outillage. Le réseau est appris sur un jeu de données issues de simulations, afin d'éviter de réaliser une longue campagne d'essais. À contrario, dans nos travaux, nous avons choisi de mesurer des données réelles sur des machines industrielles. L'essai est réalisé pour un unique matériau et la caractéristique cible de la pièce est sa longueur.

En 2004, dans ces travaux de doctorat, YANG introduit l'utilisation du contrôle prédictif pour ajuster la vitesse d'avance de la vis lors de l'injection [Yan04]. Le contrôleur s'appuie sur la littérature du *Global Process Control* (aussi appelée *Engineering Process Control*) qui consiste à utiliser des régulateurs Proportionnels Intégrateur Dérivateurs pour asservir les caractéristiques du produit. La masse des pièces est prédite à partir des mesures sur le procédé ( $C_2, C_1$  §1.5). À partir de la prédiction, la vitesse d'injection ( $P_3$ ) est ajustée cycles après cycles. Ce travail montre la limite que pose l'utilisation des servovalves pour ajuster les paramètres du procédé. Les servovalves ont un temps de réponse lent, de l'ordre de la seconde. Ainsi, il est nécessaire d'anticiper leurs actionnements, notamment lorsqu'il s'agit d'imposer des consignes temporelles précises. À l'aide de l'ensemble des ces techniques, ce travail montre qu'il est possible de détecter les variations de la masse pendant le cycle, sans nécessiter une pesée de la pièce. Il est également possible d'ajuster les conditions de maintien et de refroidissement en conséquence, afin de compenser les variations prédites. L'auteur conclut sur la possibilité de piloter d'autres caractéristiques du produits, en utilisant la même démarche.

### Ajustement à partir des caractéristiques mesurées sur le produit

Dans l'ensemble de la littérature, la diversité des caractéristiques qualité mesurées sur les pièces est limitée. La masse est la mesure la plus mesurée, car sa mesure peut facilement être réalisée pendant le cycle d'injection-moulage. Il suffit de disposer d'un bras robotique préhenseur et d'une balance instrumentée. Les mesures complémentaires de la qualité des produits sont effectuées après de la production d'un lot, mais non pendant le cycle d'injection. Il serait intéressant de réaliser une caractérisation complète de la pièce produite, pendant le temps de cycle, afin de pouvoir optimiser les caractéristiques des pièces suivantes, à partir de ces mesures. La limite à cette démarche est la durée des mesures, qui doit être inférieure au temps de cycle.

En 1998, FOURNIER et al. proposent un ajustement des paramètres du procédé à partir de la mesure de la masse du produit [Fou+06]. Le matériau utilisé est le Poly(Téréphthalate de Butylène). C'est un semi-cristallin sensible aux phénomènes de retrait. Le système utilise un modèle physique simplifié du procédé pour réguler les paramètres en boucle fermé. L'écart entre la masse mesurée et la masse cible est ajouté à la boucle de régulation. Deux perturbations sont appliquées pendant

le cycle : un changement de la matière et un changement de la température dans le fourreau. Afin de limiter les contraintes internes, les pièces passent par un recuit. Les mesures des retraits volumiques et de résiliences (chocs Charpy) montrent une dispersion de masse réduite de 24% avec l'utilisation du correcteur sur la masse. La dispersion du retrait volumique, mesurée deux jours après la production des pièces, est réduite de moitié, mais l'épaisseur des échantillons augmente. Après recuit, le retrait volumique est diminué de 12% par rapport à un procédé non régulé. La résilience après recuit est quant à elle augmentée de 23%. L'étape de recuit peut être assimilé aux étapes de peintures qu'une pièce est susceptible de subir. Une piste de travail importante est de considérer les caractéristiques des produits finaux ( $C_0$  §1.5) pour ajuster le procédé. Cela nécessite nécessairement l'utilisation d'un modèle prédictif, car la pièce finale sera réalisée plusieurs jours après son moulage. Les résultats obtenus sont encourageants. L'étude montre qu'il y a un intérêt à ajuster les paramètres du procédé en fonction des caractéristiques du produit mesurées dès la sortie de la machine ( $P_1$ ). Nos travaux s'inscrivent dans cette démarche.

La même année, IVESTER, DANAI et KAZMER proposent une méthode d'ajustement qui s'appuie sur un modèle du procédé qui est construit au fur et à mesure de la production, à partir de la mesure des caractéristiques et des paramètres [IDK98]. Ce modèle permet de réaliser une optimisation des caractéristiques en ajustant les paramètres du procédé, par descente de gradient (cette méthode est détaillée dans le Chapitre 3 3.3.1.3). L'étude indique que la caractérisation de la qualité des pièces est réalisée par un opérateur humain. Dans le cadre de nos travaux, nous chercherons à automatiser ce contrôle. Cette méthode est néanmoins dépendante des points de fonctionnement initiaux, qui sont choisis aléatoirement. C'est pourquoi, YANG, DANAI et KAZMER propose l'ajustement des paramètres du procédé par optimisation d'un modèle qui lie les paramètres avec les caractéristiques du produit, préalablement établi [YDK00]. C'est ce modèle empirique multivarié qui détermine le point de fonctionnement initial. Par la suite, le modèle initial est ajusté au fur et à mesure de la production. Les caractéristiques prédictes par le modèle sont comparées aux mesures réelles, et le modèle est ajusté pour diminuer l'écart de la prédiction. Une comparaison expérimentale est réalisée entre le système proposé et un point de fonctionnement obtenue par résolution d'un plan d'expériences Central Composite (§3.4.2). 7 caractéristiques qualités sont mesurées sur le produit. 5 paramètres sont pilotés sur la machine. Le protocole expérimental est détaillé. Les 20 premières pièces de chaque essai ne sont pas prises en compte, afin de stabiliser le procédé. Par la suite, 5 pièces sont mesurées 15 à 25 minutes après leur sortie du moule pour que leurs caractéristiques géométriques soient également stabilisées. Nous pensons que l'amplitude temporelle de 10 minutes est trop grande et qu'elle génère une dispersion des mesures. En effet, si les pièces sont conservées dans le même atelier de production, alors nous pouvons estimer que leur vitesse de refroidissement est identique, aussi pour une même durée de refroidissement, les caractéristiques seront identiques. Or cette durée de refroidissement varie ici beaucoup. Nous recommandons de mesurer les pièces quelques secondes après la sortie du moule, en maîtrisant le délai de mesure, afin de limiter cette dispersion.

Les résultats de l'étude montrent que le plan Central Composite n'inclue qu'un petit espace de la plage de réglage du procédé. Il est alors difficile de trouver un point de fonctionnement optimal, à partir du modèle issu du plan d'expérience. L'étude aurait pu positionner les bornes du plan d'expériences aux limites du procédé. Le système d'apprentissage proposé est capable de trouver un point de fonctionnement satisfaisant en 10 séries d'essais, tandis que le plan Central Composite demande 45 séries d'essais. Enfin, nous notons que la mesure des caractéristiques qualités du produit est ici facilité car ce sont des disques DVD. Les mesures ont une résolution élevée puisqu'un micromètre laser mesure un profil sur la surface des disques. La capacité du moyen de mesure est excellente. Une bonne capacité du système de mesures est un pré-requis, pour la caractérisation des pièces, et pour le pilotage du procédé.

En 2001, LAU et al. propose de réaliser un modèle inverse du procédé, à multiples entrées et multiples sorties, par apprentissage d'un réseau de neurones [Lau+01]. L'objectif est que le réseau suggère des ajustements des paramètres du procédé pour obtenir une pièce cible aux dimensions

spécifiées. Les caractéristiques dimensionnelles portent sur trois longueurs et l'épaisseur de la pièce. L'apprentissage du réseau est réalisé sur les mesures récoltés pour 100 pièces, en faisant varier 6 paramètres du procédé : la température du fourreau centrale et la température de l'arrière du fourreau, la vitesse de fermeture de l'outillage, la vitesse d'avance de la vis, la force de fermeture, la durée d'injection, la durée de refroidissement, la vitesse d'avance de la vis lors de l'injection et la pression d'injection. L'étude mentionne que, malgré la répétition de paramètres identiques, les quatre dimensions étudiées varient entre différentes journées de réalisation des essais. Les valeurs des dimensions attendues sont appliquées en entrée du modèle. Ce dernier indique en sortie les valeurs du point de fonctionnement permettant de produire une pièce avec ces dimensions. La démarche du système est alors d'obtenir l'écart entre les paramètres pour les caractéristiques actuelles et pour les caractéristiques cibles, puis d'ajuster les paramètres en soustrayant ces écarts, afin de compenser les variations des caractéristiques pièces. En connaissant le réglage utilisé pour produire les pièces, l'étude propose de régler les paramètres dans des proportions inverses afin de compenser les variations de pièces. Cette démarche suppose que les paramètres sont indépendants, ce qui n'est pas le cas pour le procédé d'injection-moulage.

Récemment, en 2015, [Joh+15] réalise une étude à l'aide d'un plan d'expériences à 12 facteurs de 15 essais, pendant 30 minutes de production par essais, totalisant 942 cycles d'injection-moulage, soit 60 pièces par essais. Pour chaque cycle, une boucle d'optimisation évalue et minimise une contrainte imposée, par un ajustement des paramètres du procédés ( $P_1, P_2, P_3$  §1.5). Le contrôleur qui est proposé est capable de détecter les variations du procédé par comparaison avec un modèle de référence préétabli. Un objectif de réduction du temps de cycle est ajouté grâce à une pénalité temporelle dans la fonction à minimiser. Un double objectif est atteint : le temps de cycle est réduit et la variabilité des dimensions des pièces est réduite. L'étude conclut qu'une optimisation multi-paramètres, par exemple sur les caractéristiques dimensionnelles, est possible.

#### 1.1.4 La mesure de la qualité du produit : un verrou technologique au pilotage

Nous concluons cette revue de la littérature par un tableau récapitulatif de l'ensemble des articles cités, Tableau 1.8). On observe les travaux initiaux qui utilisent la théorie Automatique pour réguler le procédé. Par la suite, la modélisation non-linéaire a été perfectionnée par l'utilisation des réseaux de neurones Dernièrement, les travaux s'appuient sur des démarches de Maîtrise Statistique des Procédés et sur la simulation numérique. Pour chacune des publications, nous identifions les variables du procédé étudiées selon la représentation Zig Zag, présenté dans la Figure 1.5, ainsi que les résultats obtenus. Le procédé d'injection possède plusieurs phases interdépendantes, mais la littérature ne s'intéresse pas à l'étude de l'ensemble des variables et paramètres du procédé. De nombreuses études présentées se focalisent sur une unique phase du procédé. Les études de la littérature sont centrées sur les paramètres de la phase d'injection ( $P_3$ ).

La mesure des caractéristiques qualité des produits, dès leurs sorties de l'outillage est peu étudié. La caractéristique des pièces produites qui est la plus étudiée dans la littérature est la masse. Cela s'explique par la facilité de la réalisation de la pesée, dès la sortie de la machine, ainsi que par la bonne précision des balances. Dans l'industrie, la masse est souvent considérée comme une caractéristique nécessaire pour valider les pièces. Si les pièces ont des manques, la masse sera plus faible. La quantité de matière compactée afin de produire une pièce fait varier les caractéristiques mécaniques de celle-ci. Il y a également un lien entre le taux d'humidité de la matière et la masse des pièces.

Les caractéristiques d'aspects des pièces sont en revanche peu étudiées. Les cahiers des charges posent pourtant des contraintes en termes d'aspect : pas de trace, pas de tâche, une couleur uniforme, une texture définie. Pour les grandes séries, le contrôle qualité dimensionnel est souvent réalisé par imagerie de profils. Il demande un équipement calibré et automatisé. Il prend place après la production et il permet d'exclure les pièces non conformes. Ce contrôle dimensionnel ne s'intéresse pas à l'aspect des pièces, seulement aux profils.

Année	Auteurs	Méthodes	Régulation	Ajustement								Résultats
				DPO	PV1	DP1	PV2	DP2	PV3	DP3	PV4	
1975	[Laczko, 1975]	Automatique										
1983	[Sanschagrin, 1983]	Automatique		■	■		■	■		■		dispersion de masses -50%
1985	[Fara et al., 1985]	Automatique										
1986	[Gomes et al., 1986]	Automatique						■	■	■	■	
1987	[Agrawal et al., 1987]	Automatique				■	■	■	■			
1987	[Kamal et al., 1987]	Automatique										
1987	[Gustafson et Lebow, 1987]	Automatique							■	■		
1988	[Fara et I, 1988]	Automatique										
1990	[Fara et al., 1990]	Automatique							■	■		
1990	[Devos, 1990]	Automatique		■	■							
1991	[Chiu et al., 1991]	Automatique							■	■	■	
1991	[Srinivasan et al., 1991]	Automatique							■	■		
1993	[Haeussler et Wortberg, 1993]	Réseaux Neur	■	■	■	■	■	■	■	■		corrél. 80% masse/88% dimens.
1994	[Blyskal et Meheran, 1994]	Statistique		■	■	■	■	■	■	■		
1994	[Gao et al., 1994]	Automatique						■	■	■		
1995	[Kazmer et Barkan, 1995]	Automatique			■	■						Cpk amélioré de 0,52 à 1,67
1996	[Thomas et al., 1996]	Automatique					■	■				
1996	[Woll et al., 1996]	Réseaux Neur	■	■	■	■	■	■	■	■		prédiction de masse
1997	[Demirci et al., 1997]	Réseaux Neur										
1997	[Woll et Cooper, 1997]	Réseaux Neur				■	■	■	■	■		
1997	[Tsoi, 1997]	Automatique							■	■		
1998	[Ryckebusch et al., 1998]	Automatique	■	■	■	■	■	■	■	■		dispersion de masses -43%
1998	[Ivester et al., 1998]	Statistique	■	■	■	■	■	■	■	■		
1999	[Tsoi et Gao, 1999]	Automatique	■	■	■	■	■	■	■	■		
2000	[Huang et Lee, 2000]	Automatique	■	■	■	■	■	■	■	■		
2000	[Schnerr-Häselbarth et Mi	Réseaux Neur	■	■	■	■	■	■	■	■		prédiction masse de 86 à 95,2%
2000	[Yang et al., 2000]	Statistique	■	■	■	■	■	■	■	■		
2001	[Lau et al., 2001]	Réseaux Neur	■	■	■	■	■	■	■	■		variations dimens. de 0,07RMS
2002	[Chen, 2002]	Automatique	■	■	■	■	■	■	■	■		
2002	[Lu et al., 2002]	Statistique			■	■	■	■	■	■		détection de défauts
2004	[Béreaux et al., 2004]	Simulation										
2004	[Yang, 2004]	Automatique	■	■	■	■	■	■	■	■		vitesses ajustées cycle à cycle
2009	[Moguedet et al., 2009]	Simulation										
2009	[Michaeli et Schreiber, 2009]	Réseaux Neur	■	■	■	■	■	■	■	■		dispersion de masse réduite
2013	[Fei et al., 2013]	Statistique		■	■	■	■	■	■	■		
2015	[Johnston et al., 2015]	Statistique	■	■	■	■	■	■	■	■		optimisation des réglages
2016	[Zhang et al., 2016]	Statistique										91% détection défaut masse

Figure 1.7 – Cartographie bibliographique de la maîtrise du procédé d'injection-moulage.

De plus, les caractéristiques d'aspect contiennent des informations importantes pour l'analyse des causes de la variabilité d'une pièce. Le technicien-régleur utilise prioritairement l'aspect des pièces pour régler la machine. Pourtant, aucun système de pilotage automatique n'a aujourd'hui pris en compte les caractéristiques visuelles des pièces. La maîtrise de l'aspect dans les productions industrielles est une thématique de recherche importante. Le manque de spécifications et la difficulté à mesurer, en cycle, limite l'utilisation de la mesure des caractéristiques d'aspect.

## 1.2 Objectifs de recherche : enjeux du déploiement du contrôle qualité en ligne de production

L'étude de la littérature nous permet d'affirmer que la mesure de la qualité est cruciale pour optimiser le procédé d'injection-moulage. En étant réalisée de manière automatique au sein de la ligne de production, le contrôle de la qualité permet d'ajuster les réglages de la machine dès la pièce suivante. Dans nos travaux, nous chercherons à proposer un dispositif de mesure adapté aux contraintes industrielles.

Nous avons identifié trois verrous au déploiement massif du contrôle qualité à cent pour-cent sur les lignes de production :

**Économique** : c'est le coût du dispositif de mesure,

**Technologique** : le respect des contraintes industriel,

**Humain** : la formalisation de la notion de qualité.

### 1.2.1 Viabilité économique du déploiement de la mesure de la qualité en ligne de production

L'automatisation du contrôle des produits est principalement déployée pour des productions à gros volumes. Les productions industrielles ont pour caractéristiques des cadences élevées (temps de cycle court) et une marge sur le produit finale faible. Spécifique au secteur de la plasturgie, la Figure [Dir19] indique une marge moyenne de 27% pour le secteur, en France en 2016.

Contrairement à la transformation de la matière première, le contrôle de la qualité n'apporte pas de valeur au produit final ; c'est une sécurité pour respecter le cahier des charges et palier aux dérives du procédé. C'est pourquoi le coût d'intégration d'un dispositif de mesure de la qualité se doit d'être faible.

Le contrôle qualité à cent pour-cent présente un coût d'installation et de fonctionnement. Il est nécessaire de définir les caractéristiques d'un système de mesure économiquement viable, pour les productions industrielles.

#### 1.2.1.1 Intérêt économique du contrôle de la qualité en ligne de production

La mesure par prélèvement de pièces, sur un procédé qui possède des étapes successives, ne permet pas d'identifier l'étape à l'origine du défaut. Pour un produit qui nécessite un niveau de qualité élevé, un contrôle qualité final de chaque pièce sera réalisé (contrôle à cent pour-cent). Dans le cas de l'injection-moulage de pièces plastiques, le contrôle qualité final est effectué bien après l'étape de moulage. Le contrôle final est réalisé après l'ajout de valeur des étapes de finitions, ce qui prend parfois plusieurs jours. Si la pièce était de mauvaise qualité juste après le moulage, alors de la valeur a été ajoutée à une pièce mauvaise. Ce coût peut être très important pour une production industrielle. Les coûts des étapes de finitions, telles que la peinture ou la galvanoplastie, sont bien plus importants que le coût de l'injection-moulage. Le temps de cycle du procédé permet

Secteurs d'activité*	Nombre d'entreprises	Chiffre d'affaires HT (en Md€)	Valeur ajoutée - y compris produits et autres charges (en Md€)
Industries agroalimentaires	59 765	181	41
Réparation, installation et produits manufacturés divers	56 356	53	20
Bois, papier et imprimerie	28 542	39	11
Métallurgie et produits métalliques	22 007	81	25
Textile, habillement, cuir et chaussures	19 950	21	6
Plastique, caoutchouc et produits minéraux non métalliques	13 529	60	18
Machines et équipements	4 238	48	14
Industrie chimique	3 042	67	18
Produits informatiques, électroniques et optiques	2 814	32	12
Équipements électriques	2 299	29	8
Industrie automobile	1 925	118	19
Matériels de transport hors automobile	895	110	19
Industrie pharmaceutique	644	37	12
Cokéfaction et raffinage	37	34	5
<b>Industrie manufacturière</b>	<b>216 045</b>	<b>908</b>	<b>229</b>

\* nomenclature NAF, niveau A38, sauf matériels de transport (A88).

Champ : ensemble des entreprises marchandes, y compris auto-entrepreneurs.

Source : Insee, Ésane 2016.

*Figure 1.8 – Figure issue du rapport annuel de la DIRECTION GÉNÉRALE DES ENTREPRISES : Chiffres clés de l'industrie manufacturière, Édition 2018 [Dir19]*

de produire une pièce toutes les 30 secondes, ce qui correspond pour une journée à mille pièces. Si le défaut de qualité est détecté un jour après son apparition, cela correspond à 1000 pièces rebuts qui ont été finit inutilement. L'objectif du contrôle qualité en ligne est de détecter les dérives de la qualité le plus tôt possible, afin d'éviter de produire. Dans un second temps, il pourrait être possible d'ajuster les paramètres du procédé pour corriger les dérives qualité.

### 1.2.1.2 Coût acceptable du dispositif de mesure

Dans le cas de l'injection plastique, le coût de la matière première et de l'étape d'injection-moulage est très faible en comparaison du coût des étapes de finitions. Produire des pièces mauvaises à un coût négligeable en comparaison du coût des étapes de peinture et de finitions de pièces mauvaises. Détecter les pièces mauvaises dès la sortie de la presse d'injection, afin de les écarter de la chaîne de production, permet de diminuer les rebuts finaux et de supprimer ainsi les coûts associés à leurs finitions. Le coût économique du dispositif de mesure est à rapporter au coût économique des rebuts. Des industriels nous ont confirmés que le coût de la production de rebuts est faible en comparaison du coût de déploiement d'un dispositif de mesure sur chaque machine. Si le coût du système de mesure dépasse la dizaine de milliers d'euros, il n'est pas intéressant. Un coût élevé restera néanmoins intéressant à plus long terme, or nous cherchons à obtenir une adoption rapide.

La méthode la plus utilisée afin de maîtriser le procédé d'injection-moulage est de positionner des capteurs à l'intérieur de l'outillage, directement en contact avec la matière. Une analyse multivariée permet ensuite de détecter les dérives du procédé. Les capteurs les plus répandus sont les sondes de températures et les sondes de pressions [Kur+09]. Le coût de l'instrumentation invasive des moules est élevé. D'après nos chiffres régionaux, l'intégration de capteurs dans l'outillage représente un surcoût de vingt pour-cent de la conception et de la fabrication de l'outillage, pour un coût initialement compris entre cinquante et cent mille euros. Intégrer les conduits nécessaires au passage des fils et connecteurs est compliqué sachant que les moules intègrent également des canaux de refroidissements indispensables au procédé. Les capteurs subissent de fortes contraintes ce qui peut limiter leur durée de vie. Enfin, un capteur intégré à un outillage est associé un outillage, et donc à une unique pièce. Il n'est pas viable de démonter l'instrumentation du capteur pour le positionner sur un autre outillage. À contrario, un dispositif de mesure extérieur au procédé ne serait pas lié à un outillage. Cela permettrait de contrôler la qualité quand le cahier des charges le nécessite, quel que soit la pièce produite. L'installation du dispositif de mesure devra être simple, d'une durée inférieure à une heure, pour que sa mobilité soit réelle. Un tel dispositif de mesure s'inscrit dans la démarche de l'industrie agile, où les lignes de productions sont fréquemment reconfigurées, en fonction des besoins.

### 1.2.2 Faisabilité technologique de la mesure de la qualité en ligne de production

Pour identifier un procédé, il est nécessaire de réaliser des mesures afin de connaître l'état de celui-ci à différents instants discrets. Afin de pouvoir analyser l'évolution du procédé, l'échantillonnage des mesures doit être supérieure au temps caractéristique d'évolution du procédé. Dans le cas de l'injection-moulage, la périodicité court-terme du procédé est celle du cycle de production d'une pièce, soit le temps de cycle du procédé, de 10 à 60 secondes.

Pour respecter cette contrainte, nous distinguons deux grands types de mesures :

- les mesures dites "invasives", qui nécessite l'installation de capteurs de grandeurs physiques à l'intérieur de l'outillage,
- les mesures "non-invasives", qui s'intéressent à la pièce produite dès qu'elle sort de l'outillage.

Les mesures "invasives" nécessitent un travail de conception de l'outillage compliqué. L'outillage intègre des canaux de refroidissement et le câblage des capteurs doit être conçu

pour les éviter. C'est pourquoi, GAO et KAZMER proposent un capteur qui communiquent sans connectivité filaire [Kaz+11 ; GK12]. L'outillage en acier conducteur est une cage de Faraday idéale. La communication radiofréquence est impossible. C'est pourquoi ce capteur utilise des ondes acoustiques pour communiquer. L'alimentation en énergie est assurée par le capteur, en récupérant les variations de pressions créées lors du cycle l'injection. Cependant, cette solution n'est pour l'instant pas commercialisée. De plus, la maintenance des capteurs intégrés à l'outillage est compliquée. Les capteurs sont fragiles et les contraintes physiques qui s'exercent sur eux lors du cycle d'injection sont élevées. En particulier, nous avons observé que les défaillances proviennent le plus souvent des câbles et non des capteurs eux-mêmes. En cas de défaillance d'un capteur, c'est l'ensemble de l'outillage qui est immobilisé, ce qui impacte fortement la production industrielle. À contrario, la mesure des pièces, dès leur sortie de l'outillage, ne demande pas ce travail de conception avancée. Le système de mesure est éloignée des contraintes physiques du procédé d'injection-moulage, ce qui limite le besoin de maintenance.

Pour être intégrée à la ligne de production, la mesure doit être réalisée dans une durée inférieure au temps de cycle du procédé d'injection-moulage. Cette contrainte permet de réaliser le contrôle de cent pour-cent des pièces. Si la pièce doit être déplacée devant le moyen de mesure, la durée du convoyage doit également être inclue. De plus, si un ajustement des paramètres du procédé est attendu dès la pièce suivante, il est alors nécessaire d'effectuer l'analyse de la mesure pendant le temps de cycle. Ces contraintes réduisent d'autant plus la durée disponible pour réaliser la mesure. Nous estimons la durée maximale de la mesure à 10 secondes. La Figure 1.9 inscrit la mesure dans le chronogramme du cycle du procédé (§1.4). Cette limite sera une contrainte forte pour l'ensemble des choix technologiques de nos travaux.

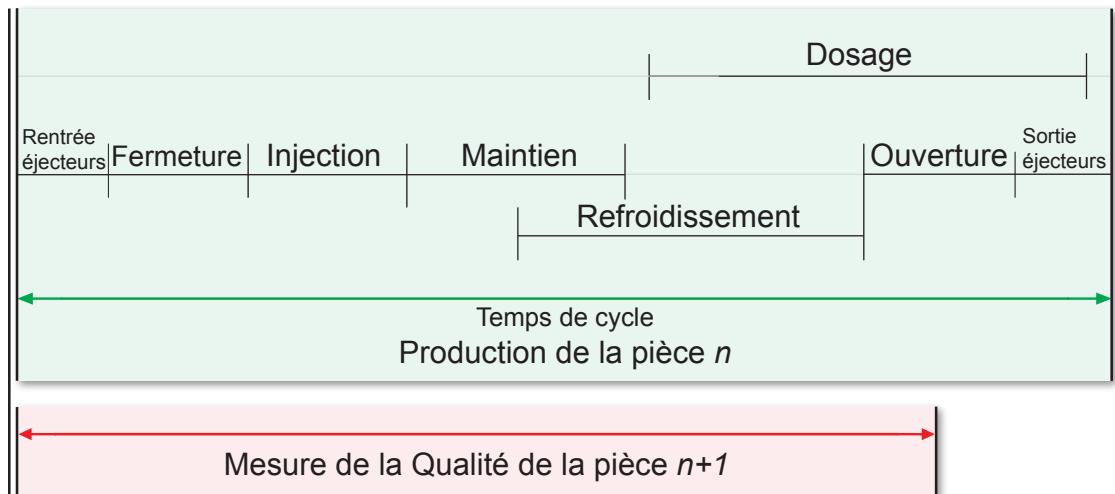


Figure 1.9 – Intégration de la mesure dans le cycle d'injection-moulage.

Différentes mesures sont aujourd'hui compatibles, de par leurs courtes durées, avec une utilisation en cycle industriel : pesée de la masse, imagerie par thermographie infrarouge, contrôle d'aspect par caméras, imagerie tridimensionnelle [Sch+02], durométrie, imagerie des champs de contraintes par l'imagerie des phénomènes de photoélasticité sur les pièces plastiques transparentes. Pour obtenir des mesures fiables et répétables la mesure doit être automatisée. C'est un des objectifs de nos travaux. Nous discuterons de manière approfondie des techniques de mesure existantes dans le Chapitre 2.

### 1.2.3 Formalisation de la notion de qualité et intégration de l'expert humain

Aucune norme ne statue actuellement sur la qualité d'aspect. Plusieurs travaux de recherches sont en cours. On distinguera la sensation tactile [Alb+16a ; Alb+16b ; Alb+17 ; Alb+19 ; Alb19]

## 1.2. Objectifs de recherche : enjeux du déploiement du contrôle qualité en ligne de production

de l'aspect esthétique [Des+15]. Ces travaux associent une démarche de spécifications de la qualité, aux développements de nouveaux moyens de mesures [Des15 ; Pit16 ; Lac18]. Nos travaux s'inscrivent dans cette démarche de recherche.

Les pièces produites en injection-moulage des thermoplastiques possèdent une grande variété de dimension, de forme et d'état de surface. Au sein d'une même entreprise, un système de contrôle de la qualité doit être capable de mesurer une grande variété de pièces. Nous chercherons à simplifier au maximum l'utilisation de notre système de mesure afin de maximiser son adoption, tout en répondant à la problématique du contrôle de la qualité en ligne de production.

Le principal objectif est d'exploiter l'information issues des capteurs, afin de proposer une mesure de la qualité d'aspect. La qualité d'aspect est une notion subjective qui est souvent définie dans les cahiers des charges sous la forme de défautthèques. L'interprétation de ces définitions nécessite le travail d'un expert qualité humain. L'expert qualité est généralement le référent de la notion de qualité pour un atelier de production. Il forme les opérateurs, afin de leur transmettre les exigences qui sont interprétées à partir du cahier des charges. Malheureusement, la transmission du savoir et la répétabilité de la mesure de la qualité par différents opérateurs, sont difficiles à mettre en œuvre. Dans nos travaux, nous chercherons à automatiser cette interprétation. Le système réalisera l'interprétation des mesures à partir d'un modèle de la qualité construit par apprentissage (Chapitre 3). À terme, l'objectif est de proposer un modèle de la qualité qui serait générique à tous types de pièces, ; tel qu'un humain formé à l'expertise qualité est capable d'inspecter tous types de pièces.



# 2

## Mesure industrielle non-invasive de la qualité

---

*La mesure des caractéristiques d'une pièce est un pré-requis indispensable à la maîtrise de la qualité. Dans ce chapitre, nous discuterons du choix des capteurs pour mesurer des pièces plastiques en ligne de production, juste après la sortie du moule. Deux expérimentations permettent de justifier de la pertinence de la mesure par thermographie et de la mesure par polarimétrie. Nous exploiterons l'information de ces mesures afin de classifier les pièces conformes et les pièces non-conformes, à l'aide de méthodes par apprentissage.*

---

### Sommaire

---

<b>2.1</b>	<b>Qualité d'un produit en injection-moulage des thermoplastiques . . . . .</b>	<b>32</b>
<b>2.2</b>	<b>Mesures en ligne de production . . . . .</b>	<b>33</b>
2.2.1	Mesures indirectes des caractéristiques du produit . . . . .	36
2.2.2	Mesures directes des caractéristiques du produit . . . . .	38
2.2.3	Définition du degré d'invasivité d'un système de mesure pour un procédé	39
2.2.4	Corrélation entre les caractéristiques des pièces chaudes et des pièces finales . . . . .	41
<b>2.3</b>	<b>Apport de l'imagerie non-conventionnelle pour la mesure de la qualité . . . . .</b>	<b>43</b>
2.3.1	Mesure géométrique par triangulation laser . . . . .	44
2.3.2	Apport de la thermographie pour le contrôle de la géométrie des pièces	46
2.3.3	Apport de la polarimétrie pour le contrôle des défauts d'aspect . . . . .	52
<b>2.4</b>	<b>Traitemen et fusion de l'information issue de mesures multimodales . . . . .</b>	<b>59</b>
2.4.1	Nécessité de l'extraction de l'information pertinente . . . . .	60
2.4.2	Traitemen d'images amont . . . . .	60
2.4.3	Descripteurs statistiques . . . . .	61
2.4.4	Sélection automatique de descripteurs pertinents . . . . .	65
2.4.5	Description d'images par apprentissage statistique . . . . .	65
<b>2.5</b>	<b>Synthèse et perspectives . . . . .</b>	<b>67</b>
2.5.1	Synthèse : mesure non-invasive en cycle industriel . . . . .	67
2.5.2	Perspectives de recherche . . . . .	67

---

## 2.1 Qualité d'un produit en injection-moulage des thermoplastiques

La norme ISO9000 [ISO15] spécifie la notion de qualité comme l'*aptitude d'un ensemble de caractéristiques intrinsèques d'un objet à satisfaire des exigences*. Dans notre cadre des produits industriels, les exigences sont celles du client finale ou du donneur d'ordre. Elles sont spécifiées dans le cahier des charges du produit. Nous distinguons cinq grands types de caractéristiques :

- géométriques : normes ISO GPS *Global Product Specifications* [ISO11],
- mécaniques,
- sensoriels : visuelles, haptiques, olfactives,
- chimiques,
- économiques.

Les pièces industrielles sur lesquelles nous avons travaillées sont fabriquées par injection-moulage de thermoplastiques. Dans le cadre de ce procédé, les caractéristiques chimiques sont définies par le choix de la matière initiale. Les caractéristiques économiques sont les coûts associés à la production des pièces. Les caractéristiques mécaniques dépendent du choix du matériau, de la géométrie des pièces et des paramètres du procédé d'injection-moulage.

Dans le cadre de ce travail, nous nous intéressons en particulier aux caractéristiques géométriques et aux caractéristiques d'aspects. La maîtrise de ces deux types de caractéristiques est un enjeu important pour le secteur de la plasturgie. Les tolérances dimensionnelles sont de plus en plus élevées. À celles-ci s'ajoutent l'exigence d'aucune non-conformité de l'aspect des pièces. Enfin, l'amélioration de la qualité est un facteur important de rentabilité et de compétitivité des entreprises. Dès 1950, DEMING [Dem82] décrit la réaction en chaîne que permet d'initier l'amélioration de la qualité : réduction des coûts des rebuts, diminution du prix de revient, gain de marchés et, à terme, création d'emplois.

Non-conformité	Description	Causes probables
Homothétie	dimension modifiée	retrait de la matière
Gauchissement	défaut de surfaces	refroidissement non homogène
Retassure	réduction de dimension	pression de maintien, vitesse d'injection
Incomplet	trou dans la pièces	quantité de matière injectée, clapet anti-retour défaillant
Ligne de soudure	ligne décoloré	pression d'injection, température matière
Jet libre	trace décolorée	viscosité matière, conception moule, pression d'injection
Éjection	trace de l'éjecteur	vitesse éjecteur, température matière, durée refroidissement
Combustion	trace noire	conception du moule, température de la matière
Givrage	bulle ou trace blanche	matière humide
Pollution	trace colorée	usure de l'outil, pollution de la matière

Tableau 2.1 – Principaux défauts d'aspects en injection-moulage des thermoplastiques.

Le Tableau 2.1 présente les non-conformités d'aspect usuelles en injection-moulage des thermoplastiques. Sur des pièces qui seront utilisées en contraintes et en fatigues, un défaut de ligne de soudure entraîne une fragilité. La majorité des non-conformités d'aspect sont des défauts de surface géométriques. Seules la combustion, le givrage et la pollution sont entraînées pas la modification des propriétés du polymère. Ces imperfections géométriques sont souvent dues au mauvais maintien du polymère contre les parois de l'outil, ou bien à la solidification trop rapide de la matière. Les défauts d'incomplets sont les plus simples à détecter par vision

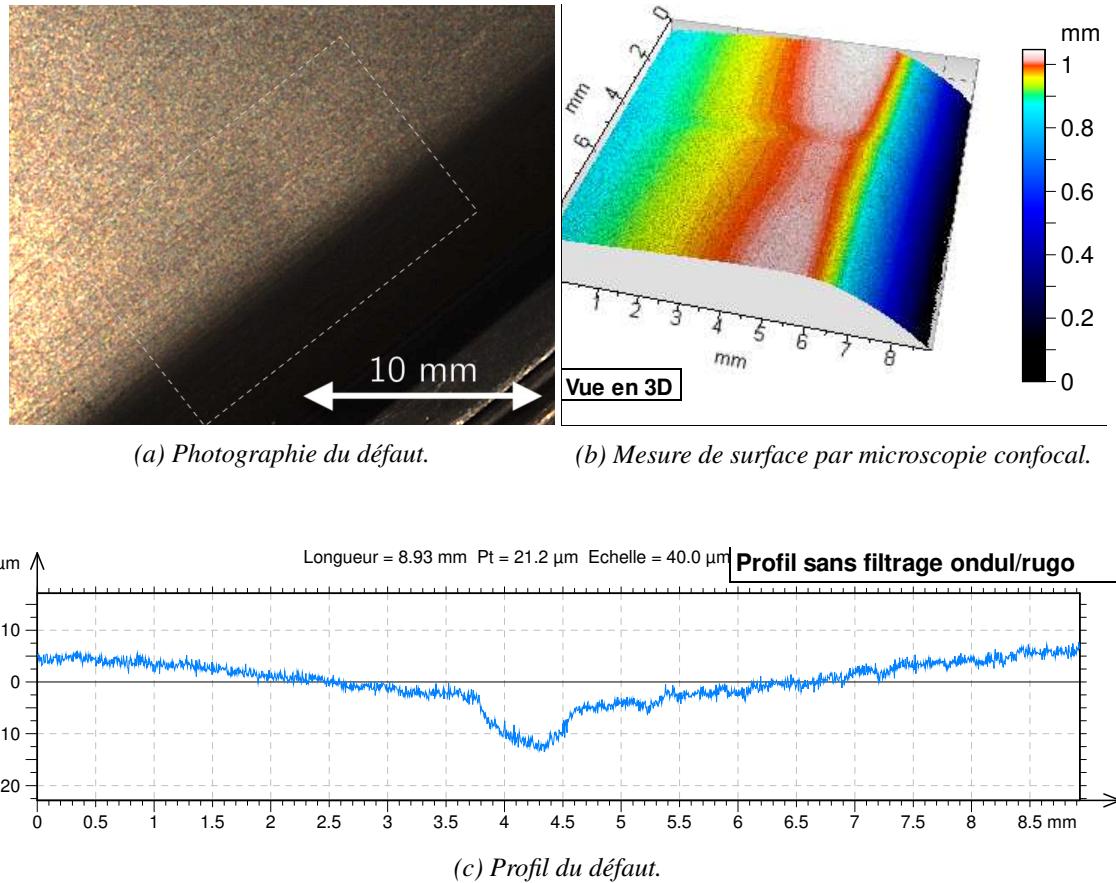


Figure 2.1 – Défaut de retassure mesurée par microscope confocal, AltiMet 520<sup>0</sup>.

informatique : un seuillage puis une détection de contours permet de mettre en évidence le manque de matière. La détection des autres non-conformité est plus compliquée. C'est notre objectif principal de recherche.

La Figure 2.1 présente une mesure par microscopie confocal d'un défaut de retassure, sur une pièce plastique automobile de deux mètres. Ce défaut est caractéristiques des pièces automobiles qui sont peintes afin de réfléchir la lumière de manière harmonieuse. La légère inflexion de la surface entraîne une "cassure" de la réflexion de la lumière, ce qui attire le regard et amplifie le défaut. L'écart géométrique de l'inflexion est de l'ordre de la dizaine de micromètres. La mesure de la géométrie à l'échelle micrométrique, pendant le temps de cycle du procédé, est un véritable défi. Dans la suite de ce chapitre, nous présenterons les moyens de mesure existants ; puis nous orienterons notre étude sur les moyens de mesure sans contact ; en particulier nous nous intéresserons aux moyens de mesure compatibles avec le temps de cycle du procédé d'injection-moulage.

## 2.2 Mesures en ligne de production

Dans le cadre de productions industrielles par injection-moulage des thermoplastiques, les contrôles qualité sont traditionnellement effectués en sortie des lignes de production de pièces. Il

0. Spécifications techniques de l'Altisurf 520 sur le site Internet de la société Altimet.

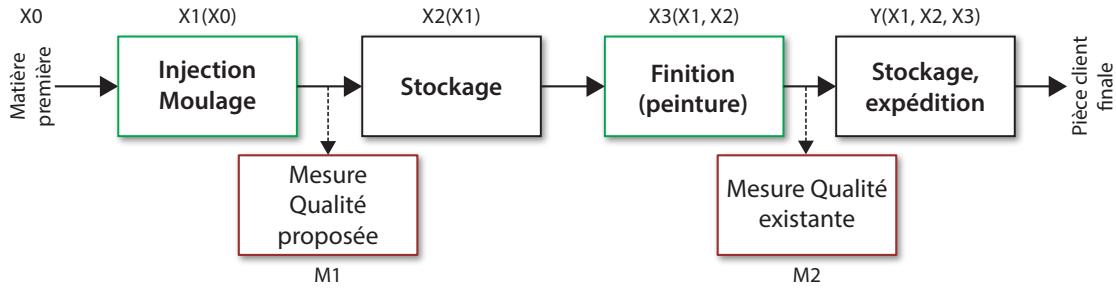


Figure 2.2 – Intégration du contrôle de la qualité

s'agit de s'assurer que les pièces sont conformes aux exigences spécifiées par le cahier des charges. Les pièces non-conformes sont mises au rebut. Le contrôle qualité final est réalisé à l'issue des étapes de finitions, telles que la peinture. Pour des productions de pièces à faible valeur ajoutée, le contrôle est réalisé par prélèvement aléatoire de pièces. Il est effectué par un opérateur humain. Le contrôle peut nécessiter la mesure de dimensions géométriques, ou de propriétés mécaniques de la pièces. La mesure géométrique est le plus souvent réalisée à l'aide d'un pied à coulisse. Dans le cas de pièces techniques à forte valeur ajoutée, les exigences du client peuvent atteindre 5 pièces défectueuses par million de pièces (*5 ppm* pour les secteurs automobile, aéronautique, de la santé). Il est alors indispensable de réaliser le contrôle qualité sur cent pourcent des pièces, avant l'envoi au client. Ce contrôle présente un coût et il devient intéressant de l'automatiser.

La Figure 2.2 présente l'intégration du contrôle de la qualité au sein de la ligne de production. Le contrôle de la qualité des produits est de manière classique réalisé après la finition de la pièce. L'état de la pièce dépend alors des paramètres des étapes précédentes de finitions  $X_3$ , stockage  $X_2$ , transport et en premier lieu de l'injection-moulage  $X_1$ . Des défauts peuvent apparaître lors des phases de transports, stockages et de finitions. Plus on multiplie le nombre d'étapes successives, plus il est difficile de trouver l'étape qui est la cause du défaut. De plus, si la pièce est défectueuse à l'étape de l'injection-moulage  $X_1$ , il est coûteux de stocker cette pièce  $X_2$  et il est encore plus coûteux de réaliser sa finition  $X_3$ . Le contrôle qualité en fin de chaîne de production, à l'instant  $M_2$ , ne permet d'éviter ces coûts. Pour l'ensemble de ces raisons, nous proposons de réaliser la mesure de la qualité du produit à l'instant  $M_1$ , dès la sortie du moule d'injection, alors que les pièces sont encore chaudes. En plus d'éliminer les rebuts de la chaîne de production, il deviendrait alors possible de construire un modèle de causalité entre les paramètres du procédé d'injection-moulage et la qualité de la pièce. Ce modèle permettrait d'ajuster les paramètres du procédé, pièces après pièces, pour optimiser la qualité. La construction de ce modèle et le pilotage du procédé d'injection-moulage ne sont pas l'objet de notre travail. C'est une perspective de recherche qui pourra s'inscrire à sa suite.

Le contrôle qualité est souvent réalisé de manière binaire : pièces conformes ou pièces non-conformes. Des grilles de notations peuvent être utilisées pour attribuer une note de la qualité. On réalise alors la moyenne géométrique de la note pour chacun des critères. Ce processus de notation est néanmoins beaucoup plus long. Il est réservé aux productions de pièces dont le prix de revient est élevé. Ainsi, la mesure de la qualité est rarement réalisée et on parle de contrôle (binaire) de la qualité. Les causes sont discutées dans le Chapitre 1 §1.2. Afin de réaliser une mesure de la qualité, notre travail propose également de définir une métrique de la qualité par apprentissage, à partir des données issues des capteurs, Chapitre 3.3.

Dans ce chapitre, nous discuterons du choix des capteurs à mettre en œuvre, pour mesurer en ligne de production. La littérature a proposé de réaliser des mesures indirectes<sup>1</sup> de la qualité,

1. On parle d'une mesure indirecte lorsqu'il ne s'agit pas de mesurer directement la caractéristique de la pièce. La mesure de pression dans l'outillage est une mesure indirecte des caractéristiques finales de la pièces. C'est un indicateur

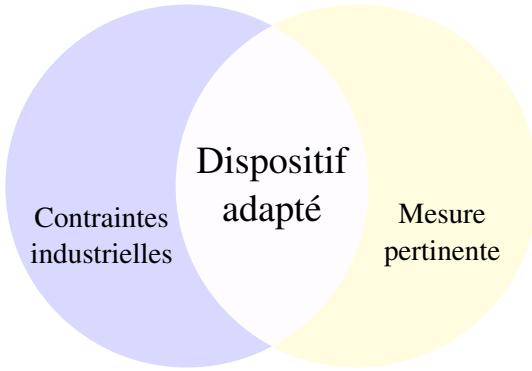


Figure 2.3 – Diagramme de Venn d'un dispositif de mesure adapté.

Contraintes industrielles	Limites pour la mesure
Flux continu	Mesure de toutes les pièces
Durée du cycle	Durée de 10 secondes
Coût du moyen de mesure	Inférieur à 10 000 €
Type de défauts	Géométrie et couleur
Dimensions du défaut	10 micromètres à 1 mètre
Dimensions de la pièce	10 à 200 centimètres
Distance à la pièce	0 à 200 centimètres
Dimensions du moyen de mesure	Inférieure à $1m^3$
Taux de faux positifs	1 à 8 pièces par journée
Taux de faux négatifs	100 à 5 pièces par million

Tableau 2.2 – Contraintes industrielles pour la contrôle en ligne de la qualité.

comme par exemple des mesures de température ou de pression sur le moyen de production. C'est pourquoi nous discuterons également de la pertinence des mesures. La Figure 2.3 positionne le dispositif de mesure adéquate, comme l'intersection entre les contraintes industrielles et la pertinence de la mesure, pour les caractéristiques qualité de la pièce. Les contraintes qui s'appliquent à la réalisation de la mesure du produit, au sein de la ligne de production, sont synthétisées dans le Tableau ??.

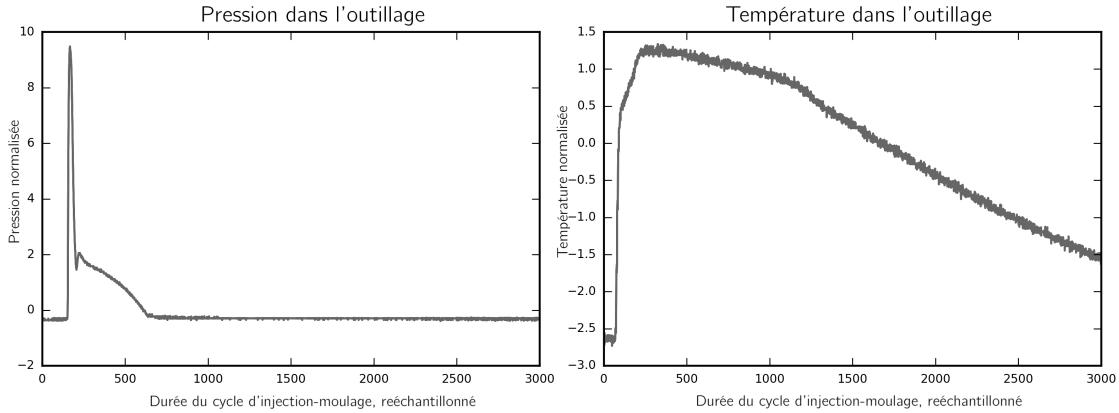
Une des contraintes du contrôle en ligne de production est que, si une mesure par prélèvement est réalisée, alors la pièce prélevée ne peut pas être ré-intégrée au flux de production. À cette contrainte s'ajoute les bénéfices de réaliser un contrôle à cent pourcent :

- du point de vue de la traçabilité des pièces pour le client, qui dispose d'une fiche de qualité pour chaque pièce,
- pour l'analyse des performances du procédé et son optimisation,
- du point de vue de l'assurance de respecter les exigences du nombres de pièces défectueuses par lot.

Le contrôle à cent pour-cent nécessite alors d'être réalisé pendant le temps de cycle du procédé. La durée de la mesure est d'autant plus réduite que la pièce doit être amenée devant le dispositif

---

des caractéristiques du produit.



*Figure 2.4 – Mesures de variables physiques dans l’outillage.*

de mesure. Il est nécessaire d’inclure la durée des convoiages aller et retour. Enfin, il n’est pas envisageable d’augmenter la durée du cycle du procédé pour pouvoir réaliser la mesure, car cela impacterait très fortement la rentabilité du moyen de production. C’est pourquoi nous estimons que la durée de la mesure doit être au maximum de 10 secondes. Pour un temps de cycle du procédé de 30 secondes, la durée de convoyage devra elle aussi être inférieure à 10 secondes. Ainsi, le moyen de mesure doit être situé à proximité de la ligne de production. En admettant une vitesse de déplacement du convoyeur de 1 mètre par seconde (ce qui est élevée), le dispositif de mesure devra être situé à moins de 5 mètres de l’emplacement initial de la sortie de la pièce. Cette contrainte de proximité du moyen de mesure dans la ligne de production implique une contrainte de dimensions du moyen de mesure. À partir de nos essais réalisés dans des ateliers de production, nous estimons l’encombrement acceptable à 1 mètre cubique. Il est également recommandé de limiter la complexité du câblage nécessaire à l’installation du moyen de mesure. Enfin, nous avons pour objectif de couvrir une large gamme de pièces plastiques. C’est pourquoi nous définissons des critères sur la dimension des pièces pouvant être mesurées, ainsi que sur la distance du moyen de mesure à la pièce. Dans le cas où le moyen de mesure nécessite une solution optique, cela permet de dimensionner la solution optique : il devra couvrir un angle de champ allant de  $40^\circ$  à  $90^\circ$ . Nous ajoutons une dernière recommandation qui concerne le contrôle dimensionnelles des pièces plastiques souples : il est difficile de toucher ces pièces sans les déformer ; d’autant plus à la sortie du moule, quand elles sont encore chaudes. C’est pourquoi un moyen de mesure sans contact est à privilégier pour le contrôle des pièces souples.

### 2.2.1 Mesures indirectes des caractéristiques du produit

La littérature §1 étudie peu la mesure directe des caractéristiques du produit. La majorité des travaux proposent la mesure de variables physiques sur le procédé, à partir de capteurs positionnés à l’intérieur de l’outillage. En général, un modèle est construit afin de prédire une caractéristique du produit (qui est souvent la masse de la pièce) à partir des mesures dans le moule (souvent la pression et la température). L’argumentaire de ces travaux est que la mesure de variables au plus porche de la matière fondue permet de prédire au mieux les caractéristiques finales de la pièce.

Nous avons réalisé une étude afin de vérifier la pertinence de ce type de mesures indirectes. L’étude “Quality Prediction in Injection Molding” propose de construire un modèle qui détermine la valeur d’cotation géométrique de la pièce produite, à partir de mesures sur le procédé [Nag+17]. À partir des signaux de capteurs de pressions et de températures mesurés dans le moule pendant le cycle du procédé, une valeur géométrique continue est prédite. Un capteur de pression et un capteur de température sont positionnées au contact de la matière. La position optimale des capteurs est définie dans l’étude de AGAZZI et al. [Aga+13]. Les signaux sont acquis à une fréquence de 100

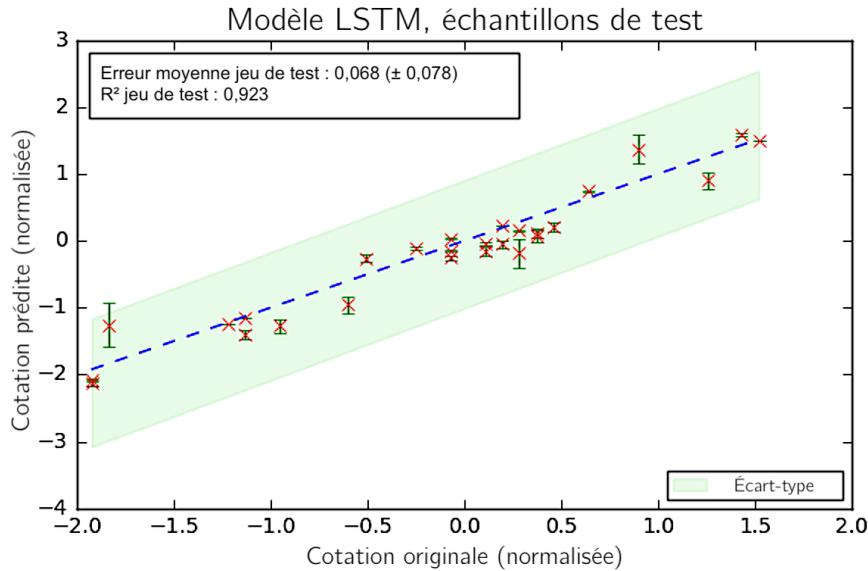


Figure 2.5 – Résultats du modèle *LSTM* sur le jeu de données de test.

Hertz, sans filtrage<sup>2</sup>. La Figure 2.4 présente ces signaux pour une pièce. Ce travail s'appuie sur une production de 204 pièces, où les paramètres de la machine ont été variés suivant un plan d'expérience de criblage, discuté dans le Chapitre 3.1 §3.1.1. Le modèle qui prédit au mieux la géométrie à partir des mesures s'appuie sur des réseaux de neurones récurrents *LSTM*. La Figure 2.5 présente les résultats normalisés obtenus par validation croisée §3.3.1.2, pour 27 pièces de la base de données de tests. L'erreur de prédiction est faible. Cela correspond à une erreur moyenne de prédiction de la dimension de 1,3 millimètres sur une dimension de 90 millimètres, soit une erreur relative de 1,54%. Ces résultats montrent l'intérêt de ce type de mesures indirectes des caractéristiques du produit. Cette erreur est inférieure aux attentes dimensionnelles de la norme ISO20457 [ISO18], qui exige entre 2% et 4% d'écart dimensionnelles. Toutefois, si elle est généralisée pour des pièces dont les dimensions font un mètre, une erreur relative de 1,54% est équivalente à une erreur de 1,54 centimètres. Aussi, l'erreur obtenue par le modèle *LSTM* est à la limite de faisabilité de l'employer comme moyen de contrôle prédictif de la cotation. Il faudrait une erreur dix fois moindre pour garantir la possibilité d'utilisation de ce modèle prédictif. Un jeu de données de pièces plus conséquent, ainsi qu'un modèle plus raffiné, sont des pistes pour améliorer ces performances.

Dans une démarche de surveillance de l'état du procédé d'injection-moulage, la mesure de variables sur la machine est importante. Une mise sous surveillance statistique MSP §1.1.3.3 permet de détecter les situations hors-contrôle et les défaillances du procédé. Nous invitons le lecteur qui s'intéresse aux mesures réalisées sur le procédé, plutôt que sur le produit, à consulter l'état de l'art récent et exhaustif de AGEYEVA, HORVÁTH et KOVÁCS [AHK19].

Dans notre objectif, nous cherchons à mesurer les caractéristiques d'aspects des pièces. À la différence de notre travail sur la prédiction d'une cotation géométrique, les caractéristiques d'aspect sont difficilement représentées par une unique variable continue. La représentation des caractéristiques d'aspect la plus commune est l'image numérique. Elle se compose de plusieurs milliers de valeurs entières, encodées sur 8 bits, entre 0 et 254. Nous n'avons pas réussi à construire un modèle qui prédit les caractéristiques de l'aspect des pièces à partir des seules mesures dans le moule. Aussi, nous pensons qu'il est nécessaire de capturer une image de l'aspect des pièces, dès leur sortie du moule, afin d'établir une mesure de l'aspect. La mesure directe est l'objet du

2. Acquisition réalisée à l'aide du système *PSP* de la société *SISE*.

paragraphe suivant.

### 2.2.2 Mesures directes des caractéristiques du produit

Notre approche consiste à mesurer les caractéristiques des pièces dès la sortie de la presse d'injection-moulage. La réalisation de ces mesures nécessite de convoyer la pièce devant le banc de mesure. Un tapis roulant ou un bras robotique permet de déplacer la pièce. La durée des déplacement est à prendre en compte, vis à vis de la durée du cycle et de la durée de la mesure.

De nombreuses méthodes ont été développées pour mesurer les propriétés mécaniques des pièces. En particulier, la mesure des contraintes résiduelles dans la pièce est un enjeu important. Le travail de doctorat de GIROUD mesure les contraintes internes d'une pièce à l'aide de la méthode de l'enlèvement de couches fines [Gir01]. C'est une méthode qui détruit la pièce et qui doit être réalisée en dehors de la ligne de production. La mesure des caractéristiques cristallographiques est importante car ce sont ces caractéristiques qui déterminent les propriétés mécaniques des pièces. Cette structure est dépendante des paramètres du procédé. MENDOZA et al. utilise par exemple la polarimétrie sur couche mince (micrographie) pour mettre en évidence les axes d'orientation du polymère semi-cristallin [Men+03]. Dans ces travaux de thèses, MALHAB introduit l'utilisation de la diffraction des rayons X à grands et petits angles, à l'aide d'un rayonnement synchrotron hautement énergétique (12 KeV) [Mal12]. La cohérence du faisceau synchrotron (30 micromètres) permet de mettre en évidence la structure cristallographique dans l'épaisseur des échantillons plastiques. Cela permet d'étudier l'évolution de la structure en fonction des paramètres du procédé d'injection. La diffraction à grands angles (*Wide-Angle X-ray Scattering*) permet de déterminer le degré de cristallinité du polymère semi-cristallin, en observant les figures de diffraction de la structure cristalline. La diffraction à petits angles permet de déterminer les dimensions nanométriques de la structure du polymère, telles que la longueur de la période de répétition du motif cristallin, la largeur et la longueur de la lamelle cristalline. En 2009, le projet européen *poliMATIC*<sup>3</sup> a proposé de robotiser l'utilisation de la microscopie confocal ([capteur Altimet](#) de la société Altisurf) pour contrôler la rugosité de pièces. La position du capteur est asservie en distance pour mesurer des pièces courbes. Un interféromètre portatif<sup>4</sup> a également été développé pour améliorer la résolution, et atteindre une précision inférieure au nanomètre [BR12]. La durée de la mesure est élevée, car le capteur mesure une surface de  $4 \times 4$  centimètres ; il doit être déplacé sur toute la surface. En 2010, le projet *PLASTX* du Pôle Européen de la Plasturgie<sup>5</sup> propose de réaliser une mesure par tomographie rayons X des pièces dès la sortie du moule<sup>6</sup>. La source de rayons X est fixe et un bras robotique déplace la pièce dans un mouvement de rotation. La tomographie permet de mesurer le champ de densité de la pièce. Les résultats du projet *PLASTX* montrent que cette mesure est compatible avec le temps de cycle industriel. Cependant, la mise en place d'une source de rayons X sur une ligne de production reste difficile en raison des blindages nécessaires. Dernièrement, dans ses travaux de doctorat, LACOMBE utilise un dôme de mesure de la réflectance (*Reflectance Transformation Imaging*) pour contrôler les caractéristiques d'aspect de pièces plastiques [Lac18]. Ce dôme de réflectance n'est actuellement pas intégré à la ligne de production, à cause de ses dimensions. Des travaux sont actuellement en cours pour proposer une méthode d'intégration en ligne de production.

Dans la suite de nos travaux, nous n'utiliserons pas de méthodes de mesure invasives, ni de mesures réalisées hors-ligne. L'étude de ces méthodes peuvent néanmoins servir d'inspiration pour concevoir des méthodes de mesure non invasives, sans destruction des pièces. Ainsi, nous nous intéresserons dans nos travaux à la polarimétrie. En revanche, nous n'étudierons pas l'utilisations

---

3. Projet de recherche de la Commission européenne FP7-NMP-2009-SME-3 : [automated-polishing.eu](http://automated-polishing.eu)

4. Interféromètre désormais commercialisé par la société [QiSab](#)

5. Le Pôle Européen de la Plasturgie est devenu le [Centre Technique Industriel de la Plasturgie et des Composites \(CTI\)](#) en 2017.

6. [Vidéo de présentation](#) du démonstrateur du projet PLASTX.

des rayonnements X et Térahertz, car les machines restent coûteuses ; mais nous pensons qu'ils seraient très intéressant de les mettre en œuvre en ligne de production.

Le Tableau 2.3 présente des mesures qui peuvent être réalisées sur la pièce. Afin d'évaluer la compatibilité du moyen de mesure avec le temps de cycle industriel, nous précisons la durée totale de la mesure en incluant la durée du déplacement de la pièce vers l'appareil de mesure. Plus l'appareil de mesure est volumineux et nécessite une précision de positionnement, plus cette durée sera longue. Les durées de mesure sont calculées pour une pièce de dimension  $10 \times 10$  centimètres. Les chiffres économiques sont indicatifs. Ils sont issues de notre analyse du marché, réalisée en 2019. Le coût par pièce correspond au coût d'utilisation d'une machine pour réaliser le contrôle à cent pourcent des pièces dès la sortie du moule. En particulier, le coût horaire est calculé pour un amortissement du coût d'achat de la machine de mesure de dix années et une production de 120 pièces par heures, 300 journée ouvrées annuelles, soit 3 millions de pièces par an. Dans le cas où un opérateur humain est requis, le coût chargé est intégré : 15€ par heures, en 2019. Dans le cas où la mesure est automatisé, le coût de l'installation inclut le dispositif de convoyage (bras robotique ou tapis roulant) en plus du dispositif de mesure.

Ce tableau permet de mettre en valeur les moyens compatibles avec les contraintes industrielles ; les contraintes majeures sont la durée de la mesure et le coût de la mesure. De plus, les productions industrielles doivent aujourd'hui être réactives aux changements ; on parle d'agilité. Aussi, dans une démarche d'évolution continue de la ligne de production, au fil des changements de séries, nous recommandons fortement l'utilisation d'un dispositif de mesure qui puisse être adapté rapidement à de nombreux types de pièces. C'est ici un avantage considérable des moyens de mesure sans contact physique avec la pièce. Les moyens de mesure légers et mobiles sont à recommander.

Nous sommes en présence de pièces chaudes ; aussi, la thermographie est un moyen de mesure intéressant du point de vue du coût et de la rapidité. Nous détaillons l'emploi de la thermographie dans la Section 2.3.2. La polarimétrie est une méthode d'imagerie non-conventionnelle qui est aussi intéressante, en particulier pour les pièces transparentes, mais aussi pour la mise en valeur des défauts de surfaces. Nous détaillons son emploi dans la Section 2.3.3.

De manière général, nous recommandons un coût de la mesure de la qualité inférieur à 1% du prix de revient des pièces. Pour des pièces plastiques dont le coût de production est proche de 10€, le coût de la mesure doit être inférieur à 10 centimes d'euros. À partir du Tableau 2.3, nous retenons les moyens de mesures suivants :

- plusieurs comparateurs instrumentés,
- profilomètre laser,
- scanner 3D §2.3.1,
- thermographie §2.3.2,
- imagerie couleur,
- polarimétrie §2.3.3.

Ce travail de doctorat a permis d'évaluer expérimentalement la faisabilité de ces mesures. Dans la section suivante, nous détaillerons leur utilisation.

### 2.2.3 Définition du degré d'invasivité d'un système de mesure pour un procédé

Afin de conclure notre étude sur les différents moyens de mesure, nous proposons de définir une échelle du degré d'invasivité d'un système de mesure, au regard du procédé industriel. La notion d'invasivité d'un moyen de mesure est lié à la quantité de travail nécessaire pour intégrer le moyen de mesure au procédé. Ces modifications sont effectuées sur le procédé, soit à posteriori de la fabrication du moyen de production, soit en amont pendant la conception du procédé. Il est intéressant de remarquer que le degré d'invasivité ne prend pas en compte le coût d'achat du

Caractéristique	Moyen de mesure	Résolution	Durée	Coût horaire	Coût initial	Coût / pièce
<b>Dimensions géométriques</b>						
1 dimension	Pied à coulisse + opérateur	50 µm	20 s	15 €/h	200 €	12,50 c€/p
1 dimension	Comparateur instrumenté	30 µm	10 s	0,20 €/h	5 000 €	0,16 c€/p
1 dimension	Comparateur laser	10 µm	5 s	0,40 €/h	10 000 €	0,33 c€/p
Profil	Profilomètre laser	1 µm	30 s	0,80 €/h	20 000 €	0,67 c€/p
Profil	Microscopie confocal	0,5 µm	60 s	4 €/h	100 000 €	3,33 c€/p
Surface	Photogrammétrie	50 µm	30 s	0,80 €/h	20 000 €	0,67 c€/p
Surface	Lumière structurée	30 µm	30 s	4 €/h	100 000 €	3,33 c€/p
Surface	Microscopie confocal	0,5 µm	600 s	40 €/h	100 000 €	3,33 c€/p
Volume	Profilomètre/tomographie	10 µm	60 s	2 €/h	50 000 €	0,16 c€/p
Volume	Scanner volumique	100 µm	30 s	2 €/h	50 000 €	0,16 c€/p
<b>Dimensions et densité</b>						
Densité plane	Radiographie	10 µm	20 s	12 €/h	300 000 €	9,99 c€/p
Densité volumique	Tomographie rayons X	10 µm	120 s	40 €/h	1 000 000 €	33,33 c€/p
<b>Champ de température</b>						
Champ °C	Thermographie	1000 µm	1 s	0,40 €/h	10 000 €	0,33 c€/p
Évolution du champ	Thermographie dynamique	1000 µm	120 s	1,20 €/h	30 000 €	1,00 c€/p
<b>Aspect visuel</b>						
Image couleur	Caméra couleur	200 µm	1 s	0,04 €/h	1 000 €	0,03 c€/p
Réflectance	Dôme de réflectance	50 µm	10 s	0,40 €/h	10 000 €	0,33 c€/p
<b>Propriétés du matériau</b>						
Champ de contraintes	Photoélasticité/polarimétrie	200 µm	1 s	0,10 €/h	3 000 €	0,10 c€/p
Résilience	Essai Charpi	/	60 s	15 €/h	10 000 €	12,50 c€/p
Dureté	Indentation	/	120 s	30 €/h	10 000 €	22,50 c€/p
Élasticité/Plasticité	Traction/Compression	/	120 s	30 €/h	20 000 €	22,50 c€/p
Composition chimique	Spectromètre (UV-VIS-NIR)	/	10 s	0,40 €/h	10 000 €	0,33 c€/p
Structure cristalline	Diffraction X/polarimétrie	/	120 s	30 €/h	20 000 €	22,50 c€/p

Tableau 2.3 – Moyens de mesure des caractéristiques d'une pièce.

Degré d'invasivité	Type de mesure	Description de la modification
0	Pas de mesure	Aucune modification sur le procédé.
1	Sans contact >10s	Positionnement d'un capteur vers la pièce.
2	Sans contact de la pièce >30s	Positionnement capteur et convoyage spécifique.
3	Avec contact de la pièce	Convoyage de la pièce pour le contact.
4	Mesure sur la machine	Intégration de capteurs sur la machine.
5	Mesure dans l'outillage	Intégration de capteurs dans l'outillage.
6	Avec manutention humaine	Intégration de l'humain dans la prod.
7	Mesure hors-ligne	Extraction de la pièce de la ligne de prod. pour mesurer.

Tableau 2.4 – Degré d'invasivité de la mesure sur le procédé d'injection-moulage des thermoplastiques.

moyen de mesure qui peut être amorti. En revanche, il prend en compte le coût de la modification du procédé, qui est un élément de décision important.

Le Tableau 2.4 présente notre définition du degré d'invasivité d'un moyen de mesure, pour un moyen de production.

**Degré 0** Atteindre le degré 0 nécessite de ne pas réaliser de mesure. Disposer d'une simulation informatique (modèle mathématique) du procédé est alors indispensable. Ce modèle doit être suffisamment représentatif du procédé pour prédire les caractéristiques des pièces réellement produites. En 2019, il n'existe pas à notre connaissance de modèle capable de prédire l'aspect des pièces.

**Degré 1-2** Les mesures ne nécessitent pas de contact. Nous classons les deux premiers degrés en fonction de la durée de la mesure.

**Degré 3** Le contact avec la pièce requiert une programmation précise du moyen de convoyage. Il nécessite l'utilisation d'un bras robotique, car un tapis n'est pas suffisamment précis.

**Degré 4-5** L'intégration de capteurs dans l'outillage entraîne un travail de conception supplémentaire et une maintenance spécifique.

**Degré 6** Une mesure qui nécessite une manutention humaine pour être réalisée est très invasive. En effet, l'humain est difficilement automatisable. De manière générale, une mesure industrielle doit être automatisée pour éviter que les erreurs et les retards n'impactent le flux de production.

**Degré 7** La mesure hors-ligne entraîne la perte de la pièce, en plus de toutes les contraintes précédentes. C'est ainsi la mesure la plus invasive possible pour un procédé de production industriel.

Dans notre travail, nous nous intéresserons à mettre en œuvre des moyens de mesures possédant une invasivité minimale ; ce qui correspond au degré 1 de notre échelle : on parlera de mesure non-invasive pour le procédé industriel. Le degré 0 n'est pas atteignable, à cause de la non-existence d'une simulation réaliste de la qualité des pièces produites.

#### 2.2.4 Corrélation entre les caractéristiques des pièces chaudes et des pièces finales

La mesure pendant le cycle du procédé est réalisée sur des pièces chaudes, qui viennent de sortir de l'outillage. Leur température est comprise entre 50°C et 150°C. Les caractéristiques de la pièce chaude ne sont pas celles de la pièce finale. En effet, le refroidissement de la pièce inclut la relaxation des contraintes internes à la pièce. Pendant le refroidissement de la pièce, les

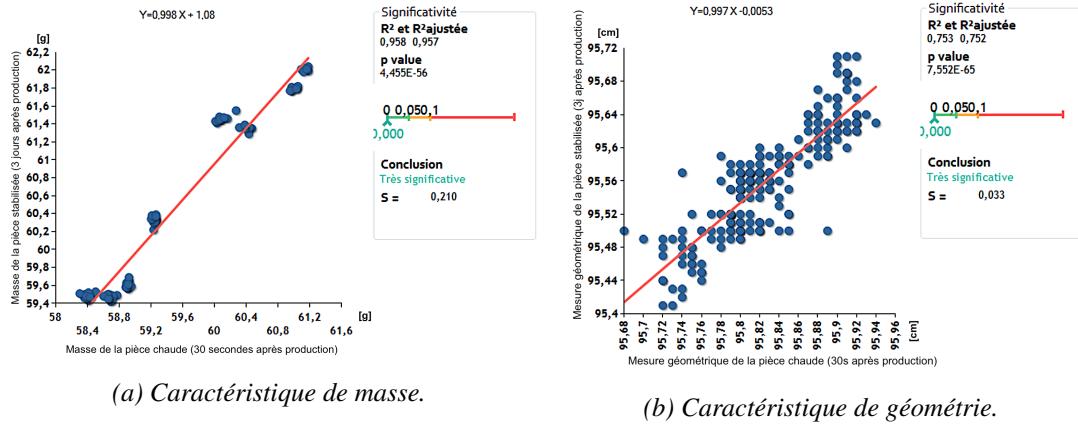


Figure 2.6 – Modèle linéaire entre les caractéristiques des pièces chaudes (30 secondes après production) et des pièces stabilisées (3 jours après production), Logiciel Elistat<sup>6</sup>.

contraintes internes s'équilibrivent. La géométrie de la pièce est modifiée : c'est le retrait. Le retrait est équivalent à une réduction des dimensions de la pièce. La réduction des dimensions varie de l'ordre de 0,01% à 1%, selon le matériau.

GALESKI et REGNIER étudient les liens entre les propriétés mécaniques macroscopiques et la structure cristallines des pièces moulées [GR09]. La structure cristalline d'une pièce varie en fonction de l'épaisseur. Cela dépend des conditions des phases de maintien et de refroidissement. Les propriétés mécaniques, dont les contraintes internes, dépendent des conditions d'injection et de la géométrie des pièces. Dans notre cas, les pièces chaudes ont déjà subit les phases de maintien et de refroidissement. Aussi, nous considérons que la stabilisation des contraintes internes est identique pour toutes les pièces d'un même lot. Cette hypothèse est forte. Il est probable que les conditions d'hygrométrie et de température lors du stockage des pièces impactent la stabilisation des contraintes internes, et donc modifient le retrait dimensionnel. Nous avons stocké toutes nos productions d'études dans les mêmes conditions, c'est pourquoi nous n'avons pas observé de différences entre les lots. Aussi dans notre cas, la réduction des dimensions (le retrait) peut être approximée par un modèle linéaire.

Nous avons étudié l'évolution des caractéristiques sur une production de 204 pièces. Cette étude a été réalisée pour un unique matériau et une unique pièce. Afin d'étudier l'ensemble de la plage de fonctionnement du procédé d'injection-moulage, nous avons ajusté les réglages de la machine selon une plan d'expériences de criblage. L'organisation de cet essai expérimental est détaillé au Chapitre 3.1 §3.1.1. L'étude propose la mesure d'une cotation géométrique et de la masse des pièces, 30 secondes après la production et 3 jours après. Les pièces mesurées 3 jours après la production sont considérées comme refroidies et leur géométrie stabilisée. Un modèle linéaire permet de modéliser le phénomène de retrait avec une justesse  $R^2 = 0,75$  et une significativité statistique élevée, au sens de la *valeur-p* inférieure à 0,05. La Figure ?? présente les résultats obtenus. Il est intéressant de noter que la masse des pièces réduit de 0,2% lors de la relaxation. Il pourrait s'agir d'une perte de composés organiques volatils.

De plus, les caractéristiques d'aspect des pièces sont très proches, entre les deux instants de mesures. Les pièces qui possèdent des givrures ou des marques de brûlure les conservent. Enfin, nous avons observé que le retrait dimensionnel et la relaxation des contraintes internes, lors du refroidissement, entraînent une diminution des défauts visuels d'aspect de type retassure. Ces défauts sont causés par la courbure géométrique des surfaces ; ce sont des défauts de planéité. La Figure 2.7 présente des photographies d'une même pièce, avant et après le refroidissement.

6. Site Internet de la société Elistat.



(a) 10 secondes après production.

(b) 3 jours après production.

*Figure 2.7 – Photographie de l'aspect des pièces.*

On observe que la planéité du fond de la boîte (défaut de retassure, "tâche") est visuellement similaire entre la pièce chaude et la pièce stabilisée. On observe également l'effet de la relaxation des contraintes internes, ainsi que du retrait : les parois de la boîte sont incurvées après le refroidissement. En revanche, lors du refroidissement, les défauts de planéité diminuent, car la géométrie de la pièce se stabilise sur une position moyenne. C'est pourquoi, les défauts d'aspect de retassure sont plus prononcés quelques secondes après la production des pièces, que sur des pièces stabilisées. L'évolution de la pièce lors du refroidissement est simple à modéliser. Les caractéristiques des pièces finales sont proches des caractéristiques des pièces chaudes. Notre démarche de mesure des caractéristiques du produit, dès la sortie de la machine, s'appuie sur ce constat.

## 2.3 Apport de l'imagerie non-conventionnelle pour la mesure de la qualité

Dans cette section, nous discuterons de l'intérêt de différentes méthodes d'imageries non conventionnelles pour mesurer les caractéristiques des pièces. Nous présenterons nos travaux expérimentaux afin de valider l'intérêt de mesures par imagerie non-conventionnelle : triangulation laser §2.3.1, thermographie §2.3.2 et polarimétrie §2.3.3.

L'imagerie conventionnelle concerne l'acquisition et le traitement de signaux sous forme d'images, dans le domaine spectrale du visible (longueur d'ondes dans le vide de 380 à 780 nanomètres). À contrario, l'imagerie non-conventionnelle concerne les domaines spectraux qui ne sont pas visibles, ainsi que les méthodes d'acquisition et de traitement qui ne forment pas une image. En particulier, nous discuterons de l'imagerie dans les spectres infrarouges intermédiaires et lointains (3 000 à 15 000 nanomètres), qui permettent de mesurer l'émission thermique des pièces. La faisabilité industrielle de la mesure géométrique par triangulation laser (scanner tridimensionnel) sera discutée. Nous discuterons également de l'acquisition d'images polarisées des pièces. Enfin, nous discuterons des méthodes de fusions et d'exploitation des données de ces

capteurs, afin de qualifier les caractéristiques des pièces.

### 2.3.1 Mesure géométrique par triangulation laser

Il existe un grand nombre de moyens de mesure pouvant être qualifiés de scanner tridimensionnel. Nous distinguons les méthodes avec contact physique de la pièce (Machines de Mesure Tridimensionnelle) et sans contact (utilisant la réflexion d'une onde électromagnétique). Dans le cadre de la mesure sans contact, les moyens les plus employés sont la stéréoscopie, la photogrammétrie, la lumière structurée, la tomographie, temps de vol, la triangulation laser.

La triangulation laser, communément appelée "scanner 3D", s'appuie sur la mesure d'un écart de parallaxe entre une source laser balayée linéairement par un galvanomètre, et l'image de la réflexion du laser sur l'objet à mesurer. Une source lumineuse laser est choisie car la largeur du faisceau lumineux sur la pièce doit être la plus petite possible. Le faisceau laser cohérent garantit une dispersion angulaire faible à grande distance.

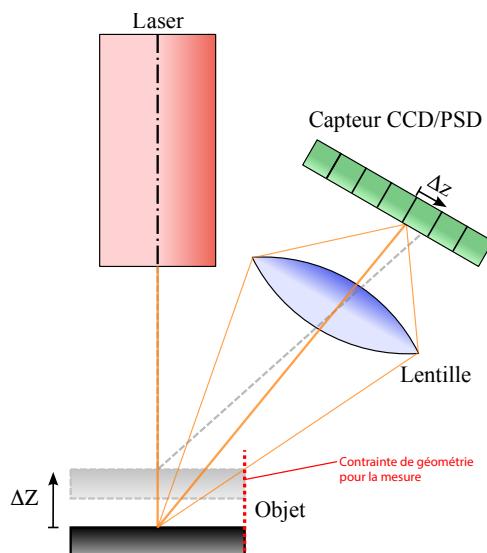


Figure 2.8 – Mesure de la géométrie d'un objet par triangulation laser. Figure dérivée du travail original de Georg Wiora (Dr. Schorsch), CC BY-SA 3.0, via Wikimedia Commons.

La Figure 2.8 présente le principe de la mesure par triangulation. Un écart géométrique vertical  $\Delta Z$  entraîne un écart proportionnel, sur le plan du capteur. La méthode de triangulation laser possède quatre contraintes principales :

- L'intensité lumineuse du laser doit être supérieure à la luminosité ambiante. C'est pourquoi la longueur d'onde du laser est souvent choisie hors du spectre visible.
- La fréquence de balayage du laser doit être très supérieure au déplacement de la pièce. Cela conditionne la fréquence d'acquisition du capteur.
- Le matériau qui est mesuré doit avoir un absorption faible pour la longueur d'onde du laser et les réflexions spéculaires directes doivent être évitées.
- La géométrie de la pièce doit permettre la mesure : elle ne doit pas former d'occlusion pour le couple source laser-capteur. L'écart angulaire de ce couple est souvent proche de 45°. C'est pourquoi les géométries aigus doivent être supérieure à 45°.

La Figure 2.9 présente notre mise en place cette mesure, en cycle industriel. Le scanner tridimensionnel que nous utilisons est le *Handyscan 3D ZSCANNER 700*<sup>7</sup>. Nous avons évalué la

7. Caractéristiques techniques du scanner Handyscan 3D ZSCANNER 700 sur le site du fabricant Créaform.

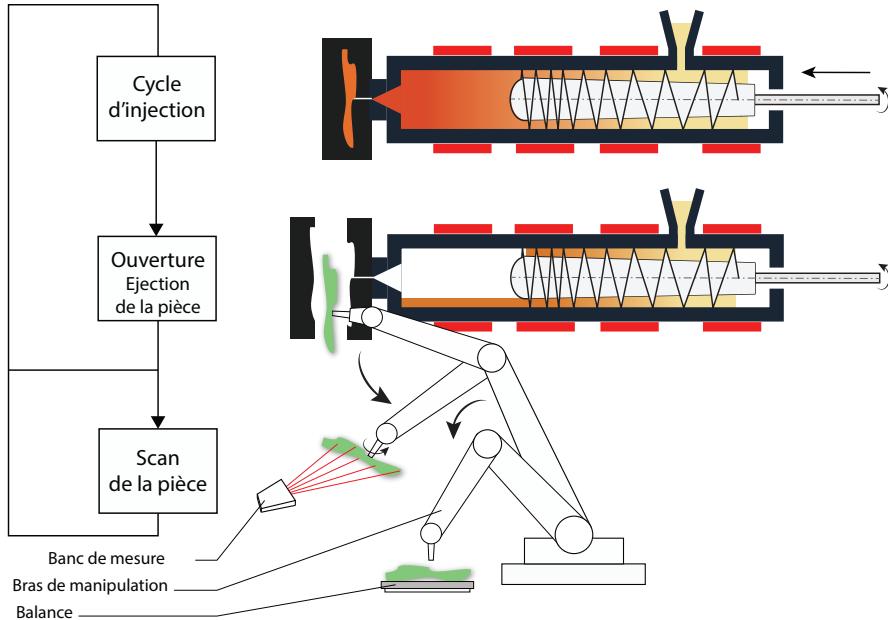


Figure 2.9 – Mesure de la géométrie d'un objet par triangulation laser en cycle industriel.

répétabilité de ce scanner à 100 micromètres. Une pièce étalon a été mesurée sur une machine de mesures tridimensionnelle, puis elle est mesurée à l'aide du scanner. Les résultats ont été comparés à l'aide du logiciel GéoVérif [Pai16]. La durée de la mesure est de 40 secondes. Le scanner est fixe et une rotation complète de la pièce est exécutée devant le scanner, à l'aide d'un bras robotique, Figure 2.9. La vitesse de cette rotation est imposée par la durée du temps de cycle de la machine. Un temps de cycle court impose de réaliser des déplacements rapides devant le scanner.

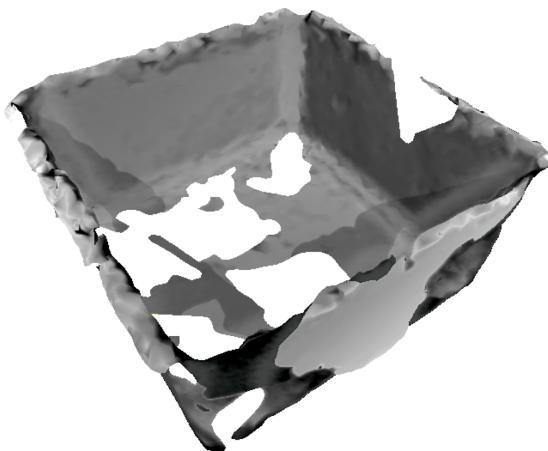


Figure 2.10 – Nuage de points de la pièce obtenu en cycle industriel.

Dans notre cas d'application, la pièce mesurée possède des cotés qui occultent la réflexion. Il est nécessaire de réaliser des rotations sous différents angles pour mesurer toute les surfaces de la pièce, ce qui nécessiterait plus d'un tour complet. De plus, la vitesse de déplacement que nous avons dû utiliser est trop élevée en comparaison de la fréquence de balayage du laser. La Figure 2.10 présente le résultat de la mesure. On observe la présence de trous dans le maillage du nuage de points. D'une pièce à l'autre, les trous n'apparaissent pas aux mêmes localisations. Aussi, l'interprétation d'une telle mesure est compliquée. Dans la suite de nos travaux, nous avons choisis de ne pas exploiter cette mesure à cause des performances non-adaptées de notre matériel.

L'étude de cette démarche de mesure devra être évaluée à l'aide d'un profilomètre laser ou d'un scanner tridimensionnel qui possède une fréquence de balayage laser plus rapide, afin de limiter les trous dans le nuage de points.

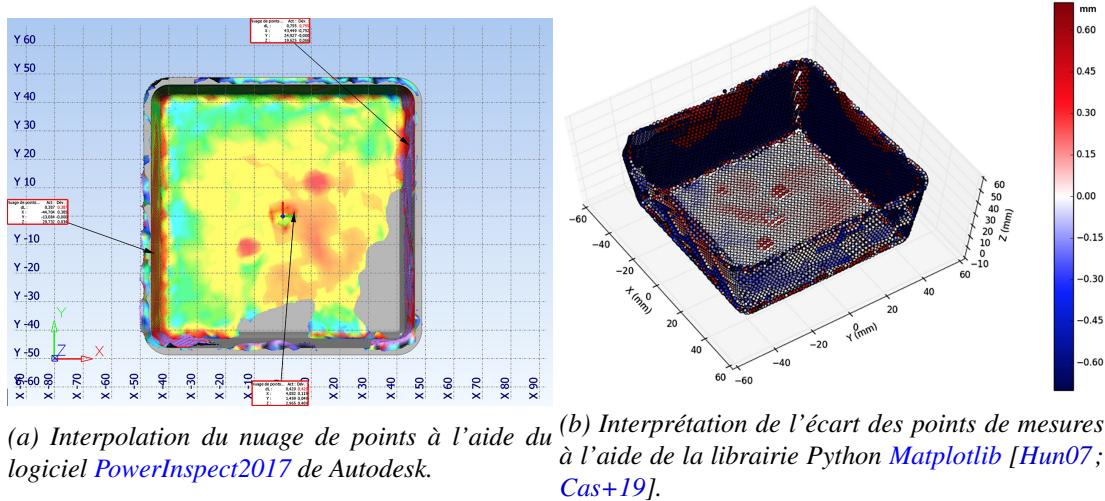


Figure 2.11 – Écart géométrique entre la pièce mesurée et la géométrie de référence de l’outillage.

La mesure d'un maillage de points de la pièce dès sa production permet de comparer la géométrie produite à la géométrie de référence de l'outil. La Figure 2.11 présente l'écart de la pièce chaude à la géométrie de référence. Il est ainsi possible de caractériser le phénomène de retrait. Ce travail de doctorat n'a pas cherché à étudier le retrait. Nous supposons que le retrait est constant pour toutes les pièces d'un même lot. C'est néanmoins une information importante à exploiter pour la conception des outillages. Enfin, cette information est également très intéressante pour ajuster les paramètres du procédé afin d'obtenir une géométrie cible.

En comparaison de l'imagerie thermique et polarimétrique, le coût du scanner tridimensionnel et la difficulté d'intégration de la mesure en ligne de production, sont bien plus importants. De plus, la répétabilité du scanner de 100 micromètres, rend difficile l'utilisation des mesures obtenues. Les déformations géométriques que nous observons sont de l'ordre de 100 micromètres. Les phénomènes de trous aléatoires dans le nuage de points compliquent l'interprétation automatique des mesures. Un traitement manuel des nuages de points est souvent requis. Or nous souhaitons automatiser la mesure de la qualité des produits. Pour toutes ces raisons, dans la suite de ce travail, nous avons choisi de ne pas exploiter le scanner tridimensionnel. L'utilisation d'un tel moyen de mesure reste une piste de recherche très intéressante.

### 2.3.2 Apport de la thermographie pour le contrôle de la géométrie des pièces

L'imagerie thermographie d'un objet capture son émission thermique dans les domaines de l'infrarouge moyen et lointain (longueur d'onde de 3 000 à 15 000 nanomètres). Les caméras thermiques les plus courantes utilisent un grand nombre de capteurs de type micro-bolomètres non-refroidis (une image de  $320 \times 240$  pixels utilise 76 800 micro-bolomètres). Un micro-bolomètre est composé d'un matériau semi-conducteur thermo-sensible. Lorsqu'il est soumis à un rayonnement infrarouge, le matériau chauffe et la résistance électrique du matériau qui compose le pixel varie. Les micro-bolomètres des caméras thermiques FLIR que nous utilisons sont précis à  $0,05^\circ\text{C}$  en moyenne à une température ambiante de  $30^\circ\text{C}$ . Cependant, la mesure de température nécessite de réaliser une calibration rigoureuse de l'évolution de la sensibilité du capteur à une émission thermique. Nous choisissons de nous affranchir de cette calibration et d'analyser les données brutes issues du capteur.

Une méthode très répandue de contrôle non-destructif de l'intégrité des pièces mécaniques

est la thermographie dynamique (thermographie active). La norme ISO18434 [ISO08] spécifie son utilisation. Il s'agit de chauffer localement la pièce à l'aide d'une source de chaleur dirigée (par exemple un laser), puis de mesurer précisément l'évolution de la température de refroidissement. Cette mesure dure quelques secondes. Elle permet de mettre en évidence les différences de capacités calorifiques ; ce qui indique en particulier les changements de densité causés par des défauts structurels. Dans ses travaux de doctorat, LEGRAND propose une méthode de thermographie multi-spectrale afin de s'affranchir de la connaissance de l'émissivité de la surface contrôlée [Leg02]. De nombreuses applications industriels de la thermographie dynamique mettent en évidence l'intérêt de la méthode pour le contrôle des pièces et des assemblages. Dernièrement, HERRMANN, MIGNOT et AUBRETON propose d'utiliser cette méthode afin de contrôler rapidement des pièces transparentes [HMA19] ; c'est la technique du *Scanning From Heating*, proposée pour la première fois par EREN et al. [Ere+09].

Dans notre cas d'application, les pièces sont déjà chaudes. C'est pourquoi la mesure thermique a un fort intérêt. L'utilisation de l'imagerie thermique pour le suivi des procédés de production se résume, aujourd'hui, au suivi de l'évolution de la température de quelques régions d'intérêts, qui ont été préalablement prédéfinies.

Dans le cadre de notre travail, nous ne chercherons pas à mesurer une température absolue. Nous n'avons pas non plus étudier son évolution (nous discutons de cette perspective de recherche à la fin de ce chapitre §??). Nous souhaitons obtenir le champ de températures sur la surface de la pièce. Ainsi, la connaissance de l'émissivité du matériau n'est pas nécessaire. La principale difficulté de notre approche est l'analyse pertinente de l'ensemble des pixels de l'images. Il s'agit d'extraire une information pertinente à propos des caractéristiques de la pièces, à partir du champ de température.

### 2.3.2.1 Mesures expérimentales

Nous avons étudié l'intérêt de la mesure thermique sur une production de 204 pièces. Les paramètres de la machine ont été variés suivant un plan d'expérience de criblage, discuté dans le Chapitre 3.1 §3.1.1. Nous avons réalisé une mesure par imagerie thermique des pièces chaudes, 10 secondes après la production. La caméra thermique utilisée était sensible sur une plage de longueur d'ondes de 7,5 à 13 micromètres. Le matériau utilisé est un polypropylène automobile, au degré de cristallinité supérieur à 70%. Nous n'avons pas caractérisé l'émissivité thermique du polypropylène utilisé. Une émissivité  $\epsilon = 0,97$ , "classique" pour le polypropylène, est utilisée pour calculer la température de surface, dans les figures suivantes.

L'instant de mesure a été rendu répétable par l'utilisation d'un bras robotique et la synchronisation temporelle de la mesure. 3 jours après la production, nous avons mesuré la surface qui a été imagée à l'aide d'une Machine de Mesures Tridimensionnel (MMT). 1000 points de mesures ont été réalisés à l'aide d'un capteur laser temps de vol. La Figure 2.12 présente le résultats des deux mesures réalisées :thermique et géométrique. Une soustraction des deux mesures montre qu'il n'y a pas de correspondance directe entre la géométrie et l'image thermique. Il semble exister une relation entre le point chaud central de la pièce (buse d'injection) et l'inflexion géométrique du plan.

### 2.3.2.2 Extraction de la géométrie à partir de l'image thermique

Dans l'étude "Generative Adversarial Networks for Geometric Surfaces Prediction in Injection Molding", nous avons cherché à déterminer si une information sur la géométrie de la pièce pouvait être obtenue par la thermographie [Nag+18]. La technique du *Scanning From Heating* permet par exemple de mesurer la géométrie à partir d'une image thermique. Cette technique suit le même mode opératoire que la triangulation laser. Elle nécessite le même investissement, la même complexité, et une source laser plus puissante pour chauffer la pièce. Dans notre étude, nous avons évalué la faisabilité de l'utilisation d'une unique image thermique, pour obtenir la géométrie de la

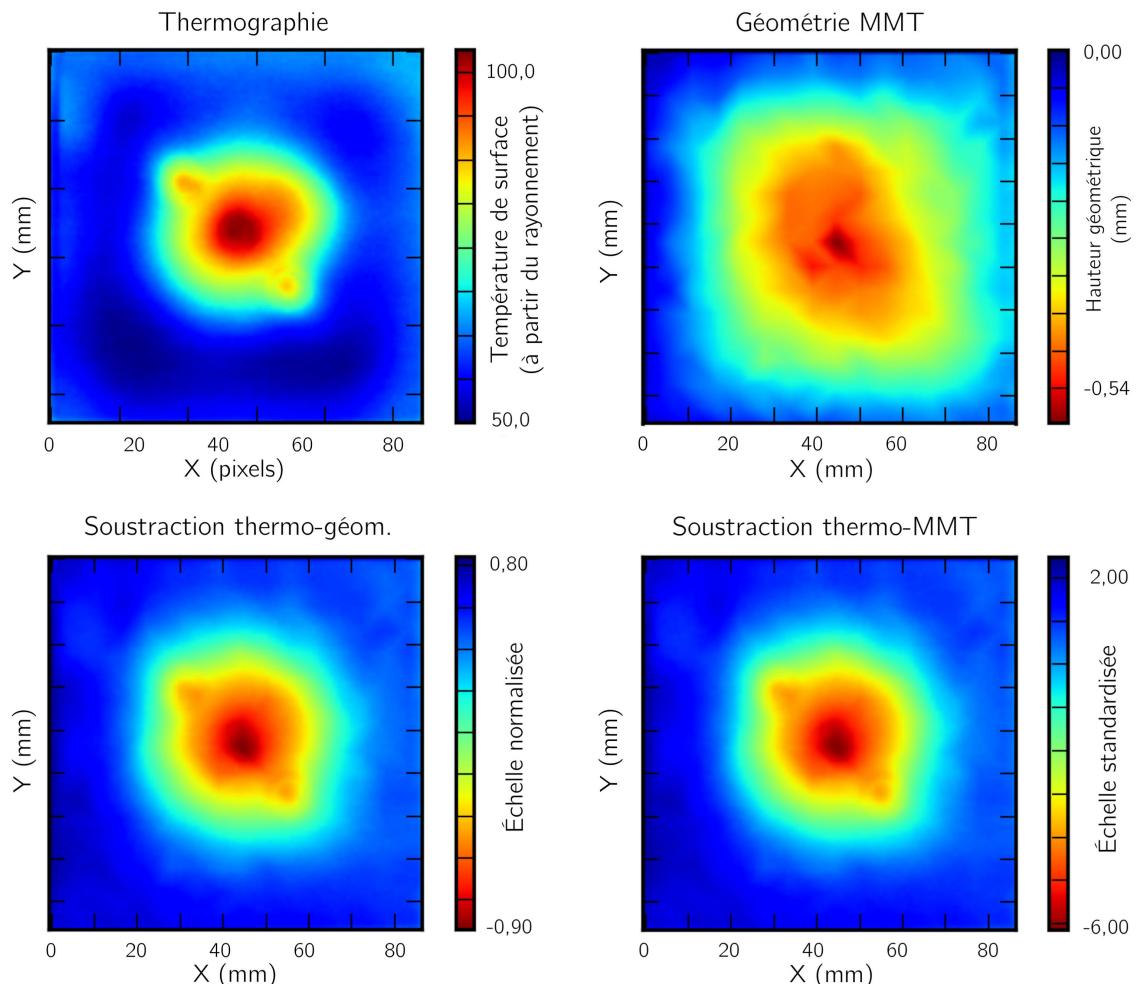


Figure 2.12 – Mesure thermique et mesure géométrique d'une pièce thermo-moulée.

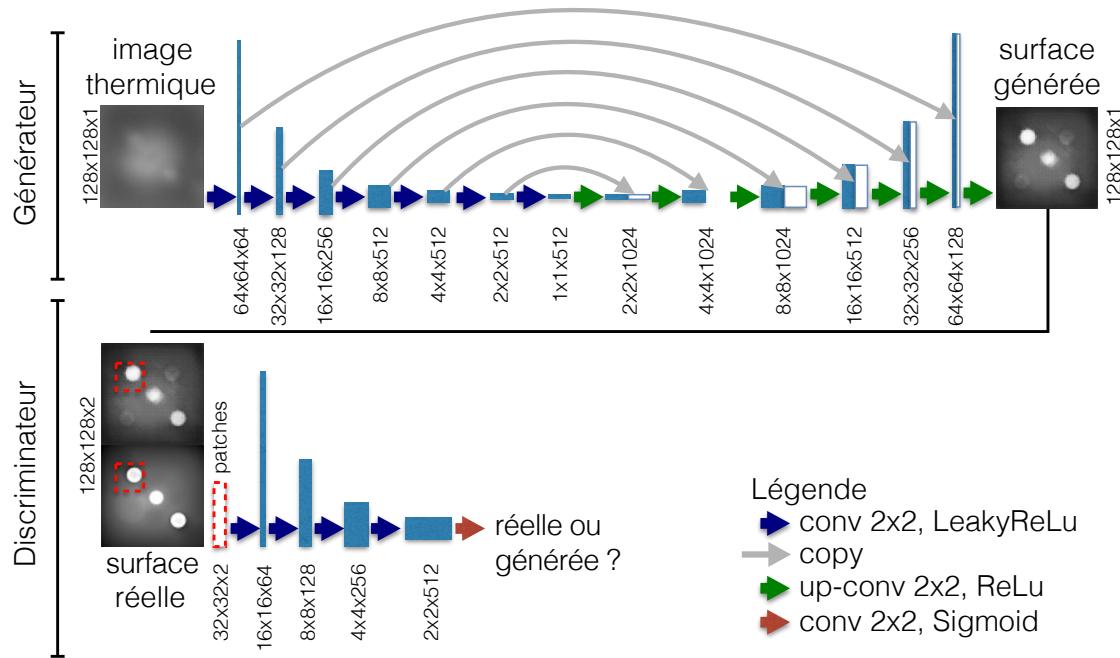


Figure 2.13 – Architecture du réseau antagoniste génératif U-Net, utilisée par le modèle pix2pix.

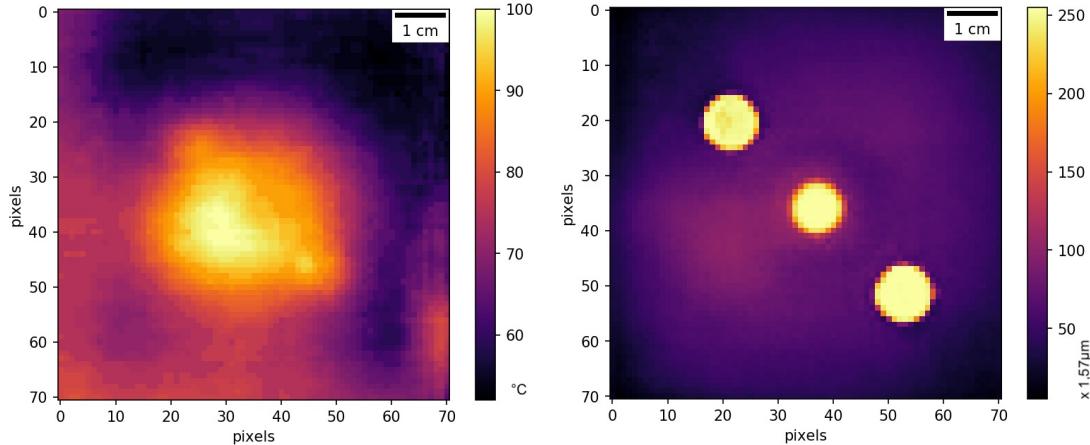
pièce. Sous réserve qu'une information sur la géométrie soit présente dans l'image thermique, il s'agit alors de construire un modèle qui relie l'image thermique et la géométrie.

Nous avons évalué que l'utilisation de modèles polynomiaux ne permet pas de modéliser la relation entre une image thermique et une image géométrique. Il est nécessaire de disposer d'un modèle qui est adapté au traitement d'images. Les récentes avancées des réseaux de neurones profonds pour la transformation d'images nous ont poussé à évaluer l'utilisation de réseaux génératifs antagonistes (GAN, détaillé dans le Chapitre 3 §3.3.4.8). Nous avons utilisé la méthode *pix2pix*, proposée par ZHU et al. [Zhu+17]. La Figure 2.13 présente l'architecture du réseau de neurones *U-Net*, qui est proposée par RONNEBERGER, FISCHER et BROX [RFB15]. Un réseau extrait l'information pertinente dans un vecteur de petites dimensions, appelé espace latent. Le réseau génératif produit ensuite une image réaliste à partir de cette espace latent. L'espace latent contient uniquement l'information pertinente pour reconstruire la géométrie de la pièce.

Le jeu de données de notre étude est composé de couples d'images thermiques et d'images géométriques, présenté Figure 2.14. La valeur des pixels d'une image "géométrique" encode la mesure de hauteur de la surface. La géométrie de la surface de la pièce est obtenue à l'aide d'un microscope confocal Altisurf 520<sup>8</sup>, de la société Altimet. Le réseau de neurones que nous utilisons est spécialisé dans le traitement d'images numériques de 8 bits de profondeur. Aussi, les images que nous utilisons sont normalisées sur 8 bits, soit sur une plage de valeurs entières de 0 à 255. La perte de résolution de cette conversion est importante : le capteur thermique est précis à 0,05°C, le microscope confocal à 0,1 micromètres. L'image thermique que nous utilisions a une résolution de 0,902°C, tandis que la carte d'élévation de la surface a une résolution de 1,57 micromètres. Une perspective à nos travaux sera d'utiliser un réseau de convolutions qui travaille avec des valeurs flottantes. La durée de l'apprentissage sera cependant nettement augmentée.

La durée de l'apprentissage du modèle est de 3 heures. Les poids de l'ensemble du réseau de neurones sont ajustés par apprentissage sur un jeu de données (méthode détaillée dans le Chapitre 3 §3.3.1.3). Le jeu de données comporte 37 pièces. Le nombre d'échantillons est très faible car la durée de la mesure d'une pièce à l'aide du microscope confocal était de 45 minutes.

8. Spécifications techniques de l'Altisurf 520 sur le site Internet de la société Altimet.



(a) Image thermique de la pièce 10s après prod. (b) Image de l élévation de la surface de la pièce.

Figure 2.14 – Présentation d'un échantillon du jeu de données, en fausse couleur.

Mesure de similarité	Moyenne	Écart-type
Distance cosinus	0,9	0,10
Corrélation $R^2$	0,8	0,15
PSNR	18	3,95
SSIM	0,86	0,07
Distance cosinus sur DMD	0,99	0,08
Corrélation $R^2$ sur DMD	0,98	0,04

Tableau 2.5 – Performance du modèle pix2pix : similarité entre images générées et images originales.

Une perspective directe est de réaliser cette étude sur un plus grand nombre d'échantillons et une variété de pièces.

Par la suite, nous évaluons la capacité du modèle appris à transformer une image thermique en image géométrique. La Figure 2.15 présente les résultats obtenus sur un jeu de données de test, qui n'a pas été utilisé pendant l'apprentissage. Afin de mesurer la similarité entre les images originales et les images générées, nous calculons les indicateurs les plus utilisés dans la littérature de la compression d'images. Nous n'avons pas réalisé de comparaison pixel à pixel, car les images générées peuvent être légèrement translatées en comparaison des images originales. Nous calculons les mesures de Similarité Structurelle SSIM proposée par WANG et al. [Wan+04] et Peak Signal to Noise Ratio pour la qualité de reproduction d'une image. La comparaison d'histogrammes est robuste aux transformations des images. Enfin, nous introduisons la Décomposition Modale Discrète comme méthode de comparaison de la similarité entre deux images. La Décomposition Modale Discrète décrit une image à partir de ses modes de vibrations spatiaux. C'est une méthode de décomposition qui produit un spectre représentatif de la géométrie, car les modes de vibrations sont ceux de plaques physiques. Nous détaillons le principe de la Décomposition Modale Discrète dans la Section 2.4.3.2. Le Tableau 2.5 présente les résultats de cette comparaison.

La similarité moyenne sur les 14 pièces du jeu de données de test est élevée et l'écart-type global est faible. Le jeu de données étant très petit, nous avons vérifié que la p-valeur est inférieure à 5%, pour toutes les mesures de similarité. La méthode pix2pix permet d'extraire de l'image

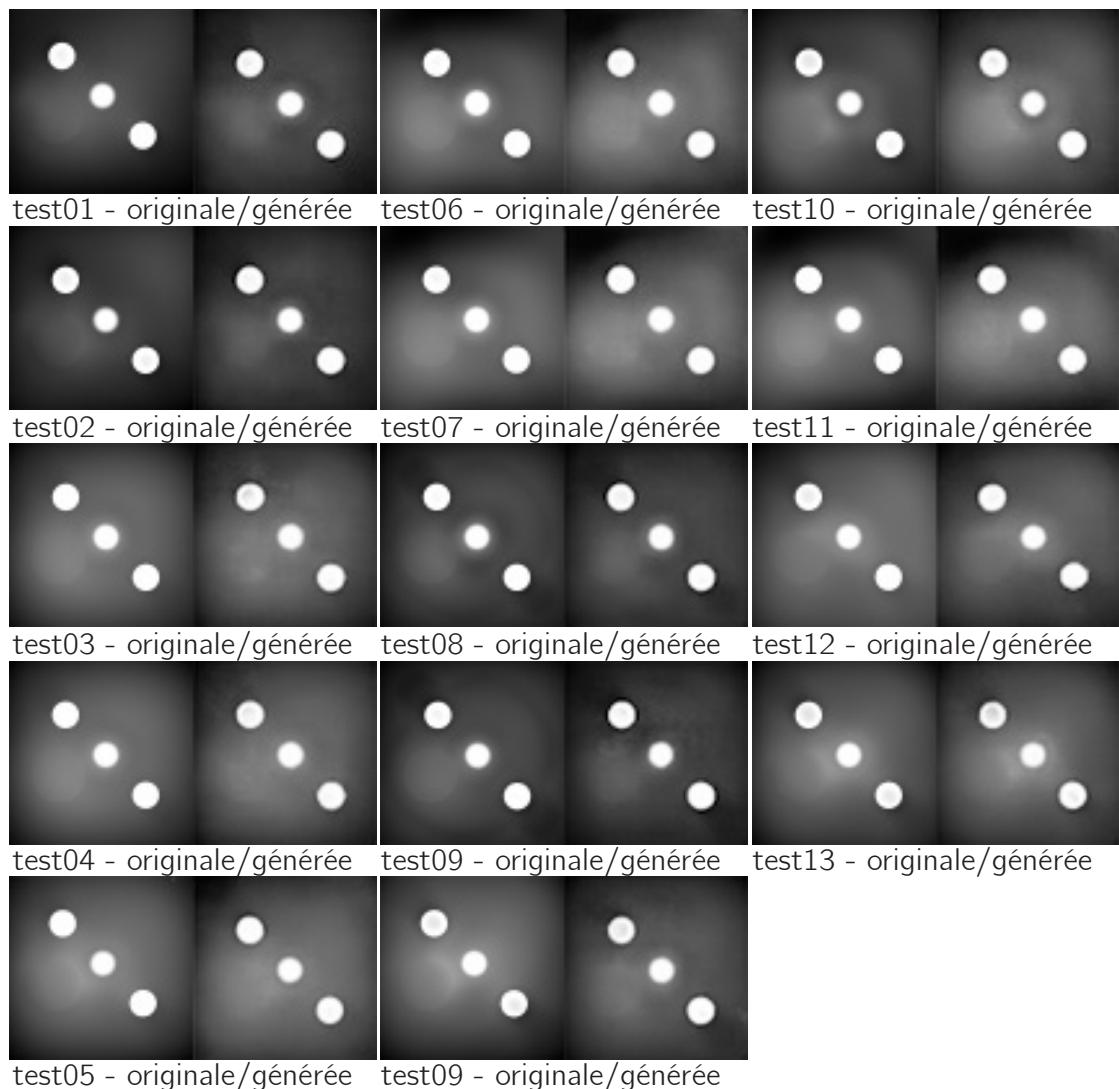


Figure 2.15 – Comparaison entre les images générées par le modèle à partir de l'image thermique et les images "géométriques" originales.

thermique de l'information sur la géométrie des pièces.

Bien que ces résultats soient encourageants, la durée de la phase d'apprentissage du modèle est de 3 heures sur une carte graphique puissante (*Nvidia GTX 1080Ti*). Il n'est pour l'instant pas envisageable de construire ces modèles en ligne de production. Il semble possible de pouvoir mettre à jour ce type de modèle à une fréquence journalière. Enfin, la robustesse de ce modèle fortement non-linéaire au perturbation est également difficile à évaluer, faute de quantité de données suffisantes et de cas d'études. Dernièrement, WU et al. montre que les réseaux génératifs antagonistes permettent de reconstruisent des représentations volumiques d'objet à partir d'une simple image [Wu+16]. La précision de la représentation de ces modèles est faible, car l'objectif est de représenter une grande diversité d'objets. Spécialiser ces modèles sur une unique pièce, pour augmenter la précision de la représentation volumique est aussi une perspective de recherche intéressante.

Dans notre recherche d'une méthode de contrôle des caractéristiques de la pièces, il reste à interpréter cette géométrie prédite. Il s'agit de comparer la géométrie obtenue à la géométrie de référence. Différent critère de distance géométrique existe. Nous retiendrons l'apport du critère inertiel total, qui s'appuie sur le tolérancement inertiel généralisé aux formes tridimensionnelles par ADRAGNA, SAMPER et PILLET [ASP10]. Si l'écart inertiel à la géométrie de référence de la pièce est grand, une alerte de non-qualité peut être déclenchée. Cependant, cette démarche nécessite de disposer de la géométrie de référence sous un format informatisé. Cette information de référence est rarement disponible en injection-moulage ; en particulier car les outillages sont souvent retouchés au cours de leur vie, sans forcement numériser les retouches effectuées. Nous souhaitons proposer une méthode de contrôle de la qualité des pièces la plus simplifiée possible, c'est pourquoi nous choisissons pour la suite de notre travail de ne pas utiliser de modèle prédictif de la géométrie. En revanche, nous utiliserons la démarche d'extraction de l'information à l'aide de réseau de neurones profonds, qui a ici montré des possibilités intéressantes.

### 2.3.3 Apport de la polarimétrie pour le contrôle des défauts d'aspect

La mesure du Degré de POlarisation Linéaire (DOLP) est utilisée dans de nombreux domaines techniques, dont le contrôle de la qualité de pièces composites<sup>9</sup> et de bouteilles en verre [AE18]. Dans le cadre du moulage de polymères transparents, l'imagerie polarimétrique est couramment utilisée pour visualiser les contraintes mécaniques. Les polymères transparents ont des propriétés de biréfringence qui sont induites par l'organisation interne des chaînes des polymères [DVA95]. Les contraintes mécaniques appliquées à la pièce modifieront la biréfringence locale du matériau : c'est le phénomène optique de photoélasticité [Bre33]. Il est ainsi possible d'observer les contraintes mécaniques qui s'exercent. Cependant, il n'existe pas à notre connaissance de travaux dans le domaine de la détection des défauts sur les pièces moulées en plastique. C'est l'objet de l'étude présentée dans cette section.

#### 2.3.3.1 Polarimétrie pour mettre en valeur les défauts des pièces plastiques

Les réflexions de la lumière à la surface d'un matériau peuvent être décomposées en une composante spéculaire et une composante diffuse [HK93]. La composante diffuse est causée par le phénomène de transluminescence (*subsurface scattering*). La lumière entre dans le matériau ; où elle est absorbée et dispersée ; avant de sortir du matériau. La lumière est absorbée et diffusée en fonction de sa longueur d'onde et de la couleur du matériau. De plus, le rayon lumineux est dispersé plusieurs fois. La lumière devient isotrope et la direction de la sortie du rayon est aléatoire.

---

9. Le Fraunhofer Institute for Integrated Circuits IIS a présenté la solution de polarimétrie POLKA pour le contrôle de pièces composites, lors du salon [Control Stuttgart 2018](#).

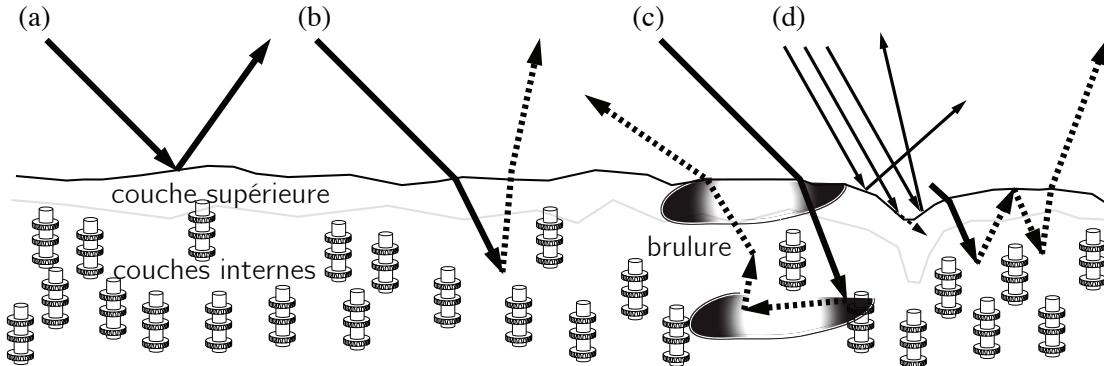


Figure 2.16 – Différentes composantes de réflexion d'un polymère semi-cristallin. Figure inspirée de [Deb+00].

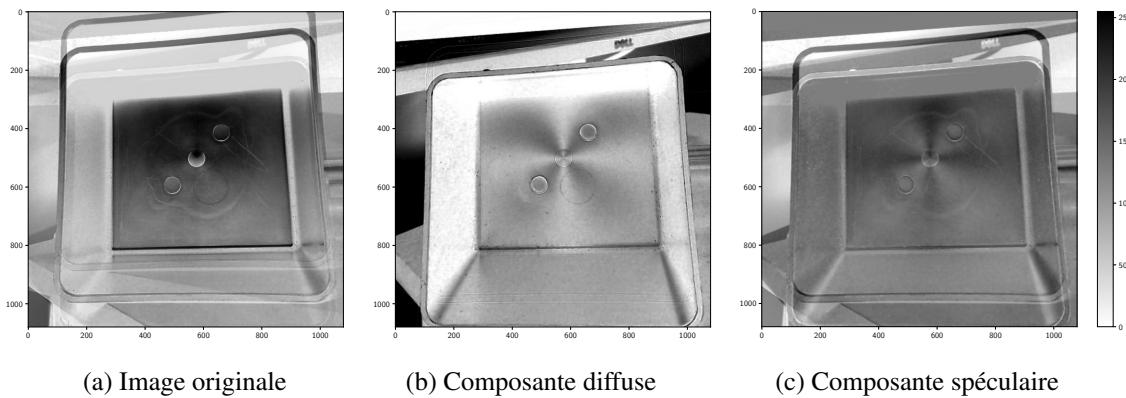


Figure 2.17 – Acquisition des composantes de réflexions spéculaire et diffuse.

À partir de ces connaissances physiques, nous cherchons à utiliser la mesure de la polarisation pour la détection de défauts sur les pièces plastiques. La Figure 2.16 présente l'effet de différents défauts sur la polarisation de la lumière réfléchie. Dans le cadre des matériaux semi-cristallins, nous posons deux hypothèses importantes pour la polarisation des réflexions :

- Un défaut géométrique (forme, retassure, rayure) modifie la courbure de la surface, ainsi il modifie l'angle de polarisation de la réflexion spéculaire : réflexion (d) Fig. 2.16.
- Un défaut qui modifie la structure de la matière (brûlure, givrage, pollution) diffuse la lumière incidente, ainsi la lumière que le défaut réfléchie est diffuse ; elle ne possède pas de composante spéculaire : réflexion (c) Fig. 2.16.

Il n'existe pas à notre connaissance d'étude qui concernent l'utilisation de la polarimétrie pour le contrôle des défauts sur des pièces plastiques. Dans la suite de cette section, nous évaluerons l'intérêt de la polarimétrie pour la mise en valeur des défauts, sur un cas d'application industriel. Si nos hypothèses sont vérifiées, cela permettra de mettre en valeur séparément les défauts géométriques à l'aide de la composante spéculaire et les défauts de la structure de la matière à l'aide de la composante diffuse. Pour des questions de coût de l'équipement, nous utiliserons un capteur polarimétrique que nous avons conçu.

La mesure des composantes de réflexions spéculaire et diffuse d'un matériau est un enjeu important pour les rendus d'images photo-réalistes. En 1997, NAYAR, FANG et BOULT [NFB97] propose de mesurer la composante spéculaire, séparément de la composante diffuse, en utilisant un filtre polarisant. Selon les équations de Fresnel, une réflexion spéculaire est linéairement polarisée selon son plan d'incidence : réflexion (a) Fig. 2.16. À contrario, une réflexion diffuse

n'est pas polarisée : réflexion (b) Fig. 2.16. En positionnant un filtre polarisant linéaire devant une caméra et en l'orientant selon un angle orthogonal à l'angle d'incidence, on capture uniquement la composante diffuse [Deb+00]. La même démarche, réalisée avec le polarisant orienté selon le même angle que l'angle d'incidence, permet de capturer la composante spéculaire, ainsi que la moitié de la composante diffuse. Il est alors possible de séparer la réflexion spéculaire et la réflexion diffuse, telle que la Figure 2.17 le présente. Sur ces images, on observe que, selon l'hypothèse émise, les défauts de forme géométrique du fond de la boîte sont mis en valeur dans l'image spéculaire, mais pas dans l'image diffuse. En revanche, les défauts qui ont modifiés la structure de la matière (ici des tâches de givrage et des impuretés) apparaissent peu dans l'image spéculaire, mais ils sont présents, de couleur sombre, dans l'image diffuse. Dans la suite de ce travail, nous chercherons à évaluer quantitativement l'apport de la polarimétrie pour améliorer les performances de la détection des défauts.

### 2.3.3.2 Conception d'un capteur polarimétrique bas-coût

Les capteurs polarimétriques commerciaux actuellement disponibles sont coûteux ( $> 5000\text{€}$ ). Cela s'explique par les technologies complexes qu'ils emploient pour obtenir une résolution élevée de la mesure du Degré de Polarisation Linéaire (DoLP). Cela nécessite un processus de calibration long. Aussi, ces capteurs sont principalement utilisés à des fins de recherches ou d'applications de surveillance civile et militaire.

Différentes méthodes de conception d'un capteur polarimétrique existent [Tyo+06]. La méthode la plus classique utilise 3 ou 4 caméras, qui mesurent chacune un angle de polarisation. Plusieurs caméras sont superposées et un chemin optique spécialement conçu aligne précisément l'image des différentes caméras pour compenser la différence de parallaxe. Les miroirs sans-teints sont couramment utilisés pour supprimer la parallaxe entre deux caméras [WA95]. Dans le cas où 3 caméras, ou plus sont utilisées, des miroirs et des prismes deviennent nécessaires. C'est le principe de la *division d'amplitude*.

La méthode de la *division du temps* utilise une caméra haute vitesse et un filtre polarisant, en rotation synchronisée avec la fréquence d'acquisition du capteur. Cela limite le nombre d'images par secondes de l'acquisition et cela limite la durée de vie à cause du système rotatif. Dans la même démarche, le filtre polarisant peut être remplacé par des cristaux liquides [GFB11].

Les capteurs polarimétriques peuvent aussi être construits avec un capteur unique. On parle de *division de l'ouverture* lorsque la surface du capteur est séparée pour acquérir différents degrés de polarisation. Plusieurs lentilles focalisent une image, qui a été polarisée par un filtre, sur une région différente du capteur. Cette méthode nécessite une compensation de l'écart de parallaxe créé par l'écart des différentes lentilles [Rio+15].

Enfin, on parle de *division du plan focal* lorsque ce sont les pixels du capteur qui sont directement polarisés. Des filtres polarisés directement intégrés sur le plan du capteur [Yam+16] (CMOS Sony IMX250-MZR, 5 millions de pixels<sup>10</sup>). Cela supprime la complexité optique, mais réduit la résolution finale de l'image par un facteur 4. Il est possible d'interpoler entre les angles de polarisation pour limiter cette perte, à la manière de la reconstruction d'une image couleur à partir de la matrice de filtre de Bayer. Le capteur IMX250-MZR dispose néanmoins d'une résolution d'acquisition de  $2448 \times 2048$  pixels, soit une résolution de  $1223 \times 1023$  pixels, ce qui est suffisant pour utiliser nos algorithmes.

Nous avons étudié la possibilité de réalisation d'un capteur polarimétrique au coût limité ( $< 1000\text{€}$ ), avec une résolution satisfaisante de  $1920 \times 1080$  pixels [Nag+19]. En comparaison, le coût des caméras basées sur un capteur à division de plan focal est de l'ordre de  $2000\text{€}$ .

Nous avons choisi de ne pas utiliser de système optique pour simplifier la construction. Notre capteur repose sur la fusion de trois caméras polarisées par des filtres linéaires à  $0^\circ$ ,  $60^\circ$  et  $120^\circ$

---

10. Informations sur le capteur polarimétrique **IMX250-MZR**, sur le site Internet du fabricant Sony.

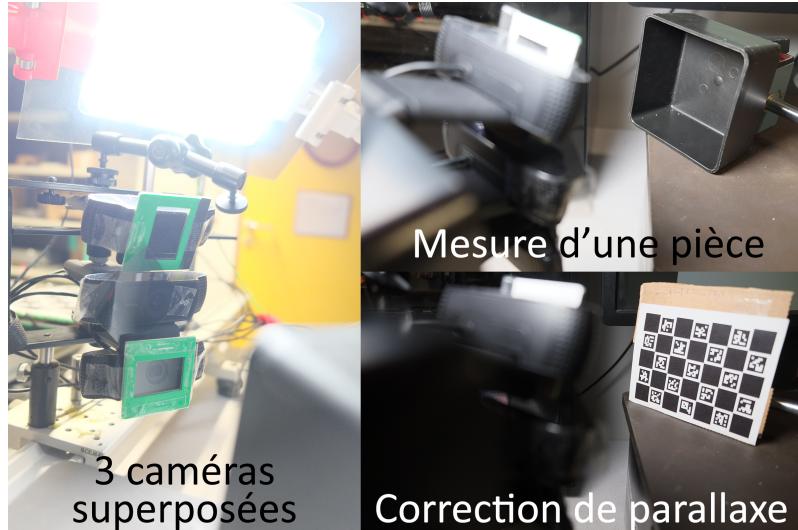


Figure 2.18 – Notre polarimètre et le procédé de calibration par homographie.

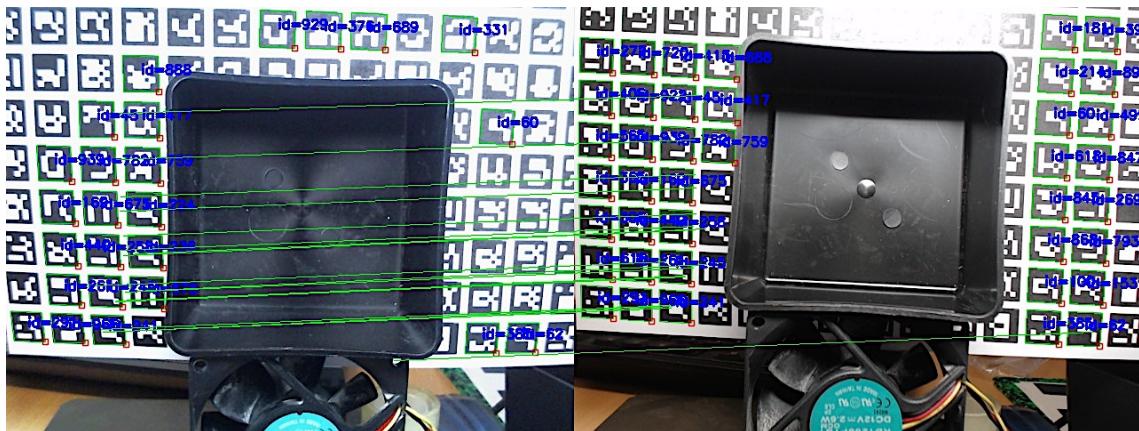


Figure 2.19 – Détection et association de marqueurs ChAruco entre deux caméras.

(Figure 2.18). Ces trois angles, espacés de  $60^\circ$ , permettent de modéliser l'ensemble du disque de polarisation linéaire dans la sphère de Poincaré<sup>11</sup>. Une source de lumière polarisée linéairement à  $0^\circ$  éclaire les pièces.

### Choix des capteurs et correction de parallaxe

Nous utilisons des webcams grands publics Logitech C922<sup>12</sup>, pour leurs simplicités d'implémentation en USB et leurs qualités en condition de basse luminosité. Des caméras industrielles plus robustes peuvent être utilisées avec la même méthode.

Afin de compenser la parallaxe, sans système optique, nous utilisons une transformation logicielle. Dans un premier temps, nous calibrons la position du plan de fusion à l'aide d'une planche de référence composée de marqueurs ChAruco [Gar+14; Gar+16; RMM18]. Ensuite, nous utilisons la bibliothèque OpenCV [Bra00] pour calculer une transformation homographique entre les trois caméras. La détection des marqueurs et l'association entre deux images est présentée

11. Pour reproduire la sphère de Poincaré, il est nécessaire de mesurer également la composante de polarisation circulaire, ce que nous ne faisons pas ici

12. Informations sur la webcam C922 sur le site Internet de la société Logitech.

dans la Figure ?? . Le logiciel de fusion polarimétrique est distribué sous licence libre MIT, sous la forme d'un paquet Python. Le code source du paquet [camera-fusion est disponible en ligne](#).

Il est important de noter que la parallaxe est compensée sur un plan unique. Ainsi, le vecteur de Stokes ne peut être reconstruit que dans ce plan. La parallaxe est une limite importante de notre système lorsque la surface de l'objet que l'on mesure n'est pas située sur le plan de la calibration. La différence de parallaxe dans notre système pourrait être exploitée. Une perspective intéressante à nos travaux sera d'utiliser la théorie *light-field*, à partir des 3 caméras qui possèdent un point de vue différent [Wil+05]. Il sera alors possible d'obtenir une information de polarimétrie en fonction de la distance, tel que le travail de GENDRE et al. [Gen+18] le propose. La reconstruction d'une carte de profondeur de la scène est également possible. La précision attendue n'est pas de l'ordre du micromètre, mais cette information complémentaire pourrait être utile afin d'évaluer la qualité.

Cependant, afin de classifier la qualité des pièces, notre approche est de ne pas corriger la parallaxe. Nous analyserons les données brutes issues des capteurs.

### Dimensionnement de la source lumineuse polarisée

La Figure 2.20 présente les images brutes pour une unique pièce. Dans ce cas spécifique, on observe que la dernière image est plus sombre que les autres. Cela est dû au fait que l'angle du filtre polarisant situé devant la caméra est perpendiculaire à l'angle de polarisation de la lumière. La Figure 2.17 utilisait cette propriété pour séparer les réflexions diffuses et les réflexions spéculaires. Cela peut être limitant pour la définition de cette image, dans le cas où la sensibilité du capteur est trop limitée. Il est recommandé de ne pas orienter les filtres des caméras perpendiculaires à la polarisation de la source lumière.



Figure 2.20 – Images de la même pièce par les 3 caméras, sous 3 angles de polarisation différents.

Enfin, nous avons évalué empiriquement dans des ateliers de production l'effet de la lumière ambiante, non polarisée, sur la qualité de nos mesures. Nous employons une source lumineuse puissante de 8000 lumens pour pouvoir négliger l'effet de la lumière ambiante. Nous estimons que la moitié de la luminosité est perdue par la polarisation. Grâce au dimensionnement de notre source lumineuse, nous n'avons pas observé de problème lors des essais en ateliers. En effet, les zones situées à proximité directe des machines sont souvent peu éclairées, car l'éclairage des ateliers est traditionnellement au plafond ; il est éloigné de plusieurs mètres des machines, tandis que notre dispositif mesure des pièces à une distance maximale d'un mètre. Nous recommandons néanmoins

de vérifier qu'aucune source lumineuse directe ne risque de polluer les mesures. Une perspective à nos travaux sera d'évaluer quantitativement l'effet de la lumière ambiante non polarisée sur les performances de notre système.

### Reconstruction du vecteur partiel de Stokes

L'équation de STOKES 2.1 permet de connaître l'intensité lumineuse de la réflexion, pour un degré de polarisation  $\theta$  [Sto51] :

$$I(\theta, \Phi) = \frac{1}{2} [S_0 + S_1 \cos 2\theta + S_2 \cos \Phi \sin 2\theta + S_2 \sin \Phi \sin 2\theta] \quad (2.1)$$

À l'aide de notre dispositif, nous mesurons l'intensité lumineuse pour trois angles de polarisation linéaires  $\theta$  espacés de  $60^\circ$ ; mais nous ne mesurons pas l'angle de polarisation circulaire  $\Phi$  que nous définissons arbitrairement à zéro. Nous supposons que nos filtres polarisants sont idéaux et polarisent complètement la lumière. Aussi, nous négligeons la transmittance perpendiculaire à leur angle de polarisation effective. À partir de 2.1, nous obtenons les Équations 2.2, 2.3, 2.4 :

$$I(\theta = 0^\circ) = \frac{1}{2} [S_0 + S_1] \quad (2.2)$$

$$I(\theta = 60^\circ) = -\frac{1}{4}S_0 + \frac{1}{2}\sqrt{\frac{3}{2}}S_2 \quad (2.3)$$

$$I(\theta = 120^\circ) = -\frac{1}{4}S_0 - \frac{1}{2}\sqrt{\frac{3}{2}}S_2 \quad (2.4)$$

On obtient le vecteur partiel de STOKES de la lumière réfléchie 2.5 [Li+14] :

$$S_{partiel} = \begin{pmatrix} S_0 \\ S_1 \\ S_2 \end{pmatrix} = \begin{pmatrix} I \\ Q \\ U \end{pmatrix} = \begin{pmatrix} \frac{1}{3}I_0 + \frac{1}{3}I_{60} + \frac{1}{3}I_{120} \\ \frac{2}{3}I_0 - \frac{1}{3}I_{60} - \frac{1}{3}I_{120} \\ \frac{1}{2\sqrt{\frac{3}{2}}}I_{60} - \frac{1}{2\sqrt{\frac{3}{2}}}I_{120} \end{pmatrix} \quad (2.5)$$

La Figure 2.21 présente les résultats de reconstruction obtenus. On observe les erreurs dues à la parallaxe sur une pièce qui n'est ici pas plane. Les mesures sur le fond de la boîte sont néanmoins exploitables. Cette évaluation est actuellement qualitative. Une perspective de ce travail sera d'évaluer la justesse de notre reconstruction sur des échantillons optiques de références.

#### 2.3.3.3 Apport de la polarimétrie sur les performances d'un classifieur de la qualité

Dans "Polarimetric Imaging for Quality Control in Injection Molding", nous évaluons l'apport de la polarimétrie pour la détection des défauts sur les pièces plastiques [Nag+19]. Nous réalisons notre étude sur le jeu de donnée de 204 pièces "boîte noires", qui est présenté dans le Chapitre 3.1 §3.1.1. Les paramètres de la machine ont été variés suivant un plan d'expérience de criblage; aussi, le jeu de donnée est composé de pièces qui possèdent une variété de défauts. La qualité de chaque pièce a été évaluée selon la méthode présentée dans le Chapitre 3.1 §3.2.1. À partir de cette évaluation, nous définissons deux types de pièces sur notre échelle de qualité : les pièces conformes dont l'indice de qualité est inférieur à 3, et les pièces non-conformes dont l'indice est supérieur. Les images enregistrées ont une résolution de  $1920 \times 1200$  et nous avons dû sous-échantillonner cette résolution à  $331 \times 331$  pixels pour réaliser l'analyse sur notre infrastructure informatique.

Nous comparons les performances d'algorithmes de classification de la qualité dans trois cas :

- Sans éclairage polarisé, ni filtres polarisants sur les caméras (*dataset\_bench*),

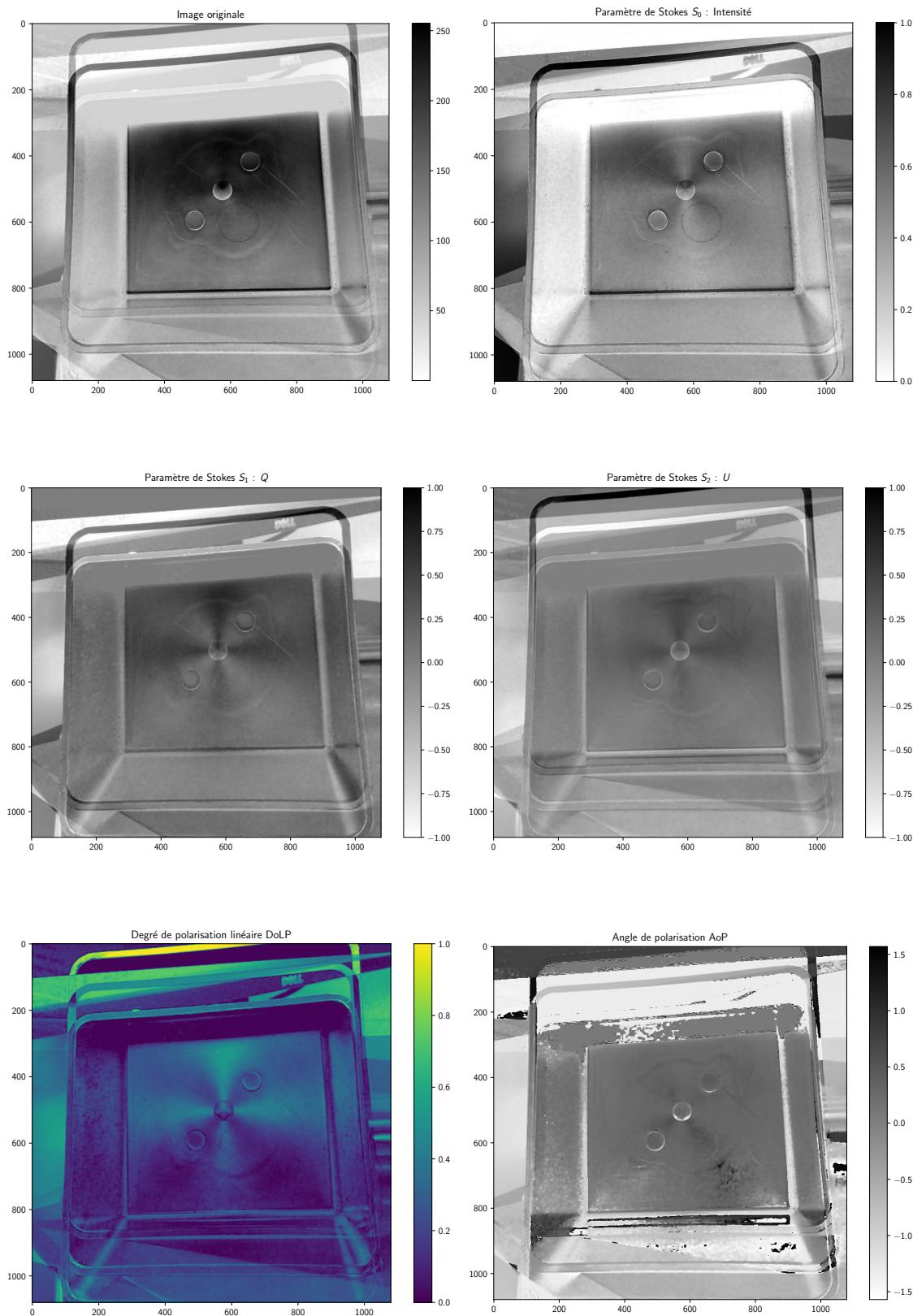


Figure 2.21 – Reconstruction des paramètres de Stokes partiels, du Degré et de l'Angle de Polarisation, à partir des images de 3 caméras.

- sans éclairage polarisé, mais avec des filtres polarisés sur les caméras (*dataset\_nopola*),
- avec éclairage polarisé et filtres polarisés sur les caméras (*dataset\_pola*).

La Figure 2.20 présente les images obtenues dans les trois cas par notre système. La classification est réalisée sur les images brutes issues des capteurs. Nous n'avons pas compensé la parallaxe, ni pré-traité les images. Nous discutons de l'intérêt du pré-traitement d'images dans la section suivante §2.4.2.

Dans le cadre de cette étude, l'algorithme d'apprentissage est considéré comme le meilleur possible, afin de négliger son influence. L'algorithme d'apprentissage est choisi et il est optimisé automatiquement par la méthode *TPOT*, qui est détaillée dans le Chapitre 3 §3.4.6. La durée de l'apprentissage du meilleur classifieur est de 15 heures.

Angle	<i>dataset_pola</i>	<i>dataset_nopola</i>	<i>dataset_bench</i>
3 angles	<b>0,921</b>	0,851	<b>0,948</b>
0°	0,841	0,831	<b>0,923</b>
60°	0,853	0,851	0,820
120°	0,812	0,830	<b>0,889</b>

Tableau 2.6 – Comparaison des scores de justesse de classification de la non-conformité.

Le Tableau ?? présente les scores de justesse obtenus avec le meilleur algorithme de classification, pour différentes conditions d'éclairage et de polarisation. La première ligne indique le score dans le cas où les images des trois caméras ont été utilisées. Les images brutes des capteurs sont concaténées dans une matrice de dimensions  $993 \times 331$ . Les lignes suivantes indiquent le score dans le cas où seule une caméra polarisée à un certain angle est utilisée. La première colonne est le score obtenu avec un éclairage polarisé. La seconde colonne est le score obtenu avec un éclairage non polarisé. Enfin, la dernière colonne est une mesure de référence : aucune caméra n'est polarisée et l'éclairage n'est pas polarisé.

On observe que l'utilisation des 3 caméras polarisées et de l'éclairage polarisée permet d'obtenir un score de classification plus élevé (justesse de 0,921). La polarimétrie semble apporter de l'information pour améliorer la détection des pièces conformes.

En revanche, l'utilisation des 3 caméras sans polarisation permet d'obtenir un score plus élevé que le système polarisé. Nous expliquons ce résultat par le fait qu'il existe de l'information supplémentaire pertinente dans la fusion des trois points de vue. De plus, les filtres polarisants pourraient dégrader une partie de cette information issue des trois caméras. Enfin, il est possible que notre système à faible coût occasionne un biais pour les données. Nous n'avons pas réussi à mettre en évidence de biais. Une perspective importante de nos travaux sera de réaliser la même étude à l'aide d'un polarimètre calibré, afin de valider définitivement si c'est la polarimétrie qui apporte de l'information, ou bien les différents points de vues des caméras.

## 2.4 Traitement et fusion de l'information issue de mesures multimodales

Dans le cadre de nos travaux, nous avons évalué l'intérêt de deux techniques d'imageries non-conventionnelles, polarimétrie et thermographie, pour la mesure des caractéristiques d'une pièce. Nous cherchons désormais à combiner ces mesures. De nombreuses méthodes de fusions

multimodales peuvent être appliquées.

Dans notre étude sur l'apport de la polarimétrie §2.3.3.3 [Nag+19], nous proposons de concaténer les images issues des capteurs et d'exploiter cette donnée en utilisant l'apprentissage statistique. Cette démarche peut être réalisée en ajoutant une image thermique aux données. On parle de "fusion amont" des données. Cette démarche est intéressante car un algorithme d'apprentissage statistique va pouvoir apprendre de la relation entre les différentes mesures. C'est la démarche que nous choisissons d'employer dans nos travaux.

Sur ces données concaténées, il est possible d'appliquer un méthode de réduction de l'information §2.4.3. Il est important que cette méthode ne détruise pas l'information pertinente contenue dans les mesures. C'est pourquoi nous avons privilégié dans nos travaux les méthodes de réduction de l'information par apprentissage non-supervisé. Dans le Chapitre §3, nous comparons différentes méthodes d'apprentissage non-supervisées sur nos données.

Dans cette section, nous présenterons en particulier les méthodes de filtrages que nous avons évaluées sur nos données. Enfin, nous détaillerons l'intégration logicielle des méthodes de fusion sur système embarqué, dans le Chapitre 4.

#### 2.4.1 Nécessité de l'extraction de l'information pertinente

Nous supposons dans cette section que l'information pertinente est contenue dans les mesures issues des capteurs. Aucun algorithme ne peut extraire de l'information si elle n'existe pas présente. Nous avons discuté de la pertinence des mesures thermique et polarimétrique dans la Section 2.3.

Une perspective de nos travaux est de proposer l'ajustement en boucle fermée de la qualité des pièces. C'est pourquoi notre système de contrôle de la qualité doit respecter le temps de cycle du procédé, afin de pouvoir corriger les paramètres du procédé dès la pièce suivante. Doivent être inscrits dans le temps de cycle de production d'une pièce : le déplacement de la pièce devant le système de mesure, la durée de la mesure et la durée de l'analyse de la mesure. Si la durée de la mesure est de 10 secondes, la durée de l'analyse des mesures doit être proche de la seconde. Le contexte industriel demande une grande fiabilité et le respect de cette durée d'analyse de la mesure.

Nous avons évalué la faisabilité industrielle d'un traitement des mesures sur serveur distant (*Cloud*), qui disposent de moyens de calcul performants. La durée des calculs sur ces serveurs est de l'ordre de la milliseconde. En pratique, ce sont les problèmes de latence des réseaux informatiques qui ne permettent pas d'être certain d'une réponse inférieure à la seconde. Sans évolution de l'infrastructure réseau dans les ateliers de production, au vue de cette contrainte temporelle, la réalisation de l'analyse des mesures doit être effectuée sur un système de calcul embarqué, au plus proche des moyens de mesure.

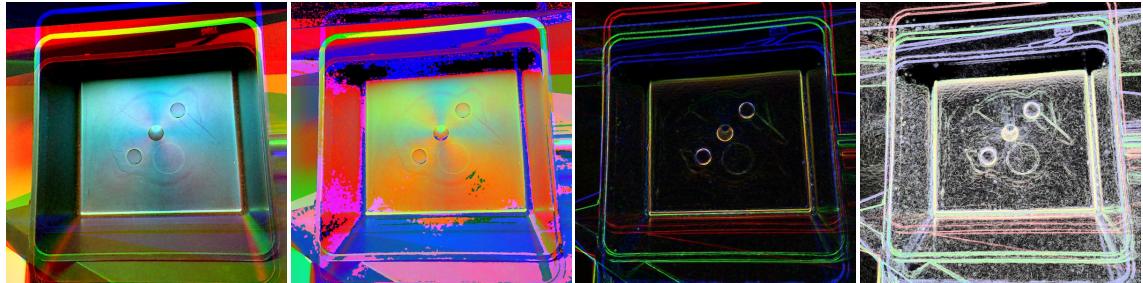
Aussi, plus la dimension de l'information issue des capteurs est réduite, plus l'inférence est rapide. La réduction de dimensions est aussi appelée "extraction de l'information" ou "compression de l'information".

Dans cette section, nous discutons de l'intérêt du traitement des images obtenues par nos capteurs, en amont de l'analyse des mesures §2.4.2. Puis nous présentons notre méthode d'extraction de l'information pertinente, à l'aide d'une sélection automatique de descripteurs statistiques §2.4.4.

#### 2.4.2 Traitement d'images amont

Nous avons étudié la possibilité de mettre en valeur la présence des défauts d'aspect que nous souhaitons détecter. Notre hypothèse est simple : si un traitement d'image fait ressortir les défauts pour la vision d'un humain, ce traitement améliorera les performances du système informatique de détection des défauts.

Nous avons étudié l'effet de différentes méthodes de traitement d'image sur la performance de la classification de la qualité des pièces. La Figure 2.22 présente les différents traitements étudiés.



(a)  $0^\circ, 45^\circ, 90^\circ \rightarrow \text{RGB}$  (b)  $\text{Différence à } 0^\circ \rightarrow \text{RGB}$  (c)  $\text{Diff.} + \text{gaussien} + \text{Sobel}$  (d) (c) avec histo. optimisé

Figure 2.22 – Illustration de différentes méthodes de pré-traitements des images issues des trois caméras polarisées.

En pratique, nous n'avons pas réussi à valider notre hypothèse. Toutes les méthodes de traitement que nous avons implémentées ont diminuées les performances de classification. Le traitement d'image semble détruire une partie de l'information pertinente. Aussi, à défaut de trouver un traitement idéal, nous recommandons de ne pas réaliser de traitement d'images et d'analyser les données brutes issues des capteurs. C'est l'intérêt des méthodes d'apprentissage statistique pour exploiter les données brutes. Ils construisent la méthode de traitement la plus appropriée de manière automatique.

### 2.4.3 Descripteurs statistiques

Nous utilisons des descripteurs de référence de la littérature. À contrario des méthodes présentées dans le Chapitre 3, il n'y a pas d'apprentissage. C'est l'humain qui conçoit les fonctions mathématiques des descripteurs. La limite à cette démarche est le manque de variété des descripteurs disponibles. À contrario, les méthodes par apprentissage de réseau de neurones permettent de créer des descripteurs spécifiquement adaptés aux problèmes à résoudre. Le réseau de neurones va apprendre la méthode : c'est la démarche de l'apprentissage de bout-en-bout (*end to end learning*) : à partir des données brutes obtient le résultat attendu.

Dans notre étude “Quality Prediction in Injection Molding”, nous avons choisi de balayer un large spectre de la vision par ordinateur [Nag+17]. 25 descripteurs statistiques ont été évalués :

- statistiques d'ordre inférieur à 2 :
  - minimum,
  - maximum,
  - moyenne,
  - médiane,
  - mode,
  - écart-type,
  - quantile 75 et 90,
- statistiques d'ordre supérieur :
  - quantile 75% et 90%,
  - assymétrie (*skewness*),
  - coefficient d'aplatissement (*kurtosis*),
  - 14 paramètres de HARALICK, SHANMUGAM et DINSTEIN,
  - 100 premiers modes de la Décomposition Modale Discrète.

Nous détaillerons le principe des paramètres de Haralick §2.4.3.1 et de la Décomposition Modale Discrète §2.4.3.2 dans les paragraphes suivants.

### 2.4.3.1 Description de la texture d'une image : paramètres de Haralick

En 1973, HARALICK, SHANMUGAM et DINSTEIN proposent 14 paramètres pour décrire la texture d'une image [HSD73]. Les travaux de doctorat de LACOMBE utilisent avec succès les paramètres de Haralick pour décrire des images, dans un objectif d'automatisation du contrôle de la qualité de pièces plastiques. Les paramètres de Haralick ont été largement utilisés dans la littérature du traitement d'images. C'est pourquoi nous avons étudié leurs apports pour la description de nos images thermiques et polarimétriques.

La méthode de Haralick nécessite le calcul de la matrice des co-occurrences des pixels voisins de l'image, appelée *Gray-Level Co-occurrence Matrix*. La matrice de co-occurrence  $G$  est carrée, de dimension  $N_g$  le nombre de niveau de luminance dans l'image. Pour une image classique 8 bits,  $N_g = 256$ . La matrice  $G$  est constituée de couples de valeur  $P_{i,j}$  qui indiquent le nombre d'occurrence où un pixel de valeur  $i$  est adjacent à un pixel de valeur  $j$ , Équation 2.6. Pour une image 8 bits, on aura  $255^2$  couples  $p$ . HARALICK, SHANMUGAM et DINSTEIN proposent d'évaluer les co-occurrences entre les pixels selon 4 directions  $\theta = 0, \frac{\pi}{2}, \frac{3\pi}{4}, \pi$ .

$$\mathbf{G} = \begin{bmatrix} p(1,1) & p(1,2) & \cdots & p(1,N_g) \\ p(2,1) & p(2,2) & \cdots & p(2,N_g) \\ \vdots & \vdots & \ddots & \vdots \\ p(N_g,1) & p(N_g,2) & \cdots & p(N_g,N_g) \end{bmatrix} \quad (2.6)$$

Une distance  $d$  entre les couples de pixels comparées peut également être introduite. En pratique on choisit souvent  $d = 1$  pour que les couples soient les voisins directs. La matrice finale est divisée par le nombre total  $R$  de comparaisons  $P$  effectuées  $p_{i,j} = \frac{P_{i,j}}{R}$ ;  $R$  est fonction des dimensions de l'image. Chaque valeur de la matrice peut être considérée comme la probabilité qu'un pixel de valeur  $i$  soit adjacent à un pixel de valeur  $j$ , selon l'angle  $\theta$ . Ainsi, pour une image évaluée sous 4 angles, la matrice de co-occurrence  $G$  a pour dimension  $256^2 \times 4 = 262\,144$ . Pour une image couleur à 3 canaux, 3 matrices de co-occurrences sont calculées, une matrice par canal. Une image d'une profondeur de  $B$  bits à 3 canaux, de dimensions  $N \times N$  pixels, possède une matrice de co-occurrences de dimension  $N^2 \times 3B^2 \times 4$  valeurs.

Afin d'exploiter cette matrice qui est beaucoup plus volumineuse que l'image originale, HARALICK, SHANMUGAM et DINSTEIN proposent 14 descripteurs statistiques qui encodent une information sur la texture de l'image.

1. le second moment angulaire  $\sum_i \sum_j p(i,j)^2$ , qui mesure l'homogénéité de l'image,
2. le contraste  $\sum_{k=0}^{N_g-1} k^2 p_{x-y}(k)$ , qui mesure la quantité de variation locale,
3. la corrélation,  $\frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (ij)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$ , qui mesure la relation entre les niveaux des pixels voisins,
4. la variance, l'inertie)  $\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu)^2 p(i,j)$ ,
5. le moment différentiel inverse  $\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{1}{1+(i-j)^2} p(i,j)$  qui mesure l'homogénéité locale : il est élevé lorsque les niveaux des pixels locaux sont uniformes et réciproquement,
6. la moyenne des sommes  $\sum_{i=2}^{2N_g} i p_{x+y}(i)$ ,
7. la variance des sommes  $\sum_{i=2}^{2N_g} (i - \text{entropie})^2 p_{x+y}(i)$ ,
8. l'entropie des sommes  $-\sum_{i=2}^{2N_g} p_{x+y}(i) \log(p_{x+y}(i))$ ,

9. l'entropie  $-\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \log(p(i,j))$ ,
10. la variance des différences  $\sum_{i=0}^{N_g-1} i^2 p_{x-y}(i)$ ,
11. l'entropie des différences  $-\sum_{i=0}^{N_g-1} p_{x-y}(i) \log(p_{x-y}(i))$ ,
12. la première information sur la mesure de corrélation  $\frac{\text{entropie-HXY1}}{\max\{\text{HX},\text{HY}\}}$ ,
13. la seconde information sur la mesure de corrélation  $[1 - \exp(-2(\text{HXY2} - \text{entropie}))]^{1/2}$ ,
14. le coefficient de corrélation maximum, qui est la seconde plus grande valeur propre de  $Q(i,j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$ .

#### 2.4.3.2 Filtrage d'image par décomposition modale

La problématique de la mesure des surfaces complexes a amené aux développements de méthodes de description de la géométrie des objets. Ces méthodes s'appuient principalement sur les théories du traitement de signal. Les méthodes de description géométrique les plus répandues sont :

- La transformée en cosinus discrète qui utilise des fonctions harmoniques en cosinus pour décrire les surfaces par leurs modes et leurs amplitudes. Cette méthode est souvent utilisée en traitement d'images et pour la compression de données.
- La décomposition en séries de Fourier, qui s'appuie sur la transformation de Fourier discrète. Cette méthode permet de décomposer une surface en un espace harmonique composé de fonctions sinus et cosinus. La décomposition obtenue est ainsi plus générale que celle obtenue par la Transformée en cosinus discrète.
- La transformée en ondelettes discrète, qui décompose la surface tridimensionnelle mesurée en fonctions d'une base composée d'ondelettes discrètes.
- La décomposition en harmoniques sphériques qui décrit des formes complexes dans un espace harmonique sphérique. Cette technique est notamment utilisée pour caractériser des formes dans un espace en trois dimensions.

En complément de ces travaux, la décomposition d'une forme par ses modes de vibrations a été proposée en 2007 par SAMPER et FORMOSA [SF07]. Puis, les travaux de doctorat de ADRAGNA ; FAVRELIERE [Adr07 ; Fav09] ont cherché à utiliser la Décomposition Modal Discrète (DMD) comme méthode de tolérancement et de spécification des formes géométriques complexes [Fav09].

Une décomposition modale décompose un signal dans une base spectrale, qui est construite à partir d'une base de modes propres. La base des modes propres est définie par ses vecteurs propres, aussi appelés "vecteurs modaux". La surface mesurée est alors projeté dans cette base propre ; elle est exprimée en terme d'amplitude des différents modes discrets. La base est construite à partir de la solution à un problème de mécanique vibratoire [Goï+16]. Récemment, GOÏC et al. [Goï+ct] a montré que la DMD peut être utilisée pour caractériser des géométries globales (défauts de formes), ainsi que des géométries locales (rugosités).

Lorsque la DMD est utilisée pour un problème de géométrie plane, les modes correspondent aux modes propres de la vibration d'une plaque plane. Contrairement aux publications précédentes qui s'appuient sur des logiciels d'analyse par éléments finis pour obtenir les modes de vibrations (théorie des poutres épaisses de Timoshenko [Tim37]), nous construisons une base modale à partir de la théorie de Euler–Bernoulli. La solution analytique de la valeur des modes  $w_n$  de vibration d'une poutre aux conditions aux limites libre-libre est donnée par l'Équation 2.7<sup>13</sup>.

13. Afin de réaliser l'application numérique, nous fixons arbitrairement les valeurs de l'amplitude  $A = 1\text{m}$ , module de Young  $E = 1.0E^{+10}\text{Pa}$ , du moment d'inertie  $I_{xx} = 1.0E^{-08}\text{kg}\cdot\text{m}^2$ , la masse par unité de longueur  $\mu = 3.0E^{+02}\text{kg}\cdot\text{m}^{-1}$ .

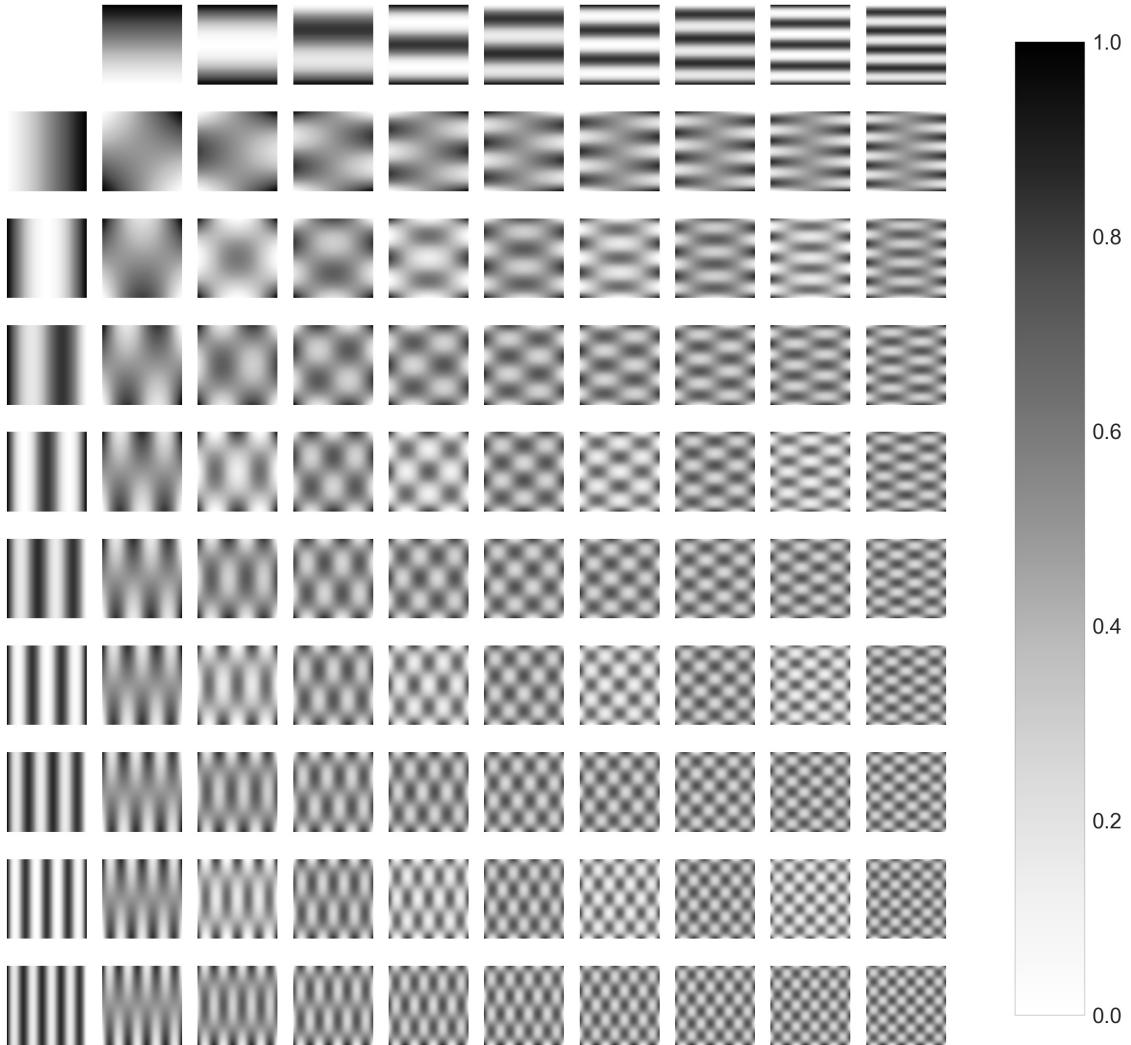


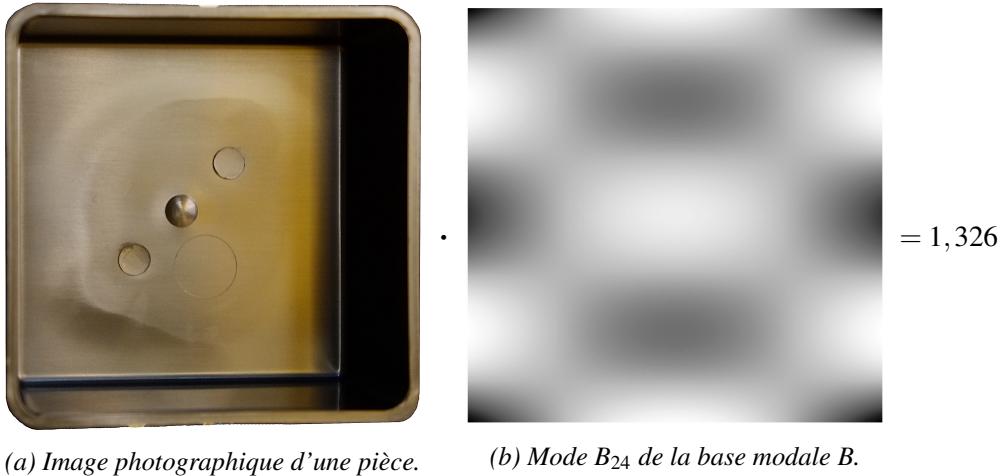
Figure 2.23 – Représentation des 100 premiers modes de notre base  $B$ .

$$\hat{w}_n = A \left[ (\sin \beta_n x + \sinh \beta_n x) + \frac{\sin \beta_n L - \sinh \beta_n L}{\cosh \beta_n L - \cos \beta_n L} (\cos \beta_n x + \cosh \beta_n x) \right] \quad \text{avec } \beta_n := \left( \frac{\mu \omega_n^2}{EI} \right)^{1/4} \quad (2.7)$$

Les modes obtenues pour une poutre sont combinés par un simple produit afin d'obtenir la base modale  $B_{ij} = w_i \cdot w_j$ . La Figure 2.23 présente les modes plans obtenus. Il est intéressant de noter que nous construisons une représentation erronée des modes de vibrations d'une plaque, pour des conditions aux limites libre-libre : les modes à la symétrie horizontale et verticale (colonne 1 et ligne 1) n'existent pas<sup>14</sup>. L'application de la transformée modale au filtrage d'images est directe : on décrit l'image en la projetant dans la base modale. On peut également reconstituer l'image originale par la transformée inverse.

Les travaux de doctorat de PITARD [Pit16] ont montré l'intérêt de la Décomposition Modale Discrète pour la reconstruction du champ de luminance d'un objet, à partir de mesures par un dôme multi-éclairages. Par la suite, les travaux de doctorat de LACOMBE [Lac18] exploitent la DMD pour décrire les mesures d'un dôme multi-éclairages pour le contrôle de l'aspect de produits.

<sup>14</sup>. La théorie des plaques de Kirchhoff-Love permettrait une modélisation réaliste des modes de vibration d'une plaque libre ; dans le cas du filtrage d'image, cela permettrait de supprimer la présence de modes orthogonaux.



*Figure 2.24 – Filtrage de la photographie d'une pièce par un mode de la base B.*

Nous proposons d'utiliser cette paramétrisation DMD sur les images polarimétrique et thermique.

Nous proposons d'appliquer la DMD à la description des images d'une pièce conforme et d'une pièce non-conforme, Figure 2.25. La base modale a été ici optimisée par validation croisée §3.3.1.2 afin de maximiser les performances de classification des pièces conformes/non-conformes ???. Les 27 meilleurs modes ont été retenus. La moyenne des modes de la base est représentée dans la première image de la Figure 2.25. Les deux dernières images de la Figure 2.25 présentent le résultat d'une reconstruction d'une image d'une pièce conforme et d'une image d'une pièce non-conforme, à partir d'une description modale. On observe que l'image originale n'est pas reconstruite. La majeure partie de l'information de l'image est perdue par le filtrage. Cependant, il est facile de distinguer visuellement une pièce conforme d'une pièce non-conforme : leurs descriptions modales possèdent des différences visuelles.

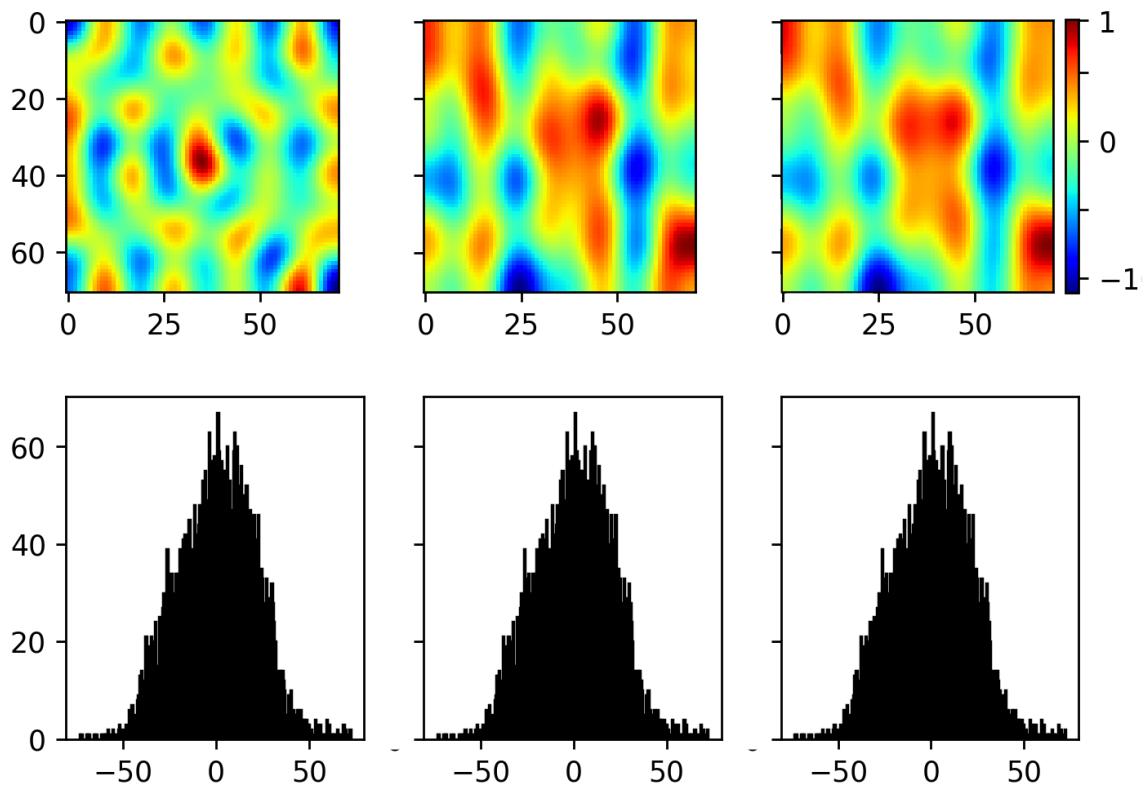
#### 2.4.4 Sélection automatique de descripteurs pertinents

Dans cette section, nous présentons la méthode de sélection de descripteurs que nous avons utilisée dans nos travaux. L'objectif de cette démarche est d'obtenir un nombre réduit de valeurs qui contiennent l'information pertinente de la mesure de la qualité. À partir de ces valeurs, un algorithme de classification est utilisé pour déterminer la qualité de la pièce. Nous discutons des méthodes existantes pour construire ce modèle classifieur dans le Chapitre 3 suivant.

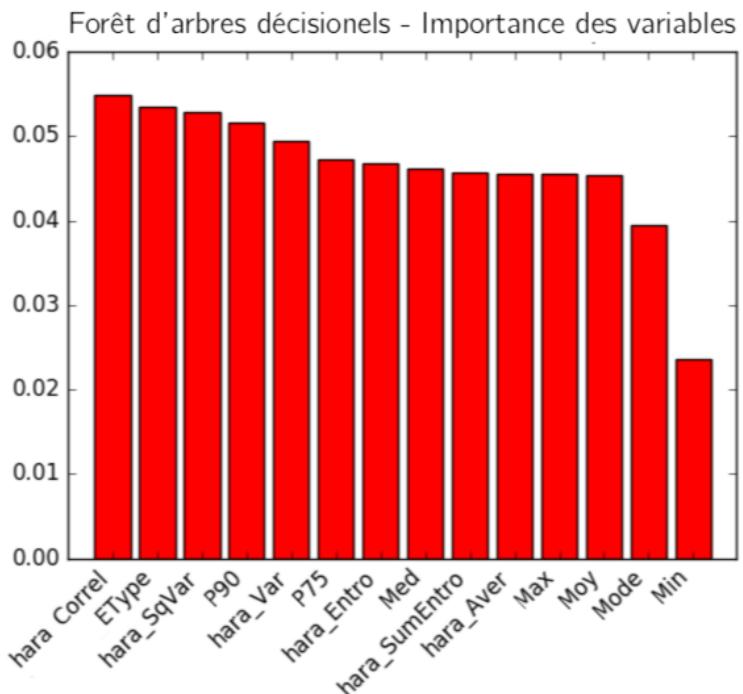
Nous avons retenu comme méthode de sélection automatique des descripteurs les Forêts d'arbres décisionnels pour leurs meilleures performances sur nos cas d'études. Pour chaque arbre, il est possible d'obtenir le nombre de variables qui sont associées. On obtient ainsi une valeur sur l'importance relative des variables. Plus une variable est associée à un grand nombre d'arbre, plus cette variable est importante. La Figure 2.26 montre les résultats de cette méthode dans le cadre de l'étude de classification de la qualité à partir des images thermiques §2.3.2 [Nag+17]. Les paramètres de Haralick §2.4.3.1 sont les plus représentés. En particulier, l'homogénéité, la variance et l'entropie sont les plus pertinents. L'homogénéité telles que définis par HARALICK, SHANMUGAM et DINSTEIN est un indicateur de l'uniformité d'une image. L'entropie évalue la présence de motifs aléatoires. Dans notre cas, la présence de défauts entraîne un écart plus important de ces valeurs, par rapport à une image conforme, ce qui permet de classifier.

#### 2.4.5 Description d'images par apprentissage statistique

L'intérêt de l'utilisation de méthodes d'apprentissage pour la description d'images est que c'est l'algorithme qui va construire la méthode de traitement optimale des données brutes. Nous



*Figure 2.25 – Base modale moyenne obtenue par sélection des 27 modes par validation croisée. Reconstruction d'une image d'une pièce conforme et d'une image d'une pièce non-conforme, après description modale.*



*Figure 2.26 – Importance des descripteurs par Forêt d'arbres décisionnels.*

présentons les méthodes que nous avons évaluées sur notre problématique dans le Chapitre §3 §3.3.4.

En particulier, nous avons obtenu des résultats encourageants à l'aide des auto-encodeurs variationnels §3.3.4.7. Un auto-encodeur cherche à reproduire la donnée entrée, avec le minimum de paramètres possibles. Ainsi, l'auto-encodeur apprend à compresser l'information dans un vecteur de petite dimension. C'est une méthode d'apprentissage non-supervisée. Aucune annotation des données n'est requise.

Dans le Chapitre §3, nous comparons différentes méthodes d'apprentissage non-supervisées. Nous discutons en particulier du coût des calculs des différentes méthodes et de la possibilité d'intégration de ces algorithmes sur des systèmes embarqués, au plus près des capteurs.

## 2.5 Synthèse et perspectives

Pour conclure ce chapitre, nous réalisons une synthèse de l'apport de notre travail au contrôle de la qualité des pièces plastiques. Nous proposons ensuite plusieurs pistes de recherche à privilégier.

### 2.5.1 Synthèse : mesure non-invasive en cycle industriel

Dans ce chapitre, nous avons mis en évidence l'importance de la sélection d'une moyen de mesure adapté aux contraintes industrielles. L'invasivité d'un moyen de mesure pour le procédé est un critère important à prendre en compte. Nous avons sélectionné des moyens de mesures sans contacts qui s'appuient sur l'imagerie non-conventionnelle. Afin d'évaluer la pertinence de l'imagerie thermique et de la polarimétrie pour le contrôle de la qualité, nous avons réalisé des études sur une production de 192 pièces. Une perspective importante à nos travaux est la reproduction de nos résultats sur des cas d'études différents et sur un nombre de pièces beaucoup plus important.

La triangulation laser permet d'obtenir un nuage de point de la géométrie de la pièce. Cependant, les difficultés de mises en place et la faible répétabilité de la mesure nous ont amené à écarter ce moyen de mesure. La thermographie contient une information sur la géométrie de la pièce. C'est une mesure rapide qui peut être déployée facilement. Une perspective de recherche est l'étude du refroidissement dynamique des pièces. La polarimétrie permet de mettre en valeur différents types de défauts : les défauts géométriques et les défauts sur la structure de la matière. C'est également une méthode de mesure rapide que nous avons mis en place avec succès en cycle industriel. Nous avons conçu un polarimètre bas-coût pour réaliser cette mesure et il sera nécessaire de confronter nos mesures à un capteur calibré.

Enfin, nous avons discuté de différentes méthodes de traitements de ces mesures multimodales. Dans le Chapitre 3, nous discuterons en particulier de l'exploitation de ces mesures par apprentissage statistique. L'objectif est de réaliser un système qui classe la qualité des pièces à partir des mesures brutes.

À la suite de ce travail exploratoire, nous avons intégré la mesure thermique et la polarimétrie dans un système compact, adapté aux contraintes industrielles. Nous présentons ce système dans le Chapitre 4.

### 2.5.2 Perspectives de recherche

Nous détaillons dans la suite trois perspectives majeures à nos travaux : la première concerne l'exploitation dynamique de la mesure thermographique, la seconde est l'apport de la mesure

de réflectance de la surface de la matière, et enfin l'exploitation optimale de la polarimétrie par l'imagerie de la matrice de Mueller complète.

### 2.5.2.1 Thermographie dynamique

L'intérêt de l'utilisation d'une caméra thermique est que la fréquence d'acquisition est au minimum de 9 images par secondes<sup>15</sup>. Cela permet de réaliser des mesures dynamiques. Une perspective de travail intéressante est la mesure du refroidissement de la pièce à partir de plusieurs images thermiques prisent dans le temps. La soustraction des valeurs de température donnera l'évolution de la température. L'équation analytique du transfert thermique par convection permet de dimensionner la durée de l'intervalle de capture en fonction de la sensibilité des microbolomètres (dans notre cas 0,05°C en moyenne à 30°C).

En posant l'hypothèse que le transfert de la chaleur par conduction au sein de la pièce et par convection à l'interface avec l'air sont du même ordre de grandeur, on obtient l'Équation 2.8

$$h(T_{surface} - T_{ambiante}) \approx \frac{k}{L_{épaisseur}} (T_{interne} - T_{surface}) \quad (2.8)$$

On pose l'hypothèse que la température de la pièce est uniforme, ce qui permet de négliger les gradients de température internes. On obtient  $\dot{Q}_{pièce}$ , le flux de chaleur transféré à l'air par la pièce, Équation 2.9.

$$\dot{Q}_{pièce} = Sh(T - T_{ambiante}) = -\rho V C_p \frac{dT}{dt} \quad (2.9)$$

Une solution représente l'évolution de la température de la pièce au court du temps, Équation 2.10.

$$\frac{T_t - T_{ambiante}}{T_{sortie} - T_{ambiante}} = \exp\left(-\frac{hS}{\rho c V} t\right) \quad (2.10)$$

- $\rho = 0,9210E^{-3}\text{kg}\cdot\text{m}^{-3}$ , la densité du polypropylène,
- $C_p = 2000\text{J}\cdot\text{kg}^{-1}\cdot\text{K}^{-1}$ , la capacité thermique massique approximative du polypropylène,
- $h = 15\text{W}\cdot\text{m}^{-2}\cdot\text{K}^{-1}$ , le coefficient d'échange thermique de l'air ambiant,
- $S = 503,0\text{cm}^2$ , la surface de la peau de la pièce étudiée,
- $V = 61,7\text{cm}^3$ , le volume de matière de la pièce étudiée,
- $T_{sortie} = 363\text{K}$ , la température de la pièce à sa sortie de l'outillage,
- $T_{ambiante} = 293\text{K}$ , la température ambiante de l'atelier de production.

Le modèle analytique prévoit une diminution de 0,5°C par seconde de la température de la pièce. Il est possible de réaliser une mesure dynamique en capturant des images thermographiques toutes les secondes ; ce qui est compatible avec le temps de cycle industriel. La Figure 2.27 montre une mesure préliminaire réalisé sur une pièce 10 secondes après sa production. On observe que la diminution de la température sur 2 secondes est en moyenne de 1°C, ce qui est concordant avec le résultat analytique §2.10. Le champ de l'évolution de température est difficile à interpréter visuellement. Il est possible que cette mesure dynamique contienne beaucoup de bruit. Une étude statistique sur de nombreuses pièces devra être réalisée. La plage temporelle pourrait alors être allongée et le nombre de prises d'images augmenté pour moyenner le bruit. Une perspective de recherche est l'exploitation de ces mesures, par exemple à l'aide de réseau de neurones de convolutions ou de réseaux récurrents pour prendre en compte l'évolution temporelle. Le champ

---

15. L'utilisation de fréquences d'acquisition supérieures à 9 Hertz est limitée par la politique de restriction du commerce international de la défense [International Traffic in Arms Regulations](#).

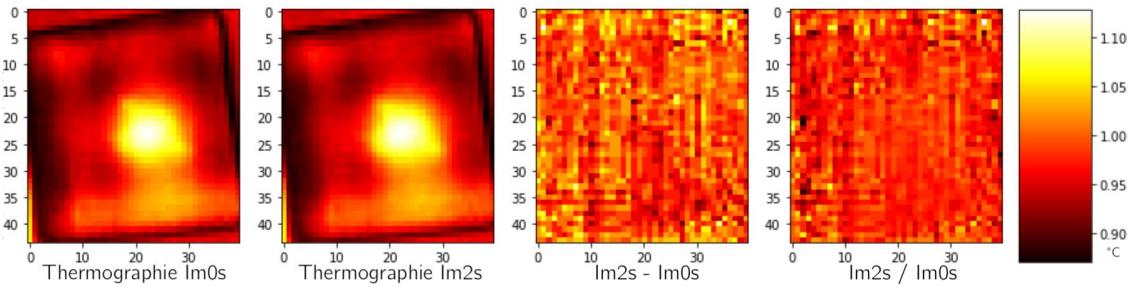


Figure 2.27 – Mesure par caméra thermique de l'évolution sur 2 secondes de la température d'une pièce.

de la dynamique de refroidissement pourrait contenir une information intéressante du point de la densité de la pièce.

### 2.5.2.2 Mesure de la réflectance en ligne de production

Ce travail de doctorat ne nous a pas permis d'explorer l'ensemble des méthodes d'imageries non-conventionnelles. En particulier, nous ne nous sommes pas intéressés à la mesure de la réflectance de la surface de la pièce. Les caractéristiques de l'apparence d'un objet peuvent être mesurées avec un dôme composé de multiples points d'éclairages (aussi appelé "dôme multi-éclairages" ou "dôme de réflectance") [Mar+00 ; Gho+07]. Ce système est plus rapide que le gonio-réflectomètre qui doit déplacer la source lumineuse avec précision. La mesure de multiple point d'éclairage permet de détecter des défauts géométriques de l'ordre du micromètre. La distribution de la luminosité des réflexions<sup>16</sup> est modifiée par la présence du défaut. Par exemple, un défaut de rayure entraîne une concentration de l'intensité lumineuse sur un petit nombre de pixels du capteur, ce qui permet de le mettre en valeur [Dés+14].

Les caractéristiques de réflectance d'un matériau possèdent un lien direct avec la rugosité de sa surface [TS67]. ZHANG, DANA et NISHINO [ZDN16] propose une méthode d'exploitation intéressante de la mesure d'un matériau sous différents angles d'éclairages, qui utilise un réseau de neurones de convolutions (le principe du réseau de convolution est présenté dans le Chapitre 3 §3.3.1.4). L'étude utilise un jeu de données composé d'une unique image sous 42 angles d'éclairage, et de la valeur du coefficient d'adhérence associé au matériau. 137 échantillons composent le jeu de données, ils sont répartis en 21 classes représentant des coefficients d'adhérence différents. Le modèle à réseau de neurones de convolution est capable de prédire le coefficient d'adhérence du matériau à partir des éclairages avec un coefficient de corrélation  $R^2 = 0,7885$ . Ce résultat encourageant motive l'utilisation de la mesure de réflectance des pièces.

De plus, la rugosité d'un matériau produit des sensations haptiques lorsqu'il est touché. La sensation haptique d'un objet est une caractéristique importante pour le client final. Certains cahiers des charges spécifient désormais des sensations haptiques pour les produits, en particulier dans le domaine du luxe. L'étude "See to Touch ?" a cherché à quantifier la relation entre les mesures réalisées à l'aide d'un dôme multi-éclairages et la sensation haptique perçue sur les échantillons [Lac+17]. Sur un nombre restreint de 12 échantillons, les auteurs établissent un modèle entre le coefficient d'adhérence et les images réalisées à l'aide d'un dôme de réflectance. L'erreur de prédiction maximale est de 7%, ce qui montre l'intérêt de la mesure de réflectance pour l'étude des sensations haptiques.

Cependant, ce système de dôme de réflectance est difficile à intégrer sur les lignes de production, en raison de sa taille. Plus l'objet à mesurer est grand, plus le diamètre du dôme

<sup>16</sup>. La distribution de la luminosité des réflexions est appelée "réflectivité bidirectionnelle" (*Bidirectional Reflectance Distribution Function*).

augmente. Pour une application de contrôle de pièces automobiles de 1 mètre de long, le dôme peut atteindre plus de 10 mètres de diamètres.

Un système d'éclairage mono-point est plus simple à intégrer en production, mais il nécessite un positionnement précis de l'équipement d'éclairage et du capteur. Enfin, il n'est pas robuste aux décalages de positionnement de la pièce. Il nécessite un positionnement précis des pièces, ce qui augmente la complexité technique de la chaîne de production. Une perspective importante à nos travaux est l'intégration de la mesure de réflectance en ligne de production, dans le respect des contraintes industrielles.

### 2.5.2.3 Polarimétrie par imagerie de Mueller

Dans le cadre de ce travail, nous n'avons pas réalisé la mesure de l'ensemble des couples de polarisations. En particulier, nous n'avons pas fait varier la polarisation de l'éclairage. Nous n'avons pas non plus utiliser la polarisation circulaire. Une perspective de recherche intéressante est de réaliser une mesure de la matrice de Mueller complète [PC95].

Pour mesurer la matrice de Mueller, il est nécessaire de réaliser 16 images en variant les conditions de polarisation :

- 4 images en variant l'angle de polarisation de la source lumineuse (*Polarization State Generator*),
- 4 images en variant l'angle de polarisation du capteur (*Polarization State Analyzer*).

Notre système réalise déjà 3 images de condition de polarisation linéaire. L'ajout d'une quatrième caméra permettra de mesurer la polarisation circulaire. Enfin, il reste à commander un éclairage afin de faire varier son angle de polarisation.

À partir de la mesure de la matrice de Mueller, QUÉAU et al. calculent des images de dépolarisation qu'ils utilisent pour identifier la classe de différents matériaux [Qué+19]. Les résultats qu'ils obtiennent sont encourageants et l'approche de l'exploitation des données brutes par apprentissage est similaire à la notre. L'introduction de la polarimétrie pour le contrôle des défauts sur les pièces plastiques ouvre un large champ de recherche.



# 3

## Modélisation de la notion de qualité par apprentissage

---

*Dans ce chapitre, nous étudierons les différentes méthodes de la littérature permettant de modéliser la notion de qualité par apprentissage statistique. L'objectif est de construire un modèle qui permet à partir d'images, d'inférer la qualité des pièces. Nous discuterons de l'optimisation de l'apprentissage de ces modèles. Le diagramme 3.1 synthétise cette démarche. C'est sous le prisme des contraintes industrielles que nous discuterons des limites au déploiement de ces méthodes sur les lignes de production.*

---

### Sommaire

---

<b>3.1 Construction du jeu de données d'apprentissage à partir de l'expertise humaine</b> . . . . .	<b>75</b>
3.1.1 Exploration du procédé par plan d'expériences . . . . .	75
3.1.2 Apprentissage en ligne . . . . .	76
<b>3.2 Annotation humaine du jeu de données en contexte industriel</b> . . . . .	<b>78</b>
3.2.1 Annotation hors ligne . . . . .	79
3.2.2 Annotation en ligne de production . . . . .	81
3.2.3 Intégration de l'humain dans le système d'apprentissage . . . . .	81
3.2.4 Assistance à l'annotation humaine . . . . .	82
<b>3.3 Construction d'une métrique de la qualité par apprentissage</b> . . . . .	<b>84</b>
3.3.1 Apprentissage supervisé . . . . .	84
3.3.2 Apprentissage par transfert de domaine . . . . .	103
3.3.3 Apprentissage semi-supervisé . . . . .	104
3.3.4 Apprentissage non-supervisé . . . . .	112
<b>3.4 Optimisation automatique des hyper-paramètres d'un modèle</b> . . . . .	<b>129</b>
3.4.1 Optimisation par recherche aléatoire . . . . .	130
3.4.2 Optimisation par plan d'expériences . . . . .	131
3.4.3 Optimisation par métamodèle bayésien . . . . .	132
3.4.4 Méthodes de réduction du coût d'évaluation du métamodèle . . . . .	134

---

3.4.5	Optimisation par algorithmes bandits . . . . .	136
3.4.6	Optimisation par algorithmes évolutionnistes . . . . .	136
3.4.7	Optimisation de l'architecture de réseaux de neurones profonds . . . . .	137
3.4.8	Conclusion : AutoML pour l'industrie . . . . .	137
<b>3.5</b>	<b>Conclusion : Modélisation par apprentissage pour l'industrie . . . . .</b>	<b>138</b>

---

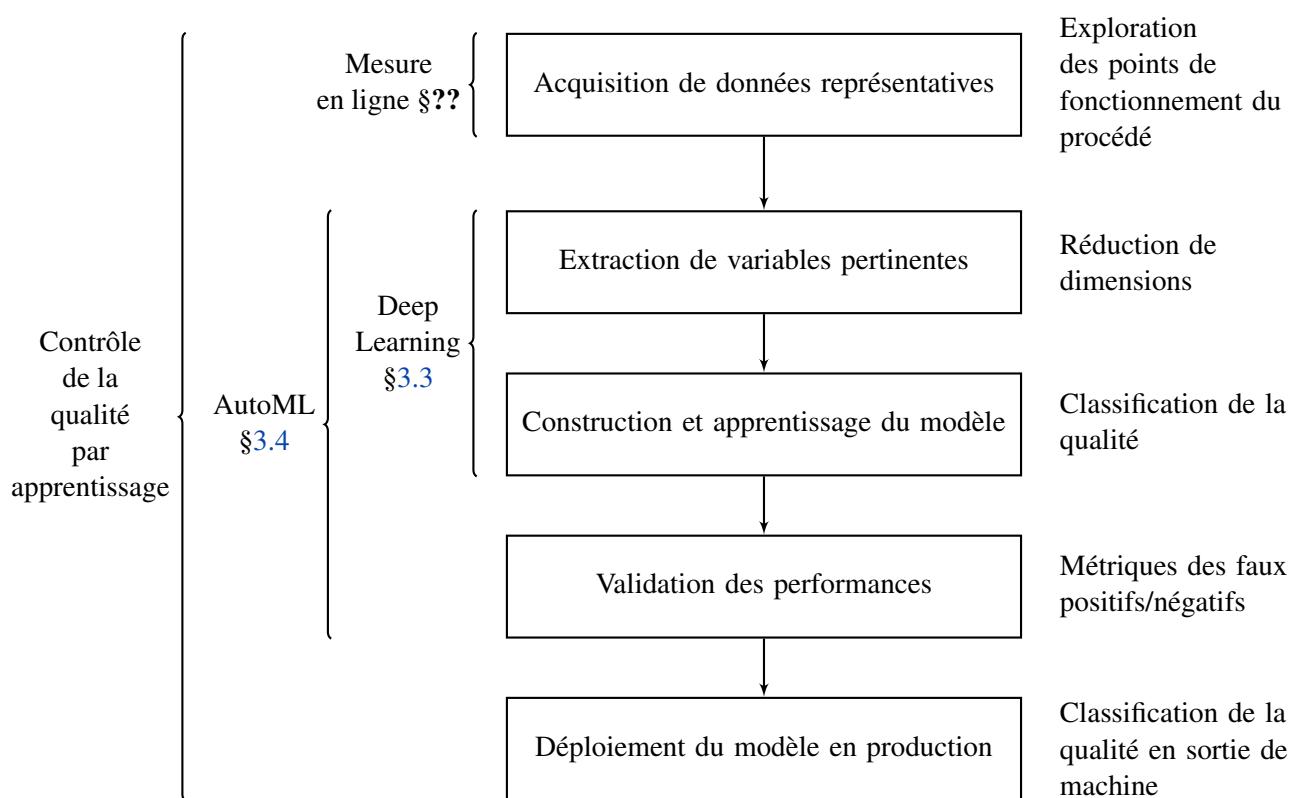


Figure 3.1 – Diagramme du processus de modélisation de la qualité par apprentissage.

### 3.1 Construction du jeu de données d'apprentissage à partir de l'expertise humaine

Tout système d'apprentissage statistique nécessite un jeu de données représentatif du modèle à construire. Ces travaux nous ont permis d'évaluer différentes méthodes de construction du jeu de données d'apprentissage vis à vis des contraintes industrielles. Le Tableau 3.1 présente les méthodes que nous avons évaluées.

Méthode	Avantage	Inconvénient
Plan d'expérience	Exploration de la plage de réglages	Difficile à intégrer dans la démarche de production
En ligne	Pendant la phase de réglage du procédé	L'exploration dépend de la variabilité du procédé
Hors ligne	Pendant la phase de réglage du procédé	Nécessite une traçabilité et un stockage rigoureux

Tableau 3.1 – Méthodes de construction du jeu de données d'apprentissage.

#### 3.1.1 Exploration du procédé par plan d'expériences

Pour construire un jeu de données représentatif du procédé d'injection-moulage et de la pièce en présence, il est nécessaire de parcourir la plus grande variété possible de réglages machine. C'est l'intérêt des plans d'expériences. L'analyse des résultats obtenus permet d'obtenir une modélisation de la réponse en fonction des variables de réglages. La réponse ici étudiée est la qualité finale de la pièce. Les variables sont les réglages disponibles sur la machine. Le modèle obtenu permet de mettre en évidence les variables les plus influentes sur la réponse. Il est alors possible de négliger les variables qui n'ont pas ou peu d'influence. Nous n'avons cependant pas choisi d'exploiter cette information car nous cherchons à produire un moyen de mesure de la qualité adapté à différentes machines. Le nombre d'essais réalisé dans le plan d'expériences est défini par le nombre de variables, le nombre de niveaux choisis et le choix de la construction du plan. Dans le cas des plans factoriels, le nombre d'essais à réaliser augmente rapidement. Dans le cadre d'une première série d'essais, il est intéressant de choisir un plan orthogonal à deux niveaux qui comporte peu d'essais. Cela permettra d'étudier les interactions sous forme linéaire, afin de pouvoir affiner le modèle dans un plan suivant.

En amont du travail de modélisation, il est nécessaire de sélectionner les variables à étudier. L'expertise des partenaires du projet FUI SAPRISTI a permis de retenir 11 variables pertinentes à prendre en compte dans l'étude du procédé d'injection-moulage. Ces variables sont celles qui influencent le plus la qualité des pièces plastiques. Le Tableau 3.2 référence les variables de réglage du procédé sélectionnées. Les variables "critiques" ne sont pas étudiées dans le plan d'expériences. En effet, leur bon réglage est nécessaire pour produire une pièce complète. C'est pourquoi le réglage est réalisé en amont du plan d'expérience. Suite à la sélection des variables étudiées, un ordre de priorité est défini, d'après les connaissances des experts. Le procédé d'injection-moulage est séquentiel, c'est pourquoi il est possible de régler certaines consignes sous la forme de séquences temporelles ; par exemple des paliers successifs de pression de maintien, en fonction de l'avancement du cycle d'injection. En pratique, sur notre pièce simple, cette possibilité n'est pas réalisée. En injection multi-buses séquentielle, le réglage des séquences temporelles est à contrario crucial pour obtenir une pièce de bonne qualité. Le nombre de valeurs de réglages de ces paramètres séquentiels est très élevé, c'est pourquoi les plans d'expériences ne sont plus adaptés à leur étude exhaustive. Un point de fonctionnement doit être initialement trouvé : c'est l'étape du réglage initial. Elle est réalisée par le technicien régleur. Par la suite, une exploration du procédé pourra être réalisée autour de ce point de fonctionnement.

Nous avons conduit deux expérimentations par plans d'expériences sur des machines

Priorité		Variable de réglages	Identifiant	Niveau -1	Niveau 0	Niveau 1
1	Critique	Vit. de rentrée des éjecteurs	/	/	optimale	/
1	Critique	Durée de maintien	/	15 s	15 s	15 s
1	Critique	Débit thermorégulateurs	/	/	optimal	/
1	Critique	Course d'ouverture	/		optimale	/
1	Testée	Température fourreau	T-four	220°C	240°C	260°C
2	Testée	Temp. thermo-régulateurs	T-therm	20°C	40°C	60°C
3	Testée	Pression de maintien	P-main	100 hPa	200 hPa	300 hPa
4	Testée	Vit. d'ouverture outillage	v-ouve	100 mm/s	110mm/s	120mm/s
5	Testée	Course d'injection	C-com	80 cm <sup>3</sup>	83 cm <sup>3</sup>	86 cm <sup>3</sup>
6	Testée	Vitesse d'injection	v-inj	50 cm <sup>3</sup> /s	70 cm <sup>3</sup> /s	90 cm <sup>3</sup> /s
7	Testée	Contre-Pression	C-ouv	30 hPa	50 hPa	70 hPa
8	Testée	Délais de dosage	t-reta	1 s	5 s	10 s
9	Testée	Vitesse de rotation de vis	v-rot	50 tr/min	100 tr/min	150 tr/min
10	Testée	Force de fermeture	P-ferm	1000 kN	1500 kN	2000 kN
16	Testée	Course de dosage	C-dos	90 cm <sup>3</sup>	100 cm <sup>3</sup>	110 cm <sup>3</sup>

Tableau 3.2 – Sélection des variables pour l'étude.

industrielles. Dans un premier temps, un plan d'expérience de criblage L12 a été utilisé : 11 variables sur 2 niveaux [PB46]. La Section 3.4.2 discute des constructions possibles de plans d'expériences et de l'intérêt des plans de criblage. Nous avons ajouté deux essais au centre pour étudier la répétabilité du procédé et deux essais de validation. Notre plan est détaillé dans le Tableau 3.3. La réalisation pratique de ce plan d'expériences a demandé 4 journées de 8 heures. Il a été choisi d'attendre la stabilisation du procédé avant de réaliser les mesures. La production de 20 à 40 pièces est requise pour être certain de la stabilité. Nous n'avons pas mesuré les 30 premières pièces produites lors du début de chaque expérience. Du point de vue de l'organisation des essais, la contrainte principale est la capacité thermique du moule. Aux vues de la masse d'acier en présence, la montée et la descente en température de l'outillage est très longue. C'est pourquoi nous avons ordonné les essais pour réaliser une montée progressive de la température de l'outillage<sup>1</sup>.

### 3.1.2 Apprentissage en ligne

Pour maximiser son adoption, un système de mesure de la qualité ne doit pas modifier le moyen de production en place : il doit être non invasif. Il doit s'intégrer au procédé et aux pratiques métiers. La réalisation de plans d'expériences, afin de modéliser le procédé, est difficile à intégrer dans la pratique industrielle. Les plans d'expériences sont conçus pour limiter le nombre d'essais, mais malgré cela, il est admis que la seule utilisation du savoir-faire du technicien-régulateur est plus avantageuse pour régler le procédé. Sur les pièces simples que nous avons pu étudier dans le cadre

1. En parallèle de ce travail de doctorat, COLLOMB a récemment proposé de réduire au maximum la capacité thermique de l'outillage et d'isoler les parois. Cela a permis de réduire le temps de cycle du moulage des composites [Col18]. Il serait intéressant d'étudier cette démarche, par exemple en tenant compte les déformations de l'outillage sous les contraintes pour l'alléger.

	T-four	T-therm	P-main	v-ouve	C-com	v-inj	C-ouv	t-reta	v-rot	P-ferm	C-dos
Essai01	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
Essai02	-1	-1	-1	-1	-1	1	1	1	1	1	1
Essai03	-1	-1	1	1	1	-1	-1	-1	1	1	1
Essai04	-1	1	-1	1	1	-1	1	1	-1	-1	1
Centré1	0	0	0	0	0	0	0	0	0	0	0
Essai05	-1	1	1	-1	1	1	-1	1	-1	1	-1
Essai06	-1	1	1	1	-1	1	1	-1	1	-1	-1
Essai07	1	-1	1	1	-1	-1	1	1	-1	1	-1
Essai08	1	-1	1	-1	1	1	1	-1	-1	-1	1
Centré2	0	0	0	0	0	0	0	0	0	0	0
Essai09	1	-1	-1	1	1	1	-1	1	1	1	-1
Essai10	1	1	1	-1	-1	-1	-1	1	1	-1	1
Essai11	1	1	-1	1	-1	1	-1	-1	-1	1	1
Essai12	1	1	-1	-1	1	-1	1	-1	1	1	-1
Centré3	0	0	0	0	0	0	0	0	0	0	0
Test1	1	1	1	-1	-1	1	-1	-1	1	1	-1
Test2	-1	-1	-1	1	1	-1	1	1	-1	-1	1

Tableau 3.3 – Plan de criblage L12 utilisé.

de ce travail, le technicien régulateur est capable de régler le procédé d'injection pour obtenir une pièce de qualité satisfaisante en une trentaine de minutes, soit 30 pièces produites pour un temps de cycle de production d'une minute par pièce. Deux solutions sont possibles :

- Utiliser la phase de réglage, pilotée par le technicien-régulateur, afin de recueillir des données variées.
- Automatiser la phase de réglages machines.

### 3.1.2.1 Exploration humaine en cycle de production

Nous avons conduit deux expérimentations sur des machines industrielles, sans planification préalable des configurations de réglages de la machine. Ces essais ont permis d'obtenir des retours sur la faisabilité et la fiabilité de notre dispositif de mesure de la qualité. Afin d'explorer les plages de fonctionnement des réglages du procédé, nous recommandons d'acquérir les données en particulier pendant ces évènements :

- pendant la phase de réglage,
- pendant les défaillances du procédé,
- pendant que certains asservissements du procédé sont arrêtés et que le procédé dérive.

Ces recommandations permettent de maximiser la variance du jeu de données. Nous notons que c'est pendant la phase de réglages que la variance des données est maximale. En effet, le réglage de la machine est un processus itératif qui nécessite d'appliquer successivement plusieurs points de fonctionnement au procédé. Les 20 pièces produites pendant le réglage possèdent des caractéristiques de qualité variées.

### 3.1.2.2 Pilotage automatique du procédé pour la création d'un jeu de données

Nous définissons le pilotage d'un procédé comme la sélection d'un point de fonctionnement initial (phase de réglage), son optimisation et sa régulation. Le réglage est aujourd'hui réalisé par un opérateur humain. Dans le cas où la phase de réglage est automatisée, il devient possible d'explorer le procédé, tout en optimisant les réglages, à l'aide des méthodes d'optimisation que nous présentons dans la Section 3.4. Les plans d'expériences §3.4.2 sont particulièrement adaptés à l'acquisition de données sur l'ensemble de l'espace. Il est possible de compléter un plan d'expériences par de nouveaux essais, dans les plages de réglages qui possèdent les meilleures performances.

## 3.2 Annotation humaine du jeu de données en contexte industriel

C'est par l'annotation d'un jeu de données que l'expert humain transfert sa connaissance au système informatique.

Nous distinguons deux méthodes d'annotation humaine d'un jeu de données d'apprentissage :

- l'annotation "hors ligne", qui est effectuée à posteriori de la production des données,
- l'annotation "en ligne", qui est effectuée pendant la production des données.

Dans le cadre de la pratique industrielle, on parlera d'annotation "en ligne" lorsqu'elle est réalisée au fur et à mesure de la production des pièces. Dans ce cas, l'annotation est effectuée directement depuis la ligne de production. À contrario, l'annotation hors ligne peut être réalisée en dehors de la ligne de production.

Méthode	Durée par classe	Qualité de localisation	Illustration de l'annotation
Label de présence	1,0 s	Aucune localisation	
Label + point	2,4 s	Localisation globale	
Label + rectangle	10 s	Localisation locale	
Label + masque	78 s	Segmentation précise	

Tableau 3.4 – Durées d’annotations moyenne d’une image pour un humain.

Une des principales limites au déploiement de système par apprentissage est le coût de l’annotation. En effet, le coût de l’annotation humaine est conséquent car il monopolise un expert du domaine. En 2015, BEARMAN et al. évalue les durées moyennes de différents types d’annotations [Bea+15]. Le Tableau 3.4 présente ces résultats. Dans le cadre industriel, la durée d’annotation définit le coût économique du travail de l’expert qualité qui la réalise. Ainsi, il n’est pas envisageable de réaliser une segmentation précise depuis une ligne de production. Nous discuterons dans la suite de la faisabilité de l’annotation d’un jeu de données dans un contexte industriel.

### 3.2.1 Annotation hors ligne

Nous proposons une méthode de notation de la qualité basée sur l’expertise humaine. De précédents travaux sur l’évaluation visuelle de la qualité par de multiples experts ont montré la difficulté d’accorder les avis d’experts humains [BPM11]. BAUDET, PILLET et MAIRE étudie l’évaluation de la qualité de 18 produits, par un panel de cinq contrôleurs humains. Les

résultats de l'évaluation de ce panel possédaient une variance maximale de 50% sur l'ensemble des niveaux de qualité. Cette variance est très élevée. Elle n'est pas acceptable pour un moyen de mesure. Cette étude montre qu'il est nécessaire de formaliser l'évaluation. Il faut également former les contrôleurs humains pour harmoniser les évaluations. Afin d'évaluer la variabilité des évaluations humaines, BAUDET [Bau12] propose une évolution de la méthode *R&R* (*Repeatability & Reproductibility*) qui a été initialement proposée par le GROUP et al. pour évaluer la variabilité des moyens de mesure [Gro+10]. La variabilité peut avoir deux causes : un problème de répétabilité de la mesure ou un problème de reproductibilité. Pour que le système de mesure soit conforme et puisse être utilisé, la pratique industrielle accepte généralement une variabilité inférieure à 30%. Dans notre cas, nous chercherons à concevoir un système de contrôle qui réalisera moins de 10% d'erreurs de classification.

### Expérimentation de l'annotation hors ligne

Sur le cas d'application "boîte noire", présenté dans le Chapitre 2, nous avons expérimenté une méthode de construction d'une métrique de la qualité des pièces. L'échelle de notation est construite au fur et à mesure, par comparaison entre les niveaux de défauts des pièces. Les pièces sont classées sur une grande table, par comparaison manuelle. Afin d'obtenir une évaluation répétable et reproductible de la qualité, ce classement est réalisée par quatre experts humains qui débattent jusqu'à être d'accord. Nous justifions cette démarche par la nécessité de limiter la dispersion inter-classes. Si nous ne sommes pas capables de différencier deux pièces, alors ces pièces appartiennent à la même classe. Nous n'avons pas réalisé d'analyse de répétabilité *R&R* du panel de quatre experts. C'est une perspective intéressante pour des travaux complémentaires.

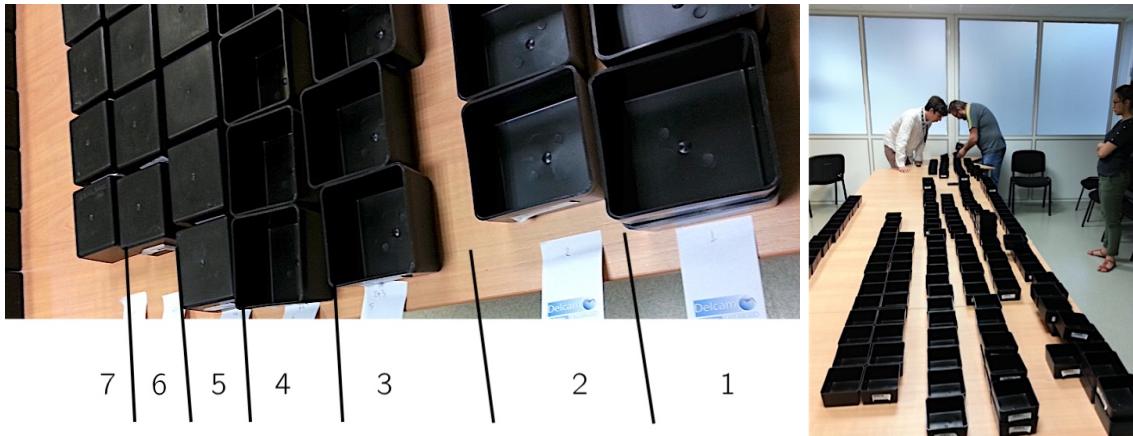


Figure 3.2 – Annotation hors-ligne par les experts humains.

Les photographies de la Figure 3.2 présentent la réalisation de l'annotation hors-ligne. Les numéros sous les colonnes de pièces indiquent la valeur attribuée à chaque colonne sur l'échelle de qualité. L'échelle a été construite de manière dynamique. Elle va de 1 —niveau de qualité le meilleur —à 7 —les pièces qui possèdent les pires caractéristiques.

Il s'est avéré difficile pour les contrôleurs humains de donner une note qui prend en compte plusieurs types de défauts. Les contrôleurs experts possèdent un a-priori de la notion de qualité qui s'appuie sur des types de défauts. C'est pourquoi nous avons identifié trois défauts présents sur les : la retassure, le givrage et les brûlures. La démarche de notation par comparaison a été réalisé trois fois de suite, pour les trois types de défauts différents. Une moyenne géométrique des notes combine les trois types de défauts et produit l'évaluation finale de la qualité de chaque pièce. Nous identifions trois limites à ce processus d'annotation humaine :

- La durée de 3 heures avec 3 experts pour évaluer 200 pièces, soit 60 pièces par heure. Cette cadence est équivalente au temps de cycle de production d'une pièce par heure.

- La surface de travail nécessaire pour organiser les pièces et que les trois experts puissent les manipuler librement.
- Les risques liés à la détérioration des pièces entre leur production et l'étape d'annotation.

C'est pourquoi, nous recommandons de réaliser l'annotation au plus près du procédé d'injection-moulage ; si possible dès la sortie des pièces du moule d'injection.

### 3.2.2 Annotation en ligne de production

Réaliser l'annotation en ligne de production permet d'éliminer deux risques d'erreurs critiques sur le jeu de données d'apprentissage :

- la modification des pièces entre leur production et leur annotation
- l'erreur d'annotation à posteriori : l'expert peut ici observer la pièce

En contrepartie, la durée disponible pour l'annotation est très limitée. En plus de la durée de l'annotation, l'expert humain doit assister les opérations de manutention des pièces et les opérations connexes. Nous estimons la disponibilité de l'expert humain à une trentaine de secondes pour un temps de cycle de procédé d'une minute.

Le Tableau 3.4 présente les durées moyenne de la tâche d'annotation. Dans le cas d'une annotation au fil de la production de pièces, il n'est pas envisageable de réaliser une segmentation précise, à cause de sa durée élevée. Nous obtenons de ces résultats l'intérêt de réaliser une annotation rapide, sans nécessairement chercher à localiser les défauts. Indiquer la présence ou l'absence de défauts est rapide. Ensuite, nous pouvons préciser le niveau de la qualité en proposant une échelle discrète simple, par exemple à cinq niveaux de qualité. Afin de répondre à ce problème de vitesse d'annotation, une interface logicielle a été spécifiquement développée. Nous discutons de cette interface dans le Chapitre 4.

### 3.2.3 Intégration de l'humain dans le système d'apprentissage

L'expertise humaine est indispensable à la modélisation par apprentissage statistique. L'humain doit transférer ses connaissances holistiques au logiciel informatique. C'est pourquoi, il nous paraît indispensable d'intégrer l'humain au plus près du processus d'apprentissage. Cela nécessite une interface adaptée. Par ce biais, l'expert humain peut effectuer deux actions :

- identifier la cause du défaut.
- alerter si le modèle statistique donne une information erronée.

Afin de motiver l'expert humain à transmettre ses connaissances au système informatique, la ludification du processus d'annotation peut aussi être un facteur important. De nombreuses sociétés informatiques l'ont bien compris et nous annotons actuellement de manière naturelle et transparente des bases de données d'apprentissage : par le biais de notre navigation sur l'Internet ou des interactions avec notre smartphone, et de toutes les données recueillies par les capteurs connectés. Le déploiement d'un système d'évaluation de la qualité dans les ateliers de production sera facilité par l'attrait de l'interaction avec l'interface d'annotation. Nous avons développé une solution logicielle afin de résoudre ce problème. Une interface sur navigateur web est accessible depuis tous les terminaux. Tout au long de la production de nouvelles pièces, celles-ci s'affichent en temps réel et l'expert humain peut annoter la qualité. L'annotation peut être réalisée depuis la ligne de production, ce qui permet le contrôle visuel direct des pièces. L'annotation peut également être effectuée à distance : dans ce cas, l'expert humain se base sur l'image enregistré par le dispositif de mesure pour évaluer la qualité. Dans le cas de pièces techniques à haute valeur ajoutée, la présence de l'expert sur la ligne est recommandée.

### 3.2.4 Assistance à l'annotation humaine

La disponibilité de l'expert humain est limitée. Il est possible de l'assister dans la tâche d'annotation. Il est également possible de limiter le nombre d'annotations à réaliser aux cas les plus critiques. Deux démarches peuvent être réalisées en parallèle : suggérer et détecter les pièces nouvelles, qui sont pertinentes à annoter.

#### 3.2.4.1 Suggestion d'annotations

Dans certains cas, l'expertise humaine ne permet pas de différencier la qualité entre deux pièces. Plusieurs limites humaines peuvent être la cause d'une telle situation :

- L'expert ne peut pas mémoriser les caractéristiques de l'ensemble des pièces produites.
- l'expert humain ne peut pas comparer les pièces car les pièces ne lui sont plus accessibles.
- L'expert humain ne dispose pas d'assez de temps pour comparer les pièces.

Il est alors intéressant de suggérer à l'expert humain une annotation. De plus, si cette suggestion est motivée par un argumentaire, elle sera d'autant plus crédible. Il ne faut néanmoins pas influencer l'expert humain. C'est pourquoi cette suggestion ne doit être effectuée que lorsque le niveau de confiance du modèle statistique est élevée. Des les autres cas, il vaudra mieux écarter la pièce incertaine du jeu de données d'apprentissage. Nous recommandons une certitude de classification supérieure à 90% pour suggérer une annotation.

Pour suggérer la catégorie de la pièce, les méthodes de *clustering* permettent d'identifier des paquets de pièces similaires. Pour suggérer, il faut également justifier. Il est nécessaire de transmettre à l'expert humain l'information. C'est pourquoi, l'affichage sous forme d'un graphique des paquets de pièces similaires est utile. Les algorithmes *ACP* 3.3.4.3, *t-SNE* 3.3.4.5 ou *UMAP* 3.3.4.6 permettent de projeter un espace de grandes dimensions dans un graphique à deux ou trois dimensions. Nous détaillons leur application dans la Section 3.3.4. Ce graphique peut être interprété par un humain. Sur ce graphique, nous ajoutons l'affichage automatique de l'image de chaque pièce lors du survol des points. Cela permet d'identifier les paquets de pièces similaires très rapidement, mais aussi les erreurs de classifications (humaine où du système automatique).

#### 3.2.4.2 Détection de nouveautés

Dans le cas spécifique des productions industrielles sur des procédés stables, la majorité des pièces produites sera de bonne qualité. C'est pourquoi, la tâche d'annotation de toute la production industrielle sera laborieuse pour l'expert humain et aura peu d'intérêt pour le modèle. Aussi, la variabilité maximale sur les procédés industriels est produite lors de la phase de réglage du procédé. Il est capital d'annoter les pièces issues de cette phase. Par la suite, les pièces seront similaires et de bonne qualité. De plus, différents points de fonctionnement du procédé peuvent amener à la production de pièces similaires, car le procédé d'injection moulage est non-bijectif, à cause de la multiplicité de ses paramètres réglables et de leurs réponses non-linéaires. Il est ainsi nécessaire de mettre en évidence les pièces qui présentent de l'intérêt pour la base d'apprentissage : c'est à dire les pièces qui sont "nouvelles". Des méthodes d'analyses de données sans supervision permettent de réaliser cette tâche. Nous détaillons l'application de ces méthodes dans la Section 3.3.4. Nous présentons ici les deux méthodes que nous avons implémentées, la mesure de distance de Mahalanobis et la machine à vecteurs de support à une classe.

#### Mesure de distance dans un espace réduit par l'Analyse en Composante Principale

Il s'agit de réduire l'espace des données en projetant les valeurs dans un espace de dimensions réduite. L'Analyse en Composantes Principales §3.3.4.3 réduit l'espace des données initial en combinant les relations linéaires existantes entre les composantes. On choisit la dimension de l'espace de donnée réduit est choisi afin de limiter la perte d'informations à partir de l'espace

initial. Cette perte d'information est mesurée par l'écart de la variance exprimée. Dans notre cas spécifique, la dimension des données d'entrés sont des images qui comportent près d'un million de pixels. Nous projetons ces valeurs dans un espace de dimensions 100. La perte de variance est alors inférieure à 0,01%.

Dans cet espace de données réduit, il devient possible de mesurer la distance entre les points. La distance euclidienne ne prend pas en compte les corrélations entre les variables, c'est pourquoi la distance de Mahalanobis [Mah36], Équation 3.1 est couramment employée pour mesurer des distances entre paquets d'échantillons. La distance de Mahalanobis entre un paquet de pièces de références et une pièce nouvelle est représenté par une ellipse.

$$D_{\text{Mahalanobis}}(x, y) = \sqrt{(x - y)^T \Sigma^{-1} (x - y)} \quad (3.1)$$

Pour obtenir un détecteur de nouveauté, il reste à définir un seuil à partir duquel la pièce est considérée comme mauvaise. Ce seuil est soit imposé, soit défini de manière automatique. Dans le cas d'un seuil automatique, on choisira un seuil égal à la distance de Mahalanobis de l'ensemble des pièces actuellement dans la base d'apprentissage ; ce sont les pièces existantes et nous cherchons à détecter les nouvelles. La distance de toute pièce nouvelle dépassera alors ce seuil et la nouveauté sera détectée. Cette démarche est employée avec succès dans les travaux de LACOMBE sur le contrôle visuel automatique par système multi-éclairage [Lac18].

### Machine à Vecteurs Supports à une classe

À partir d'un classifieur machine à vecteurs de support §3.3.1.3, SCHÖLKOPF et al. proposent de chercher à classifier l'ensemble des données existantes dans une unique classe [Sch+99 ; Sch+01]. C'est la méthode *One-Class SVM*. À l'aide de noyau gaussien, cette méthode permet de prendre en compte des données aux propriétés plus complexes, que la matrice de covariance précédente de la distance de Malhanobis. Le paramètre de marge  $b$  correspond à la probabilité de classifier un nouvel échantillon, qui serait situé en dehors de l'hyperplan de marge maximal, comme nouveau.

### Détection de modifications sur le procédé

Une approche complémentaire, pragmatique, à ce problème est de réaliser la détection de l'intervention du technicien humain sur les réglages du procédé. En effet, lorsque le technicien modifie les réglages, il cherche nécessairement à optimiser la qualité des pièces produites. Il suffit alors d'ajouter à la base de données d'apprentissage les pièces pour lesquelles une intervention humaine sur les réglages a eu lieu. De manière pratique, le système d'apprentissage doit disposer d'une mémoire tampon où quelques dernières pièces produites sont conservées. En cas d'intervention humaine sur les réglages du procédé, cette mémoire tampon est ajoutée à la base d'apprentissage. Nous ajouterons également un certain nombre de pièces produites à la suite avec les nouveaux réglages. Nous proposons d'ajouter un nombre de pièces qui correspond à la variabilité court-terme du procédé. Dans le cas de l'injection des thermoplastiques, suite à un réglage, une dizaine de pièces permettent de stabiliser le procédé. C'est pourquoi nous proposons d'enregistrer deux pièces précédant la modification des réglages, ainsi que les dix pièces suivant la modification. Cette contrainte dimensionne la capacité de mémoire informatique du système d'apprentissage. Dans le cas de douze pièces, la mémoire tampon sera de l'ordre de mille mégaoctets.

Nous n'avons pas pu évaluer notre proposition dans le cadre de ce travail, car cela nécessite l'enregistrement automatique de la valeur des réglages de la machine. Cette intégration doit être réalisée en lien avec les constructeurs des presses d'injection-plastique ; ou par rétro-ingénierie, ce qui demandera des ressources conséquentes. L'évaluation de cette proposition est une piste de travail importante pour la suite de nos travaux.

### 3.3 Construction d'une métrique de la qualité par apprentissage

La notion de qualité d'un produit est spécifiée dans le cahier des charges. Cependant, lorsque les spécifications ne sont pas normées, elles entraînent souvent une traduction subjective. C'est pourquoi la notion de qualité d'un produit, lorsqu'elle intègre les aspects sensoriels en plus des tolérances géométriques, est souvent subjective. Elle s'appuie sur une connaissance holistique d'un expert qualité humain. La notion de qualité d'aspect est difficile à formaliser. Des travaux de recherches ont permis des avancées dans ce domaine. L'analyse ontologique de corpus de textes de références permet par exemple de hiérarchiser les descriptions formelles afin de normaliser ces notions de qualité du point de vue visuel [Bau12] et sensorimoteur [Alb+17]. Une fois normée, le contrôle de la qualité peut être automatisé [Dés15; Pit16; Lac18; Alb19] afin de garantir la répétabilité.

Construire une métrique de la qualité par apprentissage revient à apprendre la fonction  $f_\theta$  qui a pour paramètres  $\theta$  tel que  $f_\theta(\mathbf{x}) : \mathbb{R}^F \rightarrow \mathbb{R}^D$  transforme les points de l'espace d'origine  $\mathbb{R}^F$  dont les propriétés sont similaires (par exemple un niveau de qualité de pièces équivalent) vers des points proches au sens de la métrique, dans l'espace  $\mathbb{R}^D$ . Réciproquement, la fonction  $f_\theta$  doit transformer les points d'origines dont les propriétés sont opposées, vers des points distants de l'espace  $\mathbb{R}^D$ . La transformation  $f : X \times X \rightarrow \mathfrak{R}_0^+$  dans un espace vectoriel  $X$  est une métrique si  $\forall \vec{\mathbf{x}}_i, \vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k \in \chi$  :

1.  $f(\vec{\mathbf{x}}_i, \vec{\mathbf{x}}_j) + f(\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k) \geq f(\vec{\mathbf{x}}_i, \vec{\mathbf{x}}_k)$ , inégalité triangulaire
2.  $f(\vec{\mathbf{x}}_i, \vec{\mathbf{x}}_j) \geq 0$ , inégalité non négative.
3.  $f(\vec{\mathbf{x}}_i, \vec{\mathbf{x}}_j) = f(\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_i)$ , symétrie.
4.  $f(\vec{\mathbf{x}}_i, \vec{\mathbf{x}}_j) = 0 \iff \vec{\mathbf{x}}_i = \vec{\mathbf{x}}_j$ , identité des indiscernables.

Dans le cas de la construction d'une métrique par apprentissage, il est probable que l'identité des indiscernables ne soit pas remplie. Aussi, nous cherchons dans ce travail à apprendre une pseudo-métrique et non pas une véritable métrique. La littérature sur ce sujet est vaste, c'est pourquoi ce travail de doctorat n'a pas permis d'explorer de manière exhaustive l'ensemble des méthodes d'apprentissage de métriques proposées dans la littérature. Un excellent état de l'art sur le sujet est le doctorat de BELLET [Bel12]. Nous nous sommes intéressés aux méthodes les plus pertinentes compte-tenu de notre problématique d'analyse d'images. Ce travail de doctorat a permis de comparer différents algorithmes sur notre problème de classification binaire de la qualité des pièces. Pour cette étude, nous nous sommes uniquement intéressés à l'utilisation de données de mesure sous forme d'images numériques. Nous présenterons les algorithmes supervisés étudiés, puis les algorithmes semi-supervisés qui permettent de limiter le temps d'annotation nécessaire. Enfin, nous présenterons les moyens disponibles pour optimiser les performances des classificateurs utilisés par sélection automatique des hyper-paramètres.

#### 3.3.1 Apprentissage supervisé

Sur notre problématique du contrôle qualité, nous cherchons ici à résoudre un problème de classification (aussi appelé "discrimination"). C'est à dire que nous cherchons à classer automatiquement dans des groupes les échantillons qui ont des propriétés similaires, mesurées sur des observations. Ici, nous cherchons à classer les pièces qui sont observées comme bonnes dans le groupe des pièces bonnes, et réciproquement classer les pièces observées mauvaises dans le groupe des pièces mauvaises. Cette fonction de classification est modélisée par l'Équation 3.2. Pour un vecteur d'observations  $\vec{\mathbf{x}}$  dans  $\mathbb{R}^N$ , la prédiction  $\hat{y}$  du classifieur  $f$  est donnée par :

$$\hat{y} = f(\vec{\mathbf{x}}) \quad (3.2)$$

Pour un classifieur,  $y$  est souvent un vecteur de la probabilité d'appartenance de l'échantillon à chacune des classes. La dimension de  $y$  est donc le nombre de classes possibles dans notre

problème. Le même principe peut être employé pour résoudre un problème de régression.  $y$  est alors un scalaire continu. Un cas particulier est le classifieur dit "linéaire", Équation 3.3. La fonction est alors le produit scalaire des observations avec la matrice de poids. Le terme linéaire signifie que la fonction est une combinaison linéaire.  $\mathbf{W}$  est alors une matrice de poids et  $\vec{\mathbf{b}}$  est un vecteur de biais. L'apprentissage supervisé consiste à trouver les valeurs de la matrice de poids  $\mathbf{W}$  et du biais  $\vec{\mathbf{b}}$  tel que la valeur de  $\hat{y}$  s'approche le plus possible de la valeur connue du couple  $(\mathbf{x}, y)$ .

$$\hat{y} = f(\mathbf{W}^T \cdot \mathbf{x} + \vec{\mathbf{b}}) = f\left(\sum_{i=1}^N \mathbf{w}_i \cdot \mathbf{x}_i + b_i\right) \quad (3.3)$$

Plusieurs distances mathématiques peuvent être utilisées pour estimer l'écart entre deux populations. La distance de MAHALANOBIS [Mah36] a été utilisée dans les travaux de doctorat de LACOMBE sur une problématique similaire de contrôle automatique de la qualité à partir de mesures visuelles [Lac18]. Cependant l'utilisation de la distance de MAHALANOBIS est limitée aux espaces linéaires, ce qui rend difficile la séparation de motifs visuels complexes. Il est alors nécessaire de transformer les données pour les rendre linéairement séparables, en amont de la mesure de distance. Nous détaillons l'utilisation de la distance de Mahalanobis dans la Section 3.3.3.3. Nous avons choisi dans ce travail de comparer des approches traditionnelles d'apprentissage supervisé, avec des approches par réseaux de neurones de convolutions profonds. L'intérêt de ces méthodes est la non-linéarité des modèles obtenus. Les réseaux de neurones de convolutions sont organisés en multiples couches successives, ce qui crée des architectures dites "profondes" (d'où l'expression *Deep Learning*). Ces types de classificateurs représentent aujourd'hui l'état de l'art académique et industriel en analyse d'images.

### 3.3.1.1 Métrique de performances d'un classifieur

Pour exprimer la performance d'un classifieur, de nombreuses métriques ont été proposées dans la littérature. Une métrique rend compte de l'écart entre la réponse donnée par un classifieur et la vérité. La métrique la plus couramment employée est la justesse définie par METZ [Met78], Équation 3.4. Cette métrique est adaptée au cas où de multiples classes  $j$  sont présentes dans le problème. Nous rapporterons ici toutes les métriques sur le nombre d'échantillons total.

$$\text{justesse}(y, \hat{y}) = \frac{1}{n_{\text{échantillons}}} \sum_{i=0}^{n_{\text{échantillons}}-1} \mathbb{1}(\hat{y}_j = y_j) \quad (3.4)$$

Le procédé d'injection-moulage des thermoplastiques est stable. Lorsqu'un point de fonctionnement satisfaisant pour obtenir une pièce de bonne qualité est obtenu. La dérive du procédé au cours du temps est de l'ordre de la journée. Ainsi, peu de pièces mauvaises sont produites. C'est pourquoi une population traditionnelle de pièce moulées comportera peu d'individus de mauvaise qualité et une majorité d'individus de bonne qualité. Ainsi, le choix de la métrique d'évaluation des performances doit tenir compte de la répartition non égalitaire des classes dans l'échantillon. La justesse ne prend pas en compte les répartitions inégalitaires de classes dans la population, c'est pourquoi ce n'est pas une métrique adaptée à notre problème [JS02].

#### Justesse équilibrée

La justesse équilibrée, Équation 3.5, est une métrique qui tient compte de ce problème en intégrant une matrice de poids  $\mathbf{W}$  qui compense les déséquilibres inter-classes [Bro+10; Mos13]. Dans le cas où la répartition des classes est équilibrée, la justesse équilibrée est égale à la justesse.

$$\text{justesse}_{\text{éq}}(y, \hat{y}, \mathbf{W}) = \frac{1}{\sum \hat{\mathbf{w}}_i} \sum_i 1(\hat{y}_i = y_i) \hat{\mathbf{w}}_i \quad (3.5)$$

avec  $\hat{\mathbf{w}}_i = \frac{\mathbf{w}_i}{\sum_j 1(y_j = y_i) > j}$

### Matrice de confusion

Dans notre cas industriel, il est crucial de limiter le nombre de faux positifs, c'est à dire le nombre de pièces mauvaises détectées comme bonnes. A contrario, les faux négatifs correspondent aux pièces bonnes détectées comme mauvaises. Sur des pièces moulées en plastique dont le coût de revient est faible, les faux négatifs ne sont pas aussi critiques que les faux positifs pour notre problématique industrielle car la relation client entre en jeu. En effet, les taux de pièces mauvaises acceptés dans les lots par le client sont, pour les plus drastiques des cahiers des charges, de 5 pièces défectueuses par millions. Dans ce cas, il vaudra mieux jeter des pièces bonnes que prendre le risque d'envoyer un lot contenant des pièces mauvaises. Si le coût de revient des pièces est élevé, alors faux positifs et faux négatifs sont à limiter : c'est le zéro défaut. Si la dimension écologique est prise en compte, on cherchera également à atteindre le zéro défaut : jeter des pièces bonnes est une hérésie écologique, qui plus est : produire des pièces mauvaises.

La littérature présente les nombres de faux positifs et faux négatifs sous la forme d'un Tableau 3.5, la matrice de confusion.

		Classes vraies	
		Vrais positifs (VP)	Faux Positifs (FP)
Classes prédites	Vrais négatifs (FN)	Faux Négatifs (VN)	

Tableau 3.5 – Matrice de confusion.

### Précision et rappel

La précision et le rappel sont définis à partir de ces valeurs, Équations 3.6. Elles renseignent sur la pertinence des résultats.

$$\begin{aligned} \text{précision} &= \frac{|y \cap \hat{y}|}{|y|} = \frac{\text{VP}}{\text{VP} + \text{FP}} \\ \text{rappel} &= \frac{|y \cap \hat{y}|}{|\hat{y}|} = \frac{\text{VP}}{\text{VP} + \text{FN}} \end{aligned} \quad (3.6)$$

### F-mesure

La moyenne harmonique de la précision et du rappel est appelé F-mesure. Le facteur  $\beta$  permet d'ajuster le rapport entre précision et rappel. Lorsque  $\beta = 1$ , précision et rappel sont également pondérés, on parle alors de F1-mesure.

$$\begin{aligned} F_\beta &= (1 + \beta^2) \frac{\text{précision} \cdot \text{rappel}}{\beta^2 \text{précision} + \text{rappel}} \\ F_1 &= 2 \frac{\text{précision} \cdot \text{rappel}}{\text{précision} + \text{rappel}} \end{aligned} \quad (3.7)$$

L'ensemble de ces métriques est généralisable par moyennage, dans le cas où plus de deux classes existent dans la population :

- micro-moyenne : les métriques sont calculées d'après les Équations combinatoires 3.6.
- macro-moyenne : les métriques sont calculées pour chaque classes, puis moyennées par l'ensemble des classes.

### Courbe ROC

La courbe ROC (*Receiver Operating Characteristic*) représente la caractéristique de performance d'un classifieur. Elle a été proposée en 1940 pour distinguer les vrais positifs du bruit de fond sur les relevés radars. La courbe compare sensibilité (taux de vrais positifs) et spécificité (taux de vrais négatifs) d'un classifieur. La performance est comparée à la réponse d'un classifieur aléatoire. Cette courbe peut servir à choisir la valeur optimale de séparation entre deux classes. Dans notre cas, une valeur milieu, non biaisée est satisfaisante. La courbe 3.3 est issue du meilleur classifieur de la qualité obtenue. L'écart-type est calculé par validation croisée, §3.3.1.2.

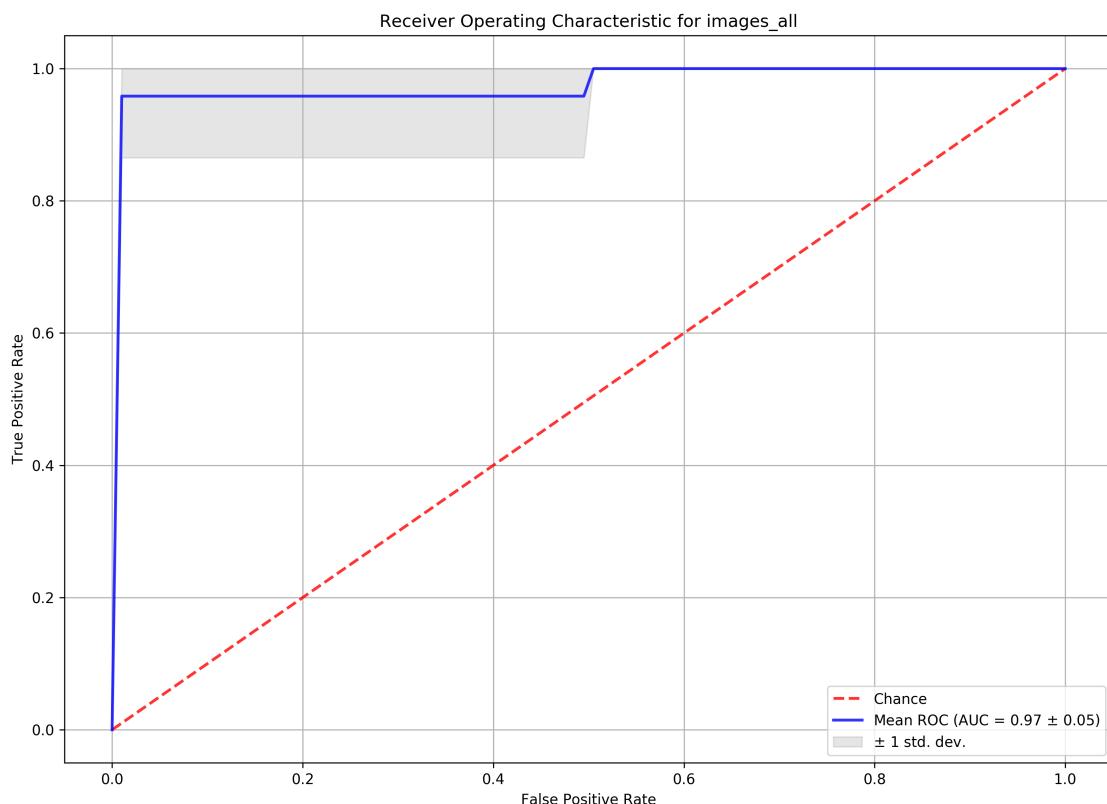


Figure 3.3 – Courbe ROC du meilleur classifieur de la qualité des pièces.

#### 3.3.1.2 Validation croisée pour la comparaison des performances de classificateurs

Pour comparer les performances de classificateurs, nous comparons la valeur des métriques définies précédemment. Les classificateurs sont appris sur un jeu de données d'apprentissage et la performance est mesurée sur un jeu de données de test, qui n'a pas été utilisé lors de l'apprentissage. On réserve traditionnellement 20 pour-cent du jeu de données pour le test. Une limite de cette méthode est le biais potentiel des deux portions. Dans le cas de grands jeux de données ( $> 2000$  par classes), le biais est naturellement limité par la variabilité des données. Dans le cas de jeu de données de plus petites dimensions, il est nécessaire de limiter ce biais. On utilise des méthodes de validations croisées. Il s'agit d'entraîner de multiples modèles sur des portions différentes des données, puis d'observer la moyenne et l'écart-type des performances. Nous présenterons ici la méthode des K-partitions stratifiées par classes (*Stratified K-Folds*),

utilisée dans nos travaux. D'autres méthodes existent mais nous retenons cette dernière par sa robustesse au jeu de données où les classes ne sont pas égales. Cette méthode est également extensible aux problèmes à classes multiples. La stratification consiste à associer dans chaque partition une proportion d'échantillons de chaque classe équivalente. KOHAVI conclue ainsi sur l'intérêt de la stratification, en matière de réduction du biais des modèles [Koh95]. Le nombre de partition  $k$  est choisi en fonction du nombre d'échantillons disponible. Dans nos problèmes où les jeux de données sont inférieurs à mille échantillons, nous fixons  $k = 3$ . Dans nos travaux, nous utiliserons l'implémentation de la librairie *Scikit-Learn* [Ped+11]. La Figure 3.4 représente l'effet de cette méthode sur un jeu de données.

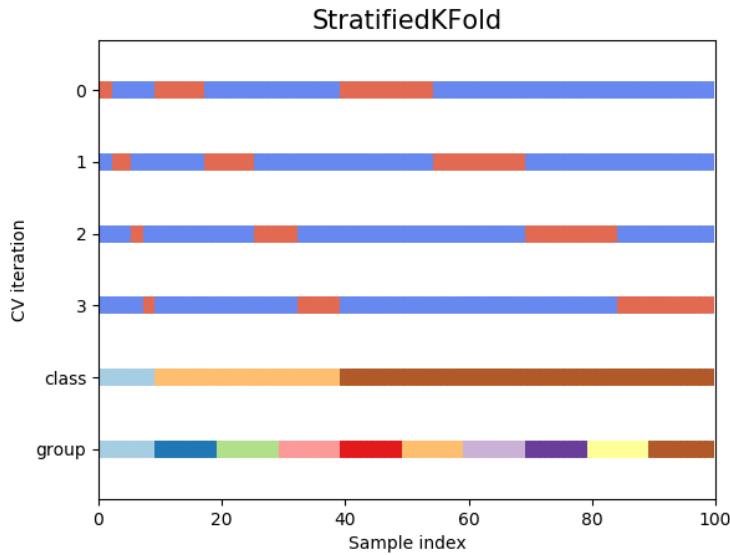


Figure 3.4 – Stratified 4-Folds. Figure extraite de la [doc SciKit-Learn, License BSD 3-clauses](#).

Lorsque l'on souhaite réaliser un optimisation du classifieur en lui-même (hyper-paramètres), une partition de données de validation est introduite. Chaque classifieur sera entraîné sur le jeu d'apprentissage, puis validé sur le jeu de validation. Enfin, le meilleur classifieur obtenu sera évalué sur le jeu de test, qui n'aura jamais été utilisé ni pendant l'apprentissage, ni pendant l'optimisation. Dans ce cas, le jeu de données est découpé en trois portions présentées dans le Tableau 3.6. Nous emploierons cette répartition dans l'ensemble de nos travaux.

Apprentissage	Validation	Test
60%	20%	20%
Apprentissage du modèle	Validation du modèle	Validation du meilleur modèle

Tableau 3.6 – Répartition du jeu de données : apprentissage, validation et test.

### 3.3.1.3 Apprentissage statistique traditionnel

Nous présenterons dans cette section les méthodes de classification les plus utilisées dans la littérature. Dans ce travail, nous discuterons de l'intérêt de ces différents classificateurs sur notre application industrielle. La méthode des k-plus-proches voisins est la plus utilisée sur des problèmes de vision par ordinateur. Nous distinguons les classificateurs dits "traditionnels", des classificateurs *Deep Learning* qui utilisent des réseaux de neurones profonds : leur spécificité est

de synthétiser dans un seul modèle la méthode d'extraction de l'information et la classification. Ces derniers seront présentés et discutés dans la Section 3.3.1.4.

### Machine à vecteurs de support

Soit les couples d'échantillons et leurs annotations de classe binaire associées  $(\mathbf{x}_i, y_i)$ . VAPNIK et LERNER [VL63] proposent une solution au problème de classification binaire dans un espace à grandes dimensions, permettant de prédire l'annotation  $y_i$  à partir des données  $\mathbf{x}_i$ . Une machine à vecteurs de support (*Support Vector Machine*) cherche à trouver l'hyperplan qui sépare au mieux les deux classes (Figure 3.5).  $\vec{w}$  est orthogonal à l'hyperplan. L'hyperplan est choisi tel que la distance  $b$  entre le plan et les échantillons de  $\mathbf{x}_i \in \mathbf{X}$  les plus proches de ce dernier, soit maximale : c'est la recherche de la marge maximale  $\frac{b}{\|\vec{w}\|}$ . L'hyperplan obtenu est alors défini par ses vecteurs supports  $\mathbf{x}_i$ . CORTES et VAPNIK proposent l'utilisation d'une fonction de marge dite *soft* basée sur le coût de *hinge* [CV95 ; VGS97]. Ils introduisent un terme de régularisation  $C$ . Trouver le classifieur optimal revient alors à la résolution du problème d'optimisation de l'Équation 3.8.

$$f_{SVM}(\mathbf{x}_i, y_i) = \arg \min \left[ \frac{1}{n} \sum_{i=1}^n \max(0, 1 - y_i (\vec{w} \cdot \vec{x}_i - b)) \right] + C \|\vec{w}\|^2 \quad (3.8)$$

avec  $\mathcal{L}_{Hinge}(\mathbf{x}_i - b, y_i) = \max(0, 1 - y_i \cdot (\mathbf{x}_i - b))$

Lorsque les échantillons ne sont pas linéairement séparables, BOSER, GUYON et VAPNIK [BGV92] proposent d'effectuer une transformation  $\varphi$  de l'espace des données, afin de trouver un espace où les échantillons sont linéairement séparables. La transformation  $\varphi$  a pour noyau une forme choisie (c'est le *kernel trick*, proposé par AŽZERMAN, BRAVERMAN et ROZONOÉR [ABR64]). Soit  $\mathbf{c}$  et  $\varphi$  tel que  $\vec{w} = \sum_{i=1}^n c_i \cdot y_i \cdot \varphi(\vec{x}_i)$ , on a le noyau  $K(X, X') = \varphi(X) \cdot \varphi(X')$ . Cette démarche de transformation vectorielle est utilisée dans de nombreux autres algorithmes. Les noyaux non linéaires les plus employés sont les noyaux polynomiaux et les noyaux gaussiens de bases radiales (*Radial Basis Function*), Équations 3.9. En utilisant le *kernel trick*, un classifieur SVM est alors formalisé par l'Équation 3.10.

$$K_{polynômial}(X, X') = (X \cdot X' + c)^d, \text{ avec } d \text{ le degré et } c \text{ une constante.} \quad (3.9)$$

$$K_{RFB}(X, X') = \exp\left(-\frac{\|X - X'\|^2}{2\sigma^2}\right) = \exp\left(-\gamma \|X - X'\|^2\right), \text{ avec } \gamma > 0$$

$$f_{SVM,RBF}(X_i, y_i) = \arg \min \left[ \frac{1}{n} \sum_{i=1}^n \max(0, 1 - y_i (\vec{w} \cdot \varphi(\vec{x}_i) - b)) \right] + C \|\vec{w} \cdot \varphi\|^2 \quad (3.10)$$

Le choix des hyper-paramètres  $C$  et  $\gamma$  est critique pour obtenir de bonnes performances en matière de justesse et de durée de calcul.  $C$  influence la largeur de la marge entre les deux classes. Ce facteur régularise le classifieur et permet sa généralisation à de nouveaux échantillons.  $C$  est un facteur qui équilibre la nécessité de trouver l'hyperplan le plus simple possible, et le droit aux erreurs de classifications. Une petite valeur de  $C$  encouragera une marge large, ce qui produira un hyperplan simple, mais acceptera des erreurs de classifications. Enfin,  $\gamma$  pondère l'influence d'un unique échantillon sur l'hyperplan.

La librairie *LIBSVM* [CL11] permet de résoudre le problème avec différents noyaux non linéaires, dans un ordre de grandeur de  $O(N^2)$  à  $O(N^3)$ . Aussi, elle n'est plus utilisable lorsque le nombre d'échantillon dépasse dix mille. Dans nos travaux, nous utilisons la librairie *LIBLINEAR* de FAN et al. [Fan+08], au travers de l'interface de la librairie *SciKit-Learn* de PEDREGOSA et al. [Ped+11]. La complexité de l'apprentissage, avec des noyaux linéaires, est de l'ordre de

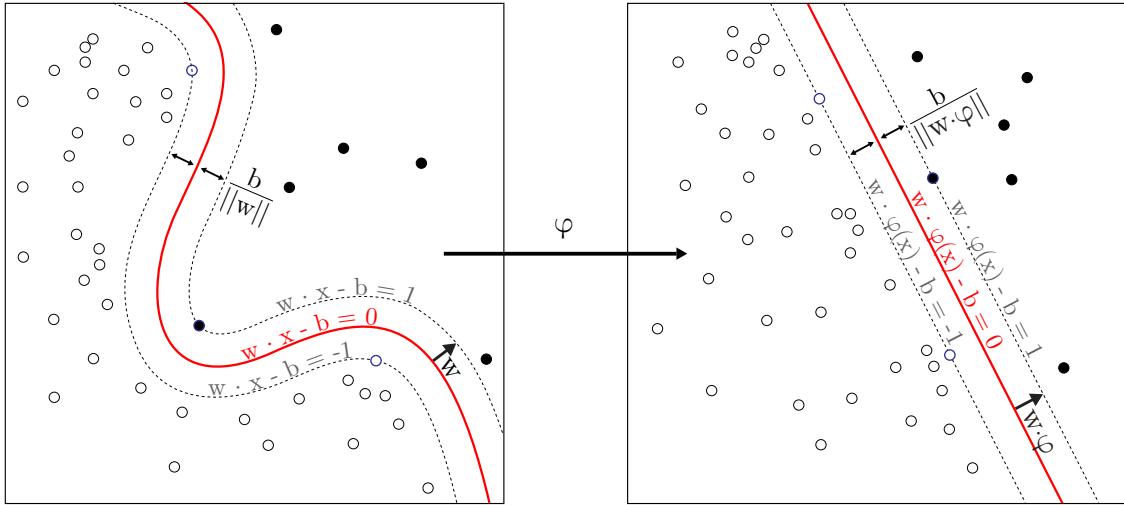


Figure 3.5 – SVM avec kernel trick  $\varphi$ . Figure originale de Alisneaky 0, Wikimedia Commons.

$O(N)$ , avec  $N$  le nombre d'échantillons ce qui rend cette méthode intégrable sur des systèmes embarqués modernes aux vues des puissances de calculs aujourd'hui disponibles. L'apprentissage d'un modèle peut être effectué pendant le cycle de production et le système de calcul peut être intégré directement au cœur de l'atelier de production, au plus près du procédé d'injection-moulage.

### K-plus-proches voisins

La méthode des k-plus-proches voisins a été proposée par FIX et HODGES [FH51] et développée par COVER et HART [CH67]. Cette méthode est à l'origine de méthodes d'apprentissages non supervisées, par exemple celles basées sur les variétés et l'analyse spectrale. Cette méthode ne cherche pas à construire de modèle générique au problème de classification. Elle enregistre les positions des échantillons du jeu d'apprentissage dans l'espace des données. La classification d'un nouvel échantillon est la classe majoritaire sur l'ensemble des échantillons parmi les  $k$  plus proches voisins. Le nombre de voisins  $k$  peut être prédéfini ou bien varier à partir de la densité de points (la métrique de distance entre points est souvent la norme Euclidienne  $\ell_2$ ). Cette méthode est particulièrement efficace lorsque la séparation entre les échantillons est irrégulière.

### Bagging

Le *bagging* (*Bootstrap aggregating*) est une méthode d'association ensembliste de modèles simples afin d'obtenir un méta-modèle aux performances supérieures. Il est défini par BREIMAN [Bre96a]. La technique permet de limiter le sur-apprentissage. Pour un jeu de données  $\mathbf{X}$  de taille  $n$ , on génère  $b$  lots  $\mathbf{X}_b \in \mathbf{X}$  contenant  $\sim 10 < m < n$  échantillons tirés aléatoirement selon une distribution uniforme. Les échantillons, uniques dans  $X$ , peuvent être répétés dans  $\mathbf{X}_b$ . Pour chaque lot  $b$ , un modèle de l'algorithme choisi est entraîné. La prédiction du méta-modèle final est la moyenne (ou le vote pondéré) de tous les modèles  $f_b$ , Équation 3.11. Cette méthode d'agrégation de modèles est efficace pour limiter le sur-apprentissage. La méthode est employée dans les *Random Forests* et elle est souvent combinée avec le *boosting* 3.3.1.3. La méthode de référence pour calculer le nombre optimal  $b$  de fonctions est la sélection du meilleur score par validation croisée 3.3.1.2.

Une métrique intéressante se base sur la performance individuelle de chaque modèle  $f_b$  vis à vis de son jeu de données partiel  $\mathbf{X}_b$ . BREIMAN [Bre96a] définit l'erreur *Out-of-Bag* comme la moyenne, de l'erreur de prédiction pour chaque échantillon  $\mathbf{x}_{i,b}$ , des modèles  $f_b$  avec  $\mathbf{x}_{i,b} \notin \mathbf{X}_b$  qui n'ont pas été entraînés avec les échantillons  $x_i$ . Cette erreur est similaire à une erreur de prédiction

qui serait calculée sur des échantillons de validation. Elle permet ainsi de se passer de jeu de données de validation. Elle est utilisée par les *Random Forests* 3.3.1.3 pour mesurer l'influence des variables sur la prédiction.

$$f_{\text{boosted}} = \frac{1}{b} \sum_{b=1}^B f_b(x_b) \quad (3.11)$$

### Forêt d'arbres décisionnels

L'arbre décisionnel est une méthode de classification adaptée aux grands jeux de données. Un arbre se compose de branches avec entre chacune d'elle un nœud. Chaque nœud répond à une condition sur une variable d'entrée. Ainsi, le modèle obtenu peut être facilement interprété par l'humain. Le modèle obtenu est cependant peu robuste aux perturbations des variables d'entrée. La complexité du modèle est souvent égale au nombre de variables d'entrée ce qui est une cause de sur-apprentissage. Son utilisation pour des échantillons d'entrée qui comporte une dimension supérieure à dix mille est difficile (par exemple les pixels d'une image). C'est pourquoi Ho [Ho95; Ho98] propose un modèle comportant de multiples arbres de décisions (une forêt), chacun entraînés sur une unique partie des variables des échantillons. La prédiction du modèle est donnée par la moyenne des prédictions de l'ensemble des arbres. BREIMAN [Bre01] fait évoluer la méthode : les arbres sont choisis pour ne pas être corrélés entre eux et leurs sorties sont agrégées par *bagging* 3.3.1.3.

Une utilité indirecte de la notion stochastique des forêts d'arbres décisionnels est la possibilité de mesurer l'influence des variables sur le modèle. Lors de l'entraînement, l'erreur *Out-of-Bag* 3.3.1.3 (équivalente à une erreur de prédiction sur un échantillon de validation) est moyennée pour l'ensemble des arbres. C'est l'erreur *Out-of-Bag* de référence, pour chacune des variables  $j$ ,  $e_{\text{ref},j}$ . À chaque itération, on permute aléatoirement la position des variables  $v$  dans le jeu de données et on calcule à nouveau l'erreur *Out-of-Bag*. Les permutations sont réalisées un grand nombre de fois ( $p > 1000$ ). On obtient l'importance de la  $j^{\text{ème}}$  variable en calculant la moyenne de l'erreur sur toutes les permutations  $e_{\text{perm}}$  de la différence avec l'erreur de référence de la variable :  $\text{importance}_j = e_{\text{perm},j} - e_{\text{ref},j}$ . Il est enfin possible de normaliser la valeur de l'importance par l'écart-type sur l'ensemble des variables pour classer les variables sur une échelle continue [0; 1]. Nous utiliserons cette méthode pour sélectionner les variables pertinentes.

### Boosting

KEARNS [Kea88] pose l'hypothèse de l'existence d'un méta-classifieur qui utiliserait une association de classificateurs aux performances non optimales (légèrement supérieure à une réponse aléatoire) pour obtenir de bonnes performances : le *boosting*. Dans un tel ensemble, chaque classificateur est ainsi faiblement corrélé avec la totalité des données. SCHAPIRE [Sch90] vérifie l'hypothèse et par la suite BREIMAN [Bre96b; Bre97] développe la méthode. De nombreux algorithmes sont dérivés de cette méthode. Par la suite, MASON et al.; FRIEDMAN [Mas+99; Fri01] proposent le *boosting* avec descente de gradient (*Gradient Boosting*) et FRIEDMAN [Fri02] ajoute une dimension stochastique : *Stochastic Gradient Boosting*. Cette idée est motivée par le *bagging* 3.3.1.3. Chaque classificateur est entraîné sur des partitions aléatoires différentes du jeu de données, mais sans aucune répétition des échantillons initiaux. Une partition qui a pour dimensions la moitié du jeu de données initial produit une augmentation de performance significative du méta-classificateur. Un des algorithmes le plus populaire dans des compétitions de classification est *XGBoost*. Il est basé sur le *Stochastic Gradient Boosting* d'arbres de décisions binaires. Cependant, les performances de cet algorithme pour des jeux de données où les échantillons possèdent de grandes dimensions ( $> 1000$ ), comme des images, sont inférieures aux méthodes à réseaux de neurones profonds. La performance de ces algorithmes dépend de la pertinence de la réduction de la dimension des échantillons, qui doit être implémentée par un humain. Les réseaux de *Deep Learning* dépassent cette limite en apprenant l'algorithme de réduction de dimensions adapté en

même temps que l'algorithme de classification. Enfin, les idées issues du *bagging* et du *boosting* sont utilisées dans les méthodes d'apprentissages des réseaux de neurones profonds.

### Réseaux de neurones

Le travail de MCCULLOCH et PITTS [MP43] interroge sur la méthode de fonctionnement du cerveau humain : de nombreuses cellules élémentaires connectées entre elles permettent de résoudre des problèmes complexes, ce qui pose la problématique de la recherche Connexionniste. Ils proposent de concevoir des neurones, sous la forme de portes logiques élémentaires. Pour des valeurs d'entrée  $\mathbf{x}_i$ , un neurone calcule la somme pondérée par une matrice de poids  $\mathbf{W}$ , puis si la valeur obtenue est supérieure à une valeur  $b$ , la sortie est zéro, sinon la sortie est un, Équation 3.12. Cet algorithme permet de convertir des variables continues en valeurs booléennes discontinues. Par la suite, ROSENBLATT [Ros58] propose le *perceptron*, en s'appuyant sur la connaissance de l'époque du système de la vision biologique : de la rétine au cerveau. Le *perceptron* permet d'agréger à l'aide d'un réseau de neurones à une couche, des données représentant des variables pertinentes extraites d'images (*pre-processing*). La thématique de recherche Connexionnisme sera peu développée par la suite car les ressources informatiques nécessaires sont importantes et les résultats obtenus sont limités en comparaison des attentes de l'époque. Les autres méthodes présentées dans cette section seront préférées, jusqu'aux récents succès du *Deep Learning*, §3.3.1.4.

$$f_{\text{perceptron}}(\mathbf{x}_i) = \begin{cases} 1 & \text{si } \mathbf{w}_i \cdot \mathbf{x}_i + b_i > 0 \\ 0 & \text{sinon} \end{cases} \quad (3.12)$$

Suite à la proposition de réseaux de neurones à multiples couches successives de FUKUSHIMA [Fuk80], RUMELHART, HINTON et WILLIAMS [RHW85] applique une méthode d'ajustement itératif des poids en utilisant la rétro-propagation de l'erreur. La rétro-propagation de l'erreur s'appuie sur les travaux sur la dérivation automatique des fonctions composées, initiés par LINNAINMAA [Lin76]. De plus, RUMELHART, HINTON et WILLIAMS [RHW85] étend les capacités du réseau à répondre aux problèmes où les données ne sont pas linéairement séparables, en introduisant des fonctions d'activations non linéaires, en lieu et place de la simple condition inférieure/supérieure. La particularité des réseaux de neurones est leur capacité à modéliser des problèmes fortement non linéaires. La prochaine étape est la construction de réseaux aux architectures plus complexes et l'utilisation de fonctions de convolution, §3.3.1.4.

#### 3.3.1.4 Réseaux profonds de convolutions, Deep Learning

Les réseaux de neurones de convolutions, proposés par LECUN et al. [LeC+89], ont un comportement non linéaire particulièrement adapté à l'analyse de données complexes, telles que des images, et à l'apprentissage de représentations de ces données. La couche de convolution est l'élément principal du réseau. Les couches de convolutions sont associées les unes à la suite des autres pour former un réseau profond (d'où le nom de *Deep Learning*). Entre chaque couche peuvent être ajoutées des couches composées de fonctions d'activations non linéaires, des couches de moyennes globales ou locales, ou d'autres fonctions avancées. Le réseau résultant est une composée de fonctions de convolutions avec des fonctions intermédiaires. L'Équation 3.13 détaille la forme d'un modèle complet  $F$  et de l'application d'une couche  $f_n$  : c'est un modèle affine composé d'une matrice de poids  $\mathbf{W}$  et d'un vecteur de biais  $\mathbf{b}$ .

$$F(\mathbf{x}) = (f_n \circ f_{n-1} \circ \dots \circ f_2 \circ f_1)(\mathbf{x}), \text{ avec } f_n(\mathbf{x}) = \mathbf{w}_n \cdot \mathbf{x} + \mathbf{b}_n \quad (3.13)$$

De plus, l'opération de convolution est très efficiente pour des problèmes en grandes dimensions [GBC16]. Pour une entrée de dimensions  $m$ , et une sortie de dimensions  $n$ , l'application matricielle complète  $x \rightarrow y$  aura une complexité  $O(m \cdot n)$ . L'opération de convolution

pour un noyau de dimensions  $k$  aura une complexité de  $O(k \cdot n)$ . Lorsque la dimension du noyau  $k$  est très petite devant la dimension de l'entrée  $m$ , l'opération de convolution permet un gain d'occupation mémoire et ainsi de coût de calcul. Le gain est considérable lorsque les données d'entrée ont de grandes dimensions, comme dans notre cas avec des images. Il est bien plus efficient de créer des architectures profondes, supérieures à  $c$  couches de convolutions successives avec des noyaux de petite dimension  $k$ , car  $c$  est très petit devant  $m$ .

Ces réseaux ont profité de la disponibilité croissante de la puissance de calcul des processeurs graphiques. En 2006, le premier réseau de convolution est implémenté sur processeur graphique par CHELLAPILLA, PURI et SIMARD [CPS06], avec des performances quatre fois supérieures à l'implémentation sur processeur classique. En 2011, un réseau de convolution accéléré sur processeur graphique (60 fois plus rapide) dépasse les performances de l'humain sur une tâche de classification de panneaux signalétiques routiers [Cir+11]. En 2012, le réseau AlexNet de KRIZHEVSKY, SUTSKEVER et HINTON [KSH12] gagne le challenge annuel de classification d'images *ImageNet* [Den+09] avec un très grand écart de 10,8% sur le score, avec l'ensemble des concurrents. La performance d'*AlexNet* (8 couches) est dépassée en 2013 une version identique aux hyper-paramètres optimisés [ZF13], puis en 2014 par *VGG16-19* (16 et 19 couches) [SZ14] et *GoogleNet* (22 couches). Ce dernier est dépassé en 2015, par le réseau *ResNet101* [He+15a] contenant plus de 150 couches. L'architecture massive de ces réseaux évolue depuis 2011 vers des architectures de plus en plus profondes [He+15a] et qui sont plus efficientes sur le plan de la mémoire à allouer et puissances de calculs, mais aussi de plus en plus complexes, §3.3.1.5. Les architectures de l'état de l'art de la littérature dépassent aujourd'hui les capacités d'ingénierie humaine ; elles sont générées automatiquement, §3.4. La dernière couche du réseau est appelée couche de classification (ou classifieur). Si le problème est binaire (comporte uniquement deux classes), la sortie du réseau est une probabilité d'appartenance à la première classe. Pour un problème multi-classes, la sortie du réseau est un vecteur de probabilité d'appartenance à chacune des classes. Pour un problème binaire, une fonction d'activation sigmoïde, Équation 3.14, est utilisée afin de transformer les valeurs de la précédente couche vers une probabilité :  $s_{[-\infty; +\infty]} \rightarrow P_{[0;1]}$ .

$$\text{Sigmoïde } f(s_i) = \frac{1}{1 + e^{-s_i}} \quad (3.14)$$

Pour un problème multi-classes, une fonction *SoftMax* (Équation 3.15), inverse de la sigmoïde, est utilisée.  $\text{SoftMax} = \vec{s}_{[-\infty; +\infty]} \rightarrow \vec{P}_{[0;1]}$  Chaque classe a préalablement été encodée sous la forme d'un vecteur à la dimension égale au nombre total de classes, c'est le *one-hot encoding*.

$$\text{SoftMax } f(s_i) = \frac{e^{s_i}}{\sum_c^{\text{nb classes}} e^{s_c}} \quad (3.15)$$

Ce modèle à réseau de neurones est pleinement différentiable. C'est pourquoi la méthode traditionnelle d'ajustement des poids des neurones du réseau est la rétro-propagation de l'erreur proposée, par RUMELHART, HINTON et WILLIAMS [RHW85] qui est adaptée aux réseaux de convolutions à couches multiples par LECUN et al. [LeC+89]. Par la suite, LECUN et al. [LeC+98] développe l'apprentissage progressif stochastique des poids du réseau, aussi appelée descente de gradient stochastique : l'apprentissage est réalisé successivement sur une portion des données choisie aléatoirement (les *mini-batch*), car calculer le gradient pour la totalité du jeu de données serait trop coûteux en matière de puissance de calcul.

[Vap92] définit dans la théorie Vapnik–Tchervonenkis la notion d'apprentissage statistique comme un problème d'optimisation, Équation 3.16, où on cherche à minimiser l'erreur empirique  $E$ , en fonction des poids  $\mathbf{W}$  du réseau et d'un terme de régularisation  $\lambda$  sur les poids  $\mathbf{W}$  (régularisation de Tikhonov §3.3.1.7).

$$E(\mathbf{W}, \lambda_p) = \frac{1}{\ell} \sum_{i=1}^{\ell} L(y_i, f(x_i, \mathbf{W})) + \lambda_p \|\mathbf{W}\|_2 \quad (3.16)$$

A chaque itération, une fonction de coût  $\mathcal{L}$  calcule l'erreur entre les prédictions du réseau  $\hat{y}$  et les valeurs de références  $y$ . La fonction de coût est choisie pour être adaptée au problème de classification en présence. Nous détaillerons dans ce paragraphe le cas de la classification binaire. L'entropie  $H$  est une métrique importante de la théorie de l'information, proposée par [Sha48]. Pour une distribution de probabilité  $p$  donnée, l'entropie est exprimée par l'Équation 3.17. Elle mesure la quantité de données minimale nécessaire pour encoder l'information de la distribution.

$$\text{Entropie } H(p) = - \sum_{i=1}^N p(x_i) \cdot \log p(x_i) \quad (3.17)$$

Dans le cas de l'apprentissage supervisé, on cherche à trouver la fonction  $F$  qui minimise la perte de l'information entre l'approximation finale et le jeu de données initial. Aussi, en comparant les entropies d'une distribution source  $y$  et d'une approximation de  $\hat{y}$ , on mesure la quantité d'informations perdue par la méthode d'approximation  $F$  apprise (aussi appelée méthode d'encodage). C'est la divergence de KULLBACK et LEIBLER [KL51], Équation 3.18 et nous cherchons ici à la minimiser.

$$\text{Divergence}_{KL}(y \parallel \hat{y}) = \sum_{i=1}^N y(\mathbf{x}_i) \cdot (\log y(\mathbf{x}_i) - \log \hat{y}(\mathbf{x}_i)) \quad (3.18)$$

Dans le cas d'une classification binaire, on utilise ainsi la fonction de coût logistique (Équation 3.19), aussi appelée entropie-croisée binaire. Cette fonction est généralisable aux problèmes multi-classes avec  $i > 1$ .

$$\mathcal{L}_{logit}(\hat{y}, y) = - \sum_{i=1}^2 y_i \log(\hat{y}_i) = -y_1 \log(\hat{y}_1) - (1-y_1) \log(1-\hat{y}_1) \quad (3.19)$$

Le calcul est simplifié en combinant la fonction d'activation sigmoïde et la fonction de coût logistique, Équation 3.20.

$$\begin{aligned} (\mathcal{L}_{logit} \circ f_{sig})(X, y_1) &= -y_1 \log(f_{sig}(X)) - (1-y_1) \log(1-f_{sig}(X)) = -y_1 \cdot X + \log(1+e^X) \\ \text{d'où } (\mathcal{L}_{logit} \circ f_{sig})(X, y_1) &= \begin{cases} -X + \log(1+e^X) & \text{si } y_1 = 1 \\ \log(1+e^X) & \text{si } y_1 = 0 \end{cases} \end{aligned} \quad (3.20)$$

A chaque itération, la dérivée de cette erreur est rétro-propagée pour ajuster chacun des poids du réseau. Le lecteur pourra se référer à la note de référence de SADOWSKI pour le détail du calcul de la dérivée [Sad16]. De plus, la dérivée est pondérée par un facteur  $0 < \eta < 1$  qui est appelée le facteur d'apprentissage. Le facteur  $\eta$  est critique pour obtenir la convergence de l'erreur vers zéro. Plus le réseau est complexe, plus le problème d'optimisation en présence est non-linéaire et non-convexe ; il possède de nombreux minima locaux. C'est pourquoi, un facteur  $\eta$  trop petit demandera un grand nombre d'itérations avant d'obtenir la convergence et l'équilibre sera souvent atteint sur un minima local. A contrario, un facteur  $\eta$  trop grand empêchera la convergence car le gradient oscillera. Les deux problèmes critiques à éviter lors de l'ajustement par descente de gradients sont l'effondrement du gradient qui occasionne une convergence globale des poids du réseau vers zéro et l'explosion du gradient qui occasionne une divergence globale des poids du réseau vers l'infini [BSF94; Hoc+01]. Des algorithmes avancés existent pour limiter ces problèmes.

La vaste littérature sur le sujet ne nous permettra pas de détailler l'intérêt des différentes méthodes proposées, aussi, nous utiliserons dans l'ensemble de nos travaux l'algorithme d'optimisation récent *Adam* proposé par KINGMA et BA [KB14]. Ce dernier permet l'ajustement automatique du facteur d'apprentissage  $\eta$  au fur et à mesure des itérations. De plus, des facteurs de régularisation sur la dérivée seconde (aussi appelé "moment") sont ajoutés pour accélérer la convergence et éviter les minimums locaux.

Nous récapitulons dans le Tableau 3.7 les différents hyper-paramètres critiques de l'apprentissage par descente de gradient stochastique des réseaux de neurones profonds. Dans ce travail, nous ne présenterons pas de manière exhaustive les contraintes qui existent sur l'architecture des réseaux de convolutions. Le choix des nombres de couches, des dimensions des noyaux de convolutions, des décalages, remplissages (*padding*), réduction (*pooling*, §3.3.1.7). Nous invitons le lecteur à se référer à l'excellente synthèse de DUMOULIN et VISIN [DV16].

Hyper-paramètres	Valeurs de la littérature	Description
Nombre de couches	[2; 200]	Nombre de couches dans le réseau.
Type de couche	Convolution, Connectés, Pooling	Type de couche, varie à chaque couche.
Neurones par couche	[1; 4000]	Varié à chaque couche successive.
Fonction d'activation	tanh, sigmoïde, SoftMax	Transforme la sortie en probabilités.
Fonction de coût	Logistique, Entropie-croisée, $\ell_1$ , $\ell_2$	Erreur entre référence et prédiction.
Initialisation	Loi uniforme, normale ...	Méthode d'initialisation des poids.
Régularisation des poids	$\ell_1$ , $\ell_2$	Limite l'explosion des poids.
Facteur d'apprentissage	$[10^{-7}; 10^{-1}]$	Pas du gradient, vitesse de convergence.
Lots d'apprentissage/ <i>mini-batch</i>	[16; 128]	Nombre d'échantillons par lot

Tableau 3.7 – Hyper-paramètres critiques pour l'apprentissage de réseaux de neurones profonds.

Dans ce travail, nous avons étudié un panel représentatif de différentes architectures récentes de réseaux de neurones profonds, sur notre cas d'application de contrôle de la qualité. Le Tableau 3.8 présente les architectures étudiées avec leurs occupations de mémoire de travail, le nombre de paramètres (Degrés De Libertés) et les scores de justesse obtenus sur le jeu de données de validation d'*ImageNet*. L'erreur *Top-1%* correspond à l'erreur de justesse pour l'annotation qui est prédite avec la plus grande probabilité par le réseau. L'erreur *Top-10* correspond à l'erreur de justesse pour les dix annotations qui sont prédites avec la plus grande probabilité par le réseau. Ainsi, si l'annotation qui devait être prédite est "chat", mais que le réseau prédit l'annotation "chat" en neuvième position, alors l'erreur *Top-10%* sera nulle, mais l'erreur *Top-1* sera maximale. Dans notre problématique de classification binaire de la qualité, nous nous intéressons uniquement à l'erreur *Top-1*. Nous distinguons également les ressources informatiques nécessaires à entraîner ces modèles : la durée du calcul et la mémoire de travail nécessaire. De plus, pour réaliser l'inférence à l'aide de ces modèles, ils doivent être chargés en mémoire de travail (*RAM* du processeur, ou *RAM* du *GPU*). Les modèles qui possèdent un grand nombre de poids nécessitent de grandes quantités de mémoires ce qui limite leur utilisation dans des systèmes embarqués positionnés au cœur des ateliers de production. La faible quantité de données annotées dont nous disposons est une problématique industrielle qui a orienté nos travaux vers le transfert de modèles profonds, préalablement appris sur de grandes quantités de données annotées, §3.3.2.

Architecture	Année	Mémoire	Paramètres	Erreur Top-1%	Erreur Top-5%
MobileNetV2	2018	14 Mo	3,5 M	28,12	9,71
VGG19	2014	549 Mo	143,6 M	25,76	9,12
DenseNet-201	2016	80 Mo	20,2 M	22,80	6,43
ResNet-152	2015	232 Mo	60,3 M	21,69	5,94
ResNext-50_32x4d	2016	85 Mo	25,1 M	22,38	6,30
ResNext-101_32x8d	2016	170 Mo	44,3 M	20,69	5,47
ResNeXt-101_32x48d_wsl	2018	3100 Mo	829 M	14,6	2,4

Tableau 3.8 – Architecture des réseaux de convolutions étudiés.

### 3.3.1.5 Réseaux de convolutions résiduels

Construire des réseaux de neurones aux architectures profondes permet de limiter l'utilisation de couches pleinement connectées, dont les poids sont très coûteux à optimiser. La problématique principale des réseaux profonds est l'entraînement. La rétro-propagation a une faible incidence sur les poids situés les plus en amont du réseau. Cela rend l'entraînement des réseaux aux couches supérieures à dix très difficile. C'est pourquoi des connections intermédiaires ont été introduites par SRIVASTAVA, GREFF et SCHMIDHUBER [SGS15a; SGS15b] : les *Highway Networks*. Elles permettent à la rétro-propagation d'agir sur toutes les couches, y compris les couches en amonts. Cette architecture est inspirée des connections intermédiaires pondérées des réseaux récurrents *Long Short-Term Memory*. Soit  $y_{i+1}$  la sortie d'une couche  $i$  qui est composée d'une fonction affine et d'une activation non linéaire  $f_{i+1}$  dont les poids sont  $\mathbf{W}_f$ . On introduit une connexion intermédiaire  $T$  pondérée par une matrice de poids  $\mathbf{W}_T$ , Équation 3.21. Le poids  $\mathbf{W}_T$  détermine si la couche laissera passer l'information ou la transformera. Les auteurs notent que  $\mathbf{W}_T$  doit être initialisée à des valeurs négatives afin de débuter nécessairement l'apprentissage de  $\mathbf{W}_f$ . Par la suite,  $f$  pourra être ignoré au fur et à mesure de l'apprentissage si cela entraîne de meilleures performances. De manière simplifiée, cette architecture permet au réseau d'ajuster sa profondeur optimale, lors de l'apprentissage : si les connections intermédiaires sont utilisées, le réseau est moins profond.

$$y_{i+1} = f_{i+1}(y_i, \mathbf{W}_f) \cdot T_{i+1}(y_i, \mathbf{W}_T) + y_i \cdot (1 - T_{i+1}(y_i, \mathbf{W}_T)) \quad (3.21)$$

Par la suite, HE et al. [He+15a] propose l'architecture *ResNet* qui est un cas particulier de *Highway Network* où  $\mathbf{W}_T \equiv 0,5$ , Équation 3.22. Les données sont transmises de manière équivalente entre la transformation et la connexion intermédiaire. [He+16] compare les performances des deux méthodes et montre que la plus simple architecture *ResNet* offre de meilleures performances. Cependant, SRIVASTAVA, GREFF et SCHMIDHUBER [SGS15a] montrent que la valeur apprise de  $\mathbf{W}_t$  est différente de 0.5. On a particulièrement souvent  $\mathbf{W}_T > 0.5$  : le réseau a appris à utiliser en priorité les connections intermédiaires, ce qui simplifie le réseau mais impacte les performances. De plus, les poids  $\mathbf{W}_T$  augmentent le nombre de paramètres à apprendre, ce qui pourrait diminuer la performance.

$$y_{i+1} = f_{i+1}(y_i, \mathbf{W}_f) + y_i \quad (3.22)$$

Par la suite, ces idées amènent à l'architecture *DenseNet* [Hua+16] : les connections intermédiaires sont présentes entre toutes les couches et non plus seulement d'une couche à l'autre. Toutes les couches peuvent alors avoir accès à l'information de bas niveau, comme par exemple l'image originale et combiner cette information pour résoudre le problème. La Figure 3.6 présente

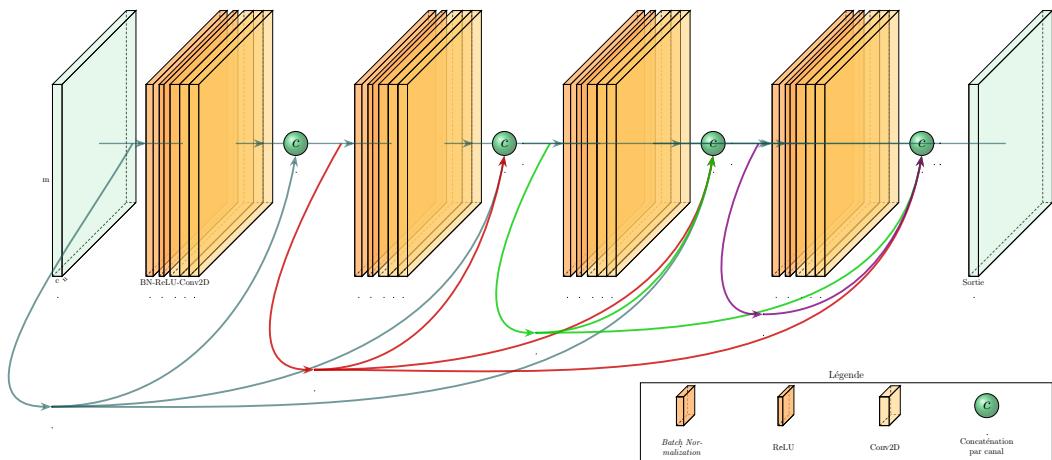


Figure 3.6 – Architecture d'un bloc DenseNet.

l'architecture d'un bloc *Dense*.<sup>2</sup> Un réseau complet enchaîne successivement quatre blocs *dense*.

La dernière évolution des réseaux résiduels est l'introduction de multiples blocs dans chaque neurones classiques. C'est l'architecture *ResNext* [Xie+16], une dimension  $C$  dite de "cardinalité" qui définit le nombre de bloc présent dans chaque couche. Les auteurs montrent l'intérêt d'une architecture où  $c = 32$  : chaque bloc est répété 32 fois par neurones.

Les performances sont améliorées bien que cela représente une augmentation conséquente des paramètres du réseau. Dernièrement, MAHAJAN et al. [Mah+18] ont réussi à augmenter les performances sur *ImageNet* en réalisant un pré-apprentissage sur 940 millions d'images issues d'*Instagram*. La capacité (au sens de la théorie VC [Vap92]) du modèle *ResNext* est bien supérieure à la seule base *ImageNet* et un plus grand nombre de données réelles permet d'augmenter les performances, par la généralisation du problème.

Cependant, l'augmentation du nombre de paramètres des réseaux n'est pas nécessairement significative de meilleures performances. Dernièrement, le réseau *EfficientNet* dépasse l'état de l'art sur *ImageNet* en diminuant drastiquement le nombre de paramètres ajustables et optimisant sous contrainte l'occupation mémoire et le nombre d'opérations nécessaires. Précédemment, les réseaux *MobileNetV1* [How+17] et *MobileNetV2* [San+18] cherchent à obtenir les meilleures performances sur dispositifs embarqués tels que les smartphones. Il s'agit de limiter le nombre de paramètres et le nombre d'opérations à réaliser (additions et multiplications). *MobileNetV1* obtient une performance supérieure à *GoogLeNet*, avec trois fois moins d'opérations à réaliser. Il égale *VGG19* avec trente fois moins d'opérations et dix fois moins d'occupation mémoire. *MobileNetV1* utilise un bloc de convolution efficient, proposé dans l'architecture *Xception* de CHOLLET [Cho16] : la convolution est effectuée indépendamment sur chaque canal, avec un poids différent pour chaque canal, puis les résultats sont concaténés. Pour une convolution traditionnelle de fenêtre  $3 \times 3$ , la réduction du nombre d'opérations de calcul à effectuer est d'un facteur 9. *MobileNetV2* ajoute ensuite l'utilisation de connexions intermédiaires entre chaque couche, à la manière des réseaux résiduels. Les performances de ces réseaux sont inférieures à l'état de l'art mais ils permettent d'exécuter l'inférence avec des performances satisfaisantes sur des systèmes embarqués.

2. Les architectures présentées dans ce travail sont générées avec l'aide des scripts développés par IQBAL [Iqb18].

### 3.3.1.6 Fonctions d'activations

Pour éviter les deux problèmes de l'explosion ou de l'effondrement de la valeur du gradient pendant l'apprentissage, nous utiliserons dans nos travaux des fonctions d'activation *LeakyReLU*, *PreLU* ou *ELU*, récemment proposées dans la littérature. La fonction d'activation est une non-linéarité qui suit traditionnellement l'opération de convolutions. Elle permet de modéliser une réponse à un problème non linéaire. Les fonctions d'activation historiques sont la tangente hyperbolique et la sigmoïde. NAIR et HINTON [NH10] proposent la fonction *ReLU* (*Rectified Linear Unit*) qui une approximation de la sigmoïde par trois et obtient de meilleures performances, en diminuant le coût du calcul.

Fonction	Équation	Dérivée
Sigmoïde	$f(x) = \frac{1}{1+e^{-x}}$	$f(x)(1-f(x))$
Tanh	$tanh(x) = \frac{2}{1+e^{-2x}} - 1$	$1 - f(x)^2$
ReLU	$\begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$	$\begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases}$
LeakyReLU	$\begin{cases} 0,01x & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$	$\begin{cases} 0,01 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases}$
PreLU	$\begin{cases} \alpha x & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$	$\begin{cases} \alpha & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases}$
ELU	$\begin{cases} \alpha(e^x - 1) & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$	$\begin{cases} f(x) + \alpha & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases}$

Tableau 3.9 – Fonctions d'activation.

Le principal problème est que la fonction *ReLU* vaut zéro pour des valeurs négatives. C'est pourquoi MAAS, HANNUN et NG [MHN13] introduisent le *LeakyReLU* : une succession de deux droites affines, dont la pente de la droite pour les valeurs négatives est non nulle, mais très faible ( $0,01 \cdot x$ ). Par la suite, HE et al. [He+15b] introduisent le *PreLU* (*Parametric ReLU*) dont un paramètre  $\alpha \geq 0$  définit la pente de la droite pour les valeurs négatives.  $\alpha$  est appris au cours de l'apprentissage tout comme les matrices de poids du réseau. La valeur de  $\alpha$  peut également être indépendante pour chaque canal. Ces fonctions sont discontinues et ne sont pas dérivables en zéro. Cela peut entraîner des problèmes de convergence lors de la rétro-propagation du gradient. C'est pourquoi dernièrement, CLEVERT, UNTERTHINER et HOCHREITER proposent le *ELU* (*Exponential Linear Unit*) qui est dérivable [CUH15]. L'ensemble des graphes de ces fonctions est présenté dans la Figure 3.7, leurs dérivées dans le Tableau 3.9.

### 3.3.1.7 Méthode de régularisation pour limiter le sur-apprentissage

Le nombre de degrés de liberté d'un modèle à réseaux de neurones profonds est supérieur à un million. C'est pourquoi, il est courant que le modèle se spécialise afin d'obtenir un très bon score de justesse sur les échantillons d'apprentissages, mais un mauvais score sur de nouveaux échantillons : c'est le problème du *sur-apprentissage*. Plusieurs méthodes de régularisation permettent de limiter ce problème et nous détaillerons dans cette section les méthodes que nous utilisons. GOODFELLOW, BENGIO et COURVILLE proposent une définition du terme "régularisation", spécifique à l'alchimie du *Deep Learning* [GBC16] :

« La régularisation est toute modification faite à un algorithme d'apprentissage afin que l'erreur de généralisation soit réduite, mais pas l'erreur d'apprentissage. »

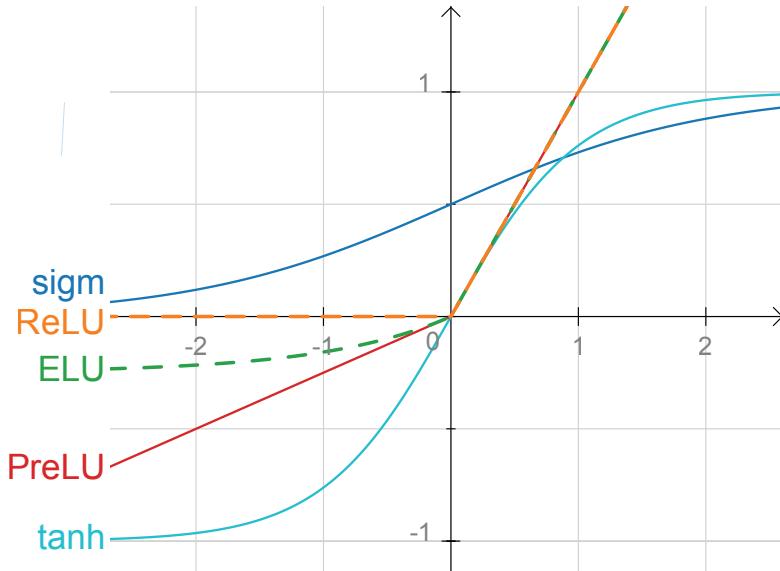


Figure 3.7 – Fonctions d'activation.

### Dropout

SRIVASTAVA et al. [Sri+14] proposent l'utilisation d'une couche dite de *dropout* : la valeur de sortie de cette couche est aléatoirement, selon un pourcentage de chances  $\alpha$  choisi, fixée à zéro. Le *dropout* est actuellement la méthode de régularisation la plus performante pour limiter le sur-apprentissage. Ainsi, l'optimisation est réalisée sur une mauvaise valeur (zéro) de sortie du modèle pour un pourcentage  $\alpha$  d'échantillons d'un *mini-batch*. La valeur de  $\alpha$  est souvent fixée entre  $[0; 0,5]$ . Lors de l'utilisation du modèle pour inférence, la valeur de  $\alpha$  est nulle : la couche de *dropout* n'est jamais activée, pour ne pas fausser les prédictions.

### Normalisation de lots

La couche de normalisation de lots (*Batch Normalization*) a été proposée par IOFFE et SZEGEDY [IS15]. Elle nécessite l'emploi de l'entraînement par descente de gradient stochastique qui découpe le jeu de données d'apprentissage en *mini batch* répartis aléatoirement. On définit les valeurs  $x_i$  en entrée de la couche pour un *mini batch*  $B \in X$ . On calcule la variance  $\sigma_B^2$  de l'entrée et la moyenne  $\mu_B$ . La couche *BN* normalise les valeurs des entrées  $x_i$  en  $\bar{x}_i$  par la variance et la moyenne de la distribution du *mini batch*, Équations 3.23.

$$\begin{aligned} \mu_B &\leftarrow \frac{1}{m} \sum_{i=1}^m x_i, && \text{Moyenne} \\ \sigma_B^2 &\leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2, && \text{Variance} \\ \bar{x}_i &\leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}, \text{ Batch normalization} && (3.23) \end{aligned}$$

La couche *BN* limite les variations importantes de la covariance des entrées. S'il n'y a pas de normalisation, lors de la rétro-propagation, l'ajustement des poids devra être fait pour chaque distribution différente de *mini-batch*. Avec la normalisation, la sortie aura toujours une distribution normalisée sur zéro et ainsi la valeur des poids à ajuster ne sera pas extrême. IOFFE et SZEGEDY [IS15] suggère d'utiliser en priorité des couches *BN* plutôt que le *dropout*, car ce dernier fait perdre la totalité de l'information du réseau lorsqu'elle est active, ce qui pénalise fortement la

vitesse d'apprentissage. Les architectures de réseaux profonds antérieures à 2015 (par exemple VGG16) n'utilisent pas de couches  $BN$ , alors que la majorité des architectures postérieures à 2015 en utilisent. Lors de l'utilisation du modèle pour inférence, la valeur de  $\mu_B$  et  $\sigma_B^2$  est fixée pour tous les *mini batch* : ce sont leurs moyennes respectives sur tous les *mini batch* de l'entraînement.

### Arrêt prématurée de l'apprentissage

Cette méthode permet de limiter le sur-apprentissage en stoppant la phase d'apprentissage dès lors qu'il n'y a plus d'amélioration des performances du modèle. Aussi appelé *early stopping* [YRC07], il s'agit de trouver le nombre optimal d'itérations successives pour ne pas sur-apprendre le jeu de données d'apprentissage. Pour cela, le jeu de données initial est séparé en une partition d'apprentissage et une partition de validation. L'apprentissage du modèle est effectué en utilisant la partition d'apprentissage. À chaque itération, la performance du modèle est évaluée sur la partition de validation, inconnue du modèle. Si la performance de validation a été améliorée, alors l'itération d'apprentissage a amélioré le modèle. En revanche, si la performance de validation est moins bonne qu'à l'itération précédente, le modèle a été dégradé pour la partition de validation. C'est pourquoi, il aurait été judicieux de ne pas effectuer cette itération qui a dégradé le modèle, et de conserver le modèle tel que précédemment : c'est ce que réalise le *early stopping*. L'algorithme de descente de gradient peut souvent produire une mauvaise performance pendant quelques itérations avant d'améliorer le modèle, c'est pourquoi on introduit le terme de patience  $p$  : le nombre d'itérations qui n'améliore pas le modèle, avant de terminer l'apprentissage. La performance de cette méthode dépendra du choix judicieux de l'hyper-paramètre  $p$ , c'est pourquoi il est obligatoire de trouver la valeur de  $p$  optimale pour le problème à résoudre, §3.4.

### Réduction de dimensions par interpolation

L'idée principale de l'architecture des réseaux de neurones profonds (avant la proposition des réseaux résiduels) est de réduire successivement, à chaque couche, la dimension des vecteurs, afin de réduire le nombre de paramètres des couches. Cela réduit les coûts des calculs et le risque de sur-apprentissage. Les couches de convolutions appliquent un filtre mais ne réduisent pas la dimension. C'est pourquoi elles sont suivies de couches dites de *pooling*. Parmi les nombreuses fonctions de *pooling* proposées, on distingue les deux principales : le sous-échantillonnage par moyenne, et la seule conservation de la valeur maximale, le *max pooling*, Tableau 3.10. Cette couche est appliquée sur un certain nombre de valeurs adjacentes. Des dimensions courantes sont de  $2 \times 2$  à  $4 \times 4$ , en fonction de la dimension des caractéristiques des échantillons que l'on cherche à mettre en valeur. Une fenêtre de  $2 \times 2$  réduit la dimension d'un facteur 4.

0	255	10	0
10	20	0	200
50	0	0	0
10	0	20	10

→

255	200
50	20

Tableau 3.10 – Illustration d'une couche de max pooling.

BOUREAU, PONCE et LECUN [BPL10] analysent l'influence du choix de la fonction et montre la supériorité du *max pooling* lorsque les valeurs à détecter sont peu fréquentes. SCHERER, MÜLLER et BEHNKE [SMB10] obtient également des performances supérieures avec le *max pooling* qui permet d'extraire les caractéristiques de petites dimensions spatiales. Une particularité du *max pooling* est d'être très simple à calculer lors de la rétro-propagation : seul le poids associé à la valeur maximale est ajusté. La Figure 3.8 illustre l'effet du *max pooling* et de l'*average pooling* pour quatre filtres de convolutions différents.

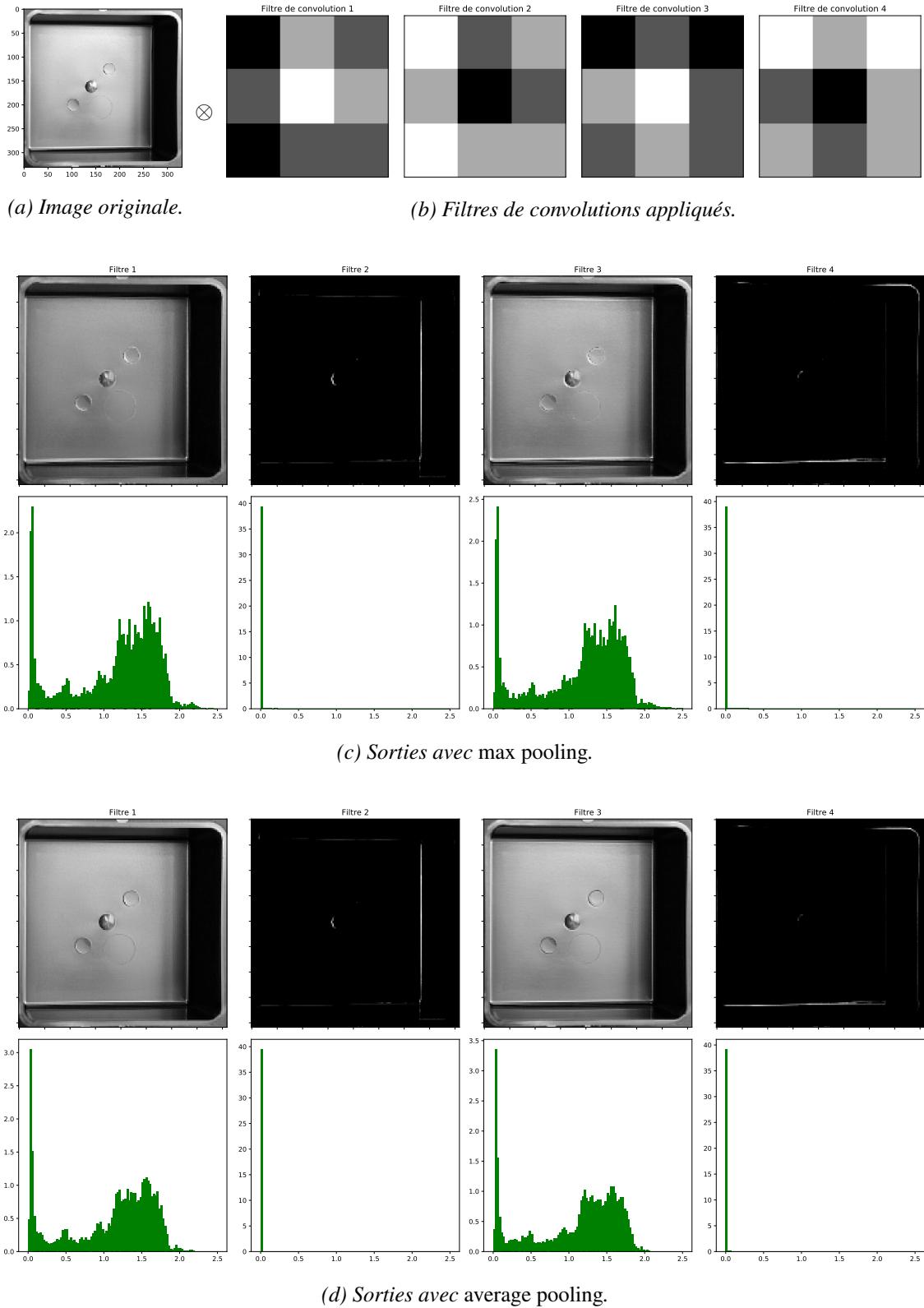


Figure 3.8 – Comparaison max pooling **c** et average pooling **d** pour 4 filtres de convolutions **b**.

OQUAB et al. [Oqu+15] étudie l'utilisation du *global max pooling* : seule une unique valeur maximale de la couche présente  $z$  est conservée, Équation 3.24. Cela fait émerger une propriété intéressante : la possibilité d'utiliser la valeur de la couche précédente  $z$  comme une carte de saillance de la localisation des différentes classes. Cette carte est générée à partir d'un jeu de données sans information sur la localisation des objets, de manière semi-supervisée. La couche de *global max pooling* force le réseau à détecter la position de la meilleure classe dans l'image.

$$y_{i+1} = \max_{j,k} z_{j,k}(y_i) \quad (3.24)$$

De récentes analyses montrent que le *max pooling* est une source majeure de perte d'information, ce qui peut limiter les performances, particulièrement pour les réseaux génératifs. C'est pourquoi les architectures récentes cherchent à le remplacer par des couches de sous-échantillonnage par moyenne (*average pooling*). Dans son réseau composé de sous blocs denses, [LCY13] introduit le *Global Average Pooling*, Équation 3.25. [Zho+15] montre l'intérêt du *Global Average Pooling* lors de l'apprentissage, ce qui permet d'obtenir des cartes de saillance de la localisation des objets dans l'image. Cette démarche est similaire aux travaux de OQUAB et al. [Oqu+15] avec le *global max pooling*, mais l'utilisation de la moyenne force le réseau à détecter l'ensemble des régions où la classe est présente. Cette carte de saillance par classes  $z$  est nommée *Class Activation Map*.

$$y_{i+1} = \frac{1}{N} \sum_{j,k} z_{j,k}(y_i) \quad (3.25)$$

### Pénalisation sur les poids

Proposé par KROGH et HERTZ [KH91], il s'agit d'ajouter un terme de pénalisation  $E$  de la valeur des poids à la fonction de coût  $\mathcal{L}$  que l'on cherche à optimiser. Un facteur  $\lambda$  pondère la valeur des poids  $\mathbf{W}$  du réseau. Cette méthode permet de rendre le réseau robuste aux perturbations ce qui améliore les performances. C'est une méthode simple et efficace, dont l'utilisation est indispensable pour atteindre les performances de l'état de l'art de la littérature. La pondération peut être effectuée sur la norme  $\ell_1$ , ce qui pénalisera tous les poids différents de zéro. Mais la pondération est généralement réalisée sur la norme  $\ell_2$ , ce qui pénalisera les grandes valeurs des poids. Dans ce cas, cette démarche est équivalente à la régularisation de TIKHONOV [Tik43 ; TA77] par une matrice  $\Gamma$ , avec  $\Gamma = \lambda I$ . On cherche à amener un maximum de poids du réseau à zéro. La méthode est d'autant plus critique que les erreurs d'arrondis, inhérentes à la représentation des nombres informatiques, doivent être évitées lors de l'optimisation par descente du gradient.

$$E(\mathbf{W}) = E_0(\mathbf{W}) + \frac{1}{2} \lambda \sqrt{\sum_i \mathbf{w}_i^2} \quad (3.26)$$

### Augmentation artificielle du jeu de données

La plus simple des méthodes existantes pour limiter le sur-apprentissage est de disposer d'un jeu de données comportant une grande variété d'échantillons par classe. *ImageNet* propose 2000 échantillons par classe. Récemment, *ResNext\_wsl* [Mah+18] est entrainé sur plusieurs millions d'images par classe. Lorsque que l'on ne dispose pas de suffisamment de données, il est possible d'augmenter les données existantes. Il s'agit d'effectuer des transformations affines ou non linéaires, rotations, cadrages ou d'ajouter du bruit dans les images. L'objectif est d'augmenter la variance du jeu de données initial. Les taux de modifications optimaux doivent être choisis par validation croisée pour obtenir les meilleures performances possibles sur le problème spécifique. Nous observons de bonnes performances avec des rotations jusqu'à cinq degrés et des cadrages qui varient de dix pour-cent des dimensions originales, illustrés dans la Figure 3.9.

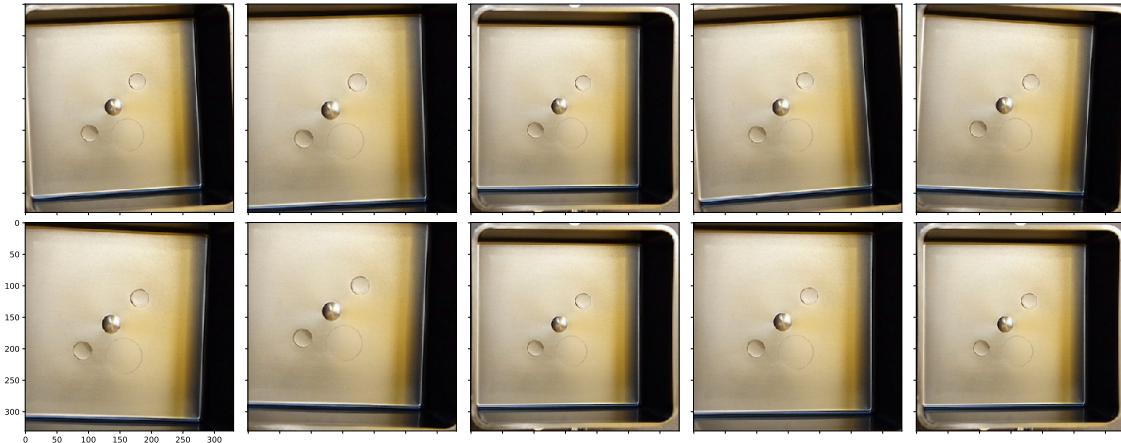


Figure 3.9 – Augmentation d'une image par translations et rotations.

Le choix des paramètres d'augmentation peut être effectué par validation croisée afin d'optimiser les performances. On préférera se rapprocher des situations réelles de production des données. Dans notre cas d'application industrielle du contrôle qualité, de légères variations de cadrage, rotation ou de perspectives peuvent exister. En revanche, nous travaillons avec un éclairage contrôlé, donc les perturbations d'éclairage ou de colorimétrie sont peu probables. Les paramètres de l'augmentation du jeu de données dépendent de la connaissance du métier. Si la connaissance du métier n'est pas certaine, il est préférable d'acquérir un plus grand nombre d'images réelles, plutôt que de générer des données artificielles avec de mauvais a priori.

### 3.3.2 Apprentissage par transfert de domaine

Le faible nombre d'échantillons disponibles (inférieur à mille pièces) nous a conduit à utiliser des réseaux de neurones profonds préalablement entraînés sur des très grands jeux de données. Le jeu de données *ImageNet* [Den+09] est composé de 2 millions d'images réparties en 1000 catégories différentes. L'apprentissage par transfert de domaine, aussi résumé "apprentissage par transfert", cherche à réutiliser un modèle spécialisé sur un certain problème, pour résoudre un autre problème. En pratique, les meilleures approches proposées dans la littérature nécessitent de conserver des caractéristiques fortes entre les deux domaines. Nous recommandons l'utilisation du même format de données initiales. Nous nous contenterons ici d'utiliser des modèles spécialisés sur l'analyse d'images à trois canaux de couleurs, et de les utiliser sur notre propre problématique.

Dans ce cadre d'apprentissage par transfert de domaine, nous n'avons pas la liberté de modifier l'architecture du réseau pré-apris, car il faudrait alors réapprendre les poids des neurones du réseau. La technique répandue est alors de supprimer la ou les dernières couches du réseau pré-apris et d'ajouter de nouvelles couches spécifiques à notre problème. Les dernières couches du réseau sont considérées comme étant le véritable classifieur. Les couches précédentes se contentent de compresser l'information en une représentation dans un espace de dimensions réduites. Nous travaillons uniquement sur l'architecture du classifieur, c'est à dire de la fin du réseau. Le choix de l'architecture et de la typologie des couches influence fortement la performance du classifieur, c'est pourquoi des méthodes de sélections automatiques de ces paramètres doivent être employées. Elles sont discutées dans la Section 3.4.

On distingue deux étapes successives à l'entraînement d'un réseau de neurones par transfert : l'entraînement des nouvelles dernières couches, puis la spécialisation de l'ensemble du réseau. Il s'agit dans un premier temps de garantir la convergence du classifieur sur notre problème. On cherche à obtenir un score de classification satisfaisant. C'est pourquoi dans un premier, seuls les poids des neurones des nouvelles dernières couches sont ajustés par rétro-propagation du gradient

de l'erreur. Une fois que la convergence des dernières couches a été obtenue, il est possible de spécialiser finement l'ensemble du réseau en réalisant un ajustement de l'ensemble des poids des couches afin d'améliorer le score de classification. Nos travaux ont montré que cette étape de spécialisation n'est pas nécessaire lorsque le jeu de données d'apprentissage est petit (inférieur à mille échantillons). En effet, nous n'observons pas d'augmentation du score de classification après la phase de spécialisation. De plus, la durée de l'apprentissage de spécialisation est deux ordres de grandeur plus longue que la durée de l'apprentissage des dernières couches du réseau, car le nombre de variables à ajuster sur l'ensemble du réseau est bien plus grande. La durée de l'apprentissage des dernières couches du réseau est de dix secondes sur un processeur graphique récent et puissant (*NVIDIA GTX1080Ti*, 2017). La durée de l'apprentissage de spécialisation est de dix minutes. Cette durée rend l'apprentissage de spécialisation compatible avec une mise à jour horaire du classifieur, sous réserve de disposer d'un processeur graphique adapté. L'apprentissage des dernières couches peut en revanche être réalisé en cycle de production, pièce après pièce. Un classifieur de la qualité qui utilise l'apprentissage par transfert sera adapté aux productions dans lesquels les changements de séries ou bien la dérive du procédé est supérieure à l'heure. Ce qui est compatible avec une mise à jour régulière du modèle de la qualité.

### 3.3.3 Apprentissage semi-supervisé

Le coût de l'annotation humaine est élevé. De plus, la problématique qui oppose qualité de l'annotation contre quantité d'annotation est critique dans le contexte du contrôle industriel. Les annotations doivent avoir un niveau de certitude élevé afin d'éviter les erreurs de types faux-positifs et faux-négatifs. Il vaut mieux que l'expert humain passe du temps pour annoter une pièce avec certitude, plutôt qu'il annote rapidement plusieurs pièces, avec nécessairement moins de certitudes. C'est pourquoi, les méthodes d'apprentissage qui nécessitent peu d'exemples par classe (*few-shot learning*), voire un unique exemple par classe (*one-shot learning*), sont très utiles. De plus, l'apprentissage semi-supervisé permet d'utiliser des échantillons qui n'ont pas été annotés. Le jeu de données d'apprentissage est alors plus grand et les performances en classification sont augmentées. L'apprentissage semi-supervisé repose sur plusieurs hypothèses sur la structure des données §3.3.3, synthétisées dans la monographie de référence sur le sujet *Semi-supervised learning* [CSZ10]. Plus récemment, la thèse de DURAND [Dur17] propose trois méthodes de classification semi-supervisées qui utilisent des réseaux de convolutions profonds.

**Hypothèse de continuité** : les points qui sont proches dans l'espace des données ont plus de chances d'appartenir à la même classe.

**Hypothèse d'ensembles** : les points sont souvent regroupés en ensembles discrets (paquets ou *clusters*). Les points dans le même ensemble ont plus de chances d'appartenir à la même classe.

**Hypothèse de densité faible** : du point de vue de la séparation des classes, la limite entre deux classes devrait se situer dans une région de l'espace où la densité de présence d'ensembles de différentes classes est faible.

**Hypothèse de variétés** : pour un espace de données initial, il existe un espace latent de dimensions réduites qui conserve les propriétés des données, pour les trois hypothèses précédentes.

L'hypothèse de variétés rend possible l'apprentissage de la transformation vers l'espace latent à l'aide de l'ensemble des données (annotées ou non), puis d'apprendre le classifieur dans l'espace latent. Ainsi, les données non annotées permettent de mieux capturer la distribution des données dans l'espace initial, puis dans l'espace latent. De plus, utiliser une plus grande quantité d'échantillons annotés partiellement permet d'obtenir un classifieur qui sera généralisable pour une plus grande diversité de nouveaux échantillons. Nous détaillerons dans cette partie les

méthodes les plus utiles à notre problème de classification de la qualité à l'aide d'un jeu de données partiellement annoté.

### 3.3.3.1 Propagation d'annotations dans l'espace latent d'un classifieur SVM

Cette approche propose d'utiliser l'hypothèse d'ensembles et l'hypothèse de variétés, §3.3.3. Les annotations sont propagées de manière itérative depuis les échantillons annotés, vers les échantillons non annotés. À chaque itération, les échantillons nouvellement annotés sont utilisés pour continuer à propager les annotations. Cette approche nécessite que les prédictions des nouvelles annotations soient correctes, ainsi que l'hypothèse d'ensembles. Des méthodes de régularisation et la nature itérative de la méthode permettent de moyenner les erreurs d'annotations. De plus, l'hypothèse des variétés est utilisée pour réaliser la propagation des annotations dans un espace latent, à partir des annotations des échantillons annotés, vers les échantillons non-annotés, qui sont voisins. Nous évaluons sur notre problème deux algorithmes implémentés dans la librairie *Scikit-Learn* [Ped+11].

Afin de réduire la dimension de l'espace des données initiales, nous utiliserons un réseau de neurones *DenseNet121* [Hua+16] (pré-entraîné sur le jeu de données *ImageNet*), comme extracteur de descripteurs pertinents. Le choix de ce réseau ainsi que de cette méthode est détaillé dans la Section 3.3.2. A partir des images brutes non modifiées, nous obtenons alors un jeu de données initiales avec pour chaque échantillon un vecteur de dimensions  $1920 \times 7 \times 7 = 94080$ . Nous utilisons ensuite l'Analyse par Composante Principale [Jol02] afin de réduire les dimensions à 20 composantes. La moyenne de la variance perdue sur l'ensemble du jeu de données lors de cette ACP est de deux pour-cent ce qui est un compromis acceptable. Une alternative à l'ACP est l'utilisation de réseau de neurones pour réduire la dimension. On se rapporte alors à l'apprentissage par transfert de domaine présenté dans la Section 3.3.2. Le coût d'utilisation de ces deux méthodes est élevé :  $O(N^3)$ , avec  $N$  le nombre d'échantillons en présence.

Dès 2002, ZHU et GHAHRAMANI [ZG02] proposent de construire le graphe qui relie l'ensemble des échantillons par le chemin le plus court. L'apprentissage est réalisé de manière itérative en propageant les annotations à partir des échantillons annotés, vers les échantillons non annotés. La convergence de la méthode est obtenue lorsque tous les échantillons sont annotés. Soit  $Y_L = \{y_1 \dots y_l\}$  les annotations des  $l$  échantillons annotés et  $Y_U = \{y_{l+1} \dots y_{l+u}\}$  les annotations des  $u$  échantillons non annotés. Enfin,  $X = \{x_1 \dots x_{l+u}\}$  les valeurs de l'ensemble des échantillons, annotés et non annotés. Le problème cherche à trouver les valeurs de  $Y_U$  pour tous les  $X$  et les  $Y_L$ . Un graphe pondéré  $w_{ij}$  entre les échantillons dans l'espace latent est construit. Il est pondéré en fonction de la distance Euclidienne entre chaque échantillon, les échantillons proches ayant un poids plus élevé.  $w$  est aussi appelé la matrice d'adjacence. Enfin, le graphe de transition entre les échantillons est construit :  $T_{ij} = P(j|i) = \frac{w_{ij}}{\sum_{k=1}^{l+u} w_{kj}}$ . L'algorithme consiste alors à appliquer  $T$  à  $Y$  ( $Y = Y \cdot T$ ) jusqu'à obtenir le critère de convergence. À chaque itération de l'algorithme, la pondération des échantillons annotés est augmentée par un facteur  $\alpha$  afin de rester supérieure aux probabilités des échantillons non annotés. Dans le cas où  $\alpha = 0.1$ , seul 10% des annotations seront ajustées et 90% des annotations initiales seront conservées à la suite d'une itération. Cette description succincte omet les différents termes de régularisation et le lecteur pourra se référer à la publication initiale pour plus de détails.

Par la suite, ZHOU et al. [Zho+03] font évoluer la méthode de construction du graphe de transition en utilisant une matrice laplacienne normalisée :  $T_{laplacienne} = I - D^{-1/2} T D^{-1/2}$ , avec  $T$  la matrice d'adjacence précédemment utilisée,  $I$  l'identité et  $D$  la matrice des degrés, diagonale, qui renseigne le nombre de liens de chaque sommets. Cette matrice de transition modélise les distances entre les échantillons.

Enfin, la matrice d'adjacence  $w_{ij}$  peut être approximée par la méthode des K-plus-proches-voisins. Le gain en performance est significatif puisque les dimensions de la matrice d'adjacence réduit de  $(N^2, N^2)$  à  $(k, N)$ . Sur notre problème qui peut comporter 1000 échantillons, le gain est

déjà d'un facteur 1000.

### 3.3.3.2 Réseau de neurones siamois profond

Un réseau siamois (*Siamese Networks*) est un réseau de neurones qui contient deux sous-réseaux identiques. Le premier réseau siamois a été appliqué à la vérification d'empreintes digitales, en 1992 [BC93]. Deux réseaux de convolutions identiques sont utilisés pour l'analyse de similitude entre deux échantillons. Puis la distance entre la sortie des deux réseaux est calculée. Les distances employées sont souvent la norme, la norme Euclidienne ou la fonction de coût de contraste proposée par HADSELL, CHOPRA et LECUN [HCL06], Équation 3.27.

Enfin, une fonction exponentielle normalisée (*SoftMax*) sur ces valeurs retourne la probabilité de similarité entre les deux échantillons. Ce travail détaille la méthode employée pour prétraiter les images à partir des données enregistrées. En parallèle, une architecture similaire a aussi été appliquée à la vérification de signatures [Bro+93]. Un modèle à réseau de neurones est spécifiquement utilisé pour exploiter pleinement les informations temporelles obtenues lors de l'enregistrement de la signature par la tablette *5990 Signature Capture Devices* : position cartésienne, distance du stylo au support, pression sur la pointe, le tout inscrit dans le temps ; ce qui rend toute falsification très difficile. Ces deux travaux montrent des résultats supérieurs à 95% de bonnes réponses en classification, dès 1992.

La montée en performance du calcul sur processeur graphique a permis de multiplier les couches de convolutions. L'utilisation de réseaux siamois profonds a ainsi été proposée en 2015 et est devenue une thématique de recherche prolifique. KOCH, ZEMEL et SALAKHUTDINOV [KZS15] les utilisent par exemple pour la détection de caractère sur de multiples alphabets. La Figure 3.10 montre l'architecture d'un tel réseau. L'architecture présentée est celle utilisée par RAO, WANG et COTTRELL [RWC16] pour la détection de similarité entre différentes expressions de visages.

La fonction de coût de contraste, Équation 3.27, se compose de  $D_W$  qui est une distance entre les vecteurs en sortie des deux branches du réseau. Dans le cas le plus simple, c'est la norme Euclidienne  $\ell_2$  entre les vecteurs  $s_A$  et  $s_B$ . Dans cette équation,  $D_W$  correspond à la distance entre les matrices de poids  $W$ . La matrice de poids  $W$  est l'ensemble des poids de la branche du réseau précédent cette fonction. Ainsi, cette fonction de coût de contraste permet de réaliser l'apprentissage des poids du réseau afin de rapprocher dans l'espace latent les images similaires, et au contraire d'éloigner les images non similaires.  $\mu \in [0; 1]$  est un terme qui permet de balancer les effets d'attraction et de répulsion.

$$\begin{aligned} \mathcal{L}_{\text{contraste}}(0, \mathbf{x}_i, \mathbf{x}_j, \mu) &= (1 - \mu) \frac{1}{2} (D_W)^2 + (\mu) \frac{1}{2} \{ \max(0, m - D_W) \}^2 \\ \text{avec } D_W &= \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|_2^2 \end{aligned} \quad (3.27)$$

Afin que la fonction de coût réagisse en particulier aux paires d'images non similaires, on introduit une marge  $m$ , de valeur supérieure à zéro, qui permet aux paires d'images dont la distance est inférieure à cette marge de ne pas être prises en compte dans le résultat du calcul du coût. La valeur de la marge peut être choisie selon les résultats de précision et de résidus souhaités. Dans notre cas de contrôle qualité, nous souhaitons par exemple limiter les faux positifs : les pièces mauvaises qui sont classées comme étant bonnes. Minimiser le coût de contraste revient à positionner toutes les images d'une même classe en un unique point de l'espace. Cette propriété peut devenir une limite lorsque des images d'une même classe possèdent des caractéristiques d'autres classes. C'est pourquoi la fonction de coût de triplet a été proposée et nous la détaillons dans la Section suivante 3.3.3.3.

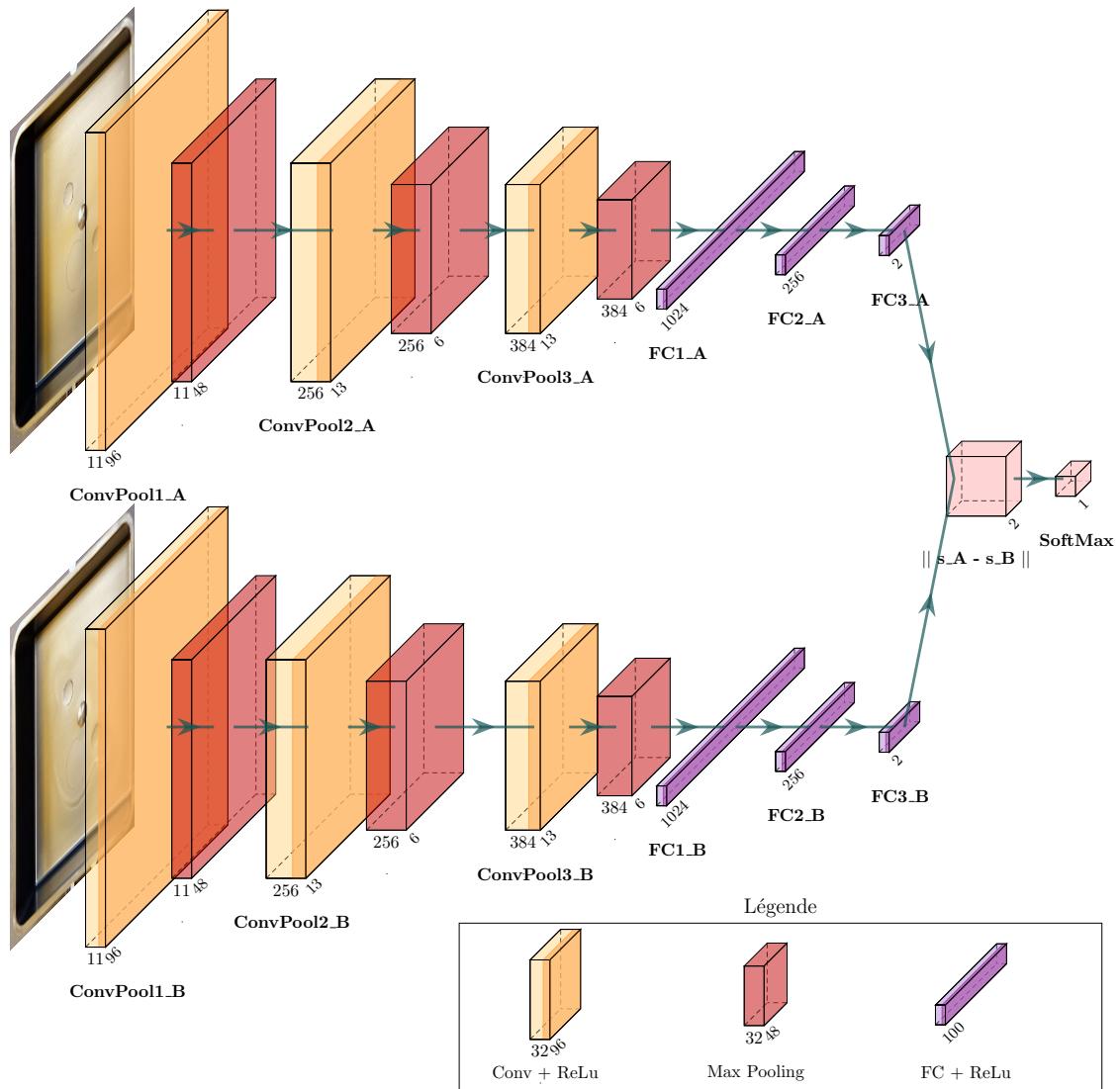


Figure 3.10 – Architecture d'un réseau siamois.

### 3.3.3.3 Réseau de neurones de triplets profond

Basé sur la même idée que le réseau siamois, le réseau de triplets (*Triplet Networks*) utilise trois réseaux de convolutions en parallèles. Les trois sorties des convolutions sont utilisées par la fonction de coût de triplets (*Triplet Loss*, Équation 3.29), proposée par [Wan+14] WANG et al., appliquée à la détection des images au contenu similaire.

La fonction de coût de triplets est dérivée de la fonction de coût *LMNN* (*Large Margin Nearest Neighbor*), Équations 3.28, proposées par WEINBERGER, BLITZER et SAUL [WBS06]. *LMNN* s'appuient sur la distance de MAHALANOBIS [Mah36] (ici notée  $\mathbf{L}_{Maha}$ ) pour la classification par la méthode des K-plus-proches voisins. Le terme  $\varepsilon_{\text{éloignement}}$  pénalise les petites distances entre les points. Il garantit également que les exemples négatifs sont éloignés des exemples de la classe  $i$ .  $\mu \in [0; 1]$  permet de balancer les deux termes (en pratique  $\mu = 0,5$  est satisfaisant).

$$\begin{aligned} \mathcal{L}_{\text{LMNN}}(0, \mathbf{x}_i, \mathbf{x}_j, y, \mu) &= (1 - \mu)\varepsilon_{\text{rapprochement}} + \varepsilon_{\text{éloignement}} \\ \varepsilon_{\text{rapprochement}}(\mathbf{L}_{Maha}) &= \sum_{j \sim i} \left\| \mathbf{L}_{Maha}(\vec{\mathbf{x}}_i - \vec{\mathbf{x}}_j) \right\|^2 \\ \varepsilon_{\text{éloignement}}(\mathbf{L}_{Maha}) &= \sum_{i, j \sim i} \sum_n (1 - y_n) \max \left( 0, 1 + \left\| \mathbf{L}_{Maha}(\vec{\mathbf{x}}_i - \vec{\mathbf{x}}_j) \right\|^2 - \left\| \mathbf{L}_{Maha}(\vec{\mathbf{x}}_i - \vec{\mathbf{x}}_n) \right\|^2 \right) \end{aligned} \quad (3.28)$$

Cette méthode est proche de la méthode employée par les Machines à Vecteurs Supports (*Support Vector Machines*) : maximiser la marge séparant les classes dans l'espace métrique appris. La fonction de coût  $\varepsilon_{\text{éloignement}}$  est composée du coût de *hinge* (*Hinge loss*,  $\mathcal{L}_{\text{Hinge}}(z) = \max(0, z)$ ) aussi utilisé par les *SVM*. Cependant, contrairement aux *SVM*, la méthode peut être utilisée pour de la classification multi-classes, sans modification et avec de meilleures performances que l'algorithme *SVM*.

Récemment, SCHROFF, KALENICHENKO et PHILBIN [SKP15] ont fait évoluer le coût  $\mathcal{L}_{\text{LMNN}}$  et propose la fonction de coût de triplet  $\mathcal{L}_{\text{triplet}}$  Équation 3.29. Trois images sont présentées en entrée au réseau : une image A de référence, une image Positive d'une catégorie et une image Négative d'une autre catégorie. La distance entre l'image R de Référence et l'image P Positive doit être plus petite que la distance entre l'image R et l'image N Négative. La fonction de coût de triplets cherche à entraîner le réseau afin séparer la paire positive (image R, image P) de la paire négative (image A, image N) par une distance de marge : c'est à dire minimiser la distance entre (images R, image P) et maximiser la distance entre (image R, image N).

Les travaux de WANG et al. [Wan+14] et SCHROFF, KALENICHENKO et PHILBIN [SKP15] proposent également l'utilisation d'architectures de réseaux de convolutions profonds issues de la littérature sans les modifier (*AlexNet* modifié [ZF13] et *GoogleLeNet-InceptionV1* [Sze+14]), à la manière de boîtes noires, et d'entraîner l'ensemble du réseau ainsi constitué. À la suite du réseau, la norme Euclidienne entre les trois branches est calculée. Ainsi, la distance entre toutes les images qui sont de la même catégorie sera petite, tandis que la distance entre les paires d'images de différentes catégories sera grande. WU et al. [Wu+17] justifient l'utilisation de la norme Euclidienne au lieu de la norme Euclidienne élevée au carré utilisée dans SCHROFF, KALENICHENKO et PHILBIN [SKP15].

A la différence de la fonction de coût de contraste (Équation 3.27) qui cherche à positionner toutes les images d'une même classe en un unique point de l'espace latent, la fonction de coût de triplet cherche à maximiser une marge entre chaque paire d'images de catégories différentes. Ainsi, les images similaires peuvent être positionnées librement à des points différents de l'espace latent, sous réserve que la condition sur la marge entre les catégories différentes soit remplie.

$$\mathcal{L}_{\text{triplet}}(\mathbf{x}_r, \mathbf{x}_p, \mathbf{x}_n) = \max(0, \|f(\mathbf{x}_r^R) - f(\mathbf{x}_i^P)\|_2 - \|f(\mathbf{x}_i^R) - f(\mathbf{x}_i^N)\|_2 + \alpha) \quad (3.29)$$

$\mathbf{x}_i^R$  représente l'ensemble des images de référence,  $\mathbf{x}_i^P$  l'ensemble des images de catégorie positive et  $\mathbf{x}_i^N$  de catégorie négative.

Le problème majeur de cette fonction de coût est que sur l'ensemble d'un jeu de données, la majorité des triplets d'images répondront à cette contrainte. Dans ce cas, le réseau ne pourra pas optimiser rapidement ses poids sur cette condition, car la fonction de coût sera le plus souvent constante. C'est pourquoi il est nécessaire de sélectionner judicieusement des triplets d'images qui sont difficiles à discriminer, en commençant au début par des triplets faciles, puis en raffinant la difficulté au fur et à mesure de l'apprentissage par descente de gradient.

Pour l'entraînement optimale du réseau, il s'agit de sélectionner, pour toutes images  $\mathbf{x}_i^R$  de référence, les images  $\mathbf{x}_i^P$  et  $\mathbf{x}_i^N$  telles que les Équations 3.30 soient remplies.

$$\begin{aligned}\mathbf{x}_i^P &= \operatorname{argmax}_{\mathbf{x}_i^P} \|f(\mathbf{x}_i^R) - f(\mathbf{x}_i^P)\|_2 \\ \mathbf{x}_i^N &= \operatorname{argmin}_{\mathbf{x}_i^N} \|f(\mathbf{x}_i^R) - f(\mathbf{x}_i^N)\|_2\end{aligned}\quad (3.30)$$

Les auteurs de [SKP15] appellent ces paires d'images "négatives-difficiles" et "positives-difficiles". Ainsi, les images "négatives-difficiles" sont celles qui possèdent le plus de similarités avec l'image de référence, mais qui devront être les plus éloignées de l'image de référence après l'apprentissage. Réciproquement, les images "positives-difficiles" sont celles qui sont les moins similaires avec l'image de référence, mais qui devront être les plus rapprochées de l'image de référence après l'apprentissage. La figure 3.11 présente les trois cas d'images Négatives  $\mathbf{x}_i^N$  possibles pour un couple d'images (Référence, Positive) :

- Négatives-faciles :  $\|f(\mathbf{x}_i^R) - f(\mathbf{x}_i^N)\|_2 > \|f(\mathbf{x}_i^R) - f(\mathbf{x}_i^P)\|_2 + \alpha$ , dans ce cas  $\mathcal{L}_{triplet} = 0$ .
- Négatives-modérées :  $\|f(\mathbf{x}_i^R) - f(\mathbf{x}_i^P)\|_2 < \|f(\mathbf{x}_i^R) - f(\mathbf{x}_i^N)\|_2 < \|f(\mathbf{x}_i^R) - f(\mathbf{x}_i^N)\|_2 + \alpha$ , avec ici  $0 < \mathcal{L}_{triplet} < \alpha$ .
- Négatives-difficiles :  $\|f(\mathbf{x}_i^R) - f(\mathbf{x}_i^N)\|_2 < \|f(\mathbf{x}_i^R) - f(\mathbf{x}_i^P)\|_2$ , on a  $\mathcal{L}_{triplet} > \alpha$ .

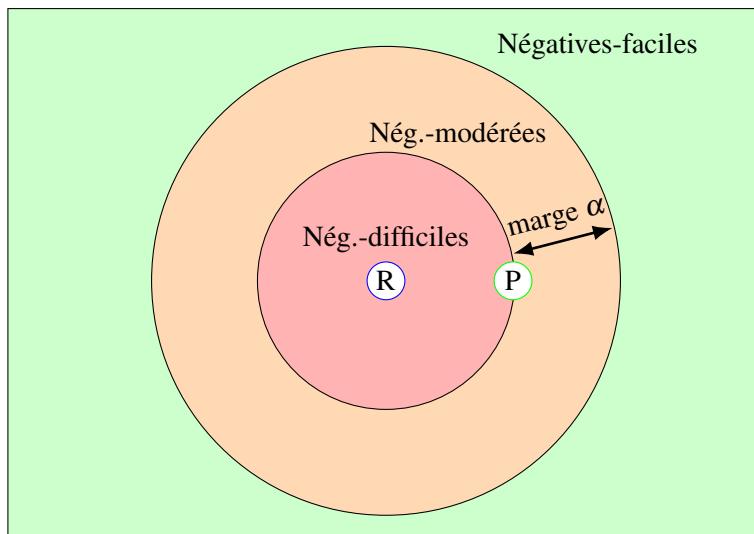


Figure 3.11 – Trois différents types d'images Négatives possibles pour un couple d'images (Référence, Positive). Figure inspirée de l'[article de blog de O. Moindrot](#).

La condition de similarité maximale sur les images négatives peut cependant créer un minima local lors de l'apprentissage, en ne sélectionnant que certaines images du jeu de données, ce qui empêche la convergence. Il est nécessaire d'introduire une certaine variabilité dans les images négatives. C'est pourquoi [SKP15] introduit une troisième condition, Équation 3.31.

$$\|f(\mathbf{x}_i^R) - f(\mathbf{x}_i^P)\|_2 < \|f(\mathbf{x}_i^R) - f(\mathbf{x}_i^N)\|_2 \quad (3.31)$$

Il s'agit de sélectionner des images "négatives-modérées", c'est à dire des images qui sont plus éloignées de l'image de référence que les images positives, mais qui ont une similarité élevée avec l'image de référence. Dans la même démarche, SHI et al. [Shi+16] proposent de sélectionner en plus les images "positives-modérées". La Figure 3.12 récapitule la méthode d'apprentissage d'un réseau de triplets.

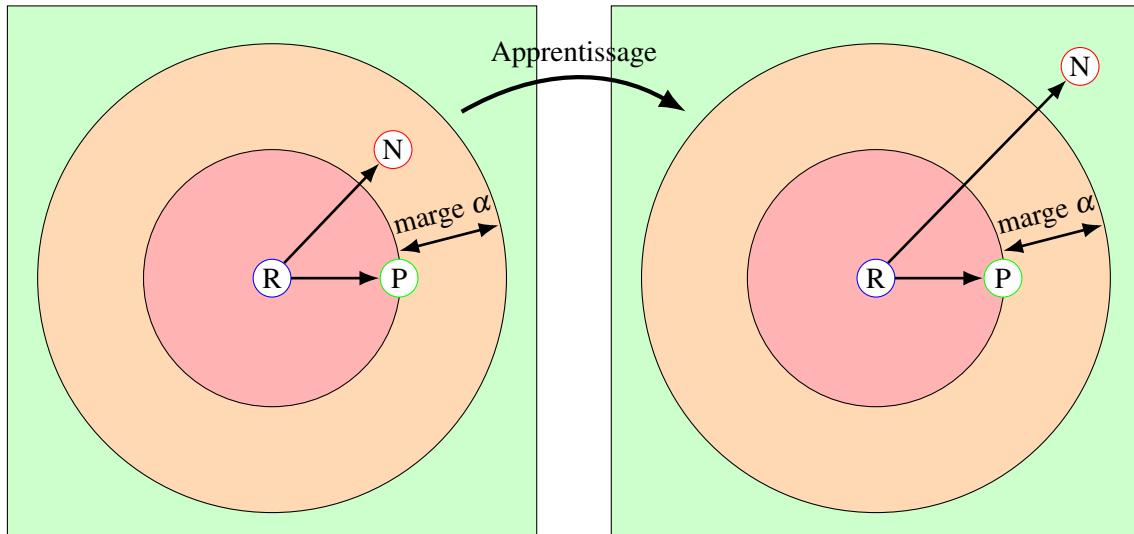


Figure 3.12 – Principe de l'apprentissage d'un réseau de triplets. Figure inspirée de [SKP15].

Le calcul des Équations 3.30 pour la sélection d'images sur l'ensemble du jeu de données est très coûteux. C'est pourquoi ce calcul est traditionnellement réalisé en amont de l'apprentissage. De plus, pour sélectionner les images parmi l'ensemble du jeu de données, il est nécessaire que le jeu de données puisse être chargé dans la mémoire de travail de la machine qui effectue le calcul des conditions. Les jeux de données actuels sont souvent trop volumineux pour pouvoir être chargés en entier dans la mémoire. C'est pourquoi il a été proposé de réaliser ce calcul pendant l'apprentissage [Wan+14; SKP15]. Pour maximiser les performances du calcul de la méthode d'optimisation des poids par descente de gradient sur processeur graphique, les données d'apprentissage sont séparées en paquets (*mini-batch*) avant d'être présentées en entrée du réseau pour l'apprentissage. Pour chaque paquet, les couples d'images "négatives-modérées" sont alors sélectionnées de manière itérative, en parallèle de l'apprentissage. Récemment, HERMANS, BEYER et LEIBE [HBL17; Yu+18] proposent de sélectionner les images négatives-difficiles dans les paquets. Sur leur problème spécifique de la détection de visages, ils montrent qu'il n'est pas nécessaire d'introduire la condition de marge modérée car le fait de découper le jeu de données en paquets aléatoires garantie une certaine variété d'images dans les paquets.

La Figure 3.13 présente les valeurs de la dernière couche d'un réseau de triplets, de 200, projetée en deux dimensions par l'algorithme *t-SNE* §3.3.4.5. Les différentes classes annotées par les experts humains sont bien séparées. Les pièces non annotées sont également positionnées dans les bonnes classes. Enfin, il est intéressant d'observer qu'un paquet correspondant à une erreur de mesure, comme par exemple l'absence de pièces et la capture du bras robotique préhenseur est mis en évidence.

L'apprentissage de métrique est un domaine de recherche actif. SOHN [Soh16] généralisent par exemple ces méthodes à N-paires d'images et montre l'intérêt d'une fonction de coût à 72-paires en comparaison des triplets (3-paires). SONG et al. [Son+15] proposent une méthode d'adaptation de domaine : un modèle pré-appris peut être complété à l'aide de nouveaux échantillons. [Wan+17] WANG et al. proposent d'utiliser une fonction de coût basée sur une projection dans un espace angulaire. DUAN et al. [Dua+18] proposent la méthode *DAML* (*Deep Adversarial*

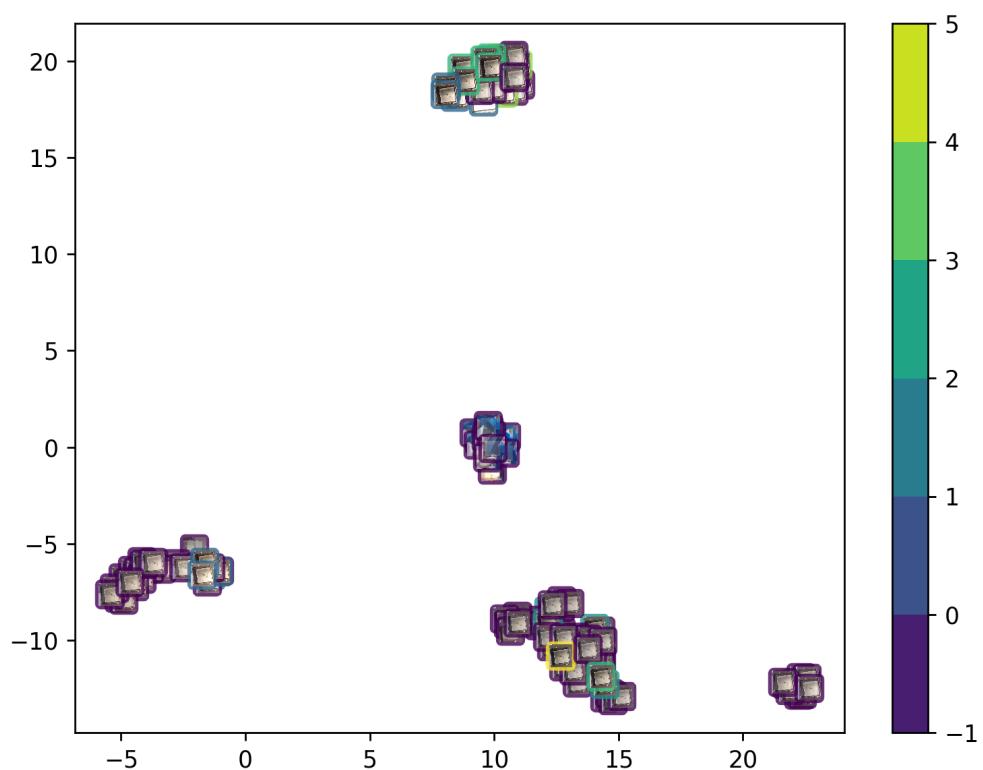


Figure 3.13 – Projection par t-SNE à partir des valeurs de la dernière couche d'un réseau de triplet. L'échelle 1-5 correspond aux niveaux de qualité évalués par les experts humains ; -1 correspond aux pièces non annotées.

*Metric Learning*) qui utilise un réseau antagoniste génératif (détaillé dans la Section 3.3.4.8) afin de produire une image négative-difficile artificielle, à partir d'un triplet d'images défini. Les résultats obtenus montrent l'intérêt de la méthode dans le cas où le nombre d'échantillons est faible. Dernièrement, ZHENG et al. [Zhe+19] proposent une comparaison de ces méthodes (N-paires, métrique angulaire et *DAML*) sur trois jeux de données de références.

### 3.3.3.4 Intérêt des méthodes issues de l'apprentissage par renforcement

Les récents succès des méthodes d'apprentissage par renforcement peuvent également être inclus dans cette section. À chaque itération, l'algorithme choisit de réaliser une action et il obtient, pour cette action, une récompense, positive en cas d'amélioration de l'objectif, ou négative dans le cas inverse. L'apprentissage par renforcement est particulièrement adapté au réglage optimal des procédés. Il est nécessaire de disposer d'un environnement capable de recevoir une action et de produire une récompense. Cet environnement peut être une simulation réaliste de l'environnement, ou bien l'environnement réel. Dans notre cas, les actions seraient la modification des réglages du procédé et la récompense, la pièce produite. Cependant, nous ne disposons actuellement pas de simulation informatique réaliste de notre problème. De plus, l'évaluation en conditions réelles nécessite de disposer d'une machine pilotée par l'algorithme, qui produit un minimum de mille pièces. Néanmoins, dans le contexte d'une production en quantité industrielle, l'utilisation de l'apprentissage par renforcement devient possible. C'est une piste de recherche qui devra être évaluée à la suite de ces travaux.

### 3.3.4 Apprentissage non-supervisé

Contrairement aux méthodes précédemment présentées, nous cherchons ici à classer sans aucune assistance d'un expert humain. Supprimer le besoin d'annotation d'un expert amène deux bénéfices : une disponibilité totale du système et une suppression des risques d'erreurs humaines. Ces méthodes permettent de réduire la dimension des données et de classifier en utilisant uniquement les hypothèses sur la structure des données, précitées dans la Section 3.3.3, sans annotation. La réduction de dimensions est un enjeu important dans notre problématique industrielle. En effet, les données enregistrées par les capteurs ont un ordre de grandeur de l'ordre du million de valeurs scalaires par pièces. Le traitement de la mesure doit être effectué sur un système embarqué, intégrable à la ligne de production, qui a une puissance de calcul faible. Plus la dimension de l'information issue des capteurs est réduite, plus l'inférence de la qualité sera rapide. De plus, la réduction de l'information est un objectif crucial afin de pouvoir piloter en boucle fermée le procédé. En effet, le pilotage multivarié des procédés industriels est possible dans un espace de dimensions faibles (de dimensions dix à cent). Il existe deux grandes classes de méthodes de réduction de dimensions. Celles-ci s'appuient sur deux démarches distinctes : la factorisation de matrices (ACP §3.3.4.3, ACP à noyau §3.3.4.4, Auto-encodeurs) et les graphes de voisnages (K-moyennes §3.3.4.1, §3.3.4.5, §3.3.4.2, §3.3.4.6). Nous évaluons dans cette section les algorithmes les plus importants que nous utilisons pour répondre à notre problème de contrôle automatique de la qualité des produits. La réduction de dimensions et le *clustering*<sup>3</sup> spectral (aussi appelé "apprentissage de variétés") sont souvent combinées. Dans un premier temps on réduit la dimension de l'information, puis on regroupe par paquet les échantillons (*clustering*) à partir de l'espace réduit.

---

3. Nous employons le terme anglais *clustering* en lieu et place du mot français "regroupement", afin de préciser qu'il s'agit de réaliser un regroupement algorithmique des données en paquets de données similaires, et non pas d'autres formes de regroupement.

### 3.3.4.1 K-moyennes

L'algorithme des K-moyennes (*K-means*), proposé par MACQUEEN [Mac67; Llo82] cherche à séparer les échantillons  $X$  en  $C$  groupes de classes telle que la variance soit égale entre les groupes. C'est une méthode de *clustering* spatial dans l'espace des données initiales. Soit  $c_j$  le barycentre d'un groupe dans l'espace des données, l'algorithme cherche à minimiser la distance Euclidienne des échantillons au point  $c$ . Il s'agit de minimiser l'inertie<sup>4</sup> du groupe d'échantillon  $c_j \in C$ . L'algorithme est itératif. Le nombre de groupes doit être préalablement connu. La position initiale des points  $c_j$  est initialisée aléatoirement. À chaque itération, les échantillons  $\mathbf{x}_i \in X$  sont associés au point  $c_j$  le plus proche. La position de  $c_j$  est alors mise à jour comme barycentre des points  $\mathbf{x}_i$  qui lui sont associés, Équation 3.32.

$$C(X) = \arg \min_{c_j \in C} \sum_{j=1}^{n_C} \sum_{i=1}^n (\|\mathbf{x}_i - c_j\|^2) \quad (3.32)$$

La convergence est nécessairement atteinte mais elle peut être un minimum locale. Le résultat final est fortement dépendant de la position initiale des points centroïdes  $c$ . C'est pourquoi l'algorithme complet est répété de nombreuses fois avec des initialisations de  $c_j$  différentes. La méthode d'initialisation *k-means++*, plus performante, a été proposée par ARTHUR et VASSILVITSKII [AV07]. Elle propose de positionner les points centroïdes  $c$  de manière à les éloigner au maximum entre eux. Les résultats obtenus sont ainsi meilleurs. Enfin, la distance Euclidienne  $\ell_2$  est une métrique qui répond mal aux problèmes en très grandes dimensions ou aux formes de groupes allongées ou irrégulières. Cette distance ne prend pas en compte la variation de la densité des points dans l'espace. C'est pourquoi une transformation de l'espace des données est souvent réalisée en amont de cet algorithme (ACP, *kernel trick* ...). Des méthodes de *clustering* spectral ont alors été proposées pour répondre à cette limite.

La complexité de l'algorithme K-moyennes est de l'ordre de  $O(N \log(N))$ . C'est la méthode de *clustering* la plus efficiente qui existe.

### 3.3.4.2 HDBSCAN

ESTER et al. [Est+96] proposent la méthode *DBSCAN* (*Density-Based Spatial Clustering of Applications with Noise*) qui s'appuie sur l'organisation spatiale, en matière de densité, des données. L'algorithme distingue en amont trois catégories d'échantillons dans les données. Pour chaque point  $c$ , on distingue : les points centraux s'ils sont entourés d'un nombre minimal *MinPts* de voisins et sont situés dans un rayon  $\varepsilon$  du point  $c$ ; les points de bordures s'ils ont moins de *nb\_min\_points* voisins et sont situés dans un rayon  $\varepsilon$ ; et les points de bruit, dont la distance avec  $c$  est supérieure à  $\varepsilon$ . Ainsi, le nombre de paquets est défini par *nb\_min\_points* et les points de bordures sont associés au paquet par leur point  $c$  correspondant. À la différence de l'algorithme K-moyennes §3.3.4.1, l'algorithme *DBSCAN* ne cherche pas à produire des paquets circulaires et les points trop éloignés sont ignorés ce qui le rend robuste aux perturbations. Néanmoins, les performances de la méthode *DBSCAN* sont limitées sur les problèmes où les échantillons ont de grandes dimensions. On cherchera à réduire la dimension de l'information, par exemple lorsque l'on traite des images, avant d'appliquer cette méthode. Les hyper-paramètres *nb\_min\_points* et  $\varepsilon$  doivent être optimisés. Trouver un couple optimal peut être difficile, dans le cas où les données sont réparties de manière fortement non homogène ; c'est à dire lorsque les différences de densités de la répartition des points sont grandes.

Récemment, CAMPELLO, MOULAVI et SANDER [CMS13] introduisent la méthode *HDBSCAN* (*Density-Based Clustering Based on Hierarchical Density Estimates*) qui fait évoluer

4. L'inertie est la somme des distances Euclidiennes entre les échantillons du groupe.

*DBSCAN* en prenant en compte l'organisation spatiale des données. *HDBSCAN* ne nécessite pas de paramètre  $\varepsilon$ .

De nombreuses méthodes de *clustering* s'appuient sur l'organisation hiérarchique des données. C'est le *clustering* hiérarchique, qui utilise les arbres binaires de tri. Le résultat est souvent visualisé sous forme de dendrogrammes. Une évolution du *clustering* hiérarchique ajoute l'utilisation de l'analyse spectrale. Ce sont les méthodes de *clustering* spectral qui proposent des solutions à ces limites. Il s'agit d'utiliser les relations entre les échantillons au sein des paquets, par la construction de graphes et par l'analyse spectrale, c'est à dire l'utilisation des vecteurs propres des données. Nous invitons le lecteur à consulter l'excellent tutoriel “A Tutorial on Spectral Clustering” [von07] sur les méthodes de *clustering* spectrales, qui ne peuvent être approfondies dans le cadre de ce travail.

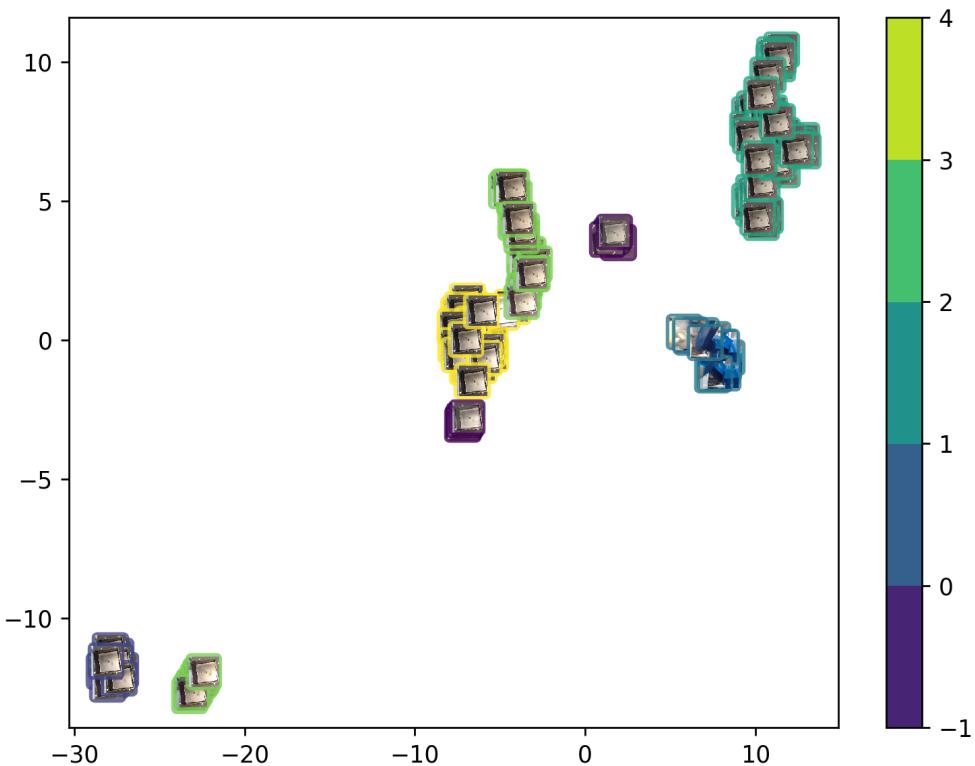


Figure 3.14 – Projection par UMAP à partir de la valeur des pixels des images. L'échelle -1 à 4 correspond aux clusters obtenus par HDBSCAN.

La Figure 3.14 présente les clusters obtenus par l'algorithme *HDBSCAN*. La dimension des données a préalablement été réduite par l'algorithme *UMAP* §3.3.4.6. Les différents niveaux de qualité annotés par l'expertise humaine sont détectés.

### 3.3.4.3 Analyse en Composante Principale

Dès 1901, PEARSON; HOTELLING [Pea01; Hot33] proposent la méthode d'Analyse en Composante Principale (*Principal Component Analysis*). C'est la méthode de réduction de dimensions de référence qui s'appuie sur une factorisation de la matrice représentative des données. Pour un jeu de données  $X$  de  $N$  échantillons, on calcule la matrice de covariance  $\bar{C}$ , Équation 3.33.

$$\bar{C} = \text{cov}(X) = \frac{1}{N} \sum_{n=1}^N (x_n - \bar{x}) \cdot (x_n - \bar{x})^\top \quad (3.33)$$

$$\lambda(X \cdot \mathbf{V}) = (X \cdot \bar{C} \cdot \mathbf{V}) \quad (3.34)$$

Soit  $m$  le nombre de variables par échantillons. Les vecteurs propres  $\mathbf{v}_k$  (avec  $1 \leq k \leq N$ , de la matrice de covariance  $\bar{C}$ ) sont appelés les composantes principales de  $X$ , Équation 3.34. Les valeurs propres  $\lambda_k$  sont utilisées pour ordonner les vecteurs propres  $\mathbf{v}_k$ , dans l'ordre croissant de la variance des données exprimées par chaque vecteur propre. On obtient un espace de dimensions réduites où l'on peut projeter les données. La réalisation d'une ACP entraîne nécessairement une réduction de la variance, en comparaison de la variance de l'ensemble des échantillons du jeu de données initial. La valeur de cette réduction indique la perte d'informations liée à l'ACP. Dans ce travail, nous serons satisfaits des ACP qui ne font perdre que 1% de la variance des données initiales. On choisira le nombre  $k$  de vecteurs supports en fonction de ce critère. Dans notre application, cinquante vecteurs supports sont suffisants.

L'Analyse en Composante Principale peut être employée comme méthode de réduction de l'information, par transformation linéaire. Nous pouvons appliquer une ACP à  $k$  composantes sur l'image brute, ce qui compresse l'information dans un vecteur de dimension  $k$ . Puis nous appliquons la transformée inverse pour obtenir une reconstruction. La Figure 3.15 montre la capacité de l'ACP à encoder les caractéristiques d'une image, en fonction du nombre de composantes. Des artefacts de couleurs (ici vertes) sont présents. L'erreur de reconstruction est la différence des valeurs des pixels entre l'image originale et l'image reconstruite. Le graphique 3.16 montre l'évolution de la qualité de reconstruction en fonction de la proportion de la variance conservée. On observe que l'erreur diminue lorsque le nombre de composantes et la variance conservée augmentent. Il est intéressant d'observer que pour 5 composantes, ce sont les détails des côtés de la pièce et des défauts du fond de la pièce qui ne sont pas représentés. Nous nous intéressons ; dans le cadre de notre problématique de contrôle de la qualité, à ces défauts, mais l'ACP prend en compte en priorité les caractéristiques globales de l'image. Lorsque l'on utilise 100 composantes, l'erreur de reconstruction est négligeable. Des artefacts restent néanmoins présents. Dans les sections suivantes, nous étudierons des méthodes qui permettent de supprimer la présence d'artefacts et de prendre en compte les caractéristiques locales des images.

La Figure 3.17 présente les valeurs des deux dimensions obtenues pour chaque pièce en appliquant une ACP sur la valeur des pixels bruts de l'image. On observe une séparation intéressante des images qui ne contiennent pas de pièce, mais par exemple le bras robotique préhenseur. En revanche, seuls deux paquets sont mis en évidence, parmi lesquels des pièces de bonne qualité et des pièces de mauvaise qualité sont mélangées. L'ACP réalisée sur les pixels bruts est limitée aux motifs globaux de l'image. Il est nécessaire de réaliser un prétraitement des images pour extraire les motifs pertinents sur notre problème avant d'appliquer une ACP. L'expertise humaine est requise pour choisir les méthodes judicieuses de prétraitement à appliquer. Nous détaillons des approches de ce type dans la Section ???. L'ACP utilise ici deux uniques composantes. Comme présenté dans les figures 3.15 et 3.16, une centaine de composantes permettent de modéliser la variance de notre jeu de données de manière satisfaisante. Avec deux composantes, la perte de variance est de 36 %, ce qui est important. Il sera judicieux d'utiliser un nombre de composantes plus important, ou bien d'utiliser une méthode plus robuste tels que le *t-SNE* 3.3.4.5 ou *UMAP* 3.3.4.6.

Le calcul de la matrice de covariance  $\bar{C}$  a une complexité de l'ordre de  $O(m^2 \cdot N)$ . La décomposition en valeur propre a une complexité  $O(m^3)$ . La complexité de l'ACP est  $O(m^2 \cdot N + m^3)$ . Soit, pour un nombre de variables constant  $m \ll N$ , l'ACP a une complexité linéaire  $O(N)$ . Ainsi, le coût de son calcul est négligeable sur les ressources informatiques modernes. C'est une méthode très efficiente de réduction de l'information.

Nombre de composantes	Reconstruction	Différence	Reconstruction	Différence	Variance conservée	Erreur de reconstruction
n=1					41%	0,015
n=2					64%	0,0093
n=5					87%	0,0035
n=100					99.7%	0,00025

Figure 3.15 – Compression par ACP.

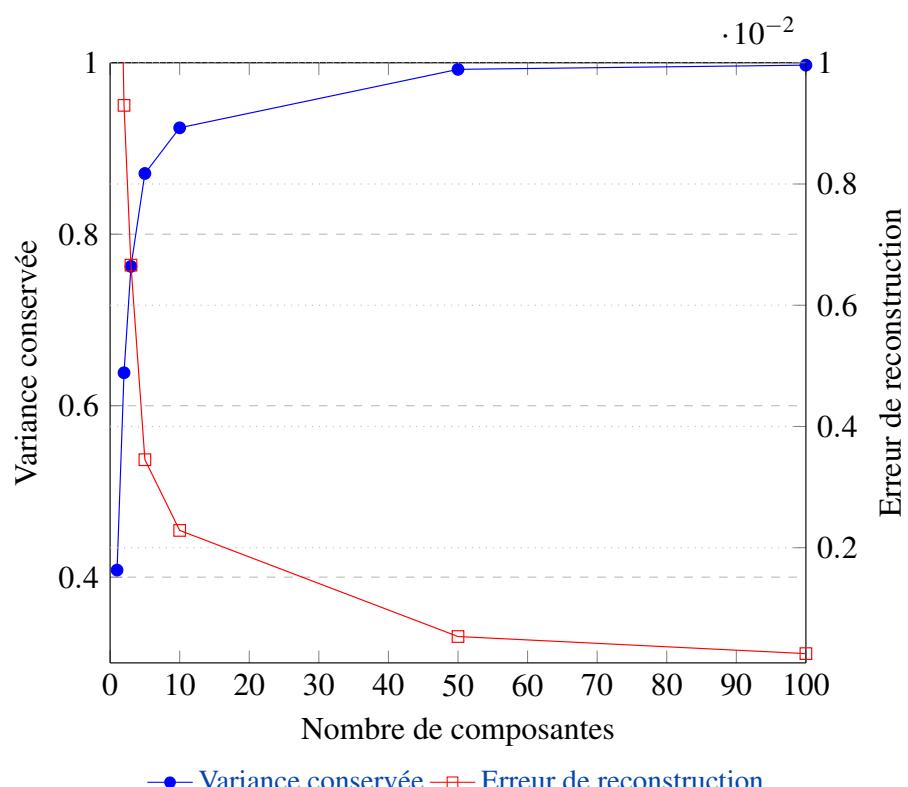


Figure 3.16 – Qualité de la compression par ACP en fonction du nombre de composantes.

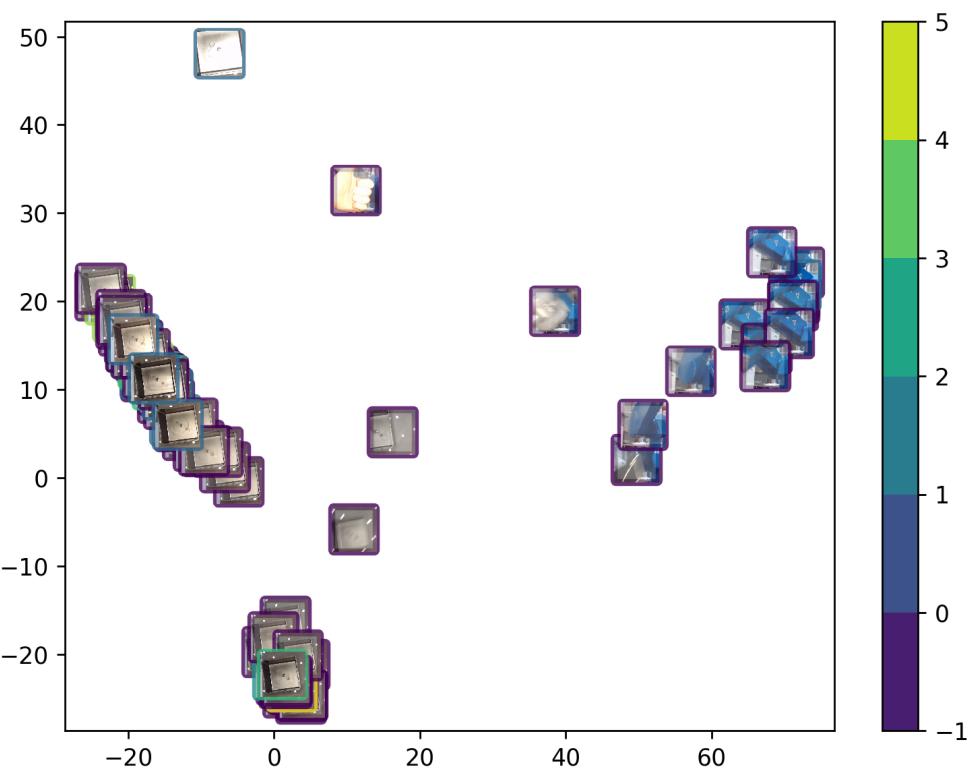


Figure 3.17 – Projection par ACP à partir de la valeur des pixels des images. L'échelle 1-5 correspond aux niveaux de qualité évalués par les experts humains; -1 correspond aux pièces non annotées.

### 3.3.4.4 ACP à noyau

L'ACP à noyau (*Kernel-PCA*) réalise une ACP dans l'espace transformé par un noyau  $K$  (voir la méthode de construction d'un noyau, §3.3.1.3). Elle a été proposée par SCHÖLKOPF, SMOLA et MÜLLER [SSM97; SSM98]. La méthode de la transformée par noyau utilise une transformation  $\varphi$  de l'espace initial des données, vers un espace où les données sont linéairement séparables. On applique alors la méthode de l'ACP dans ce nouvel espace pour trouver les  $k$  vecteurs propres  $\mathbf{V}$ , Équation 3.35.

$$\begin{aligned}\bar{C}_{ker} &= \text{cov}(\varphi(X)) = \frac{1}{N} \sum_{n=1}^N \varphi(\mathbf{x}_n - \bar{\mathbf{x}}) \cdot \varphi(\mathbf{x}_n - \bar{\mathbf{x}})^T \\ \lambda(\varphi(X) \cdot \mathbf{V}) &= (\varphi(X) \cdot \bar{C}_{ker} \mathbf{V})\end{aligned}\quad (3.35)$$

L'optimisation des hyper-paramètres d'une ACP à noyau (choix du noyau et de ses paramètres) peut être réalisée par recherche exhaustive d'hyper-paramètres et validation croisée. On cherche à obtenir la transformée qui limite au maximum la corruption de l'information du jeu de données initial. Ainsi, on cherche l'ACP qui, suite à une projection des données dans l'espace des vecteurs propres, puis à la transformée inverse dans l'espace initial, conserve les données. La métrique utilisée est la norme Euclidienne, entre les données initiales et les données transformées puis transformées inverses, par l'ACP à noyau.

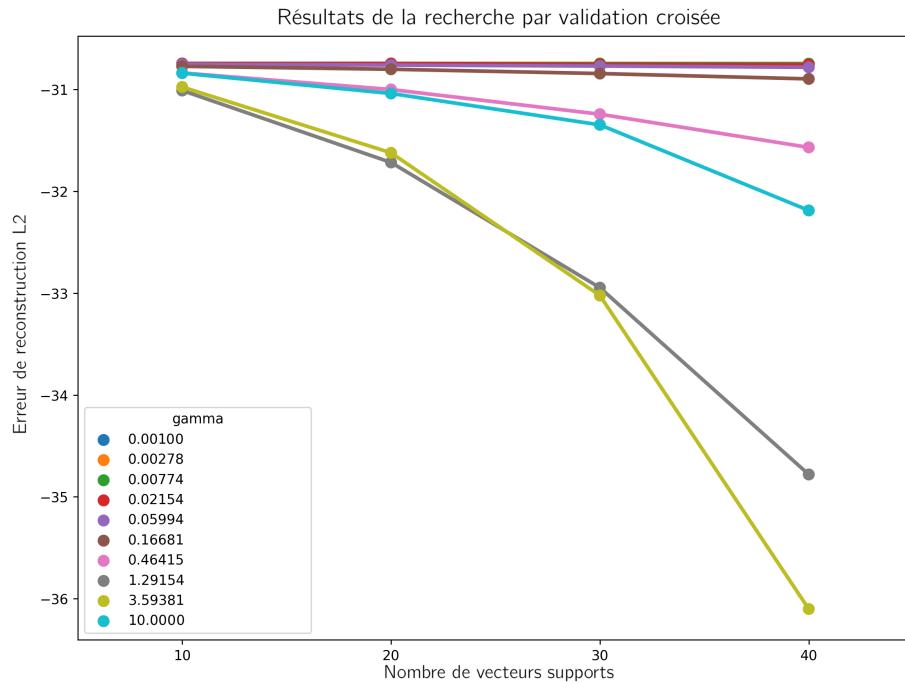


Figure 3.18 – Optimisation des paramètres d'une ACP à noyaux. Métrique : Norme  $\ell_2$ .

La Figure 3.18 présente le résultat d'une optimisation sur  $k$  : le nombre de paramètres et  $\gamma$  : coefficient d'un noyau *RBF*. On choisira ici  $k = 20$  et  $\gamma = 0,02154$ .

Nous obtenons des performances inférieures à l'ACP, Figure 3.19. La transformation par un noyau gaussien est ici inutile. De manière générale, le traitement d'images aux textures complexes nécessite l'utilisation de modèles fortement non linéaires, ce qui n'est pas le cas du

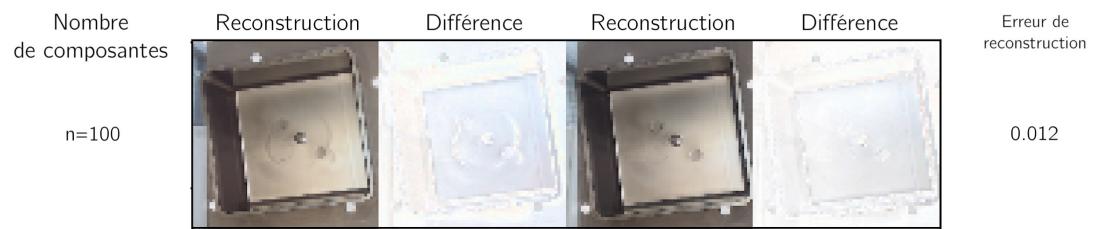


Figure 3.19 – Compression par ACP à noyau.

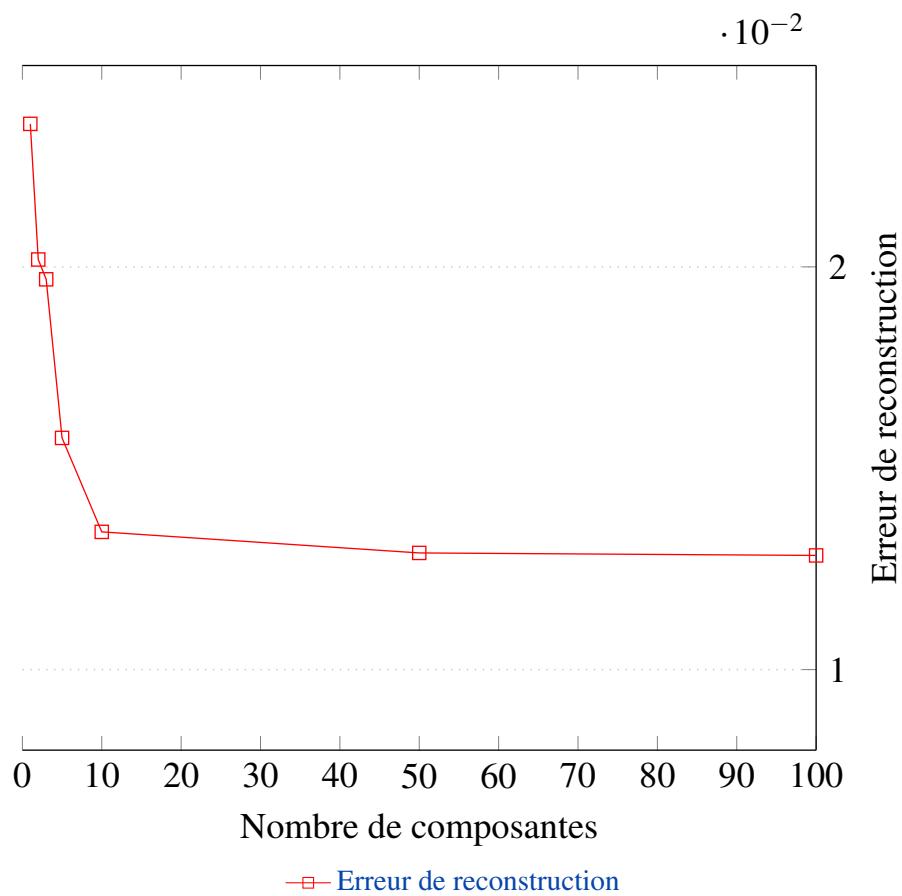


Figure 3.20 – Qualité de la compression par ACP à noyau en fonction du nombre de composantes.

noyau gaussien. C'est pourquoi les méthodes par réseaux de neurones non linéaire obtiennent de meilleurs résultats. Nous les étudierons dès la section suivante, §3.3.4.7.

La complexité de l'ACP à noyau est  $O(m^2 \cdot n^2 + m^3)$  pour un noyau linéaire et  $O(m^2 \cdot n^3 + m^3)$  pour un noyau gaussien (*RBF*). Dans le cas  $m \ll n$ , les complexités sont quadratiques  $O(n^2)$  pour un noyau linéaire et cubique  $O(n^3)$  pour un noyau gaussien. Lorsqu'on utilise l'ACP à noyau, on limitera le nombre de variables  $m$  des échantillons. De plus, la difficulté apparait dans le cas où le jeu de données est grand ( $n > 1000$ ) : le coût du stockage du noyau  $K$  en mémoire devient significatif.

### 3.3.4.5 t-SNE

La méthode *t-SNE* (*t-Distributed Stochastic Neighbor Embedding*) est proposée par VAN DER MAATEN et HINTON [vH08]. C'est une méthode de réduction de l'information itérative particulièrement adaptée à la visualisation des données dans un espace à deux ou trois dimensions. *t-SNE* est une évolution de la méthode *SNE* (*Stochastic Neighbor Embedding*) [HR03]. Les deux méthodes utilisent la même démarche. Une distribution de probabilité  $P$  est générée afin de modéliser la distribution des distances entre les points, pour le jeu de données initial. Le nombre de points voisins à prendre en compte est défini par l'hyper-paramètre de *perplexité*. Une petite valeur de perplexité permet de mettre en évidence les structures locales, alors qu'une grande valeur mettra en évidence les structures globales. La valeur de perplexité définira la variance de la distribution. En pratique, on choisira une valeur dans  $[5; 50]$ . Une distribution  $Q$  est générée afin de modéliser les mêmes propriétés des données, mais dans un espace de dimensions réduites. On cherche alors à minimiser l'écart entre les deux distributions  $P$  et  $Q$  afin de trouver la distribution  $Q$  optimale. Cette démarche permet de modéliser la structure non-linéaire des données. Les paramètres de  $Q$  sont ajustés par descente de gradient stochastique, afin de minimiser la divergence de Kullback-Leibler entre les deux distributions, Équation 3.18. *SNE* utilise une distribution de probabilité gaussienne, alors que *t-SNE*, une distribution qui suit une loi de Student (*t-distribution*). La loi de Student est à queue-longue. Elle tend plus lentement vers zéro qu'une distribution gaussienne. Aussi, elle est plus adaptée à la modélisation des distances entre de petits nombres d'échantillons. Enfin, *t-SNE* propose de simplifier les calculs de l'optimisation de  $Q$  en utilisant une loi de Student à un degré de liberté, qui est une loi de Cauchy,  $f(t) = \frac{1}{\pi(1+t^2)}$ .

Trois conséquences de la méthode limitent son utilisation à la seule visualisation de données. La réduction de dimensions obtenue n'est pas une fonction. Aussi, de nouveaux échantillons ne pourront pas être projetées dans l'espace réduit. L'algorithme est stochastique et chaque exécution produit un résultat différent. Enfin, les résultats dépendent fortement de la valeur de l'hyper-paramètre *perplexité*.

La Figure 3.21 présente la projection obtenue par *t-SNE* sur la valeur des pixels bruts de l'image de chaque pièce. On observe une séparation en quatre paquets, dont un paquet qui contient les images "erronées" (qui ne contiennent pas de pièce). À la différence de la projection par ACP, Figure 3.17, trois paquets de différente qualité de pièces sont mis en évidences. L'algorithme *t-SNE* propose une réduction de dimensions plus pertinente que l'ACP, pour les problèmes où des motifs complexes doivent être distingués. Cependant, appliquer cet algorithme sur les valeurs des pixels bruts limite les résultats à la séparation de motifs globaux. Il est nécessaire de réaliser un prétraitement des images pour extraire les motifs pertinents sur notre problème. Une approche par apprentissage supervisé et réseau de neurones de convolution est la plus pertinente. L'apprentissage par transfert de domaine (§3.3.2) peut également être utilisé, comme méthode d'extraction des motifs pertinents.

Une dernière limite de cette méthode est posée par le calcul de la distance Euclidienne entre les points que l'on cherche à modéliser dans les distributions  $P$  et  $Q$ . La distance Euclidienne nécessite de vérifier l'hypothèse de linéarité de l'espace entre les points. C'est une hypothèse forte qui est rarement vérifiée pour les données complexes, notamment lorsque la répartition des

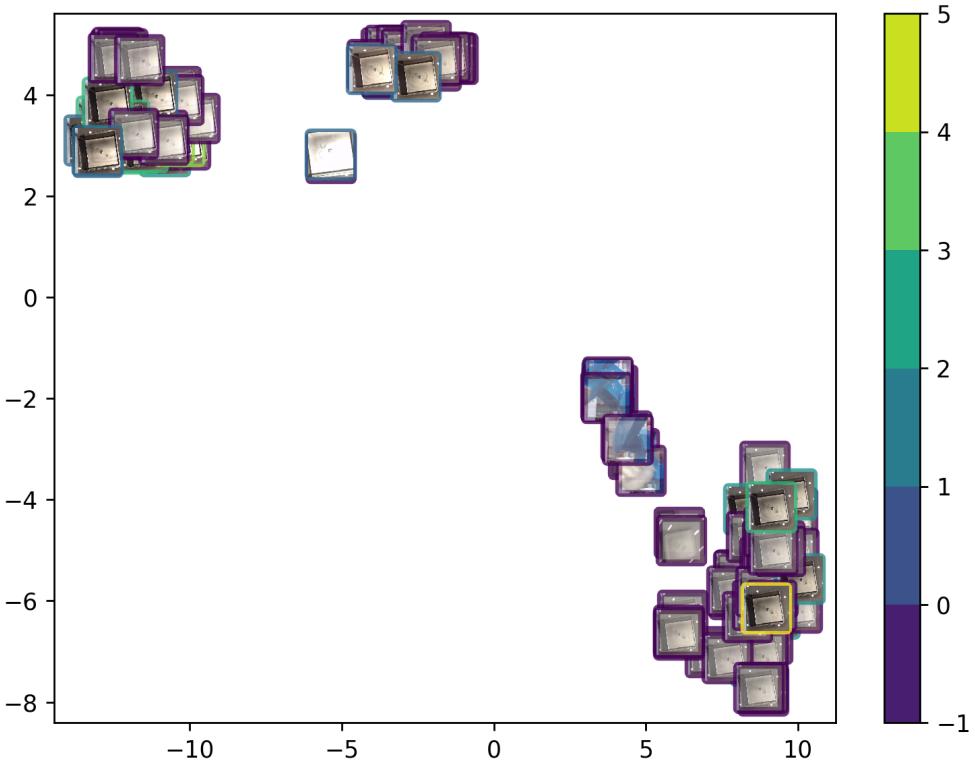


Figure 3.21 – Projection par t-SNE à partir de la valeur des pixels des images. L'échelle 1-5 correspond aux niveaux de qualité évalués par les experts humains ; -1 correspond aux pièces non annotées.

points n'est pas uniforme. C'est pourquoi de nouvelles méthodes de réduction de l'information ont été proposées. Nous nous intéresserons en particulier à la méthode *UMAP* §3.3.4.6, qui utilise la géométrie riemannienne pour résoudre ces limites.

La complexité de l'algorithme t-SNE est de l'ordre de  $O(dN^2)$ , avec  $d$  la dimension de l'espace réduit. [Maa14] propose d'accélérer l'algorithme à une complexité de  $O(dN \log N)$ , dans le cas où  $d \leq 3$  en approximant le gradient par l'algorithme à arbres binaires de recherche de Barnes et Hut [BH86].

### 3.3.4.6 UMAP

La méthode de réduction de l'information *UMAP* est proposée par McINNES, HEALY et MELVILLE [MHM18 ; McI+18]. Elle s'appuie sur l'apprentissage de variété (*manifold learning*). On cherche, dans l'espace de dimensions initiales, une variété (des surfaces topologiques), pour laquelle les données sont au plus proche de celle-ci.

La variété obtenue est alors une représentation des données dans un espace de dimensions réduites. Alors que l'ACP réalise une projection dans un espace euclidien, *UMAP* projette les données dans un espace riemannien. L'espace riemannien possède les hypothèses de connexité et de complétude, ce qui limite la perte d'informations. L'algorithme *UMAP* cherche une variété  $\mathbb{V}$  dans cet espace, telle que les données soient situées au plus proches de celle-ci. Afin d'utiliser cette démarche, trois hypothèses sont faites :

- La distribution des données est uniforme sur la variété  $\mathbb{V}$  utilisée.
- La métrique riemannienne peut être localement approximée comme une constante.
- L'espace riemannien utilisé est connexe.

Le théorème du nerf [Zis72] affirme qu'il est possible de représenter la variété topologique  $\mathbb{V}$  par un n-simplexe<sup>5</sup>. Dans ce cas,  $\mathbb{V}$  et sa représentation en simplexes sont équivalentes. Il est alors possible de modéliser ces simplexes par une distribution combinatoire  $P$ , en appliquant la logique floue (topologie-floue). Nous invitons le lecteur qui souhaite approfondir ce raisonnement à consulter la publication originale des auteurs de l'algorithme UMAP [MHM18]. Par la suite, représenter les données dans un espace de petites dimensions, revient à trouver la variété de petites dimensions qui possède la distribution  $Q$  la plus similaire possible, avec la distribution  $P$  de la variété des données initiales. Cela revient à trouver la distribution  $Q$  telle que l'entropie croisée entre les deux distributions  $P$  et  $Q$  soit minimale. Cette démarche est similaire à la minimisation de la divergence de Kullback-Leibler de t-SNE §3.3.4.5 qui met en évidence la structure locale des données, mais elle permet également de préserver la structure globale.

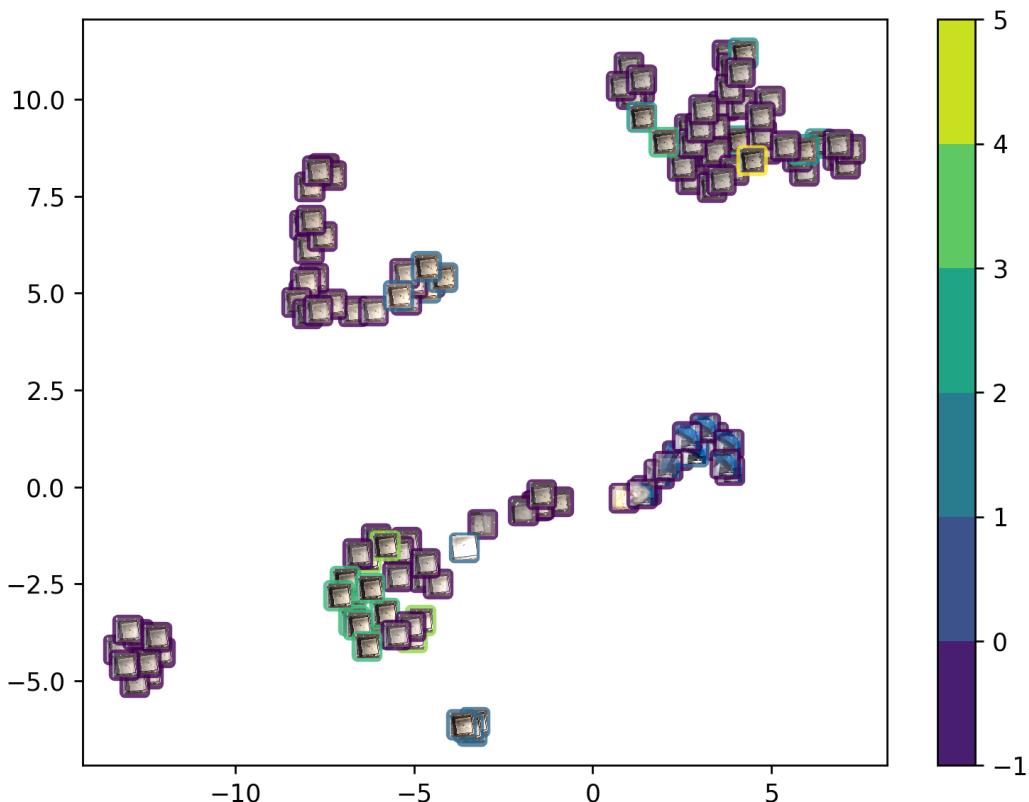


Figure 3.22 – Représentation par UMAP à partir de la valeur des pixels des images. L'échelle 1-5 correspond aux niveaux de qualité évalués par les experts humains ; -1 correspond aux pièces non annotées.

La Figure 3.22 présente la représentation en deux dimensions obtenues en appliquant UMAP sur la valeur des pixels brutes de l'image. En comparaison de t-SNE, UMAP permet de mettre en évidence la structure globale des données dans les deux dimensions, tout en conservant la structure locale des paquets d'échantillons. Les images qui ne contiennent pas de pièces sont séparées, ainsi que le paquet de pièces annotées d'un niveau de qualité 0 (mauvais). De plus, le niveau de qualité 1 est séparé. Les niveaux 2 à 5 sont représentés dans deux paquets distincts. Il apparaît un paquet de pièces de très bonne qualité (niveau 5). Ce paquet contient quelques pièces des niveaux 3 et 4; aussi cela indique que l'annotation humaine n'est pas robuste lorsque l'on approche de la qualité maximale. En comparaison des résultats obtenus avec un réseau de triplets, Figure 3.13,

5. Un simplexe est l'élément topologique le plus simple possible.

cette méthode présente une séparation des paquets identiques, mais sans annotation humaine et avec un coût de calcul plus de cent fois moindre. L'approche topologique permet d'améliorer les performances et la méthode *UMAP* montre son intérêt dans le cadre d'une approche non-supervisée de suggestion d'annotation à l'expert humain.

Afin de trouver le graphe de la représentation en petites dimensions, *UMAP* utilise l'algorithme (*Nearest-Neighbor-Descent*) d'optimisation de graphes proposé par DONG, MOSES et LI [DML11]. Les auteurs évaluent empiriquement la complexité de leur méthode de l'ordre de  $O(N^{1.14})$ . Aussi, la complexité de la méthode *UMAP* est limitée par cet algorithme. Elle est de l'ordre de  $O(N \log N)$ .

### 3.3.4.7 Auto-encodeurs variationnels

Un auto-encodeur cherche à reproduire la donnée entrée, avec le minimum de paramètres (ou Degrés De Liberté). Aucune annotation des données n'est requise. C'est une méthode qui permet d'apprendre la représentation des données dans un espace de dimensions réduites (aussi appelé *feature learning* ou *representation learning*). En 1987, LE CUN, MILGRAM et UNIVERSITÉ PIERRE ET MARIE CURIE (PARIS); GALLINARI et al. proposent un réseau auto-encodeur utilisant des réseaux de neurones, afin de supprimer le bruit d'images [LMU87; Gal+87]. Dans leurs travaux, ils étudient la capacité du réseau à mémoriser les caractéristiques des images initiales, en fonction de son architecture. En parallèle, BALLARD propose un auto-encodeur (appelé dans ce texte *systèmes auto-associatifs*) à trois couches [Bal87]. Dans ces travaux, les poids des neurones sont ajustés par rétro-propagation du gradient, qui a été récemment proposée.

Un auto-encodeur se compose de deux parties : un encodeur  $z = f(x)$  qui compresse l'information initiale dans un espace de dimensions réduites (appelé *espace latent*) et un décodeur  $\hat{x} = g(z)$  qui reconstruit ensuite une approximation de  $x$  à partir de  $z$ .  $f$  et  $g$  sont généralement des réseaux de convolutions profonds, dont les poids respectifs sont  $\mathbf{W}_f$  et  $\mathbf{W}_g$ . La Figure 3.23 présente l'architecture d'un auto-encodeur qui utilise des réseaux de convolutions pour ces fonctions.

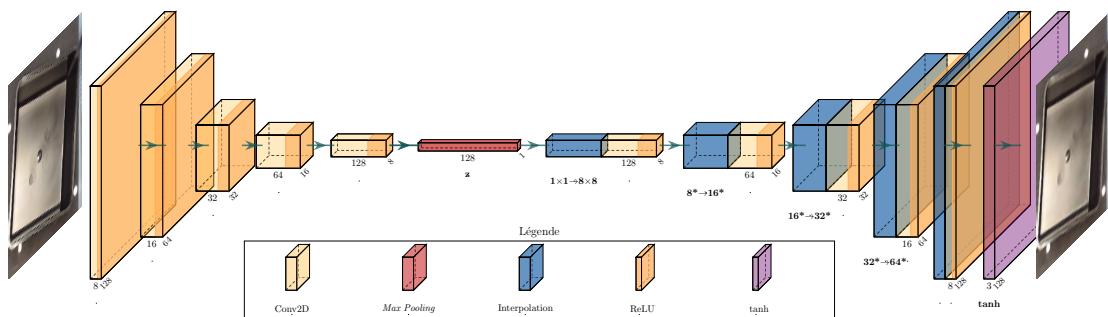


Figure 3.23 – Architecture d'un auto-encodeur.

Afin de mesurer la similitude telle que  $x \simeq \hat{x}$ , la fonction de coût à minimiser lors de l'apprentissage du modèle est souvent l'erreur quadratique entre  $x$  et  $\hat{x}$ , Équation 3.36.

$$\min_{\mathbf{W}_f, \mathbf{W}_g} \mathcal{L}_{\text{génération}}(x, \hat{x}) = \min_{\mathbf{W}_f, \mathbf{W}_g} \|x - g(f(x, \mathbf{W}_f), \mathbf{W}_g)\|_2^2 \quad (3.36)$$

Enfin, une contrainte indispensable est appliquée à la dimension de  $z$ , qui doit être très petite devant la dimension de  $x$ . Ainsi, seules les propriétés les plus importantes des données sont prises en compte. Un cas particulier apparaît si le décodeur  $g$  est une application linéaire et si la fonction de coût est l'erreur quadratique : alors les valeurs de  $z$  sont équivalentes aux valeurs propres d'une Analyse en Composante Principale [BK88].

La Figure 3.24 présente une image entrée et sa reconstruction après le passage dans un auto-encodeur dont  $z$  a une dimension de 128 valeurs. On observe la perte d'information visuelle :

la marque de l'éjecteur a, par exemple, disparu. En revanche, les défauts d'aspect qui nous intéressent sont reproduits. Le taux de compression est ici de 120 :  $124 \times 124$  pixels  $\rightarrow$  128 valeurs. À la différence de la compression par ACP, Figure 3.15, la reconstruction ne possède pas d'artefacts. En revanche, le coût de calcul de cette méthode est beaucoup plus grand que pour une ACP. Il est nécessaire de disposer d'une accélération des calculs massivement parallèle, par exemple sur processeur graphique, pour pouvoir construire un modèle à partir d'un grand nombre d'échantillons.

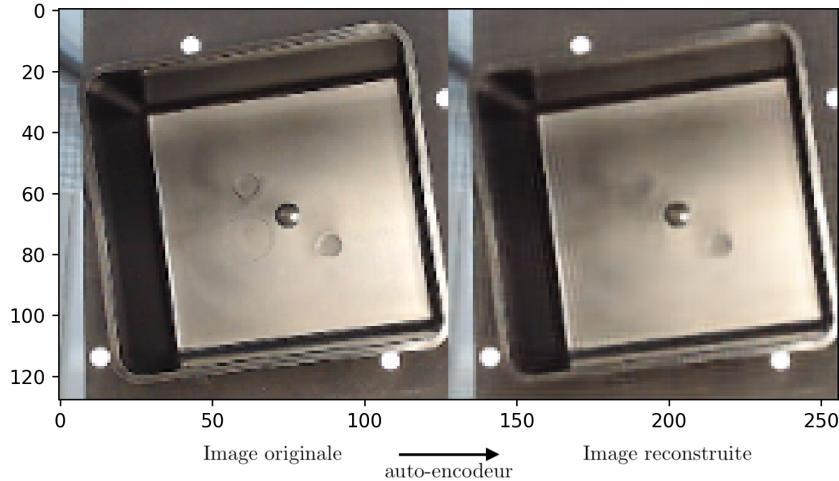


Figure 3.24 – Reconstruction d'une image par un auto-encodeur.

La Figure 3.25 présente les valeurs des 128 dimensions de  $z$  projetées par l'algorithme  $t$ -SNE, présenté dans la Section 3.3.4.5. Les différentes classes exprimées dans l'espace latent  $z$  sont mieux séparées que dans la Figure 3.21 qui utilisait les données brutes : l'ensemble des valeurs des pixels. On observe également que l'évaluation des cinq classes de qualité, évaluées par les experts humains, est respectée. Les pièces non annotées sont positionnées dans les bonnes classes. Enfin, il est intéressant d'observer qu'un paquet correspondant à une erreur de mesure, comme par exemple l'absence de pièces et la capture du bras robotique préhenseur est mis en évidence.

Un auto-encodeur modélise une application (encodeur  $x \rightarrow z$ ) et l'application inverse (générateur  $z \rightsquigarrow \hat{x}$ ). Cependant, il est difficile d'interpoler les valeurs de  $z$  afin de générer des  $\hat{x}$ , car l'espace  $z$  n'a pas de contrainte de continuité ; aussi  $z$  est généralement discontinu. Récemment, l'auto-encodeur variationnel (*Variational Auto-Encoder*) est proposé par KINGMA et WELLING ; REZENDE, MOHAMED et WIERSTRA [KW13; RMW14]. Il utilise une méthode d'inférence variationnelle. À la différence d'un auto-encodeur, un VAE modélise le générateur  $g$  sous la forme d'une distribution de probabilité  $p(x|z)$  sur les données. De même, l'encodeur  $f$  sera une distribution  $q(z|x)$ . Un a priori fort sur la distribution  $q$  est introduit : on cherche généralement  $q(z|x)$  qui soit proche d'une gaussienne centrée  $\mathcal{N}(\mu = 0, \sigma = 1)$ . Ainsi,  $q$  sera continue. Cette condition revient à minimiser la divergence de Kullback-Leibler entre la distribution de l'encodeur  $q(z|x)$  et la distribution du générateur  $p(z|x)$ . L'auto-encodeur variationnel nécessite une fonction de coût composite, Équation 3.37.

$$\begin{aligned} \mathcal{L}_{VAE} &= \mathcal{L}_{\text{génération}} + \mathcal{L}_{\text{latent}} \\ \text{avec } \mathcal{L}_{\text{latent}}(\hat{y}, y) &= \text{Divergence}_{KL}(p(z|x) \| N(\mu = 0; \sigma = 1)) \end{aligned} \quad (3.37)$$

Lors de l'apprentissage, les distributions  $p$  et  $q$  sont modélisées par des réseaux de neurones. Les sorties du générateur seront les paramètres de la distribution : la moyenne et l'écart-type, tels que  $q(z|x) = \mathcal{N}(z; \vec{\mu}(x), \text{diag}(\vec{\sigma}(x))^2)$ . La Figure 3.26, inspirée de [Doe16], représente le graphe de calcul de ce modèle. Il devient possible de générer de nouvelles sorties à partir de la distribution

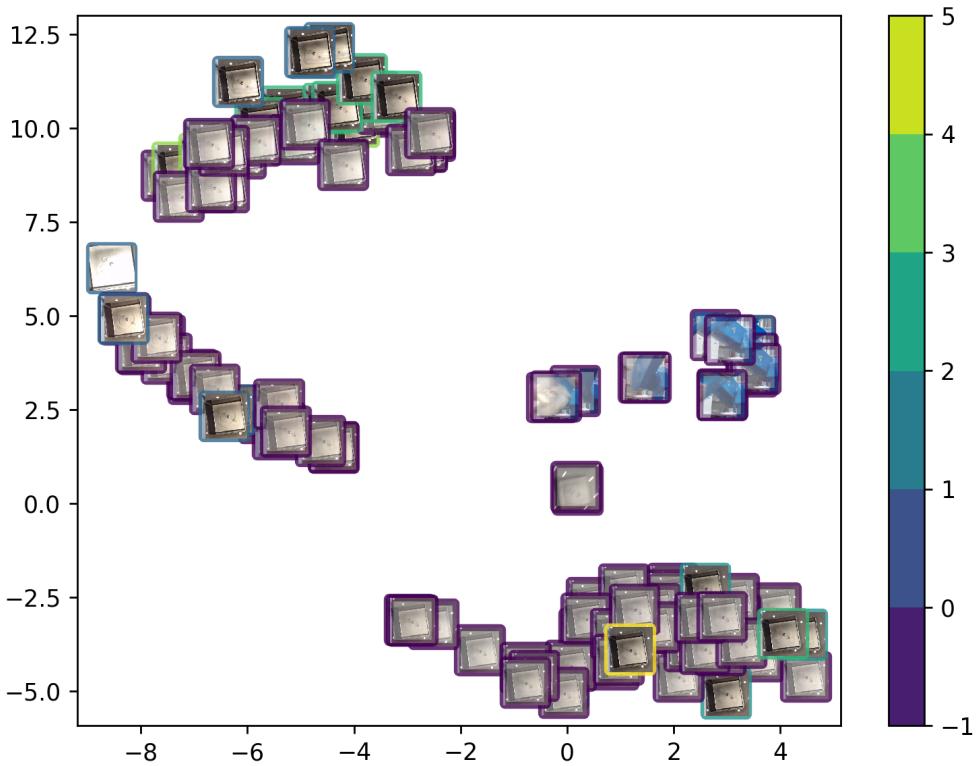


Figure 3.25 – Projection par t-SNE à partir des valeurs de l'espace latent  $z$  d'un VAE. L'échelle 1-5 correspond aux niveaux de qualité évalués par les experts humains; -1 correspond aux pièces non annotées.

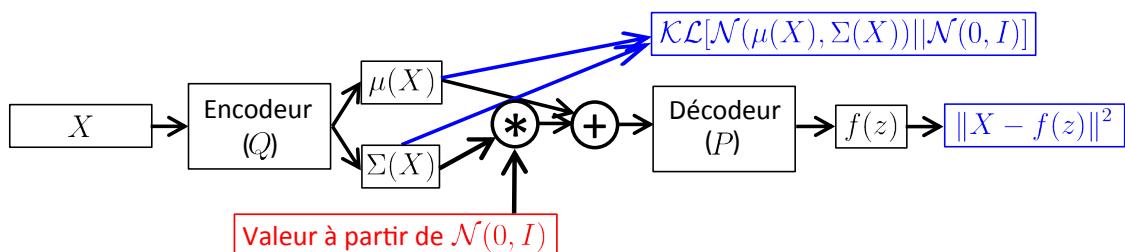


Figure 3.26 – Graphe d'un auto-encodeur variationnel, inspiré de [Doe16].

$p(z)$ , en tirant les valeurs de  $z$  à partir de la distribution  $\mathcal{N}(\mu = 0, \sigma = 1)$ . Ces nouvelles sorties ne sont pas nécessairement similaires aux données d'apprentissages de  $x$ . Ainsi, l'auto-encodeur variationnel généralise l'auto-encodeur et le rend *génératif*.

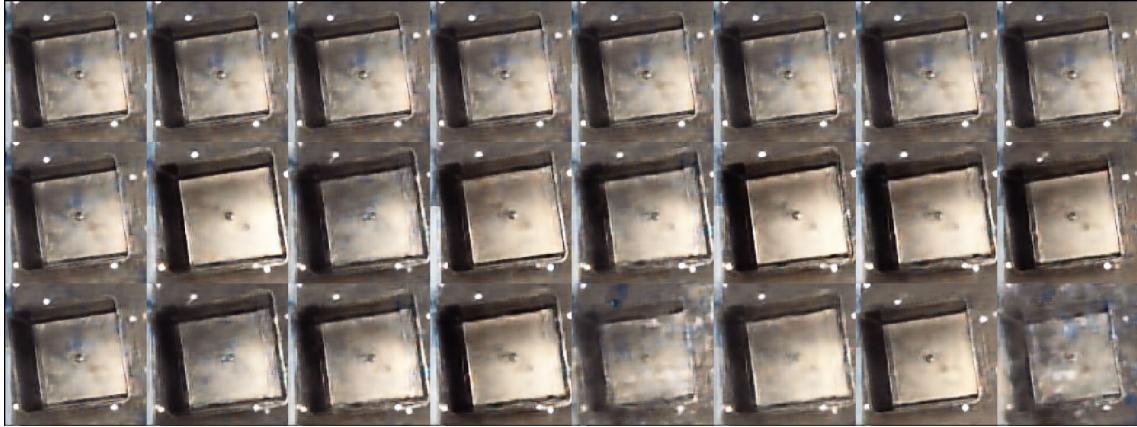


Figure 3.27 – Génération d'images à partir de la distribution modélisant l'espace latent d'un auto-encodeur variationnel.

La Figure 3.27 montre des images générées à partir de tirages aléatoires sur la distribution  $P$  de l'espace latent  $z$ . On observe la diversité des défauts modélisés dans  $z$ , mais également le réalisme limité que produit le générateur d'un VAE. Cela est dû à la compression élevée et au petit jeu de données dont nous disposons ce qui entraîne du bruit. Il est intéressant d'observer des artefacts de couleur bleue, qui correspondent à la présence d'images dans le jeu de données où il n'y a pas de pièce, mais où le bras robotique de couleur bleue est présent. Le bras robotique est néanmoins très peu présent dans  $z$  en comparaison des pièces : le VAE est robuste.

Les réseaux génératifs permettent de palier de manière intéressante au manque de variétés des échantillons dans les petits jeux de données. Cependant, les images générées sont souvent floues, ce qui est souvent dû à la fonction de coût, l'erreur moyenne entre  $x$  et  $\hat{x}$ , utilisée lors de l'apprentissage. En effet, l'erreur moyenne ne permet pas la prise en compte dans  $z$  des motifs complexes qui sont peu présents dans le jeu de données. Dernièrement, KINGMA et WELLING [KW19] propose une synthèse exhaustive des solutions à cette problématique : construire des modèles plus profonds, ajouter des contraintes plus pertinentes dans la fonction de coût, comme par exemple des tâches de classification auxiliaires, et mettre à jour de manière plus judicieuse les poids lors de l'apprentissage du modèle. Un autre type de modèle génératif propose également d'améliorer le réalisme des images générées. L'apprentissage n'est pas réalisé sur l'erreur entre  $x$  et  $\hat{x}$ , mais de manière indirecte à partir de la valeur retournée par un second modèle. Ce principe permet d'améliorer le réalisme en prenant en compte des motifs complexes lors de la reconstruction. Il s'agit des modèles antagonistes génératifs que nous détaillons dans la Section 3.3.4.8.

L'apprentissage d'un modèle VAE est réalisé par descente de gradient stochastique. Les poids du modèle sont ajustés sur un sous-ensemble du jeu de données de taille  $B$  échantillons (appelé mini-batch), pendant  $T$  itérations. La complexité de l'apprentissage est de l'ordre de  $O(TB^2D + TB^2d)$ , avec  $T$  le nombre d'itérations,  $B$  le nombre d'échantillons dans un mini-batch,  $D$  la dimension d'un échantillon et  $d$  la dimension de l'espace latent. De plus, la quantité de mémoire de travail nécessaire au stockage des mini-batch est de l'ordre de  $O(B^2D)$ . Ces coûts de calcul sont importants. Dans nos expériences, c'est la quantité de mémoire de travail sur le processeur graphique dont nous disposons (11 Giga octets) qui conditionne le nombre d'échantillons par mini-batch (ici,  $B = 64$  et  $D = 128^2$ ), d'où la complexité des calculs ( $T = 100$ ,  $B = 64^2$ ,  $D = 128^2$ ). Nous obtenons une complexité de l'ordre de 7 Milliards d'opérations et une occupation mémoire de l'ordre de 10 Giga octets. La durée de l'apprentissage de notre

modèle VAE est d'une heure sur un processeur graphique dédié (*NVIDIA GTX 1080 Ti*). Avec les ressources informatiques actuelles, une mise à jour journalière de ce modèle peut être envisagée, lors de son utilisation au sein d'un procédé industriel.

### 3.3.4.8 Réseaux antagonistes génératifs

Proposés par GOODFELLOW et al. [Goo+14], les réseaux antagonistes génératifs (*Generative Adversarial Networks*) permettent de raffiner le modèle de l'espace latent. En comparaison avec les auto-encodeurs, les images issues du générateur sont plus proches des images originales. Les *GAN* s'appuient sur une association de deux modèles qui sont entraînés en concurrence : un générateur  $g$  génère  $\hat{x} = g(z)$  et un classifieur  $d$  (appelé *discriminateur*) calcule la probabilité  $p$  que  $\hat{x}$  appartienne au jeu de données réel,  $p = d(x)$ . L'objectif du générateur est de générer  $\hat{x}$  tel que  $\hat{x} \simeq x$ , tandis que l'objectif du discriminateur est de distinguer les  $\hat{x}$  générés par le générateur, des  $x$  réels, issues du jeu de données initial. La phase d'apprentissage de ce modèle est similaire à un duel de la théorie du jeu où le générateur et son adversaire le discriminateur s'affrontent. Les deux modèles sont entraînés en parallèle afin d'apprendre les poids  $w_g$  et  $w_d$  qui maximiseront leurs performances. Un équilibre de Nash [Nas50; Nas51] est théoriquement atteint lorsque le générateur produit des  $\hat{x}$  qui ne peuvent plus être distingués des  $x$  par le discriminateur. Trouver le générateur  $g$  optimal revient à résoudre le problème d'optimisation de l'Équation 3.38 pour toute donnée d'entrée  $x$ . Cependant, comme le problème d'optimisation cherche à minimiser un terme et minimiser un autre terme, la solution optimale est difficile à obtenir lors de l'ajustement des poids par descente de gradients stochastique car de nombreux optimums locaux (optimum pour un seul des membres) existent [Goo+14].

$$g_{optimal} = \min_g \max_d = \mathbb{E}_{\mathbf{x} \sim p_{\text{réelle}}} \log d(\mathbf{x}) + \mathbb{E}_{\mathbf{x} \sim p_{\text{générée}}} \log(1 - d(\mathbf{x})) \quad (3.38)$$

Le modèle *GAN* fait partie des nombreux modèles proposés depuis les années 1990, qui s'appuient sur un apprentissage non-supervisé par affrontement de deux modèles qui s'opposent dans un jeu de minimisation-maximisation. SCHMIDHUBER propose une revue des travaux fondateurs sur le sujet [Sch19]. Le dynamisme actuel de la littérature sur les modèles de type *GAN*, ainsi que la disponibilité des implémentations ouvertes par leurs auteurs, nous a permis d'évaluer ces modèles sur notre problématique. De nombreux modèles utilisant des idées proches restent à évaluer. Ils pourraient permettre d'obtenir de meilleures performances sur notre problématique d'analyse de la qualité.

À la suite de la proposition originale des *GAN*, RADFORD, METZ et CHINTALA [RMC15] proposent des règles pour adapter les *GAN* à la reconstruction d'images : remplacer les couches de moyenne (*pooling*, §3.3.1.7) par des convolutions de noyaux  $1 \times 1$ , utiliser la normalisation de lot 3.3.1.7, ne pas utiliser de couches pleinement connectées, utiliser l'activation *ReLU* 3.3.1.6 pour le générateur sauf pour la dernière couche *tanh*, utiliser *LeakyReLU* pour le discriminateur et enfin des règles de construction de l'architecture du réseau qu'ils nomment *DCGAN*. Ce travail évalue ces règles avec comme objectif d'encoder des jeux de données d'images complexes dans un espace latent de dimensions 100. Enfin, le travail montre l'intérêt des *GAN* pour réaliser une classification d'images de manière non-supervisée (pas d'annotation des données d'apprentissage). Il s'agit d'utiliser le générateur comme fonction d'extraction des valeurs pertinentes d'une image (espace latent), puis ensuite d'utiliser un classifieur *SVM* 3.3.1.3 sur ces dimensions de l'espace latent. Les performances sont légèrement inférieures à celles obtenues par une approche supervisée, ce qui montre l'intérêt de la méthode puisqu'il n'a pas été nécessaire d'annoter les données pour obtenir ce résultat. Le *GAN*, puisqu'il est entraîné pour reconstruire une image à partir des dimensions réduites de l'espace latent, permet également d'extraire l'information pertinente des images. Cette approche est particulièrement intéressante et nous chercherons à l'utiliser afin de répondre aux problèmes industriels où les jeux de données sont petits.

Par la suite, SALIMANS et al. [Sal+16] complètent les règles du modèle *DCGAN*. ODENA ; SALIMANS et al. [Ode16; Sal+16] proposent en parallèle de réaliser l'apprentissage de manière semi-supervisée. L'objectif est de créer un discriminateur qui soit également capable de répondre au problème de classification d'un jeu de données annotées. Il s'agit d'introduire dans le jeu de données annotées une nouvelle classe "image générée". Pour un jeu de données comportant  $K$  classes, on obtient un jeu de données de  $K + 1$  classes. On utilise alors le générateur pour produire les images correspondant à cette nouvelle classe  $K + 1$ . Cette démarche est alors introduite dans la fonction de coût du discriminateur, Équation 3.39. Ainsi, on obtient un discriminateur qui répond au problème de classification supervisée, tout en étant capable de distinguer les images générées par le générateur. L'intérêt est double : augmentation artificielle du jeu de données et robustesse aux échantillons adversaires (les "fausses" images générées).

$$g_{optimal} = \min_g \max_d = \mathbb{E}_{\mathbf{x} \sim p_{\text{réelle}}} \log d(\mathbf{x}) + \mathbb{E}_{\mathbf{x} \sim p_{\text{générée}}} \log(1 - d(\mathbf{x})) \quad (3.39)$$

Par la suite, des améliorations importantes sont apportées à la méthode. ARJOVSKY, CHINTALA et BOTTOU [ACB17] proposent de modifier la fonction de coût du discriminateur qui évalue la qualité de la reconstruction en utilisant la distance de Wasserstein. [Gul+17], ainsi que la normalisation spectrale [Miy+18].

L'ensemble de la littérature évalue les performances de ces modèles sur la base de données *ImageNet*. En comparaison, dans notre cas d'application, le nombre de classes est très faible, de deux classes à une dizaine de classes : pour représenter une échelle qualité bonne — qualité mauvaise, ou pour représenter différentes typologies de défauts. C'est pourquoi un réseau génératif simple est capable de réaliser une synthèse fidèle des nos images. En revanche, sur une base de données qui comporte une grande variété de classes, comme *ImageNet*, il apparaît que les modèles génératifs ne permettent pas de modéliser les associations entre des éléments géométriques qui doivent nécessairement être présents dans les images d'une classe. Un exemple est celui de la représentation d'un chien, auquel la texture de pelage est bien appliquée par le modèle *GAN*, mais qui ne possède pas quatre pattes attachées au corps. Cette limite provient de l'opérateur de convolution qui est une opération locale spatialement contrainte. Par exemple, l'opérateur de convolution ne peut pas associer la valeur du pixel située en haut à gauche, avec celle en bas à droite. De plus, lors de l'ajustement des poids par rétro-propagation du gradient, il n'y a pas de relation spatiale entre la valeur en sortie de convolution et les valeurs en entrées. Cette problématique est critique dans le domaine du traitement de textes et la traduction (*Natural Language Processing*) : la sémantique d'une phrase dépend de la position des mots dans la phrase. C'est pourquoi le mécanisme d'*attention* a été proposé par BAHDANAU, CHO et BENGIO [BCB14] pour la traduction de texte, puis étendu à l'analyse d'images par un réseau de convolutions par [Xu+15]. Par la suite, VASWANI et al. [Vas+17] propose l'architecture de couche *transformer*, qui généralise la méthode en éliminant l'usage de réseaux récurrents. Ils montrent qu'un réseau constitué uniquement de couches d'*attention* permet d'atteindre les performances de l'état de l'art en traduction de texte. L'*attention* est depuis utilisée dans toutes les méthodes de l'état de l'art en traduction automatique. Une couche d'*attention* pondère la valeur d'entrée par une distribution de probabilité, dont les paramètres sont appris, et qui dépend de la position spatiale de la valeur entrée. Cette démarche est similaire aux portes d'activations des réseaux *LSTM*, ainsi qu'à l'architecture des *Highway Networks*, §3.3.1.5, §3.21.

Dernièrement, ZHANG et al. [Zha+18] propose le modèle *SAGAN* (*Self-Attention Generative Adversarial Networks*) qui associe l'utilisation de multiples méthodes et introduit l'utilisation du mécanisme d'*attention*. Il s'agit pour la couche d'*attention* de modéliser une relation de dépendance pour un pixel, entre tous les autres pixels. Il n'y a ainsi plus de restriction spatiale aux convolutions.

La Figure 3.28 montre des images générées à partir de tirages aléatoires dans l'espace latent du générateur d'un modèle *SAGAN*. On observe le meilleur réalisme des images obtenues, en

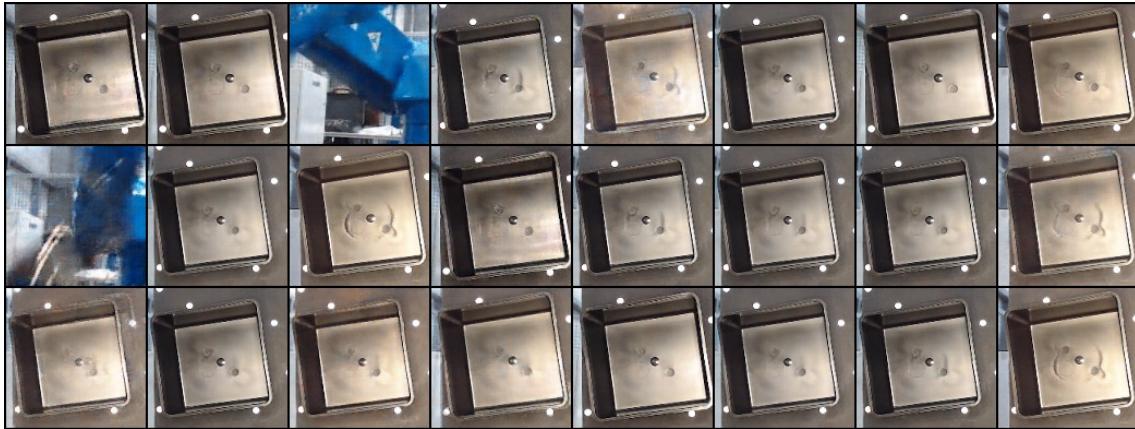


Figure 3.28 – Génération d'images à partir de la distribution modélisant l'espace latent du générateur d'un GAN.

comparaison des résultats donnés par un VAE 3.27. La diversité des défauts est bien représentée. De plus, il est intéressant de remarquer que malgré le très faible nombre d'images représentant un bras robotique dans le jeu de données initial (1 erreur de capture pour 100 images), ces derniers sont représentés avec beaucoup de réalisme dans l'espace latent du générateur. Un modèle GAN est peu robuste. Il est sensible aux échantillons extrêmes d'un jeu de données. Dans notre cas, cela peut être un avantage pour modéliser des défauts qui n'apparaissent pas souvent; mais cela demande également une rigueur dans la construction du jeu de données, pour ne pas introduire d'images erronées, comme ici le bras robotique.

De même que pour le modèle VAE, §3.3.4.7, la complexité de l'apprentissage d'un GAN est très importante. L'apprentissage du modèle est réalisé par descente de gradients stochastiques. Les poids du modèle sont ajustés sur un sous-ensemble du jeu de données de taille  $B$  échantillons (appelé mini-batch), pendant  $T$  itérations. La complexité de l'apprentissage est de l'ordre de  $O(TB^2D + TB^2d)$ , avec  $T$  le nombre d'itérations,  $B$  le nombre d'échantillons dans un mini-batch,  $D$  la dimension d'un échantillon et  $d$  la dimension de l'espace latent. Nous sommes limités par la capacité de mémoire de travail de notre processeur graphique à  $B = 64$  et  $D = 128^2$ , d'où la complexité des calculs ( $T = 5000$ ,  $B = 64^2$ ,  $D = 128^2$ ). Nous obtenons une complexité de l'ordre de 350 Milliard d'opérations. La durée de l'apprentissage de notre modèle GAN est de 10 heures sur un processeur graphique dédié (*NVIDIA GTX 1080 Ti*). Avec les ressources informatiques actuelles, seule une mise à jour hebdomadaire de ce modèle peut être envisagée. Malgré le réalisme de la méthode, il est nécessaire de disposer d'importantes ressources de calculs (au minimum 10 processeurs graphiques comme le notre) pour mettre à jour le modèle à une fréquence journalière.

## 3.4 Optimisation automatique des hyper-paramètres d'un modèle

Les algorithmes présentés dans la section précédente possèdent plusieurs dizaines d'hyper-paramètres. Leur ajustement est souvent crucial pour obtenir des performances satisfaisantes. On parle d'optimisation des hyper-paramètres : on cherche à optimiser la performance du modèle pour la tâche à effectuer. Il est également nécessaire de sélectionner les bonnes méthodes de préparation des données. Le Tableau 3.7 recense, par exemple, les hyper-paramètres des modèles à réseaux de neurones de convolutions profonds. Des méthodes d'optimisation automatique de ces hyper-paramètres ont été proposées. Cette optimisation est souvent appelée *AutoML* (*Automated Machine Learning*). Dans le cas des réseaux de neurones profonds, l'architecture du réseau

est également importante. Elle doit être considérée comme de nombreux hyper-paramètres. Le nombre de paramètres devient alors très grand et les algorithmes d'optimisation traditionnels deviennent limités ; la métadescente de gradients et les algorithmes évolutionnistes deviennent plus adaptés §3.4.7. Cette section propose de parcourir les méthodes les plus avancées et de discuter de leurs applications sur notre problématique de modélisation du contrôle de la qualité.

L'ensemble des expériences d'optimisation de modèles réalisées dans ces travaux ont utilisé la librairie *Sacred* [Gre+17], afin de conserver un historique des hyper-paramètres et des performances associés. Certaines de ces optimisations peuvent prendre plusieurs jours, c'est pourquoi il est nécessaire de sauvegarder l'évolution des performances pour éviter toute perte de ressources temporelles en cas d'erreur. *Sacred* permet également de rendre compte de la progression des optimisations et de surveiller toute erreur.

### 3.4.1 Optimisation par recherche aléatoire

Afin de réaliser l'optimisation, il s'agit d'évaluer la performance de la méthode sur la plage de valeurs de chacun des hyper-paramètres. Il est possible d'évaluer l'ensemble de la plage de valeurs de manière exhaustive, par une grille d'essais régulière, ce qui est coûteux sur le plan du calcul. Une démarche plus économique est le tirage aléatoire sur la plage de valeurs, proposée par BERGSTRA et BENGIO [BB12].

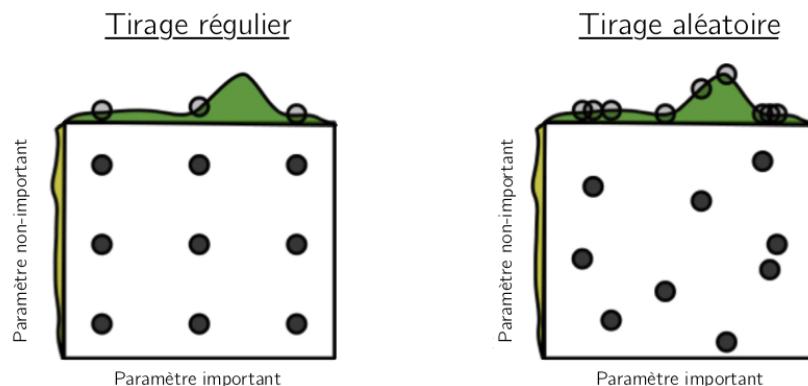


Figure 3.29 – Optimisation d'une fonction à deux variables avec neuf essais. Figure reproduite de [BB12].

La Figure 3.29 présente un tirage de 9 essais pour trouver le maximum d'une fonction à deux dimensions. Dans le cas de la recherche exhaustive, seuls trois essais permettent d'étudier la dimension horizontale. Dans la recherche aléatoire, les 9 essais permettent d'étudier la dimension horizontale, ce qui est plus efficient. Cette limite de la recherche exhaustive est d'autant plus importante que le nombre de dimensions étudiées est grand. Des méthodes plus évoluées qui ont été proposées.

En 2011, BERGSTRA et al. [Ber+11] réalisent un état de l'art des méthodes d'optimisations d'hyper-paramètres, pour les modèles à réseaux de neurones profonds. Ce travail montre l'intérêt de l'optimisation itérative, basée sur le critère de la prédiction de l'amélioration de performance du modèle (*Expected Improvement*, proposée par JONES [Jon01]). L'étude introduit deux méthodes d'optimisation. Une méthode cherche à modéliser le problème d'optimisation par des processus stochastiques gaussiens et la seconde méthode *TPE* propose une modélisation par noyaux, §3.4.3.2. Ces méthodes s'appuient sur la construction de métamodèles. L'étude montre la supériorité de ces deux méthodes sur l'optimisation par tirage aléatoire.

### 3.4.2 Optimisation par plan d'expériences

Afin de sélectionner les essais à réaliser pour modéliser un phénomène, on peut employer les plans d'expériences. Il s'agit de méthodes permettant d'ordonner la réalisation d'essais expérimentaux et de spécifier les modifications à réaliser sur les variables. L'objectif est de limiter le nombre d'essais à réaliser. Il est également possible d'associer la complexité (ou le degré polynomial) du modèle, avec le nombre d'essais pouvant être réalisé. Un petit nombre d'essais permettra de construire un modèle linéaire, tandis qu'un plus grand nombre d'essais pourrait permettre de réaliser un modèle quadratique. En 1935, FISHER [Fis74] propose un ouvrage de référence sur le sujet de la construction et de l'analyse des plans d'expériences. Nous ne discuterons pas des critères d'optimalité des plans que nous présentons. Nous invitons le lecteur à se référer à cet ouvrage pour étudier cette notion importante pour la construction de plans utilisables. La Figure 3.30 représente l'espace de la valeur de trois variables comme un cube. Les valeurs des trois niveaux des variables, obtenues pour différents plans d'expériences, sont représentées par les cercles.

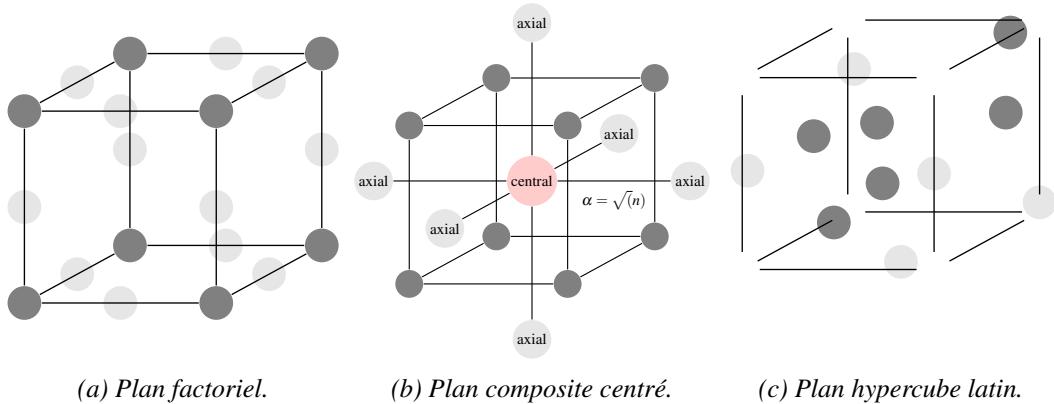


Figure 3.30 – Essais prévus pour des plans d'expériences à trois niveaux pour trois variables.

#### Plan factoriel

Le *plan factoriel* propose une recherche exhaustive de l'effet de la variation de chacune des variables. Lorsque la variation des variables est réalisée sur deux niveaux distincts, il permet de construire un modèle linéaire. Il permet également de rendre compte des interactions entre les variables. Lorsque la variation des variables est réalisée sur trois niveaux distincts, il permet de construire un modèle quadratique. Des modèles d'ordre supérieur peuvent être réalisés mais l'augmentation du  $k$  nombre d'essais suit alors une loi factorielle  $2^k$ .

#### Plan central-composite

Le *plan central-composite* (plan de Box-Wilson) permet de construire des modèles d'ordre 3. Il a été proposé dès 1950. Il s'appuie sur un *plan factoriel* à deux niveaux. Ce modèle est complété en réalisant des essais pour les valeurs médianes des variables (appelés *essais au centre*). Enfin, on réalise des essais où chaque variable est ajustée indépendamment des autres (*essais axiaux*). Un coefficient  $\alpha$  définit la valeur de la variation de la variable par rapport à l'essai au centre. Pour un nombre  $n$  de variables à étudier, on choisit généralement  $\alpha = \sqrt{n}$ . D'autres méthodes de choix de  $\alpha$  sont possibles et nous invitons le lecteur à consulter un ouvrage de référence sur la théorie des plans de surfaces de réponses tel que *Response Surface Methodology* [Mye71]. Le *plan central-composite* réduit le nombre d'essais à réaliser en comparaison du *plan factoriel*. Le plan de Box-Behnken propose de réduire le nombre d'essais en n'utilisant pas un plan factoriel comme plan de référence. Les plans de PLACKETT et BURMAN [PB46] proposent de réduire le nombre d'essais pour une étude à deux niveaux, dans le cas où le nombre d'essais est un multiple

de 4 ( $N = 12, N = 20, N = 24 \dots$ ). Ces plans sont idéaux pour réaliser une étude de nombreuses variables en peu d'essais et construire un modèle linéaire.

### Plan hypercube latin

L'objectif de ce plan est de couvrir la plage de valeurs des variables de manière quasi-aléatoire. La méthode de construction de ce plan a été proposée par MCKAY, BECKMAN et CONOVER [MBC79].

Soit une grille à deux dimensions qui représente les valeurs de deux variables que l'on étudie. On positionne sur cette grille des essais. Cette grille est un *carré latin* si elle ne contient qu'un unique essai par colonnes et par lignes. Un *hypercube latin* est la généralisation de ce principe pour une grille à  $d$  dimensions. Cette méthode garantit un positionnement quasi-aléatoire des essais dans l'espace de dimensions  $d$ . À la différence des plans factoriels §3.4.2 ou composite centré §3.4.2, le nombre d'essais doit être connu pour construire le plan. Cette méthode garantit que les essais soient réalisés sur l'ensemble de la variabilité du phénomène à modéliser ; à la différence d'un tirage aléatoire de la valeur des variables où il n'y a pas cette garantie. Les plans d'hypercubes latins sont les plus utilisés dans la littérature de l'optimisation d'hyper-paramètres.

Il est également possible d'utiliser une séquence quasi-aléatoire pour réaliser le choix des essais. La séquence de SOBOL' [Sob67] est la plus utilisée dans la littérature. Lorsque le nombre de variables est inférieur à 100, il permet une répartition plus intéressante que l'hypothèse des hypercubes latins.

### 3.4.3 Optimisation par méta-modèle bayésien

Cette démarche cherche dans un premier temps à modéliser le lien entre les hyper-paramètres  $\mathbf{x}$  et les performances  $y = f(\mathbf{x}, \mathbf{D})$  du modèle  $f$  appris sur le jeu de données  $\mathbf{D}$ . Il s'agit ensuite de résoudre le problème d'optimisation des hyper-paramètres  $\mathbf{x}$  dans l'espace de valeurs des hyper-paramètres  $\mathcal{X}$ , afin de trouver les hyper-paramètres optimaux  $\mathbf{x}_*$ . L'Équation 3.40 présente ce problème dans le cas où la sortie de  $f$  est l'erreur d'apprentissage du modèle.

$$\mathbf{x}_* \in \arg \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) \quad (3.40)$$

$f$  est approximée par un méta-modèle  $\mathcal{M}$ , sous la forme de procédés gaussiens<sup>6</sup>, qui permet de prendre en compte l'incertitude sur les futures évaluations de  $m$ , Équation 3.41.

$$y = f(\mathbf{x}) \sim \mathcal{M}(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \quad (3.41)$$

On cherche à modéliser la distribution de probabilité de  $y$  connaissant les hyper-paramètres  $\mathbf{x}$  :  $p(y|\mathbf{x})$ . Le procédé gaussien  $\mathcal{M}$  est défini par sa moyenne  $\mu$  et son noyau  $k$  (différents noyaux sont présentés dans la Section précédente §3.9). Le coût de l'optimisation des procédés gaussiens est néanmoins de l'ordre de  $O(N^3)$ . L'optimisation de  $f$  nécessite généralement de 10 à 100 évaluations de la fonction. Chaque évaluation de la fonction possède le coût de l'apprentissage du modèle sur le jeu de données complet. Ce coût est très grand et il est nécessaire de limiter au maximum le nombre d'évaluations du méta-modèle  $\mathcal{M}$ . Des modèles différents des coûteux procédés gaussiens ont été proposés. Nous les présentons dans les Sections suivantes §3.4.3.1, §3.4.3.2.

Il est également nécessaire de choisir judicieusement les points d'évaluations  $\mathbf{x}_{n+1}$  de  $f$  à la prochaine évaluation. Il n'est pas judicieux de choisir  $\mathbf{x}_{n+1} = \mu \mathbf{x}_n$  dans le cas où des minima locaux existeraient. Le dilemme de l'optimisation itérative apparait : c'est le choix entre l'*exploration* de

6. La théorie des procédés gaussiens a été développée par MATHERON [Mat63]. Cette démarche est aussi appelée *krigeage*, d'après les travaux originaux de KRIGE [Kri51].

$f$  pour construire un modèle  $\mathcal{M}$  pertinent qui représente tous les minima (dont le minimum de  $f$ ), et l'*exploitation* des valeurs de  $y = f(\mathbf{x}_{n+1})$  qui permet d'atteindre le minimum  $y_*$  et trouver  $\mathbf{x}^*$ . Dans le cas de l'optimisation bayésienne, l'*exploration* cherche à choisir un point qui diminuera les régions où l'incertitude du modèle est grande. C'est pourquoi JONES, SCHONLAU et WELCH [JSW98] proposent le critère de la prédiction de l'amélioration de performance prédictive (*Expected Improvement*), Équation 3.42.

$$EI_{p(f|D)} [\max(y_* - f(\mathbf{x}), 0)] \quad (3.42)$$

Lors de la prochaine évaluation de  $f$ , les valeurs  $\mathbf{x}_{n+1}$  seront choisies pour maximiser ce critère. Cela permet de choisir les valeurs des hyper-paramètres optimales pour l'évaluation suivante de  $f$ . D'autres critères, que nous ne pouvons détailler ici, existent. Ils sont étudiés en réponse aux problèmes de bandits à N bras.

### 3.4.3.1 Forêts d'Arbres décisionnels

La méthode *SMAC* (*Sequential Model-Based Algorithm Configuration*) est proposée par HUTTER, HOOS et LEYTON-BROWN [HHL11]. La démarche est identique à la modélisation gaussienne, mais elle utilise ici des Forêts d'Arbres décisionnels (§3.3.1.3). La sélection des premiers essais à réaliser peut être effectuée à partir des plans d'expériences, §3.4.2. Un plan hypercube latin ou une séquence de SOBOL' §3.4.2 est généralement utilisé.

La librairie *auto-sklearn* [Feu+15] utilise également les Forêts d'arbres décisionnels. Cette librairie utilise la majorité des algorithmes de la librairie *Scikit-Learn* [Ped+11], afin de construire automatiquement le modèle optimal. Cela représente la possibilité d'utiliser 15 classificateurs et 14 méthodes de prétraitement des données, ce qui introduit un nombre conséquent de 110 hyper-paramètres. Cet algorithme utilise un méta-modèle préalablement appris sur un grand nombre de jeux de données, afin de sélectionner les valeurs des hyper-paramètres initiaux. Cette démarche a été proposée par FEURER, SPRINGENBERG et HUTTER [FSH15]. À la manière d'un expert humain qui a appris à prioriser certains algorithmes en fonction du jeu de données auquel il fait face, un méta-modèle est construit sur 140 jeux de données. Ce méta-modèle permet de sélectionner les hyper-paramètres initiaux en fonction du jeu de données en présence. Ce travail montre la pertinence du méta-modèle pour la sélection des essais initiaux, en comparaison d'une sélection basée, par exemple, sur un plan latin hypercube §3.4.2. Enfin, *auto-sklearn* propose de construire un modèle qui agrège plusieurs méthodes différentes en fonction de leurs performances, dans une démarche ensembliste §3.3.1.3.

La Figure 3.31 montre le résultat de l'exploration des multiples classificateurs par *auto-sklearn*. Le classifieur le plus performant est ici le *Stochastic Gradient Boosting* §3.3.1.3, à contrario d'une durée d'apprentissage importante.

### 3.4.3.2 Estimateur par noyaux structuré en arbre

La librairie *hyperopt* [BYC13] (*HYPER-parameter OPTimization*) implémente la recherche aléatoire §3.4.1 et l'algorithme *TPE* (*Tree-structured Parzen Estimator*) [Ber+11] qui réalise une approximation de la fonction  $f$  par estimation par noyau<sup>7</sup>. La modélisation de chaque hyper-paramètre est structurée dans un arbre de recherche. Le parcours de l'arbre permet de réaliser l'optimisation entre de multiples paramètres. Une limite forte de cette méthode est que les interactions entre chaque hyper-paramètre ne sont pas prises en compte dans le modèle car chaque

7. L'estimation par noyau (*kernel density estimator*) est aussi appelée méthode de Parzen-Rosenblatt, car elle a été initiée par Rosenblatt en 1956 et développée par Parzen en 1962.

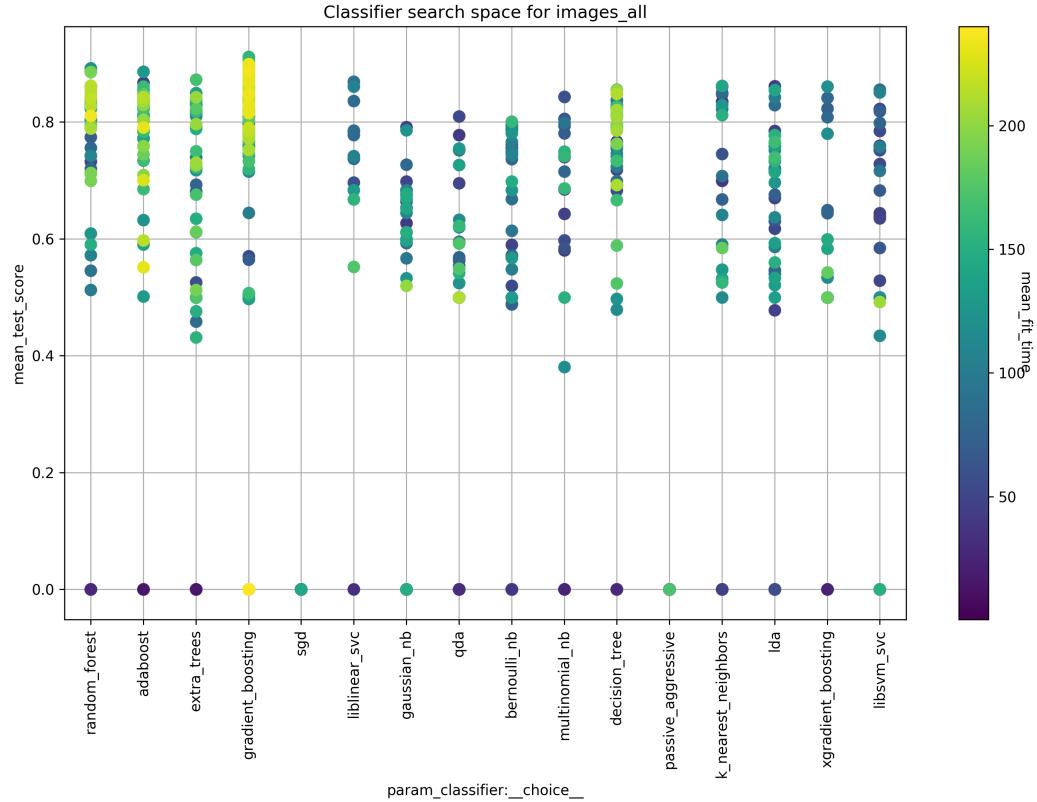


Figure 3.31 – Optimisation d'un modèle appris de manière supervisée, à l'aide de la librairie auto-sklearn [Feu+15].

hyper-paramètres disposent de leur branche indépendante sur l'arbre. Dans le cas où deux hyper-paramètres sont inter-dépendants, cette méthode est moins performante que les procédés gaussiens §3.4.3.

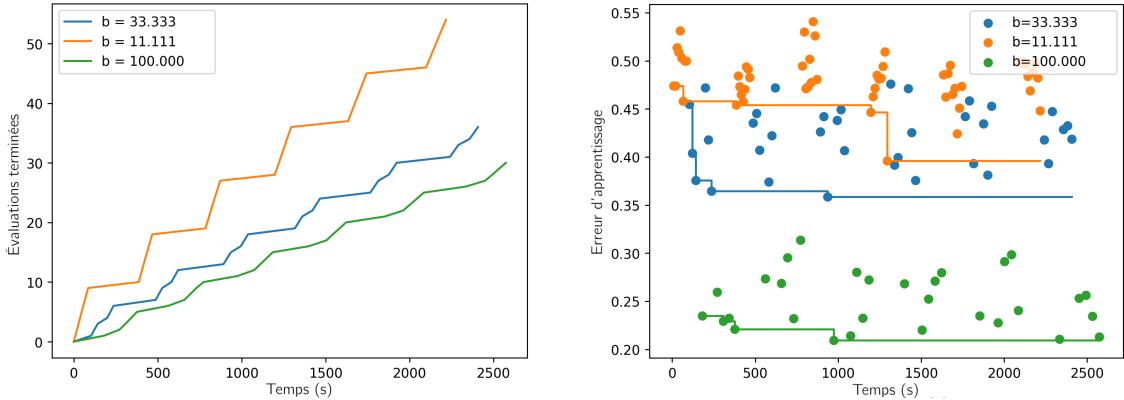
Dans le cas de l'optimisation bayésienne, on cherche à modéliser la distribution de probabilité de  $y$  connaissant les hyper-paramètres  $\mathbf{x}$  :  $p(y|\mathbf{x})$ . Dans le cas de l'estimation par densité de noyaux, on cherche à modéliser la distribution de probabilité de  $p(\mathbf{x}|y)$  et la distribution  $p(y)$ . Le théorème de Bayes permet d'obtenir  $p(\mathbf{x}|y) = p(y)p(y|\mathbf{x})$ . À chaque évaluation de  $f$ , on sépare les combinaisons d'hyper-paramètres qui obtiennent les meilleurs scores  $y \leq y^*$  pour  $\mathbf{x}^{top}$  et on définit la distribution  $m(\mathbf{x}) = p(\mathbf{x}^{meilleurs}|y)$ . Les combinaisons, dont le score est inférieur à  $y^*$ , définissent la distribution  $b(\mathbf{x}) = p(\mathbf{x}^{bof}|y)$ . On choisit généralement la limite  $y^*$  comme le premier quartile, tel que  $p(y < y^*) = 0.75$ . [Ber+11] montre que le rapport  $m/b$  est équivalent au critère de la prédiction de l'amélioration de performance (§3.4.3), ce qui permet de choisir les prochains essais en minimisant  $m/b$ .

### 3.4.4 Méthodes de réduction du coût d'évaluation du métamodèle

L'évaluation de  $f(\mathbf{x}, \mathcal{D})$  nécessite l'apprentissage d'un modèle sur le jeu de données  $\mathcal{D}$ , ce qui peut prendre plusieurs heures dans le cas de l'utilisation de réseaux de neurones profonds (*Deep Learning* 3.3.1.4). C'est pourquoi des démarches ont récemment été proposées.

#### 3.4.4.1 Évaluations partielles des configurations dans une durée limitée

Dans le cadre des algorithmes bandits, JAMIESON et TALWALKAR [JT15] propose la méthode *Successive Halving* qui permet de réduire le coût de l'optimisation de  $f$ . Il s'agit d'évaluer uniquement les configurations d'hyper-paramètres qui présentent un apprentissage de  $f$  rapide.



(a) Nombre d'évaluations de  $f$  réalisées en fonction de la limite de temps  $b$ .

(b) Erreur lors de l'apprentissage de  $f$  en fonction de la limite de temps  $b$ .

Figure 3.32 – Évaluation partielle de  $f$  en fonction de  $b$ .

Dans cet objectif, plusieurs évaluations de  $f$  sont lancées en parallèle. Une durée maximale  $b$  pour les évaluations est définie. À la fin de cette durée  $b$ , l'erreur d'apprentissage de chaque  $f$  est évaluée. La moitié des modèles qui possèdent les erreurs les plus petites est conservée. La durée maximale  $b$  est alors incrémentée et l'évaluation des modèles restants est à nouveau lancée. Cette démarche est équivalente à l'évaluation itérative de la dérivée de l'erreur d'apprentissage des modèles, tout en étant massivement parallélisable.

Cette méthode s'appuie sur l'hypothèse que le meilleur modèle  $f$  aura une vitesse d'apprentissage plus rapide que les autres. La limite de cette méthode est qu'il y a un risque de ne pas explorer des configurations d'hyper-paramètres qui donneraient de meilleures performances, mais qui sont plus lentes que de mauvaises configurations. C'est une hypothèse forte qui a été vérifiée dans ce travail sur des bases de données d'apprentissages conséquentes.

La Figure 3.32a présente l'évolution du nombre d'évaluations de  $f$  en fonction de  $b$ . L'augmentation de  $b$  limite le nombre d'évaluations de  $f$  réalisées. Il est nécessaire de sélectionner les bonnes configurations d'hyper-paramètres, ce que fait le métamodèle. La Figure 3.32b présente l'erreur d'apprentissage de modèles  $f$ , obtenue pour différentes valeurs de  $b$ . L'augmentation de  $b$  permet d'augmenter les performances du modèle.

L'utilisation de l'arrêt prématuré §3.3.1.7 pendant l'apprentissage de chaque modèle  $f$  est également proposée par DOMHAN, SPRINGENBERG et HUTTER [DSH15]. Ce travail introduit l'utilisation d'un modèle prédictif de l'évolution de l'erreur d'apprentissage. Pour cela, pendant l'apprentissage de  $f$ , un modèle bayésien (méthode Monte-Carlo par chaînes de Markov) est construit pour prédire l'évolution future de l'erreur d'apprentissage. Si l'évolution est défavorable, l'apprentissage est stoppé avant la fin du nombre d'itérations prédefinies. Dans le cas d'une utilisation de cette méthode pour l'optimisation des hyper-paramètres de réseaux de neurones profonds, les auteurs évaluent le gain à un facteur deux.

#### 3.4.4.2 Évaluation de $f$ sur un sous-ensemble des données

KLEIN et al. [Kle+16] propose d'évaluer  $f$  sur une partition réduite du jeu de données, (cette méthode est prénommée *Fabolas*). La dimension du sous-ensemble du jeu de données est progressivement augmentée au fur et à mesure que des configurations d'hyper-paramètres satisfaisantes sont trouvées. La vitesse de l'optimisation est augmentée d'un facteur 10 en comparaison d'une optimisation bayésienne ou de la méthode *Hyperband* §3.4.5. Associée à l'évaluation partielle par arrêt prématuré, cette méthode est très efficace pour limiter le coût de l'évaluation.

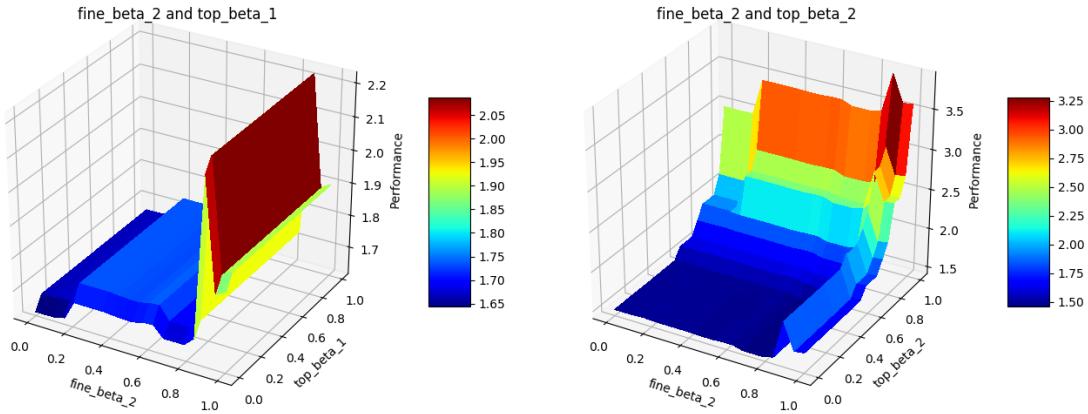


Figure 3.33 – fANOVA des hyper-paramètres suite à l’optimisation avec BOHB.

### 3.4.5 Optimisation par algorithmes bandits

Les méthodes bandits à N bras consistent en l’évaluation massivement parallèle de multiples modèles, et en la sélection des meilleurs, à chaque itération. La méthode *Hyperband* est proposée par LI et al. [Li+16]. Elle utilise la démarche d’arrêt prématuré de l’apprentissage §3.4.4.1. L’étude des auteurs démontre que, dans le cas où l’espace des hyper-paramètres est organisé de la pire des façons pour l’algorithme, la recherche d’une solution optimale amènera à un coût d’optimisation 5 fois plus grand qu’avec une recherche aléatoire §3.4.1. À contrarior, dans le cas où l’espace des hyper-paramètres est organisé dans le meilleur cas, la méthode à un coût 10 fois moins grand qu’une recherche aléatoire. La méthode utilise également l’apprentissage progressif sur des sous-ensembles des données §3.4.4.2. Les auteurs concluent sur la possibilité d’utiliser les plans d’expériences §3.4.2 pour sélectionner les configurations d’hyper-paramètres à évaluer en premier.

Récemment, la méthode *BOHB* (*Bayesian Optimization and HyperBand*) a été proposée par FALKNER, KLEIN et HUTTER [FKH17 ; FKH18]. Elle associe les techniques d’*Hyperband* pour la sélection du nombre de configurations à évaluer dans un temps limité, et ajoute l’optimisation d’un méta-modèle bayésien (similaire aux noyaux structurés en arbre *TPE* §3.4.3.2) pour sélectionner les meilleures configurations à évaluer. À la différence de *Hyperband* qui sélectionne les configurations de manière aléatoire, le méta-modèle permet de prendre en compte l’information sur les configurations acquise au fur et à mesure des évaluations. La sélection des configurations à évaluer est plus pertinente, ce qui permet d’atteindre la configuration optimale avec environ deux fois moins d’évaluations que *Hyperband*.

Afin d’observer l’influence des différents hyper-paramètres sur la performance du modèle, nous réalisons une analyse de la variance des hyper-paramètres pour  $f$  (ANAVAR/ANOVA), en utilisant la librairie HUTTER, HOOS et LEYTON-BROWN [HHL14]. Lorsque le nombre d’évaluations du modèle est assez grand ( $> 100$ ), cela permet de rendre compte des interactions possibles entre les hyper-paramètres. La Figure 3.33 présente les résultats pour deux couples d’hyper-paramètres ( $\beta$  est ici un terme de pénalisation des poids §3.3.1.7). On observe que les surfaces de réponses sont complexes. De plus, le nombre d’hyper-paramètres présent est ici de 38 et des interactions existent entre les hyper-paramètres. C’est pourquoi un expert humain aurait eu des difficultés à réaliser cette optimisation manuellement, dans un espace continu de dimensions 38.

### 3.4.6 Optimisation par algorithmes évolutionnistes

Proposée en 2016 dans OLSON et al. ; OLSON et al. [Ols+16a ; Ols+16b], *TPOT* (*Tree-Based Pipeline Optimization Tool*) s’appuie sur un algorithme génétique pour explorer les configurations

d'hyper-paramètres et sur une Forêt d'Arbres décisionnels §3.3.1.3 pour optimiser le meilleur modèle.

Nous avons appliqué cette méthode dans “Polarimetric Imaging for Quality Control in Injection Molding” [Nag+19], sur notre problématique de l'apprentissage supervisé d'un modèle pour le contrôle de la qualité. Sur des images de dimensions  $331 \times 331$  pixels, l'évaluation de 300 modèles  $f$  est effectuée en 14 heures. Le gain de performances suite à l'optimisation des hyper-paramètres est de 2%. L'algorithme a trouvé le meilleur modèle possible en une demi-journée. Un expert humain n'aurait certainement pas réussi à trouver le meilleur modèle en si peu de temps. Malgré la durée qui est importante, *TPOT* permet de remplacer plusieurs journées d'expertise d'un ingénieur en sciences des données.

### 3.4.7 Optimisation de l'architecture de réseaux de neurones profonds

Afin qu'un modèle à réseaux de neurones profonds soit fonctionnel, son architecture doit être optimisée. Il est nécessaire d'effectuer une recherche de différentes architectures. Deux solutions apparaissent : une étude bibliographique des meilleurs modèles proposés sur le problème (aussi appelée *Grad-Student Descent*), ou l'automatisation du processus de recherche (aussi appelée *Neural Architecture Search*), que nous présentons dans la suite. Le coût de la recherche et de l'optimisation de ces modèles est très important, à cause de l'infinité d'architecture possibles et des nombreux hyper-paramètres. C'est pourquoi, il est nécessaire de disposer de ressources de calculs conséquentes. Dans le cas où ces ressources ne sont pas disponibles, on préférera utiliser un démarche d'apprentissage par transfert de domaine §3.3.2.

ZOPH et LE [ZL16] propose de réaliser un apprentissage automatique de l'architecture de réseaux profonds : c'est le *Neural Architecture Search*. La méthode permet de proposer l'architecture *NASNet* [Zop+17] qui établit un nouvel état de l'art en matière de performance de classification sur les jeux de données *ImageNet* et *CIFAR-10* [Kri09].

Nous avons étudié l'intérêt de cette démarche d'optimisation de l'architecture d'un réseau de neurones. Dans notre étude [Nag+19], nous avons utilisé la librairie *Auto-Keras* [JSH18] sur notre problématique. Cette méthode propose d'utiliser des architectures de la littérature comme solutions initiales, puis de les modifier progressivement, en utilisant les principes de *Neural Architecture Search* [ZL16]. Un méta-modèle bayésien §3.4.3 est construit pour parcourir et optimiser les hyper-paramètres associés au réseau. Les modèles à réseaux de neurones profonds nécessitent d'être optimisé sur processeur graphique. Aussi, l'architecture du modèle est contrainte à la dimension de la mémoire de travail du processeur graphique sur lequel le modèle est stocké. Dans le cas de notre expérimentation, les 11 Giga octets de notre processeur graphique était limitant. Il n'a pas été possible d'obtenir de meilleures performances que celles obtenues par une optimisation manuelle du réseau.

La recherche d'architecture de réseaux de neurones possède de grands enjeux, car il est difficile pour les humains d'inventer des architectures plus performantes, aux vues de leurs complexités grandissantes. Dernièrement, [EMH18] propose un état de l'art exhaustif de la littérature *AutoML*. Les dernières avancées s'appuient sur l'apprentissage par renforcement pour proposer une recherche d'architecture efficiente. C'est une perspective de recherche importante pour la modélisation par apprentissage.

### 3.4.8 Conclusion : AutoML pour l'industrie

L'utilisation de solutions d'optimisation automatique de modèles rejoint une double question économique. Est-il avantageux d'optimiser le modèle pour le procédé industriel ? C'est à dire d'augmenter les performances, par exemple dans le cadre du contrôle de la qualité. De même, est-il plus coûteux d'optimiser manuellement les hyper-paramètres d'un modèle, ou d'investir dans la puissance de calcul nécessaire à l'automatisation ? Il est aujourd'hui plus avantageux d'utiliser les méthodes *AutoML*.

Dernièrement, GIJSBERS et al. [Gij+19] réalise une comparaison de 4 méthodes sur 39 jeux de données. Ils montrent que sur certaines tâches, l'optimisation n'apporte pas nécessairement de gain, en comparaison d'un classifieur à Arbres de Forêts décisionnels. Les méthodes *auto-sklearn* et *TPOT* obtiennent des performances similaires dans la majorité des cas. Dans de nombreux cas d'applications, il n'est pas nécessaire d'utiliser un modèle compliqué.

Notre problématique industrielle impose des contraintes en matière de puissance de calcul disponible. C'est pourquoi dans notre cadre, l'optimisation de modèles complexes, tels que les réseaux de neurones profonds, peut être effectuée à une fréquence hebdomadaire. Le coût d'une optimisation journalière est encore trop conséquent pour être envisagé pour l'industrie. En revanche, sur des cas d'applications à fortes valeurs ajoutées, le modèle doit avoir les meilleures performances possibles. Il est alors envisageable d'investir dans de plus grandes infrastructures de calculs, afin de réaliser une optimisation journalière des modèles.

## 3.5 Conclusion : Modélisation par apprentissage pour l'industrie

Nous avons choisi dans ce chapitre d'étudier une diversité représentative des méthodes et algorithmes existants pour résoudre le problème du contrôle automatique de la qualité. Maîtriser l'ensemble du vaste domaine de l'apprentissage statistique est un processus difficile qui requiert plusieurs années. Les thématiques de l'architecture logicielle, l'application des calculs sur les ressources matérielles, l'efficience des algorithmes et la large thématique de l'optimisation non-convexe sont au cœur de ce domaine. Dernièrement, la littérature de ce domaine est très dynamique, ce qui rend également difficile le suivi des avancées. Les moyens de production industrielle ont une inertie beaucoup plus grande que cette littérature, c'est pourquoi il est nécessaire de concevoir des systèmes modulaires, qui pourront être mis à jour, au fur et à mesure des progrès de l'apprentissage statistique.

### Contraintes industrielles

Nous avons choisi de proposer une vision globale appliquée à notre problématique industrielle. Le choix des méthodes et des algorithmes à utiliser dépend fortement des contraintes industrielles. La Figure 3.34 positionne les méthodes présentées dans l'ensemble des méthodes d'apprentissage existantes. Pour chaque méthode, nous avons discuté du coût des calculs. C'est une des principales limites au déploiement de solutions d'apprentissage dans l'industrie. C'est notamment la limite du coût du calcul, en rapport à sa durée, qui est encore aujourd'hui critique. Nous n'avons pas abordé la problématique de l'utilisation des modèles appris, notamment de la vitesse de l'inférence. En effet, compte tenu des puissances de calculs disponibles aujourd'hui dans les systèmes embarqués, l'inférence à partir d'images peut être réalisée en quelques millisecondes. C'est l'apprentissage des modèles et leurs mises à jour, qui sont coûteux.

Le Tableau 3.11 synthétise les discussions sur le coût des calculs. Le choix de la méthode utilisée est dépendant de la fréquence de mise à jour du modèle de la qualité et de la disponibilité des moyens de calculs. Si les calculs doivent être effectués au plus près de la production, sur des systèmes embarqués, les réseaux de neurones les plus profonds ne sont pas envisageables. Enfin, il est possible de s'appuyer sur les services de solutions de calculs distribués distantes (*Cloud*), qui limitent la nécessité de maintenance locale des moyens de calculs. Elle nécessite néanmoins le déploiement d'une infrastructure de réseau informatique fiable.

### Vers un modèle robuste de la qualité

La dernière limite est certainement la plus importante. Il s'agit du nombre d'échantillons présent

Algorithme	Complexité	Référence
Machine à vecteurs de support	$O(N^2)$	§3.3.1.3
Réseaux de convolutions	$O(N T W)$ , $T$ nb. itérations, $W$ dim. des poids	§3.3.1.4
Analyse en Composantes Principales	$O(N)$	§3.3.4.3
ACP à noyau	$O(N^3)$	§3.3.4.4
k-moyennes	$O(N \log(N))$	§3.3.4.1
t-SNE	$O(N^2)$	§3.3.4.5
UMAP	$O(N \log N)$	§3.3.4.6
Auto-encodeur variationnel	$O(T D)$ , $T$ nb. itérations, $D$ dim. échantillon	§3.3.4.7
Réseaux antagonistes adversaires	$O(T D)$ , $T$ nb. itérations, $D$ dim. échantillon	§3.3.4.8

Tableau 3.11 – Complexité de l'apprentissage des méthodes étudiées en fonction du nombre d'échantillons du jeu de données  $N$ .

dans le jeu de données. Nous disposons actuellement de jeux de données qui contiennent moins de mille échantillons. L'état de l'art en *Deep Learning*, nous permet d'envisager la quantité de données nécessaire afin de rendre robuste le système, pour toute machine, avec dix mille pièces dans le jeu de données d'apprentissage (le jeu de données *CIFAR-10* [Kri09] propose 6000 images par classes).

La faisabilité du transfert de la modélisation de la qualité apprise sur une machine, vers une autre machine produisant la même pièce, est en cours d'évaluation. À terme, cette démarche permettra de réaliser l'apprentissage supervisé de la notion de qualité sur une unique machine, puis de déployer le modèle sur un parc machines complet. Cela implique que le modèle doit avoir acquis dans sa base d'apprentissage des images prises pour une même pièce, mais dans des configurations de machines différentes.

La littérature renseigne des cas de biais, où le modèle produit des résultats satisfaisants en apparence, alors que le modèle ne s'intéresse pas aux véritables causes. Dans notre cas, il pourrait s'agir d'une machine qui possède une grille de protection de couleur rouge, qui produirait uniquement des pièces mauvaises, tandis qu'une machine de couleur verte produirait uniquement des pièces bonnes. Le modèle aura alors généralisé par apprentissage qu'aucune pièce ne pourrait être bonne si produite à côté d'une grille rouge. Ce résultat sera alors très difficile à comprendre. L'interprétation des résultats produite par un modèle est un enjeu important. Les visualisations par apprentissage non-supervisé permettent dans notre cas d'identifier des paquets d'échantillons similaires. Nous obtenons de 4 à 6 classes sur l'étude de la pièce "boîte noire", qui correspondent à différents types de défauts.

Enfin, nous pouvons également envisager la construction d'un modèle de qualité qui serait robuste à toutes pièces. Ce modèle rivaliseraient avec l'humain afin de classifier les défauts de n'importe quelle pièce. Construire ce modèle nécessiterait une base de données d'apprentissage supérieure à cent mille pièces (*ImageNet* propose quatorze millions d'images pour vingt mille catégories). Il nécessiterait aussi un modèle à réseaux de neurones profonds ayant la capacité de prendre en compte ces données.

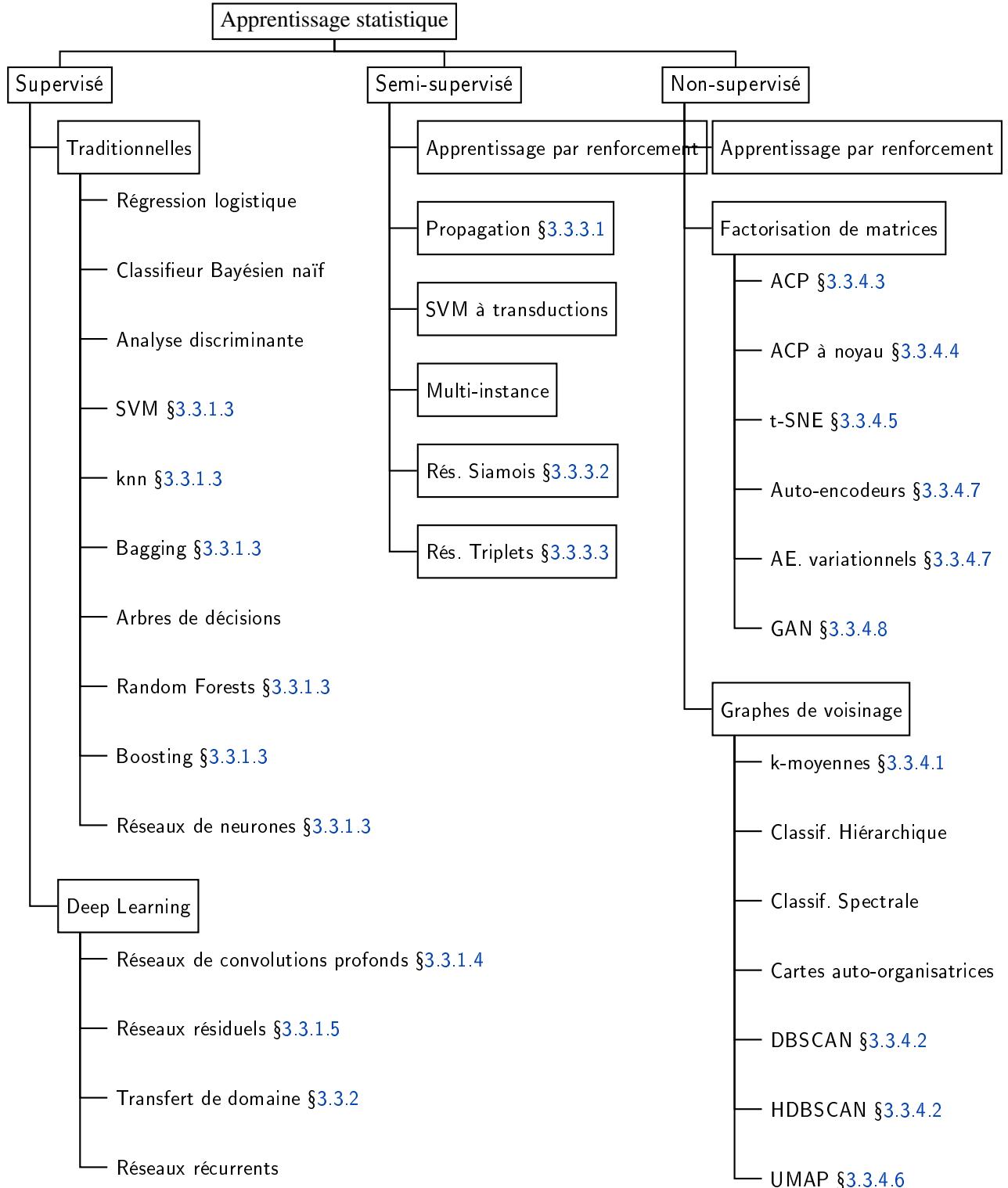


Figure 3.34 – Taxonomie de l'apprentissage statistique.



# 4

## Conception d'un système de mesure de la Qualité

---

*Résumé du chapitre possible ici.*

---

### Sommaire

---

<b>4.1</b>	<b>Conception d'un dispositif de mesure adapté aux contraintes industrielles</b>	<b>143</b>
4.1.1	Imagerie non-conventionnelle pour la mesure non-invasive de la Qualité	143
4.1.2	Détection de pièce robuste	143
<b>4.2</b>	<b>Solution logicielle</b>	<b>143</b>
4.2.1	Architecture logicielle	143
4.2.2	Interface Homme-Machine	143
4.2.3	Apprentissage de la notion de Qualité	144
<b>4.3</b>	<b>Application industrielle</b>	<b>144</b>
4.3.1	Contrôle de la qualité par apprentissage en industrie	144
4.3.2	Résultats obtenus par le dispositif de mesure TheEye	144
<b>4.4</b>	<b>Évaluation de la performance du système de mesure de la Qualité</b>	<b>144</b>
4.4.1	Comparaison avec un expert humain	144
4.4.2	Métrique de confidence industrielle	144

---

## 4.1 Conception d'un dispositif de mesure adapté aux contraintes industrielles

Poser les hypothèses Justifier les choix Polarisation pk ? Deep Learning pk ? Architecture pk ?  
A chaque fin de chapitre : ouverture, ce qu'il reste à faire  
Brevet [NPP19]

### 4.1.1 Imagerie non-conventionnelle pour la mesure non-invasive de la Qualité

Qu'est que c'est?. C'est une phrase français avant le lorem ipsum. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdier mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

### 4.1.2 Détection de pièce robuste

Qu'est que c'est?. C'est une phrase français avant le lorem ipsum. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdier mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

## 4.2 Solution logicielle

### 4.2.1 Architecture logicielle

Qu'est que c'est?. C'est une phrase français avant le lorem ipsum. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdier mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

### 4.2.2 Interface Homme-Machine

Qu'est que c'est?. C'est une phrase français avant le lorem ipsum. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdier mi nec ante. Donec ullamcorper, felis non sodales

commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

#### 4.2.3 Apprentissage de la notion de Qualité

- Des capteurs partout - Des caméras thermiques, mais l'humain n'est pas capable de les exploiter. - IA = information synthétisée, exploitée

### 4.3 Application industrielle

#### 4.3.1 Contrôle de la qualité par apprentissage en industrie

[Got+19] T de Hotelling [FC19] driven by features [Meg+19] Deflectometry Based Inline Inspection of Linearly Moving Plastic Parts

#### 4.3.2 Résultats obtenus par le dispositif de mesure TheEye

-> Perspectives, TheEye, nouveaux champs de recherche : - à la Qualité, au suivi en temps réel des dérives de la production, à l'ajustement du procédé - à d'autres mesures - à d'autres sources d'infos - à d'autres produits

-> Perspectives, TheEye, acquisisiton massive de données pour améliorer les modèles : - maximiser l'adoption - commercialisation pas chère = récupérer un maximum de données pour améliorer les modèles. - voir même commercialisation gratuite pour maximiser la récolte de données.

### 4.4 Évaluation de la performance du système de mesure de la Qualité

Coût industrielle de la non-qualité

#### 4.4.1 Comparaison avec un expert humain

Gage R et R avec expert humain

#### 4.4.2 Métrique de confidence industrielle

faux positif, faux négatif ppm Limite au déploiement : 2-3 x meilleur qu'un humain Coût industriel de la non-qualité vs coût du déploiement d'un tel système, fonction de la valeur ajouté du produit

À propos des attaques par exemple adversaire : détecter l'attaque.



## Conclusion générale

Conclusion ici.

Conclusion Throughout this thesis, we have made contributions to the four areas of transfer learning in natural language processing : domain adaptation (§4), cross-lingual learning (§5), multi-task learning (§6), and sequential transfer learning (§7). In this final chapter, we will recapitulate the proposed methods (§8.1), summarize our findings (§8.2), and provide an outlook into the future (§8.3).

8.1 Synopsis In this dissertation, we have studied the problem of automatically learning transferable representations for a variety of NLP tasks. In Chapter 3, we have presented a taxonomy for transfer learning for NLP and provided a comprehensive overview of each area. Chapter 4 presented approaches for domain adaptation that overcome the discrepancy between domains by automatically selecting relevant and informative examples. In particular, it presented a novel automatic data selection method and a novel semi-supervised learning approach that combines the strengths of both classic and neural techniques. We evaluated our methods across three diverse NLP tasks and multiple domains. Chapter 5 focused on cross-lingual learning, particularly unsupervised and weakly supervised approaches. We first extensively analyzed the limitations of unsupervised approaches and proposed a weakly supervised method that mitigates them. We then proposed a different latent-variable view of existing neural methods and a novel cross-lingual word embedding model that combines both traditional and current approaches. We evaluated our methods on bilingual dictionary induction where they outperform the state of the art.

250 Conclusion 251 In Chapter 6, we studied the problem of multi-task learning across heterogeneous tasks. We proposed two novel models that enable more flexible sharing than existing architectures. We evaluated our methods on diverse sets of tasks and domains where they outperformed hard parameter sharing and strong single-task models. Chapter 7 tackled the most impactful sequential transfer learning scenario involving transfer to an arbitrary target task. We proposed ULMFiT, a novel framework that pretrains language model representations and then transfers them to a target task using novel techniques. Our method achieved significant error reductions over the state-of-the-art across a range of text classification datasets. We additionally compared feature extraction and fine-tuning of two state-of-the-art pretrained representations across a diverse set of tasks. We found that the similarity between pretraining and target task is indicative of the relative performance between feature extraction and fine-tuning. We finally provided practical guidelines to practitioners.

8.2 Summary of findings This dissertation investigated the hypothesis that deep neural networks that leverage relevant information using transfer learning techniques outperform their supervised counterparts across a wide range of tasks. Over the course of this thesis, we have presented multiple novel methods for different transfer learning scenarios and evaluated them across a diversity of settings where they outperformed single-task learning as well as competing transfer learning methods. We now recapitulate how our methods addressed the desiderata we laid out initially and summarize our contributions and findings.

Research objectives Overcoming a discrepancy between source and target setting Transferring between dissimilar or distant settings is one of the biggest challenges in transfer learning and a key reason for the failure of current systems. Domain adaptation methods deteriorate when domains are dissimilar; cross-lingual embedding methods fail when languages are distant; and multi-task learning approaches break down when tasks are unrelated. We have proposed methods that seek to address this challenge in every setting : For domain

adaptation, we have proposed a method that takes into account the relevancy and informativeness of examples using Bayesian Optimization (§4.1). When learning cross-lingual word embeddings between distant languages such as English and Finnish, we characterized this distance with a novel eigenvector similarity metric and found weak supervision in the form of identically spelled strings to provide an effective bridge (§5).

Conclusion 252 Our proposed architectures for multi-task learning aim to be as flexible as possible so that beneficial information can be shared, but unrelated tasks do not result in worse performance (§6). In particular, using subspaces and allowing for sharing across the hierarchy proved beneficial. Our new framework for sequential transfer learning learns general-purpose representations that can be transferred to any target task (§7.1). In this framework, we proposed several techniques that facilitate transfer to a different target task. We found these techniques key in adapting LSTM-based representations and observed that task similarity plays a role in adaptation. When tasks are dissimilar, feature extraction is preferred, while fine-tuning works better for related tasks (§7.2). Inducing an inductive bias Much of the work in this thesis has focused on leveraging an inductive bias that narrows a model’s hypothesis space and improves its ability to generalize. To this end, we employed a wide array of inductive biases. We used a model’s or auxiliary models’ predictions on unlabelled data as additional learning signal via semi-supervised learning (§4.2, §6.2). Multi-task learning similarly was a pervasive inductive bias that we employed across many models (§4.2, §6). Another common bias that we found helpful is an orthogonality constraint, which allowed us to encourage multi-task learning models to focus on learning different features (§4.2, §6.1). For cross-lingual learning, we found weak supervision in the form of identically spelled strings to be a useful inductive bias (§5). As another inductive bias for cross-lingual learning, we employed a 1 : 1 matching prior that enabled us to learn a better matching between words and their translations and mitigate the hubness problem (§5.2). In our multi-task learning model, we found a hierarchical bias in the form of a latent mixture to be helpful (§6.1). Finally, pretrained language model representations allowed us to achieve state-of-the-art performance across a wide range of tasks (§7). Combining traditional and current approaches In this thesis, we have extensively reviewed the existing literature (§3). At the same time, many current approaches reinvent the wheel, replacing classic feature-based approaches with a different toolkit rather than seeking inspiration and drawing on lessons of the past. While classic approaches were an inspiration for this thesis in general, in two cases we explicitly combine them with current methods. For semi-supervised learning, we adapt classic bootstrapping approaches to neural networks and propose a novel method that combines both the agreement-based learning of tri-training and the efficiency of neural multi-task learning (§4.2). For cross- lingual learning, we combine the theoretically motivated bipartite matching dictionary prior of Haghghi et al. [2008] with a state-of-the-art embedding-based approach (§5.2).

Conclusion 253 Transfer across the hierarchy of NLP tasks Tasks in natural language processing are naturally hierarchical and range from tasks that require morphological information, to syntactic and semantic tasks. This hierarchy offers challenges, as it complicates sharing of layers, but also brings opportunities, if information related to different levels of meaning can be organized and accessed reliably. Our multi-task meta-architecture was designed with this hierarchy in mind. It enables sharing across all levels of the hierarchy, from low-level to high-level tasks (§6.1). Our second multi-task architecture focuses on the highest layer, the label embedding. It leverages new mechanisms to better integrate information across related label spaces to transfer between coarse-grained and fine-grained sentiment tasks (§6.2). Finally, ULMFiT introduces new techniques that seek to minimize catastrophic forgetting during adaptation. It employs gradual unfreezing and discriminative fine-tuning to retain the general information in the lowest layers of the model and transfer these to the target task (§7.1). Generalization across many settings We were interested in methods that do not depend on any particular task and in models that generalize across a wide range of settings. We thus evaluated our models on a wide range of tasks including POS tagging, dependency parsing, and sentiment analysis (§4); chunking, NER, semantic role

labelling, and POS tagging (§6.1); pairwise classification tasks (6.2); text classification tasks (§7.1); and seven diverse NLP tasks (§7.2). In general, we were interested in models that not only perform well on in-domain data but also achieve good performance on data outside of the training distribution. To this end, we evaluated our models on product review, movie review, and Twitter domains for sentiment analysis (§4, §6.2, §7.1), web domains for sequence labelling tasks (§4, §6.1), and multiple domains for topic classification (§7.1). We finally evaluated our approaches on a diverse set of languages, including Estonian, Finnish, Greek, Hungarian, Polish, and Turkish (§5.1) as well as English, German, Italian, Finnish, Turkish, Bengali, and Hindi (§5.2). Contributions We now summarize this thesis' contributions with regard to transfer learning overall and each transfer learning area in particular : Transfer learning We presented a taxonomy for transfer learning for NLP (§3.1.3) by adapting the widely used taxonomy of Pan and Yang [2010]. The taxonomy reflects the most common transfer learning scenarios encountered in current natural language processing. We extensively reviewed each of the scenarios (§3.2, §3.3, §3.4, §3.5) and highlighted connections and similarities between the approaches of different areas. For

Conclusion 254 instance, we made it clear where sequential transfer learning and domain adaptation approaches build on multi-task learning techniques. Domain adaptation We have presented a model that automatically learns to select training examples that are relevant for a particular target domain (§4.1). We have evaluated this model across three tasks where it outperformed existing domain similarity metrics and demonstrated the transferability of the learned policy across different settings. We additionally re-evaluated a range of classic self-labelling approaches to neural networks in the context of a domain shift (§4.2). We found that tri-training works best and even outperforms a recent state-of-the-art method. We addressed the drawback of tri-training, its time and space complexity, with a more efficient multi-task tri-training method. Our novel method outperformed both traditional tri-training and recent alternatives on sentiment analysis. For POS tagging, classic tri-training still was superior, particularly on OOV and low frequency tokens, which suggests it is less affected by error propagation. Cross-lingual learning We presented a taxonomy for cross-lingual word embedding models that organizes models based on the main sources of variation, the type of alignment and the comparability of the training data (§3.5.3). We additionally showed that cross-lingual word embedding models that learn on the word level optimize similar objectives by reducing each model to a combination of monolingual losses and regularization terms (§3.5.4). We theoretically and empirically analyzed the limitations of unsupervised cross-lingual embedding models and showed that unsupervised methods perform much worse on distant language pairs, with monolingual corpora from different domains, and different embedding algorithms (§5.1). We proposed a novel eigenvector-based metric to characterize the distance between monolingual embedding spaces and gauge the potential of unsupervised methods. We furthermore presented a novel latent-variable model for bilingual lexicon induction that combines a bipartite matching prior with a state-of-the-art embedding-based approach, which we evaluated on three standard and three low-resource language pairs where it outperformed the state-of-the-art (§5.2). We finally showed how existing approaches can be viewed as a similar latent variable model with a different prior. Multi-task learning We presented sluice networks, a novel multi-task learning framework that automatically learns which layers to share between different tasks and detailed how this model generalizes existing transfer learning and multi-task learning architectures (§6.1). We evaluated the model on synthetic data and data from OntoNotes 5.0 covering multiple tasks and domains where it consistently outperformed single-task and multi-task baselines. We additionally proposed a novel multi-task learning model that integrates

Conclusion 255 information from disparate label spaces and propose methods to leverage unlabelled and auxiliary data (§6.2). We evaluated the model on a variety of pairwise sequence classification tasks where it outperformed single and multi-task baselines and achieved a new state of the art for aspect- and topic-based sentiment analysis. Sequential transfer learning We proposed a novel framework for sequential transfer learning using pretrained language models

and novel adaptation techniques including novel ways to fine-tune a language model (§7.1). Our method convincingly outperformed the state-of-the-art on six text classification tasks with an error reduction of 18–248.3 Future directions In this section, we will give an outlook into the future for each of the respective transfer learning areas in particular and for transfer learning in general. Domain adaptation Robustness to out-of-distribution data As we become more cognizant of the brittleness of our current methods, we hope that more work will focus on making them more robust. To test robustness, evaluation on out-of-domain data will become more common. The recent CoQA [Reddy et al., 2018] dataset already includes out-of-domain evaluation. We hope that in the future when new datasets are created, in addition to a random test set, dedicated out-of-domain test sets will become common practice. This will enable us to test whether our models truly generalize.

Conclusion 256 Challenge sets A related direction is to evaluate models on challenge sets, which have recently become more popular [Linzen et al., 2016, Gulordava et al., 2018]; see [Belinkov and Glass, 2019] for an overview. Such challenge sets probe particular aspects such as certain linguistic phenomena on which current models fail. Challenge sets have been mainly created for tasks such as question answering and machine translation and English. In the future, we expect the creation of challenge sets for other tasks and non-English languages. Adversarial examples One way to induce robustness is via adversarial examples [Goodfellow et al., 2015, Wang and Bansal, 2018]. Adversarial examples are an ongoing research topic and are closely related to understanding the generalization properties of current models. Adversarial examples have become more common in NLP, but generating useful adversarial examples that work across tasks and domains is still a challenge. Semi-supervised learning We also expect more effective approaches that combine explicit semi-supervised learning such as self-labelling (§4.2) with transfer learning. While transfer learning allows us to leverage unlabelled data via an unsupervised pretraining or auxiliary task, semi-supervised learning uses unlabelled data directly for our target task, for instance by making the model’s predictions more consistent with itself. Current semi-supervised learning methods, however, still struggle with a domain shift (§3.4.4.2) and need to be made more robust to be used at their full potential and as a complement to transfer learning methods. Cross-lingual learning Cross-lingual resources In order to bridge the digital language divide, NLP models must be applied to non-English languages. To this end, resources will need to be created not just for English but for the world’s other 6,000 languages. This can be done comparatively inexpensively in many cases by releasing small test datasets in multiple languages that can be used to evaluate cross-lingual models. A recent example of such a data set is XNLI [Conneau et al., 2018b], which contains NLI test sets for 15 languages. Unsupervised cross-lingual methods As training data is not available for most of these languages, unsupervised and weakly supervised cross-lingual methods will be required to deal with such low-resource scenarios (§5). As we have shown, current unsupervised methods are still unstable in many settings, such as in the context of different domains, dissimilar languages, or different algorithms. Developing more robust unsupervised methods is thus an important research direction.

Conclusion 257 Deep cross-lingual embedding models Similar to the progression from shallow word embeddings to deep language models, we expect training of deep cross-lingual models to be a fruitful research direction. The main challenge here is to enable models to go beyond learning from sentence-level parallel data and incorporate non-parallel monolingual corpora. To this end, unsupervised and weakly-supervised techniques will also be useful. A shared representation at the input layers such as cross-lingual word embeddings or a shared word piece vocabulary [Lample et al., 2018b] may often be sufficient to kick-start the training of a deep model. Multi-task learning Massively multi-task models Hard parameter sharing will likely still be used as the default when learning between two tasks as the method is simple and efficient. However, models that learn from more tasks will likely move away from this paradigm as it will be too restrictive. In particular, we expect the development of ‘massively multi-task models’ that are heavily modularized and update only a comparatively small number of parameters for each task. Such architectures will

be more flexible and allow for variable degrees of sharing (§6). A challenge will be to allow for such flexibility without the number of parameters blowing up. Convergence of multi-task learning and sequential transfer learning Given a larger number of tasks, sequential transfer learning and multi-task learning will likely converge, as it will often be prohibitive to train all tasks jointly. Instead, many tasks will be pretrained and optionally equipped with regularizers to encourage positive backwards transfer. A related direction is the development of scalable lifelong learning approaches that are not memory-based and scale sublinearly with the number of tasks. Understanding task relationships Finally, we hope for more efforts focusing on understanding the relationships between tasks. So far, most of the insight in NLP has come from anecdotal evidence and small-scale empirical studies that investigated the pairwise relationships between a few tasks (§3.2.6). A large-scale study such as the one done by Zamir et al. [2018] in computer vision is necessary to obtain more generally useful insights. A confounding factor is the role of the domain, which is harder to isolate in NLP. As tasks are diverse, ranging from natural language generation, to sequence tagging, natural language understanding, and text classification, it is harder to obtain annotations for each instance across many tasks.

Conclusion 258 Sequential transfer learning Specialized pretraining tasks We expect deep pretrained representations to become a common tool for the NLP practitioner, in their utility similar to pretrained word embeddings in the past years. As transfer learning becomes more pervasive, we expect the emergence of dedicated pretraining tasks that encode specialized knowledge. This is already evidenced in approaches that report gains by combining different pretrained word embeddings [Kiela et al., 2018b]. Eventually, downstream models will be able to choose from a large range of pretrained representations. Some of these will be specialized to particular domains such as biomedical or legal documents, while others will be specialized to specific tasks such as particular forms of reasoning and natural language understanding, e.g. anaphora resolution, arithmetic reasoning, etc. By sparsely selecting among a set of relevant modules, approaches will both be more computationally efficient and use richer representations compared to classic NLP pipelines. Adaptation We similarly expect more work focusing on effective adaptation strategies that allow us to go from the pretrained initialization to the final trained model more effectively. When adapting to a few number of examples, meta-learning might be a viable method, though it still requires access to labelled data from a diverse distribution of tasks for training the meta-learner. As pretrained models get deeper, we will see best practices emerge regarding the adaptation of many layers that leverage different learning rates or freeze part of the model, such as the techniques we proposed (§7). In addition, methods that flexibly augment pretrained parameters with task-specific ones—similar to residual adapters [Rebuffi et al., 2017]—and that extend these connections to multiple tasks will become more commonplace. Transfer learning On a broader note, in the long-term we expect transfer learning to be an integral part of NLP systems. It will become the exception that models are trained from scratch; particularly, as we tackle more challenging problems such as applications that require common sense reasoning, we will need to develop components that can acquire and integrate world knowledge from disparate sources into our models. Transfer learning will play an important role to bridge the digital language divide and to bring NLP capabilities to the rest of the world’s languages. As NLP systems become more mature and are applied to many more real-world—and often low-resource—scenarios, leveraging information from related domains, tasks, and languages will be key.

Conclusion 259 Language is a reflection of our context and experience. Training from a blank slate deprives our models of experience and the ability to interpret context. Ultimately, in order to come closer to the elusive goal of true natural language understanding, we need to equip our models with as much relevant knowledge and experience as possible.



# Bibliographie

- [Abu88] Deeb ABU FARA. “Control of Nozzle and Cavity Pressure during Filling and Packing in Thermoplastics Injection Molding”. English. Thèse de doct. Août 1988 (cf. p. 16).
- [AKP85] Deeb ABU FARA, M. R. KAMAL et W. I. PATTERSON. “Evaluation of Simple Dynamic Models and Controllers for Hydraulic and Nozzle Pressure in Injection Molding”. en. In : *Polymer Engineering & Science* 25.11 (août 1985), p. 714-723. ISSN : 1548-2634. DOI : [10.1002/pen.760251109](https://doi.org/10.1002/pen.760251109) (cf. p. 16).
- [AKP90] Deeb ABU FARA, M.R. KAMAL et W.I. PATTERSON. “Comprehensive Strategies for Sequential Closed Loop Pressure Control Throughout the Injection Molding Cycle”. English. In : *SPE ANTEC Papers*. T. 36. Society of Plastics Engineers, 1990, p. 239 (cf. p. 16).
- [Adr07] Pierre-Antoine ADRAGNA. “Tolérancement des Systèmes Assemblés, une approche par le Tolérancement Inertiel et Modal”. fr. Thèse de doct. Université de Savoie, déc. 2007 (cf. p. 63).
- [ASP10] Pierre-Antoine ADRAGNA, Serge SAMPER et Maurice PILLET. “A Proposition of 3D Inertial Tolerancing to Consider the Statistical Combination of the Location and Orientation Deviations”. In : *International Journal of Product Development* 10.1/2/3 (2010), p. 26. ISSN : 1477-9056, 1741-8178. DOI : [10.1504/IJPD.2010.029985](https://doi.org/10.1504/IJPD.2010.029985). arXiv : [1002.0253](https://arxiv.org/abs/1002.0253) (cf. p. 52).
- [Aga+13] Alban AGAZZI, Vincent SOBOTKA, Ronan LEGOFF et Yvon JARNY. “Optimal Cooling Design in Injection Moulding Process – A New Approach Based on Morphological Surfaces”. In : *Applied Thermal Engineering* 52.1 (avr. 2013), p. 170-178. ISSN : 1359-4311. DOI : [10.1016/j.applthermaleng.2012.11.019](https://doi.org/10.1016/j.applthermaleng.2012.11.019) (cf. p. 36).
- [AHK19] Tatyana AGEYEVA, Szabolcs HORVÁTH et József Gábor KOVÁCS. “In-Mold Sensors for Injection Molding : On the Way to Industry 4.0”. en. In : *Sensors* 19.16 (jan. 2019), p. 3551. DOI : [10/gf78qn](https://doi.org/10.3390/s19163551) (cf. p. 37).
- [APP87] A. R. AGRAWAL, I. O. PANDELIDIS et M. PECHT. “Injection-Molding Process Control—A Review”. en. In : *Polymer Engineering & Science* 27.18 (oct. 1987), p. 1345-1357. ISSN : 1548-2634. DOI : [10.1002/pen.760271802](https://doi.org/10.1002/pen.760271802) (cf. p. 12, 16).
- [ABR64] Mark Aronovich AÏZERMAN, É.M. BRAVERMAN et L. I. ROZONOÉR. “Theoretical Foundations of the Potential Function Method in Pattern Recognition Learning”. In : *Automation and Remote Control* 25.6 (fév. 1964), p. 821-837 (cf. p. 89).
- [Alb+16a] ALBERT, BRUNO, KNECHT, CHRISTOPHE, DE BERTRAND DE BEUVRON, FRANÇOIS, CHARRIER JULIEN, PILLET MAURICE, MAIRE JEAN-LUC et ZANNIMERK CECILIA. “Formalisation Du Contrôle Qualité Haptique : Structuration Sémantique Des Sensations Haptiques”. In : (2016). DOI : [10.13140/RG.2.2.15316.60806](https://doi.org/10.13140/RG.2.2.15316.60806) (cf. p. 28).
- [Alb19] Bruno ALBERT. “Maîtrise de la qualité haptique des produits”. fr. Thèse de doct. Université Grenoble Alpes, mar. 2019 (cf. p. 28, 84).

## Bibliographie

---

- [Alb+19] Bruno ALBERT, François de BERTRAND DE BEUVRON, Cecilia ZANNI-MERK, Jean-Luc MAIRE, Maurice PILLET, Julien CHARRIER et Christophe KNECHT. “A Smart System for Haptic Quality Control : A Knowledge-Based Approach to Formalize the Sense of Touch In”. In : *Knowledge Discovery, Knowledge Engineering and Knowledge Management. IC3K 2016. Communications in Computer and Information Science, Vol 914. Springer, Cham.* Fred A., Dietz J., Aveiro D., Liu K., Bernardino J., Filipe J. (eds), 2019, p. 173-190. DOI : [10.1007/978-3-319-99701-8\\_8](https://doi.org/10.1007/978-3-319-99701-8_8). HAL : [hal-01952452](https://hal.archives-ouvertes.fr/hal-01952452) (cf. p. 28).
- [Alb+16b] Bruno ALBERT, Jean-Luc. MAIRE, Maurice PILLET, Cecilia ZANNI-MERK, François DE BERTRAND DE BEUVRON, Christophe KNECHT et J. CHARRIER. “Generic and Structured Description of Tactile Sensory Perceptions”. In : *6th Kansei Engineering and Emotion Research*. Août 2016 (cf. p. 28).
- [Alb+17] Bruno ALBERT, Cecilia ZANNI-MERK, Francois de Bertrand DE BEUVRON, Maurice PILLET, Jean-Luc MAIRE, Christophe KNECHT et Julien CHARRIER. “A Smart System to Standardize the Specifications of Haptic Quality Control”. In : *Procedia Computer Science. Knowledge-Based and Intelligent Information & Engineering Systems : Proceedings of the 21st International Conference, KES-2017-8 September 2017, Marseille, France* 112 (jan. 2017), p. 723-730. ISSN : 1877-0509. DOI : [10/gf4j4k](https://doi.org/10/gf4j4k) (cf. p. 28, 84).
- [ACB17] Martin ARJOVSKY, Soumith CHINTALA et Léon BOTTOU. “Wasserstein GAN”. In : *arXiv :1701.07875 [cs, stat]* (jan. 2017). arXiv : [1701 . 07875 \[cs, stat\]](https://arxiv.org/abs/1701.07875) (cf. p. 128).
- [AV07] David ARTHUR et Sergei VASSILVITSKII. “K-Means++ : The Advantages of Careful Seeding”. In : *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '07. Philadelphia, PA, USA : Society for Industrial and Applied Mathematics, 2007, p. 1027-1035. ISBN : 978-0-89871-624-5 (cf. p. 113).
- [AE18] Gary A. ATKINSON et Jürgen D. ERNST. “High-Sensitivity Analysis of Polarization by Surface Reflection”. en. In : *Machine Vision and Applications* 29.7 (oct. 2018), p. 1171-1189. ISSN : 1432-1769. DOI : [10/gf8ztt](https://doi.org/10/gf8ztt) (cf. p. 52).
- [BR12] Lars BÅÄTH et Bengt-Göran ROSÉN. “Towards New Interferometer Technology for Surface Metrology”. eng. In : *12th International Conference of the European Society for Precision Engineering and Nanotechnology (EUSPEN 2012), June 4th-7th, 2012, Stockholm, Sweden*. T. 1. EUSPEN, 2012, p. 158-161 (cf. p. 38).
- [BCB14] Dzmitry BAHDANAU, Kyunghyun CHO et Yoshua BENGIO. “Neural Machine Translation by Jointly Learning to Align and Translate”. In : *arXiv :1409.0473 [cs, stat]* (sept. 2014). arXiv : [1409 . 0473 \[cs, stat\]](https://arxiv.org/abs/1409.0473) (cf. p. 128).
- [BC93] Pierre BALDI et Yves CHAUVIN. “Neural Networks for Fingerprint Recognition”. In : *Neural Computation* 5.3 (mai 1993), p. 402-418. ISSN : 0899-7667. DOI : [10 . 1162/neco.1993.5.3.402](https://doi.org/10.1162/neco.1993.5.3.402) (cf. p. 106).
- [Bal87] Dana H. BALLARD. “Modular Learning in Neural Networks”. In : *Proceedings of the Sixth National Conference on Artificial Intelligence - Volume 1*. AAAI'87. AAAI Press, 1987, p. 279-284. ISBN : 978-0-934613-42-2 (cf. p. 123).
- [BH86] Josh BARNES et Piet HUT. “A Hierarchical O(N Log N) Force-Calculation Algorithm”. en. In : *Nature* 324.6096 (déc. 1986), p. 446-449. ISSN : 1476-4687. DOI : [10/fvcc2m](https://doi.org/10/fvcc2m) (cf. p. 121).
- [Bau12] Nathalie BAUDET. “Maîtrise de la qualité visuelle des produits - Formalisation du processus d’expertise et proposition d’une approche robuste de contrôle visuel humain”. fr. Thèse de doct. Université de Grenoble, déc. 2012 (cf. p. 80, 84).

- [BPM11] Nathalie BAUDET, Maurice PILLET et Jean-Luc MAIRE. “Visual Inspection of Products : A Comparison of the Methods Used to Evaluate Surface Anomalies.” en. In : (juin 2011). HAL : [hal-00712918](https://hal.archives-ouvertes.fr/hal-00712918) (cf. p. 79).
- [Bea+15] Amy BEARMAN, Olga RUSSAKOVSKY, Vittorio FERRARI et Li FEI-FEI. “What’s the Point : Semantic Segmentation with Point Supervision”. In : *arXiv :1506.02106 [cs]* (juin 2015). arXiv : [1506.02106 \[cs\]](https://arxiv.org/abs/1506.02106) (cf. p. 79).
- [Bel12] Aurélien BELLET. “Supervised Metric Learning with Generalization Guarantees”. en. In : (déc. 2012) (cf. p. 84).
- [BSF94] Y. BENGIO, P. SIMARD et P. FRASCONI. “Learning long-term dependencies with gradient descent is difficult”. In : *IEEE Transactions on Neural Networks* 5.2 (mar. 1994), p. 157-166. ISSN : 1045-9227. DOI : [10.1109/72.279181](https://doi.org/10.1109/72.279181) (cf. p. 94).
- [Bér+04] Y. BÉREAUX, M. MOGUEDET, X. RAOUL, J. Y. CHARMEAU, J. BALCAEN et D. GRAEBLING. “Series Solutions for Viscous and Viscoelastic Fluids Flow in the Helical Rectangular Channel of an Extruder Screw”. In : *Journal of Non-Newtonian Fluid Mechanics* 123.2–3 (nov. 2004), p. 237-257. ISSN : 0377-0257. DOI : [10.1016/j.jnnfm.2004.08.011](https://doi.org/10.1016/j.jnnfm.2004.08.011) (cf. p. 13).
- [Ber+11] James S. BERGSTRA, Rémi BARDENET, Yoshua BENGIO et Balázs KÉGL. “Algorithms for Hyper-Parameter Optimization”. In : *Advances in Neural Information Processing Systems* 24. Sous la dir. de J. SHAWE-TAYLOR, R. S. ZEMEL, P. L. BARTLETT, F. PEREIRA et K. Q. WEINBERGER. NIPS’11. USA : Curran Associates, Inc., 2011, p. 2546-2554. ISBN : 978-1-61839-599-3 (cf. p. 130, 133, 134).
- [BB12] James BERGSTRA et Yoshua BENGIO. “Random Search for Hyper-Parameter Optimization”. In : *Journal of Machine Learning Research* 13.Feb (2012), p. 281-305. ISSN : ISSN 1533-7928 (cf. p. 130).
- [BYC13] James BERGSTRA, Daniel YAMINS et David COX. “Making a Science of Model Search : Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures”. en. In : *International Conference on Machine Learning*. Fév. 2013, p. 115-123 (cf. p. 133).
- [Ber91] Michael L. BERINS. *SPI Plastics Engineering Handbook of the Society of the Plastics Industry*. Kluwer Academic, 1991. ISBN : 978-1-4615-7606-8 (cf. p. 18).
- [Bet76] Guy BETHOUX. “Approche unitaire des méthodes d’identification et de commande adaptative des procédés dynamiques”. Français. Thèse 3ème cycle. Grenoble, France : Institut National Polytechnique de Grenoble, 1976 (cf. p. 17).
- [BM94] P. J. BLYSKAL et P. J. MEHERAN. “Applying Design Of Experiment Analysis Techniques to the Injection Moulding”. English. In : *SPE ANTEC PROCESS*. Chicago, USA : Society of Plastics Engineers, avr. 1994, p. 729-732 (cf. p. 13).
- [BGV92] Bernhard E. BOSER, Isabelle M. GUYON et Vladimir N. VAPNIK. “A Training Algorithm for Optimal Margin Classifiers”. In : *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. COLT ’92. New York, NY, USA : ACM, 1992, p. 144-152. ISBN : 978-0-89791-497-0. DOI : [10/bwgqr8](https://doi.org/10/bwgqr8) (cf. p. 89).
- [BPL10] Y-Lan BOUREAU, Jean PONCE et Yann LECUN. “A Theoretical Analysis of Feature Pooling in Visual Recognition”. In : *Proceedings of the 27th International Conference on International Conference on Machine Learning*. ICML’10. USA : Omnipress, 2010, p. 111-118. ISBN : 978-1-60558-907-7 (cf. p. 100).
- [BK88] H. BOURLARD et Y. KAMP. “Auto-Association by Multilayer Perceptrons and Singular Value Decomposition”. en. In : *Biological Cybernetics* 59.4 (sept. 1988), p. 291-294. ISSN : 1432-0770. DOI : [10/fm3xqd](https://doi.org/10/fm3xqd) (cf. p. 123).

- [BE02] A. Tolga BOZDANA et Ömer EYERCIOĞLU. "Development of an Expert System for the Determination of Injection Moulding Parameters of Thermoplastic Materials : EX-PIMM". In : *Journal of Materials Processing Technology* 128.1–3 (oct. 2002), p. 113-122. ISSN : 0924-0136. DOI : [10.1016/S0924-0136\(02\)00436-3](https://doi.org/10.1016/S0924-0136(02)00436-3) (cf. p. 15).
- [Bra00] Gary BRADSKI. "The OpenCV Library". In : *Dr. Dobb's Journal of Software Tools* 25.11 (2000), p. 120-125 (cf. p. 55).
- [Bre96a] Leo BREIMAN. "Bagging Predictors". en. In : *Machine Learning* 24.2 (août 1996), p. 123-140. ISSN : 0885-6125, 1573-0565. DOI : [10.1023/A:1018054314350](https://doi.org/10.1023/A:1018054314350) (cf. p. 90).
- [Bre96b] Leo BREIMAN. *Bias, Variance, and Arcing Classifiers*. Rapp. tech. 1996 (cf. p. 91).
- [Bre97] Leo BREIMAN. "Arcing the Edge". In : 1997 (cf. p. 91).
- [Bre01] Leo BREIMAN. "Random Forests". en. In : *Machine Learning* 45.1 (oct. 2001), p. 5-32. ISSN : 0885-6125, 1573-0565. DOI : [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324) (cf. p. 91).
- [Bre33] David BREWSTER. "Experiments on the Depolarization of Light as Exhibited by Various Mineral, Animal, and Vegetable Bodies, with a Reference of the Phenomena to the General Principles of Polarization. By David Brewster, LL. D. F. R. S. Edin. and F. S. A. In a Letter Addressed to the Right Hon. Sir Joseph Banks, Bart. K. B. P. R. S". In : *Abstracts of the Papers Printed in the Philosophical Transactions of the Royal Society of London* 2 (jan. 1833), p. 4-6. DOI : [10.1098/rspl.1815.0003](https://doi.org/10.1098/rspl.1815.0003) (cf. p. 52).
- [Bro+10] K. H. BRODERSEN, C. S. ONG, K. E. STEPHAN et J. M. BUHMANN. "The Balanced Accuracy and Its Posterior Distribution". In : *2010 20th International Conference on Pattern Recognition*. Août 2010, p. 3121-3124. DOI : [10.1109/ICPR.2010.764](https://doi.org/10.1109/ICPR.2010.764) (cf. p. 85).
- [Bro+93] Jane BROMLEY, Isabelle GUYON, Yann LECUN, Eduard SÄCKINGER et Roopak SHAH. "Signature Verification Using a "Siamese" Time Delay Neural Network". In : *Proceedings of the 6th International Conference on Neural Information Processing Systems*. NIPS'93. event-place : Denver, Colorado. San Francisco, CA, USA : Morgan Kaufmann Publishers Inc., 1993, p. 737-744 (cf. p. 106).
- [CMS13] Ricardo J. G. B. CAMPELLO, Davoud MOULAVI et Joerg SANDER. "Density-Based Clustering Based on Hierarchical Density Estimates". en. In : *Advances in Knowledge Discovery and Data Mining*. Sous la dir. de Jian PEI, Vincent S. TSENG, Longbing CAO, Hiroshi MOTODA et Guandong XU. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2013, p. 160-172. ISBN : 978-3-642-37456-2 (cf. p. 113).
- [Cas+19] Thomas A CASWELL, Michael DROETTBOOM, John HUNTER, Eric FIRING, Antony LEE, Jody KLYMAK, David STANSBY, Elliott Sales De ANDRADE, Jens Hedegaard NIELSEN, Nelle VAROQUAUX, Benjamin ROOT, Tim HOFFMANN, Phil ELSON, Ryan MAY, Darren DALE, JAE-JOON LEE, Jouni K. SEPPÄNEN, Damon McDougall, Andrew STRAW, Paul HOBSON, Christoph GOHLKE, Tony S YU, Eric MA, Adrien F. VINCENT, Steven SILVESTER, Charlie MOAD, Jan KATINS, Nikita KNIAZEV, Federico ARIZA et Elan ERNEST. *Matplotlib/Matplotlib v3.1.0*. Zenodo. Mai 2019. DOI : [10.5281/zenodo.2893252](https://doi.org/10.5281/zenodo.2893252) (cf. p. 46).
- [CL11] Chih-Chung CHANG et Chih-Jen LIN. "LIBSVM : A Library for Support Vector Machines". In : *ACM Trans. Intell. Syst. Technol.* 2.3 (mai 2011), 27 :1-27 :27. ISSN : 2157-6904. DOI : [10.1145/1961189.1961199](https://doi.org/10.1145/1961189.1961199) (cf. p. 89).

- [CSZ10] Olivier CHAPELLE, Bernhard SCHÖLKOPF et Alexander ZIEN, éd. *Semi-supervised learning*. eng. Adaptive computation and machine learning series. OCLC : 845655904. Cambridge, Mass. [u.a] : MIT Press, 2010. ISBN : 978-0-262-03358-9 978-0-262-51412-5 (cf. p. 104).
- [CPS06] Kumar CHELLAPILLA, Sidd PURI et Patrice SIMARD. “High Performance Convolutional Neural Networks for Document Processing”. en. In : (oct. 2006) (cf. p. 93).
- [Che+02] Shia-Chung CHEN, Wei-Lianq LIAW, Hsin-Shu PENG, Lei-Ti HUANG et Yi-Fong CHANG. “Characteristics of Mould Separation and Its Effects on Qualities of Thin Wall Injection Moulded Parts”. In : *Plastics, Rubber and Composites* 31.1 (jan. 2002), p. 12-19. ISSN : 1465-8011. DOI : [10.1179/146580101125000457](https://doi.org/10.1179/146580101125000457) (cf. p. 14).
- [CSW91] Cheng-Ping CHIU, Laming-Chang SHIH et Jong-Hwei WEI. “Dynamic Modeling of the Mold Filling Process in an Injection Molding Machine”. en. In : *Polymer Engineering & Science* 31.19 (oct. 1991), p. 1417-1425. ISSN : 1548-2634. DOI : [10.1002/pen.760311908](https://doi.org/10.1002/pen.760311908) (cf. p. 13).
- [Cho16] François CHOLLET. “Xception : Deep Learning with Depthwise Separable Convolutions”. In : *arXiv :1610.02357 [cs]* (oct. 2016). arXiv : [1610.02357 \[cs\]](https://arxiv.org/abs/1610.02357) (cf. p. 97).
- [Cir+11] Dan C. CIREŞAN, Ueli MEIER, Jonathan MASCI, Luca M. GAMBARDELLA et Jürgen SCHMIDHUBER. “Flexible, High Performance Convolutional Neural Networks for Image Classification”. In : *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two*. IJCAI’11. event-place : Barcelona, Catalonia, Spain. AAAI Press, 2011, p. 1237-1242. ISBN : 978-1-57735-514-4. DOI : [10.5591/978-1-57735-516-8/IJCAI11-210](https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-210) (cf. p. 93).
- [CUH15] Djork-Arné CLEVERT, Thomas UNTERTHINER et Sepp HOCHREITER. “Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)”. In : *arXiv :1511.07289 [cs]* (nov. 2015). arXiv : [1511.07289 \[cs\]](https://arxiv.org/abs/1511.07289) (cf. p. 98).
- [Col18] Jean COLLOMB. “Optimisation du processus de dimensionnement thermomécanique de Moule Autonome à Transfert Thermique Efficient pour la transformation rapide des matériaux composites à renforts continus”. Thèse de doct. Communauté Université Grenoble Alpes, 2018, p. 214. HAL : [tel-02022616](https://tel-02022616) (cf. p. 76).
- [CV95] Corinna CORTES et Vladimir VAPNIK. “Support-Vector Networks”. en. In : *Machine Learning* 20.3 (sept. 1995), p. 273-297. ISSN : 1573-0565. DOI : [10/cv7fn6](https://doi.org/10/cv7fn6) (cf. p. 89).
- [CH67] T. COVER et P. HART. “Nearest Neighbor Pattern Classification”. In : *IEEE Transactions on Information Theory* 13.1 (jan. 1967), p. 21-27. ISSN : 0018-9448. DOI : [10/dn5msp](https://doi.org/10/dn5msp) (cf. p. 90).
- [Deb+00] Paul DEBEVEC, Tim HAWKINS, Chris TCHOU, Haarm-Pieter DUIKER, Westley SAROKIN et Mark SAGAR. “Acquiring the Reflectance Field of a Human Face”. In : *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH ’00. New York, NY, USA : ACM Press/Addison-Wesley Publishing Co., 2000, p. 145-156. ISBN : 978-1-58113-208-3. DOI : [10/d329r5](https://doi.org/10/d329r5) (cf. p. 53, 54).

## Bibliographie

---

- [Del06] E. DEL CASTILLO. “Statistical Process Adjustment : A Brief Retrospective, Current Status, and Some Opportunities for Further Work”. en. In : *Statistica Neerlandica* 60.3 (août 2006), p. 309-326. ISSN : 1467-9574. DOI : [10.1111/j.1467-9574.2006.00328.x](https://doi.org/10.1111/j.1467-9574.2006.00328.x) (cf. p. 20).
- [Del+00] D. DELAUNAY, P. LE BOT, R. FULCHIRON, J. F. LUYE et G. REGNIER. “Nature of Contact between Polymer and Mold in Injection Molding. Part II : Influence of Mold Deflection on Pressure History and Shrinkage”. en. In : *Polymer Engineering & Science* 40.7 (juil. 2000), p. 1692-1700. ISSN : 1548-2634. DOI : [10.1002/pen.11301](https://doi.org/10.1002/pen.11301) (cf. p. 13).
- [Dem82] William Edwards DEMING. *Quality, Productivity, and Competitive Position*. English. 04180. Cambridge, MA : Massachusetts Institute of Technology, Center for Advanced Engineering Study, 1982 (cf. p. 32).
- [DCG97] H. H. DEMIRCI, John P. COULTER et S. I. GÜÇERİ. “A Numerical and Experimental Investigation of Neural Network-Based Intelligent Control of Molding Processes”. In : *Journal of Manufacturing Science and Engineering* 119.1 (fév. 1997), p. 88-94. ISSN : 1087-1357. DOI : [10.1115/1.2836559](https://doi.org/10.1115/1.2836559) (cf. p. 17).
- [Den+09] J. DENG, W. DONG, R. SOCHER, L. LI, Kai LI et Li FEI-FEI. “ImageNet : A large-scale hierarchical image database”. In : *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Juin 2009, p. 248-255. DOI : [10.1109/CVPR.2009.5206848](https://doi.org/10.1109/CVPR.2009.5206848) (cf. p. 93, 103).
- [DVA95] O. DENIZART, M. VINCENT et J. F. AGASSANT. “Thermal Stresses and Strains in Injection Moulding : Experiments and Computations”. en. In : *Journal of Materials Science* 30.2 (jan. 1995), p. 552-560. ISSN : 1573-4803. DOI : [10.1007/BF00354424](https://doi.org/10.1007/BF00354424) (cf. p. 52).
- [Des15] Simon-Frédéric DESAGE. “Constraints and Opportunities for Automation of Visual Inspection with Regard to the Human Process”. PhD Thesis. Université Grenoble Alpes, nov. 2015 (cf. p. 29).
- [Dés15] Simon-Frédéric DÉSAGE. “Contraintes et opportunités pour l’automatisation de l’inspection visuelle au regard du processus humain”. fr. In : (nov. 2015) (cf. p. 84).
- [Dés+14] Simon-Frédéric DÉSAGE, Gilles PITARD, Maurice PILLET, Hugues FAVRELIÈRE, Fabrice FRELIN, Serge SAMPER et Gaëtan LE GOÏC. “Visual Quality Inspection and Fine Anomalies : Methods and Application”. In : *IFIP Advances in Information and Communication Technology* 435 (nov. 2014), p. 94-106 (cf. p. 69).
- [Des+15] Simon-Frédéric DESAGE, Gilles PITARD, Maurice PILLET, Hugues FAVRELIÈRE, Jean Luc MAIRE, Fabrice FRELIN, Serge SAMPER et Gaëtan LE GOÏC. “Syntactic Texture and Perception for a New Generic Visual Anomalies Classification.” In : *The International Conference on Quality Control by Artificial Vision 2015*. T. 9534. Le Creusot, France, juin 2015. HAL : [hal-01246517](https://hal-01246517) (cf. p. 29).
- [Dev90] Pierre DEVOS. “Contribution à l’optimisation de la conduite d’une presse à injecter les polymères thermoplastiques”. Français. Doctorat. Université des Sciences et Techniques de Lille - Flandres - Artois : Université des Sciences et Techniques de Lille - Flandres - Artois, juil. 1990 (cf. p. 17).
- [Dir19] DIRECTION GÉNÉRALE DES ENTREPRISES. *Chiffres clés de l’industrie manufacturière, Édition 2018*. Français. Étude économique ISSN : 2492-7090. Ivry-sur-Seine, France : Ministère de l’Économie et des Finances, 2019, p. 8 (cf. p. 25, 26).
- [Doe16] Carl DOERSCH. “Tutorial on Variational Autoencoders”. In : *arXiv :1606.05908 [cs, stat]* (juin 2016). arXiv : [1606.05908 \[cs, stat\]](https://arxiv.org/abs/1606.05908) (cf. p. 124, 125).

- [DSH15] Tobias DOMHAN, Jost Tobias SPRINGENBERG et Frank HUTTER. “Speeding Up Automatic Hyperparameter Optimization of Deep Neural Networks by Extrapolation of Learning Curves”. In : *Proceedings of the 24th International Conference on Artificial Intelligence*. IJCAI’15. AAAI Press, 2015, p. 3460-3468. ISBN : 978-1-57735-738-4 (cf. p. 135).
- [DML11] Wei DONG, Charikar MOSES et Kai LI. “Efficient K-Nearest Neighbor Graph Construction for Generic Similarity Measures”. In : *Proceedings of the 20th International Conference on World Wide Web*. WWW ’11. New York, NY, USA : ACM, 2011, p. 577-586. ISBN : 978-1-4503-0632-4. DOI : [10/dh93jt](https://doi.org/10/dh93jt) (cf. p. 123).
- [Dua+18] Yueqi DUAN, Wenzhao ZHENG, Xudong LIN, Jiwen LU et Jie ZHOU. “Deep Adversarial Metric Learning”. In : 2018, p. 2780-2789 (cf. p. 110).
- [DV16] Vincent DUMOULIN et Francesco VISIN. “A Guide to Convolution Arithmetic for Deep Learning”. In : *arXiv :1603.07285 [cs, stat]* (mar. 2016). arXiv : [1603.07285 \[cs, stat\]](https://arxiv.org/abs/1603.07285) (cf. p. 95).
- [Dur17] Thibaut DURAND. “Weakly supervised learning for visual recognition”. en. Thèse de doct. Université Pierre et Marie Curie, sept. 2017 (cf. p. 104).
- [Ega79] Bo EGARDT. *Stability of Adaptive Controllers*. Sous la dir. de M. THOMA et A. V. BALAKRISHNAN. Berlin, Heidelberg : Springer-Verlag, 1979. ISBN : 978-0-387-09646-9 (cf. p. 17).
- [EMH18] Thomas ELSKEN, Jan Hendrik METZEN et Frank HUTTER. “Neural Architecture Search : A Survey”. In : *arXiv :1808.05377 [cs, stat]* (août 2018). arXiv : [1808 . 05377 \[cs, stat\]](https://arxiv.org/abs/1808.05377) (cf. p. 137).
- [Ere+09] Gonen EREN, Olivier AUBRETON, Fabrice MERIAUDEAU, L. A. Sanchez SECADES, David FOFI, A. Teoman NASKALI, Frederic TRUCHETET et Aytul ERCIL. “Scanning from Heating : 3D Shape Estimation of Transparent Objects from Local Surface Heating”. EN. In : *Optics Express* 17.14 (juil. 2009), p. 11457-11468. ISSN : 1094-4087. DOI : [10.1364/OE.17.011457](https://doi.org/10.1364/OE.17.011457) (cf. p. 47).
- [Est+96] Martin ESTER, Hans-Peter KRIEGEL, Jörg SANDER et Xiaowei XU. “A Density-Based Algorithm for Discovering Clusters a Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”. In : *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. KDD’96. AAAI Press, 1996, p. 226-231 (cf. p. 113).
- [FKH17] Stefan FALKNER, Aaron KLEIN et Frank HUTTER. “Combining Hyperband and Bayesian Optimization”. In : *NIPS 2017 Bayesian Optimization Workshop (Dec 2017)*. Long Beach, USA, déc. 2017 (cf. p. 136).
- [FKH18] Stefan FALKNER, Aaron KLEIN et Frank HUTTER. “BOHB : Robust and Efficient Hyperparameter Optimization at Scale”. In : *arXiv :1807.01774 [cs, stat]* (juil. 2018). arXiv : [1807.01774 \[cs, stat\]](https://arxiv.org/abs/1807.01774) (cf. p. 136).
- [Fan+08] Rong-En FAN, Kai-Wei CHANG, Cho-Jui HSIEH, Xiang-Rui WANG et Chih-Jen LIN. “LIBLINEAR : A Library for Large Linear Classification”. In : *J. Mach. Learn. Res.* 9 (juin 2008), p. 1871-1874. ISSN : 1532-4435 (cf. p. 89).
- [FC19] Riccardo FANTINEL et Angelo CENEDESE. “Visual Inspection for Metallic Surfaces : CNN Driven by Features”. In : *Fourteenth International Conference on Quality Control by Artificial Vision*. T. 11172. International Society for Optics and Photonics, juil. 2019, 111720H. DOI : [10/gf8qgt](https://doi.org/10/gf8qgt) (cf. p. 144).
- [Fav09] Hugues FAVRELIERE. “Modal Tolerancing : From Metrology to Specifications”. In : (oct. 2009) (cf. p. 63).

- [FMK13] Ng Chin FEI, Nik Mizamzul MEHAT et Shahrul KAMARUDDIN. “Practical Applications of Taguchi Method for Optimization of Processing Parameters for Plastic Injection Moulding : A Retrospective Review”. en. In : *International Scholarly Research Notices* 2013 (juin 2013), e462174. ISSN : , DOI : [10.1155/2013/462174](https://doi.org/10.1155/2013/462174) (cf. p. 13).
- [Feu+15] Matthias FEURER, Aaron KLEIN, Katharina EGGENSPERGER, Jost SPRINGENBERG, Manuel BLUM et Frank HUTTER. “Efficient and Robust Automated Machine Learning”. In : *Advances in Neural Information Processing Systems* 28. Sous la dir. de C. CORTES, N. D. LAWRENCE, D. D. LEE, M. SUGIYAMA et R. GARNETT. Curran Associates, Inc., 2015, p. 2962-2970 (cf. p. 133, 134).
- [FSH15] Matthias FEURER, Jost Tobias SPRINGENBERG et Frank HUTTER. “Initializing Bayesian Hyperparameter Optimization via Meta-Learning”. en. In : *Twenty-Ninth AAAI Conference on Artificial Intelligence*. Fév. 2015 (cf. p. 133).
- [Fis74] Ronald Aylmer FISHER. *The Design of Experiments*. English. OCLC : 924773451. New York, 1974. ISBN : 978-0-02-844690-5 (cf. p. 131).
- [FH51] Evelyn FIX et Jr HODGES. *Discriminatory Analysis - Nonparametric Discrimination : Consistency Properties*. en. Rapp. tech. CALIFORNIA UNIV BERKELEY, fév. 1951 (cf. p. 90).
- [Fou+06] Jean-Étienne FOURNIER, Nicolas HAVARD, Marie-France LACRAMPE, Marc RYCKEBUSCH et Patricia KRAWCZAK. “Conduite Adaptative Du Procédé d’injection Des Thermoplastiques”. In : *Techniques de l’ingénieur Procédés d’injection des thermoplastiques* base documentaire : TIB151DUO.ref. article : in53 (avr. 2006) (cf. p. 21).
- [Fri01] Jerome H. FRIEDMAN. “Greedy Function Approximation : A Gradient Boosting Machine”. In : *The Annals of Statistics* 29.5 (oct. 2001), p. 1189-1232. ISSN : 0090-5364, 2168-8966. DOI : [10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451) (cf. p. 91).
- [Fri02] Jerome H. FRIEDMAN. “Stochastic Gradient Boosting”. In : *Computational Statistics & Data Analysis*. Nonlinear Methods and Data Mining 38.4 (fév. 2002), p. 367-378. ISSN : 0167-9473. DOI : [10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2) (cf. p. 91).
- [Fuk80] Kunihiko FUKUSHIMA. “Neocognitron : A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position”. en. In : *Biological Cybernetics* 36.4 (avr. 1980), p. 193-202. ISSN : 1432-0770. DOI : [10/fhrjjv](https://doi.org/10/fhrjjv) (cf. p. 92).
- [GR09] A. GALESKI et G. REGNIER. “Nano- and Micromechanics of Crystalline Polymers”. In : *Nano- and Micromechanics of Polymer Blends and Composites*. Carl Hanser Verlag GmbH & Co. KG, mai 2009, p. 1-58. ISBN : 978-3-446-41323-8 (cf. p. 12, 42).
- [Gal+87] P. GALLINARI, Yann LECUN, S. THIRIA et F. Fogelman SOULIE. “Mémoires associatives distribuées : Une comparaison (Distributed associative memories : A comparison)”. English (US). In : *Proceedings of COGNITIVA 87, Paris, La Villette, May 1987* (1987) (cf. p. 123).
- [GK12] Robert X. GAO et David Owen KAZMER. “Multivariate Sensing and Wireless Data Communication for Process Monitoring in RF-Shielded Environment”. In : *CIRP Annals - Manufacturing Technology* 61.1 (2012), p. 523-526. ISSN : 0007-8506. DOI : [10.1016/j.cirp.2012.03.014](https://doi.org/10.1016/j.cirp.2012.03.014) (cf. p. 28).

- [GTW08] Yuehua GAO, Lih-Sheng TURNG et Xicheng WANG. “Adaptive Geometry and Process Optimization for Injection Molding Using the Kriging Surrogate Model Trained by Numerical Simulation”. en. In : *Advances in Polymer Technology* 27.1 (mar. 2008), p. 1-16. ISSN : 1098-2329. DOI : [10.1002/adv.20116](https://doi.org/10.1002/adv.20116) (cf. p. 13).
- [Gar+14] S. GARRIDO-JURADO, R. MUÑOZ-SALINAS, F. J. MADRID-CUEVAS et M. J. MARÍN-JIMÉNEZ. “Automatic Generation and Detection of Highly Reliable Fiducial Markers under Occlusion”. In : *Pattern Recognition* 47.6 (juin 2014), p. 2280-2292. ISSN : 0031-3203. DOI : [10.1016/j.patcog.2014.01.005](https://doi.org/10.1016/j.patcog.2014.01.005) (cf. p. 55).
- [Gar+16] S. GARRIDO-JURADO, R. MUÑOZ-SALINAS, F. J. MADRID-CUEVAS et R. MEDINA-CARNICER. “Generation of Fiducial Marker Dictionaries Using Mixed Integer Linear Programming”. In : *Pattern Recognition* 51 (mar. 2016), p. 481-491. ISSN : 0031-3203. DOI : [10.1016/j.patcog.2015.09.023](https://doi.org/10.1016/j.patcog.2015.09.023) (cf. p. 55).
- [Gen+18] L. GENDRE, S. BAZEILLE, L. BIGUÉ et C. CUDEL. “Interest of Polarimetric Refocused Images Calibrated in Depth for Control by Vision”. In : *Unconventional Optical Imaging*. T. 10677. International Society for Optics and Photonics, mai 2018, 106771W. DOI : [10.1117/12.2306939](https://doi.org/10.1117/12.2306939) (cf. p. 56).
- [GFB11] Luc GENDRE, Alban FOULONNEAU et Laurent BIGUE. “Full Stokes Polarimetric Imaging Using a Single Ferroelectric Liquid Crystal Device”. In : *Optical Engineering* 50.8 (2011), p. 1-9. DOI : [10.fg4bfk](https://doi.org/10.fg4bfk) (cf. p. 54).
- [Gho+07] Abhijeet GHOSH, Shruthi ACHUTHA, Wolfgang HEIDRICH et Matthew O’TOOLE. “BRDF Acquisition with Basis Illumination”. In : *2007 IEEE 11th International Conference on Computer Vision*. Oct. 2007, p. 1-8. DOI : [10.1109/ICCV.2007.4408935](https://doi.org/10.1109/ICCV.2007.4408935) (cf. p. 69).
- [Gij+19] Pieter GIJSBERS, Erin LEDELL, Janek THOMAS, Sébastien POIRIER, Bernd BISCHL et Joaquin VANSCHOREN. “An Open Source AutoML Benchmark”. In : *arXiv :1907.00909 [cs, stat]* (juil. 2019). arXiv : [1907 . 00909 \[cs, stat\]](https://arxiv.org/abs/1907.00909) (cf. p. 138).
- [Gir01] Thomas GIROUD. “Mesure et calcul des contraintes résiduelles dans les pièces injectées en thermoplastiques avec et sans fibres de renfort”. fr. Thèse de doct. École Nationale Supérieure des Mines de Paris, déc. 2001 (cf. p. 20, 38).
- [Goï+16] Gaëtan Le GOÏC, Maxence BIGERELLE, Serge SAMPER, Hugues FAVRELIÈRE et Maurice PILLET. “Multiscale Roughness Analysis of Engineering Surfaces : A Comparison of Methods for the Investigation of Functional Correlations”. In : *Mechanical Systems and Signal Processing* 66-67. Supplement C (jan. 2016). 00009, p. 437-457. ISSN : 0888-3270. DOI : [10.1016/j.ymssp.2015.05.029](https://doi.org/10.1016/j.ymssp.2015.05.029) (cf. p. 63).
- [Goï+ct] Gaëtan Le Goïc, Hugues Favrelière, Serge Samper et Fabien Formosa. “Multi Scale Modal Decomposition of Primary Form, Waviness and Roughness of Surfaces”. eng. In : *Scanning* 33.5 (2011 Sep-Oct), p. 332-341. ISSN : 1932-8745. DOI : [10.1002/sca.20253](https://doi.org/10.1002/sca.20253) (cf. p. 63).
- [GPK86] V. G. GOMES, W. I. PATTERSON et M. K. KAMAL. “An Injection Molding Study. Part II : Evaluation of Alternative Control Strategies for Melt Temperature”. en. In : *Polymer Engineering & Science* 26.12 (juil. 1986), p. 867-876. ISSN : 1548-2634. DOI : [10.1002/pen.760261206](https://doi.org/10.1002/pen.760261206) (cf. p. 16).
- [Goo+14] Ian J. GOODFELLOW, Jean POUGET-ABADIE, Mehdi MIRZA, Bing XU, David WARDE-FARLEY, Sherjil OZAIR, Aaron COURVILLE et Yoshua BENGIO. “Generative Adversarial Networks”. In : *arXiv :1406.2661 [cs, stat]* (juin 2014). arXiv : [1406 . 2661 \[cs, stat\]](https://arxiv.org/abs/1406.2661) (cf. p. 127).

- [GBC16] Ian GOODFELLOW, Yoshua BENGIO et Aaron COURVILLE. *Deep Learning*. eng. Adaptive Computation and Machine Learning. Cambridge, Massachusetts London, England : The MIT Press, 2016. ISBN : 978-0-262-03561-3 (cf. p. [92](#), [98](#)).
- [Got71] Marcel GOTLIB. *Rubrique-à-brac, Taume 2*. fre. Dargaud. T. Taume 2. Rubrique-à-brac. Paris : Dargaud, jan. 1971. ISBN : 978-2-205-00554-7 (cf. p. [iii](#)).
- [Got+19] Keisuke GOTO, Kunihito KATO, Shunsuke NAKATSUKA, Takaho SAITO et Hiroaki AIZAWA. “Anomaly Detection of Solder Joint on Print Circuit Board by Using Adversarial Autoencoder”. In : *Fourteenth International Conference on Quality Control by Artificial Vision*. T. 11172. International Society for Optics and Photonics, juil. 2019, 111720T. DOI : [10/gf8v9j](https://doi.org/10/gf8v9j) (cf. p. [144](#)).
- [Gre+17] Klaus GREFF, Aaron KLEIN, Martin CHOVAEC, Frank HUTTER et Jürgen SCHMIDHUBER. “The Sacred Infrastructure for Computational Research”. In : *Proceedings of the 15th Python in Science Conference*. 2017, p. 49-56 (cf. p. [130](#)).
- [Gro+10] Automotive Industry Action GROUP, CHRYSLER, FORD et General MOTORS. *Measurement Systems Analysis (MSA) Reference Manual*. eng. 4<sup>e</sup> éd. OCLC : 729994003. Southfield, Mich : Automotive Industry Action Group, 2010. ISBN : 978-1-60534-211-5 (cf. p. [80](#)).
- [Gul+17] Ishaan GULRAJANI, Faruk AHMED, Martin ARJOVSKY, Vincent DUMOULIN et Aaron COURVILLE. “Improved Training of Wasserstein GANs”. In : *arXiv :1704.00028 [cs, stat]* (mar. 2017). arXiv : [1704 . 00028 \[cs, stat\]](https://arxiv.org/abs/1704.00028) (cf. p. [128](#)).
- [GL87] Donald GUSTAFSON et William LEBOW. “Model Predictive Control (MPC) of Injection Molding Machines”. In : *26th IEEE Conference on Decision and Control*. IEEE, déc. 1987, p. 2017-2026. DOI : [10 . 1109/CDC . 1987 . 272888](https://doi.org/10.1109/CDC.1987.272888) (cf. p. [16](#)).
- [HCL06] R. HADSELL, S. CHOPRA et Y. LECUN. “Dimensionality Reduction by Learning an Invariant Mapping”. In : *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)*. T. 2. New York, NY, USA : IEEE, 2006, p. 1735-1742. ISBN : 978-0-7695-2597-6. DOI : [10 . 1109/CVPR . 2006 . 100](https://doi.org/10.1109/CVPR.2006.100) (cf. p. [106](#)).
- [HW93] J. HAEUSSLER et J. WORTBERG. “Quality Assurance in Injection Molding with Neural Networks”. English. In : *Conference Proceedings*. T. 1. NEW ORLEANS : Society of Plastics Engineers, mai 1993, p. 123-129 (cf. p. [21](#)).
- [HK93] Pat HANRAHAN et Wolfgang KRÜGER. “Reflection from Layered Surfaces Due to Subsurface Scattering”. In : *SIGGRAPH*. 1993. DOI : [10/b4tw3j](https://doi.org/10.b4tw3j) (cf. p. [52](#)).
- [HSD73] R. M. HARALICK, K. SHANMUGAM et I. DINSTEIN. “Textural Features for Image Classification”. In : *IEEE Transactions on Systems, Man, and Cybernetics SMC-3*.6 (nov. 1973), p. 610-621. ISSN : 0018-9472. DOI : [10 . 1109/TSMC . 1973 . 4309314](https://doi.org/10.1109/TSMC.1973.4309314) (cf. p. [61](#), [62](#), [65](#)).
- [He+15a] Kaiming HE, Xiangyu ZHANG, Shaoqing REN et Jian SUN. “Deep Residual Learning for Image Recognition”. In : *arXiv :1512.03385 [cs]* (déc. 2015). arXiv : [1512 . 03385](https://arxiv.org/abs/1512.03385) (cf. p. [93](#), [96](#)).
- [He+15b] Kaiming HE, Xiangyu ZHANG, Shaoqing REN et Jian SUN. “Delving Deep into Rectifiers : Surpassing Human-Level Performance on ImageNet Classification”. In : *arXiv :1502.01852 [cs]* (fév. 2015). arXiv : [1502 . 01852 \[cs\]](https://arxiv.org/abs/1502.01852) (cf. p. [98](#)).
- [He+16] Kaiming HE, Xiangyu ZHANG, Shaoqing REN et Jian SUN. “Identity Mappings in Deep Residual Networks”. In : *arXiv :1603.05027 [cs]* (mar. 2016). arXiv : [1603 . 05027 \[cs\]](https://arxiv.org/abs/1603.05027) (cf. p. [96](#)).

- [HBL17] Alexander HERMANS, Lucas BEYER et Bastian LEIBE. “In Defense of the Triplet Loss for Person Re-Identification”. In : *arXiv :1703.07737 [cs]* (mar. 2017). arXiv : [1703.07737](https://arxiv.org/abs/1703.07737) (cf. p. 110).
- [HMA19] Thomas HERRMANN, Cyrille MIGNOT et Olivier AUBRETON. “Cracks Detection on Glass Object Based on Active Thermography Approach”. In : *Fourteenth International Conference on Quality Control by Artificial Vision*. T. 11172. International Society for Optics and Photonics, juil. 2019, 111721E. DOI : [10 / gf8qfn](https://doi.org/10.1117/12.311721) (cf. p. 47).
- [HR03] Geoffrey E HINTON et Sam T. ROWEIS. “Stochastic Neighbor Embedding”. In : *Advances in Neural Information Processing Systems 15*. Sous la dir. de S. BECKER, S. THRUN et K. OBERMAYER. Vancouver, British Columbia, Canada : MIT Press, 2003, p. 857-864 (cf. p. 120).
- [Ho95] Tin Kam HO. “Random Decision Forests”. In : *Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1) - Volume 1*. ICDAR ’95. Washington, DC, USA : IEEE Computer Society, 1995, p. 278-. ISBN : 978-0-8186-7128-9 (cf. p. 91).
- [Ho98] Tin Kam HO. “The Random Subspace Method for Constructing Decision Forests”. In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20.8 (août 1998), p. 832-844. ISSN : 0162-8828. DOI : [10 / bn3jhv](https://doi.org/10.1109/TPAMI.1998.67269) (cf. p. 91).
- [Hoc+01] Sepp HOCHREITER, Yoshua BENGIO, Paolo FRASCONI et Jürgen SCHMIDHUBER. *Gradient Flow in Recurrent Nets : the Difficulty of Learning Long-Term Dependencies*. 2001 (cf. p. 94).
- [Hot33] Harold HOTELLING. “Analysis of a Complex of Statistical Variables into Principal Components”. In : *Journal of Educational Psychology* 24.6 (1933), p. 417-441. ISSN : 1939-2176(Electronic),0022-0663(Print). DOI : [10 / fb5435](https://doi.org/10.1037/0022-0663.24.6.417) (cf. p. 19, 114).
- [How+17] Andrew G. HOWARD, Menglong ZHU, Bo CHEN, Dmitry KALENICHENKO, Weijun WANG, Tobias WEYAND, Marco ANDREETTO et Hartwig ADAM. “MobileNets : Efficient Convolutional Neural Networks for Mobile Vision Applications”. In : *arXiv :1704.04861 [cs]* (avr. 2017). arXiv : [1704.04861 \[cs\]](https://arxiv.org/abs/1704.04861) (cf. p. 97).
- [Hua+16] Gao HUANG, Zhuang LIU, Laurens van der MAATEN et Kilian Q. WEINBERGER. “Densely Connected Convolutional Networks”. In : *arXiv :1608.06993 [cs]* (août 2016). arXiv : [1608.06993](https://arxiv.org/abs/1608.06993) (cf. p. 96, 105).
- [HL00] S.-J. HUANG et T.-H. LEE. “Fuzzy Logic Controller for a Retrofitted Closed-Loop Injection Moulding Machine”. en. In : *Proceedings of the Institution of Mechanical Engineers, Part I : Journal of Systems and Control Engineering* 214.1 (fév. 2000), p. 9-22. ISSN : 0959-6518, 2041-3041. DOI : [10 . 1243 / 0959651001540483](https://doi.org/10.1243/0959651001540483) (cf. p. 18).
- [Hun07] J. D. HUNTER. “Matplotlib : A 2D Graphics Environment”. In : *Computing in Science Engineering* 9.3 (mai 2007), p. 90-95. DOI : [10 / drbjhg](https://doi.org/10.1109/MCSE.2007.42) (cf. p. 46).
- [HHL11] Frank HUTTER, Holger H. HOOS et Kevin LEYTON-BROWN. “Sequential Model-Based Optimization for General Algorithm Configuration”. In : *Proceedings of the 5th International Conference on Learning and Intelligent Optimization*. LION’05. Berlin, Heidelberg : Springer-Verlag, 2011, p. 507-523. ISBN : 978-3-642-25565-6. DOI : [10 / fb9tjc](https://doi.org/10.1007/978-3-642-25565-6_43) (cf. p. 133).
- [HHL14] Frank HUTTER, Holger HOOS et Kevin LEYTON-BROWN. “An Efficient Approach for Assessing Hyperparameter Importance”. In : *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*. ICML’14. JMLR.org, 2014, p. I-754-I-762 (cf. p. 136).

- [IS15] Sergey IOFFE et Christian SZEGEDY. “Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift”. In : *arXiv :1502.03167 [cs]* (fév. 2015). arXiv : [1502.03167 \[cs\]](https://arxiv.org/abs/1502.03167) (cf. p. 99).
- [Iqb18] Haris IQBAL. *HarisIqbal88/PlotNeuralNet v1.0.0*. Zenodo. Déc. 2018. DOI : [10.5281/zenodo.2526396](https://doi.org/10.5281/zenodo.2526396) (cf. p. 97).
- [ISO08] ISO/TC 108. *ISO 18434-1 :2008 Condition Monitoring and Diagnostics of Machines — Thermography — Part 1 : General Procedures*. English. Standard ISO 18434-1 :2008. 00000. Geneva, Switzerland : International Organization for Standardization / International Organization for Standardization, mar. 2008, p. 24 (cf. p. 47).
- [ISO15] ISO/TC 176. *ISO 9000 :2015 : Quality Management Systems – Fundamentals and Vocabulary*. English. Standard ISO 9000 :2015. Geneva, Switzerland : International Organization for Standardization / International Organization for Standardization, sept. 2015, p. 51 (cf. p. 32).
- [ISO11] ISO/TC 213. *ISO 8015 :2011 : Geometrical Product Specifications (GPS) – Fundamentals – Concepts, Principles and Rules*. English. Standard ISO 9000 :2015. Geneva, Switzerland : International Organization for Standardization / International Organization for Standardization, juin 2011, p. 10 (cf. p. 32).
- [ISO18] ISO/TC 261. *ISO 20457 :2018 : Plastics moulded parts – Tolerances and acceptance conditions*. English. Standard ISO 20457 :2018(E). Geneva, Switzerland : International Organization for Standardization / International Organization for Standardization, sept. 2018, p. 28 (cf. p. 18, 37).
- [IDK98] R. IVESTER, K. DANAI et David Owen KAZMER. “Automatic Tuning of Injection Molding by the Virtual Search Method”. eng. In : *The Journal of injection molding technology* 2.3 (1998), p. 103-108. ISSN : 1533-905X (cf. p. 22).
- [JT15] Kevin JAMIESON et Ameet TALWALKAR. “Non-Stochastic Best Arm Identification and Hyperparameter Optimization”. In : *arXiv :1502.07943 [cs, stat]* (fév. 2015). arXiv : [1502.07943 \[cs, stat\]](https://arxiv.org/abs/1502.07943) (cf. p. 134).
- [Jan92] Tzy-Cherng JAN. “Expert System for the Injection Molding of Engineering Thermoplastics”. English. Thèse de doct. Newark, NJ, USA : New Jersey Institute of Technology, oct. 1992 (cf. p. 15).
- [JS02] Nathalie JAPKOWICZ et Shaju STEPHEN. “The class imbalance problem : A systematic study”. en. In : *Intelligent Data Analysis* 6.5 (jan. 2002), p. 429-449. ISSN : 1088-467X. DOI : [10.3233/IDA-2002-6504](https://doi.org/10.3233/IDA-2002-6504) (cf. p. 85).
- [JSH18] Haifeng JIN, Qingquan SONG et Xia HU. “Auto-Keras : An Efficient Neural Architecture Search System”. In : *arXiv :1806.10282 [cs, stat]* (juin 2018). eprinttype : arxiv eprintclass : cs.LG, cs.AI, stat.ML eprint : <http://arxiv.org/abs/1806.10282>. arXiv : [1806.10282 \[cs, stat\]](https://arxiv.org/abs/1806.10282) (cf. p. 137).
- [Joh+15] Stephen JOHNSTON, Christopher MCCREADY, Daniel HAZEN, Darin VANDERWALKER et David Owen KAZMER. “On-Line Multivariate Optimization of Injection Molding”. en. In : *Polymer Engineering & Science* 55.12 (déc. 2015), p. 2743-2750. ISSN : 00323888. DOI : [10.1002/pen.24163](https://doi.org/10.1002/pen.24163) (cf. p. 23).
- [Jol02] Ian T. JOLLIFFE. *Principal component analysis*. eng. 2. ed. Springer series in statistics. New York, NY : Springer, 2002. ISBN : 978-0-387-95442-4 (cf. p. 105).
- [Jon01] Donald R. JONES. “A Taxonomy of Global Optimization Methods Based on Response Surfaces”. en. In : *Journal of Global Optimization* 21.4 (déc. 2001), p. 345-383. ISSN : 1573-2916. DOI : [10/c229dg](https://doi.org/10/c229dg) (cf. p. 130).

- [JSW98] Donald R. JONES, Matthias SCHONLAU et William J. WELCH. “Efficient Global Optimization of Expensive Black-Box Functions”. en. In : *Journal of Global Optimization* 13.4 (déc. 1998), p. 455-492. ISSN : 1573-2916. DOI : [10 / fg68nc](https://doi.org/10.1007/s10898-000-0003-5) (cf. p. 133).
- [Kam+87] M. R. KAMAL, W. I. PATTERSON, N. CONLEY, Deeb ABU FARA et G. LOHFINK. “Dynamics and Control of Pressure in the Injection Molding of Thermoplastics”. en. In : *Polymer Engineering & Science* 27.18 (oct. 1987), p. 1403-1410. ISSN : 1548-2634. DOI : [10 . 1002 / pen . 760271809](https://doi.org/10.1002/pen.760271809) (cf. p. 16).
- [KPG86] M. R. KAMAL, W. I. PATTERSON et V. G. GOMES. “An Injection Molding Study. Part I : Melt and Barrel Temperature Dynamics”. en. In : *Polymer Engineering & Science* 26.12 (juil. 1986), p. 854-866. ISSN : 1548-2634. DOI : [10 . 1002 / pen . 760261205](https://doi.org/10.1002/pen.760261205) (cf. p. 16).
- [KHS93] Seiji KAMEOKA, Nobuhiro HARAMOTO et Tadamoto SAKAI. “Development of an Expert System for Injection Molding Operations”. en. In : *Advances in Polymer Technology* 12.4 (déc. 1993), p. 403-418. ISSN : 1098-2329. DOI : [10 . 1002 / adv . 1993 . 060120407](https://doi.org/10.1002/adv.1993.060120407) (cf. p. 15).
- [KB95] David Owen KAZMER et P BARKAN. “Dynamic Feed Control : A New Method for Injection Molding of High Quality Plastic Parts”. Thèse de doct. Citeseer, 1995 (cf. p. 16).
- [KD01] David Owen KAZMER et Kourosh DANAI. “Control of Polymer Processing”. en. In : *The Mechanical Systems Design Handbook*. Sous la dir. d’Osita NWOKAH et Yildirim HURMUZLU. T. 24. CRC Press, déc. 2001. ISBN : 978-0-8493-8596-4 978-1-4200-3674-9 (cf. p. 20).
- [KH99] David Owen KAZMER et David HATCH. “Towards Controllability of Injection Molding”. In : *Journal of Materials Processing and Manufacturing Science* 8 (oct. 1999), p. 127-140. DOI : [10 . 1106 / DAD5 - PNGG - U87B - 08Y5](https://doi.org/10.1106/DAD5-PNGG-U87B-08Y5) (cf. p. 9, 11).
- [Kaz+11] David Owen KAZMER, S. P. JOHNSTON, R. X. GAO et Z. FAN. “Feasibility Analysis of an In-Mold Multivariate Sensor”. In : *International Polymer Processing* 26.1 (mar. 2011), p. 63-72. ISSN : 0930-777X. DOI : [10 . 3139 / 217 . 2397](https://doi.org/10.3139/217.2397) (cf. p. 28).
- [Kaz+10] David Owen KAZMER, Suganya VELUSAMY, Sarah WESTERDALE, Stephen JOHNSTON et Robert X. GAO. “A Comparison of Seven Filling to Packing Switchover Methods for Injection Molding”. en. In : *Polymer Engineering & Science* 50.10 (oct. 2010), p. 2031-2043. ISSN : 1548-2634. DOI : [10 . 1002 / pen . 21731](https://doi.org/10.1002/pen.21731) (cf. p. 18).
- [KWH08] David Owen KAZMER, S. WESTERDALE et D. HAZEN. “A Comparison of Statistical Process Control (SPC) and On-Line Multivariate Analyses (MVA) for Injection Molding”. In : *International Polymer Processing* 23.5 (nov. 2008), p. 447-458. ISSN : 0930-777X. DOI : [10 . 3139 / 217 . 2192](https://doi.org/10.3139/217.2192) (cf. p. 19).
- [Kea88] Michael KEARNS. “Thoughts on Hypothesis Boosting.” English. In : *Non publiée, Machine Learning class project* (déc. 1988) (cf. p. 91).
- [KB14] Diederik P. KINGMA et Jimmy BA. “Adam : A Method for Stochastic Optimization”. In : *arXiv :1412.6980 [cs]* (déc. 2014). arXiv : [1412.6980](https://arxiv.org/abs/1412.6980) (cf. p. 95).
- [KW13] Diederik P. KINGMA et Max WELLING. “Auto-Encoding Variational Bayes”. In : *arXiv :1312.6114 [cs, stat]* (déc. 2013). arXiv : [1312.6114 \[cs, stat\]](https://arxiv.org/abs/1312.6114) (cf. p. 124).
- [KW19] Diederik P. KINGMA et Max WELLING. “An Introduction to Variational Autoencoders”. In : *arXiv :1906.02691 [cs, stat]* (juin 2019). arXiv : [1906 . 02691 \[cs, stat\]](https://arxiv.org/abs/1906.02691) (cf. p. 126).

- [Kle+16] Aaron KLEIN, Stefan FALKNER, Simon BARTELS, Philipp HENNIG et Frank HUTTER. “Fast Bayesian Optimization of Machine Learning Hyperparameters on Large Datasets”. In : *arXiv :1605.07079 [cs, stat]* (mai 2016). arXiv : [1605.07079 \[cs, stat\]](https://arxiv.org/abs/1605.07079) (cf. p. 135).
- [KZS15] Gregory KOCH, Richard ZEMEL et Ruslan SALAKHUTDINOV. “Siamese Neural Networks for One-Shot Image Recognition”. English. In : *ICML Deep Learning Workshop*. T. 2. Lille, France, juil. 2015 (cf. p. 106).
- [Koh95] Ron KOHAVI. “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection”. In : *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*. IJCAI’95. San Francisco, CA, USA : Morgan Kaufmann Publishers Inc., 1995, p. 1137-1143. ISBN : 978-1-55860-363-9 (cf. p. 88).
- [Kri51] D. G. KRIGE. “A Statistical Approach to Some Mine Valuation and Allied Problems on the Witwatersrand”. en. Master Thesis. Johannesburg, South Africa : University of the Witwatersrand, 1951 (cf. p. 132).
- [Kri09] Alex KRIZHEVSKY. *Learning Multiple Layers of Features from Tiny Images*. English. Technical. University of Toronto, 2009 (cf. p. 137, 139).
- [KSH12] Alex KRIZHEVSKY, Ilya SUTSKEVER et Geoffrey E. HINTON. “ImageNet Classification with Deep Convolutional Neural Networks”. In : *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. NIPS’12. event-place : Lake Tahoe, Nevada. USA : Curran Associates Inc., 2012, p. 1097-1105 (cf. p. 93).
- [KH91] Anders KROGH et John A. HERTZ. “A Simple Weight Decay Can Improve Generalization”. In : *Proceedings of the 4th International Conference on Neural Information Processing Systems*. NIPS’91. San Francisco, CA, USA : Morgan Kaufmann Publishers Inc., 1991, p. 950-957. ISBN : 978-1-55860-222-9 (cf. p. 102).
- [KL51] S. KULLBACK et R. A. LEIBLER. “On Information and Sufficiency”. EN. In : *The Annals of Mathematical Statistics* 22.1 (mar. 1951), p. 79-86. ISSN : 0003-4851, 2168-8990. DOI : [10.1214/aoms/1177729694](https://doi.org/10.1214/aoms/1177729694) (cf. p. 94).
- [Kur+09] Mustafa KURT, O. SABAN KAMBER, Yusuf KAYNAK, Gurcan ATAKOK et Oguz GIRIT. “Experimental Investigation of Plastic Injection Molding : Assessment of the Effects of Cavity Pressure and Mold Temperature on the Quality of the Final Products”. In : *Materials & Design* 30.8 (sept. 2009), p. 3217-3224. ISSN : 0261-3069. DOI : [10.1016/j.matdes.2009.01.004](https://doi.org/10.1016/j.matdes.2009.01.004) (cf. p. 27).
- [Lac18] Thomas LACOMBE. “Exploitation d’une information multiéclairages pour une approche générique de l’inspection automatique de la qualité visuelle des produits en industrie.” fr. PhD Thesis. Annecy, France : UGA - Université Grenoble Alpes, sept. 2018. HAL : [tel-01959270](https://tel.archives-ouvertes.fr/tel-01959270) (cf. p. 29, 38, 62, 64, 83-85).
- [Lac+17] Thomas LACOMBE, Bruno ALBERT, Halise DERELI, Mélanie TOMCZYK, Youva NAIT-YOUCEF, Jocelyn PHILPPA et Hughes FAVRELIÈRE. “Voir pour Toucher ? Caractérisation Optique de la Sensation d’Adhérence”. In : *QUALITA 2017*. Bourges, France, août 2017 (cf. p. 69).
- [Lac75] Franck LACZKO. “Controller for Injection Molding Machine”. English. US 3893792 A (Westwood, Massachusets). Juil. 1975 (cf. p. 15).
- [Lan79] Ioan Doré LANDAU. *Adaptive Control : The Model Reference Approach*. New York, NY, USA : Marcel Dekker, Inc., 1979. ISBN : 978-0-8247-6548-4 (cf. p. 17).

- [Lau+01] H. C. W. LAU, A. NING, K. F. PUN et K. S. CHIN. “Neural Networks for the Dimensional Control of Molded Parts Based on a Reverse Process Model”. In : *Journal of Materials Processing Technology* 117.1–2 (nov. 2001), p. 89-96. ISSN : 0924-0136. DOI : [10.1016/S0924-0136\(01\)01086-X](https://doi.org/10.1016/S0924-0136(01)01086-X) (cf. p. 22).
- [LMU87] Yann LE CUN, Maurice MILGRAM et UNIVERSITÉ PIERRE ET MARIE CURIE (PARIS). “Modèles connexionnistes de l’apprentissage”. French. PhD Thesis. Paris : Université Paris 6, 1987 (cf. p. 123).
- [LeC+89] Y. LECUN, B. BOSEN, J. S. DENKER, D. HENDERSON, R. E. HOWARD, W. HUBBARD et L. D. JACKEL. “Backpropagation Applied to Handwritten Zip Code Recognition”. In : *Neural Comput.* 1.4 (déc. 1989), p. 541-551. ISSN : 0899-7667. DOI : [10.1162/neco.1989.1.4.541](https://doi.org/10.1162/neco.1989.1.4.541) (cf. p. 92, 93).
- [LeC+98] Yann LECUN, Léon BOTTOU, Genevieve B. ORR et Klaus-Robert MÜLLER. “Efficient BackProp”. In : *Neural Networks : Tricks of the Trade, This Book is an Outgrowth of a 1996 NIPS Workshop*. London, UK, UK : Springer-Verlag, 1998. ISBN : 978-3-540-65311-0 (cf. p. 93).
- [Leg02] Anne-Claire LEGRAND. “Thermographie Multispectrale Haute et Basse Température : Application Au Contrôle Non Destructif”. Thesis. Dijon, jan. 2002 (cf. p. 47).
- [Li+14] L. LI, Z. LI, K. LI, L. BLAREL et M. WENDISCH. “A Method to Calculate Stokes Parameters and Angle of Polarization of Skylight from Polarized CIMEL Sun/Sky Radiometers”. In : *Journal of Quantitative Spectroscopy and Radiative Transfer* 149 (déc. 2014), p. 334-346. ISSN : 0022-4073. DOI : [10/gf8zt7](https://doi.org/10/gf8zt7) (cf. p. 57).
- [Li+16] Lisha LI, Kevin JAMIESON, Giulia DESALVO, Afshin ROSTAMIZADEH et Ameet TALWALKAR. “Hyperband : A Novel Bandit-Based Approach to Hyperparameter Optimization”. In : *arXiv :1603.06560 [cs, stat]* (mar. 2016). arXiv : [1603 . 06560 \[cs, stat\]](https://arxiv.org/abs/1603.06560) (cf. p. 136).
- [LCY13] Min LIN, Qiang CHEN et Shuicheng YAN. “Network In Network”. In : *arXiv :1312.4400 [cs]* (déc. 2013). arXiv : [1312 . 4400 \[cs\]](https://arxiv.org/abs/1312.4400) (cf. p. 102).
- [Lin76] Seppo LINNAINMAA. “Taylor Expansion of the Accumulated Rounding Error”. en. In : *BIT Numerical Mathematics* 16.2 (juin 1976), p. 146-160. ISSN : 1572-9125. DOI : [10/cv4d63](https://doi.org/10/cv4d63) (cf. p. 92).
- [LLZ16] Jingxiang LIU, Tao LIU et Jie ZHANG. “Window-Based Stepwise Sequential Phase Partition for Nonlinear Batch Process Monitoring”. In : *Industrial & Engineering Chemistry Research* 55.34 (août 2016), p. 9229-9243. ISSN : 0888-5885. DOI : [10 . 1021/acs.iecr . 6b01257](https://doi.org/10.1021/acs.iecr.6b01257) (cf. p. 19).
- [Llo82] S. LLOYD. “Least Squares Quantization in PCM”. In : *IEEE Transactions on Information Theory* 28.2 (mar. 1982), p. 129-137. ISSN : 0018-9448. DOI : [10 . 1109 / TIT . 1982 . 1056489](https://doi.org/10.1109/TIT.1982.1056489) (cf. p. 113).
- [Lu+04] Ningyun LU, Yi YANG, Furong GAO et Fuli WANG. “Stage-Based Multivariate Statistical Analysis for Injection Molding”. In : *IFAC Proceedings Volumes*. 7th International Symposium on Advanced Control of Chemical Processes (ADCHEM 2003), Hong-Kong, 11-14 January 2004 37.1 (jan. 2004), p. 439-444. ISSN : 1474-6670. DOI : [10/gf7jxc](https://doi.org/10/gf7jxc) (cf. p. 19).
- [Ma74] Carl Y. W. MA. “A Design Approach to a Computer-Controlled Injection-Molding Machine”. en. In : *Polymer Engineering & Science* 14.11 (nov. 1974), p. 768-772. ISSN : 1548-2634. DOI : [10 . 1002 / pen . 760141106](https://doi.org/10.1002/pen.760141106) (cf. p. 9, 20).

- [MHN13] Andrew MAAS, Awni HANNUN et Andrew NG. “Rectifier Nonlinearities Improve Neural Network Acoustic Models”. In : *Proceedings ICML* (juin 2013) (cf. p. [98](#)).
- [Maa14] Laurens van der MAATEN. “Accelerating T-SNE Using Tree-Based Algorithms”. In : *Journal of Machine Learning Research* 15 (2014), p. 3221-3245 (cf. p. [121](#)).
- [Mac67] J. MACQUEEN. “Some Methods for Classification and Analysis of Multivariate Observations”. EN. In : *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1 : Statistics*. The Regents of the University of California, 1967 (cf. p. [113](#)).
- [Mah+18] Dhruv MAHAJAN, Ross GIRSHICK, Vignesh RAMANATHAN, Kaiming HE, Manohar PALURI, Yixuan LI, Ashwin BHARAMBE et Laurens VAN DER MAATEN. “Exploring the Limits of Weakly Supervised Pretraining”. In : *arXiv :1805.00932 [cs]* (mai 2018). arXiv : [1805.00932 \[cs\]](https://arxiv.org/abs/1805.00932) (cf. p. [97](#), [102](#)).
- [Mah36] PC MAHALANOBIS. “On the Generalised Distance in Statistics”. In : *Proceedings National Institute of Science, India*. T. 2. Avr. 1936, p. 49-55 (cf. p. [83](#), [85](#), [108](#)).
- [Mal12] Nada Bou MALHAB. “Moulage par microinjection des polymères semi-cristallins”. fr. PhD Thesis. Paris, France : École Nationale Supérieure d’Arts et Métiers, déc. 2012 (cf. p. [38](#)).
- [Mar+00] Stephen R. MARSCHNER, Stephen H. WESTIN, Eric P. F. LAFORTUNE et Kenneth E. TORRANCE. “Image-Based Bidirectional Reflectance Distribution Function Measurement”. EN. In : *Applied Optics* 39.16 (juin 2000), p. 2592-2600. ISSN : 1539-4522. DOI : [10.1364/AO.39.002592](https://doi.org/10.1364/AO.39.002592) (cf. p. [69](#)).
- [Mas+99] Llew MASON, Jonathan BAXTER, Peter BARTLETT et Marcus FREAN. “Boosting Algorithms As Gradient Descent”. In : *Proceedings of the 12th International Conference on Neural Information Processing Systems*. NIPS’99. Cambridge, MA, USA : MIT Press, 1999, p. 512-518 (cf. p. [91](#)).
- [Mat63] Georges MATHERON. “Principles of Geostatistics”. en. In : *Economic Geology* 58.8 (déc. 1963), p. 1246-1266. ISSN : 0361-0128. DOI : [10/fdsdjx](https://doi.org/10/fdsdjx) (cf. p. [132](#)).
- [MP43] Warren S. MCCULLOCH et Walter PITTS. “A Logical Calculus of the Ideas Immanent in Nervous Activity”. en. In : *The bulletin of mathematical biophysics* 5.4 (déc. 1943), p. 115-133. ISSN : 1522-9602. DOI : [10/djsbj6](https://doi.org/10/djsbj6) (cf. p. [92](#)).
- [MHM18] Leland MCINNES, John HEALY et James MELVILLE. “UMAP : Uniform Manifold Approximation and Projection for Dimension Reduction”. In : *arXiv :1802.03426 [cs, stat]* (fév. 2018). arXiv : [1802.03426 \[cs, stat\]](https://arxiv.org/abs/1802.03426) (cf. p. [121](#), [122](#)).
- [McI+18] Leland MCINNES, John HEALY, Nathaniel SAUL et Lukas GROSSBERGER. “UMAP : Uniform Manifold Approximation and Projection”. In : *Journal of Open Source Software* 3.29 (sept. 2018), p. 861. ISSN : 2475-9066. DOI : [10/gf6k3s](https://doi.org/10/gf6k3s) (cf. p. [121](#)).
- [MBC79] M. D. MCKAY, R. J. BECKMAN et W. J. CONOVER. “A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code”. In : *Technometrics* 21.2 (1979), p. 239-245. ISSN : 0040-1706. DOI : [10/bp9p63](https://doi.org/10/bp9p63) (cf. p. [132](#)).
- [Meg+19] Anis MEGUENANI, Jean-Pierre CHAMBARD, Yvan MAILLOT et Christophe CUDEL. “Deflectometry Based Inline Inspection of Linearly Moving Plastic Parts”. In : (juil. 2019) (cf. p. [144](#)).

- [Men+03] R. MENDOZA, G. RÉGNIER, W. SEILER et J. L. LEBRUN. “Spatial Distribution of Molecular Orientation in Injection Molded iPP : Influence of Processing Conditions”. In : *Polymer* 44.11 (mai 2003), p. 3363-3373. ISSN : 0032-3861. DOI : [10/bdg6zh](https://doi.org/10/bdg6zh) (cf. p. 38).
- [Met78] C. E. METZ. “Basic principles of ROC analysis”. eng. In : *Seminars in Nuclear Medicine* 8.4 (oct. 1978), p. 283-298. ISSN : 0001-2998 (cf. p. 85).
- [MS09] Walter MICHAELI et Andreas SCHREIBER. “Online Control of the Injection Molding Process Based on Process Variables”. en. In : *Advances in Polymer Technology* 28.2 (juin 2009), p. 65-76. ISSN : 1098-2329. DOI : [10.1002/adv.20153](https://doi.org/10.1002/adv.20153) (cf. p. 17).
- [Miy+18] Takeru MIYATO, Toshiki KATAOKA, Masanori KOYAMA et Yuichi YOSHIDA. “Spectral Normalization for Generative Adversarial Networks”. In : *arXiv :1802.05957 [cs, stat]* (fév. 2018). arXiv : [1802.05957 \[cs, stat\]](https://arxiv.org/abs/1802.05957) (cf. p. 128).
- [MNB09] M. MOGUEDET, P. NAMY et Y. BÉREAUX. “Use of Comsol Multiphysics to Understand and Optimize the Filling Phase in Injection and Micro-Injection Molding Process”. In : *Popular Plastics & Packaging* 54.11 (nov. 2009), p. 29-32. ISSN : 09710078 (cf. p. 13).
- [Mos13] Lawrence MOSLEY. “A balanced approach to the multi-class imbalance problem”. In : *Graduate Theses and Dissertations* (jan. 2013). DOI : <https://doi.org/10.31274/etd-180810-3375> (cf. p. 85).
- [Mye71] Raymond H MYERS. *Response Surface Methodology*. English. OCLC : 655239528. Boston : Allyn and Bacon, 1971 (cf. p. 131).
- [Nag+18] Pierre NAGORNY, Thomas LACOMBE, Hugues FAVRELIÈRE, Maurice PILLET, Éric PAIREL, Ronan Le GOFF, Marlène WALI, Jérôme LOUREAUX et Patrice KIENER. “Generative Adversarial Networks for Geometric Surfaces Prediction in Injection Molding : Performance Analysis with Discrete Modal Decomposition”. In : *2018 IEEE International Conference on Industrial Technology (ICIT)*. Fév. 2018, p. 1514-1519. DOI : [10.1109/ICIT.2018.8352405](https://doi.org/10.1109/ICIT.2018.8352405) (cf. p. 5, 47).
- [Nag+19] Pierre NAGORNY, Thomas LACOMBE, Thomas MULLER, Hugues FAVRELIÈRE, Éric PAIREL et Maurice PILLET. “Polarimetric Imaging for Quality Control in Injection Molding”. In : *Fourteenth International Conference on Quality Control by Artificial Vision*. T. 11172. International Society for Optics and Photonics, juil. 2019, p. 1117217. DOI : [10/gf48kh](https://doi.org/10/gf48kh) (cf. p. 5, 54, 57, 60, 137).
- [NPP19] Pierre NAGORNY, Eric PAIREL et Maurice PILLET. “Dispositif et Procédé de Contrôle de Pièces”. English. (Annecy, France). Sept. 2019 (cf. p. 5, 143).
- [NPP17a] Pierre NAGORNY, Éric PAIREL et Maurice PILLET. “Pilotage En Injection Plastique - État de l’Art”. In : *12ème Congrès International de Génie Industriel (CIGI 2017)*. Compiègne, France, mai 2017. HAL : [hal-01551840](https://hal-01551840) (cf. p. 5, 8, 14).
- [NPP17b] Pierre NAGORNY, Éric PAIREL et Maurice PILLET. “Vers Un Pilotage de La Qualité Des Pièces Injectées”. In : *arXiv :1707.01765 [Cs]*. Juil. 2017. arXiv : [1707.01765 \[cs\]](https://arxiv.org/abs/1707.01765) (cf. p. 5).
- [Nag+17] Pierre NAGORNY, Maurice PILLET, Éric PAIREL, Ronan Le GOFF, Jérôme LOUREAUX, Marlène WALI et Patrice KIENER. “Quality Prediction in Injection Molding”. In : *2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*. Juin 2017, p. 141-146. DOI : [10.1109/CIVEMSA.2017.7995316](https://doi.org/10.1109/CIVEMSA.2017.7995316) (cf. p. 5, 36, 61, 65).

- [NH10] Vinod NAIR et Geoffrey E. HINTON. “Rectified Linear Units Improve Restricted Boltzmann Machines”. In : *Proceedings of the 27th International Conference on International Conference on Machine Learning*. ICML’10. USA : Omnipress, 2010, p. 807-814. ISBN : 978-1-60558-907-7 (cf. p. 98).
- [Nas51] John NASH. “Non-Cooperative Games”. In : *Annals of Mathematics* 54.2 (1951), p. 286-295. ISSN : 0003-486X. DOI : [10/dzj4db](https://doi.org/10/dzj4db) (cf. p. 127).
- [Nas50] John F. NASH. “Equilibrium Points in N-Person Games”. en. In : *Proceedings of the National Academy of Sciences* 36.1 (jan. 1950), p. 48-49. ISSN : 0027-8424, 1091-6490. DOI : [10/b6hm7s](https://doi.org/10/b6hm7s) (cf. p. 127).
- [NFB97] Shree K. NAYAR, Xi-Sheng FANG et Terrance BOULT. “Separation of Reflection Components Using Color and Polarization”. en. In : *International Journal of Computer Vision* 21.3 (fév. 1997), p. 163-186. ISSN : 1573-1405. DOI : [10/cp36j8](https://doi.org/10/cp36j8) (cf. p. 53).
- [Nor87] Association Française de NORMALISATION (AFNOR). *NF T58-000*. English. Standard. AFNOR, oct. 1987 (cf. p. 18).
- [Ode16] Augustus ODENA. “Semi-Supervised Learning with Generative Adversarial Networks”. In : *arXiv :1606.01583 [cs, stat]* (juin 2016). arXiv : [1606.01583 \[cs, stat\]](https://arxiv.org/abs/1606.01583) (cf. p. 128).
- [Ols+16a] Randal S. OLSON, Nathan BARTLEY, Ryan J. URBANOWICZ et Jason H. MOORE. “Evaluation of a Tree-Based Pipeline Optimization Tool for Automating Data Science”. In : *Proceedings of the Genetic and Evolutionary Computation Conference 2016*. GECCO ’16. New York, NY, USA : ACM, juil. 2016. ISBN : 978-1-4503-4206-3. DOI : [10.1145/2908812.2908918](https://doi.org/10.1145/2908812.2908918) (cf. p. 136).
- [Ols+16b] Randal S. OLSON, Ryan J. URBANOWICZ, Peter C. ANDREWS, Nicole A. LAVENDER, La Creis KIDD et Jason H. MOORE. “Automating Biomedical Data Science through Tree-Based Pipeline Optimization”. In : *arXiv :1601.07925 [cs]* (jan. 2016). eprinttype : arxiv eprintclass : cs.LG, cs.NE eprint : <http://arxiv.org/abs/1601.07925>. arXiv : [1601.07925 \[cs\]](https://arxiv.org/abs/1601.07925) (cf. p. 136).
- [Oqu+15] Maxime OQUAB, Leon BOTTOU, Ivan LAPTEV et Josef SIVIC. “Is Object Localization for Free ? - Weakly-Supervised Learning With Convolutional Neural Networks”. In : *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, p. 685-694 (cf. p. 102).
- [Pai16] Eric PAIREL. “Maîtrise de La Qualité Géométrique Des Produits”. Habilitation à Diriger Des Recherches. ComUE Grenoble-Alpes, déc. 2016 (cf. p. 45).
- [PA88] I. O. PANDELIDIS et A. R. AGRAWAL. “Optimal Anticipatory Control of Ram Velocity in Injection Molding”. en. In : *Polymer Engineering & Science* 28.3 (fév. 1988), p. 147-156. ISSN : 1548-2634. DOI : [10.1002/pen.760280305](https://doi.org/10.1002/pen.760280305) (cf. p. 16).
- [Pas94] Pierre PASTRÉ. “Le rôle des schèmes et concepts dans la formation des compétences”. Français. In : *Performances humaines et techniques* 71 (1994). 00039, p. 21-28 (cf. p. 15).
- [PS04] Pierre PASTRÉ et R. SAMURÇAY. “Le Rôle Des Concepts Pragmatiques Dans La Gestion Des Situations Problèmes : Le Cas Des Régleurs En Plasturgie”. In : *Recherches En Didactique Professionnelle*. Octarès. 2004, p. 17-49 (cf. p. 15).
- [Pea01] Karl PEARSON. “On Lines and Planes of Closest Fit to Systems of Points in Space”. English. In : *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.11 (nov. 1901), p. 559-572. ISSN : 1941-5982. DOI : [10/dd63n4](https://doi.org/10/dd63n4) (cf. p. 114).

- [Ped+11] Fabian PEDREGOSA, Gaël VAROQUAUX, Alexandre GRAMFORT, Vincent MICHEL, Bertrand THIRION, Olivier GRISEL, Mathieu BLONDEL, Peter PRETTENHOFER, Ron WEISS, Vincent DUBOURG, Jake VANDERPLAS, Alexandre PASSOS, David COURNAPEAU, Matthieu BRUCHER, Matthieu PERROT et Édouard DUCHESNAY. “Scikit-learn : Machine Learning in Python”. In : *Journal of Machine Learning Research* 12 (nov. 2011). ISSN : 1532-4435 (cf. p. 88, 89, 105, 133).
- [PC95] Joseph PEZZANITI Larry et Russell A CHIPMAN. “Mueller matrix imaging polarimetry”. English (US). In : *Optical Engineering* 34.6 (juin 1995), p. 1558-1568. ISSN : 0091-3286 (cf. p. 70).
- [PMB03] Maurice PILLET, Jean Luc MAIRE et Vincent BRONET. “Maîtrise Statistique Des Procédés (MSP) - Cas Particulier de l’injection Plastique”. In : oct. 2003 (cf. p. 19).
- [Pit16] Gilles PITARD. “Métrologie et modélisation de l’aspect pour l’inspection qualité des surfaces”. fr. Thèse de doct. Université Grenoble Alpes, mai 2016 (cf. p. 29, 64, 84).
- [PB46] R. L. PLACKETT et J. P. BURMAN. “The Design of Optimum Multifactorial Experiments”. In : *Biometrika* 33.4 (juin 1946), p. 305. ISSN : 00063444. DOI : [10.2307/2332195](https://doi.org/10.2307/2332195) (cf. p. 76, 131).
- [Qué+19] Yvain QUÉAU, Florian LEPORCQ, Alexis LECHERVY et Ayman ALFALOU. “Learning to Classify Materials Using Mueller Imaging Polarimetry”. In : *Fourteenth International Conference on Quality Control by Artificial Vision (QCAV)*. Mulhouse, France, mai 2019. DOI : [10/gf8twj](https://doi.org/10/gf8twj) (cf. p. 70).
- [RMC15] Alec RADFORD, Luke METZ et Soumith CHINTALA. “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks”. In : *arXiv :1511.06434 [cs]* (nov. 2015). arXiv : [1511.06434 \[cs\]](https://arxiv.org/abs/1511.06434) (cf. p. 127).
- [RWC16] Sanjeev Jagannatha RAO, Yufei WANG et Garrison W. COTTRELL. “A Deep Siamese Neural Network Learns the Human-Perceived Similarity Structure of Facial Expressions Without Explicit Categories”. In : *CogSci.* 2016 (cf. p. 106).
- [RMW14] Danilo Jimenez REZENDE, Shakir MOHAMED et Daan WIERSTRA. “Stochastic Backpropagation and Approximate Inference in Deep Generative Models”. In : *arXiv :1401.4082 [cs, stat]* (jan. 2014). arXiv : [1401.4082 \[cs, stat\]](https://arxiv.org/abs/1401.4082) (cf. p. 124).
- [Ric+09] Jean-François RICHARD, Pierre PASTRÉ, Pierre PARAGE, Emmanuel SANDER, Michel FUTTERSACK et Jean-Marc LABAT. “Analyse des stratégies de correction de défauts en plasturgie à l'aide d'un modèle de résolution de problème à base de contraintes”. fr. In : *Le travail humain* 72.3 (sept. 2009), p. 267-292. ISSN : 1243-1370 (cf. p. 15).
- [Rio+15] Cécile RIOU, Bruno COLICCHIO, Jean-Philippe LAUFFENBURGER, Olivier HAEBERLE et Christophe CUDEL. “Calibration and Disparity Maps for a Depth Camera Based on a Four-Lens Device”. In : *Journal of Electronic Imaging* 24 (2015), p. 1-11. DOI : [10.1117/1.JEI.24.6.061108](https://doi.org/10.1117/1.JEI.24.6.061108) (cf. p. 54).
- [RMM18] Francisco J. ROMERO-RAMIREZ, Rafael MUÑOZ-SALINAS et Rafael MEDINA-CARNICER. “Speeded up Detection of Squared Fiducial Markers”. In : *Image and Vision Computing* 76 (août 2018), p. 38-47. ISSN : 0262-8856. DOI : [10.1016/j.imavis.2018.05.004](https://doi.org/10.1016/j.imavis.2018.05.004) (cf. p. 55).
- [RFB15] Olaf RONNEBERGER, Philipp FISCHER et Thomas BROX. “U-Net : Convolutional Networks for Biomedical Image Segmentation”. In : *arXiv :1505.04597 [cs]* Medical Image Computing and Computer-Assisted Intervention (MICCAI) (mai 2015). arXiv : [1505.04597 \[cs\]](https://arxiv.org/abs/1505.04597) (cf. p. 49).

- [Ros58] F. ROSENBLATT. “The Perceptron : A Probabilistic Model for Information Storage and Organization in the Brain”. In : *Psychological Review* 65.6 (1958), p. 386-408. ISSN : 1939-1471(Electronic),0033-295X(Print). DOI : [10/fg6wr5](https://doi.org/10/fg6wr5) (cf. p. 92).
- [RHW85] David E. RUMELHART, Geoffrey E. HINTON et Ronald J. WILLIAMS. *Learning Internal Representations by Error Propagation*. en. Rapp. tech. Sept. 1985 (cf. p. 92, 93).
- [Sad16] Peter SADOWSKI. *Notes on backpropagation*. English. Rapp. tech. Irvine, CA 92697 : University of California, 2016, p. 1-4 (cf. p. 94).
- [Sal+16] Tim SALIMANS, Ian GOODFELLOW, Wojciech ZAREMBA, Vicki CHEUNG, Alec RADFORD et Xi CHEN. “Improved Techniques for Training GANs”. In : *arXiv :1606.03498 [cs]* (juin 2016). arXiv : [1606.03498 \[cs\]](https://arxiv.org/abs/1606.03498) (cf. p. 128).
- [SF07] Serge SAMPER et Fabien FORMOSA. “Form Defects Tolerancing by Natural Modes Analysis”. In : *Journal of Computing and Information Science in Engineering* 7.1 (mar. 2007), p. 44-51. ISSN : 1530-9827. DOI : [10.1115/1.2424247](https://doi.org/10.1115/1.2424247) (cf. p. 63).
- [San+18] Mark SANDLER, Andrew HOWARD, Menglong ZHU, Andrey ZHMOGINOV et Liang-Chieh CHEN. “MobileNetV2 : Inverted Residuals and Linear Bottlenecks”. In : *arXiv :1801.04381 [cs]* (jan. 2018). eprinttype : arxiv eprintclass : cs.CV eprint : [http://arxiv.org/abs/1801.04381](https://arxiv.org/abs/1801.04381). arXiv : [1801.04381 \[cs\]](https://arxiv.org/abs/1801.04381) (cf. p. 97).
- [San83] Bernard SANSCHAGRIN. “Process Control of Injection Molding”. en. In : *Polymer Engineering & Science* 23.8 (juin 1983), p. 431-438. ISSN : 1548-2634. DOI : [10.1002/pen.760230804](https://doi.org/10.1002/pen.760230804) (cf. p. 17).
- [Sch90] Robert E. SCHAPIRE. “The Strength of Weak Learnability”. In : *Mach. Learn.* 5.2 (juil. 1990), p. 197-227. ISSN : 0885-6125. DOI : [10.1023/A:1022648800760](https://doi.org/10.1023/A:1022648800760) (cf. p. 91).
- [SMB10] Dominik SCHERER, Andreas MÜLLER et Sven BEHNKE. “Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition”. en. In : *Artificial Neural Networks – ICANN 2010*. Sous la dir. de Konstantinos DIAMANTARAS, Wlodek DUCH et Lazaros S. ILIADIS. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010, p. 92-101. ISBN : 978-3-642-15825-4 (cf. p. 100).
- [Sch19] Juergen SCHMIDHUBER. “Unsupervised Minimax : Adversarial Curiosity, Generative Adversarial Networks, and Predictability Minimization”. In : *arXiv :1906.04493 [cs]* (juin 2019). arXiv : [1906.04493 \[cs\]](https://arxiv.org/abs/1906.04493) (cf. p. 127).
- [SM00] O. SCHNERR-HÄSELBARTH et W. MICHAELI. “Automation of Online Quality Control in Injection Moulding”. en. In : *Macromolecular Materials and Engineering* 284-285.1 (déc. 2000), p. 81-85. ISSN : 1439-2054. DOI : [10.1002/1439-2054\(20001201\)284:1<81::AID-MAME81>3.0.CO;2-Q](https://doi.org/10.1002/1439-2054(20001201)284:1<81::AID-MAME81>3.0.CO;2-Q) (cf. p. 14).
- [Sch+01] Bernhard SCHÖLKOPF, John C. PLATT, John C. SHawe-TAYLOR, Alex J. SMOLA et Robert C. WILLIAMSON. “Estimating the Support of a High-Dimensional Distribution”. In : *Neural Comput.* 13.7 (juil. 2001), p. 1443-1471. ISSN : 0899-7667. DOI : [10.1162/089976601750264965](https://doi.org/10.1162/089976601750264965) (cf. p. 83).
- [SSM97] Bernhard SCHÖLKOPF, Alexander SMOLA et Klaus-Robert MÜLLER. “Kernel Principal Component Analysis”. en. In : *Artificial Neural Networks — ICANN’97*. Sous la dir. de Wulfram GERSTNER, Alain GERMOND, Martin HASLER et Jean-Daniel NICoud. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 1997, p. 583-588. ISBN : 978-3-540-69620-9 (cf. p. 118).
- [SSM98] Bernhard SCHÖLKOPF, Alexander SMOLA et Klaus-Robert MÜLLER. “Nonlinear Component Analysis as a Kernel Eigenvalue Problem”. In : *Neural Computation* 10.5 (juil. 1998), p. 1299-1319. ISSN : 0899-7667. DOI : [10/dpn6x6](https://doi.org/10/dpn6x6) (cf. p. 118).

- [Sch+99] Bernhard SCHÖLKOPF, Robert WILLIAMSON, Alex SMOLA, John SHawe-Taylor et John PLATT. “Support Vector Method for Novelty Detection”. In : *Proceedings of the 12th International Conference on Neural Information Processing Systems*. NIPS’99. Cambridge, MA, USA : MIT Press, 1999, p. 582-588 (cf. p. 83).
- [SKP15] Florian SCHROFF, Dmitry KALENICHENKO et James PHILBIN. “FaceNet : A Unified Embedding for Face Recognition and Clustering”. In : *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (juin 2015), p. 815-823. DOI : [10.1109/CVPR.2015.7298682](https://doi.org/10.1109/CVPR.2015.7298682). arXiv : [1503.03832](https://arxiv.org/abs/1503.03832) (cf. p. 108-110).
- [Sch+02] Heinrich SCHWENKE, Ulrich NEUSCHAEFER-RUBE, Tilo PFEIFER et Horst KUNZMANN. “Optical Methods for Dimensional Metrology in Production Engineering”. In : *CIRP Annals - Manufacturing Technology* 51.2 (jan. 2002), p. 685-699. ISSN : 0007-8506. DOI : [10.1016/S0007-8506\(07\)61707-7](https://doi.org/10.1016/S0007-8506(07)61707-7) (cf. p. 28).
- [Sha78] A. SHANKAR. “Dynamic Modeling and Control of Injection Molding Machines”. English. Doctoral Dissertation. Pittsburgh : Carnegie-Mellon University, 1978 (cf. p. 13).
- [SP82] A. SHANKAR et F. W. PAUL. “A Mathematical Model for the Evaluation of Injection Molding Machine Control”. In : *Journal of Dynamic Systems, Measurement, and Control* 104.1 (mar. 1982), p. 86-92. ISSN : 0022-0434. DOI : [10.1115/1.3149636](https://doi.org/10.1115/1.3149636) (cf. p. 13).
- [Sha48] C. E. SHANNON. “A Mathematical Theory of Communication”. In : *The Bell System Technical Journal* 27.3 (juil. 1948), p. 379-423. ISSN : 0005-8580. DOI : [10.1002/j.1538-7305.1948.tb01338.x](https://doi.org/10.1002/j.1538-7305.1948.tb01338.x) (cf. p. 94).
- [SS97] K. SHELESH-NEZHAD et E. SIORES. “An Intelligent System for Plastic Injection Molding Process Design”. In : *Journal of Materials Processing Technology* 63.1 (jan. 1997), p. 458-462. ISSN : 0924-0136. DOI : [10.1016/S0924-0136\(96\)02664-7](https://doi.org/10.1016/S0924-0136(96)02664-7) (cf. p. 15).
- [She+97] Gal SHERBELIS, Emlyn GARVEY, Colchester ROAD et David Owen KAZMER. “The Methods and Benefits of Establishing a Process Window”. In : *Submitted for the 1997 Annual Technical Conference of the Society of Plastics Engineers*. Avr. 1997 (cf. p. 18).
- [She30] Walter Andrew SHEWHART. “Economic Quality Control of Manufactured Product”. en. In : *Bell System Technical Journal* 9.2 (avr. 1930), p. 364-389. ISSN : 1538-7305. DOI : [10.1002/j.1538-7305.1930.tb00373.x](https://doi.org/10.1002/j.1538-7305.1930.tb00373.x) (cf. p. 18, 19).
- [She31] Walter Andrew SHEWHART. *Economic Control of Quality of Manufactured Product*. en. New York, NY, USA : Van Nostrand, jan. 1931. ISBN : 978-0-87389-076-2 (cf. p. 19).
- [Shi+16] Hailin SHI, Yang YANG, Xiangyu ZHU, Shengcai LIAO, Zhen LEI, Weishi ZHENG et Stan Z. LI. “Embedding Deep Metric for Person Re-identification A Study Against Large Variations”. In : *arXiv :1611.00137 [cs]* (nov. 2016). arXiv : [1611.00137](https://arxiv.org/abs/1611.00137) (cf. p. 110).
- [SZ14] Karen SIMONYAN et Andrew ZISSERMAN. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. In : *arXiv :1409.1556 [cs]* (sept. 2014). arXiv : [1409.1556](https://arxiv.org/abs/1409.1556) (cf. p. 93).
- [Sob67] Ilya Meyerovich SOBOL’. “On the Distribution of Points in a Cube and the Approximate Evaluation of Integrals”. en. In : *USSR Computational Mathematics and Mathematical Physics* 7.4 (jan. 1967), p. 86-112. ISSN : 00415553. DOI : [10/crdj6j](https://doi.org/10/crdj6j) (cf. p. 132, 133).

- [Soh16] Kihyuk SOHN. “Improved Deep Metric Learning with Multi-class N-pair Loss Objective”. In : *Proceedings of the 30th International Conference on Neural Information Processing Systems*. NIPS’16. event-place : Barcelona, Spain. USA : Curran Associates Inc., 2016, p. 1857-1865. ISBN : 978-1-5108-3881-9 (cf. p. [110](#)).
- [Son+15] Hyun Oh SONG, Yu XIANG, Stefanie JEGELKA et Silvio SAVARESE. “Deep Metric Learning via Lifted Structured Feature Embedding”. In : *arXiv :1511.06452 [cs]* (nov. 2015). arXiv : [1511.06452](#) (cf. p. [110](#)).
- [Sri+14] Nitish SRIVASTAVA, Geoffrey HINTON, Alex KRIZHEVSKY, Ilya SUTSKEVER et Ruslan SALAKHUTDINOV. “Dropout : A Simple Way to Prevent Neural Networks from Overfitting”. In : *Journal of Machine Learning Research* 15 (2014), p. 1929-1958 (cf. p. [99](#)).
- [SGS15a] Rupesh Kumar SRIVASTAVA, Klaus GREFF et Jürgen SCHMIDHUBER. “Highway Networks”. In : *arXiv :1505.00387 [cs]* (mai 2015). arXiv : [1505.00387 \[cs\]](#) (cf. p. [96](#)).
- [SGS15b] Rupesh Kumar SRIVASTAVA, Klaus GREFF et Jürgen SCHMIDHUBER. “Training Very Deep Networks”. In : *arXiv :1507.06228 [cs]* (juil. 2015). arXiv : [1507.06228 \[cs\]](#) (cf. p. [96](#)).
- [Sto51] G. G. STOKES. “On the Composition and Resolution of Streams of Polarized Light from Different Sources”. In : *Transactions of the Cambridge Philosophical Society* 9 (1851), p. 399 (cf. p. [57](#)).
- [Suh90] Nam P. SUH. *The Principles of Design*. en. Oxford University Press, 1990. ISBN : 978-0-19-504345-7 (cf. p. [12](#)).
- [Sze+14] Christian SZEGEDY, Wei LIU, Yangqing JIA, Pierre SERMANET, Scott REED, Dragomir ANGUELOV, Dumitru ERHAN, Vincent VANHOUCKE et Andrew RABINOVICH. “Going Deeper with Convolutions”. In : *arXiv :1409.4842 [cs]* (sept. 2014). arXiv : [1409.4842](#) (cf. p. [108](#)).
- [Thy01] Peter THYREGOD. “Modelling and Monitoring in Injection Molding”. eng. Doctorat. Department of Mathematical Modelling, Technical University of Denmark, jan. 2001 (cf. p. [9](#)).
- [Tik43] Andrey Nikolayevich TIKHONOV. “On the stability of inverse problems”. Russian. In : *Doklady Akademii Nauk SSSR* 39.5 (1943), p. 195-198 (cf. p. [102](#)).
- [TA77] Andrey Nikolayevich TIKHONOV et V. ARSENIN. *Solutions of Ill-Posed Problems*. eng. Scripta Series in Mathematics. Washington : New York : Winston ; distributed solely by Halsted Press, 1977. ISBN : 978-0-470-99124-4 (cf. p. [102](#)).
- [Tim37] S. TIMOSHENKO. *Vibration Problems In Engineering*. eng. 2nd. Stanford University Palo Alto, CA, USA : D. Van Nostrand Company Inc., New-York, 1937 (cf. p. [63](#)).
- [TS67] K. E. TORRANCE et E. M. SPARROW. “Theory for Off-Specular Reflection From Roughened Surfaces\*”. EN. In : *JOSA* 57.9 (sept. 1967), p. 1105-1114. DOI : [10/dw6f9n](#) (cf. p. [69](#)).
- [Tso97] Hoi-Pang TSOI. “A Fuzzy Logic Controller (FLC) for Ram Velocity in Injection Molding”. English. Thesis. 1997 (cf. p. [18](#)).
- [Tyo+06] J. Scott TYO, Dennis L. GOLDSTEIN, David B. CHENAULT et Joseph A. SHAW. “Review of Passive Imaging Polarimetry for Remote Sensing Applications”. EN. In : *Applied Optics* 45.22 (août 2006), p. 5453-5469. ISSN : 2155-3165. DOI : [10/bx8tpw](#) (cf. p. [54](#)).

- [vH08] Laurens VAN DER MAATEN et Geoffrey HINTON. “Visualizing Data Using T-SNE”. In : *Journal of Machine Learning Research* 9.Nov (2008), p. 2579-2605. ISSN : ISSN 1533-7928 (cf. p. 120).
- [vW06] Nees Jan VAN ECK et Ludo WALTMAN. “VOS : A New Method for Visualizing Similarities between Objects”. en. In : *ERIM report series research in management Erasmus Research Institute of Management* (avr. 2006). ISSN : 1566-5283 (cf. p. 8).
- [van+10] Nees Jan VAN ECK, Ludo WALTMAN, Rommert DEKKER et Jan VAN DEN BERG. “A Comparison of Two Techniques for Bibliometric Mapping : Multidimensional Scaling and VOS”. en. In : *Journal of the American Society for Information Science and Technology* 61.12 (déc. 2010), p. 2405-2416. ISSN : 1532-2890. DOI : [10.1002/asi.21421](https://doi.org/10.1002/asi.21421) (cf. p. 8).
- [Vap92] V. VAPNIK. “Principles of Risk Minimization for Learning Theory”. In : *Advances in Neural Information Processing Systems* 4. Sous la dir. de J. E. MOODY, S. J. HANSON et R. P. LIPPMANN. Morgan-Kaufmann, 1992, p. 831-838 (cf. p. 93, 97).
- [VL63] Vladimir Naumovitch VAPNIK et Alexander Yakovlevich LERNER. “Patterns recognition using Generalized Portraits method”. Russian. In : *Automation and Remote Control* 24.6 (1963), p. 774-780 (cf. p. 89).
- [VGS97] Vladimir VAPNIK, Steven E. GOLOWICH et Alex J. SMOLA. “Support Vector Method for Function Approximation, Regression Estimation and Signal Processing”. In : *Advances in Neural Information Processing Systems* 9. Sous la dir. de M. C. MOZER, M. I. JORDAN et T. PETSCHE. MIT Press, 1997, p. 281-287 (cf. p. 89).
- [Vas+17] Ashish VASWANI, Noam SHAZEER, Niki PARMAR, Jakob USZKOREIT, Llion JONES, Aidan N. GOMEZ, Lukasz KAISER et Illia POLOSUKHIN. “Attention Is All You Need”. In : *arXiv :1706.03762 [cs]* (juin 2017). arXiv : [1706.03762 \[cs\]](https://arxiv.org/abs/1706.03762) (cf. p. 128).
- [von07] Ulrike VON LUXBURG. “A Tutorial on Spectral Clustering”. In : *arXiv :0711.0189 [cs]* (nov. 2007). arXiv : [0711.0189 \[cs\]](https://arxiv.org/abs/0711.0189) (cf. p. 114).
- [Wan+17] Jian WANG, Feng ZHOU, Shilei WEN, Xiao LIU et Yuanqing LIN. “Deep Metric Learning with Angular Loss”. In : *arXiv :1708.01682 [cs]* (août 2017). arXiv : [1708.01682](https://arxiv.org/abs/1708.01682) (cf. p. 110).
- [Wan+14] Jiang WANG, Yang SONG, Thomas LEUNG, Chuck ROSENBERG, Jinbin WANG, James PHILBIN, Bo CHEN et Ying WU. “Learning Fine-grained Image Similarity with Deep Ranking”. In : *arXiv :1404.4661 [cs]* (avr. 2014). arXiv : [1404.4661](https://arxiv.org/abs/1404.4661) (cf. p. 108, 110).
- [Wan+04] Zhou WANG, A. C. BOVIK, H. R. SHEIKH et E. P. SIMONCELLI. “Image Quality Assessment : From Error Visibility to Structural Similarity”. In : *IEEE Transactions on Image Processing* 13.4 (avr. 2004), p. 600-612. ISSN : 1057-7149. DOI : [10.1109/TIP.2003.819861](https://doi.org/10.1109/TIP.2003.819861) (cf. p. 50).
- [WBS06] Kilian Q WEINBERGER, John BLITZER et Lawrence K. SAUL. “Distance Metric Learning for Large Margin Nearest Neighbor Classification”. English. In : *Advances in Neural Information Processing Systems* 18. Sous la dir. d’Y. WEISS, B. SCHÖLKOPF et J. C. PLATT. Vancouver, Canada : MIT Press, 2006, p. 1473-1480 (cf. p. 108).
- [Wil+05] Bennett WILBURN, Neel JOSHI, Vaibhav VAISH, Eino-Ville TALVALA, Emilio ANTUNEZ, Adam BARTH, Andrew ADAMS, Mark HOROWITZ et Marc LEVOY. “High Performance Imaging Using Large Camera Arrays”. In : *ACM SIGGRAPH 2005 Papers. SIGGRAPH ’05*. New York, NY, USA : ACM, 2005, p. 765-776. DOI : [10.b7wkxc](https://doi.org/10.b7wkxc) (cf. p. 56).

- [WA95] Lawrence B WOLFF et Andreas G ANDREOU. “Polarization Camera Sensors”. In : *Image and Vision Computing* 13.6 (août 1995), p. 497-510. ISSN : 0262-8856. DOI : [10.1016/0262-8856\(95\)94383-B](https://doi.org/10.1016/0262-8856(95)94383-B) (cf. p. 54).
- [WC97] Suzanne L. B. WOLL et Douglas J. COOPER. “Pattern-Based Closed-Loop Quality Control for the Injection Molding Process”. en. In : *Polymer Engineering & Science* 37.5 (mai 1997), p. 801-812. ISSN : 1548-2634. DOI : [10.1002/pen.11723](https://doi.org/10.1002/pen.11723) (cf. p. 17).
- [WCS96] Suzanne L. B. WOLL, Douglas J. COOPER et Blair V. SOUDER. “Online Pattern-Based Part Quality Monitoring of the Injection Molding Process”. en. In : *Polymer Engineering & Science* 36.11 (mai 1996), p. 1477-1488. ISSN : 1548-2634. DOI : [10.1002/pen.10542](https://doi.org/10.1002/pen.10542) (cf. p. 21).
- [Wu+17] Chao-Yuan WU, R. MANMATHA, Alexander J. SMOLA et Philipp KRÄHENBÜHL. “Sampling Matters in Deep Embedding Learning”. In : *arXiv :1706.07567 [cs]* (juin 2017). arXiv : [1706.07567](https://arxiv.org/abs/1706.07567) (cf. p. 108).
- [Wu+16] Jiajun WU, Chengkai ZHANG, Tianfan XUE, William T. FREEMAN et Joshua B. TENENBAUM. “Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling”. In : *arXiv :1610.07584 [cs]* (oct. 2016). arXiv : [1610.07584 \[cs\]](https://arxiv.org/abs/1610.07584) (cf. p. 52).
- [Xie+16] Saining XIE, Ross GIRSHICK, Piotr DOLLÁR, Zhuowen TU et Kaiming HE. “Aggregated Residual Transformations for Deep Neural Networks”. In : *arXiv :1611.05431 [cs]* (nov. 2016). arXiv : [1611.05431 \[cs\]](https://arxiv.org/abs/1611.05431) (cf. p. 97).
- [Xu+15] Kelvin XU, Jimmy BA, Ryan KIROS, Kyunghyun CHO, Aaron COURVILLE, Ruslan SALAKHUTDINOV, Richard ZEMEL et Yoshua BENGIO. “Show, Attend and Tell : Neural Image Caption Generation with Visual Attention”. In : *arXiv :1502.03044 [cs.LG]* (fév. 2015). arXiv : [1502.03044 \[cs.LG\]](https://arxiv.org/abs/1502.03044) (cf. p. 128).
- [Yam+16] Tomohiro YAMAZAKI, Yasushi MARUYAMA, Yusuke UESAKA, Motoaki NAKAMURA, Yoshihisa MATOBA, Takashi TERADA, Kenta KOMORI, Yoshiyuki OHBA, Shinichi ARAKAWA, Yasutaka HIRASAWA, Yuhi KONDO, Jun MURAYAMA, Kentaro AKIYAMA, Yusuke OIKE, Shuzo SATO et Takayuki EZAKI. “Four-Directional Pixel-Wise Polarization CMOS Image Sensor Using Air-Gap Wire Grid on 2.5-Mm Back-Illuminated Pixels”. In : *2016 IEEE International Electron Devices Meeting (IEDM)* (2016), p. 8-8. DOI : [10/gf8zsc](https://doi.org/10.1109/iedm.2016.7800002) (cf. p. 54).
- [YDK00] Dongzhe YANG, Kourosh DANAI et David Owen KAZMER. “A Knowledge-Based Tuning Method for Injection Molding Machines”. In : *Journal of Manufacturing Science and Engineering* 123.4 (déc. 2000), p. 682-691. ISSN : 1087-1357. DOI : [10.1115/1.1382596](https://doi.org/10.1115/1.1382596) (cf. p. 22).
- [Yan04] Yi YANG. “Injection Molding Control : From Process to Quality”. eng. PhD Thesis. Hong Kong University of Science et Technology, jan. 2004 (cf. p. 21).
- [YRC07] Yuan YAO, Lorenzo ROSASCO et Andrea CAPONNETTO. “On Early Stopping in Gradient Descent Learning”. en. In : *Constructive Approximation* 26.2 (août 2007), p. 289-315. ISSN : 1432-0940. DOI : [10/d3qqdj](https://doi.org/10.1007/s00365-006-0303-0) (cf. p. 100).
- [Yu+18] Baosheng YU, Tongliang LIU, Mingming GONG, Changxing DING et Dacheng TAO. “Correcting the Triplet Selection Bias for Triplet Loss”. In : 2018, p. 71-87 (cf. p. 110).
- [ZF13] Matthew D. ZEILER et Rob FERGUS. “Visualizing and Understanding Convolutional Networks”. In : *arXiv :1311.2901 [cs]* (nov. 2013). arXiv : [1311.2901 \[cs\]](https://arxiv.org/abs/1311.2901) (cf. p. 93, 108).

- [Zha+18] Han ZHANG, Ian GOODFELLOW, Dimitris METAXAS et Augustus ODENA. “Self-Attention Generative Adversarial Networks”. In : *arXiv :1805.08318 [cs, stat]* (mai 2018). arXiv : [1805.08318 \[cs, stat\]](https://arxiv.org/abs/1805.08318) (cf. p. 128).
- [ZDN16] Hang ZHANG, Kristin DANA et Ko NISHINO. “Friction from Reflectance : Deep Reflectance Codes for Predicting Physical Surface Properties from One-Shot In-Field Reflectance”. In : *arXiv :1603.07998 [cs]* (mar. 2016). arXiv : [1603.07998 \[cs\]](https://arxiv.org/abs/1603.07998) (cf. p. 69).
- [Zha+16] Yun ZHANG, Ting MAO, Zhigao HUANG, Huang GAO et Dequn LI. “A Statistical Quality Monitoring Method for Plastic Injection Molding Using Machine Built-in Sensors”. en. In : *The International Journal of Advanced Manufacturing Technology* 85.9-12 (août 2016), p. 2483-2494. ISSN : 0268-3768, 1433-3015. DOI : [10.1007/s00170-015-8013-2](https://doi.org/10.1007/s00170-015-8013-2) (cf. p. 19).
- [Zhe+19] Wenzhao ZHENG, Zhaodong CHEN, Jiwen LU et Jie ZHOU. “Hardness-Aware Deep Metric Learning”. In : *arXiv :1903.05503 [cs]* (mar. 2019). arXiv : [1903.05503](https://arxiv.org/abs/1903.05503) (cf. p. 112).
- [Zho+15] Bolei ZHOU, Aditya KHOSLA, Agata LAPEDRIZA, Aude OLIVA et Antonio TORRALBA. “Learning Deep Features for Discriminative Localization”. In : *arXiv :1512.04150 [cs]* (déc. 2015). arXiv : [1512.04150 \[cs\]](https://arxiv.org/abs/1512.04150) (cf. p. 102).
- [Zho+03] Dengyong ZHOU, Olivier BOUSQUET, Thomas Navin LAL, Jason WESTON et Bernhard SCHÖLKOPF. “Learning with Local and Global Consistency”. In : *Proceedings of the 16th International Conference on Neural Information Processing Systems*. NIPS’03. event-place : Whistler, British Columbia, Canada. Cambridge, MA, USA : MIT Press, 2003, p. 321-328 (cf. p. 105).
- [Zhu+17] Jun-Yan ZHU, Taesung PARK, Phillip ISOLA et Alexei A. EFROS. “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks”. In : *arXiv :1703.10593 [cs]* (mar. 2017). arXiv : [1703.10593 \[cs\]](https://arxiv.org/abs/1703.10593) (cf. p. 49).
- [ZG02] Xiaojin ZHU et Zoubin GHAHRAMANI. *Learning from Labeled and Unlabeled Data with Label Propagation*. Rapp. tech. 2002 (cf. p. 105).
- [Zis72] Michel ZISMAN. *Topologie algébrique élémentaire : Maîtrise de mathématiques*. français. Paris, France : Armand Colin, 1972 (cf. p. 122).
- [ZL16] Barret ZOPH et Quoc V. LE. “Neural Architecture Search with Reinforcement Learning”. In : *arXiv :1611.01578 [cs]* (nov. 2016). arXiv : [1611.01578 \[cs\]](https://arxiv.org/abs/1611.01578) (cf. p. 137).
- [Zop+17] Barret ZOPH, Vijay VASUDEVAN, Jonathon SHLENS et Quoc V. LE. “Learning Transferable Architectures for Scalable Image Recognition”. In : *arXiv :1707.07012 [cs, stat]* (juil. 2017). eprinttype : arxiv eprintclass : cs.CV, cs.LG, stat.ML eprint : <http://arxiv.org/abs/1707.07012>. arXiv : [1707.07012 \[cs, stat\]](https://arxiv.org/abs/1707.07012) (cf. p. 137).



# Table des figures

1.1	Cartographie bibliographique de la maîtrise du procédé d'injection-moulage. . . . .	8
1.2	Schéma-bloc du procédé d'injection-moulage. . . . .	10
1.3	Vue systémique du procédé d'injection-moulage, adaptée de KAZMER et HATCH [KH99]. . . . .	11
1.4	Chronogramme du cycle d'injection-moulage. . . . .	11
1.5	Représentation Zig Zag du procédé d'injection-moulage. . . . .	12
1.6	Vers le pilotage du point de fonctionnement du procédé pour optimiser la qualité du produit. . . . .	14
1.7	Cartographie bibliographique de la maîtrise du procédé d'injection-moulage. . . . .	24
1.8	Figure issue du rapport annuel de la DIRECTION GÉNÉRALE DES ENTREPRISES : <i>Chiffres clés de l'industrie manufacturière, Édition 2018</i> [Dir19] . . . . .	26
1.9	Intégration de la mesure dans le cycle d'injection-moulage. . . . .	28
2.1	Défaut de retassure mesurée par microscope confocal, <i>AltiMet 520<sup>0</sup></i> . . . . .	33
2.2	Intégration du contrôle de la qualité . . . . .	34
2.3	Diagramme de Venn d'un dispositif de mesure adapté. . . . .	35
2.4	Mesures de variables physiques dans l'outillage. . . . .	36
2.5	Résultats du modèle LSTM sur le jeu de données de test. . . . .	37
2.6	Modèle linéaire entre les caractéristiques des pièces chaudes (30 secondes après production) et des pièces stabilisées (3 jours après production), Logiciel Elistat <sup>6</sup> . . . . .	42
2.7	Photographie de l'aspect des pièces. . . . .	43
2.8	Mesure de la géométrie d'un objet par triangulation laser. Figure dérivée du travail original de Georg Wiora (Dr. Schorsch), CC BY-SA 3.0, via Wikimedia Commons. . . . .	44
2.9	Mesure de la géométrie d'un objet par triangulation laser en cycle industriel. . . . .	45
2.10	Nuage de points de la pièce obtenu en cycle industriel. . . . .	45
2.11	Écart géométrique entre la pièce mesurée et la géométrie de référence de l'outillage. . . . .	46
2.12	Mesure thermique et mesure géométrique d'une pièce thermo-moulée. . . . .	48
2.13	Architecture du réseau antagoniste génératif <i>U-Net</i> , utilisée par le modèle <i>pix2pix</i> . . . . .	49
2.14	Présentation d'un échantillon du jeu de données, en fausse couleur. . . . .	50
2.15	Comparaison entre les images générées par le modèle à partir de l'image thermique et les images "géométriques" originales. . . . .	51
2.16	Différentes composantes de réflexion d'un polymère semi-cristallin. Figure inspirée de [Deb+00]. . . . .	53
2.17	Acquisition des composantes de réflexions spéculaire et diffuse. . . . .	53
2.18	Notre polarimètre et le procédé de calibration par homographie. . . . .	55
2.19	Détection et association de marqueurs <i>ChAruco</i> entre deux caméras. . . . .	55
2.20	Images de la même pièce par les 3 caméras, sous 3 angles de polarisation différents. . . . .	56
2.21	Reconstruction des paramètres de Stokes partiels, du Degré et de l'Angle de Polarisation, à partir des images de 3 caméras. . . . .	58
2.22	Illustration de différentes méthodes de pré-traitements des images issues des trois caméras polarisées. . . . .	61
2.23	Représentation des 100 premiers modes de notre base <i>B</i> . . . . .	64

2.24	Filtrage de la photographie d'une pièce par un mode de la base $B$ . . . . .	65
2.25	Base modale moyenne obtenue par sélection des 27 modes par validation croisée. Reconstruction d'une image d'une pièce conforme et d'une image d'une pièce non-conforme, après description modale. . . . .	66
2.26	Importance des descripteurs par Forêt d'arbres décisionnels. . . . .	66
2.27	Mesure par caméra thermique de l'évolution sur 2 secondes de la température d'une pièce. . . . .	69
3.1	Diagramme du processus de modélisation de la qualité par apprentissage. . . . .	74
3.2	Annotation hors-ligne par les experts humains. . . . .	80
3.3	Courbe ROC du meilleur classifieur de la qualité des pièces. . . . .	87
3.4	<i>Stratified 4-Folds.</i> Figure extraite de la doc <i>SciKit-Learn</i> , License BSD 3-clauses.	88
3.5	SVM avec <i>kernel trick</i> $\varphi$ . Figure originale de Alisneaky © 0, Wikimedia Commons.	90
3.6	Architecture d'un bloc DenseNet. . . . .	97
3.7	Fonctions d'activation. . . . .	99
3.8	Comparaison <i>max pooling</i> c et <i>average pooling</i> d pour 4 filtres de convolutions b.	101
3.9	Augmentation d'une image par translations et rotations. . . . .	103
3.10	Architecture d'un réseau siamois. . . . .	107
3.11	Trois différents types d'images Négatives possibles pour un couple d'images (Référence, Positive). Figure inspirée de l'article de blog de O. Moindrot. . . . .	109
3.12	Principe de l'apprentissage d'un réseau de triplets. Figure inspirée de [SKP15]. .	110
3.13	Projection par <i>t-SNE</i> à partir des valeurs de la dernière couche d'un réseau de triplet. L'échelle 1-5 correspond aux niveaux de qualité évalués par les experts humains ; -1 correspond aux pièces non annotées. . . . .	111
3.14	Projection par <i>UMAP</i> à partir de la valeur des pixels des images. L'échelle -1 à 4 correspond aux clusters obtenus par <i>HDBSCAN</i> . . . . .	114
3.15	Compression par <i>ACP</i> . . . . .	116
3.16	Qualité de la compression par <i>ACP</i> en fonction du nombre de composantes. . . .	116
3.17	Projection par <i>ACP</i> à partir de la valeur des pixels des images. L'échelle 1-5 correspond aux niveaux de qualité évalués par les experts humains ; -1 correspond aux pièces non annotées. . . . .	117
3.18	Optimisation des paramètres d'une <i>ACP</i> à noyaux. Métrique : Norme $\ell_2$ . . . . .	118
3.19	Compression par <i>ACP</i> à noyau. . . . .	119
3.20	Qualité de la compression par <i>ACP à noyau</i> en fonction du nombre de composantes.	119
3.21	Projection par <i>t-SNE</i> à partir de la valeur des pixels des images. L'échelle 1-5 correspond aux niveaux de qualité évalués par les experts humains ; -1 correspond aux pièces non annotées. . . . .	121
3.22	Représentation par <i>UMAP</i> à partir de la valeur des pixels des images. L'échelle 1-5 correspond aux niveaux de qualité évalués par les experts humains ; -1 correspond aux pièces non annotées. . . . .	122
3.23	Architecture d'un auto-encodeur. . . . .	123
3.24	Reconstruction d'une image par un auto-encodeur. . . . .	124
3.25	Projection par <i>t-SNE</i> à partir des valeurs de l'espace latent $z$ d'un <i>VAE</i> . L'échelle 1-5 correspond aux niveaux de qualité évalués par les experts humains ; -1 correspond aux pièces non annotées. . . . .	125
3.26	Graphe d'un auto-encodeur variationnel, inspiré de [Doe16]. . . . .	125
3.27	Génération d'images à partir de la distribution modélisant l'espace latent d'un auto-encodeur variationnel. . . . .	126
3.28	Génération d'images à partir de la distribution modélisant l'espace latent du générateur d'un <i>GAN</i> . . . . .	129

3.29 Optimisation d'une fonction à deux variables avec neuf essais. Figure reproduite de [BB12]. . . . .	130
3.30 Essais prévus pour des plans d'expériences à trois niveaux pour trois variables. . . . .	131
3.31 Optimisation d'un modèle appris de manière supervisée, à l'aide de la librairie <i>auto-sklearn</i> [Feu+15]. . . . .	134
3.32 Évaluation partielle de $f$ en fonction de $b$ . . . . .	135
3.33 <i>fANOVA</i> des hyper-paramètres suite à l'optimisation avec <i>BOHB</i> . . . . .	136
3.34 Taxonomie de l'apprentissage statistique. . . . .	140



# Liste des tableaux

2.1	Principaux défauts d'aspects en injection-moulage des thermoplastiques . . . . .	32
2.2	Contraintes industrielles pour la contrôle en ligne de la qualité. . . . .	35
2.3	Moyens de mesure des caractéristiques d'une pièce. . . . .	40
2.4	Degré d'invasivité de la mesure sur le procédé d'injection-moulage des thermoplastiques. . . . .	41
2.5	Performance du modèle pix2pix : similarité entre images générées et images originales. . . . .	50
2.6	Comparaison des scores de justesse de classification de la non-conformité. . . . .	59
3.1	Méthodes de construction du jeu de données d'apprentissage. . . . .	75
3.2	Sélection des variables pour l'étude. . . . .	76
3.3	Plan de criblage $L12$ utilisé. . . . .	77
3.4	Durées d'annotations moyenne d'une image pour un humain. . . . .	79
3.5	Matrice de confusion. . . . .	86
3.6	Répartition du jeu de données : apprentissage, validation et test. . . . .	88
3.7	Hyper-paramètres critiques pour l'apprentissage de réseaux de neurones profonds.	95
3.8	Architecture des réseaux de convolutions étudiés. . . . .	96
3.9	Fonctions d'activation. . . . .	98
3.10	Illustration d'une couche de <i>max pooling</i> . . . . .	100
3.11	Complexité de l'apprentissage des méthodes étudiées en fonction du nombre d'échantillons du jeu de données $N$ . . . . .	139



# A

## Annexe 1 : Titre

---

*Résumé de l'annexe ici.*

---

## A.1 Section 1

### A.1.1 Sous section 1

Qu'est que c'est?. C'est une phrase français avant le lorem ipsum. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdier mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

### A.1.2 Sous section 2

Qu'est que c'est?. C'est une phrase français avant le lorem ipsum. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdier mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

# B

## Annexe 2 : Titre

---

*Résumé ici.*

---

## B.1 Section 1

### B.1.1 Sous section 1

Qu'est que c'est?. C'est une phrase français avant le lorem ipsum. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdier mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

### B.1.2 Sous section 2

Qu'est que c'est?. C'est une phrase français avant le lorem ipsum. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdier mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

# C

## Annexe 3 : Titre

---

*Résumé ici.*

---

## C.1 Section 1

### C.1.1 Sous section 1

Qu'est que c'est?. C'est une phrase français avant le lorem ipsum. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdier mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

### C.1.2 Sous section 2

Qu'est que c'est?. C'est une phrase français avant le lorem ipsum. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdier mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.



# Contrôle automatique non-invasif de la Qualité des produits

Application au procédé d'injection-moulage des thermoplastiques

## Résumé

Le procédé d'injection moulage des thermoplastiques permet la production industrielle de produits, dont les exigences de qualités géométriques et d'aspects sont croissantes. Maîtriser la qualité du produit nécessite de maîtriser le procédé d'injection. Les contraintes posées par le respect du cycle industriel sont nombreuses. Ces travaux de doctorats s'inscrivent dans la thématique du contrôle des produits, spécifiquement intégré sur la ligne de production. Les technologies de mesures non invasives permettent le respect des contraintes industrielles, ainsi qu'une simplification de la mise en place et une mobilité dans l'atelier. L'apport de l'imagerie non-conventionnelle, avec la thermographie et la polarimétrie, est étudiée. La classification automatique des pièces selon leur qualité est effectuée à l'aide d'un modèle de la qualité de la pièce construit par apprentissage supervisé. Un dispositif de contrôle de la qualité des produits, dès la sortie du moule, est présenté. Ce dernier utilise l'information issue de multiples capteurs et l'apprentissage supervisé itératif du modèle, au fil de la production. Plusieurs cas d'études industriels permettent d'évaluer les performances et la viabilité de la solution retenue.

## Mots-clés

Qualité, Métrologie, Imagerie non-conventionnelle, Thermographie, Apprentissage Supervisé, Deep Learning

## Abstract

The injection molding process of thermoplastics allows the industrial production of products, whose requirements of geometric qualities and aspects are increasing. Mastering the quality of the product requires mastering the injection process. The constraints posed by the respect of the industrial cycle are numerous. This doctoral work is part of the product control theme, specifically integrated into the production line. Non-invasive measurement technologies allow compliance with industrial constraints, simplification of implementation and mobility in the workshop. The contribution of non-conventional imaging, with thermography and polarimetry, is studied. The automatic classification of parts according to their quality is carried out using a model of the quality of the part built by supervised learning. A device for checking the quality of the products, right out of the mold, is presented. The latter uses information from multiple sensors and the iterative supervised learning of the model, during production, during the process adjustment phase. Several cases of industrial studies make it possible to evaluate the performance and the viability of the chosen solution.

**Keywords :** Quality, Metrology, Unconventional imaging, Thermography, Supervised learning, Deep Learning

