

Master 2
Observation de la Terre et Géomatique
Année 2017 - 2018

**L'apport des méthodes d'apprentissage
profond pour l'exploitation des photographies
aériennes anciennes**

Développement de techniques de colorisation automatique

Quentin Poterek
Septembre 2018

Structure d'accueil	Faculté de Géographie et d'Aménagement Université de Strasbourg 3, rue de l'Argonne - 67000 Strasbourg
Maître de stage	Pierre-Alexis HERRAULT, Maître de conférences Laboratoire Image, Ville, Environnement UMR 7362 CNRS - Université de Strasbourg 3, rue de l'Argonne - 67000 Strasbourg
Tutrice universitaire	Anne PUISSANT, Maître de conférences HDR Laboratoire Image, Ville, Environnement UMR 7362 CNRS - Université de Strasbourg 3, rue de l'Argonne - 67000 Strasbourg



Remerciements

J'exprime ici toute ma gratitude aux personnes qui ont participé, de près ou de loin, à la réalisation de ce travail de recherche.

Mes remerciements vont tout d'abord à Grzegorz Skupinski qui s'est donné beaucoup de mal pour m'obtenir ce stage et a rendu ce travail possible. Missa remercie big boss pour soutien et écoute. Les choses auraient été autrement plus difficiles sans sa présence ; son t-shirt "Père-fect" est tout à son honneur. Je remercie également Pierre-Alexis Herrault pour son encadrement et nos entrevues quasi-quotidiennes. Merci pour sa confiance dans le cadre de ce travail. Outre ses conseils, j'ai été très heureux de pouvoir apprendre à le connaître, et ce déjà depuis le début de l'année.

Au groupe de la Zone Atelier Environnemental Urbain, qui a accepté de financer ce projet, et plus particulièrement à Nadège Blond et Sandrine Glatron.

Aux différentes personnes qui m'ont écouté, aidé puis épaulé dans la recherche d'un stage, notamment Dominique Badariotti, Thierry Rosique, Dominique Schwartz, Alain Clappier et Adine Hector. Leur soutien m'aura permis d'avancer plus sereinement vers ce deuxième semestre.

Je remercie également l'ensemble de l'équipe enseignante de la Faculté de Géographie et d'Aménagement de Strasbourg. Ces cinq dernières années auront été certes éprouvantes, mais avant tout très intéressantes, du fait de chacun et chacune. Je ne m'imaginais pas un jour pouvoir trouver une passion, et cela a finalement été le cas avec la géomatique. Je pense pour cela plus particulièrement à Aziz Serradj, qui nous a initiés à la cartographie et à la télédétection. Merci pour son intégrité, sa rigueur et ses *punchlines* parfois étonnantes. Une parmi tant d'autres : "Vous et les connaissances, c'est un peu comme quand on égoutte des pâtes dans une passoire... Sauf que les pâtes partent avec l'eau, n'est-ce pas ?". Merci à Anne Puissant, avec laquelle j'ai beaucoup appris. Parmi les points qui m'auront marqués se trouvent sa bonne humeur, son humour, et le soutien qu'elle a pu apporter à de nombreuses personnes dans la promotion. Enfin, je tiens à remercier Arnaud Piombini, qui n'est certes pas géomaticien, mais que j'ai toujours beaucoup apprécié pour son caractère et son sarcasme bien pesé. Je n'oublierai pas les quelques mots que nous avons échangés, ses blagues, et l'aide procurée en L3 lorsque je cherchais à découvrir le monde de la recherche.

Mes remerciements au personnel de la faculté et du LIVE, pour leur aide et/ou les discussions

Remerciements

que nous avons eues, notamment Ali Saidi, Taraneh Saidi et Estelle Baehrel.

Ce stage et ces dernières années n'auraient également pas été les mêmes sans mes ami-e-s et camarades. Tout d'abord, un grand merci à Anaïs, Émilie, Léa, Louis, Morgane, Paul et Ronan pour le cadeau qu'ils m'ont offert au cours du stage, cela m'a beaucoup touché.

Toute ma gratitude va aux promotions de Licence et de Master. A Lucie, qui m'accompagne depuis la L1, et avec laquelle nous nous sommes soutenus mutuellement à de nombreuses reprises. A Aurore et Hugo, pour nos mardis burger et soirées jeux de société. Le GEO-Lab aura su renforcer le lien qui existe entre nous, Aurore. L'adversité est effectivement un terreau propice au lien social. Tous deux avez ma gratitude pour le rituel que vous avez mené au musée vaudou, puisse-t-il se concrétiser.

Mes pensées vont également à mes camarades de stage. A Émilie, pour nos discussions fréquentes et interminables, mais si intéressantes. Merci d'avoir essayé de me donner du courage là où je n'en avais pas. A Ronan, pour son enthousiasme (*Allez Brest!*) et sa bonne humeur, et qui a également réussi à supporter mes taquineries incessantes. A Bruno, Morgane, Louis, Anaïs et Xuefei qui ont parfois participé aux discussions un peu trop fréquentes dans la salle Master.

A Berge, Alexis pour les intimes, qui sait toujours stimuler mon intellect et mon imagination. Ces dernières années n'auraient pas eu la même saveur sans lui ni son humour (merci de ne pas m'avoir jugé pour mon *Tumblr*). A mon Paul, "wow", qui m'a accompagné dans bien des aventures, à la fois réelles et virtuelles. Je suis content d'avoir pu trouver une personne avec laquelle je partage tant de centres d'intérêt. Ses périples sur R auront définitivement su divertir la populace de la salle Master. A Mathou, Mathilde pour les non intimes, qui même si elle est lyonnaise sait aujourd'hui prononcer fièrement "Wacken" et autres localités alsaciennes. Le bretzel n'est malheureusement pas d'origine monégasque, c'est aujourd'hui pardonné mais pas oublié. Merci pour tous les bons moments passés ensemble, les repas et soirées jeux de société chez elle, ses conseils souvent avisés... Sa personne d'une manière générale. A Léa, qui m'aura aidé dans des instants où je me sentais isolé. Merci pour toutes les onomatopées qu'elle m'a enseigné, ce sont aujourd'hui des compétences que j'espère pouvoir valoriser dans le monde du travail. Merci pour nos discussions enrichissantes et son ouverture d'esprit.

Merci à mes amies de longue date, Colette, Laura et Martine, qui auront appris à me supporter. Les temps du lycée sont lointains, mais nous avons malgré tout réussi à préserver cette belle relation.

Enfin, toute ma gratitude va à ma mère, mon frère et mon oncle, qui m'auront toujours soutenu et sans qui je n'en serais pas là aujourd'hui.

Acronymes

ACP	Analyse en Composantes Principales
API	<i>Application Programming Interface</i> , Interface de programmation
BEGAN	<i>Boundary Equilibrium Generative Adversarial Network</i>
C, c	Image ou photographie en couleurs
cGAN	<i>Conditional Generative Adversarial Network</i> , Réseau génératif antagoniste conditionnel
CNN	<i>Convolutional Neural Network</i> , Réseau de neurones convolutif
CPU	<i>Central Processing Unit</i> , Unité centrale de traitement
CUS	Communauté Urbaine de Strasbourg
D	Discriminateur
dB	Décibel
DCGAN	<i>Deep Convolutional Generative Adversarial Network</i> , Réseau génératif antagoniste convolutif profond
DNN	<i>Deep Neural Network</i> , Réseau de neurones profond
DRAGAN	<i>Deep Regret Analytic Generative Adversarial Network</i>
EMS	Eurométropole de Strasbourg
G	Générateur
GAN	<i>Generative Adversarial Network</i> , Réseau génératif antagoniste
GPU	<i>Graphics Processing Unit</i> , Processeur graphique
IRC, I	Image en composition colorée infrarouge-couleur
IS	<i>Inception Score</i>
LBP	<i>Local Binary Patterns</i> , Motifs binaires locaux
ReLU	<i>Rectified Linear Unit</i>
lr	<i>Learning Rate</i> , Taux d'apprentissage
LIVE	Laboratoire Image, Ville et Environnement
MS	Multi-spectral

Acronymes

MSE	<i>Mean Square Error</i> , Erreur quadratique moyenne
PAN, P, p	Image ou photographie (pseudo-)panchromatique
PSNR	<i>Peak Signal-to-Noise Ratio</i>
RF	<i>Random Forest</i> , Forêt aléatoire
SGD	<i>Stochastic Gradient Descent</i> , Algorithme du gradient stochastique
SIG	Système d'Information Géographique
SSIM	<i>Structural Similarity Index Measure</i>
THRS	Très Haute Résolution Spatiale
UMC	Unité Minimale de Cartographie
VAE	<i>Variational Autoencoder</i> , Autoencodeur variationnel
WGAN	<i>Wasserstein Generative Adversarial Network</i> , Réseau génératif antagoniste de Wasserstein
ZAEU	Zone Atelier Environnementale Urbaine

Glossaire

Afin de simplifier la lecture du mémoire, le vocabulaire associé aux méthodes d'apprentissage profond est ici défini et précisé. Ces termes ont été regroupés en trois catégories afin de faciliter leur compréhension, mais aussi de rendre compte des étapes auxquelles chaque concept est susceptible d'intervenir. A noter que les définitions proposées sont généralement issues de sources multiples qui sont les suivantes, par ordre d'importance : Goodfellow *et al.* (2015), Ronneberger *et al.* (2015), Simonyan et Zisserman (2014), Paszke *et al.* (2017), Ioffe et Szegedy (2015), Radford *et al.* (2015), Hariharan *et al.* (2014). D'autres sources moins académiques ont également été utilisées : Howard *et al.* (2018) et Chintala *et al.* (2016). Les références ne sont renseignées que lorsqu'une seule a permis de définir un terme en particulier.

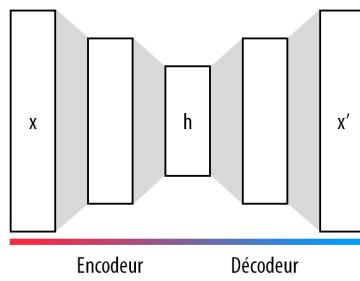
Catégories, sous-catégories de modèles et architectures

Architecture : L'architecture d'un modèle décrit les blocs, ou couches cachées, qui le constituent, ainsi que leur organisation en largeur et en profondeur. Pour une même catégorie de modèles, différentes architectures peuvent être adoptées selon les objectifs, le type de données, les méthodes d'optimisation retenues... Ces points sont détaillés dans les sections suivantes du glossaire.

Architecture entièrement convulsive : L'architecture d'un modèle définit la manière dont les entrées vont être manipulées par celui-ci. Dans le cas des produits matriciels en particulier, qui possèdent donc une dimension spatiale (pas nécessairement géographique), les architectures entièrement convolutives sont souvent privilégiées. Elles s'articulent uniquement autour de couches de convolution, qui permettent d'apprendre un corpus d'attributs discriminants. Contrairement à la plupart des architectures non convolutives, celles-ci sont capables de traiter des produits matriciels de n'importe quelle dimension, si tant est que suffisamment de mémoire est à disposition.

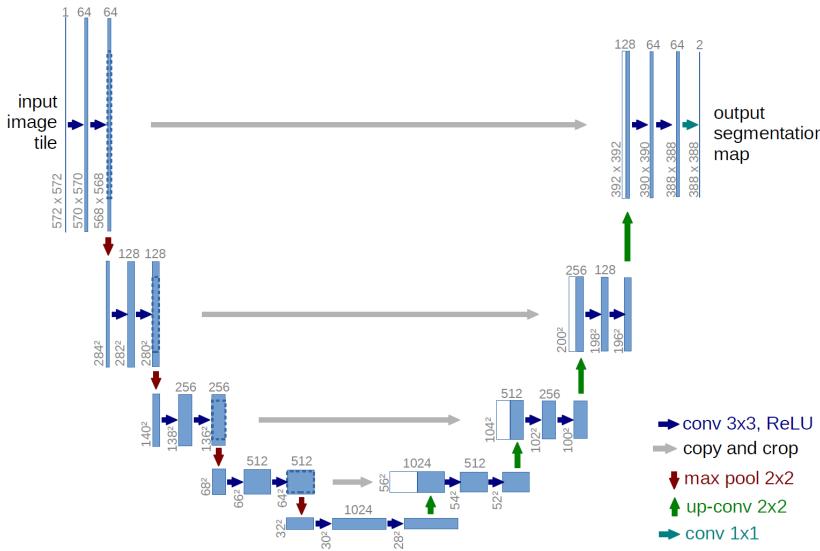
Autoencodeur : Un autoencodeur correspond à un réseau de neurones qui apprend, de façon non supervisée, à réduire la dimensionnalité de son entrée, puis à la restituer en sortie. Il s'organise en deux parties, avec (1) une fonction d'encodage $h = f(x)$ qui transforme x en une encapsulation ou code représenté par une couche cachée h , et (2) une fonction de décodage $x' = g(h)$ qui reconstitue x à partir de h . Les deux parties du réseau sont en général symétriques (Goodfellow *et al.*, 2015).

Glossaire



Exemple d'autoencodeur.

U-Net : Le réseau U-Net est un CNN organisé autour d'une architecture entièrement convective. Généralement utilisé pour la segmentation sémantique d'images, il est tout d'abord constitué d'un goulot d'étranglement qui réduit l'information spatiale à l'aide d'opérations de rééchantillonnage, tout en augmentant le corpus de sémantiques disponibles sur le produit. La deuxième partie du réseau est symétrique à la première, et restitue progressivement l'information spatiale, qu'il combine également aux attributs calculés précédemment, cela par concaténation (Ronneberger *et al.*, 2015).



Réseau U-Net utilisé pour la segmentation d'images médicales (*in* Ronneberger *et al.* (2015)).

VGG-16 : Le réseau VGG-16 est un CNN développé afin de classifier des images issues du dépôt ImageNet, ainsi que de localiser dans celles-ci des objets en particulier. Il est constitué de 16 couches cachées, avec 13 convolutions et 3 nœuds entièrement connectés. En sortie de celles-ci, l'utilisateur-trice obtient les probabilités d'appartenance de l'image à une ou plusieurs des mille classes proposées par ImageNet (Simonyan et Zisserman, 2014).

Structure des modèles et couches cachées

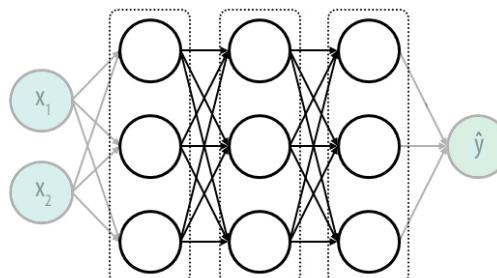
Activation : Une activation correspond à la sortie d'une couche cachée.

Batch Normalization (normalisation de batch) : La normalisation de batch est une technique utilisée directement après une couche, généralement convective. Elle prend ainsi l'ensemble des activations, puis ramène leur moyenne à 0 et variance à 1 (Ioffe et Szegedy, 2015). Cela permet de stabiliser l'apprentissage et même de faire fonctionner certains modèles, comme les GANs convolutifs (Radford *et al.*, 2015).

Couche cachée : Un réseau de neurones peut être constitué d'une ou plusieurs couches cachées, qui contiennent les nœuds n'appartenant ni aux entrées, ni aux sorties du modèle.

Couche convulsive : Une couche convulsive s'organise autour d'un noyau de convolution, ou fenêtre mouvante, généralement carrée et qui possède une taille fixe w . Outre w , d'autres hyperparamètres la définissent, principalement le nombre d'attributs appris f , le *stride* s et le *padding* p . Les paramètres d'une couche convulsive sont partagés et leur nombre est égal à $w^2 \times c \times f$, avec c le nombre d'attributs de la couche en entrée (Simonyan et Zisserman, 2014).

Couche entièrement connectée : Une couche entièrement connectée c_i est constituée d'un certain nombre de neurones, chacun étant connecté à l'ensemble des neurones des couches c_{i-1} et c_{i+1} qui lui sont directement adjacentes. Le nombre de paramètres à apprendre est donc égal au nombre de ces connexions, soit le produit de l'ensemble des dimensions.



Exemple d'un réseau avec trois couches cachées entièrement connectées.

Dropout : Le *dropout* consiste à éliminer aléatoirement un certain nombre de neurones situés dans une couche cachée du modèle. C'est une méthode de régularisation qui prévient le sur-apprentissage et améliore généralement les performances.

Fonction d'activation : Primitivement, les réseaux de neurones sont des modèles linéaires, qui apprennent une forme triviale de $f(x; \theta)$. Dans de nombreux problèmes, il est nécessaire d'utiliser des fonctions non linéaires afin d'apprendre des représentations adéquates. Pour répondre à ce problème, il existe différentes fonctions, dites d'activation, qui prennent les sorties issues d'une couche cachée, puis les transforment. Leur choix repose sur les besoins de l'étude, le type de données, la catégorie de modèle utilisée...

Glossaire

Dans le cadre de ce travail, trois fonctions d'activation ont été retenues. La première correspond à LeakyReLU, qui a été développée à partir de la fonction ReLU. Cette dernière transforme l'ensemble des activations négatives en 0, et peut donc entraîner la mort de certains neurones, qui ne sont alors plus mis à jour. Pour répondre à ce constat, la fonction LeakyReLU a été développée, et permet de transformer les activations négatives en des valeurs proches de 0. La seconde fonction correspond à tanh. Elle est ici utilisée spécifiquement pour transformer les activations de telle sorte à ce qu'elles restent dans l'intervalle $[-1, 1]$, indispensable pour permettre le passage d'un espace colorimétrique à un autre. La dernière correspond à la fonction sigmoïde, qui renvoie une probabilité comprise dans l'intervalle $[0, 1]$, et qui renseigne ici sur la plausibilité d'une colorisation. Ces fonctions sont décrites par les équations ci-dessous.

$$\text{LeakyReLU} \rightarrow f(x) = \max(0, 0.01x; x)$$

$$\text{Tanh} \rightarrow f(x) = \tanh(x)$$

$$\text{Sigmoïde} \rightarrow f(x) = \frac{1}{1 + e^{-x}}$$

Hypercolonne : Une hypercolonne consiste à concaténer en profondeur, pour chaque pixel d'une image, l'ensemble des attributs calculés par les couches cachées d'un modèle. Une telle structure permet ainsi de conserver l'information spatiale présente au début du réseau, et de profiter de la richesse sémantique décrite par les neurones plus profonds (Hariharan *et al.*, 2014).

Kernel size (taille de noyau) : La taille de noyau correspond aux dimensions, en longueur et en largeur, de la fenêtre utilisée pour réaliser une convolution.

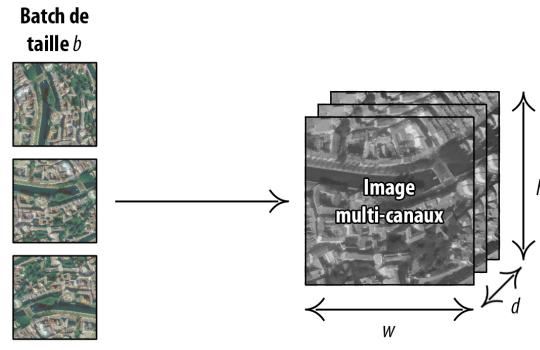
Neurone : Unité élémentaire dans un réseau de neurones, il prend une ou plusieurs entrées pondérées par des paramètres θ appris lors de la phase d'entraînement du modèle, puis génère une sortie.

Padding (remplissage) : Le remplissage permet de spécifier la manière dont sont gérées les bordures d'une image lorsqu'elle est passée à une couche convulsive dans le réseau. Il est par exemple possible d'ajouter des pixels nuls autour du produit, qui permettent de maintenir sa taille initiale au fil des convolutions.

Stride (pas) : Le pas d'une convolution correspond au nombre de pixels sur lequel s'effectue le déplacement de la fenêtre mouvante.

Tenseur : Un tenseur est une matrice multi-dimensionnelle dans laquelle sont stockés des objets possédant un seul type de données (Paszke *et al.*, 2017). Un tenseur contenant les images utilisées pour la colorisation est de la forme (b, d, w, h) et possède donc quatre dimensions. La première correspond à la taille du batch b qui est passé au réseau de neurones. Elle fait partie des hyperparamètres du modèles et peut donc influencer les résultats obtenus en sortie. En effet, la taille de batch conditionne la vitesse d'apprentissage, et permet généralement une meilleure optimisation et généralisation du modèle. La seconde dimension correspond à la

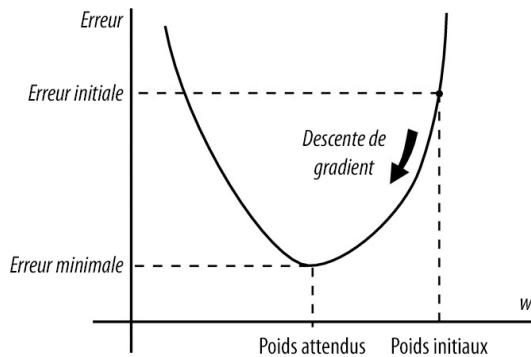
profondeur d du jeu de données, ou autrement dit, le nombre d'attributs que possède une couche, qu'elle soit cachée ou non. Les deux dernières dimensions w et h correspondent aux largeur et hauteur des images, généralement fixes lors de l'apprentissage.



Représentation d'un tenseur, constitué d'un batch d'images.

Apprentissage et optimisation

Descente de gradient : La descente de gradient est un algorithme itératif qui permet d'optimiser un modèle en cherchant le minimum global d'une fonction objectif f dans un espace continu. Pour ce faire, un déplacement est réalisé dans la direction du gradient négatif de f , soit celle pour laquelle la fonction objectif décroît le plus rapidement. A chaque étape, les paramètres θ du modèle sont alors mis-à-jour jusqu'à trouver le minimum global de f . A noter également que la vitesse de déplacement, et donc de mise-à-jour des paramètres, dépend du taux d'apprentissage, un hyperparamètre défini par l'utilisateur-trice.



Principe de la descente de gradient (modifié de Gurney (2007)).

Fonction objectif : Une fonction objectif permet de mesurer l'erreur d'un modèle, entre une valeur \hat{y} prédite par ce dernier et une référence y . Elle sert ainsi à optimiser la phase d'apprentissage, puisque le but est de la minimiser ou de la maximiser au fil des itérations, à l'aide d'un algorithme d'optimisation comme la descente de gradient. A noter qu'il existe des fonctions objectif spécifiques à différentes applications, comme pour la régression (L1, L2,

Glossaire

etc.) ou la classification (entropie croisée, logistique, etc.).

Gradient : Le gradient correspond à la pente de la fonction objectif.

Initialisation : L'initialisation consiste à donner une valeur aléatoire, empirique ou prédefinie aux paramètres θ d'un modèle. C'est une méthode systématiquement utilisée pour les applications d'apprentissage par transfert, qui peut également servir dans d'autres cas. En effet, pour les GANs en particulier, il a été montré qu'une initialisation réalisée en tirant aléatoirement des valeurs dans une loi normale offrait de bons résultats (Chintala *et al.*, 2016). D'autres méthodes aléatoires existent, comme les initialisations Xavier ou uniformes.

Label smoothing : Dans le cas des GANs, le discriminateur compare les produits générés à une référence. Il renvoie 0 lorsqu'il parvient à distinguer les deux [*'la distribution obtenue en sortie du générateur est différente de celle fournie en entrée'*], et 1 lorsqu'il n'y arrive pas [*'la distribution obtenue en sortie du générateur est similaire à celle fournie en entrée'*]. Les valeurs 0 et 1 sont en réalité des labels, qu'il est possible de brouter en les modifiant légèrement et aléatoirement, par 0,7 au lieu de 1 par exemple, opération communément appelée *label smoothing*. Il a été montré par Salimans *et al.* (2016) que cela permet d'améliorer les performances du modèle, notamment en matière de généralisation.

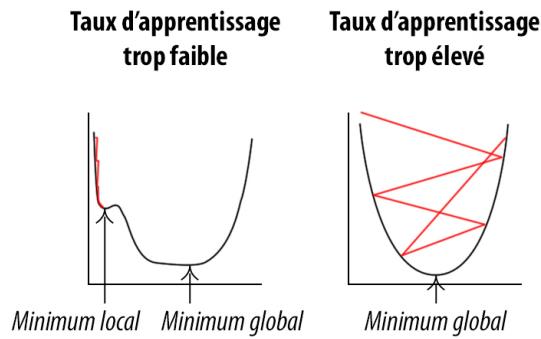
Optimiseur : Le terme d'optimiseur désigne un algorithme utilisé pour optimiser un modèle, en minimisant la fonction objectif sélectionnée pour la phase d'apprentissage. Il existe de nombreuses stratégies d'optimisation, qui présentent chacune leurs avantages. A ce jour, la descente de gradient stochastique (SGD) est l'une des plus connues, mais nécessite de fixer méticuleusement la valeur du taux d'apprentissage pour une optimisation globale du système. Depuis, d'autres algorithmes ont été développés et permettent de calculer un taux d'apprentissage adaptatif et spécifique à chaque paramètre du modèle, par exemple Adam, Adagrad, Adadelta, etc.

Paramètres : Dans le cas de l'apprentissage profond, les paramètres θ correspondent à des variables internes au modèle et associées à chaque connexion, c'est-à-dire les poids et les biais. Ils sont mis-à-jour lors de la phase d'optimisation, idéalement jusqu'à ce que la fonction objectif ait atteint son minimum local. A noter qu'il est généralement possible d'initialiser les poids et les biais, soit à partir de valeurs fixes définies par l'utilisateur-trice, soit aléatoirement, ou encore à partir des paramètres calculés pour un autre modèle ou à une itération donnée.

Propagation avant et rétropropagation : Au cours de l'apprentissage, les données sont passées au modèle durant une phase dite de propagation avant, à l'issue de laquelle des résultats \hat{y} sont produits. Une fonction objectif est alors utilisée pour mesurer l'erreur entre \hat{y} et une référence y . Celle-ci permet d'optimiser progressivement le modèle à l'aide d'un algorithme spécifique, ou optimiseur, qui mesure la pente de la fonction objectif et propage l'erreur en arrière pour mettre à jour les paramètres de l'ensemble des couches cachées du modèle.

Taux d'apprentissage : Le taux d'apprentissage est un hyperparamètre utilisé lors de la phase

d'optimisation du modèle. Il détermine la vitesse à laquelle l'optimiseur se déplace dans l'espace de la fonction objectif et met à jour les paramètres θ des couches du réseau. Certains optimiseurs sont plus sensibles que d'autres à la valeur utilisée pour le taux d'apprentissage. En effet, lorsqu'il est trop grand, la fonction objectif varie rapidement, parfois au point de ne jamais aboutir à une solution adéquate. Lorsque le taux d'apprentissage est trop petit, la convergence se fait lentement, et le modèle peut même rester coincé sur un plateau ou dans un minimum local qui n'est pas adapté au problème étudié. Il peut donc être important de le choisir correctement, soit de façon empirique, soit à l'aide de méthodes spécifiques. Il est également possible de le faire varier progressivement au cours de l'apprentissage, permettant de tirer partie des avantages de taux d'apprentissage élevés (vitesse) et faibles (finesse).



Influence du taux d'apprentissage sur la recherche d'un minimum global.

Table des matières

Remerciements	i
Acronymes	iii
Glossaire	v
Liste des figures	xv
Liste des tableaux	xvii
Introduction	1
1 État de l'art	5
1.1 Méthodes classiques de colorisation assistée par ordinateur	6
1.1.1 Les méthodes basées sur les gribouillis	7
1.1.2 Les méthodes basées sur le transfert	8
1.1.3 Les méthodes basées sur l'apprentissage	9
1.2 L'apprentissage profond pour la colorisation d'images	10
1.2.1 L'usage des DNNs pour la colorisation	11
1.2.2 L'usage des CNNs pour la colorisation	13
1.2.3 L'usage des GANs pour la colorisation	16
2 Données et méthode	19
2.1 Analyse préliminaire des produits à coloriser	19
2.1.1 Caractéristiques générales des orthophotographies anciennes	19
2.1.2 L'apport de la couleur pour les applications géographiques	20
2.2 Colorisation des orthophotographies anciennes	24
2.2.1 Développement d'une photothèque pour la colorisation	26
2.2.2 Développement d'un réseau de neurones profond	32
2.2.3 Évaluation des produits colorisés	38
2.3 Mise en œuvre de classifications historiques sur les produits colorisés	39
3 Résultats	41
3.1 Évaluation de l'apport de la couleur pour la mise en œuvre de classifications	41
3.2 Modèle local pour la colorisation des orthophotographies anciennes	44

Table des matières

3.2.1 Résultats de colorisation pour les modèles locaux dans le domaine du visible	44
3.2.2 Apprentissage par transfert et capacités de multispectralisation du modèle local	48
3.3 Modèle global pour la colorisation des orthophotographies anciennes	51
3.3.1 Évaluation du modèle global de colorisation	51
3.3.2 Visualisation des résultats et attributs appris par le modèle global	56
3.4 Résultats des classifications historiques sur les produits colorisés	59
4 Discussion	63
4.1 Recommandations et avertissements pour la mise au point de la photothèque .	63
4.2 Point sur l'apprentissage du modèle de colorisation	66
4.3 Limites des indicateurs pour l'évaluation d'une colorisation	67
4.4 Colorisation et valorisation des produits géographiques historiques	69
Conclusion	71
Bibliographie	73
Annexes	81
A Récapitulatif des étapes et tâches réalisées	81
B Préparation des mosaïques	85
C Mise en place de l'environnement de programmation	87
D Présentation des tests de colorisation réalisés	89
E La colorisation comme problème multi-modal	93

Liste des figures

1	Catégories et sous-catégories des méthodes de colorisation automatiques ou semi-automatiques	6
2	Représentation des différentes catégories de méthodes de colorisation informatiques	6
3	Méthode de classification employée pour évaluer l'apport de la couleur	22
4	Produits et attributs utilisés pour évaluer l'apport de la couleur	24
5	Résultats de colorisation pour des extraits de photographies aériennes anciennes	25
6	Structure de la photothèque utilisée pour l'apprentissage et la validation du modèle de colorisation	26
7	Représentation des espaces colorimétriques Lab et RVB	30
8	Exemple de transformations aléatoires et systématiques appliquées à une imagette, puis utilisées pour augmenter le jeu de données	31
9	Modèle général du réseau génératif antagoniste utilisé	33
10	Architectures retenues pour le générateur et le discriminateur	35
11	Résultats de classification obtenus à l'aide du scénario n°5, à partir des images et photographies en couleurs et monochromatiques	43
12	Valeurs des métriques qualité calculées entre les produits colorisés et le jeu de validation, au cours de l'apprentissage du DRAGAN local	45
13	Valeurs des fonctions objectif de G, D et L1 au cours de l'entraînement du modèle local pour l'imagerie en couleurs naturelles	45
14	Exemples de résultats de colorisation sélectionnés aléatoirement dans le jeu de validation, pour le modèle DRAGAN local à l'itération n°900	46
15	Comparaison des distributions bivariées de a et b , calculées sur les produits colorisés et le jeu de validation, pour le modèle DRAGAN local à l'itération n°900	47
16	Métriques calculées sur le jeu de validation pour les colorisations générées à l'aide du modèle DRAGAN local, à l'itération n°900	48
17	Résultats de colorisation sur Niederhausbergen avec le modèle local	49
18	Valeurs des fonctions objectif de G, D et L1 au cours de l'entraînement du modèle local pour l'imagerie en infrarouge couleur	50
19	Visualisation des NDVI historiques obtenus grâce au modèle local pour les années 1956 et 1978	51

Liste des figures

20	Valeurs des fonctions objectif de G, D et L1 au cours de l'entraînement du modèle global	52
21	Valeurs des métriques qualité calculées pour le modèle global entre les produits colorisés et le jeu de validation, au cours de l'apprentissage	53
22	Exemples de résultats de colorisation pour le jeu de données de validation, obtenus à partir du modèle global, aux itérations n°250, 350 et 950	54
23	Métriques calculées sur le jeu de validation pour les colorisations générées à l'aide du modèle DRAGAN global, à l'itération n°950	55
24	Comparaison des distributions bivariées de a et b , calculées sur les produits colorisés et le jeu de validation, pour le modèle DRAGAN global à l'itération n°950	56
25	Evolution de la qualité de la colorisation au fil des étapes d'apprentissage, pour deux emprises situées au centre-ville et en périphérie de Strasbourg	57
26	Visualisation d'un extrait des poids et attributs appris pour chaque couche du modèle globale, à l'étape n°950	58
27	Résultats de classification pour le cliché panchromatique de 1978 et son homologue colorisé, obtenu à partir du modèle global à son itération n°950	60
28	Nombre d'images utilisées par travaux pour l'apprentissage d'un modèle de colorisation	64
A.1	Répartition temporelle des tâches réalisées	81
A.2	Description générale des étapes du travail de recherche	83
B.1	Description générale des étapes du mosaïquage pour les photographies prises dans l'année 1964	86
D.1	Visualisation des résultats de colorisation obtenus pour différents paradigmes, algorithmes, modèles et architectures	90
E.1	Couleurs médianes et dominantes des toitures dans l'agglomération strasbourgeoise	93

Liste des tableaux

1	Récapitulatif des travaux menés sur la colorisation d'images à l'aide de méthodes classiques assistées par ordinateur	7
2	Récapitulatif des travaux menés sur l'apprentissage profond pour la colorisation d'images	12
3	Exemples de fonctions objectif utilisées dans les applications de régression et de classification	13
4	Description des produits raster panchromatiques constituant la base historique	20
5	Caractéristiques du couple d'images satellitaires Pléiades utilisées pour tester l'apport de la couleur	21
6	Description et valeurs des paramètres utilisés pour la grille de recherche dense	23
7	Description générale des produits utilisés pour constituer la photothèque	28
8	Description des photographies en couleurs naturelles utilisées comme complément à la photothèque	28
9	Spécifications générales de l'occupation du sol digitalisée pour l'année 1978	40
10	Valeurs des scores F1 obtenus pour le jeu de validation, pour différents scénarios de classification et classes d'occupation du sol sur l'image Pléiades de 2012	42
11	Valeurs des scores F1 obtenus pour le jeu de validation, pour différents scénarios de classification et classes d'occupation du sol	43
12	Meilleurs scores médians de MSE, PSNR et SSIM obtenus sur les produits colorisés, générés à partir des modèles locaux sur le jeu de données de validation	44
13	Valeurs des scores F1 obtenus sur un jeu de validation suite à la classification réalisée sur un extrait d'orthophotographie panchromatique et son homologue colorisé	59
D.1	Description des différents tests réalisés dans le cadre du travail de recherche	89

Introduction

Face aux processus qui façonnent aujourd’hui les territoires, tels le changement climatique ou l’artificialisation des sols, nous assistons régulièrement à un questionnement de nos pratiques de l’espace, à la fois au sein de la communauté scientifique et de l’opinion publique. En effet, la mise en œuvre des politiques aspire désormais à promouvoir l’interdisciplinarité et la pluralité des échanges. Cela incite à aller vers une meilleure intégration des connaissances sur le territoire, sur les plans économique, social et environnemental. Dans ce but se mettent en place des dispositifs sociotechniques qui facilitent l’interopérabilité des systèmes, et stimulent les interventions pluridisciplinaires. Ce sont par exemple les observatoires des territoires, infrastructures de données géographiques ou bases de données spatialisées (Noucher, 2013). La dimension diachronique est ici particulièrement importante, puisqu’il est question de procéder au suivi des espaces à enjeux. Pour répondre à ces besoins qui s’expriment à la fois spatialement et temporellement, nous disposons aujourd’hui d’une très grande quantité de données multi-sources et aux caractéristiques hétérogènes. En effet, celles-ci sont distribuées dans une variété de formats, et sont disponibles pour des usages, échelles, résolutions, publics différents…

A des fins de suivi et de gestion des territoires, les produits issus des méthodes et outils de la télédétection sont particulièrement intéressants. En effet, ils permettent (1) d’aller vers des approches prospectives en développant des modèles prédictifs à l’aide des séries temporelles Sentinel par exemple, ou (2) d’aller vers des approches rétrospectives en s’intéressant aux photographies aériennes anciennes, sur lesquelles sont représentés les précédents états des surfaces.

La mise en place d’une telle infrastructure de suivi a ainsi été envisagée par la Zone Atelier Environnementale Urbaine (ZAEU), avec l’élaboration d’un système d’information géographique (SIG) géohistorique, à l’échelle de Strasbourg. Celui-ci s’organise autour de plusieurs sphères thématiques et techniques, et devrait permettre d’analyser les évolutions qu’a connues la trame urbaine, au travers de ses composantes minérales et végétales notamment (ZAEU, 2013). Cette base de données historique a été développée par différents acteurs, principalement les services de l’EMS, le LIVE et les étudiant-e-s de la Faculté de Géographie et d’Aménagement de Strasbourg. Constituée d’orthophotographies et de divers fichiers vectoriels sur lesquels sont décrits l’habitat, les réseaux de transport et la végétation arborée pour plusieurs années de référence — 1932, 1956, 1964, 1978, 1989, 1998, 2002, 2008 et 2013 — elle a jusqu’à présent

Introduction

été utilisée pour analyser l'évolution des morphotypes urbains dans l'agglomération strasbourgeoise (Moisson, 2015; Sauter et Schwartz, 2017; Humbert, 2017; Medina Kennedy *et al.*, 2018). Outre ces travaux, la ZAEU a également exprimé un besoin concernant les photographies historiques. En effet, les campagnes aériennes réalisées sur la ville, depuis bientôt une centaine d'années, ont permis de constituer un fond photographique suffisant pour analyser les dynamiques décennales, l'artificialisation par exemple. Ces produits présentent cependant des caractéristiques hétérogènes, dont certaines en limitent les usages et leur interopérabilité.

Le besoin énoncé par la ZAEU consiste à travailler sur les produits en niveaux de gris, de sorte à disposer d'une information colorimétrique sur les surfaces. Historiquement, les photographies panchromatiques représentent un volume de données important à l'échelle de la nation, puisque collectées depuis la deuxième moitié du XIX^e siècle (Chevallier *et al.*, 1968). Ainsi, sur l'ensemble des 20 345 missions réalisées en France entre 1919 et 2018, et dont les clichés sont disponibles sur la plateforme de l'IGN, un total de 18 092 a été accompli à l'aide de capteurs panchromatiques (IGN, 2018).

Ces photographies sont cependant aujourd'hui souvent considérées comme désuètes, du fait de l'avènement des clichés numériques, mais aussi car plus difficiles à interpréter et à mobiliser que leurs homologues en couleurs naturelles par exemple. En effet, leurs caractéristiques géométriques — déformations liées à la lentille ou à la topographie, position de la plateforme... — et radiométriques — information spectrale résumée à une bande, finesse de l'émulsion pour les supports argentiques... — limitent généralement leur utilisation aux experts-interprètes (Chevallier, 1965; Muraz *et al.*, 1999).

Concernant les applications géographiques, le fait que les clichés panchromatiques ne possèdent qu'un seul canal apparaît comme une limite évidente. En effet, de nombreux travaux portent notamment sur la cartographie de l'occupation du sol, qui nécessite de disposer de suffisamment d'informations sur les surfaces pour pouvoir les discriminer. Par exemple, il est possible de cartographier différentes populations végétales, si tant est que leurs réponses spectrales dans le proche infrarouge, le rouge et le vert en particulier sont disponibles. Pour palier à ces manques, différentes méthodes d'extraction d'attributs existent sur des plans monochromatiques, mais elles ne suffisent pas forcément à remplacer une information spectrale riche (Lu et Weng, 2007). En effet, Palsson *et al.* (2012) et Cavallaro *et al.* (2016) ont montré que les canaux R, V et B génèrent de meilleurs résultats de classification sur l'urbain qu'un canal panchromatique seul. Sans avoir calculé d'attributs supplémentaires, les précisions globales obtenues en mode monochrome sont inférieures à 50%, alors que la couleur permet de dépasser les 70%.

De plus, bien que l'imagerie satellitaire ne puisse pas toujours remplacer la photographie aérienne, la télédétection spatiale apporte aujourd'hui des volumes de données d'autant plus importants, permettant de cartographier de vastes emprises à prix réduit, dans plusieurs régions du spectre électromagnétique, et à des fréquences temporelles adaptées pour des applications de cartographie en temps quasi-réel ou de l'occupation du sol (Lillesand *et al.*,

2015).

La combinaison de ces facteurs explique alors le manque d'intérêt actuel pour les fonds photographiques anciens, souvent difficiles à manipuler, outre les opérations classiques de saisie numérique.

L'objectif de ce travail est donc de proposer des algorithmes permettant de coloriser automatiquement des clichés historiques aériens, et ainsi de valoriser les produits disponibles dans la base de données constituée dans le cadre de la ZAEU. Les intérêts sont multiples, puisque la méthodologie employée devrait tout d'abord permettre de disposer d'une donnée patrimoniale plus simple à appréhender pour le grand public. Quant à la communauté scientifique, il semble raisonnable de supposer que ces produits seront plus facilement mobilisables dans les chaînes de traitements classiques, la classification notamment, pour la mise en œuvre d'analyses thématiques, comme celles de l'étalement urbain ou de la fragmentation des paysages...

Compte-tenu des points énoncés précédemment, nous pouvons alors poser plusieurs questions de recherche auxquelles ce travail tâchera de répondre :

- (1) Est-il possible de coloriser automatiquement des photographies aériennes anciennes monochromatiques et d'obtenir des résultats de colorisation corrects ?
- (2) Peut-on développer une base de données de référence suffisamment robuste et extensive pour aboutir à un résultat de colorisation plausible ? Quels sont les choix à faire pour la mise en place de cette infrastructure ? Quelles en sont les limites ?
- (3) Est-il possible d'appliquer la méthode dégagée à différents domaines : année, saison, espace géographique... ?
- (4) Peut-on utiliser les produits colorisés pour d'autres applications : visualisation, classification sémantique... ?

Afin de répondre à ces questions, nous allons tout d'abord évaluer l'apport de la couleur pour des produits issus de la télédétection spatiale et aérienne. Cela permettra de justifier ce travail ainsi que son intérêt, et de proposer une première piste de réflexion pour la suite des développements. Un état de l'art décrivant les travaux menés sur la question de la colorisation d'images est ensuite présenté. Il permet de justifier de la méthodologie employée, à la fois pour la constitution d'une photothèque de référence, et pour le développement d'un algorithme spécifique. Une fois les colorisations menées, une analyse d'indicateurs quantitatifs et sémantiques est réalisée, permettant d'évaluer la qualité des produits obtenus. Enfin, les limites et applications de la méthodologie dégagée sont discutées, permettant d'aboutir à des ouvertures et propositions d'améliorations.

1 État de l'art

Le besoin de disposer d'une chaîne de traitements automatisable s'explique avant tout par le temps, les outils et les connaissances expertes nécessaires à une colorisation correcte. En effet, les techniques manuelles développées dès la première moitié du XIX^e siècle reposent sur l'utilisation d'aquarelles, pigments ou gommes, appliqués directement sur des supports analogiques décrivant un paysage ou des individus par exemple (Lavédrine et McElhone, 2009).

Les données géographiques représentent cependant des espaces d'une toute autre dimension, avec des clichés numériques constitués de plusieurs centaines de millions de pixels et des emprises de l'ordre de $3,5 \times 3,5$ km pour chaque tuile dont nous disposons, le tout assemblé de sorte à pouvoir couvrir ici le territoire de l'Eurométropole de Strasbourg. Les techniques de colorisation manuelles se voient donc être limitées du fait de leur manque de scalabilité, justifiant la recherche de méthodes automatiques plus simples à mettre en œuvre.

Depuis les années 1970, l'avènement des systèmes de traitements informatisés pousse à l'apparition de techniques de colorisation automatiques ou semi-automatiques (Royer *et al.*, 2017). Ces méthodes, le plus souvent assistées par ordinateur au départ, se répartissent en plusieurs catégories et sous-catégories (Figure 1).

La première catégorie correspond aux méthodes non-paramétriques, qui reposent le plus souvent sur l'intervention d'un individu pour fournir une source depuis laquelle la couleur est extraite, puis transférée à une cible en noir et blanc. La seconde catégorie correspond aux méthodes paramétriques ou *learning-based*, qui permettent d'apprendre une fonction de prédiction de la couleur, à partir d'un jeu de données d'entraînement. La Figure 2 illustre le fonctionnement général des algorithmes associés.

Ces familles de méthodes sont développées dans les sous-parties suivantes. A noter qu'elles sont présentées comme un état des lieux décrivant les travaux des différents auteurs, plutôt qu'une synthèse générale. Cela permet d'avoir un panorama des algorithmes proposés dans le domaine, pour n'en retenir que quelques aspects ensuite utilisés pour nos travaux.

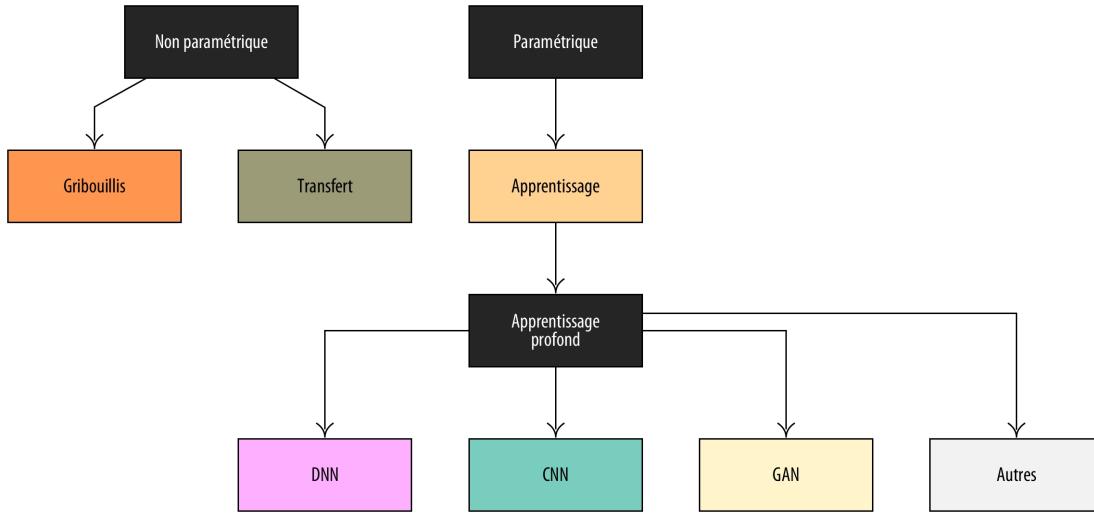


FIGURE 1 – Catégories et sous-catégories des méthodes de colorisation automatiques ou semi-automatiques. Bien que l'organigramme soit *a priori* exhaustif, seules les catégories étudiées dans le cadre de ce travail ont été présentées.

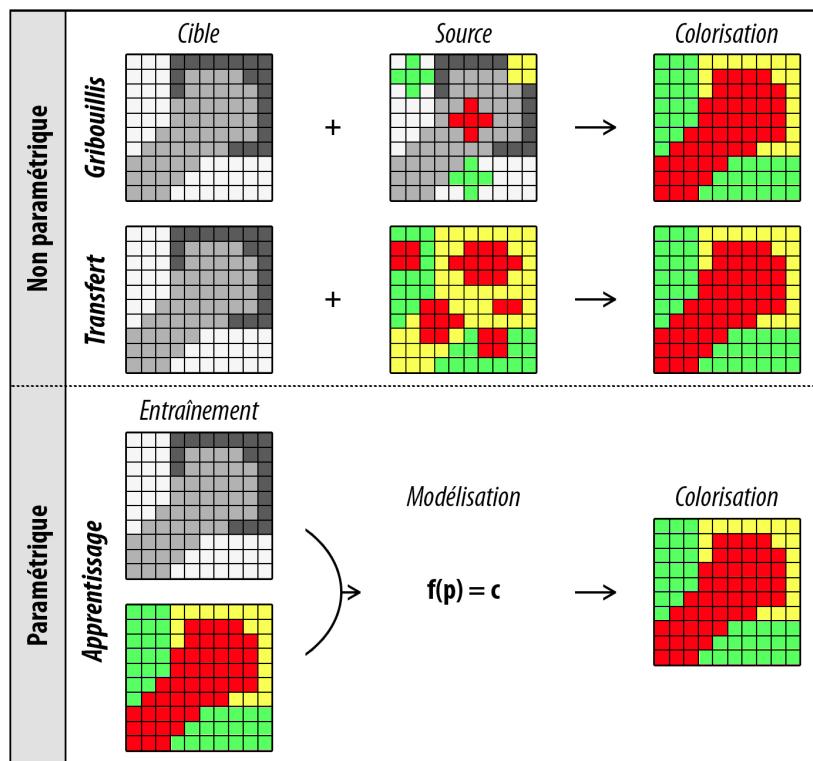


FIGURE 2 – Représentation des différentes catégories de méthodes de colorisation informatiques.

1.1 Méthodes classiques de colorisation assistée par ordinateur

Dans le cadre de ce travail, nous considérons comme méthodes classiques de colorisation assistée par ordinateur celles qui requièrent l'intervention d'un individu, ou de façon plus

1.1. Méthodes classiques de colorisation assistée par ordinateur

générale, celles qui ont été développées avant l'avènement de l'apprentissage profond. Elles se déclinent en plusieurs catégories et sous-catégories (Figure 1), avec des méthodes non paramétriques — gribouillis et transfert — et paramétriques.

Celles-ci sont détaillées dans les sous-parties à suivre, avec en complément une description de plusieurs travaux de référence menés sur la colorisation de clichés monochromatiques (Table 1).

Auteurs	Date	Couleurs	Méthode
Reinhard <i>et al.</i>	2001	Lab	Transfert
Welsh <i>et al.</i>	2002	Lab	Transfert
Irony <i>et al.</i>	2005	YUV	Transfert
Qu <i>et al.</i>	2006	YUV	Gribouillis
Yatziv et Sapiro	2006	YCbCr	Gribouillis
Luan <i>et al.</i>	2007	YUV	Gribouillis
Charpiat <i>et al.</i>	2008	Lab	Transfert
Liu <i>et al.</i>	2008	YUV	Transfert
Sýkora <i>et al.</i>	2009	?	Gribouillis
Chia <i>et al.</i>	2011	RGB	Transfert
Bugeau et Ta	2012	Lab	Apprentissage
Gupta <i>et al.</i>	2012	Lab	Transfert
Deshpande <i>et al.</i>	2015	Personnalisé	Apprentissage

TABLE 1 – Récapitulatif des travaux menés sur la colorisation d'images à l'aide de méthodes classiques assistées par ordinateur. Les codes couleurs sont les suivants :  apprentissage  gribouillis  transfert.

1.1.1 Les méthodes basées sur les gribouillis

Parmi les méthodes non paramétriques se trouvent celles basées sur l'utilisation de gribouillis ou *scribble-based*. Elles reposent sur l'intervention d'un-e opérateur-trice, en charge de positionner des échantillons de couleur (source) sur l'image en niveaux de gris, puis dont l'information colorimétrique est ensuite propagée au reste du produit (cible).

D'une façon générale, la propagation de la couleur est réalisée en fonction d'un critère de proximité, qui mesure la ressemblance entre un échantillon et le contenu numérique de la région dans laquelle il se trouve. L'information sur la chrominance est alors copiée sur les pixels homologues ou similaires, jusqu'à rencontrer les limites d'un objet par exemple, qui marquent la transition vers un espace radiométriquement différent. Ces interfaces sont ainsi problématiques, puisqu'elles combinent généralement les caractéristiques de deux régions distinctes, donnant lieu à des résultats de colorisation souvent flous.

Différentes techniques ont été proposées pour améliorer le fonctionnement des algorithmes basés sur les gribouillis, par exemple Yatziv et Sapiro (2006) qui utilisent une distance géodésique pour guider la propagation de la chrominance au reste de l'image. Il a été montré que cela améliore ainsi la qualité des résultats au niveau des zones de transition.

Qu *et al.* (2006) et Luan *et al.* (2007) ont également proposé des méthodes de colorisation

basées sur le calcul d'indicateurs de texture, utilisés pour évaluer la similarité entre plusieurs groupes de super-pixels, et réduire le nombre d'échantillons requis pour obtenir des résultats corrects.

Enfin, Sýkora *et al.* (2009) ont mis au point l'outil interactif *LazyBrush*, pour la colorisation de dessins réalisés à la main. L'algorithme est basé sur un système chargé d'assigner une couleur à chaque pixel, tout en minimisant une fonction d'énergie. Pour cela, les auteurs résolvent un problème de coupure de graphe, similaire aux approches développées en segmentation d'image.

Les améliorations apportées à l'algorithme classique consistent ainsi en de nouvelles méthodes d'optimisations, ou en l'utilisation d'attributs discriminants. De façon générale cependant, le placement d'échantillons sur l'ensemble des régions d'une image s'avère être une tâche difficile, qui pourrait être substituée par des techniques de mise en correspondance par exemple.

1.1.2 Les méthodes basées sur le transfert

C'est à cet inconvénient que répondent les méthodes basées sur le transfert ou *transfer-based*. Non paramétriques également, elles requièrent de disposer de deux clichés, l'un en couleurs (source) et l'autre en niveaux de gris (cible). La chrominance est alors transférée à la cible, à l'aide d'algorithmes de mise en correspondance de voisinage, compte-tenu de la luminance ou de la texture notamment.

Reinhard *et al.* (2001) proposent une méthode permettant de transférer l'information colorimétrique d'une source à une cible, les deux étant ici en couleurs, sur la base d'une simple analyse statistique. La distribution des valeurs du produit vers lequel s'opère le transfert est ainsi modulée en fonction des moyennes et écarts-types calculés pour chaque bande de la référence. Cette technique correspond d'avantage à du transfert de style, mais marque les premiers développements menés sur le passage de la chrominance d'une image à une autre.

Welsh *et al.* (2002) ont développé une méthode basée sur la mise en correspondance de patchs entre une source en couleurs, et une cible en niveaux de gris. Le processus repose sur l'analyse des valeurs de luminance et de texture, attributs qui fournissent des informations locales et globales sur la scène. Irony *et al.* (2005) améliorent la technique de Welsh *et al.* (2002) à l'aide d'un algorithme de classification supervisée, qui recherche sur la cible un ensemble de régions contenant une information de bas niveau adaptée à la procédure de colorisation. Les pixels appartenant à ces régions sont alors utilisés comme références, à la manière des gribouillis, avec une information colorimétrique qui est ensuite propagée au reste de l'image.

Charpiat *et al.* (2008) proposent un algorithme d'optimisation globale adapté à la résolution des problèmes multi-modaux, pour la prédiction des couleurs de chaque pixel. Une méthode de coupure de graphe est également développée, afin de maximiser la probabilité de cohérence de l'image colorisée.

1.1. Méthodes classiques de colorisation assistée par ordinateur

Gupta *et al.* (2012) ont développé une méthode de mise en correspondance de super-pixels entre une source et une cible, suite au calcul de différents attributs. La chrominance est transférée entre les paires de patchs ainsi associés, puis propagée au reste de l'image à coloriser. L'approche super-pixels permet, selon les auteurs, d'améliorer la cohérence spatiale du produit en sortie.

Liu *et al.* (2008) proposent une méthode basée sur le transfert, qui réduit le nombre d'interventions nécessaires à l'individu en charge de la colorisation. En effet, l'algorithme développé cherche un ensemble d'images de référence sur internet, similaires au produit à coloriser sur le plan sémantique. Plusieurs groupes de pixels sont ensuite mis en correspondance puis coregistrés, avec finalement un transfert direct de la couleur. Malgré les capacités d'automatisation de cette méthode, les auteurs notent des difficultés à la généraliser à autre chose que des photographies de paysages. Chia *et al.* (2011) améliorent la technique de Liu *et al.* (2008) en filtrant les images de référence et à coloriser, de sorte à n'extraire que des régions parfaitement homologues, facilitant donc le transfert de la couleur. Pour ce faire, un individu doit fournir à l'algorithme une série de labels qui décrivent la sémantique de l'image, ainsi qu'une segmentation qui renseigne sur les objets présents au premier plan de celle-ci. Selon les auteurs, ces améliorations permettent de généraliser la technique de Liu *et al.* (2008) à des scènes autres que des paysages.

Les méthodes basées sur le transfert montrent que la colorisation requiert de disposer de descripteurs globaux et locaux pour l'image. La texture apporte une première information, qu'il peut être intéressant de compléter avec une segmentation par exemple, sur laquelle les différents objets sont décrits. Cela renvoie donc à l'importance des échelles d'analyse de l'image, avec des niveaux pixel et super-pixels qui n'apportent pas la même information. L'intervention d'un individu reste cependant essentielle dans la sélection d'une référence adéquate pour orienter le processus de colorisation.

1.1.3 Les méthodes basées sur l'apprentissage

Viennent enfin les méthodes basées sur l'apprentissage ou *learning-based*, paramétriques et qui permettent d'apprendre une fonction de la forme $f(p) = c$, avec p le cliché panchromatique et c le produit colorisé correspondant. Leur fonctionnement général est décrit plus en détails dans la Partie 1.2, dédiée à l'apprentissage profond.

Plus récentes que les méthodes basées sur les gribouillis ou le transfert, elles permettent d'aller plus facilement vers une automatisation du processus de colorisation. En effet, l'apprentissage informatique appartient aux techniques développées en intelligence artificielle, qui consistent à tirer partie des capacités de calcul des CPUs ou GPUs pour mener une réflexion sur des concepts.

Sans tenir compte des méthodes d'apprentissage profond, ces techniques n'ont connu que peu de développements. Parmi celles-ci se trouvent les travaux de Bugeau et Ta (2012), qui ont

développé un algorithme de colorisation basé sur l'apprentissage d'un modèle de prédiction. Celui-ci prend en entrée un ensemble de patchs dans lesquels sont extraits la luminance des pixels ainsi qu'une variété d'attributs. Une régularisation à l'aide de la méthode de variation totale est également réalisée pour améliorer la cohérence spatiale des résultats en sortie. Enfin, Deshpande *et al.* (2015) posent la colorisation comme un problème qui peut être résolu en optimisant un système linéaire basé sur une fonction objectif quadratique. Une correction de l'histogramme en sortie est ensuite réalisée, en tenant compte des prédictions locales de la couleur, puis des cohérences spatiale et radiométrique.

Jusqu'ici, les méthodes classiques basées sur les gribouillis, le transfert ainsi que l'apprentissage, montrent que le processus de colorisation requiert une méthode d'optimisation adéquate, ainsi que des attributs suffisamment discriminants pour orienter le processus de colorisation. Les travaux portant sur ces techniques classiques mettent en exergue un manque d'automatisation, avec des approches semi-automatiques le plus souvent. Adaptées à des problèmes de petite envergure, elles sont alors difficiles à mettre en œuvre lorsqu'il s'agit de proposer une méthode opérationnelle pour le traitement d'images, dont les emprises font plusieurs millions, voire milliards de pixels. En effet, elles requièrent soit l'intervention d'un individu, soit la sélection puis le calcul d'une série d'attributs répondant au problème énoncé.

1.2 L'apprentissage profond pour la colorisation d'images

Plus récemment, l'apprentissage profond ou *deep learning*, un ensemble d'algorithmes paramétriques et basés sur les réseaux de neurones, a permis de résoudre une grande variété de problèmes avec une précision sans précédents. Les applications développées en vision artificielle à partir de ces méthodes sont diverses : transfert de style (Gatys *et al.*, 2016), vision stéréoscopique (Žbontar et Le Cun, 2015), classification et segmentation d'images (Krizhevsky *et al.*, 2012; Shelhamer *et al.*, 2016), super-résolution (Ledig *et al.*, 2016), prédiction d'un champ de profondeur (He *et al.*, 2018) ou encore colorisation...

L'ensemble de ces méthodes se base sur les réseaux de neurones profonds, définis selon l'équation $y = f(x ; \theta)$. Leur objectif est de prédire une valeur de y , en minimisant l'erreur grâce à l'apprentissage des paramètres θ , ou poids et biais, du modèle. A l'origine inspirés des neurosciences, ils ont été développés avec l'objectif de reproduire les capacités du cerveau humain à mener une réflexion sur des concepts, puis à inférer des connaissances. Si l'on parle aujourd'hui de réseaux de neurones profonds, c'est avant tout parce qu'ils s'organisent autour d'un ensemble de fonctions hiérarchisées – linéaires et non-linéaires – qui constituent les couches cachées du réseau. Ce sont elles, ou plutôt les attributs θ qui leurs sont associés, qui font l'objet d'un apprentissage. De part et d'autre de ces couches se trouvent les entrées x , avec lesquelles le modèle cherche à prédire les valeurs de y , puis les sorties \hat{y} calculées à partir des couches cachées. Afin d'évaluer l'efficacité du modèle, il existe différentes métriques, ou fonctions objectif, qui mesurent la distance entre y et \hat{y} . Les fonctions les plus souvent rencontrées en *machine learning* sont par exemple l'estimateur des moindres carrés, ou

1.2. L'apprentissage profond pour la colorisation d'images

encore l'entropie croisée. Selon les objectifs et applications, d'autres métriques peuvent être proposées. Le processus d'apprentissage des réseaux de neurones profonds consiste, selon les cas, à minimiser ou maximiser la fonction objectif, en mettant tout simplement à jour les valeurs des paramètres θ . Pour cela, l'erreur de prédiction est rétro-propagée au sein du réseau, selon un processus itératif, à l'aide d'un algorithme d'optimisation (gradient stochastique, Adam, Adagrad...), jusqu'à atteindre la représentation la plus fidèle de y (Rumelhart *et al.*, 1986).

Hérités des réseaux de neurones simples, comme le perceptron, les réseaux de neurones profonds permettent de combiner une vaste palette de couches cachées, ce qui leur confère cette aptitude à résoudre des problèmes de natures diverses, du traitement d'images aux applications en langage naturel. En effet, ils apprennent une hiérarchie de concepts imbriqués, obtenus en combinant des représentations de complexité croissante, jusqu'à générer une information pertinente et adaptée à la résolution d'un problème posé. Ainsi, sur une image avec un réseau de neurones convolutif par exemple, les concepts appris par la première couche cachée correspondent habituellement aux côtés. Combinés, ils permettent ensuite de détecter des motifs au sein de la seconde couche, puis des familles d'objets ou de surfaces... Ces représentations, parfois très abstraites, sont en général difficiles à obtenir à partir des algorithmes classiquement utilisés en traitement d'images par exemple. L'apprentissage profond permet ainsi de s'affranchir, au moins en partie, du calcul d'attributs, étape le plus souvent indispensable aux autres applications développées en *machine learning*, puisqu'ils sont ici en général appris par le réseau (Rumelhart *et al.*, 1986; Goodfellow *et al.*, 2015).

Cependant, c'est seulement depuis 2015 que se développent des méthodes d'apprentissage profond destinées à coloriser des photographies en noir et blanc, rarement issues de la télé-détection. Parmi celles-ci, trois grandes catégories d'architectures sont proposées pour les réseaux de neurones artificiels : les DNNs (*Deep Neural Networks*), les CNNs (*Convolutional Neural Networks*) et dérivés, puis plus récemment les GANs (*Generative Adversarial Networks*) et dérivés. Ces techniques sont renseignées dans la Table 2 et décrites dans les sous-parties suivantes. A noter qu'il demeure également d'autres méthodes moins fréquentes, rapidement évoquées dans le cadre de ce travail. Il est aussi important de préciser que peu de recherches, parmi celles qui existent aujourd'hui, portent sur la colorisation de produits géographiques.

1.2.1 L'usage des DNNs pour la colorisation

La première famille de modèles à avoir été utilisée pour la colorisation correspond aux DNNs, ou réseaux de neurones profonds simples. Ils s'organisent autour de couches entièrement connectées, dans lesquelles les neurones appartenant à une couche sont connectés à l'ensemble des neurones des couches directement adjacentes.

Ces modèles ont été utilisés par Cheng *et al.* (2015), qui ont développé une méthode de colorisation s'appuyant sur le calcul préliminaire d'attributs discriminants. Pour cela, ils extraient trois niveaux de représentations, avec (1) les valeurs de luminance de l'image à

Chapitre 1. État de l'art

Auteurs	Date	Couleurs	Modèle	Orientation	Fonction	Données
Cheng <i>et al.</i>	2015	YUV	DNN	Régression	MSE	SIFT Flow
Agrawal et Sawhney	2016	Lab	CNN	Classification, régression*	Entropie croisée, L2*	CIFAR-10
Deshpande <i>et al.</i>	2016	RGB	VAE	Génération	Composite	Wild FLW, LSUN-Church, ILSVRC
Iizuka <i>et al.</i>	2016	Lab	CNN	Régression, classification	MSE, entropie croisée	Places
Isola <i>et al.</i>	2016	Lab	GAN	Génération	Antagoniste + L1	Cityscapes, CMP Facades, Google Maps, etc.
Larsson <i>et al.</i>	2016	YUV	CNN	Classification	Composite	ImageNet, SUN-A, SUN-B
Limmer et Lensch	2016	RGB	CNN	Régression	MSE	Personnalisé
Varga et Szirányi	2016	YUV	CNN	Classification	Entropie croisée	ILSVRC, SUN
Zhang <i>et al.</i>	2016	Lab	CNN	Classification	Entropie croisée	ImageNet
Cao <i>et al.</i>	2017	RGB, YUV	GAN, WGAN*	Génération	Antagoniste + L1	LSUN Bedroom
Frans	2017	RGB	GAN	Génération	Antagoniste + L2	Safebooru
Fu <i>et al.</i>	2017	Lab	CNN, GAN*, WGAN*	Régression, génération*	L2, antagoniste*	Safebooru
Lal <i>et al.</i>	2017	Lab	WGAN	Génération	Composite	ImageNet
Royer <i>et al.</i>	2017	Lab	VAE	Génération	Logarithme de la vraisemblance	CIFAR-10, ILSVRC
Song <i>et al.</i>	2017	Autre	CNN	Classification	Entropie croisée	Personnalisé
Suárez <i>et al.</i>	2017	RGB	GAN	Génération	Antagoniste	Personnalisé
Varga et Szirányi	2017	YUV	CNN	Classification	Entropie croisée	SUN
Liu <i>et al.</i>	2018	Lab	CNN	Classification	Entropie croisée	AID, RSSCN7

TABLE 2 – Récapitulatif des travaux menés sur l'apprentissage profond pour la colorisation d'images. Les éléments marqués par une étoile * correspondent à des tests secondaires menés par les auteurs, en général peu concluants ou peu développés. Les codes couleurs sont les suivants : ■ DNN ■ CNN et dérivés ■ GAN et dérivés ■ Autres.

coloriser, (2) une série d'attributs DAISY et (3) des descripteurs sémantiques de haut niveau. Une fois concaténées, ces informations sont utilisées en entrée d'un réseau de neurones profond, constitué de couches entièrement connectées seulement, et entraîné à l'aide de l'erreur quadratique moyenne. Une méthode de segmentation est également présentée pour incorporer une information globale sur l'image et orienter le processus de colorisation. Enfin, la sortie est affinée à l'aide d'un filtrage bilatéral qui assure un résultat de qualité.

Bien qu'ayant obtenu des colorisations correctes, les auteurs soulignent les limites imposées par l'architecture générale des DNNs. En effet, ils sont constitués de couches entièrement connectées, avec pour chacune un nombre de paramètres θ à apprendre qui est égal au produit de l'ensemble des dimensions de l'entrée (Goodfellow *et al.*, 2015). Cela limite ainsi leur capacité d'apprentissage sur de grands jeux de données spatialisés et multi-canaux. Les couches entièrement connectées manquent également de souplesse, dans la mesure où les modèles développés fonctionnent uniquement sur une taille fixe d'image. Il existe donc un

1.2. L'apprentissage profond pour la colorisation d'images

manque d'intérêt de la part de la communauté scientifique pour cette famille de réseaux, dans le cadre d'applications portant sur la manipulation de produits matriciels.

1.2.2 L'usage des CNNs pour la colorisation

Le manque de scalabilité des DNNs pour le traitement d'images a été l'amorce nécessaire au développement des CNNs, ou réseaux de neurones convolutifs. Apparus à la fin des années 1980 sous la forme d'un neocognitron (Fukushima, 1988), puis popularisés en 1998 avec le modèle LeNet (LeCun *et al.*, 1998), il a fallu attendre jusqu'en 2012 pour que les CNNs se fassent une place dans le monde du *deep learning*, avec l'amélioration des technologies informatiques – les GPUs notamment – et le développement du réseau AlexNet (Krizhevsky *et al.*, 2012).

Leur architecture s'organise autour de fonctions de convolution, qui récupèrent successivement l'information contenue dans la couche qui leur est passée, grâce à un noyau dont la taille est fixée par l'utilisateur-trice. Les poids sont ainsi partagés à l'échelle de la fenêtre mouvante, permettant de réduire le nombre de paramètres à apprendre et d'aller vers une analyse sémantique des images (Goodfellow *et al.*, 2015).

A ce jour, les CNNs sont les architectures les plus utilisées pour la colorisation de clichés monochromatiques. La prédiction d'une couleur peut cependant être perçue comme un problème de régression, dans un espace colorimétrique continu, ou de classification, dans un espace colorimétrique discret. Ces approches diffèrent principalement du fait des fonctions objectif proposées pour minimiser l'erreur du modèle (Table 3), et ainsi converger vers une solution adéquate.

Régression	Classification
L1, Smooth L1, L2, Huber, Log Cosh, Quantile, etc.	Entropie croisée, Hinge, Logistique, Exponentielle, etc.

TABLE 3 – Exemples de fonctions objectif utilisées dans les applications de régression et de classification.

A ce jour, seuls les travaux de Iizuka *et al.* (2016) ont permis d'associer les approches de régression et de classification au sein d'un même système. Les auteurs présentent une technique de colorisation qui combine des attributs locaux et globaux. Le modèle général correspond à un CNN, organisé en quatre sous-réseaux qui apprennent des attributs de bas niveau, de niveau intermédiaire, de haut niveau, ainsi qu'une information globale sur la scène. Ces représentations sont alors fusionnées, puis passées à une série de couches convolutives dont la fonction est de coloriser le produit monochromatique, en résolvant un problème de régression. Une partie du modèle permet également de classifier la sortie, afin de mieux apprendre les descripteurs globaux de l'image lors de la phase d'entraînement.

Agrawal et Sawhney (2016) explorent quant à eux plusieurs architectures et familles de modèles. Leur effort se concentre principalement sur les CNNs, et la mise en œuvre de différentes

Chapitre 1. État de l'art

méthodes d'optimisation, en testant plusieurs fonctions objectif. Les auteurs comparent ainsi des approches de colorisation basées sur les régressions, et d'autres sur les classifications. Ils notent globalement des résultats de mauvaise qualité avec les GANs – présentés plus tard dans cette même partie – et fonctions objectifs basées sur une distance euclidienne.

D'autres travaux ont été menés sur les problèmes de régression plus spécifiquement, notamment ceux de Limmer et Lensch (2016), qui modélisent la fonction permettant de passer d'un canal proche-infrarouge vers une image en couleurs naturelles. Leur modèle se base sur un CNN multi-échelle, qui prend une pyramide normalisée en entrée. En complément, les auteurs passent également un filtre moyen de l'image d'origine à une couche entièrement connectée, permettant d'améliorer la prédiction des valeurs de chrominance. L'image d'entrée est ensuite concaténée au résultat, afin de récupérer les détails de la scène de départ. Cette approche permet ainsi de souligner les capacités de multi-spectralisation, avec un transfert qui s'opère entre différentes régions du spectre électromagnétique.

Les approches qui considèrent la colorisation comme un problème de régression sont généralement peu nombreuses, du fait des limites imposées par les fonction objectif basées sur la distance euclidienne ou les résidus $r = y - \hat{y}$. En effet, elles tendent à produire des résultats peu saturés et flous, car elles minimisent l'erreur moyenne de la prédiction.

Ce constat explique donc la pré-dominance des approches basées sur une fonction objectif de classification. En effet, une série d'auteurs a montré que les résultats obtenus à l'aide d'un tel système sont généralement de meilleure qualité, à la fois spatialement et spectralement. Pour cela, l'espace colorimétrique est discréte, chaque classe correspondant donc à un groupe de pixels, proches sur le plan radiométrique.

Le premier travail posant la colorisation comme un problème de classification correspond à celui de Larsson *et al.* (2016). Selon les auteurs, il est nécessaire de disposer de représentations sémantiques de bas niveau pour comprendre comment s'organise une image, et ainsi la coloriser. Un réseau VGG-16 pré-entraîné est tout d'abord utilisé, afin d'extraire cette information à partir de l'image monochromatique. Les attributs sémantiques sont ensuite concaténés au sein d'une hypercolonne, qui représente un continuum allant des descripteurs de haut niveau vers ceux de plus bas niveau. L'information est finalement passée à trois couches entièrement connectées, qui prédisent un histogramme des valeurs de teinte et de chroma pour chaque pixel. Cette méthode permet ainsi de tenir compte de la dimension multi-modale du problème de colorisation.

Les travaux de Larsson *et al.* (2016) font ainsi partie des premiers à poser la colorisation comme un problème multi-modal sur le plan des développements techniques. Zhang *et al.* (2016) cherchent à aller plus loin, en soumettant un CNN entraîné sur plus d'un million d'images, et dont les sorties sont ensuite pondérées afin de donner plus de poids aux couleurs les plus rares. La combinaison d'une fonction objectif orientée vers les problèmes de classification, ainsi que d'une pondération des valeurs de chrominance, permet alors d'obtenir des résultats de colorisation diversifiés et de bonne qualité.

1.2. L'apprentissage profond pour la colorisation d'images

Selon une approche similaire, Varga et Szirányi (2016) utilisent eux-aussi un modèle VGG-16 afin d'extraire une information sémantique discriminante, complétée ensuite par des attributs Multi-Sage. Ces couches sont concaténées puis passées à un CNN à deux étapes. Leur architecture permet de prédire les canaux U et V d'une image monochromatique Y, tout en disposant de représentations suffisamment riches pour coloriser une vaste palette d'objets et de scènes. Plus tard, Varga et Szirányi (2017) ont développé une méthode qui combine les avantages des méthodes paramétriques et de celles basées sur les exemples. Ils proposent une architecture constituée de deux CNNs entraînés en concert, et qui possèdent chacun la même structure. Le premier réseau prend l'image en couleurs comme entrée, et identifie les zones nécessaires à l'extraction d'une palette de couleurs. Le second récupère cette information et la combine à des descripteurs sémantiques, afin de coloriser le produit monochromatique. Cette méthode requiert donc de disposer d'une référence, point qui peut être perçu comme une limite, mais qui permet aussi de mieux orienter le processus en fournissant une palette *a priori* pertinente.

Une série de travaux a également été menée sur des produits plus atypiques. En effet, les méthodes développées jusqu'ici traitent la colorisation dans sa dimension horizontale et/ou sur des images naturelles.

Il peut cependant être plus compliqué de manipuler des produits issus de méthodes purement numériques par exemple, comme le montrent Fu *et al.* (2017), qui proposent une technique de colorisation automatisée pour les dessins au trait. Dans ce but, ils comparent trois modèles différents – CNN, GAN et WGAN – et complètent l'apprentissage par un histogramme de couleurs utilisé comme condition, afin d'orienter le processus de colorisation.

Les développements réalisés sur les produits géographiques sont eux-aussi rares, malgré le nombre d'applications évidentes qu'il serait possible d'en tirer, tant en termes de colorisation que de multi-spectralisation, classification, segmentation sémantique, super-résolution, etc.

Song *et al.* (2017) sont les premiers à proposer une méthode de colorisation pour des produits issus de la télédétection spatiale radar. Les auteurs ont travaillé sur la prédiction d'images satellitaires entièrement polarimétriques, à partir de produits SAR à polarisation simple. Ils calculent un ensemble de descripteurs multi-échelles à l'aide d'un modèle VGG-16 pré-entraîné, et dont les poids ont été adaptés aux caractéristiques spectrales des images. Ces attributs sont ensuite concaténés pour former une hypercolonne, passée à un deuxième réseau constitué de couches entièrement connectées. Les représentations ainsi apprises correspondent à une matrice de covariance, dont peuvent être dérivées les caractéristiques polarimétriques des surfaces.

Enfin, Liu *et al.* (2018) proposent un réseau multi-tâches, capable de coloriser des images satellitaires THRS et d'augmenter artificiellement leur résolution. Le modèle s'organise autour d'un auto-encodeur, ainsi que d'une structure résiduelle qui permet d'améliorer les performances de reconstruction de l'image.

Outre les questions liées à l'architecture du modèle, le choix des fonctions objectif est donc ici crucial, puisqu'il conditionne la qualité du rendu en sortie. Les auteurs sont donc souvent amenés à proposer des fonctions objectif adaptées au problème de colorisation, qui reste avant tout multi-modal, puisqu'à un même niveau de gris peut être associée une vaste palette de couleurs.

1.2.3 L'usage des GANs pour la colorisation

Compte-tenu des limites imposées par le choix d'une fonction objectif adaptée, il pourrait être plus simple de demander à un réseau de générer une image dont la distribution des valeurs de chrominance se rapproche de celle d'un jeu d'entraînement. C'est à ce type de besoin que répondent les GANs, ou réseaux génératifs antagonistes, utilisés pour générer une nouvelle information selon une approche non supervisée ou semi-supervisée (Goodfellow *et al.*, 2014a). En effet, les méthodes de classification et de régression considèrent l'espace de sortie comme étant non structuré, avec des pixels qui sont donc indépendants les uns des autres. Les GANs apprennent au contraire une prédiction structurée, avec une fonction objectif antagoniste, qui pénalise la sortie dans son ensemble, et pas seulement les unités élémentaires qui la composent (Isola *et al.*, 2016).

Ce fonctionnement s'explique par la structure même de ces modèles, avec deux réseaux mis en concurrence : (1) un générateur G qui capture une distribution réelle et génère une nouvelle information à partir de celle-ci, et (2) un discriminateur D qui évalue la possibilité que le produit généré provienne d'une distribution réelle plutôt que de G (Goodfellow *et al.*, 2014a). Ainsi, l'analogie utilisée par Goodfellow *et al.* (2014a) pour expliquer les GANs consiste à poser G comme un contrefacteur, et D comme un douanier. Le rôle du discriminateur est d'apprendre à déterminer si l'objet qui lui est passé a été contrefait ou est issu d'un distributeur officiel. Le générateur, quant à lui, apprend à produire des objets qui ne seront pas détectés par le douanier comme étant une contrefaçon. Ces deux parties se font face, et apprennent simultanément à améliorer leurs performances, jusqu'à ce que le contrefacteur parvienne à faire passer ses produits de synthèse comme étant issus d'une distribution réelle. Récents, les GANs sont mobilisés pour répondre à une grande variété de problèmes, dont la colorisation. Bien qu'instables et plus difficiles à entraîner qu'un CNN par exemple, les images en sortie sont généralement de bonne qualité et possèdent des couleurs vibrantes.

De façon traditionnelle, le générateur prend un vecteur bruit en entrée, qui est ensuite transformé en une sortie appartenant à la distribution cible. C'est le cas non supervisé. Dans le cas de la colorisation, les approches proposées sont systématiquement semi-supervisées, dans la mesure où l'utilisateur-trice fournit une condition aux réseaux G et D , le produit monochromatique depuis lequel des valeurs de chrominance seront prédites. Cette sous-catégorie de modèle correspond au cGAN, ou GAN conditionnel, proposé par Mirza et Osindero (2014). A noter également que les GANs peuvent être composés de couches entièrement connectées, de couches convolutives, d'une combinaison des deux, voire même d'autres blocs... Cela donne

1.2. L'apprentissage profond pour la colorisation d'images

lieu à différentes sous-catégories de modèles, qui ne seront pas détaillées par souci de clarté.

Les travaux d'Isola *et al.* (2016) ont permis le développement de l'outil *Pix2Pix* pour la transformation image à image. Il s'appuie sur un GAN conditionnel, ou cGAN, qui conditionne le modèle à l'aide d'une image en niveaux de gris par exemple. Une architecture U-Net est utilisée pour le générateur, afin d'extraire la sémantique de la scène et d'aller vers des abstractions de haut niveau. Ces attributs servent ainsi à générer la chrominance d'un produit monochromatique passé en entrée. La sortie du générateur est finalement envoyée au discriminateur, qui s'organise autour d'une structure PatchGAN, entièrement convolutive. L'apprentissage du modèle repose sur la combinaison de deux fonctions objectif. La première est antagoniste et représente les valeurs à haute fréquence de la distribution. La seconde correspond à une distance L1, qui capture avec précision les valeurs basse fréquence. Utilisées conjointement, ces fonctions favorisent les sorties de bonne qualité, tant spatialement que spectralement. Outre ses applications en colorisation, le modèle *Pix2Pix* est également utilisé pour de nombreuses autres tâches, comme la génération d'images à partir de labels ou de dessins au trait.

Cao *et al.* (2017) utilisent également un GAN conditionnel pour modéliser la distribution des couleurs d'un jeu de données d'entraînement, puis coloriser des produits monochromatiques. La condition d'entrée correspond ici à l'image en niveaux de gris, bruitée afin de favoriser la diversité des sorties. Le modèle proposé est entièrement convolutif, ce qui permet aux auteurs de traiter des images de n'importe quelle dimension.

Sur la base des travaux de Limmer et Lensch (2016), Suárez *et al.* (2017) ont développé un réseau générateur antagoniste qui prédit les trois canaux du visible, à partir d'images monochromatiques capturées dans le proche-infrarouge. Cet article permet de souligner les capacités de multi-spectralisation des GANs, et ouvre la porte vers la génération d'images issues de différentes régions du spectre électro-magnétique, comme pour Limmer et Lensch (2016).

Frans (2017) utilise quant à lui deux réseaux génératifs qui fonctionnent en tandem, pour la colorisation de dessins au trait. Le premier prédit une couleur, tandis que le second ajoute un ombrage. L'auteur développe une approche similaire à celle de Fu *et al.* (2017), avec un modèle qui colorise cette fois-ci en deux étapes.

Enfin Lal *et al.* (2017) présentent une méthode basée sur les WGANs conditionnels, théoriquement plus stables que les GANs traditionnels. Afin de générer une colorisation plausible, les auteurs optimisent l'apprentissage à l'aide de deux fonctions objectif. La première, antagoniste, permet de tenir compte de la dimension sémantique de l'image, mais aussi de la distribution des valeurs des pixels. La seconde, utilisée dans des applications de classification, mesure la différence entre le produit généré et une vérité terrain, d'une façon similaire à celle de Isola *et al.* (2016).

Outre les GANs, d'autres méthodes basées sur les modèles génératifs ont également été utilisées pour la colorisation, mais ne sont pas particulièrement développées dans le cadre de ce

Chapitre 1. État de l'art

travail. En effet, les sorties obtenues à l'aide de ces réseaux sont presque systématiquement générées aléatoirement. Le résultat de colorisation est donc différent à chaque fois qu'une même entrée est passée au modèle. La réalisation de mosaïques d'orthophotographies anciennes suppose cependant de disposer de tuiles radiométriquement cohérentes, d'où le manque d'intérêt pour ces méthodes dans notre cas.

Parmi celles-ci se trouvent Deshpande *et al.* (2016), qui proposent un auto-encodeur variationnel (VAE) afin d'apprendre une encapsulation à faible dimensionnalité d'un champ colorimétrique. A partir de celle-ci sont ensuite générés divers scénarios de colorisation pour une même image monochromatique. La fonction objectif du modèle a également été adaptée afin d'améliorer la qualité spatiale des sorties. Les auteurs soumettent aussi un modèle conditionnel, utilisé afin de capturer la distribution multi-modale des valeurs de chrominance lors de la colorisation. Enfin, Royer *et al.* (2017) ont développé une méthode probabiliste permettant de proposer plusieurs scénarios de colorisation pour une même image en niveau de gris. Pour cela, les auteurs ont créé un VAE, qui fonctionne à la manière d'un modèle auto-régressif conditionnel. Cette approche permet là-aussi de tenir compte de la nature multi-modale du problème de colorisation, en apprenant l'ensemble de la distribution conjointe des valeurs de luminance et de chrominance.

Ce passage en revue des techniques développées pour la colorisation permet ainsi de rendre compte de la diversité des méthodes avancées pour résoudre ce problème, ainsi que les contraintes associées. A nouveau, la question de l'optimisation est primordiale, dans la mesure où le choix d'une ou plusieurs fonctions objectif conditionne la qualité spatiale et radiométrique des résultats en sortie. Le manque de techniques visant spécifiquement les produits géographiques est également évident, ce qui participe à légitimer ce travail. Le développement d'une méthode repose donc sur ces différents points mis en avant dans le Chapitre 1, avec (1) l'automatisation du traitement, (2) la prise en charge de photographies aériennes, (3) l'extraction ou l'apprentissage d'attributs pertinents, et enfin (4) l'entraînement d'un modèle capable de traiter la colorisation dans sa dimension multi-modale, c'est-à-dire apprendre l'ensemble des couleurs associées à une même sémantique. Compte-tenu de ces besoins, les méthodes d'apprentissage profond semblent être les plus adaptées, car elles opèrent sans intervention extérieure et apprennent de façon autonome un corpus de représentations, mis à jour au cours de l'entraînement. Ces points sont expliqués plus en détails dans le Chapitre 2.

2 Données et méthode

La production d'une image en couleurs à partir d'un cliché monochromatique constitue un problème qui peut être perçu de différentes manières, à la fois sur les plans mathématique et algorithmique. Cela conditionne les développements méthodologiques envisagés, ainsi que la façon dont les données sont mobilisées. Afin de proposer des techniques de colorisation adéquates, il est donc avant tout nécessaire de comprendre la structure des données dont nous disposons, et d'évaluer les éventuels apports de la chrominance. Ces différents points permettent alors d'appuyer le choix d'une méthode en particulier.

L'Annexe A explique la manière dont le travail s'est organisé et fournit une présentation générale de la méthodologie employée.

2.1 Analyse préliminaire des produits à coloriser

Les caractéristiques des orthophotographies historiques diffèrent largement de celles des clichés actuels, et ce pour différentes raisons qui sont expliquées dans les sous-parties à suivre. Dans le cadre de ce travail, le premier objectif a été d'analyser ces produits afin de disposer d'un catalogue décrivant leurs spécificités, au regard des applications géographiques typiquement menées sur leurs homologues numériques en couleurs naturelles. Ce point suppose également d'évaluer l'apport de la couleur, permettant donc de justifier l'importance méthodologique du travail ici proposé.

2.1.1 Caractéristiques générales des orthophotographies anciennes

La série temporelle d'orthophotographies à coloriser est constituée de quatre produits panchromatiques, qui décrivent le territoire de la Communauté Urbaine de Strasbourg (CUS), ou de l'Eurométropole de Strasbourg (EMS) selon la date de référence. Les clichés se répartissent sur une période de 46 ans, entre 1932 et 1978, et rendent compte, par exemple, des dynamiques de remembrement des exploitations agricoles, ou encore de l'étalement urbain après-guerre. Cela justifie donc un intérêt pour ces produits historiques, qui permettent de comprendre

Chapitre 2. Données et méthode

en rétrospective l'organisation spatiale actuelle de l'intercommunalité. Bien que ces clichés aient tous été développés sur un support argentique à l'aide d'un procédé chimique, chacun possède des traits particuliers détaillés dans la Table 4.

Année	Résolution (cm)	Saison	Dégradations
1932	20	Estivale	Effet de vignettage. Portions de l'image floutées. Supports analogiques dégradés, avec taches, égratignures, bruit systématique...
1956	50	Hivernale	Le tableau d'assemblage des mosaïques est visible par endroit. Effet de vignettage. Quelques dégradations sont également observables, avec des traces blanches au Nord de Strasbourg, ainsi qu'un léger bruit.
1964	20	Printemps	Pas de mosaïque disponible. Les tuiles sont mal géoréférencées. Effet de vignettage. Supports analogiques dégradés, avec taches, égratignures, bruit systématique...
1978	30	Estivale	Le tableau d'assemblage des mosaïques est visible et incorrect par endroits, avec un problème de superposition de différentes tuiles. Effet de vignettage. Supports analogiques dégradés, avec taches, égratignures, bruit systématique...

TABLE 4 – Description des produits raster panchromatiques constituant la base historique.

Différents problèmes de géoréférencement et de mosaïquage ont été observés, en particulier pour les clichés disponibles sur les années 1964 et 1978. Ils ont donc été corrigés, étape présentée dans l'Annexe B. Cela n'a pas été décrit dans le corps du présent document, dans la mesure où ces traitements ne participent en rien aux développements portant sur la colorisation de produits historiques.

Les orthophotographies à disposition sont affectées par des dégradations diverses, qui questionnent chaque étape située en amont de la publication des clichés. Selon Chevallier (1965) et Warner (1995), la qualité d'une photographie dépend ainsi des caractéristiques du système objectif-émulsion, parmi lesquelles se trouvent :

- (1) Les propriétés de la scène ou de l'objet capturé (contraste, luminosité, atmosphère, ...);
- (2) Les propriétés du capteur (objectif photographique, lentille);
- (3) Les propriétés du film sur lequel est développée la photographie (granularité, pouvoir résolvant, plage tonale, ...);
- (4) La qualité d'impression (instrumentation, compétences).

Les caractéristiques des orthophotographies anciennes en font ainsi des produits plus délicats à manipuler que leurs homologues numériques en couleurs naturelles. Il est alors question de savoir si ces spécificités limitent leur utilisation dans le cadre d'applications géographiques, et si l'apport de la chrominance permet de simplifier et de généraliser leur usage.

2.1.2 L'apport de la couleur pour les applications géographiques

Afin de répondre à cette question, nous avons choisi de développer une méthode simple, reproductible et facile à mettre en œuvre, même avec des capacités de calcul limitées.

Pour évaluer l'apport de l'information colorimétrique, une approche classique de classifi-

2.1. Analyse préliminaire des produits à coloriser

cation d'image a été proposée. Pour cela, un couple de scènes Pléiades a été utilisé, l'une panchromatique et l'autre en couleurs, constituant ainsi une paire $\{C, P\}$, avec C l'image en couleurs et P son homologue panchromatique. Leurs spécifications sont renseignées dans la Table 5. La zone d'étude a été sélectionnée de sorte à ce que plusieurs catégories de surfaces soient visibles, permettant d'évaluer les capacités du modèle sous différentes conditions. Sur la même emprise, une cartographie de l'occupation du sol en huit classes, produite par le LIVE en 2012, a également été extraite puis rastérisée avec une résolution de 2m.

	Pléiades MS	Pléiades PN
Date	01/04/2012	01/04/2012
Emprise	Strasbourg	Strasbourg
Rés. spatiale (m)	2	0,5 (2*)
Rés. spectrale	4 bandes (B, V, R, PIR)	1 bande (PAN)
Type	Entier	Entier
Profondeur (bits)	16	16

TABLE 5 – Caractéristiques du couple d'images satellitaires Pléiades utilisées pour tester l'apport de la couleur. La notation * signifie que le produit PN, initialement distribué à 50cm, a été rééchantillonné à 2m afin d'obtenir une géométrie similaire à celle de l'image MS.

La méthode consiste ainsi à classifier séparément les produits MS et PN afin d'obtenir une cartographie de l'occupation du sol sur la zone d'étude. Compte-tenu du fait que les images panchromatiques ne sont constituées que d'un seul plan d'information, il est préférable d'aller vers des techniques d'extraction d'attributs pour compléter le jeu de données et mieux discriminer les surfaces. En effet, ces méthodes, qu'elles soient supervisées ou non, renseignent sur la distribution des valeurs des pixels selon des approches statistiques, structurelles ou spectrales (Yang et Zhu, 1998; Benediktsson *et al.*, 2003). C'est notamment le cas des analyses de texture, qui fournissent un contexte spatial important à la compréhension de la sémantique d'une image. En effet, elles étudient l'agencement des niveaux de gris à l'aide, par exemple, de noyaux à taille fixe pour les approches basées sur les matrices de cooccurrence, ou d'une décomposition par ondelettes pour les analyses multi-échelles (He et Wang, 1990).

Afin de disposer d'une méthode simple à mettre en œuvre, facilement reproductible et peu exigeante d'un point de vue informatique, deux approches basiques ont été retenues pour l'analyse de texture sur les images panchromatiques. La première consiste à calculer des motifs binaires locaux ou *Local Binary Patterns* (LBP) sur l'extrait à disposition. Ce descripteur, particulièrement utilisé en vision assistée par ordinateur, est théoriquement indépendant des effets de rotation et/ou d'illumination, selon la méthode retenue (Ojala *et al.*, 2002). L'implémentation du LBP proposée dans le cadre du projet Scikit-Image (van der Walt *et al.*, 2014) a été utilisée, avec un nombre de voisins $N = 24$, un rayon de recherche $R = 3$, et avec les méthodes *par défaut* et *uniforme*. La seconde analyse consiste à calculer un jeu simple de textures de Haralick grâce à l'implémentation de l'Orfeo Toolbox (Grizonnet *et al.*, 2017). Celles-ci sont basées sur les matrices de cooccurrence des niveaux de gris, générées pour chaque pixel dans un voisinage défini par un noyau w (Haralick *et al.*, 1973; Haralick, 1979). Parmi les 29 descripteurs initialement conçus par l'auteur, seulement 8 ont été retenus puis

Chapitre 2. Données et méthode

calculés avec une fenêtre de taille $w = 5$. Ces attributs sont les suivants : *energy*, *entropy*, *correlation*, *inverse difference moment*, *inertia*, *cluster shade*, *cluster prominence* et *Haralick correlation*.

Ces informations spectrales et texturales ont été combinés afin de définir cinq scénarios de classification, sur les produits panchromatiques tout d'abord, qui sont les suivants :

- (1) S1 : Niveaux de gris + 2 canaux de LBP ;
- (2) S2 : $\text{ACP}(\sigma^2 \leq 0,85)$ Niveaux de gris + 8 canaux de texture Haralick ;
- (3) S3 : Niveaux de gris + 8 canaux de texture Haralick ;
- (4) S4 : $\text{ACP}(\sigma^2 \leq 0,85)$ Niveaux de gris + 2 canaux de LBP + 8 canaux de texture Haralick ;
- (5) S5 : Niveaux de gris + 2 canaux de LBP + 8 canaux de texture Haralick.

Les analyses en composantes principales (ACP) résument l'information contenue dans les canaux, afin de réduire les redondances et d'accélérer l'apprentissage.

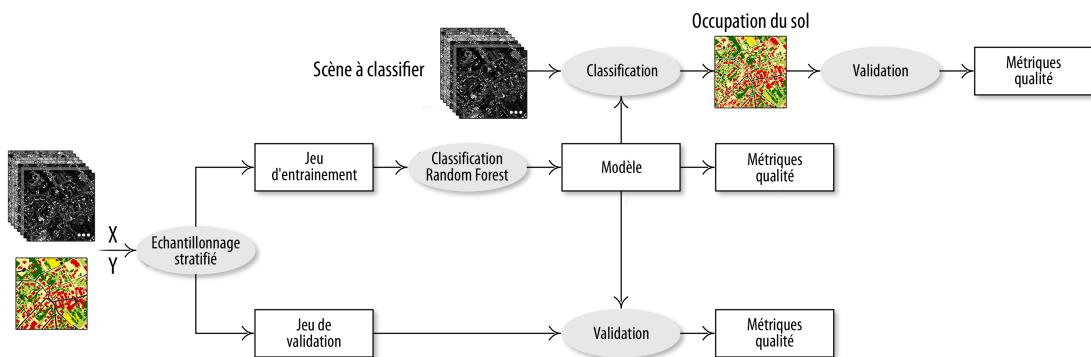


FIGURE 3 – Méthode de classification employée pour évaluer l'apport de la couleur.

Pour chaque scénario, une classification a ensuite été réalisée à l'aide d'un classifieur de Forêt Aléatoire ou *Random Forest*, grâce à l'implémentation proposée dans le cadre du projet Scikit-Learn (Pedregosa *et al.*, 2011). Cet algorithme de classification supervisée s'articule autour d'un nombre n de prédicteurs, ou arbres décisionnels, qui sélectionnent chacun de façon aléatoire un certain nombre d'échantillons et de variables dans un jeu de données d'entraînement (Breiman, 2001). Cette méthode est particulièrement robuste au sur-apprentissage et au bruit, lui permettant d'atteindre des performances intéressantes pour la classification d'images en télédétection notamment (Belgiu et Drăguț, 2016). Le classifieur Random Forest ici proposé est constitué de 100 arbres, et travaille selon une approche orientée pixel. Celui-ci prend X en entrée, les attributs radiométriques et texturaux, et cherche à prédire Y, l'occupation du sol (Figure 3).

Un échantillonnage stratifié a été effectué sur les extraits d'images, afin de récupérer 50% des pixels pour le jeu d'entraînement, et 50% pour la validation. Nous avons également veillé à obtenir une distribution équilibrée des classes, à l'aide de méthodes d'*oversampling* pour les catégories sous-représentées. Concernant l'apprentissage du modèle, ses hyper-paramètres ont été optimisés pour chaque scénario à l'aide d'une grille de recherche dense, dont les valeurs sont spécifiées dans la Table 6.

2.1. Analyse préliminaire des produits à coloriser

Enfin, le scénario avec les meilleurs résultats a été repris pour classifier les produits en couleurs, afin d'évaluer l'apport de la chrominance à rapport égal. Le canal panchromatique a ici simplement été remplacé par les bandes R, V et B.

Paramètre	Description	Valeurs testées
max_depth	Profondeur maximale de l'arbre.	None, 3
max_features	Nombre d'attributs dont il faut tenir compte pour segmenter un nœud.	None, sqrt, 1, 3
min_samples_split	Nombre minimum d'échantillons requis pour segmenter un nœud.	1, 3, 10
min_samples_leaf	Nombre minimum d'échantillons requis pour constituer une feuille.	1, 10, 25
bootstrap	Utilisation ou non d'échantillons <i>bootstrap</i> pour la construction des arbres.	True, False
criterion	Fonction utilisée pour mesurer la pureté d'une segmentation des branches.	Gini, Entropy

TABLE 6 – Description et valeurs des paramètres utilisés pour la grille de recherche dense (modifié de Pedregosa *et al.* (2011)). Les valeurs testées ont été choisies empiriquement, compte-tenu des données à disposition en particulier.

En complément, la même méthodologie a été mise en œuvre cette fois-ci sur des orthophotographies, dont les spécifications sont plus proches de celles des produits à coloriser, notamment en termes de résolution spatiale.

Pour cela, un extrait de l'orthophotographie RVB de la CUS, datant de l'année 2013, a été utilisé. D'une résolution de 13cm au départ, l'image a été rééchantillonnée à 30cm afin de disposer d'un produit géométriquement proche de ceux que nous cherchons à coloriser. Une bande pseudo-panchromatique a été calculée en moyennant les valeurs des canaux colorimétriques, pour constituer une paire $\{C, P\}$. Cette approche permet d'obtenir artificiellement deux extraits, capturés dans des conditions complètement similaires, et nous affranchit donc des changements liés aux dynamiques spatiales et conditions de prise de vue.

Afin de simuler l'apport de la couleur sur des produits historiques, nous avons également cherché à reproduire les caractéristiques des clichés développés sur un support argentique. Pour cela, une combinaison de dégradations radiométriques et spatiales a été proposée. Les produits C et P ont donc été dégradés à l'aide d'un flou gaussien — σ sélectionné aléatoirement entre 1,0 et 1,25 — puis d'un bruit gaussien — σ sélectionné aléatoirement entre 1×10^{-4} et 3×10^{-3} —. Fixer ces plages de valeurs a nécessité l'analyse de l'ensemble des orthophotographies qui constituent la série temporelle à disposition. Les bornes définies pour le bruit ont été obtenues à l'aide d'une méthode d'estimation proposée par Immerkær (1996). La plage des valeurs de variance pour le flou gaussien a quant à elle été définie empiriquement sur les produits. A noter enfin que cette méthode de dégradation est à adapter selon le nombre de canaux, la profondeur des pixels et les caractéristiques des surfaces représentées sur la scène.

L'ensemble des produits et attributs calculés, ainsi que les dégradations apportées aux images, sont présentés sur la Figure 4. Les mêmes scénarios de classification que pour les extraits d'images Pléiades ont finalement été mis en œuvre, permettant cette fois-ci d'obtenir une information sur l'apport de la couleur pour des produits actuels et historiques simulés.

Chapitre 2. Données et méthode

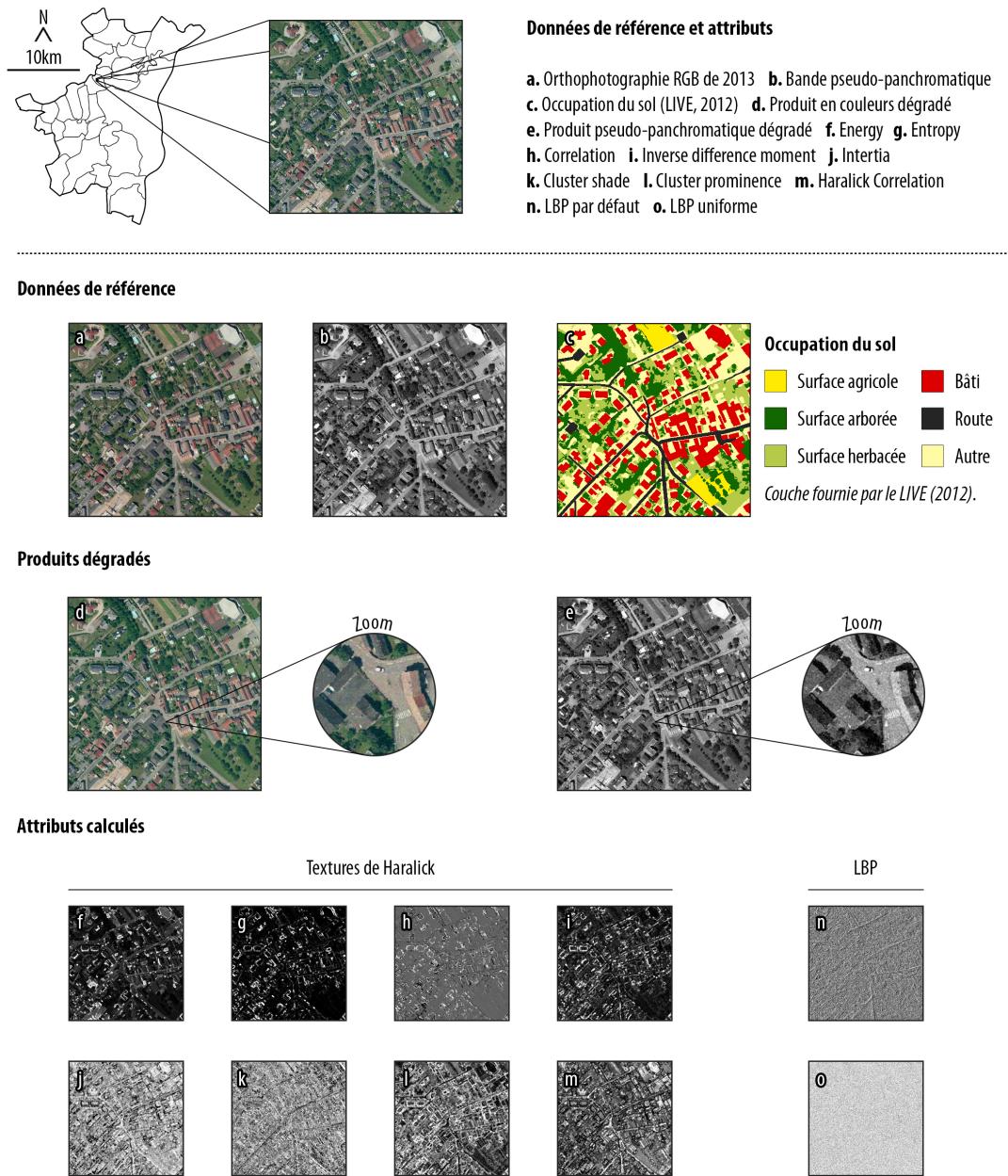


FIGURE 4 – Produits et attributs utilisés pour évaluer l’apport de la couleur.

2.2 Colorisation des orthophotographies anciennes

Une analyse des travaux décrits dans le Chapitre 1 a permis de mettre en évidence un manque dans les développements menés pour l’exploitation des données géographiques. En effet, les modèles proposés par les auteurs ont été entraînés à partir de produits issus de dépôts comme ImageNet, capturés selon des prises de vue frontales, obliques, ou en contre-plongée par exemple... Utiliser ces modèles tels quels sur des produits pris à la verticale, comme cela peut

2.2. Colorisation des orthophotographies anciennes

être le cas pour des photographies aériennes anciennes, revient donc à faire de l'apprentissage par transfert, ou *transfer learning*. Cette méthode consiste à entraîner un modèle pour acquérir une connaissance sur un problème donné, puis à le réutiliser pour réaliser une nouvelle tâche plus ou moins proche de celle pour laquelle le modèle a été entraîné au départ (Pratt, 1993). L'apprentissage par transfert est efficace quand les attributs mobilisés par le modèle sont généraux, mais devient plus difficile à mettre en œuvre lorsque les domaines ciblés sont différents. Dans ce cas, il est possible d'ajuster légèrement le modèle en réapprenant ses paramètres θ , à l'aide de données répondant au problème traité. Si les résultats ne sont toujours pas ceux escomptés, il est alors nécessaire de réapprendre complètement le modèle (Yosinski *et al.*, 2014).

Certaines méthodes état de l'art ont été testées sur des extraits d'orthophotographies aériennes anciennes, pour n'obtenir que des résultats de colorisation peu convaincants (Figure 5). Ce constat renvoie donc au fait que les domaines "prise de vue non verticale" et "prise de vue verticale" sont trop différents pour réutiliser les attributs appris sur l'un ou sur l'autre.

Acarta, Californie (1956)



Strasbourg, France (1964)



FIGURE 5 – Résultats de colorisation pour des extraits de photographies aériennes anciennes. Les images (a) et (b) sont des extraits de photographies aériennes en niveaux de gris ; (c) et (d) sont les colorisations réalisées à l'aide du modèle développé par Zhang *et al.* (2016) ; puis (e) et (f) sont les colorisations réalisées à l'aide d'un prototype de modèle non optimisé que nous avons construit et entraîné sur 25 itérations, à partir de 500 imagettes test de 64×64 pixels.

Aucune méthode état de l'art n'a donc pu être mobilisée telle quelle, signifiant qu'une série

de tests a été réalisée afin de choisir parmi différentes familles de modèles, architectures, régularisations et optimisations. Ces essais sont présentés et résumés dans les sous-parties suivantes. Ils nous ont ainsi permis de préparer une méthode d'apprentissage spécifique au problème de la colorisation des orthophotographies anciennes.

A noter que les développements ont principalement été réalisés avec PyTorch v0.4.0, une API de Torch disponible en Python (Collobert *et al.*, 2011; Paszke *et al.*, 2017). L'Annexe C présente de façon générale la manière dont l'infrastructure informatique a été mise en place.

2.2.1 Développement d'une photothèque pour la colorisation

Compte-tenu du fait qu'il n'est pas possible de réutiliser directement les travaux jusqu'à présent menés sur la colorisation, nous avons choisi d'entraîner un nouveau modèle à l'aide d'une base d'apprentissage dédiée. Nous proposons ainsi une photothèque $\Lambda = \{C, P\}$ dans laquelle sont stockés des couples couleur C et panchromatique P . Quelques tests ont également été menés localement avec une photographie infrarouge couleur I , donnant lieu à $\Lambda = \{C, P, I\}$.

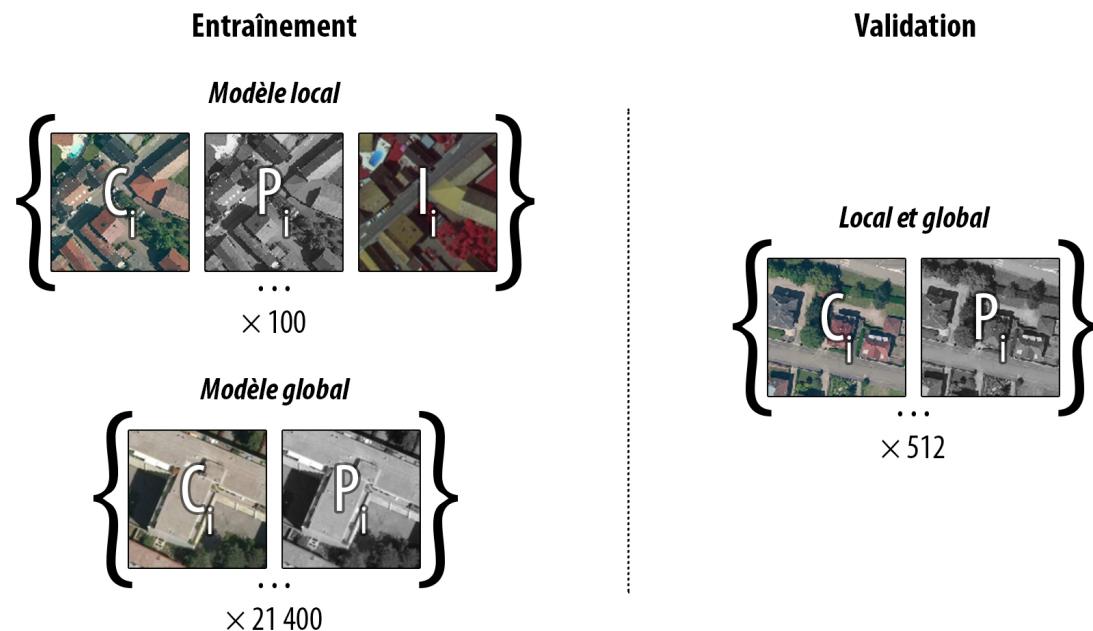


FIGURE 6 – Structure de la photothèque utilisée pour l'apprentissage et la validation du modèle de colorisation.

La structure générale de la photothèque est présentée sur la Figure 6, puis détaillée dans les sous-parties suivantes. Sans trop entrer dans les détails, puisque ces points sont expliqués ultérieurement, deux bases ont été créées pour l'entraînement de modèles de colorisation. La première, locale et modeste en termes de contenu, a été développée pour entraîner puis comparer trois algorithmes sur la commune de Niederhausbergen. La base globale est ensuite utilisée pour apprendre un nouveau modèle à partir de l'algorithme retenu. A chaque fois, les

2.2. Colorisation des orthophotographies anciennes

Résultats sont évalués à l'aide d'une base de validation, constituée de 512 clichés.

2.2.1.1 La mise au point d'une base d'entraînement globale

D'un point de vue méthodologique, la base globale est la première à avoir été développée, ce qui explique pourquoi nous avons choisi de la présenter en premier. En effet, la base locale a été constituée seulement après avoir réalisé qu'il nous faudrait un jeu d'entraînement de taille plus petite pour pouvoir comparer différents algorithmes, sans trop perdre de temps ni d'argent dans la location du serveur GPU.

Étant donné que l'objectif est ici de coloriser des clichés panchromatiques pris sur le territoire de l'EMS, une orthophotographie RVB datant de 2013 a été utilisée comme base principale. Afin de travailler sur des résolutions proches de celles des produits historiques, cette référence couleur a été rééchantillonnée à 30cm et 50cm. Cela suppose ainsi un meilleur apprentissage des sémantiques pour différentes échelles.

A partir des produits rééchantillonnés, ce sont 17100 imagettes de 128×128 pixels qui ont été extraites aléatoirement. L'échantillonnage a été réalisé compte-tenu des quatre classes d'occupation du sol représentées par la base de données CIGAL 2012 de niveau 1 (territoires artificialisés, agricoles, arborés et surfaces en eau), en cherchant à obtenir une distribution équilibrée de ces postes. Une fois collectés, les extraits ont été convertis individuellement en niveaux de gris à l'aide d'un outil proposé dans le cadre du projet Scikit-Image, afin d'aboutir aux couples $\{C, P\}$. La formule utilisée pour la conversion est décrite par l'Équation 1. A noter que les poids utilisés pour chacune des bandes sont fixes, ce qui permet d'obtenir des valeurs de pixels cohérentes d'un cliché à un autre.

$$P = 0.2125R + 0.7154V + 0.0721B \quad (1)$$

Le fait de partir d'un produit en couleurs puis de calculer une bande monochromatique nous permet ainsi d'obtenir à une image pseudo-panchromatique. L'utilisation d'un cliché panchromatique réel nécessiterait que celui-ci soit parfaitement coregistré avec le produit RVB, ceci afin d'apprendre un modèle correct sur le plan géométrique. L'absence de changements d'occupation du sol ou de phénologie des surfaces serait également un pré-requis nécessaire à l'obtention de couples adéquats. Cela explique pourquoi une simple réduction est ici réalisée, bien que ce raccourci méthodologique puisse lui aussi générer des erreurs. En effet, les produits panchromatique et pseudo-panchromatique sont *a priori* différents sur le plan radiométrique. Nous partons donc du postulat que l'information nécessaire à l'obtention d'une colorisation plausible est contenue dans la sémantique de l'image plutôt que dans la valeur des niveaux de gris.

Afin d'améliorer les capacités de colorisation du modèle, la photothèque a été complétée par des produits issus d'autres sources, puis convertis en niveaux de gris. Les informations

Chapitre 2. Données et méthode

générales de ces données sont décrites dans la Table 7. Parmi celles-ci se trouvent des images satellitaires, utilisées notamment pour des applications en classification et segmentation sémantique. Elles décrivent des objets rares et très spécifiques, par exemple les terrains de sport, piscines, échangeurs autoroutiers, etc.

Nom	Extraits	Type	Résolution (cm)	Référence
Ortho. 2013 (v1)	8550	Aérien	30	—
Ortho. 2013 (v2)	8550	Aérien	50	—
Complément photo.	900	Aérien	—	—
INRIA	2900	Aérien	30	Maggiori <i>et al.</i> (2017)
AID	300	Satellite	Multiples	Xia <i>et al.</i> (2017)
RSSCN7	100	Satellite	Multiples	Zou <i>et al.</i> (2015)
UCMerced	100	Satellite	30	Yang et Newsam (2010)

TABLE 7 – Description générale des produits utilisés pour constituer la photothèque.

A noter que les produits satellites (AID, RSSCN7 et UCMerced) n'ont finalement pas été utilisés dans l'itération finale de ce travail. En effet, bien qu'ayant des sémantiques proches de celles des orthophotographies, il semblerait que les différences d'échelles soient trop importantes. Les colorisations générées sont alors peu plausibles, avec des associations objet — chrominance incohérentes. Cette piste reste cependant à explorer pour une photothèque plus généraliste, en vue du développement d'un modèle multi-scalaire ou pyramidale par exemple.

Cliché	Numéro	Échelle	Support	Date
A	87	1/10285	Argentique	03/08/1986
B	102	1/10295	Argentique	03/08/1986
C	8	1/9969	Argentique	23/07/1990
D	9	1/9974	Argentique	23/07/1990
E	104	1/10411	Argentique	16/08/1992
F	86	1/18982	Argentique	27/06/1995
G	103	1/18992	Argentique	27/06/1995
H	103	1/10290	Argentique	03/08/1995

TABLE 8 – Description des photographies en couleurs naturelles utilisées comme complément à la photothèque (IGN, 2018).

Nous avons également constaté que les produits historiques présentent des sémantiques absentes des jeux de données raster récents. C'est notamment le cas des effets de relief en milieu urbain par exemple, avec des distorsions géométriques importantes au niveau des bâtiments hauts. Dans ce cas de figure, la façade des bâtisses est visible, phénomène qui n'est pas représenté sur l'orthophotographie de 2013 et les produits issus des autres bases de données. Pour cette raison, un complément de clichés en couleurs, distribués par l'IGN et capturés dans le centre strasbourgeois entre 1986 et 1995, a également été utilisé. Les références de ces produits sont décrites dans la Table 8. Ce sont ainsi 900 extraits, toujours d'une dimension de 128 × 128 pixels, qui ont été incorporés à la photothèque. A noter que ces photographies n'ont pas été géoréférencées, dans la mesure où seule l'information spectrale est prise en compte par le modèle de colorisation. Cela donne ainsi un total de 21 400 imagettes utilisées pour l'entraînement d'un modèle de colorisation.

2.2. Colorisation des orthophotographies anciennes

Une fois la photothèque constituée, différentes transformations — aléatoires ou non — sont appliquées aux images lorsqu'elles sont prises en charge par le modèle. Ces transformations sont également valables pour la base locale, présentée plus tard.

Les premières consistent en des transformations aléatoires, qui modifient les caractéristiques géométriques et radiométriques des images. Nous proposons ainsi d'appliquer les opérations suivantes aux couples $\{C, P\}$: (1) bruit gaussien, (2) flou gaussien, (3) rotation, (4) renversement horizontal et (5) renversement vertical. Ces transformations appartiennent aux méthodes d'augmentation de données. Elles permettent ainsi d'apprendre artificiellement de nouvelles sémantiques, en modifiant légèrement les caractéristiques des produits, et améliorent ainsi les capacités de généralisation des modèles développés (Taylor et Nitschke, 2017). A noter que les transformations radiométriques — bruit gaussien et flou gaussien — ne sont appliquées qu'aux images monochromatiques, afin de préserver l'information colorimétrique cible. Les valeurs de σ sont les mêmes que celles proposées dans la Partie 2.1.2, c'est-à-dire comprises entre 1,0 et 1,25 pour le flou, puis entre 1×10^{-4} et 3×10^{-3} pour le bruit. Ces transformations radiométriques sont ainsi supposées reproduire l'allure des photographies de la base géohistorique.

A noter que l'utilisation d'un bruit gaussien a finalement été abandonnée dans l'itération finale de ce travail. En effet, la valeur de σ est à adapter selon la donnée (encodage, qualité du produit, etc.), et selon la nature de la surface représentée sur l'image (contraste, luminosité, etc.). Son utilisation perturbait trop l'apprentissage du modèle pour être maintenue. Cette méthode reste cependant intéressante pour reproduire les caractéristiques des produits historiques. Une ouverture possible serait de développer une méthode permettant d'ajouter un bruit adaptatif aux images, selon l'occupation du sol ou une information relative à l'entropie des valeurs des pixels par exemple.

Nous appliquons également d'autres transformations aux images, systématiquement cette fois-ci, qui ne participent pas à l'augmentation des données et sont nécessaires au processus de colorisation. Celles-ci sont : (1) le passage vers l'espace colorimétrique Lab, (2) la transformation des images en tenseurs et (3) la normalisation des valeurs des pixels.

L'utilisation de l'espace colorimétrique Lab plutôt que RVB a été préconisée dans de nombreuses publications (Agrawal et Sawhney, 2016; Iizuka *et al.*, 2016; Isola *et al.*, 2016; Zhang *et al.*, 2016; Fu *et al.*, 2017; Lal *et al.*, 2017; Royer *et al.*, 2017; Liu *et al.*, 2018). Celui-ci s'organise autour de 3 canaux. Le premier correspond à la luminance $L \in [0, 100]$, que nous pouvons ici appartenir à l'image P du couple $\{C, P\}$. Viennent ensuite deux canaux de chrominance décorrélés l'un de l'autre, avec $a \in [-127, 128]$ qui correspond à l'axe *vert → rouge*, puis $b \in [-128, 127]$ qui correspond à l'axe *bleu → jaune*. Cet espace colorimétrique présente plusieurs avantages, qui justifient son utilisation. Le premier consiste évidemment à n'apprendre que deux canaux — a et b —, plutôt que trois — R, V et B —. Cela permet ainsi de réduire le nombre de paramètres θ à apprendre et d'obtenir une fonction objectif plus stable. Le fait de pouvoir séparer la luminance de la chrominance autorise également à corriger

Chapitre 2. Données et méthode

automatiquement certains défauts présents sur les orthophotographies anciennes, comme le bruit gaussien ou les effets de luminosité par exemple. A noter par ailleurs que l'analyse des avantages et inconvénients des différents espaces colorimétriques ne fait pas partie de l'analyse. Les espaces Lab et RVB sont représentés sur la Figure 7 à titre comparatif.

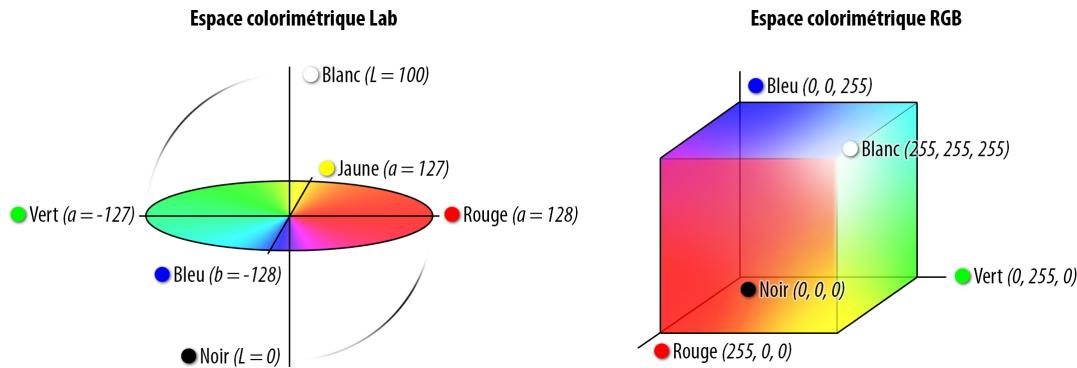


FIGURE 7 – Représentation des espaces colorimétriques Lab et RVB.

Les deux dernières transformations sont quant à elles indispensables au fonctionnement de PyTorch et des modèles développés. Les images sont en effet initialement stockées sous forme matricielle. Il est nécessaire de les transformer en tenseurs, structure de données spécifique, pour qu'elles soient lues et prises en charge par PyTorch. Une normalisation est ensuite effectuée afin de travailler dans une échelle de valeurs comprises entre -1 et $+1$, comme indiqué dans la vaste majorité des travaux portant sur la question de la colorisation. En effet, il semblerait que cette étape facilite la manipulation de données continues par le modèle.

Des exemples de transformations sont proposés sur la Figure 8. A chaque paire-mère $\{C, P\}$ sont ainsi associées des paires-filles, obtenues aléatoirement avant d'être prises en charge par le modèle de colorisation.

2.2.1.2 La mise au point d'une base d'entraînement locale

Le volume des données utilisé pour développer la photothèque globale est important, supposant donc un temps d'entraînement conséquent avant d'obtenir des modèles de colorisation robustes. Cela limite les possibilités en matière de comparaison des méthodes testées hors annexe, qui sont au nombre de trois comme présenté dans la Partie 2.2.2.

Afin de simplifier la mise en œuvre de ces tests, une petite base de données locale a été développée sur la commune de Niederhausbergen. Ce territoire a été retenu compte-tenu de sa taille modeste et de la diversité des paysages qu'il présente. La base ainsi constituée contient 100 imagettes d'une taille de 128×128 pixels, en couleur naturelles et en niveaux de gris. Elles ont été extraites à partir de l'orthophotographie RVB de 2013, rééchantillonnée à 50cm.

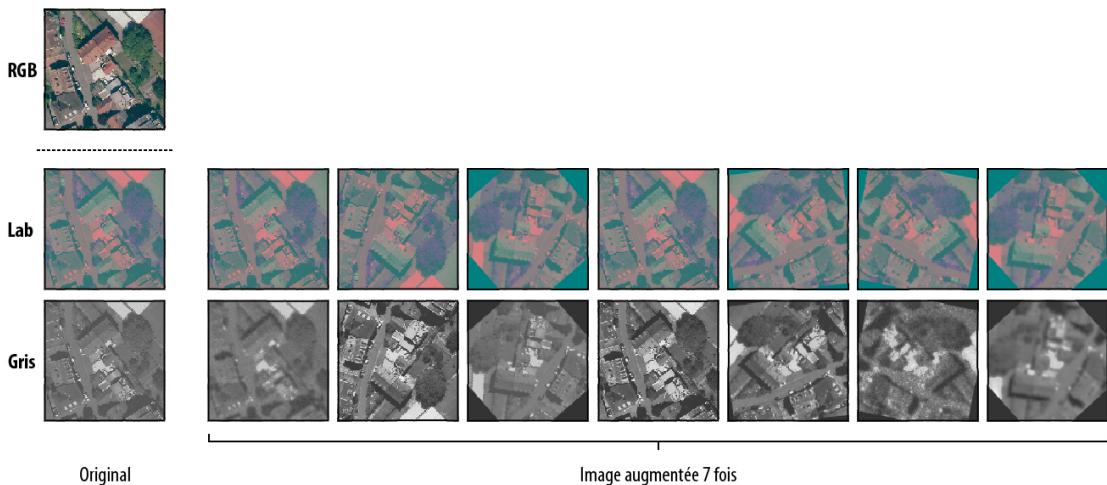


FIGURE 8 – Exemple de transformations aléatoires et systématiques appliquées à une imagette, puis utilisées pour augmenter le jeu de données.

Afin de tester les capacités d'apprentissage par transfert et de multispectralisation du modèle, des extraits de l'orthophotographie de 2012 en infrarouge couleur I , disponible sur le territoire de l'EMS et distribuée par GeoGrandEst, ont également été capturés sur les mêmes emprises. La base locale Λ' est ainsi constituée de triplets $\{C, I, P\}$. L'objectif consiste à réutiliser un modèle développé pour le visible, et pour lequel un ensemble d'attributs à déjà été appris, capable de générer des résultats de colorisation corrects. Ses paramètres θ , obtenus à l'itération retenue par l'opérateur-trice, servent donc d'initialisation de référence. L'apprentissage continue ensuite avec les couples d'images $\{I, P\}$, permettant ainsi d'affiner les attributs qui servent ici à générer une composition infrarouge couleur.

Pour cette base locale, les mêmes méthodes de préparation et d'augmentation de données que celles utilisées pour l'approche globale ont été employées.

2.2.1.3 La mise au point d'une base de validation

Afin d'évaluer la qualité des résultats de colorisation en sortie des modèles, une base de validation a été constituée. Pour ce faire, l'orthophotographie de 2013 rééchantillonnée à 30cm a servi de référence. Pour chacune des quatre classes d'occupation du sol représentées par la base de données CIGAL 2012 de niveau 1, 128 imagettes d'une taille de 128×128 pixels ont été extraites. Cela donne ainsi un total de 512 clichés. Nous pouvons en effet supposer que la qualité de la colorisation ne sera pas la même selon les caractéristiques et l'agencement des surfaces. Ces produits ont ensuite servi d'entrée au calcul des homologues pseudo-panchromatiques, obtenus de la même façon que pour la photothèque globale, et permettant finalement d'aboutir aux couples $\{C, P\}$.

Lors de l'étape de validation des modèles, l'image C est alors colorisée par la générateur, ce qui donne le produit \hat{C} . Les clichés C et \hat{C} sont finalement comparés à l'aide d'indicateurs

présentés plus tard, dans la Partie 2.2.3.

A noter qu'aucune base de validation n'a été constituée pour évaluer la technique de multispectralisation. En effet, l'objectif est seulement de montrer les capacités du modèle à apprendre des attributs transférables à une tâche de colorisation légèrement différente.

2.2.2 Développement d'un réseau de neurones profond

La mise au point d'un réseau de neurones nécessite de porter une réflexion sur différents aspects. Parmi eux se trouvent ses entrées et sorties, questions qui ont déjà été évoquées, mais aussi sa catégorie, son architecture et la méthode utilisée pour l'optimiser. D'autres points peuvent être soulevés, comme les techniques de régularisation employées pour améliorer les conditions d'apprentissage par exemple. Ces aspects méthodologiques et algorithmiques sont explorés dans les parties suivantes, ainsi que dans l'Annexe D pour les pistes non retenues.

2.2.2.1 Catégorie de modèle et fonction objectif

Parmi les familles de modèles présentées dans le Chapitre 1, seuls les GANs et leurs dérivés ont été retenus pour la colorisation de clichés anciens. Cela s'explique par leur simplicité de mise en œuvre et la qualité des résultats obtenus, avec un jeu de référence de petite taille et un temps d'entraînement limité. Différentes variations du GAN initialement proposé par Goodfellow *et al.* (2014a) ont ici été explorées : les cGANs classiques (Mirza et Osindero, 2014), les BEGANs (Berthelot *et al.*, 2017) et les DRAGANs (Kodali *et al.*, 2017). Chaque méthode possède des fonctions objectif spécifiques, pour le générateur G et le discriminateur D . A noter cependant que la base correspond systématiquement à un cGAN, avec l'image panchromatique P utilisée comme condition pour la prédiction des valeurs de chrominance de C (Figure 9). Seules les méthodes d'optimisation sont alors ajustées, pour reproduire les modes de fonctionnement des BEGANs et DRAGANs.

Concernant le cGAN ou GAN conditionnel classique, les fonctions objectif de G et D sont décrites par les Équations 2 et 3 respectivement. Les notations p et c correspondent aux extraits d'images panchromatiques et couleurs. Le produit panchromatique sert à conditionner spatialement le modèle, pour générer des valeurs de chrominance cohérentes avec la scène à coloriser.

$$L_G^{cGAN} = E[\log(D(G(p), p))] \quad (2)$$

$$L_D^{cGAN} = E[\log(D(c, p))] + E[\log(1 - D(G(p), p))] \quad (3)$$

L'instabilité des GANs, notée par de nombreux auteurs, a pu être observée à l'occasion des différents tests réalisés. En effet, les fonctions objectif antagonistes du générateur et du discriminateur oscillent fortement d'une itération à l'autre lors de l'apprentissage. Il est donc difficile d'évaluer quantitativement les performances d'un GAN, notamment en termes de

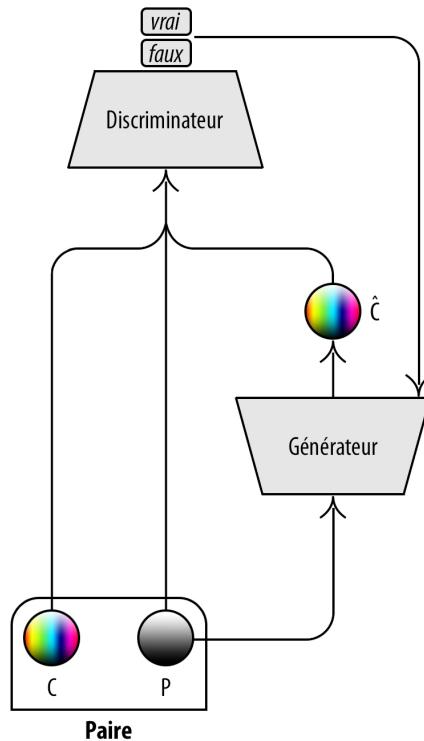


FIGURE 9 – Modèle général du réseau génératif antagoniste utilisé.

convergence vers une solution adaptée au problème. Une autre difficulté soulignée par Arjovsky et Bottou (2017) concerne cette fois-ci le discriminateur, qui atteint trop rapidement son *optimum* et parvient donc à distinguer sans faute les produits générés de ceux issus d'une distribution réelle. Cela s'explique par une disparition du gradient, le fait que les paramètres θ du modèle ne sont plus mis à jour, empêchant donc virtuellement d'aller plus loin dans l'apprentissage. D'un point de vue qualitatif également, les résultats obtenus témoignent d'un effet de *mode collapse*, le fait que le modèle ne parvienne à apprendre qu'une partie des modes de la distribution. Ce point peut être particulièrement importun dans le cas de la colorisation, puisque c'est un problème multi-modal (Annexe E). En effet, un même objet peut revêtir différentes couleurs, par exemple les toitures qui possèdent des tuiles parfois bleues, rouges, noires, etc. Plusieurs auteurs ont ainsi cherché à contourner ces problèmes, en proposant de nouvelles catégories de modèles plus robustes, les BEGANs et DRAGANs notamment.

Les BEGANs ont été développés par Berthelot *et al.* (2017), qui constatent généralement un apprentissage trop rapide du discriminateur, et un manque de diversité dans les résultats produits par le générateur. Afin de stabiliser l'entraînement et d'optimiser D , les auteurs proposent donc d'utiliser une métrique dérivée de la distance de Wasserstein, employée principalement pour les WGANs. Les Équations 4, 5 et 6 décrivent les fonctions objectif pour G , D et k , telles que proposées pour un BEGAN conditionnel. Le terme $k_t \in [0, 1]$ sert ici à équilibrer les fonctions objectif sur $L(c)$ et $L(G(p))$, afin de stabiliser l'apprentissage. Sa

Chapitre 2. Données et méthode

composante λ , dont la valeur est fixée à 1×10^{-3} , correspond au taux d'apprentissage de k_t . Le ratio de diversité γ vaut quant à lui $7,5 \times 10^{-1}$, et permet d'équilibrer les fonctions objectif de G et D . A noter que cette méthode a été proposée au départ pour une architecture auto-encodeur. Dans le cadre de ce travail, cette recommandation n'a pas été suivie, au profit d'une architecture U-Net comme présenté dans la Partie 2.2.2.2.

$$L_G^{BEGAN} = D(G(p), p) \quad (4)$$

$$L_D^{BEGAN} = D(c) - k_t(G(p), p) \quad (5)$$

$$k_{t+1} = k_t + \lambda(\gamma D(c, p) - D(G(p), p)) \quad (6)$$

Le développement des DRAGANs part quant à lui de l'hypothèse que l'effet de *mode collapse* résulte de la convergence du modèle vers des *equilibria* locaux non optimaux. La méthode présentée par Kodali *et al.* (2017) consiste ainsi à sanctionner la norme du gradient du discriminateur, afin d'éviter ces points qui restreignent le modèle à n'apprendre qu'un seul mode, ou la médiane de plusieurs modes minimisant l'erreur de prédiction. Il a aussi été montré que cette famille de modèles permet de stabiliser et d'accélérer l'apprentissage, et d'obtenir de meilleures prédictions qu'avec un GAN classique, peu importe l'architecture employée. Les Équations 7 et 8 décrivent les fonctions objectif pour G et D , telles que proposées pour un DRAGAN conditionnel. Le terme λ , dont la valeur est fixée à 10 par Kodali *et al.* (2017), est utilisé pour pondérer les gradients ∇ du discriminateur.

$$L_G^{DRAGAN} = E[\log(D(G(p), p))] \quad (7)$$

$$L_D^{DRAGAN} = E[\log(D(c, p))] + E[\log(1 - D(G(p), p))] + \lambda E[(|\nabla D| - 1)^2] \quad (8)$$

En complément, Isola *et al.* (2016) ont également proposé d'ajouter une métrique L1, pondérée par un facteur η , à la fonction objectif antagoniste du générateur G . Cette combinaison, décrite par l'Équation 9, est supposée donner des résultats de meilleure qualité, à la fois spatialement et spectralement. Les tests réalisés et rapidement présentés en Annexe D vont dans ce sens. En effet, la distance L1 donne une information sur la qualité de restitution du produit colorisé, et prévient donc l'apparition de certains artefacts.

$$L_G^{finale} = \underbrace{L_G^*}_{\text{Objectif G}} + \eta \times \underbrace{\sum_{i=1}^n |c_i - G(p_i)|}_{\text{Objectif L1}} \quad (9)$$

Nous avons ainsi suivi les recommandations de Isola *et al.* (2016), avec un paramètre $\eta = 100$. A noter que ce mode de régularisation peut cependant aboutir à des colorisations peu vibrantes. En effet, la métrique L1 minimise la différence absolue entre la référence et la prédiction. Elle peut aussi, en cas d'incertitude, trouver un *equilibrium* autour d'un point situé entre les modes de la distribution d'un même concept, afin de réduire l'erreur de prédiction du générateur. A

2.2. Colorisation des orthophotographies anciennes

l'occasion des tests réalisés, les toitures des bâtiments colorisés étaient généralement blanches, grises ou marron, alors que les deux modes de la distribution correspondent au orange et au noir pour la couleur des tuiles. Ce sont ainsi des effets que nous avons observés sur des objets pour lesquels il existe plusieurs scénarios de colorisation possibles, ou pour lesquels aucune référence n'est disponible dans la base d'entraînement utilisée lors des tests.

2.2.2.2 Architecture du modèle et méthodes de régularisation

Les fonctions objectif constituent un point critique lorsqu'il s'agit d'entraîner un réseau de neurones, puisqu'elles permettent d'orienter l'apprentissage vers une solution adaptée au problème étudié. Cependant, l'architecture du modèle conditionne elle-aussi la qualité des résultats obtenus, et nécessite une réflexion compte-tenu des sorties attendues. En effet, il faut voir les réseaux de neurones profonds comme un assemblage logique de blocs, chacun chargé de réaliser une opération donnée. Différents tests ont ainsi été menés pour trouver l'architecture la plus adaptée à la colorisation de produits géographiques, et sont succinctement présentés en Annexe D.

Les architectures que nous avons finalement retenues pour le générateur et le discriminateur sont détaillées sur la Figure 10. A noter qu'afin de simplifier la lecture de ce manuscrit, l'ensemble des fonctions utilisées n'a pas été défini dans le corps du texte mais dans le glossaire.

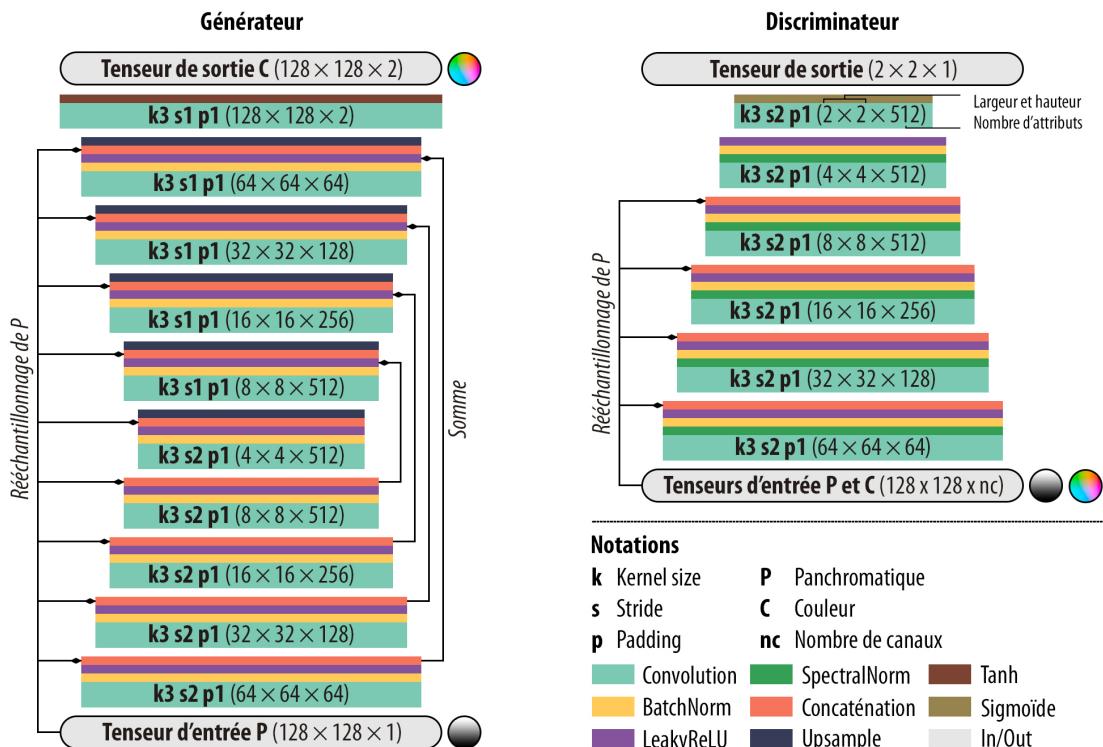


FIGURE 10 – Architectures retenues pour le générateur et le discriminateur. Les mêmes ont été utilisées pour le cGAN classique, le BEGAN et le DRAGAN.

Chapitre 2. Données et méthode

Parmi les expérimentations menées, l'architecture U-Net (Ronneberger *et al.*, 2015) est celle qui a permis d'aboutir aux résultats les plus pertinents pour le générateur. Constituée de couches entièrement convolutives, elle s'organise en deux parties. La première correspond à un goulet d'étranglement qui réduit progressivement la taille de l'image qui lui est passée en entrée, en longueur et en largeur. Elle permet de capturer la sémantique de la scène pour différentes échelles, en apprenant un nombre croissant d'attributs (64, 128, 256, 512). Toutes ces informations sont alors stockées en mémoire, pour pouvoir être utilisées plus tard. La deuxième partie est symétrique à la première, dans la mesure où la taille de l'image est progressivement restaurée par des méthodes de rééchantillonnage paramétriques ou non paramétriques, jusqu'à obtenir la sortie attendue. Lors de cette étape de restauration spatiale, les attributs stockés en mémoire sont additionnés symétriquement à ceux situés au-delà du goulet d'étranglement. Cela permet de combiner la dimension géographique aux sémantiques apprises à différentes échelles. Une dernière couche résume finalement toutes ces informations en deux attributs, qui correspondent ici à la chrominance. Les valeurs des canaux a et b prédicts sont finalement rééchelonnées dans l'intervalle $[-1, 1]$ par la fonction d'activation \tanh .

Pour le discriminateur, une architecture PatchGAN a été utilisée. Proposée par Isola *et al.* (2016) pour leur outil Pix2Pix, elle repose sur l'utilisation de noyaux de convolution pour évaluer si les patchs qui constituent l'image générée appartiennent ou non à une distribution réelle. C'est donc une approche locale, qui dénote des architectures classiquement utilisées pour les discriminateurs, puisque celles-ci travaillent à une échelle globale. L'avantage est donc de pouvoir cibler certaines régions de l'image, plutôt que d'obtenir un score qui décrit la scène dans son ensemble. En complément, nous avons également normalisé spectralement les sorties de chaque couche convulsive du discriminateur. Cette technique de normalisation permet de stabiliser l'apprentissage de D et de générer des images de haute qualité, comme l'ont montré Miyato *et al.* (2018) et Zhang *et al.* (2018).

Pour les deux modèles, l'image panchromatique est progressivement rééchantillonée puis concaténée au reste des attributs de chaque bloc afin de conditionner spatialement l'apprentissage, comme l'ont proposé Cao *et al.* (2017). Dans le cas du générateur, les attributs extraits dans le goulet d'étranglement sont conservés en mémoire, puis additionnés aux blocs symétriques. La combinaison de ces deux méthodes permet d'affiner les résultats, puisque l'information de bas niveau contenue dans les premières couches est alors restituée.

Au total, ce sont 2 434 et 1 984 attributs qui sont respectivement appris par le générateur et le discriminateur. Dans le premier cas, ils permettent la prédiction des canaux a et b d'une image en couleurs. Dans le second, ils servent à vérifier que le produit généré \hat{C} appartient à une distribution réelle, celle des produits de référence C .

Il est important de préciser enfin que l'architecture entièrement convulsive du modèle permet de traiter des images de tailles variées. Seulement, l'apprentissage doit être réalisé avec des clichés de dimension fixe, ici de 128×128 pixels, contrainte géométrique imposée par l'ensemble des bibliothèques d'apprentissage profond. Une fois le générateur entraîné, il peut

coloriser des images de n'importe quelles dimensions, du moment que celles-ci correspondent à des sommes de puissances de 2 (64×64 , 128×128 , 256×256 , ...).

A noter enfin que la sortie du modèle correspond à deux canaux a et b . Il est possible de les concaténer au tenseur de départ P , qui contient l'information panchromatique, permettant d'aboutir à une image dans l'espace Lab, P correspondant ici à la bande L . A des fins de visualisation ou pour d'autres usages, celle-ci peut ensuite être convertie en RVB. Les figures présentées dans le chapitre dédié aux résultats font donc appel à ces deux étapes de concaténation puis de conversion.

2.2.2.3 Phase d'apprentissage et optimisation du modèle

L'ensemble de la phase d'apprentissage est décrit par le Pseudo-code 1. L'optimisation du modèle suit les recommandations générales proposées par Goodfellow *et al.* (2014a), Isola *et al.* (2016), Salimans *et al.* (2016) et Miyato *et al.* (2018).

Certaines bonnes pratiques ont été adoptées afin de stabiliser l'apprentissage des modèles. Parmi celles-ci se trouvent le *label smoothing* proposé par Salimans *et al.* (2016), ainsi que l'initialisation des paramètres θ de G et D avec des valeurs tirées aléatoirement au sein d'une distribution gaussienne. Veuillez vous référer au glossaire pour plus de détails.

Pseudo-code 1 : Apprentissage du modèle de colorisation

Entrées : Batch d'imagettes homologues en couleurs et pseudo-panchromatiques
Sorties : Batch d'imagettes pseudo-panchromatiques colorisées

- 1 Transformations systématiques et aléatoires des imagettes contenues dans le batch ;
 // La préparation est réalisée par différentes classes, fonctions et méthodes d'augmentation de données, avant passage par le GAN
- 2 Définition du modèle ;
 // Les instances du générateur G et du discriminateur D sont créées
- 3 Initialisation ;
 // Les poids du générateur G et du discriminateur D sont initialisés à l'aide de valeurs tirées aléatoirement dans une loi normale
- 4 Définition des optimiseurs et fonctions objectif ;
- 5 **for** $i \leftarrow 0$ **to** n **do**
- 6 Génération d'un batch de canaux a et b des images \hat{C} avec G ;
- 7 Comparaison de \hat{C} et de C par le discriminateur ;
- 8 Calcul de l'erreur et mise-à-jour des paramètres de D ;
- 9 Génération d'un batch de canaux a et b des images \hat{C} avec G ;
- 10 Mise-à-jour des paramètres de G après calcul de l'erreur et retour de D ;
- 11 **end**

A chaque itération, les paramètres de D et G sont mis à jour séparément. Pour cela, deux optimiseurs sont utilisés. Le premier correspond à l'algorithme du gradient stochastique (SGD), qui travaille spécifiquement sur l'optimisation du discriminateur. Son utilisation a été

suggérée par Miyato *et al.* (2018), qui ont remarqué de meilleures performances lorsqu'utilisé pour la normalisation spectrale des sorties de chaque couche convective de D . Ses taux d'apprentissage lr et *momentum* ont été fixés à 2×10^{-4} et 0,9 respectivement. Concernant le générateur, un optimiseur Adam a été préféré, du fait de sa simplicité d'utilisation et de sa capacité à converger rapidement vers une solution adéquate, sans avoir à trop le paramétriser. Les valeurs de lr , β_1 et β_2 ont été fixées à 2×10^{-4} , 0,9 et 0,999 respectivement.

Une fois l'ensemble de ces points implémenté, nous avons tout d'abord entraîné les BEGAN, cGAN classique et DRAGAN à l'aide de la base de données locale. A l'issue de cette étape, nous avons ainsi retenu un modèle en particulier, compte-tenu d'une évaluation qualitative et quantitative des résultats de colorisation. Celui-ci est finalement entraîné avec la photothèque globale, afin d'apprendre une forme de $f(p) = c$ généralisable à l'ensemble du territoire de l'EMS. Après l'apprentissage, il est alors possible de passer une image P au générateur, qui prédit les valeurs de chrominance de la scène puis renvoie le cliché \hat{C} correspondant.

2.2.3 Évaluation des produits colorisés

Les modèles discriminants, utilisés pour la segmentation ou la classification d'images par exemple, disposent aujourd'hui d'une vaste palette d'indicateurs pour l'évaluation des performances d'un modèle. Dans le cas des réseaux génératifs cependant, il est question de savoir si les images obtenues en sortie du générateur sont réalistes ou non.

A ce jour, seul l'*Inception Score* (IS) proposé par Salimans *et al.* (2016) répond efficacement à ce besoin. En effet, cette métrique est fortement corrélée aux évaluations menées par des sujets humains sur des produits générés par apprentissage profond. Dans le cadre de cette étude, l'IS n'a cependant pas été retenu, car il a été formulé afin d'évaluer les capacités d'un modèle à (1) générer des images contenant chacune une classe principale et (2) à générer des images diversifiées, couvrant l'ensemble des classes du jeu d'entraînement ImageNet. Son utilisation requiert donc de disposer d'un classifieur, généralement le réseau Inception v3 (Barratt et Sharma, 2018), dont les sémantiques et thèmes ne correspondent pas à ceux des produits géographiques.

Étant dans l'incapacité d'utiliser l'IS, nous avons choisi de travailler avec des métriques classiques — MSE, PSNR et SSIM — calculées sur le jeu de validation, pour les modèles développés en local et en global. En effet, nous disposons de 512 couples $\{C, P\}$, répartis en quatre postes d'occupation du sol. L'image P est alors colorisée à l'aide d'un modèle en particulier, puis les valeurs de \hat{C} sont comparées à celles de la référence C .

La MSE permet de mesurer la dissimilarité entre deux images, selon une approche pixel à pixel. C'est une distance dont la formule est décrite par l'Équation 10. Elle correspond à la moyenne du carré des différences entre les produits en couleurs et leurs homologues obtenus à l'issue du processus de colorisation. Son domaine est compris dans l'intervalle $[0, +\infty]$. Ainsi, plus la valeur prise par la MSE diminue, meilleure est la correspondance entre la source et la cible.

2.3. Mise en œuvre de classifications historiques sur les produits colorisés

$$MSE(c, \hat{c}) = \frac{1}{n} \times \sum_{i=1}^n (c_i - \hat{c}_i)^2 \quad (10)$$

avec n le nombre de pixels dans l'image, puis c_i et \hat{c}_i les valeurs du pixel i pour les produits C et \hat{C} respectivement.

Le PSNR est un indicateur qu'il est possible de calculer à partir de la MSE, selon l'Équation 11. Il est utilisé pour évaluer la qualité de reconstruction du signal des images et s'exprime généralement en décibels (dB). Le domaine du PSNR est compris dans l'intervalle $[0, +\infty]$. Plus sa valeur augmente, plus la source et la cible sont similaires.

$$PSNR(c, \hat{c}) = 10 \log_{10} \times (255^2 / MSE(c, \hat{c})) \quad (11)$$

Enfin, le SSIM est un indicateur qui mesure la similarité entre deux images, selon une approche plus globale que les précédentes métriques cette fois-ci. Il tient compte simultanément de l'information apportée par la structure, la luminosité et le contraste (Wang *et al.*, 2004). En effet, l'erreur n'est pas calculée au niveau du pixel, mais à l'aide de noyaux de convolution qui renseignent donc d'une certaine manière sur la sémantique de l'image. La formule utilisée pour calculer le SSIM est renseignée par l'Équation 12.

$$SSIM(c, \hat{c}) = \frac{(2\mu_c\mu_{\hat{c}} + c_1)(2\sigma_c\sigma_{\hat{c}} + c_2)(2cov_{c\hat{c}} + c_3)}{(\mu_c^2 + \mu_{\hat{c}}^2 + c_1)(\sigma_c^2 + \sigma_{\hat{c}}^2 + c_2)(\sigma_c\sigma_{\hat{c}} + c_3)} \quad (12)$$

avec μ_c et $\mu_{\hat{c}}$ les espérances des valeurs de C et \hat{C} , σ_c et $\sigma_{\hat{c}}$ les écarts-types des valeurs de C et \hat{C} , $cov_{c\hat{c}}$ la covariance des valeurs de C et \hat{C} , puis c_1 , c_2 et c_3 , des constantes qui dépendent de la profondeur des images. Ces valeurs sont calculées puis comparées au niveau du noyau de convolution.

A noter que l'ensemble des indicateurs a été calculé après avoir concaténé les canaux a et b de la sortie à la bande pseudo-panchromatique P , puis réalisé la conversion de l'espace Lab vers RVB. Cela permet de simplifier l'interprétation des résultats, dans la mesure où les images Lab possèdent des pixels dont les valeurs sont négatives pour a et b .

Il est cependant important de rappeler que ces indicateurs sont purement quantitatifs et ne permettent pas, *a priori*, de rendre compte de la plausibilité d'une colorisation. Compte-tenu du fait qu'aucune des métriques ici sélectionnées ne soit réellement adaptée pour évaluer les résultats, nous proposons également de mener une évaluation plus sémantique.

2.3 Mise en œuvre de classifications historiques sur les produits colorisés

L'évaluation sémantique consiste à utiliser les produits colorisés pour réaliser une classification de l'occupation du sol. Si le fait d'incorporer une information sur la chrominance aux

Chapitre 2. Données et méthode

clichés historiques est effectivement pertinent, le classifieur devrait donc fournir de meilleurs résultats qu'avec le cliché panchromatique seulement.

Dans le cadre de cette analyse, l'évaluation porte sur la même emprise que celle présentée dans la Partie 2.1.2 (Figure 4). L'occupation du sol a été digitalisée à partir de l'orthophotographie panchromatique de l'année 1978, en suivant les spécifications générales du produit fourni par le LIVE, datant de 2012 pour le territoire de l'ancienne CUS. Les postes identifiés et les unités minimales de cartographie (UMC) sont présentés dans la Table 9. A noter également qu'aucune autre référence géographique n'a pu être utilisée, BDTopo de l'IGN par exemple, du fait de la géométrie particulière des objets, générée par l'angle de prise de vue. Cette couche a ensuite été rastérisée à la même résolution que celle du cliché historique, c'est-à-dire 30cm.

Poste	Numéro	UMC (m^2)
■ Surface agricole	2	5
■ Végétation arborée	3	0,4
■ Végétation herbacée	7	0,4
■ Bâti	4	9,5
■ Route	6	175
■ Autres surfaces	1	1,5

TABLE 9 – Spécifications générales de l'occupation du sol digitalisée pour l'année 1978.

Afin de générer une photographie en couleurs naturelles, le modèle entraîné à partir de la photothèque globale présentée dans la Partie 2.2.1.1 a été utilisé. Cela permet d'obtenir le couple $\{\hat{C}, P\}$, avec P l'orthophotographie panchromatique de 1978, et \hat{C} la colorisation correspondante. L'image \hat{C} est issue de la concaténation des canaux P , a et b , puis de la conversion du produit résultant vers l'espace colorimétrique RVB.

Les méthodes de classification et d'évaluation ensuite employées, pour C et P , sont les mêmes que celles indiquées dans la Partie 2.1.2 avec, pour rappel, l'utilisation d'un classifieur Random Forest, d'un échantillonnage stratifié et le calcul d'un score F1.

3 Résultats

La présentation des résultats suit globalement l’agencement de la partie dédiée aux données et méthodes, avec quelques modifications qu’il semble important de préciser. La première section présente l’intérêt de l’information colorimétrique pour la mise en œuvre d’applications géographiques, et permet ainsi de justifier de l’importance de ce travail. S’organisent ensuite deux sections structurellement proches mais qu’il est important de séparer. La première est vouée au modèle entraîné à partir de la base locale, puis utilisé avant tout afin de sélectionner un algorithme parmi les BEGAN, cGAN et DRAGAN proposés plus tôt. La deuxième se focalise quant à elle sur le modèle global, entraîné à partir de l’algorithme choisi précédemment. En vue de compléter les métriques utilisées pour évaluer la qualité des développements, une dernière section est finalement réservée à l’évaluation sémantique des colorisations. Elle consiste ainsi à analyser les classifications obtenues à partir de produits historiques colorisés.

3.1 Évaluation de l’apport de la couleur pour la mise en œuvre de classifications

La mise en œuvre de plusieurs scénarios de classification nous a permis de souligner l’importance de la texture pour classifier l’occupation du sol à partir des extraits d’images Pléiades.

La Table 10 montre les scores F1 moyens, ainsi que ceux obtenus pour chaque classe sur le canal panchromatique tout d’abord. Il rend compte d’une amélioration des performances du classifieur Random Forest lorsque le nombre d’attributs augmente. Le meilleur scénario correspond ainsi à S5 { Pseudo-panchromatique ; 2 canaux de LBP ; 8 canaux de texture de Haralick }, avec un score F1 moyen de 0,52. Il semblerait par ailleurs que les textures de Haralick soient suffisantes pour obtenir des résultats proches de ceux S5, avec un F1 égal à 0,51 pour S3.

D’un point de vue quantitatif, les meilleurs résultats sont obtenus pour les surfaces agricoles et minérales réfléchissantes. La route obtient systématiquement les scores F1 les plus faibles, ce qui pourrait s’expliquer avant tout par une résolution de 2m, proche de la taille de l’élément

Chapitre 3. Résultats

	Panchromatique					Couleur
	S1	S2	S3	S4	S5	S5*
■ Surface agricole	0,14	0,58	0,64	0,56	0,65	0,70 [+8%]
■ Végétation arborée	0,38	0,47	0,51	0,47	0,52	0,62 [+19%]
■ Végétation herbacée	0,44	0,46	0,50	0,46	0,51	0,61 [+20%]
■ Bâti	0,47	0,53	0,58	0,53	0,59	0,72 [+22%]
■ Route	0,01	0,01	0,07	0,00	0,07	0,47 [+571%]
■ Autres surfaces	0,48	0,52	0,56	0,53	0,57	0,62 [+9%]
Score F1 moyen	0,41	0,47	0,51	0,47	0,52	0,63 [+21%]

TABLE 10 – Valeurs des scores F1 obtenus pour le jeu de validation, pour différents scénarios de classification et classes d’occupation du sol sur l’image Pléiades de 2012. Les valeurs obtenues pour la scène en couleurs sont complétées, entre crochets, par un pourcentage décrivant l’amélioration du score F1 avec le passage NB → RGB.

à détecter. Qualitativement cette fois-ci, l’organisation spatiale de la zone d’étude ressort grossièrement, avec une mauvaise séparabilité entre le bâti, les routes et autres surfaces (Figure 11).

L’apport de la couleur, renseigné dans colonne S5* de la Table 10, montre une amélioration substantielle des résultats de classification. Le score F1 moyen passe ainsi à 0,63, soit une augmentation de 21%. Différentes classes d’occupation du sol sont également mieux classées, notamment les surfaces végétalisées, le bâti et surtout la route, avec une augmentation du F1 de 571% pour celle-ci. D’un point de vue qualitatif, l’organisation spatiale de la scène est retrouvée, avec une meilleure séparabilité entre les différentes classes (Figure 11). Dans ce cas de figure, l’information colorimétrique a alors permis d’améliorer les résultats de classification.

Les résultats obtenus à partir de l’orthophotographie de 2013 vont également dans le sens de ces observations. A nouveau, le scénario S5 fournit les meilleurs scores F1, avec des valeurs moyennes respectives de 0,57 et 0,53 pour les produits pseudo-panchromatiques non dégradé et dégradé (Table 11). D’un point de vue quantitatif, les surfaces agricoles et herbacées possèdent les scores les moins bons. Les meilleures prédictions sont quant à elles obtenues pour les surfaces artificialisées et arborées. Qualitativement, nous pouvons cependant observer une difficulté à classifier les zones urbaines, avec une mauvaise séparabilité entre le bâti, les surfaces herbacées et autres surfaces. L’apport de la couleur permet d’obtenir des scores F1 moyens respectifs de 0,67 et 0,57 pour les produits non dégradé et dégradé. Il y a donc une amélioration de la classification sur le plan quantitatif, qui se traduit spatialement par des résultats plus fins que ceux obtenus grâce à la bande pseudo-panchromatique (Figure 11). La dégradation des produits, supposée simuler les caractéristiques radiométriques et spatiales des orthophotographies anciennes, entraîne quant à elle une baisse des performances du classifieur.

Ces points confirment ainsi l’hypothèse de départ, selon laquelle les applications géographiques classiques, telles les classifications de l’occupation du sol, profitent de l’information colorimétrique. Une telle observation justifie donc de proposer des méthodes de colorisation des orthophotographies historiques, afin de disposer d’une donnée plus simple à manipuler

3.1. Évaluation de l'apport de la couleur pour la mise en œuvre de classifications

	Pseudo-panchromatique					Couleur
	S1	S2	S3	S4	S5	S5*
■ Surface agricole	0,03 (0,00)	0,23 (0,04)	0,36 (0,22)	0,25 (0,07)	0,37 (0,27)	0,53 (0,26)
■ Végétation arborée	0,49 (0,48)	0,60 (0,56)	0,65 (0,60)	0,62 (0,59)	0,65 (0,63)	0,67 (0,63)
■ Végétation herbacée	0,38 (0,37)	0,42 (0,37)	0,47 (0,42)	0,43 (0,40)	0,47 (0,46)	0,56 (0,57)
■ Bâti	0,31 (0,26)	0,49 (0,39)	0,56 (0,49)	0,49 (0,40)	0,56 (0,50)	0,80 (0,60)
■ Route	0,31 (0,19)	0,47 (0,40)	0,53 (0,46)	0,48 (0,42)	0,55 (0,51)	0,65 (0,53)
■ Autres surfaces	0,47 (0,46)	0,56 (0,54)	0,59 (0,56)	0,57 (0,55)	0,60 (0,58)	0,65 (0,59)
Score F1 moyen	0,39 (0,37)	0,51 (0,45)	0,56 (0,51)	0,52 (0,47)	0,57 (0,53)	0,67 (0,57)

TABLE 11 – Valeurs des scores F1 obtenus pour le jeu de validation, pour différents scénarios de classification et classes d'occupation du sol. Chaque cellule contient le score F1 pour le produit non dégradé, suivi entre parenthèses de celui obtenu pour le produit dégradé.

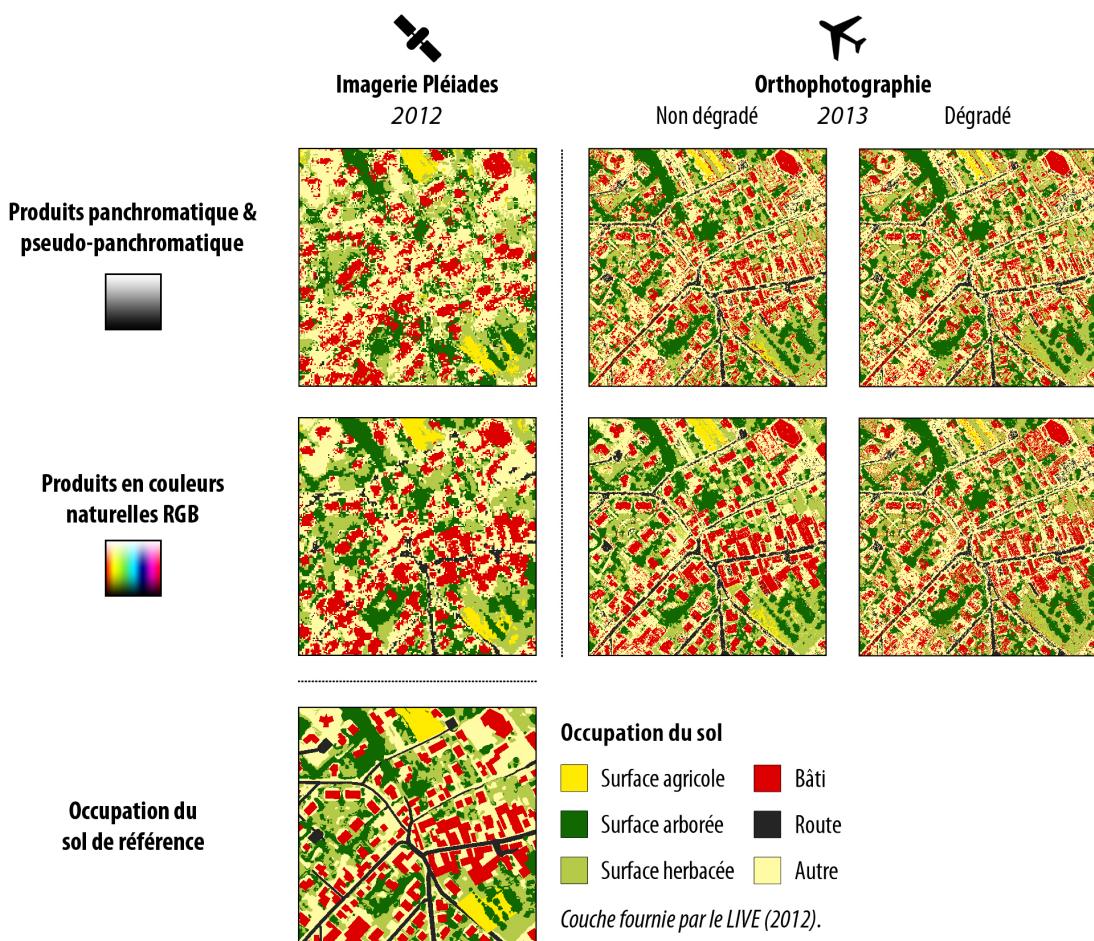


FIGURE 11 – Résultats de classification obtenus à l'aide du scénario n°5, à partir des images et photographies en couleurs et monochromatiques.

pour la communauté scientifique et autres utilisateurs-trices de ces produits. Enfin, compte-tenu des résultats obtenus suite aux dégradations, il semble raisonnable de supposer que nous pouvons nous attendre à la même qualité pour les clichés historiques colorisés.

3.2 Modèle local pour la colorisation des orthophotographies anciennes

La section portant sur le modèle local s'organise en deux sous-parties. La première consiste à sélectionner un modèle parmi les BEGAN, cGAN et DRAGAN. Une fois retenu, l'algorithme sert à tester les capacités de multispectralisation de la méthode dans une deuxième sous-section, puis à entraîner un modèle global, comme présenté dans la Partie 3.3.

3.2.1 Résultats de colorisation pour les modèles locaux dans le domaine du visible

Lors de l'entraînement des modèles locaux, trois métriques ont été calculées afin de fournir une description quantitative de la qualité des résultats de colorisation, par rapport à un jeu de validation. Ces scores ont été résumés à l'aide de la formule décrite par l'équation 13. A chaque itération, la médiane de chacune des trois métriques est calculée sur l'ensemble $\{C, \hat{C}\}$, avec C les références et \hat{C} les produits colorisés.

$$score_{med}^{it} = med(scores^{it}) \quad (13)$$

Cela permet d'obtenir des descripteurs globaux de la performances des trois modèles testés, décrits dans la Table 12. Le DRAGAN génère ici les meilleurs résultats, sur les plans quantitatif (MSE, PSNR) et structurel (SSIM). Il est aussi intéressant de noter que le GAN est plus performant que le BEGAN, malgré les optimisations apportées au modèle. Nous pouvons supposer que cela s'explique par les architectures de G et de D, qui ne respectent pas les recommandations de Berthelot *et al.* (2017).

	BEGAN	DRAGAN	GAN
MSE	239,05	<u>226,75</u>	231,93
PSNR	24,35	<u>24,58</u>	24,48
SSIM	0,933	<u>0,935</u>	0,933

TABLE 12 – Meilleurs scores médians de MSE, PSNR et SSIM obtenus sur les produits colorisés, générés à partir des modèles locaux sur le jeu de données de validation. Les valeurs optimales, ici soulignées, décrivent la meilleure correspondance entre les produits colorisés et la base de référence. Plus le score de MSE est faible, plus les produits correspondent. Plus les scores de PSNR et de SSIM sont élevés, plus les produits correspondent.

Compte-tenu de ces résultats, le DRAGAN a été retenu pour l'apprentissage d'un modèle global sur la photothèque dans son ensemble. Il est également possible de visualiser l'évolution des erreurs de prédiction au fil de l'apprentissage (Figure 12). La qualité structurelle et quantitative des produits en sortie atteint rapidement son *optimum*, avec un plateau qui apparaît dès la 150^e itération. Une analyse graphique et numérique des résultats montre cependant que la colorisation est la meilleure autour de la 900^e passe. En parallèle, les valeurs des fonctions objectif de G et de D tendent à augmenter, tout en variant fortement d'une itération à l'autre

3.2. Modèle local pour la colorisation des orthophotographies anciennes

(Figure 13). La convergence vers un optimum est donc plus difficile à évaluer du point de vue de ces métriques, hormis la distance L1 qui se rapproche fortement de la MSE du fait de sa nature euclidienne. En effet, celle-ci diminue rapidement jusqu'à atteindre un plateau également.

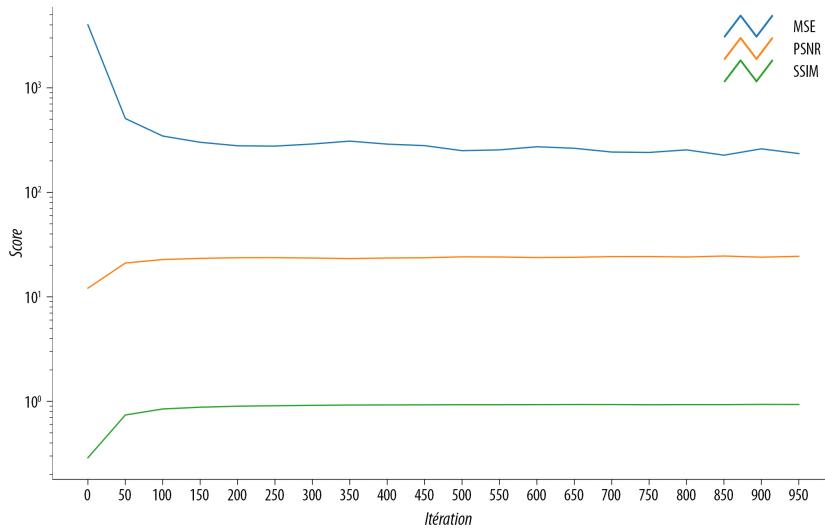


FIGURE 12 – Valeurs des métriques qualité calculées entre les produits colorisés et le jeu de validation, au cours de l'apprentissage du DRAGAN local. L'axe des y est représenté à l'aide d'une échelle logarithmique décimale.

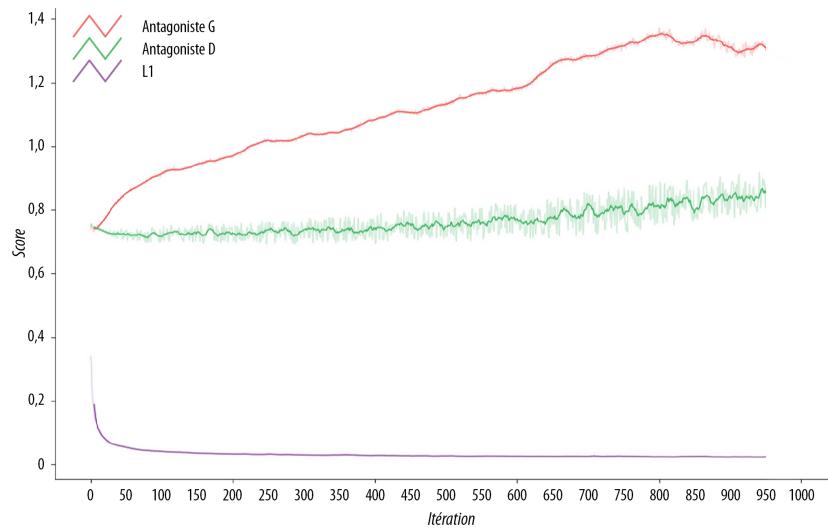


FIGURE 13 – Valeurs des fonctions objectif de G, D et L1 au cours de l'entraînement du modèle local pour l'imagerie en couleurs naturelles. Afin de mieux saisir la tendance globale de l'apprentissage, la moyenne mobile de la série a été calculée sur 11 valeurs successivement.

Compte-tenu du fait que le DRAGAN fournit les meilleures colorisations à l'itération n°900, les prochains résultats sont présentés pour celle-ci uniquement. Il est possible de comparer visuellement les produits de la base de validation à ceux générés par le modèle. La Figure

Chapitre 3. Résultats

14 montre les résultats obtenus pour quatre grandes classes d'occupation du sol, avec l'agricole, l'arboré, l'eau et l'urbain. D'un point de vue qualitatif, les colorisations donnent des résultats généralement peu plausibles, avec des associations objet — chrominance qui ne correspondent que partiellement aux références, sauf pour l'agricole. En effet, les arbres et étendues d'eau possèdent une teinte orangée, tandis que la couleur des toits des bâtiments manque de vibrance et de saturation.

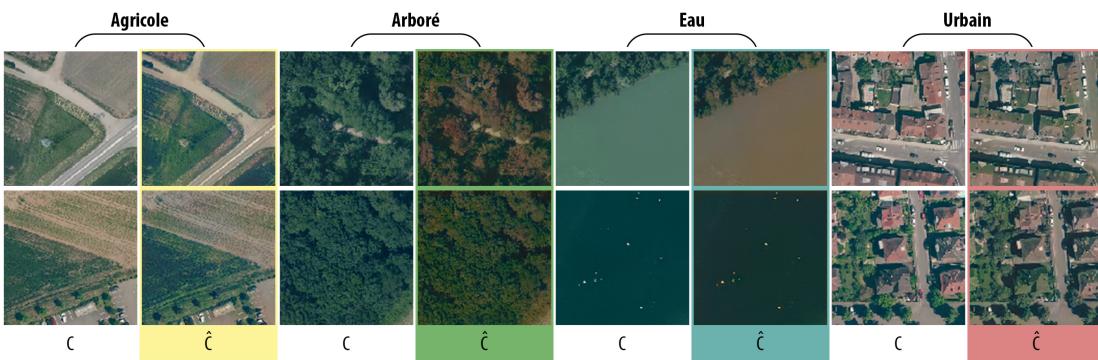


FIGURE 14 – Exemples de résultats de colorisation sélectionnés aléatoirement dans le jeu de validation, pour le modèle DRAGAN local à l'itération n°900. Les images C et \hat{C} représentent respectivement la référence en couleurs réelles, et le produit issu d'une colorisation à partir d'un cliché pseudo-panchromatique. Des extraits capturés sur différentes classes d'occupation du sol ont été montrés pour témoigner des capacités de colorisation sur plusieurs types de surfaces.

La distribution des valeurs de chrominance générées n'est donc *a priori* pas fidèle à la réalité, point qui pourrait s'expliquer par le fait que la validation présente des sémantiques trop différentes de celles du jeu d'entraînement local. La visualisation des distributions bivariées des canaux a et b pour chaque classe (Figure 15) va également dans le sens de cette observation. Les fonctions de densité de probabilité représentées dans les marges des graphiques de la Figure 15 montrent que la forme des distributions est globalement respectée, notamment l'amplitude des valeurs et les *extrema*. En revanche, les modes ne sont pas tous modélisés par le générateur, malgré l'utilisation d'un DRAGAN qui est censé répondre à ce type de problème. A nouveau, le manque de sémantiques similaires entre la validation et les produits colorisés peut expliquer cette difficulté.

Il est également important de noter la forme des distributions bivariées générées, systématiquement convexe et d'un seul tenant dans l'espace des données. Dans le cas des produits de validation pour l'arboré et l'eau en revanche, les graphiques montrent plusieurs sous-populations, séparées par des zones à très faibles densités de probabilité. De façon générale également, les régions de faible densité de probabilité sur la validation correspondent à celle de forte densité sur les images générées, et vice-versa. Le modèle parvient seulement à reproduire des distributions fidèles sur l'urbain, alors que c'est pourtant l'objet qui devrait poser le plus de difficultés, compte-tenu de la diversité des scénarios de colorisation possibles dans cet espace. En effet, coloriser des toitures peut être posé comme un problème multi-modal, dans la mesure où elles peuvent revêtir différentes couleurs (rouge, bleu, noir, gris, ...). Pour le

3.2. Modèle local pour la colorisation des orthophotographies anciennes

modèle local, sur l'eau et l'arboré spécifiquement, ce manque de cohérence peut s'expliquer par une base d'entraînement pas suffisamment riche, avec l'absence de surfaces aquatiques à proprement parler par exemple.

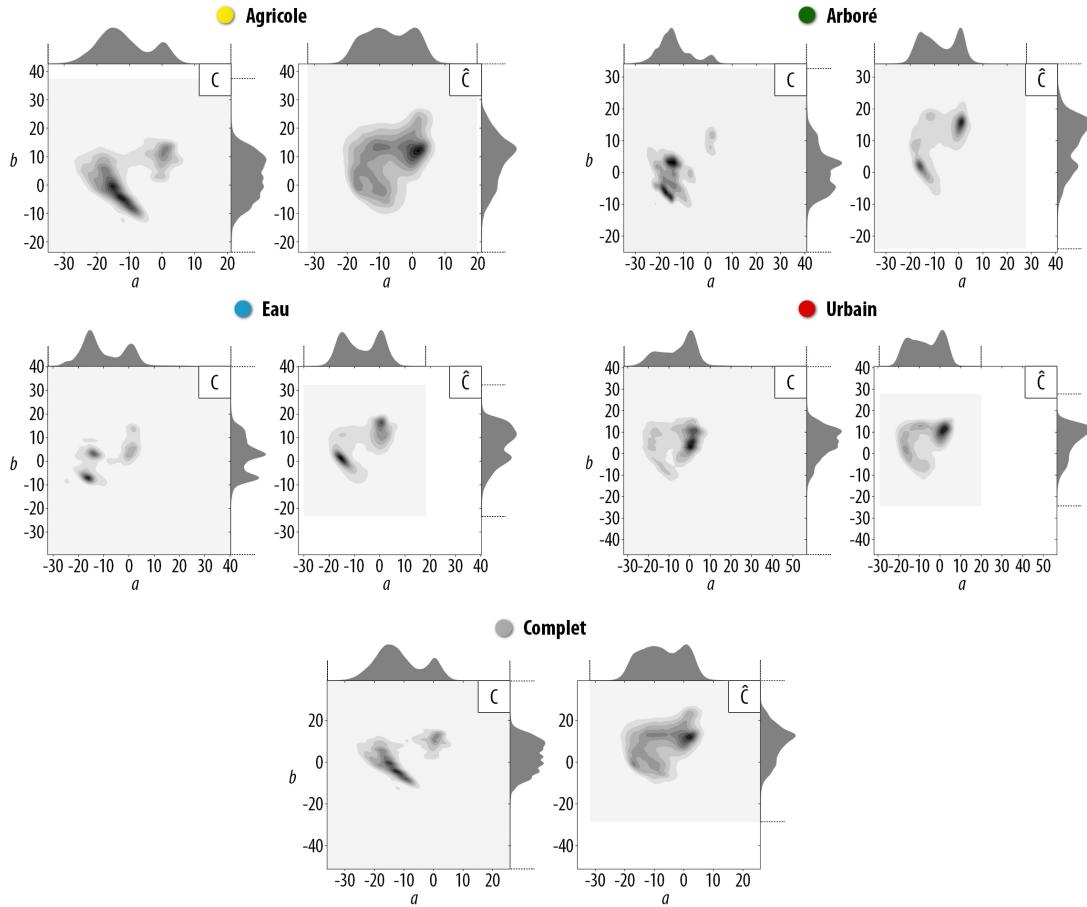


FIGURE 15 – Comparaison des distributions bivariées de a et b , calculées sur les produits colorisés et le jeu de validation, pour le modèle DRAGAN local à l'itération n°900. A noter que ces représentations ont été obtenues par échantillonnage aléatoire des valeurs des pixels. Cela explique pourquoi les bornes des axes du jeu complet sont différentes de celles des sous-ensembles spécifiques à chacune des classes d'occupation du sol.

Malgré des résultats qualitativement peu satisfaisants, les scores MSE, PSNR et SSIM révèlent que le modèle génère des colorisations quantitativement et structurellement correctes (Figure 16). En effet, les SSIM calculés pour les différentes classes d'occupation du sol sont globalement compris entre 0,85 et 0,95, signifiant que les produits colorisés et la validation sont similaires. A nouveau, l'urbain se démarque des autres postes, puisqu'il possède la MSE la plus faible et les meilleurs scores de PSNR et SSIM. En matière de performances, la qualité des colorisations décroît ensuite, dans l'ordre, avec l'eau, l'agricole puis l'arboré. Il est intéressant de noter que l'eau, pour laquelle il n'existe aucune référence dans la base d'entraînement locale, obtient de meilleurs résultats que ces autres postes. D'un point de vue qualitatif également, les colorisations générées pour l'agricole sont tout à fait acceptables, et plus plausibles

Chapitre 3. Résultats

à l'œil nu que celles obtenues sur l'urbain, malgré une qualité apparemment inférieure. Il y a donc une irrégularité entre l'analyse visuelle des résultats et les indicateurs calculés.

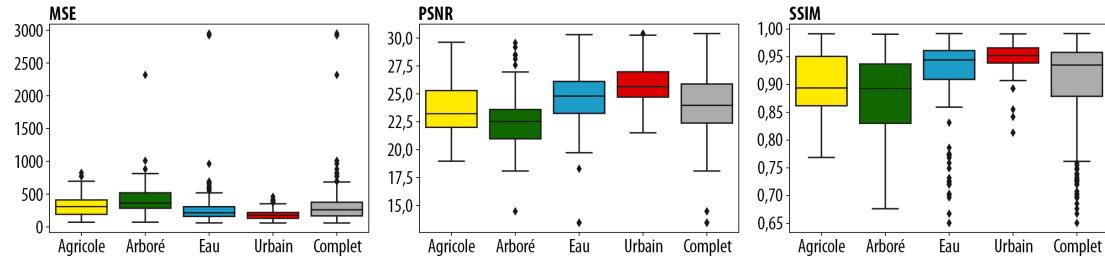


FIGURE 16 – Métriques calculées sur le jeu de validation pour les colorisations générées à l'aide du modèle DRAGAN local, à l'itération n°900. Ces métriques sont présentées pour plusieurs classes d'occupation du sol, afin de témoigner des capacités de colorisation sur différents types de surfaces.

D'une manière générale, les produits obtenus avec le modèle local sont corrects lorsque la colorisation est menée sur un extrait d'orthophotographie décrivant la commune de Niederhausbergen (Figure 17). Nous pouvons cependant noter la présence d'artefacts, qui s'explique par un manque d'exemples et un temps d'apprentissage limité avant tout. Parmi ces erreurs se trouvent des colorisations peu plausibles sur certaines toitures et ombres portées, un manque de continuité des champs colorimétriques dans les zones agricoles, des bavures, ou encore l'absence de chrominance pour certaines surfaces. D'autres artefacts liés aux étapes de mosaïquage pré-colorisation et post-colorisation sont également visibles.

Outre ces erreurs de prédiction, l'analyse réalisée à partir de la validation montre également que le générateur parvient difficilement à généraliser l'apprentissage à des sémantiques qu'il ne connaît pas ou peu. Cela se traduit qualitativement par des colorisations parfois peu plausibles, mais malgré tout correctes vis-à-vis de la structure de l'image. L'évaluation quantitative ne va cependant pas forcément dans le sens d'une analyse visuelle. Il est finalement question de savoir si le modèle global parvient à reproduire plus fidèlement les distributions réelles des valeurs de chrominance, dans la mesure où celui-ci dispose d'une photothèque à vocation plus généraliste.

3.2.2 Apprentissage par transfert et capacités de multispectralisation du modèle local

Lors de la création du jeu d'entraînement local, nous avions également collecté des extraits d'orthophotographie infrarouge couleur I , permettant de constituer des paires $\{I, P\}$ sur la commune de Niederhausbergen.

Cette base modeste a été utilisée pour tester les capacités de multispectralisation du modèle local, c'est-à-dire sa faculté à générer une information spectrale autre que celle issue du domaine du visible. Pour cela, nous avons tout d'abord initialisé les générateur et discriminateur locaux obtenus pour le visible à l'itération n°900, puisqu'il renvoyait les meilleurs résultats comme démontré dans la Partie 3.2.1. Cela suppose donc que le générateur ait appris des

3.2. Modèle local pour la colorisation des orthophotographies anciennes

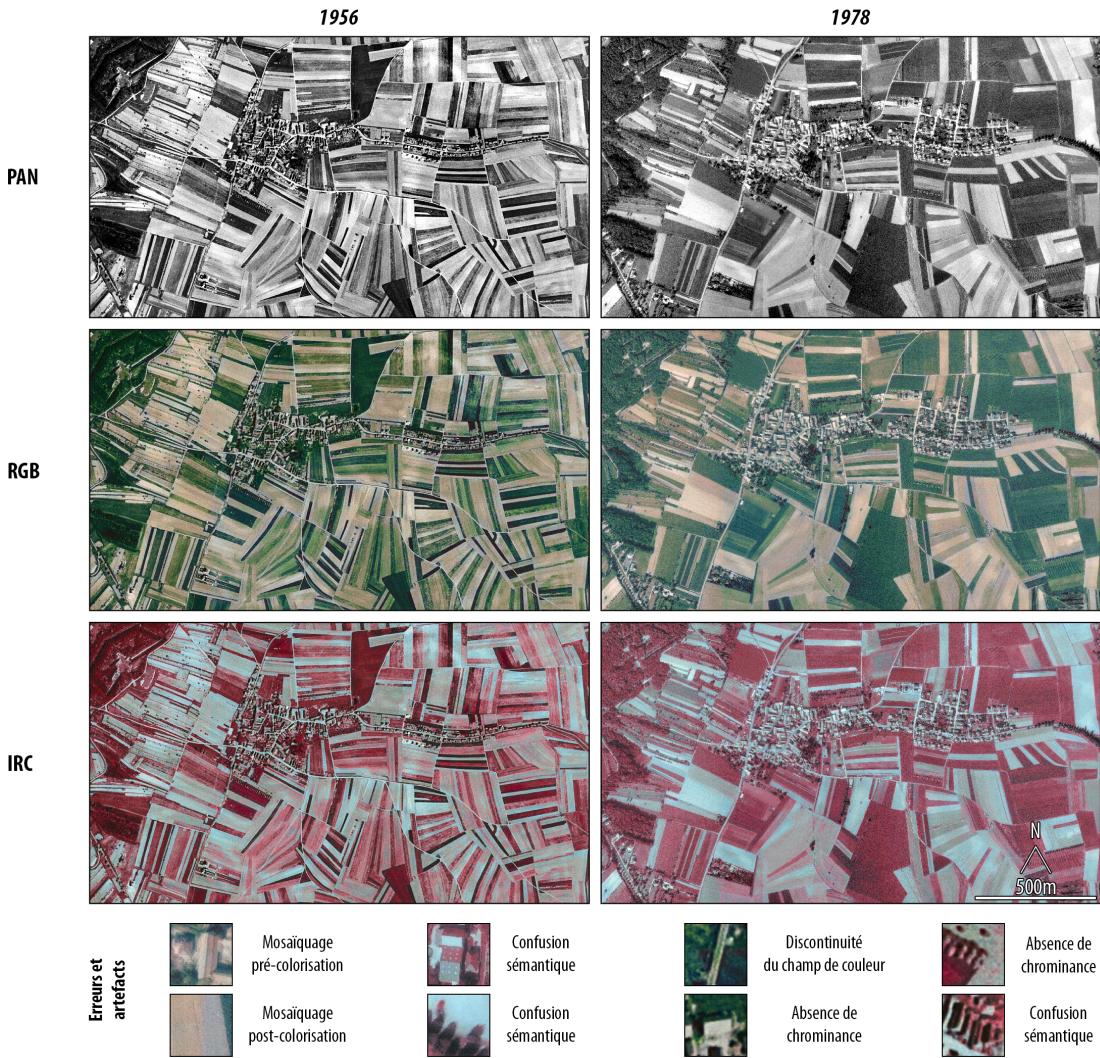


FIGURE 17 – Résultats de colorisation sur Niederhausbergen avec le modèle local. Les produits obtenus pour des compositions en couleurs naturelles et infrarouge couleur sont présentés ensemble à titre comparatif. Le canal proche-infrarouge, obtenu par *transfer learning*, est présenté dans la sous-partie suivante.

attributs suffisamment généraux pour être réutilisés dans la prédiction d'un canal proche-infrarouge. C'est le principe même de l'apprentissage par transfert, qui consiste à réutiliser comme base les paramètres θ appris par un modèle, lorsqu'il est jugé suffisamment robuste pour la tâche à accomplir.

Le code utilisé pour la colorisation n'a pas été modifié, le modèle restant donc un DRAGAN, avec les mêmes méthodes d'augmentation de données et optimisations que celles renseignées dans la Partie 2.2.2. Seul le taux d'apprentissage a été ajusté, avec une valeur de $lr = 2 \times 10^{-5}$, afin de ne pas trop modifier les paramètres θ , et donc les attributs initialement appris par le modèle. En effet, l'objectif est seulement d'apprendre une nouvelle bande, tout en conservant

Chapitre 3. Résultats

les sémantiques spatiales initiales.

D'une façon générale, l'apprentissage est plus instable que lorsque le modèle local a été entraîné pour la colorisation dans le domaine du visible (Figure 18). Cela s'explique certainement par une mise-à-jour progressive des paramètres θ , et donc des attributs, qui consignent l'ensemble des sémantiques présentes dans le jeu d'entraînement. En effet, la Figure 18 indique que les valeurs des fonctions objectif de G et D diminuent progressivement jusqu'à l'itération n°500 environ, puis évoluent ensuite rapidement et fortement. L'entraînement a ici été arrêté avant la millième étape de l'apprentissage pour des raisons financières, mais il aurait été intéressant de le poursuivre jusqu'à stabilisation du modèle.

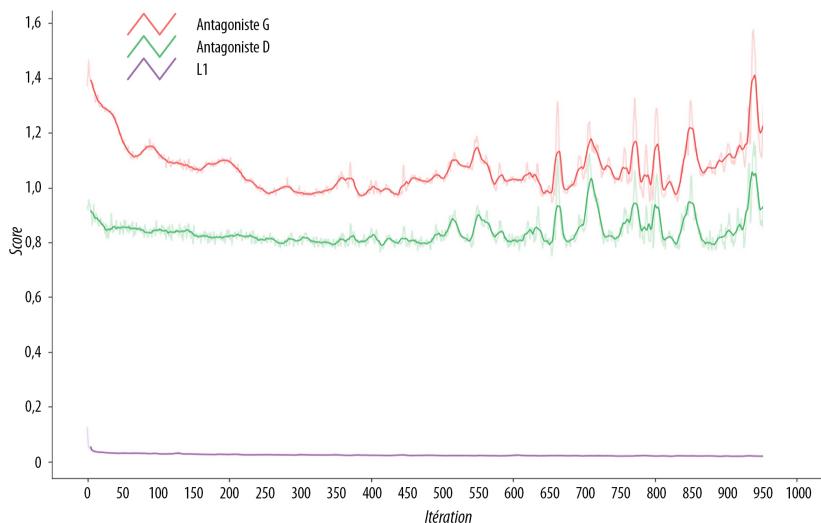


FIGURE 18 – Valeurs des fonctions objectif de G, D et L1 au cours de l'entraînement du modèle local pour l'imagerie en infrarouge couleur. Afin de mieux saisir la tendance globale de l'apprentissage, la moyenne mobile de la série a été calculée sur 11 valeurs.

D'un point de vue qualitatif, les résultats obtenus sont tout-à-fait plausibles (Figure 17), malgré la présence des mêmes artefacts et erreurs que ceux notés dans la Partie 3.2.1. Les produits sont cependant d'une qualité suffisante pour générer des NDVI historiques pertinents, pour les années 1956 et 1978 par exemple (Figure 19). Quelques incohérences sont visibles dans le centre-ville de Niederhausbergen en 1956, avec des bâtiments qui possèdent un indice de végétation particulièrement élevé, de l'ordre de 0,3. Cela s'explique en partie par la résolution spatiale de la photographie ancienne, théoriquement de 50cm, mais trop dégradée pour que le modèle puisse reconnaître le jeu de sémantiques propres aux bâtiments. Il y reconnaît ainsi une végétation basse, probablement herbacée, ce qui s'explique localement par la texture lisse des toitures par exemple.

Ainsi, il a été possible de réutiliser le modèle local développé pour la colorisation dans le domaine du visible, et de le modifier légèrement pour qu'il prédise également un canal proche-infrarouge. Cela témoigne donc des capacités d'apprentissage par transfert et de multispectralisation de la méthode ici développée.

3.3. Modèle global pour la colorisation des orthophotographies anciennes

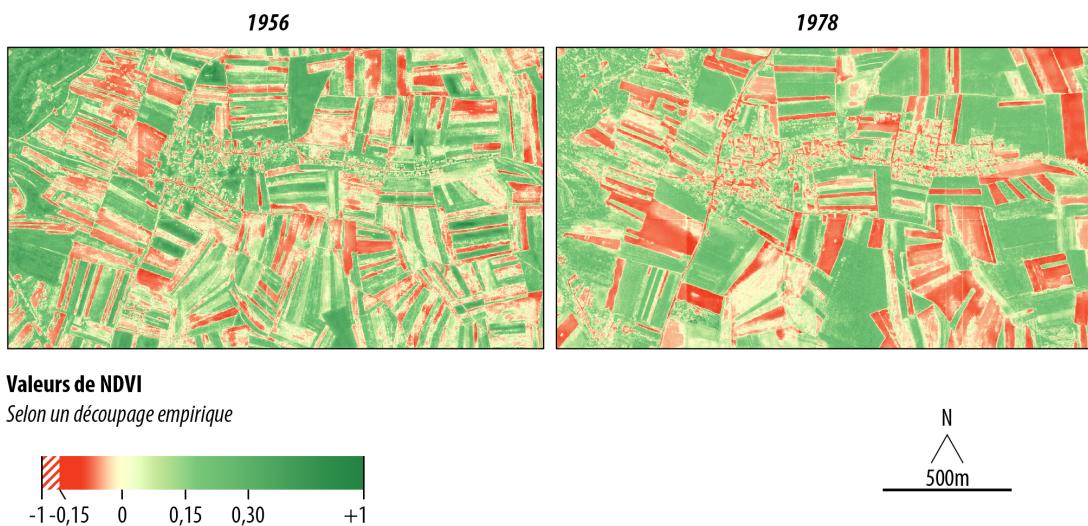


FIGURE 19 – Visualisation des NDVI historiques obtenus grâce au modèle local pour les années 1956 et 1978.

3.3 Modèle global pour la colorisation des orthophotographies anciennes

Maintenant que le modèle local a été testé, à la fois pour la prédiction des canaux du visible et du proche infrarouge, il est possible de se pencher sur l'apprentissage réalisé à partir de la photothèque globale. Pour rappel, le DRAGAN fournissait *a priori* les meilleurs résultats de colorisation. Le même algorithme a donc été utilisé pour la mise au point du modèle global, à vocation plus généraliste, pour générer cette fois-ci des clichés en couleurs naturelles sur l'ensemble de l'EMS à partir d'orthophotographies anciennes.

Le DRAGAN conditionnel a ainsi été entraîné à partir de la photothèque au complet, sur un peu plus de 950 itérations. Les résultats présentés dans les sous-sections suivantes se penchent sur une évaluation du modèle, la visualisation d'exemples de colorisation ainsi que sur les attributs appris par le générateur pour modéliser la fonction $f(p) = c$.

3.3.1 Évaluation du modèle global de colorisation

Contrairement aux modèles locaux développés précédemment, les réseaux G et D entraînés à partir de la photothèque globale parviennent à se stabiliser rapidement, à partir de la 400^e étape de l'apprentissage (Figure 20). A noter cependant que les mille itérations réalisées ne sont pas suffisantes pour dire s'il y a réellement convergence vers une solution adaptée à ce problème de colorisation. Les résultats présentés dans les sous-sections et paragraphes suivants montrent en effet que le modèle bénéficierait certainement d'un apprentissage plus long.

Chapitre 3. Résultats

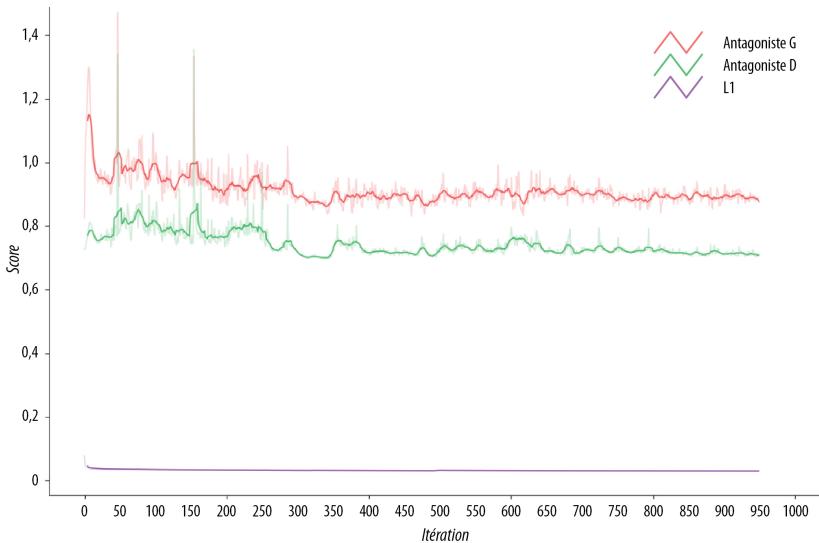


FIGURE 20 – Valeurs des fonctions objectif de G, D et L1 au cours de l’entraînement du modèle global. La moyenne mobile de la série a été calculée sur 11 valeurs successivement.

D’un point de vue quantitatif et structurel, il semblerait que la validation renvoie les meilleurs résultats de colorisation pour les itérations 250 (PSNR, MSE) et 350 (SSIM), points qui se situent donc en amont de la stabilisation des fonctions objectif antagonistes de G et D (Figure 21). Ces indicateurs se stabilisent rapidement, dès la 50^e itération, sauf pour la MSE qui oscille légèrement tout au long de l’entraînement. Il est intéressant de noter que ces métriques obtenaient leurs valeurs optimales bien plus tardivement pour le modèle local. Nous pouvons supposer ainsi que la photothèque globale, du fait de son nombre d’échantillons plus important, possède un corpus de sémantiques plus facilement et rapidement généralisable au problème étudié. Il peut alors être raisonnable de soupçonner que cela se traduira visuellement par des résultats de colorisation plus plausibles qu’avec le modèle local, au moins sur la validation.

La visualisation de quelques extraits d’images utilisés pour évaluer le modèle va effectivement dans le sens de cette hypothèse (Figure 22). Les postes décrivant l’urbain et l’agricole ne montrent pas de changement particulier par rapport au modèle local. La végétation arborée et les surfaces en eau révèlent cependant une amélioration substantielle des performances du générateur. En effet, la base locale donnait lieu à des résultats de colorisation fortement orangés pour ces deux classes, ce que nous expliquions précédemment par l’absence de sémantiques adaptées. C’est toujours le cas pour certaines portions de ces clichés, mais le phénomène est largement atténué. Le développement d’une photothèque à vocation plus généraliste permet donc de palier à ce manque, mais ne répond pas nécessairement à tous les problèmes. En effet, certains résultats en sortie revêtent toujours des couleurs peu ou moyennement plausibles, particulièrement dans l’urbain avec des toitures vertes ou peu saturées pour les itérations 250 et 350 (Figure 22). A des fins de comparaison, la Figure 22 a été complétée par les résultats obtenus à la 950^e étape, afin de voir quels sont les effets du temps

3.3. Modèle global pour la colorisation des orthophotographies anciennes

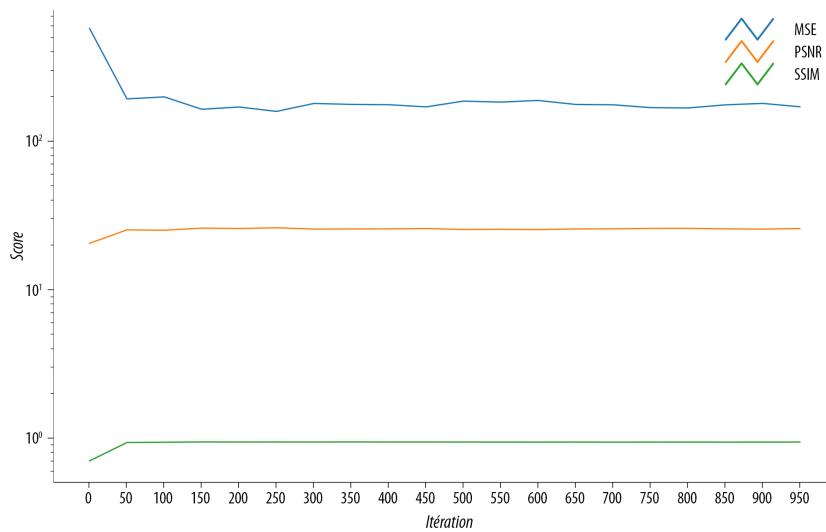


FIGURE 21 – Valeurs des métriques qualité calculées pour le modèle global entre les produits colorisés et le jeu de validation, au cours de l’apprentissage. L’axe des y est représenté à l’aide d’une échelle logarithmique décimale.

d’entraînement. Bien que ces clichés ne révèlent pas d’évolution majeure, une inspection visuelle de l’ensemble du jeu de validation souligne une amélioration globale des résultats, pour les surfaces artificialisées plus particulièrement. En effet, la couleur orangée des tuiles du centre-ville et des centres-bourgs n’apparaît que tardivement dans l’apprentissage. Cela s’explique, tout ou partie, par les propriétés des GANs, pour lesquels il est généralement difficile de capturer l’ensemble des modes d’une distribution. Certaines des toitures vertes évoquées précédemment sont aussi désormais brunes, ce qui témoigne là de l’apport supposé d’un temps d’apprentissage prolongé, des méthodes d’augmentation de données et d’une diminution progressive des confusions du modèle.

Dans la suite de ce travail, nous avons pris la décision d’aller à l’encontre des renseignements apportés par la MSE, le PSNR et le SSIM. En effet, l’étape de référence désormais utilisée pour présenter le reste des résultats correspond à la 950^e itération du modèle. Bien qu’ayant obtenu des scores objectivement moins bons, elle donne des colorisations plus naturelles et plausibles, au moins subjectivement. Comme énoncé dans la Partie 2.2.3, ces métriques ne sont pas nécessairement adaptées au problème de colorisation. En effet, il semblerait qu’elles favorisent ici les sorties peu saturées en minimisant l’erreur globale. À l’inverse, plus les couleurs sont vibrantes, plus l’erreur risque d’augmenter, simplement car l’objectif est d’obtenir une colorisation qui sera localement plausible et non pas réelle. C’est tout là la difficulté de travailler sur un problème de nature multi-modale. Il est donc important de considérer ce choix dans la suite de l’interprétation des résultats.

Compte-tenu de la Figure 22 et malgré les améliorations par rapport au modèle local, l’arboré, l’eau, puis l’urbain dans une certaine mesure, fournissent les résultats les moins bons en matière de perception visuelle. Bien que nous ayons remis en doute l’apport de la MSE, du

Chapitre 3. Résultats

PSNR et du SSIM, il reste intéressant de vérifier si ces métriques parviennent à mettre cela en avant ou non. La Figure 23 montre une erreur quadratique moyenne qui va effectivement dans ce sens, avec des scores plus importants pour les postes décrivant l’arboré et l’eau. Le PSNR et le SSIM renvoient quant à eux des résultats plus mitigés et parfois même opposés. En effet, l’urbain obtient par exemple la valeur de PSNR la plus faible, mais aussi le meilleur SSIM. La structure des images appartenant à cette classe diffère donc largement de l’information apportée par leur contenu numérique brut. Un constat similaire ressort pour les surfaces agricoles, qui possèdent les meilleurs résultats sur le plan quantitatif (MSE et PSNR), mais renvoient des colorisations de moins bonne qualité que l’eau ou l’urbain dans une dimension structurelle cette fois-ci.

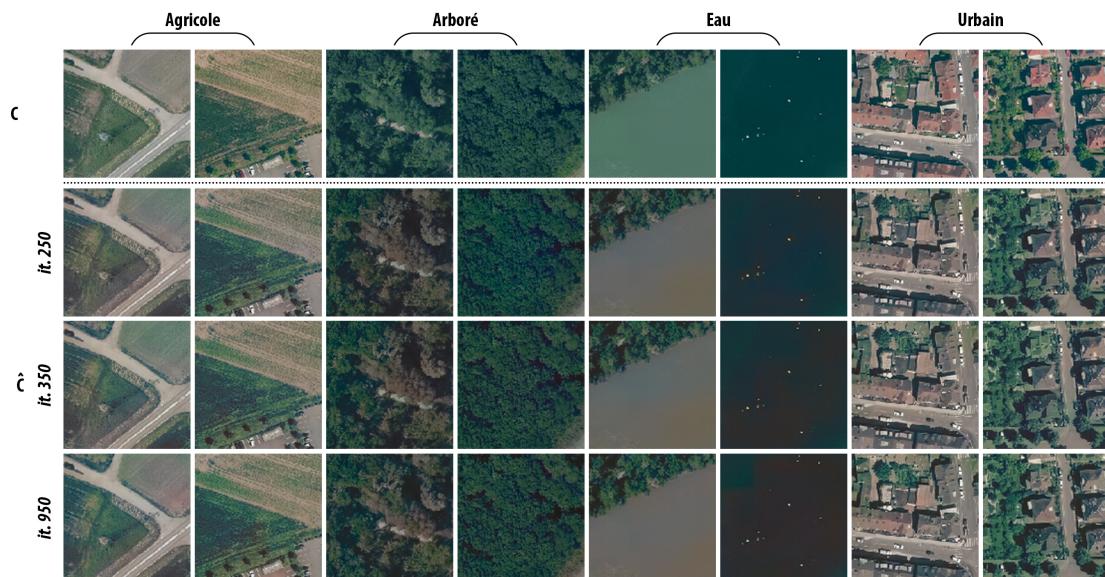


FIGURE 22 – Exemples de résultats de colorisation pour le jeu de données de validation, obtenus à partir du modèle global, aux itérations n°250, 350 et 950. Les images C et \hat{C} représentent respectivement la référence en couleurs réelles, et le produit issu d’une colorisation à partir d’un cliché pseudo-panchromatique. Des extraits capturés sur différentes classes d’occupation du sol ont été montrés pour témoigner des capacités de colorisation sur plusieurs types de surfaces.

Ces observations insinuent qu’il y a effectivement une incohérence entre la perception visuelle des résultats et ces indicateurs, mais aussi peut-être entre les indicateurs eux-mêmes.

Par rapport au modèle local (Figure 16), nous pouvons tout de même noter que l’utilisation d’une photothèque globale permet de réduire les erreurs quantitatives et structurelles pour l’ensemble des postes (Figure 23). Les variances inter-classes et intra-classes semblent également être inférieures pour ces indicateurs, par rapport à celles présentées sur la Figure 16. Cela peut s’expliquer par une meilleure généralisation, ou simplement par le fait que nous avions veillé à avoir une distribution équilibrée des classes lors du développement de la base d’entraînement globale.

Les distributions bivariées apportent quant à elles des informations complémentaires, mais

3.3. Modèle global pour la colorisation des orthophotographies anciennes

pas toujours compatibles avec la perception visuelle et les métriques présentées précédemment. La Figure 24 montre les valeurs prises par les canaux a et b des images en couleurs et de leurs homologues obtenus par colorisation, pour chaque classe et pour l'ensemble du jeu de validation. Toujours convexes, elles reflètent cependant des morphologies différentes de celles obtenues pour le modèle local (Figure 15), en particulier pour les postes décrivant l'arboré et l'eau. Bien que la photothèque possède un jeu de sémantiques plus complet que le jeu d'entraînement local, les distributions bivariées restent inexactes pour ces deux classes. D'une façon plus générale, les *extrema* sont à peu près corrects pour \hat{C} vis-à-vis de C , mais la position des modes n'est pas toujours respectée. Cela suppose donc que le générateur circule dans l'espace des données sans nécessairement trouver les sémantiques adaptées à un poste en particulier. A nouveau, les zones de faibles et fortes densités de probabilité sont souvent permutées, sauf pour l'urbain sur lequel le modèle parvient à reproduire une distribution proche de l'originale, en amplitude et en mode.

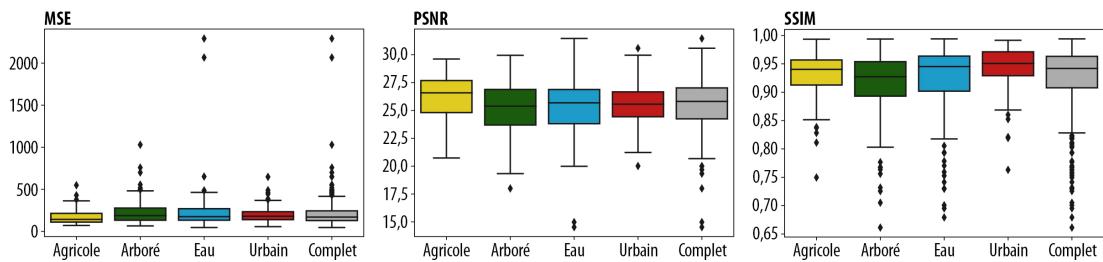


FIGURE 23 – Métriques calculées sur le jeu de validation pour les colorisations générées à l'aide du modèle DRAGAN global, à l'itération n°950. Ces métriques sont présentées pour plusieurs classes d'occupation du sol, afin de témoigner des capacités de colorisation sur différents types de surfaces.

Comme pour le modèle local, le contour des distributions bivariées des produits colorisés \hat{C} possède une forme circulaire ou ovale, d'un seul tenant. Les clichés en couleurs C montrent quant à eux, dans le cas de l'agricole, de l'arboré et de l'eau en particulier, plusieurs sous-populations séparées par des zones de faible densité de probabilité. Le phénomène est particulièrement visible sur le jeu complet. Le générateur a donc du mal à modéliser la forme de la distribution bivariée des données d'entraînement, en particulier lorsqu'elle présente des concavités ou discontinuités, et tend alors seulement à l'approximer.

Ces indicateurs et modes de représentation, bien qu'étant difficilement compatibles avec le problème de colorisation, montrent donc que le modèle parvient à apprendre des associations objet — chrominance pertinentes, mais parfois incomplètes ou confuses. Cela s'explique par la structure de la photothèque globale qui n'est pas suffisamment extensive, par un temps d'apprentissage trop court, et finalement par des confusions entre des surfaces sémantiquement proches.

Chapitre 3. Résultats

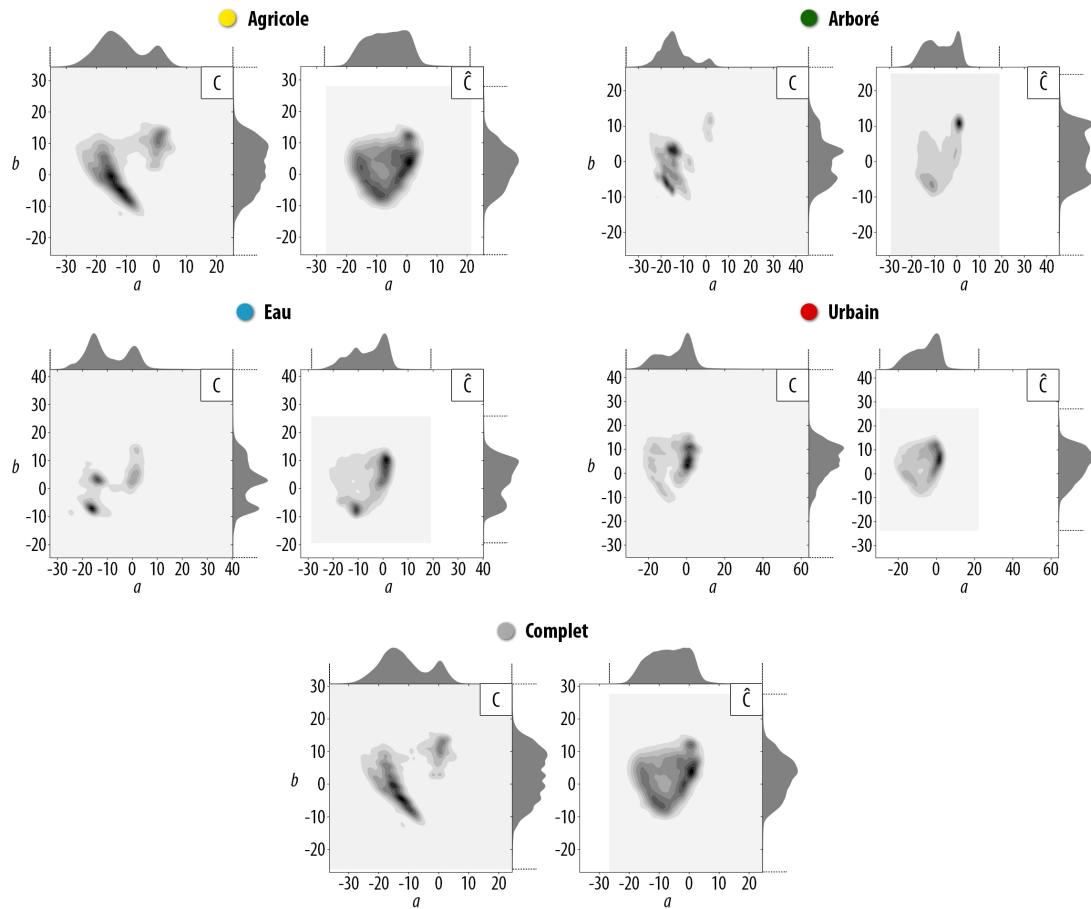


FIGURE 24 – Comparaison des distributions bivariées de a et b , calculées sur les produits colorisés et le jeu de validation, pour le modèle DRAGAN global à l’itération n°950. A noter que ces représentations ont été obtenues par échantillonnage aléatoire des valeurs des pixels. Cela explique pourquoi les bornes des axes du jeu complet sont différentes de celles des sous-ensembles spécifiques à chacune des classes d’occupation du sol.

3.3.2 Visualisation des résultats et attributs appris par le modèle global

Outre les exemples de colorisation proposés sur la Figure 22, il semblait aussi important de présenter des résultats sur des espaces plus vastes, pour différentes étapes de l’apprentissage du modèle. La Figure 25 montre des produits panchromatiques colorisés, sur des emprises de 1024×1024 pixels, dans le centre-ville et en périphérie de Strasbourg, pour les itérations 250, 350 et 950. A noter qu’un filtre de débruitage par variation totale (Chambolle, 2004) a été utilisé en amont de la colorisation, afin de retirer le grain présent sur le support argentique.

Une analyse rapide de ces clichés ne révèle pas de changement majeur dans la qualité des colorisations au fil des itérations. Bien qu’ayant décidé d’utiliser l’étape 950 comme référence pour la suite de ce travail, il semblait important de montrer qu’il y a en réalité une évolution dans la manière dont le générateur colorise les images. En effet, les étapes 250 et 350 révèlent

3.3. Modèle global pour la colorisation des orthophotographies anciennes

des couleurs ternes dans l'urbain, avec des toitures principalement grises et brunes. Cela s'améliore à l'itération 950, avec une part beaucoup plus importante des tuiles orangées, donnant donc lieu à des résultats plus plausibles. L'apprentissage des différents modes de la distribution se fait donc sur le long terme, mais s'opère parfois au détriment de la qualité des images en sortie.

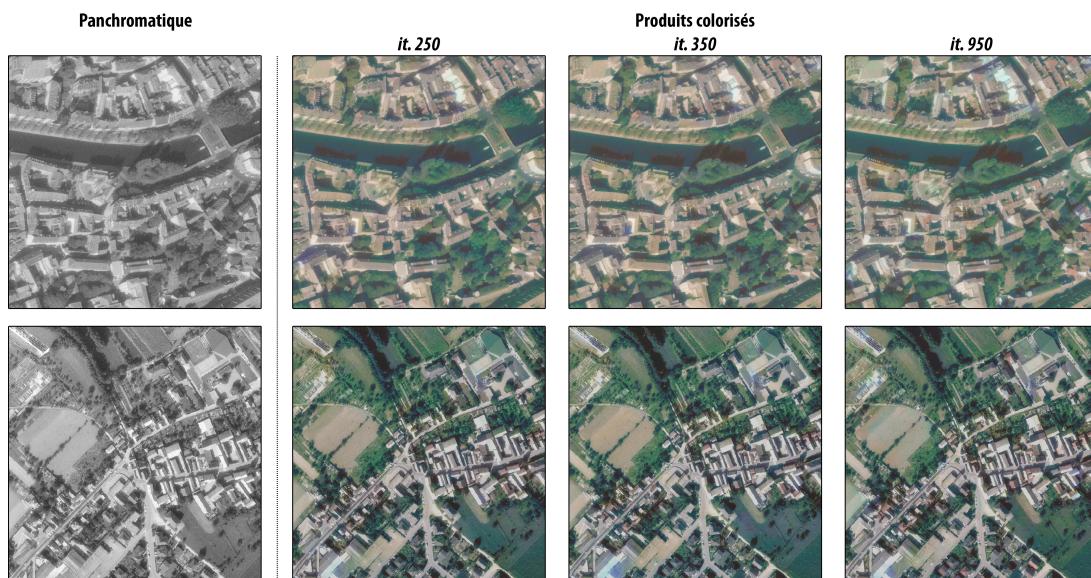


FIGURE 25 – Evolution de la qualité de la colorisation au fil des étapes d'apprentissage, pour deux emprises situées au centre-ville et en périphérie de Strasbourg. A noter que les clichés ne sont pas à l'échelle.

En effet, nous pouvons noter deux catégories d'erreurs sur les produits colorisés. La première correspond aux erreurs de confusion, et s'explique par le fait que deux surfaces pourtant différentes possèdent des sémantiques similaires. L'exemple le plus parlant correspond à certaines portions de toitures, colorisées en vert, car elles présentent des caractéristiques proches de celles de la végétation herbacée, notamment en termes de texture. Les zones affectées par ce type d'erreurs revêtent des couleurs présentes dans la base d'images, mais simplement associées à d'autres surfaces. Elles s'atténuent généralement au cours de l'apprentissage, grâce aux méthodes d'augmentation de données qui fournissent un corpus de références plus vaste pour la colorisation. La deuxième catégorie d'erreurs s'explique par l'absence totale de certaines sémantiques dans le jeu de données d'entraînement. Celles-ci persistent lors de l'apprentissage et se traduisent au départ par des patchs gris ou peu saturés. La mise à jour progressive des poids des convolutions fait cependant tendre les valeurs des pixels concernés vers des *extrema*. Ces zones revêtent alors des couleurs très saturées et absentes du jeu d'entraînement, qui correspondent aux bornes de l'espace colorimétrique considéré, comme le cyan, le bleu marine ou encore le jaune. Un exemple évident correspond à la façade située dans la partie NNE du centre-ville sur la Figure 25, qui se drapé progressivement d'un halo cyan.

Chapitre 3. Résultats

Il est aussi intéressant de se pencher sur les poids des différentes convolutions et sur les attributs extraits par le modèle. Quelques vignettes sont disponibles sur la Figure 26 à titre de visualisation seulement. Au total, ce sont 2434 attributs qui sont appris par le générateur puis utilisés pour modéliser la relation $f(p) = c$. La Figure 26 met en évidence la diminution progressive de la résolution spatiale du produit, jusqu'à la cinquième couche, qui se fait au profit d'un enrichissement des sémantiques extraites par le modèle. En effet, les attributs dégagés en premier sont particulièrement redondants, mais permettent de différencier les principales surfaces visibles sur la scène, telles le bâti, l'eau et la végétation. Viennent ensuite progressivement des abstractions plus difficiles à interpréter, issues de la combinaison des niveaux précédents. Passé le goulot, la résolution spatiale est restituée et combinée aux informations sémantiques auparavant extraites.

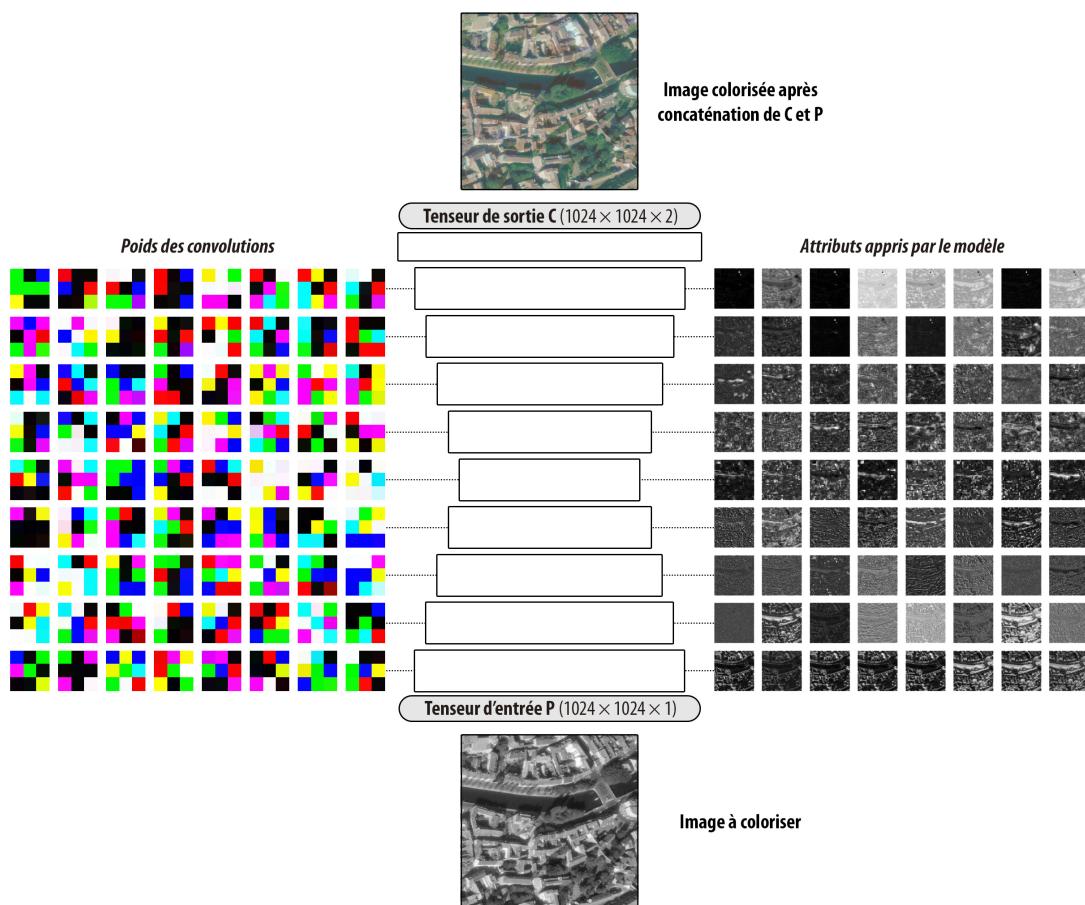


FIGURE 26 – Visualisation d'un extrait des poids et attributs appris pour chaque couche du modèle global, à l'étape n°950. L'image utilisée en entrée pour la colorisation correspond à un extrait tiré de l'orthophotographie de 1978 dans le centre-ville de Strasbourg. Seuls les huit premiers poids et attributs de chaque couche ont été représentés par souci de lisibilité.

Il est intéressant de noter que les zones pour lesquelles le modèle n'a pas su produire de colorisation plausible sont facilement identifiables sur les deux dernières couches du générateur. En

3.4. Résultats des classifications historiques sur les produits colorisés

effet, elles possèdent des valeurs extrêmes, qui expliquent ainsi pourquoi ces portions d'image prennent des couleurs si particulières sur le produit fini. En synthèse additive, la façade cyan située dans la partie NNE de la Figure 25 correspond à la combinaison du vert et du bleu. Ces deux couleurs possèdent chacune la valeur extrême de -1 dans l'espace colorimétrique Lab. Cela confirme donc l'idée précédente, dans la mesure où cette même façade est facilement identifiable par deux petites tâches blanches ou noires, selon les attributs, sur l'avant-dernière couche du générateur (Figure 26).

3.4 Résultats des classifications historiques sur les produits colorisés

La section précédente a ainsi permis de mettre en évidence les capacités du modèle global à coloriser des clichés historiques. Il existe cependant des erreurs de colorisation, qui pénalisent donc la qualité du produit dans son ensemble, phénomène difficilement palpable avec les indicateurs proposés pour l'évaluation, car ils ne rendent pas compte de la plausibilité visuelle des résultats. En substitut, une analyse déjà plus sémantique s'impose donc pour vérifier que les colorisations obtenues sont effectivement mobilisables.

La mise en œuvre d'un scénario de classification sur un produit panchromatique et son homologue colorisé a permis de révéler des résultats similaires à ceux obtenus dans la Partie 3.1, dans la mesure où la chrominance apporte toujours plus d'informations qu'une bande panchromatique et les textures associées.

	PAN	XS	Gain (%)
	S5	S5	
■ Surface agricole	0,59	0,72	22
■ Végétation arborée	0,75	0,78	4
■ Végétation herbacée	0,59	0,66	12
■ Bâti	0,48	0,66	38
■ Route	0,37	0,56	51
■ Autres surfaces	0,42	0,52	24
Score F1 moyen	0,58	0,68	17

TABLE 13 – Valeurs des scores F1 obtenus sur un jeu de validation suite à la classification réalisée sur un extrait d'orthophotographie panchromatique et son homologue colorisé.

La Table 13 montre que les gains les plus élevés sont obtenus, dans l'ordre, pour les routes, le bâti, les autres surfaces, l'agricole, l'herbacé puis l'arboré. Cela ne correspond pas aux tendances énoncées à partir des Tables 10 et 11. Nous pouvons donc supposer que les sémantiques mises en avant par les orthophotographies historiques ne sont pas compatibles avec celles de produits plus récents. En effet, l'angle de prise de vue, l'influence du relief et les dégradations sont en général absentes des images Pléiades ou photographies aériennes numériques. Les changements d'occupation du sol et la résolution spatiale des clichés participent aussi à ces discordances. Cela signifie que la méthode mise en œuvre dans les parties 2.1.2 et 3.1 pour simuler des produits historiques n'a pas été suffisamment efficace. En effet,

Chapitre 3. Résultats

celle-ci se focalisait sur la dimension radiométrique seulement (bruit et flou), en omettant la sémantique (effets de la lentille et du terrain) qui est donc plus importante, probablement du fait de l'analyse texturale menée en amont de la classification.

A noter que les statistiques présentées ici pour la Table 13 auraient pu être améliorées par l'ajout d'un poste "ombre portée". Celui-ci n'a finalement pas été créé par souci de cohérence avec la typologie proposée par le LIVE, pour sa couche d'occupation du sol de 2012.

D'un point de vue qualitatif, la qualité du résultat de classification s'améliore largement avec l'apport de la chrominance. En effet, la structure de la scène ressort de façon plus évidente, et montre moins de confusions entre les différents types de surfaces, notamment au niveau de l'espace artificialisé (Figure 27).

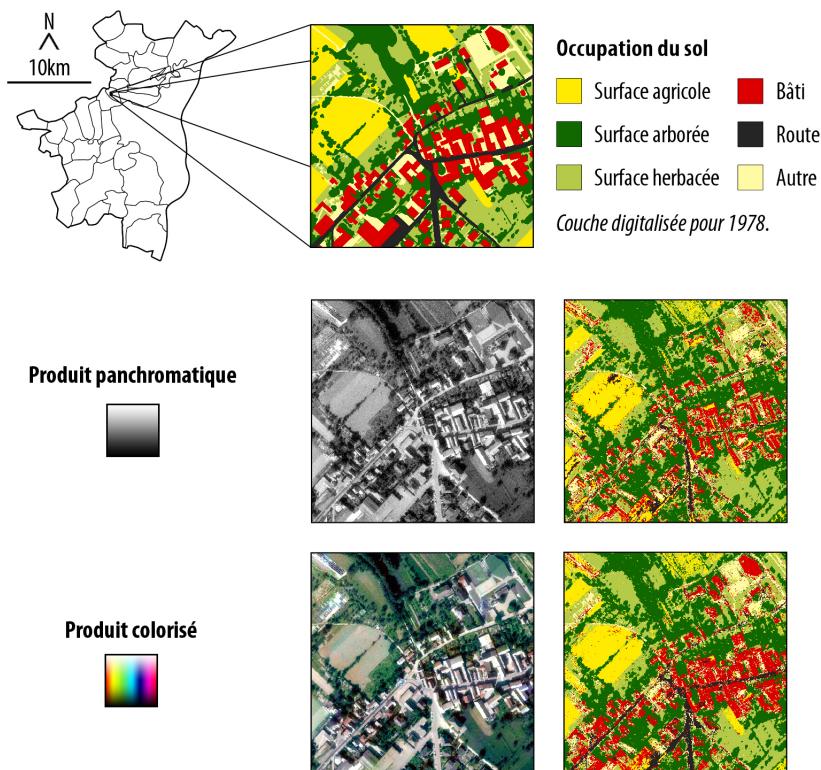


FIGURE 27 – Résultats de classification pour le cliché panchromatique de 1978 et son homologue colorisé, obtenu à partir du modèle global à son itération n°950.

De façon étonnante également, les orthophotographies récentes et images Pléiades génèrent des résultats de classification parfois moins bon que ceux obtenus à partir des produits colorisés, et ce pour des postes spécifiques. C'est notamment le cas des surfaces végétalisées et agricoles, qui revêtent des couleurs particulièrement plausibles, ou possèdent une sémantique suffisamment discriminante pour que le classifieur les sépare correctement des autres classes. L'urbain donne quant à lui généralement des valeurs de F1 inférieures à celles obtenues précédemment. Une analyse visuelle montre une confusion entre le bâti, les routes, les autres

3.4. Résultats des classifications historiques sur les produits colorisés

surfaces et la végétation herbacée principalement. Nous pouvons supposer que la qualité de la colorisation est ici directement en tors, avec une relation évidente entre des valeurs de chrominance inadaptées et des erreurs de classification. C'est ainsi l'exemple des toitures colorisées en vert, qui ressortent généralement comme étant de la végétation herbacée (Figure 27). Le classifieur fait alors la même erreur que le générateur, du fait d'un manque de sémantiques suffisamment discriminantes pour différencier les deux types de surfaces.

Les classifications de l'occupation du sol sont donc un moyen intéressant pour évaluer un modèle de colorisation. En effet, elles peuvent renseigner sur la qualité sémantique d'une image, dans la mesure où la méthode employée ici combine des informations radiométriques et texturales. Dans le cadre de ce travail, la classification ne renseigne malheureusement pas sur la qualité des colorisations au fil des itérations du modèle, mais prouve tout de même que la prédiction d'une chrominance apporte une information non négligeable pour la manipulation de produits historiques.

4 Discussion

Compte-tenu des résultats présentés dans le chapitre précédent, il est finalement possible de répondre aux différentes questions de recherche posées dans l'introduction. La première, visant à évaluer s'il est possible ou non de coloriser automatiquement des photographies aériennes anciennes, est vraisemblablement vérifiée. En effet, les modèles développés ont permis d'aboutir à des produits historiques en couleurs, au contenu pertinent même si parfois incorrect visuellement.

Le fait d'être parvenu à coloriser ces clichés répond positivement aux autres questions que nous avions posées, mais de façon indirecte. Les différents points abordés par celles-ci revêtent cependant des caractères complexes, qu'il est important de préciser et d'expliquer. Les sections présentées ultérieurement suivent ainsi la trame de la réflexion abordée plus tôt sur le problème de la colorisation, et apportent diverses pistes de réflexion à envisager pour améliorer les premières solutions développées dans le cadre de ce travail.

4.1 Recommandations et avertissements pour la mise au point de la photothèque

Avant d'aboutir à un modèle et à des colorisations, la première étape consistait à développer une photothèque, ou base d'entraînement. Une première ébauche a ici été proposée pour prédire les canaux de chrominance a et b , à partir d'un produit (pseudo-)panchromatique donné. Différentes pistes pour améliorer cette photothèque sont cependant envisageables.

La première concerne évidemment le nombre de clichés de référence utilisés pour apprendre la fonction $f(p) = c$. En effet, certaines des erreurs de colorisation notées dans la Partie 3.3.2 sont, tout ou partie, la conséquence d'un manque ou de l'absence totale de sémantiques associées à certaines catégories d'objets ou de surfaces. Pour rappel, la base développée pour le modèle global contient 21 400 paires d'imagettes seulement. A titre de comparaison, les travaux état de l'art menés sur la question de la colorisation s'articulent autour de bases d'entraînement de plusieurs centaines de milliers de clichés, avec par exemple : plus de 130

Chapitre 4. Discussion

000 pour Deshpande *et al.* (2016), 1 200 000 pour Guadarrama *et al.* (2017), 1 300 000 pour Lal *et al.* (2017), 2 327 985 pour Iizuka *et al.* (2016), etc.

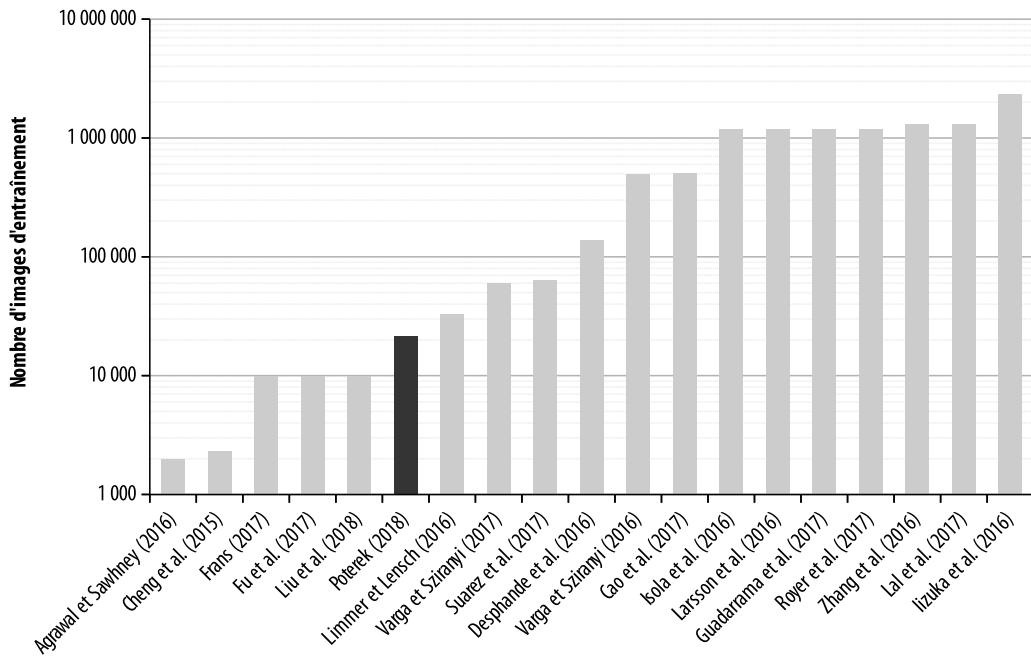


FIGURE 28 – Nombre d'images utilisées par travaux pour l'apprentissage d'un modèle de colorisation. La photothèque développée dans le cadre de ce travail est présentée sur la figure par un bâtonnet noir. L'axe des y est représenté à l'aide d'une échelle logarithmique décimale.

La Figure 28 récapitule le nombre de paires d'imagettes utilisées pour l'entraînement de différents modèles d'apprentissage profond traitant de la colorisation. Elle montre ainsi qu'il y a encore un travail important à faire dans le prototypage de la photothèque. En effet, la base d'images, en son état actuel, représente moins de 5 % de la plupart de celles qui ont été développées pour coloriser des photographies diverses, qui ne sont pas issues de la télédétection. Il y a donc certainement des améliorations à apporter, et qui permettraient d'aboutir à des résultats de meilleure qualité. Plusieurs pistes sont ainsi proposées ci-dessous.

Ce travail, comme la plupart des autres études menées sur la question, utilise des clichés en couleurs qui sont ensuite convertis en noir et blanc, ici un canal pseudo-panchromatique. Si cela ne pose pas ou peu de problèmes dans le cas des photographies issues du dépôt ImageNet par exemple, nous pouvons supposer qu'il en est autrement pour les produits obtenus à partir des méthodes et outils de la télédétection spatiale ou aérienne. Dans la mesure où chaque mission de collecte possède des objectifs précis, les capteurs utilisés pour prendre des clichés de la surface possèdent eux-aussi des caractéristiques spécifiques, en résolutions radiométrique et spatiale particulièrement. Par exemple, un canal panchromatique quelconque couvre l'ensemble du spectre visible, et peut parfois s'étendre à des longueurs d'onde légèrement plus basses ou plus hautes. Les caractéristiques spectrales des surfaces sur ce genre de cliché ne sont donc pas les mêmes que celles issues d'une bande composite

4.1. Recommandations et avertissements pour la mise au point de la photothèque

calculée à partir des canaux R, V et B. Outre les deux catégories d'erreurs renseignées dans la Partie 3.3.2, il est alors raisonnable de supposer que la colorisation souffre aussi de l'utilisation d'une image pseudo-panchromatique plutôt que panchromatique dans le couple $\{C, P\}$. Ce choix a cependant été indispensable afin de disposer de paires complètement homologues. Malgré ce biais, les résultats obtenus restent d'une qualité tout à fait correcte, supposant donc que l'information apportée par la sémantique prime sur la radiométrie. Une première piste d'amélioration serait alors tout de même d'injecter de véritables couples couleur — panchromatique dans la photothèque.

Outre la dimension radiométrique, il est aussi important de rappeler que les orthophotographies mises à disposition dans la base de données géohistorique présentent des résolutions spatiales variées, comprises entre 20 et 50cm. Cela suppose donc une évolution des sémantiques en fonction de l'échelle, qu'il faut donc prendre en compte lors de l'étape d'apprentissage. Parmi les méthodes d'augmentation de données proposées, aucune ne permet réellement de réaliser un rééchantillonnage des produits, cette étape ayant été réalisée en amont. Seule l'architecture U-Net du modèle travaille sur différentes échelles d'analyse, mais cela n'est pas forcément suffisant. Plusieurs pistes sont donc envisageables. La première consisterait simplement à implémenter une transformation aléatoire qui rééchantillonne les paires d'images en entrée selon un ratio fixe ou sélectionné dans une plage de valeurs donnée. Une autre pourrait être d'intégrer à la photothèque des produits aux résolutions variées. Dans le cadre de ce travail, le but était simplement de coloriser les orthophotographies de la base de la ZAEU. Les résolutions ont donc été sélectionnées compte-tenu de ce besoin. Il reste alors envisageable de proposer une photothèque plus extensive, avec des tailles de pixels variées. Ce travail a montré que des résolutions comprises entre 30 et 50cm parviennent à coexister dans une même base d'entraînement, mais il faudrait vérifier que cela soit aussi vrai pour des gammes plus vastes. Si cela n'est pas le cas, il pourrait être envisagé de développer un modèle pyramidal, adapté au traitement de problèmes multi-scalaires. Il s'occuperait de détecter la résolution la plus adaptée pour la colorisation d'un produit, compte-tenu de sa sémantique ou d'une information globale, puis le coloriserait à l'aide des attributs associés.

Un autre point qui a été particulièrement délicat à traiter correspond aux effets liés aux conditions de prise de vue des photographies. En effet, il a par exemple été nécessaire de compléter la photothèque par des clichés pris dans les années 1980 et 1990, afin de disposer de sémantiques propres aux déformations générées par le relief en milieu urbain. Le produit de 2013 a vraisemblablement été pris au nadir, et montre uniquement les toitures des bâtiments. L'influence du relief est pourtant particulièrement présente sur l'ensemble des images panchromatiques de la base géohistorique, avec des bâties penchées et dont les façades peuvent être aperçues. Outre l'apport de nouvelles paires d'images dans la base de données, aucune autre solution n'est *a priori* possible pour palier à ce problème. Ce n'est pas le seul phénomène dû aux conditions de prise de vue, avec par exemple des effets d'éclairement particuliers. Les transformations aléatoires implémentées, comme les rotations ou effets miroir, permettent généralement d'extraire les sémantiques associées. La dimension temporelle reste ici importante, mais pas seulement à l'échelle de la journée. En effet, la base d'images utilisée

pour la colorisation contient uniquement des clichés capturés durant la période estivale. Or, l'orthophotographie de 1956 par exemple a été prise en hiver. En effet, selon ses objectifs, une mission aérienne peut être réalisée à différents moments de l'année. La phénologie des surfaces est donc un problème dont il faut tenir compte pour la colorisation de produits issus de la télédétection. Il serait alors judicieux de compléter la photothèque avec des clichés pris à divers moments de l'année, mais aussi de la journée, et enfin selon des angles d'azimut et d'élévation différents.

4.2 Point sur l'apprentissage du modèle de colorisation

En réponse à la section précédente, les deux catégories d'erreurs de colorisation mises en évidence dans la partie 3.3.2 montrent qu'une base d'entraînement plus vaste et un temps d'apprentissage suffisant auraient permis d'obtenir des résultats de colorisation de meilleure qualité. Certaines méthodes et techniques auraient également apporté des améliorations aux sorties du modèle. Elles n'ont cependant pas été implémentées du fait de leur difficulté, ou simplement par choix suite à une série de tests qui n'ont pas été particulièrement triomphants.

Pour rappel, le modèle ici développé correspond à un DRAGAN et s'articule autour de méthodes de régularisation et d'optimisation variées, avec l'utilisation des fonctions objectif antagonistes et L1, d'une technique de *label smoothing* et d'un algorithme de normalisation spectrale. Le GitHub de Chintala *et al.* (2016) décrit un ensemble de techniques intéressantes à utiliser lors de l'apprentissage d'un GAN. Nous en avons implémenté au total 10 sur 17, certaines n'améliorant tout simplement pas la colorisation, comme l'incorporation d'un *dropout*. D'autres, comme la discrimination de batch, l'*early stopping* ou l'*experience replay* pourraient être pertinentes.

L'un des problèmes soulignés dans les travaux de Salimans *et al.* (2016) correspond au phénomène de *mode collapse*, que nous avions, pour rappel, essayé d'éviter à l'aide de la normalisation spectrale et de différentes implémentations du GAN — les BEGANs et DRAGANs ayant par exemple été retenus —. Une méthode particulièrement attrayante a été proposée par Salimans *et al.* (2016), la discrimination de batch, qui permet d'améliorer la diversité des résultats obtenus en sortie du modèle. Le discriminateur, tel qu'il est conçu dans l'implémentation classique d'un GAN, n'est capable de comparer les distributions de C et \hat{C} que pour l'image qu'il traite sur le moment. Cette nouvelle méthode l'autorise en revanche à observer le batch dans son ensemble, puis à évaluer si \hat{C} provient effectivement de C ou non. Cela serait alors susceptible d'améliorer les distributions bivariées que nous obtenions pour a et b , et qui sont décrites sur les Figures 15 et 24.

Selon Chintala *et al.* (2016), l'*experience replay* pourrait également participer à améliorer les performances des GANs. Technique particulièrement utilisée en apprentissage par renforcement, Pfau et Vinyals (2016) considèrent qu'elle serait compatible avec le mode de fonctionnement des modèles générateurs. L'idée serait ainsi de conserver un batch colorisé en mémoire, puis de le passer à D ultérieurement au cours de l'apprentissage. Ainsi, le discri-

4.3. Limites des indicateurs pour l'évaluation d'une colorisation

minateur ne parvient pas à trop rapidement déceler les produits générés, laissant le temps au générateur de s'améliorer. C'est une manière d'ajouter un bruit lors de l'apprentissage, comme cela pourrait être le cas pour un *dropout* fixé directement après une convolution. Ces techniques sont particulièrement intéressantes pour éviter les cas de sur-apprentissage, phénomène qui traduit une difficulté du modèle à généraliser ce qu'il a appris à d'autres sémantiques. A ce titre, l'*early stopping* est une technique qui sert elle-aussi à contourner ce problème. Elle consiste à utiliser deux jeux de données, l'un d'entraînement et l'autre de validation. Lors de l'apprentissage, l'erreur de prédiction est calculée toutes les n itérations sur la validation, en partant du postulat que celle-ci devrait décroître à chaque passe. En cas de sur-apprentissage, elle peut cependant augmenter puisque le modèle ne parvient plus à généraliser correctement, marquant donc le point d'arrêt de l'entraînement, ou d'*early stopping*. D'autres implémentations de cette méthode existent, mais moins répandues, et permettent de s'affranchir d'un jeu de validation, comme celle proposée par Mahsereci *et al.* (2017). Cette technique reste cependant difficile à mettre en œuvre du fait de la nature multi-modale du problème de colorisation, mais aussi des caractéristiques des produits obtenus en sortie de modèles génératifs. Même si les images colorisées sont issues d'une distribution réelle, il faudrait disposer d'indicateurs adaptés pour l'évaluation de la validation, comme présenté plus tard dans la Partie 4.3.

La question des métriques renvoie là-aussi indirectement à la façon dont l'optimisation du modèle s'opère. Même si différentes fonctions objectif ont été testées, il serait intéressant d'en proposer une qui soit adaptée au problème de la colorisation, en tenant donc compte de la dimension multi-modale du problème. En effet, il a été montré dans les Parties 3.2.1 et 3.3.1 que le générateur ne parvient pas à saisir les distributions des valeurs de a et de b dans leur ensemble. Différents auteurs montraient déjà pour les CNNs qu'une fonction objectif de classification fournit de bien meilleurs résultats de colorisation que celles utilisées en régression. La même idée pourrait donc tout aussi bien être transposée aux GANs, même s'ils ne sont pas utilisés pour la régression ou la classification, mais afin de générer de nouveaux produits.

Les différentes pistes et propositions évoquées ici sont certainement loin d'être exhaustives, mais permettent d'apporter un nouveau regard sur la manière dont le modèle a été implanté dans le cadre de cette analyse. Il semble évident que des améliorations sont à apporter sur le plan algorithmique.

4.3 Limites des indicateurs pour l'évaluation d'une colorisation

Outre les aspects liés à la mise au point d'une photothèque puis à l'entraînement d'un modèle de colorisation, il est important d'évaluer la qualité des produits obtenus en sortie.

Cette question est d'autant plus importante et compliquée à traiter dans le cas d'un problème multi-modal. En effet, comme cela a été évoqué dans les parties dédiées à la méthodologie et aux résultats, les métriques de type MSE et PSNR sont purement quantitatives et ne se

Chapitre 4. Discussion

focalisent pas sur la sémantique de l'image. Ces indicateurs ont eu tendance à favoriser les sorties peu vibrantes, généralement désaturées, puisqu'elles minimisent l'erreur globale du système. A nouveau, dans le cas d'un problème multi-modal, une même voiture par exemple peut revêtir une vaste palette de couleurs. Lorsque le modèle lui assigne une carrosserie rouge plutôt que bleue, l'erreur est plus élevée que s'il l'avait colorisée en un gris moyen, expliquant un manque de vibrance sur les sorties. Théoriquement, et comme l'ont annoncé Wang *et al.* (2004), l'indice SSIM devrait aller dans le sens d'une meilleure prise en compte de la sémantique des produits comparés. Nous avons cependant montré qu'il n'est pas adapté au problème de la colorisation, dans la mesure où il n'a pas apporté de résultat qui soit cohérent avec la perception d'un sujet humain, certes subjective. Bien qu'étant un indicateur particulièrement utilisé dans les applications de vision artificielle pour l'apport d'une dimension structurelle, il a été montré par Dosselmann et Yang (2009) qu'il existe une relation évidente entre le SSIM et la MSE. Ces différentes métriques ne sont donc pas adaptées au problème de la colorisation, qui nécessite d'aller vers des méthodes permettant de mieux tenir compte de la sémantique des images comparées. L'ensemble de ces points peut alors remettre en cause le choix du DRAGAN pour le développement des modèles, puisque MSE, PSNR et SSIM ont été utilisés pour sélectionner un algorithme parmi ceux initialement retenus. Une analyse visuelle des résultats montre cependant qu'il reste toujours supérieur au BEGAN et au cGAN, au moins subjectivement.

La question d'indicateurs pertinents est d'autant plus importante dans le cas des modèles génératifs, comme l'ont montré Theis *et al.* (2015). En effet, pour des produits matriciels destinés à la visualisation et à des applications en vision artificielle, il est nécessaire de disposer d'une information sur la plausibilité des sorties. L'idée serait donc de s'approcher au mieux du mode de fonctionnement du système visuel humain, au moins en termes de performances. Même si une évaluation sémantique a été réalisée à l'aide d'une classification, afin de comparer l'apport de la couleur vis-à-vis d'une image panchromatique simple, elle ne permet pas de réellement situer la qualité du résultat de colorisation.

Comme cela a été expliqué dans la Partie 2.2.3, l'*Inception Score* proposé par Salimans *et al.* (2016) est une métrique qui pourrait répondre à ce besoin. Etant donné que l'IS travaille généralement avec un modèle de classification Inception v3, il ne peut être utilisé que lorsque les images en sortie appartiennent aux classes du dépôt ImageNet. L'idée serait donc de reprendre cet indicateur ainsi que l'algorithme associé, puis de les modifier pour qu'ils fonctionnent avec des produits issus de la télédétection spatiale et aérienne. D'un point de vue méthodologique, cela reviendrait tout d'abord à entraîner un modèle de classification ou de segmentation sémantique pour des images déjà en couleurs, avec un ensemble de postes bien définis. Une fois entraîné, les colorisations générées lui seraient passées, donnant alors une information sur la qualité et la diversité des prédictions d'une ou plusieurs classes. Pour ce travail en particulier, un modèle de segmentation sémantique multi-poste semble plus adapté, dans la mesure où il apporterait une information locale sur la qualité des colorisations, pour chaque objet représenté sur la scène. La question d'un indicateur universel et dérivé de l'IS soulève aussi le besoin d'une échelle spatiale *a priori* globale, plutôt que limitée à l'emprise seule de l'EMS.

4.4. Colorisation et valorisation des produits géographiques historiques

Cela supposerait cependant de disposer d'une infrastructure comparable à celle d'ImageNet par exemple. Son développement reposera sur un travail important, dans la mesure où elle devrait nécessairement viser à être exhaustive, pour un territoire donné, dans les sémantiques mises à disposition sur la plateforme. Ce point suscite plusieurs interrogations, liées à l'accessibilité, la disponibilité ou encore la qualité des données par exemple. Il serait aussi question de définir un corpus des spécifications souhaitées pour les produits distribués, dans la mesure où chaque mission aérospatiale possède des objectifs précis, rarement équivalents ou interchangeables.

D'autres méthodes plus traditionnelles existent également, comme le test visuel de Turing. Cependant, nous travaillons ici sur la colorisation de produits historiques, pour lesquels il n'existe pas encore de version en couleurs. Compte-tenu des dégradations que présentent ces mêmes images, il y aurait forcément un biais dans la manière dont les sujets répondent aux produits présentés, puisque les photographies anciennes peuvent être facilement décelables. Il serait alors question de développer des techniques ou outils permettant de corriger les clichés, en enlevant les expurgeant de ces dégradations, puis en les colorisant, avant de les soumettre au test visuel.

La colorisation revêt alors ici un caractère triplement complexe, du fait de sa nature multimodale, de l'absence d'outils spécifiquement dédiés aux produits géographiques et de l'utilisation d'un modèle génératif. Le problème serait cependant le même pour un CNN par exemple, que ce soit avec une fonction objectif de classification ou de régression, puisque l'évaluation des sorties resterait compliquée. Une réflexion portant simultanément sur l'ensemble de ces points est donc nécessaire pour améliorer les sorties du modèle de colorisation.

4.4 Colorisation et valorisation des produits géographiques historiques

Les deux sections précédentes ont montré l'importance de disposer de métriques adaptées pour l'évaluation des produits colorisés. En effet les résultats obtenus montrent deux catégories d'erreurs de colorisation, décrites et mises en évidence dans la Partie 3.3.2. Ces zones pour lesquelles le modèle n'a pas été suffisamment efficace sont cependant intéressantes. En effet, elles sont facilement identifiables sur les attributs calculés par les deux dernières couches du générateur, et cela par des valeurs extrêmes. Cela signifie donc qu'il pourrait exister un moyen d'évaluer spatialement et localement la qualité d'une colorisation, si tant est que nous puissions trouver une méthode adaptée.

Malgré ces erreurs, les résultats obtenus sont pertinents, même si parfois peu plausibles. Ce succès montre qu'il est effectivement possible de coloriser des orthophotographies historiques, et même d'aller plus loin. Des termes d'avantage adaptés seraient alors la spectralisation pour la prédiction d'un canal, ou la multi-spectralisation pour la prédiction de plusieurs canaux, plutôt que la colorisation seulement. Outre les bandes associées au domaine du visible, il

Chapitre 4. Discussion

a été souligné dans la Partie 3.2.2 que l'apprentissage puis la prédiction d'un canal proche-infrarouge est une tâche tout-à-fait réalisable. Cela sous-entend qu'il est donc possible de prédire des valeurs de réflectance dans d'autres régions du spectre électromagnétique, comme l'infrarouge moyen, voire même pourquoi pas dans le domaine des hyperfréquences. Une possibilité plutôt triviale serait aussi la prédiction d'attributs radiométriques voire texturaux, comme un SAVI ou une entropie.

La colorisation n'est pas la seule force du modèle proposé. En effet, le générateur extrait un total de 2 434 attributs à partir d'un cliché panchromatique qui lui est passé. Or, le fait d'avoir travaillé sur une architecture U-Net montre que le modèle travaille sur différentes échelles, du fait d'un rééchantillonnage successif des images passées aux couches cachées. Les dimensions spatiale et sémantique sont ainsi combinées, chose qui est dès lors plus compliquée à mettre en œuvre avec des méthodes classiques d'extraction d'attributs. En effet, celles-ci travaillent en général sur une échelle unique, au niveau du pixel ou d'un noyau de convolution. Il serait donc envisageable d'utiliser le générateur pour l'extraction d'attributs, servant plus tard pour des applications plus classiques par exemple, en entrée d'un classifieur Random Forest ou d'autres modèles notamment.

Ces différentes pistes ouvrent la porte à de nouvelles solutions pour la manipulation des photographies historiques. En effet, la possibilité d'extraire aussi facilement des attributs pour la colorisation de clichés signifie que les mêmes produits peuvent aider à obtenir de nouvelles informations sur les territoires du passé, du présent, et même peut-être du futur. La discussion lancée sur les améliorations à apporter à la photothèque ou aux indicateurs peut ainsi être élargie à une réflexion plus exhaustive. En effet, nous pourrions imaginer un dépôt géographique de référence et multi-fonctions, pas seulement pour la colorisation, mais également destiné à la classification et à la segmentation sémantique de produits historiques par exemple. D'autres applications moins triviales peuvent aussi être imaginées, comme la prédiction d'un champ de profondeur lorsqu'aucun couple stéréoscopique n'est disponible, la super-résolution, la coregistration d'un ensemble de tuiles, la correction des dégradations présentes sur ces clichés, etc.

Conclusion

La ZAEU ayant exprimé le besoin de valoriser sa base de données géohistorique, un travail de recherche a été mené afin de coloriser les orthophotographies monochromatiques qui la constituent. Pour ce faire, une première évaluation de l'apport de la couleur a été réalisée. Celle-ci montre que la chrominance est effectivement supérieure au canal panchromatique et aux attributs dérivés. Plusieurs méthodes et techniques de colorisation ont donc été étudiées puis comparées. Cela nous a rapidement permis d'éliminer les procédés manuels et semi-automatiques, peu pertinents dans le cas des produits géographiques. Il fallait en effet proposer une méthode opérationnelle et simple à mettre en œuvre sur l'ensemble du territoire de l'Eurométropole de Strasbourg. Les techniques basées sur l'apprentissage profond, dont la popularité suit étroitement l'amélioration des capacités de calcul informatique, ont été finalement retenues du fait de leurs capacités d'automatisation. Un modèle génératif antagoniste a été créé puis entraîné à partir de produits matriciels, stockés dans une photothèque. Cette dernière a été constituée dans l'objectif d'apprendre un ensemble de sémantiques *ad hoc*, suffisantes pour traiter le territoire de l'EMS.

Il a été montré que les colorisations obtenues à partir de modèles génératifs sont pertinentes, à la fois dans le visible et le proche-infrarouge, même si loin d'être parfaites. En effet, il est important de rappeler que ce travail reste exploratoire avant tout. Ces produits générés ont également pu être injectés en entrée de méthodes plus classiques utilisées en télédétection, comme la classification d'images. Cela a permis d'obtenir une cartographie de l'occupation du sol sur une emprise limitée, témoignant donc de l'interopérabilité de la méthode ici développée.

Malgré les succès évoqués, les résultats de colorisation présentent certaines incohérences. Différentes pistes visant à améliorer le modèle ont ainsi été proposées. Parmi celles-ci se trouvent un temps d'apprentissage plus long, une photothèque plus exhaustive, une meilleure optimisation du modèle ainsi que le développement de métriques adaptées au problème traité. La colorisation reste en effet une thématique récente qui, même en ayant fait l'objet de plusieurs publications, suscite de nombreux questionnements. En effet, sa nature multimodale reste un frein au développement de méthodes et indicateurs adéquats.

L'originalité de ce travail est ainsi plurielle, puisqu'il s'intéresse au problème de la colorisation de produits géographiques historiques, à l'aide de modèles génératifs. Il propose également

Chapitre 4. Discussion

de nombreuses pistes pour le développement d'un dépôt à vocation multi-usage et universel, dans lequel seraient stockées des images issues des outils et méthodes de la télédétection spatiale et aérienne. Si une telle structure venait à se mettre en place et pouvait servir de base pour le développement de nouvelles méthodes génératives, il faudrait alors envisager de transformer la manière dont les données sont transmises et utilisées... En effet, Szegedy *et al.* (2013) et Goodfellow *et al.* (2014b) ont montré que les produits issus de modèles antagonistes perturbent les réseaux de neurones profonds et leurs dérivés, ce qui pourrait donc se généraliser à d'autres méthodes classiquement utilisées en géomatique. Quel impact pour les métadonnées et la qualité des produits obtenus à partir de ces méthodes génératives ? Quelle conséquence pourrait par exemple avoir un déluge d'images générées ? Est-ce que les produits obtenus à l'aide d'un processus génératif peuvent finalement être mobilisés dans les chaînes de traitements géographiques standards ? Si oui, quelles précautions prendre ? Est-il pertinent d'évaluer l'incertitude liée à l'utilisation de ces produits, et si oui, de quelle manière ?

Bibliographie

- AGRAWAL, M. et SAWHNEY, K. (2016). Exploring convolutional neural networks for automatic image colorization.
- ARJOVSKY, M. et BOTTOU, L. (2017). Towards Principled Methods for Training Generative Adversarial Networks. *ArXiv e-prints*.
- BARRATT, S. et SHARMA, R. (2018). A Note on the Inception Score. *ArXiv e-prints*.
- BELGIU, M. et DRĂGUT, L. (2016). Random forest in remote sensing : A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:24–31.
- BENEDIKTSSON, J., PESARESI, M. et ARNASON, K. (2003). Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Transactions on Geoscience and Remote Sensing*, 41(9):1940–1949.
- BERTHELOT, D., SCHUMM, T. et METZ, L. (2017). BEGAN : Boundary Equilibrium Generative Adversarial Networks. *ArXiv e-prints*.
- BREIMAN, L. (2001). Random forests. *Mach. Learn.*, 45(1):5–32.
- BUGEAU, A. et TA, V.-t. (2012). Patch-based Image Colorization To cite this version : Patch-based Image Colorization.
- CAO, Y., ZHOU, Z., ZHANG, W. et YU, Y. (2017). Unsupervised Diverse Colorization via Generative Adversarial Networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10534 LNAI:151–166.
- CAVALLARO, G., MURA, M. D., BENEDIKTSSON, J. A. et PLAZA, A. (2016). Remote sensing image classification using attribute filters defined over the tree of shapes. *IEEE Transactions on Geoscience and Remote Sensing*, 54(7):3899–3911.
- CHAMBOLLE, A. (2004). An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97.
- CHARPIAT, G., HOFMANN, M. et SCHÖLKOPF, B. (2008). Automatic image colorization via multimodal predictions. *Lecture Notes in Computer Science (including subseries Lecture*

Bibliographie

- Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 5304 LNCS(PART 3):126–139.*
- CHENG, Z., YANG, Q. et SHENG, B. (2015). Deep colorization. *Proceedings of the IEEE International Conference on Computer Vision, 2015 Inter*:415–423.
- CHEVALLIER, R. (1965). *Photographie aérienne panorama intertechnique*. Gauthier-Villars éditeur, Paris.
- CHEVALLIER, R., CARBONNELL, M. et GUY, M. (1968). *Panorama des applications de la photographie aérienne*. Ecole Pratique des Hautes Etudes 5. S.E.V.P.E.N, Paris.
- CHIA, A. Y.-S., ZHUO, S., GUPTA, R. K., TAI, Y.-W., CHO, S.-Y., TAN, P. et LIN, S. (2011). Semantic colorization with internet images. *ACM Transactions on Graphics*, 30(6):1.
- CHINTALA, S., DENTON, E., ARJOVSKY, M. et MATHIEU, M. (2016). How to Train a GAN? Tips and tricks to make GANs work. <https://github.com/soumith/ganhacks>.
- COLLOBERT, R., KAVUKCUOGLU, K. et FARABET, C. (2011). Torch7 : A matlab-like environment for machine learning. *In BigLearn, NIPS Workshop*.
- DESHPANDE, A., LU, J., YEH, M.-C., CHONG, M. J. et FORSYTH, D. (2016). Learning Diverse Image Colorization.
- DESHPANDE, A., ROCK, J. et FORSYTH, D. (2015). Learning Large-Scale Automatic Image Colorization. *Iccv*.
- DOSSELMANN, R. et YANG, X. D. (2009). A comprehensive assessment of the structural similarity index. *Signal, Image and Video Processing*, 5(1):81–91.
- FRANS, K. (2017). Outline Colorization through Tandem Adversarial Networks.
- FU, K., WANG, Y. et LIU, B. (2017). Cs229 final project automatic colorization for line arts.
- FUKUSHIMA, K. (1988). Neocognitron : A hierarchical neural network capable of visual pattern recognition. *Neural Networks*, 1:119–130.
- GATYS, L. A., ECKER, A. S. et BETHGE, M. (2016). Image style transfer using convolutional neural networks. *The IEEE conference on computer vision and pattern recognition*, pages 2414–2423.
- GOODFELLOW, I., BENGIO, Y. et COURVILLE, A. (2015). Deep Learning. *Nature Methods*, 13(1):35–35.
- GOODFELLOW, I. J., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIR, S., COURVILLE, A. et BENGIO, Y. (2014a). Generative Adversarial Networks. *ArXiv e-prints*, pages 1–9.

- GOODFELLOW, I. J., SHLENS, J. et SZEGEDY, C. (2014b). Explaining and Harnessing Adversarial Examples. *ArXiv e-prints*.
- GRIZONNET, M., MICHEL, J., POUGHON, V., INGLADA, J., SAVINAUD, M. et CRESSON, R. (2017). Orfeo ToolBox : open source processing of remote sensing images. *Open Geospatial Data, Software and Standards*, 2(1).
- GUADARRAMA, S., DAHL, R., BIEBER, D., NOROUZI, M., SHLENS, J. et MURPHY, K. (2017). PixColor : Pixel Recursive Colorization. pages 1–17.
- GUPTA, R. K., CHIA, A. Y.-S., RAJAN, D., NG, E. S. et ZHIYONG, H. (2012). Image colorization using similar images. *Proceedings of the 20th ACM international conference on Multimedia - MM '12*, page 369.
- GURNEY, K. (2007). Neural networks for perceptual processing : from simulation tools to theories. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 362(1479): 339–353.
- HARALICK, R. (1979). Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804.
- HARALICK, R. M., SHANMUGAM, K. et DINSTEIN, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621.
- HARIHARAN, B., ARBELÁEZ, P., GIRSHICK, R. et MALIK, J. (2014). Hypercolumns for Object Segmentation and Fine-grained Localization. *ArXiv e-prints*.
- HE, D. et WANG, L. (1990). Texture unit, texture spectrum, and texture analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):509–512.
- HE, L., WANG, G. et HU, Z. (2018). Learning depth from single images with deep neural network embedding focal length. *CoRR*, abs/1803.10039.
- HOWARD, J. *et al.* (2018). fastai. <https://github.com/fastai/fastai>.
- HUMBERT, P. (2017). *SIG Géohistorique et dynamiques urbaines*. Mémoire de master 1, Université de Strasbourg.
- IGN (2018). Remonter le temps. <https://remonterletemps.ign.fr/>.
- IIZUKA, S., SIMO-SERRA, E. et ISHIKAWA, H. (2016). Let there be color! *ACM Transactions on Graphics*, 35(4):1–11.
- IMMERKÆR, J. (1996). Fast noise variance estimation. *Computer Vision and Image Understanding*, 64(2):300–302.
- IOFFE, S. et SZEGEDY, C. (2015). Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift. *ArXiv e-prints*.

Bibliographie

- IRONY, R., COHEN-OR, D. et LISCHINSKI, D. (2005). Colorization by Example. *Symposium A Quarterly Journal In Modern Foreign Literatures*, pages 201–210.
- ISOLA, P., ZHU, J., ZHOU, T. et EFROS, A. A. (2016). Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004.
- KODALI, N., ABERNETHY, J., HAYS, J. et KIRA, Z. (2017). On Convergence and Stability of GANs. *ArXiv e-prints*.
- KRIZHEVSKY, A., SUTSKEVER, I. et HINTON, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, pages 1–9.
- LAL, S., GARG, V. et VERMA, O. P. (2017). Automatic Image Colorization Using Adversarial Training. *Proceedings of the 9th International Conference on Signal Processing Systems - ICSPS 2017*, pages 84–88.
- LARSSON, G., MAIRE, M. et SHAKHNAROVICH, G. (2016). Learning representations for automatic colorization. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9908 LNCS:577–593.
- LAVÉDRINE, B. et McELHONE, J. (2009). *Photographs of the past process and preservation*. Getty Conservation Institute, Los Angeles.
- LECUN, Y., BOTTOU, L., BENGIO, Y. et HAFFNER, P. (1998). Gradient-based learning applied to document recognition. *In Proceedings of the IEEE*, volume 86, pages 2278–2324.
- LEDIG, C., THEIS, L., HUSZAR, F., CABALLERO, J., CUNNINGHAM, A., ACOSTA, A., AITKEN, A., TEJANI, A., TOTZ, J., WANG, Z. et SHI, W. (2016). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network.
- LILLESAND, T., KIEFER, R. W. et CHIPMAN, J. (2015). *Remote Sensing and Image Interpretation, 7th Edition*. John wiley édition.
- LIMMER, M. et LENSCHE, H. P. A. (2016). Infrared Colorization Using Deep Convolutional Neural Networks.
- LIU, H., FU, Z., HAN, J., SHAO, L. et LIU, H. (2018). Single satellite imagery simultaneous super-resolution and colorization using multi-task deep neural networks. *Journal of Visual Communication and Image Representation*, 53:20–30.
- LIU, X., WAN, L., QU, Y., WONG, T.-T., LIN, S., LEUNG, C.-S. et HENG, P.-A. (2008). Intrinsic colorization. *ACM SIGGRAPH Asia 2008 papers on - SIGGRAPH Asia '08*, page 1.
- LU, D. et WENG, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5):823–870.

- LUAN, Q., WEN, F., COHEN-OR, D., LIANG, L., XU, Y.-Q. et SHUM, H.-Y. (2007). Natural Image Colorization. *Rendering Techniques*, pages 309–320.
- MAGGIORI, E., TARABALKA, Y., CHARPIAT, G. et ALLIEZ, P. (2017). Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE.
- MAHSERECI, M., BALLES, L., LASSNER, C. et HENNIG, P. (2017). Early Stopping without a Validation Set. *ArXiv e-prints*.
- MEDINA KENNEDY, S., POTEREK, Q. et SINDT, A. (2018). *SIG Géohistorique et dynamiques urbaines : Le cas de la moitié Nord de l'Eurométropole de Strasbourg, de 1956 à 2013*. Mémoire de master 2, Université de Strasbourg.
- MIRZA, M. et OSINDERO, S. (2014). Conditional generative adversarial nets. *CoRR*, abs/1411.1784.
- MIYATO, T., KATAOKA, T., KOYAMA, M. et YOSHIDA, Y. (2018). Spectral Normalization for Generative Adversarial Networks. *ArXiv e-prints*.
- MOISSON, S. (2015). *Analyse des évolutions urbaines à partir de données multi-sources : Application à l'Eurométropole Strasbourgeoise*. Mémoire de master 2, Université de Strasbourg.
- MURAZ, J., DURRIEU, S., LABBE, S., ANDREASSIAN, V. et TANGARA, M. (1999). Comment valoriser les photos aériennes dans les SIG? *Ingénieries - EA T*, (20):p. 39 – p. 58.
- NOUCHER, M. (2013). Infrastructures de données géographiques et flux d'information environnementale. *Networks and Communication Studies*, 2:120–147.
- OJALA, T., PIETIKAINEN, M. et MAENPAA, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987.
- PALSSON, F., SVEINSSON, J. R., BENEDIKTSSON, J. A. et AANAES, H. (2012). Classification of pansharpened urban satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(1):281–297.
- PASZKE, A., GROSS, S., CHINTALA, S., CHANAN, G., YANG, E., DEVITO, Z., LIN, Z., DESMAISON, A., ANTIGA, L. et LERER, A. (2017). Automatic differentiation in pytorch.
- PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M. et DUCHESNAY, E. (2011). Scikit-learn : Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- PFAU, D. et VINYALS, O. (2016). Connecting Generative Adversarial Networks and Actor-Critic Methods. *ArXiv e-prints*.

Bibliographie

- PRATT, L. Y. (1993). Discriminability-based transfer between neural networks. In *Advances in Neural Information Processing Systems 5, [NIPS Conference]*, pages 204–211, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- QU, Y., WONG, T.-T. et HENG, P.-A. (2006). Manga colorization. *ACM Transactions on Graphics*, 25(3):1214.
- RADFORD, A., METZ, L. et CHINTALA, S. (2015). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. pages 1–16.
- REINHARD, E., ASHIKHMEN, M., GOOCH, B. et SHIRLEY, P. (2001). Color transfer between images. *IEEE Computer Graphics and Applications*, 21(5):34–41.
- RONNEBERGER, O., FISCHER, P. et BROX, T. (2015). U-Net : Convolutional Networks for Biomedical Image Segmentation. *ArXiv e-prints*.
- ROYER, A., KOLESNIKOV, A. et LAMPERT, C. H. (2017). Probabilistic image colorization. *CoRR*, abs/1705.04258:1–12.
- RUMELHART, D. E., HINTON, G. E. et WILLIAMS, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.
- SALIMANS, T., GOODFELLOW, I., ZAREMBA, W., CHEUNG, V., RADFORD, A. et CHEN, X. (2016). Improved Techniques for Training GANs. *ArXiv e-prints*.
- SAUTER, J.-Y. et SCHWARTZ, J. (2017). *Constitution d'un SIG géohistorique et analyse des dynamiques urbaines : le cas de l'Eurométropole de Strasbourg*. Mémoire de master 2, Université de Strasbourg.
- SHELHAMER, E., LONG, J. et DARRELL, T. (2016). Fully Convolutional Networks for Semantic Segmentation. *arXiv cvpr*, pages 1–14.
- SIMONYAN, K. et ZISSERMAN, A. (2014). Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556.
- SONG, Q., XU, F. et JIN, Y. (2017). SAR image colorization : Converting single-polarization to fully polarimetric using deep neural networks. *CoRR*, abs/1707.07225.
- SUÁREZ, P. L., SAPPA, A. D. et VINTIMILLA, B. X. (2017). Learning to colorize infrared images. *Advances in Intelligent Systems and Computing*, 619:164–172.
- SÝKORA, D., DINGLIANA, J. et COLLINS, S. (2009). LazyBrush : Flexible painting tool for hand-drawn cartoons. *Computer Graphics Forum*, 28(2):599–608.
- SZEGEDY, C., ZAREMBA, W., SUTSKEVER, I., BRUNA, J., ERHAN, D., GOODFELLOW, I. et FERGUS, R. (2013). Intriguing properties of neural networks. *ArXiv e-prints*.
- TAYLOR, L. et NITSCHKE, G. (2017). Improving Deep Learning using Generic Data Augmentation. *ArXiv e-prints*.

- THEIS, L., VAN DEN OORD, A. et BETHGE, M. (2015). A note on the evaluation of generative models. *ArXiv e-prints*.
- van der WALT, S., SCHÖNBERGER, J. L., NUNEZ-IGLESIAS, J., BOULOGNE, F., WARNER, J. D., YAGER, N., GOUILLART, E. et YU, T. (2014). Scikit-image : Image processing in python. *PeerJ*, 2:e453.
- VARGA, D. et SZIRÁNYI, T. (2016). Fully automatic image colorization based on Convolutional Neural Network. *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 3691–3696.
- VARGA, D. et SZIRÁNYI, T. (2017). Twin Deep Convolutional Neural Network for Example-Based Image Colorization. volume 2, pages 184–195.
- WANG, Z., BOVIK, A., SHEIKH, H. et SIMONCELLI, E. (2004). Image quality assessment : From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.
- WARNER, W. S. (1995). *Small Format Aerial Photography*. Whittles.
- WELSH, T., ASHIKMIN, M. et MUELLER, K. (2002). Transferring color to greyscale images. *ACM Transactions on Graphics*, 21(3).
- XIA, G.-S., HU, J., HU, F., SHI, B., BAI, X., ZHONG, Y., ZHANG, L. et LU, X. (2017). AID : A Benchmark Data Set for Performance Evaluation of Aerial Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55:3965–3981.
- YANG, X. et ZHU, C. (1998). Study of remote sensing image texture analysis and classification using wavelet. *International Journal of Remote Sensing*, 19:3197–3203.
- YANG, Y. et NEWSAM, S. (2010). Bag-of-visual-words and spatial extensions for land-use classification. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS ’10, pages 270–279, New York, NY, USA. ACM.
- YATZIV, L. et SAPIRO, G. (2006). Fast image and video colorization using chrominance blending. *IEEE Transactions on Image Processing*, 15(5):1120–1129.
- YOSINSKI, J., CLUNE, J., BENGIO, Y. et LIPSON, H. (2014). How transferable are features in deep neural networks ? *ArXiv e-prints*.
- ZAEU (2013). Rapport Annuel 2012 - Zone Atelier Environnement Urbain. Rapport technique, ZAEU, Strasbourg.
- ŽBONTAR, J. et LE CUN, Y. (2015). Computing the stereo matching cost with a convolutional neural network. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June(1):1592–1599.
- ZHANG, H., GOODFELLOW, I., METAXAS, D. et ODENA, A. (2018). Self-Attention Generative Adversarial Networks. *ArXiv e-prints*.

Bibliographie

ZHANG, R., ISOLA, P. et EFROS, A. A. (2016). Colorful image colorization. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9907 LNCS:649–666.

ZOU, Q., NI, L., ZHANG, T. et WANG, Q. (2015). Deep learning based feature selection for remote sensing scene classification. *IEEE Geoscience and Remote Sensing Letters*, 12(11):2321–2325.

A Récapitulatif des étapes et tâches réalisées

Ce mémoire a été le fruit d'un travail de recherche portant sur les méthodes d'apprentissage profond, de colorisation, de comparaison d'images et de classification. La Figure A.1 décrit la répartition temporelle des différentes tâches réalisées.

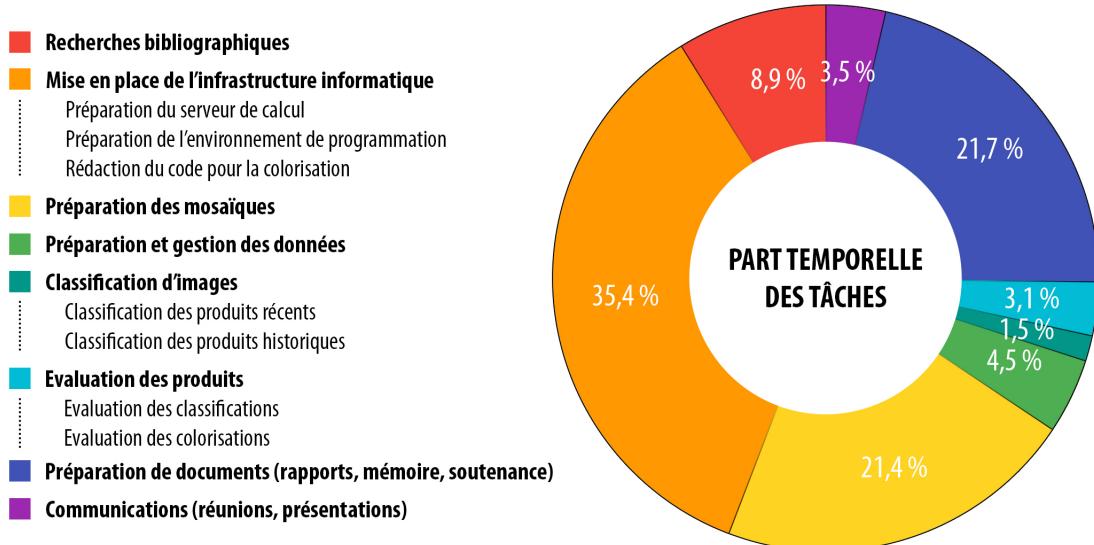


FIGURE A.1 – Répartition temporelle des tâches réalisées.

Certaines étapes nécessaires à la mise en œuvre de ce travail n'ont cependant pas été décrites, car jugées préliminaires.

La première a consisté en une comparaison de différentes offres pour la mise en place d'une infrastructure informatique. En effet, comme cela est expliqué plus en détails dans l'Annexe B, nous avions besoin de disposer d'un GPU afin de mieux paralléliser les calculs et pouvoir donc travailler avec des méthodes d'apprentissage profond. Il existe différentes plateformes qui proposent ce type d'offre, comme Google Cloud, Amazon Web Services, Azure, Crestle, Floydhub ou encore Paperspace. Elles ont été comparées compte-tenu des GPUs mis à dispo-

Annexe A. Récapitulatif des étapes et tâches réalisées

sition, de leur prix, des éventuelles réductions, et de la facilité à créer une instance de calcul puis à y accéder. Ces premières analyses nous ont permis de retenir Google Cloud, Amazon Web Services et Paperspace, qui ont ensuite été testés. Sans trop entrer dans les détails, Google Cloud a été sélectionné. Amazon Web Services était trop complexe pour l'envergure de ce projet, tandis que Paperspace n'offrait pas suffisamment de flexibilité sur le plan matériel. Le temps dédié à tout cela représente une vingtaine d'heures. Une fois la plateforme sélectionnée, celle-ci a été préparée, étape qui fait partie intégrante de la composante "Mise en place de l'infrastructure informatique".

Le temps qui a été consacré à l'apprentissage de la bibliothèque PyTorch pour le *deep learning* n'a pas non plus été représenté, mais équivaut à un peu plus de 30h. Différentes sources ont été utilisées. La principale correspond au MOOC "*Practical Deep Learning For Coders*" (Howard *et al.*, 2018), qui constitue une bonne entrée en matière pour les néophytes de l'apprentissage profond. Ce cours requiert de savoir manipuler le langage de programmation Python, et s'organise en deux parties, disponibles aux adresses suivantes : <http://course.fast.ai/> et <http://course.fast.ai/part2.html> respectivement. Les vidéos publiées sur la chaîne YouTube de Sung Kim (<https://goo.gl/haRrpG>), ainsi que les tutoriels proposés sur le site web de PyTorch (<https://pytorch.org/tutorials/>), permettent également une meilleure compréhension des concepts généraux, et sont suffisants pour créer des modules et modèles simples. Le temps passé à apprendre comment mettre en œuvre différentes techniques de colorisation est intégré à la composante "Mise en place de l'infrastructure informatique".

La Figure A.2 présente quant à elle les aspects plus techniques de ce travail. Elle résume ainsi, de façon très globale, l'ensemble des tâches qui ont été réalisées pour aboutir aux résultats de colorisation finaux. Pas toutes les étapes sont représentées, comme l'entraînement d'un modèle local puis global, ou encore l'apprentissage par transfert. En effet, elles sont suffisamment générales pour être associées à l'un des cadrants de la Figure A.2. S'il y a besoin de plus de renseignements sur chaque étape, une description détaillée est disponible dans le Chapitre 2, dédié aux données et à la méthodologie employée. Succinctement, le cadrant A décrit une tâche préliminaire au travail de colorisation, consistant à évaluer l'apport de la couleur vis-à-vis d'un produit panchromatique. La création de la photothèque ainsi que son utilisation pour l'apprentissage d'un modèle de colorisation sont décrites par la lettre B. Le cadrant C représente quant à lui le travail de géoréférencement et de mosaïquage qui a été réalisé sur certains produits disponibles dans la base de données géohistorique. Cette étape est décrite plus en détails dans l'Annexe B, car elle ne répond pas directement aux questions abordées dans ce travail, et n'a pas non plus nécessité de développements méthodologiques. Enfin, le cadrant D est réservé à la mise en œuvre des colorisations historiques, à partir des mosaïques ou de clichés pris individuellement.

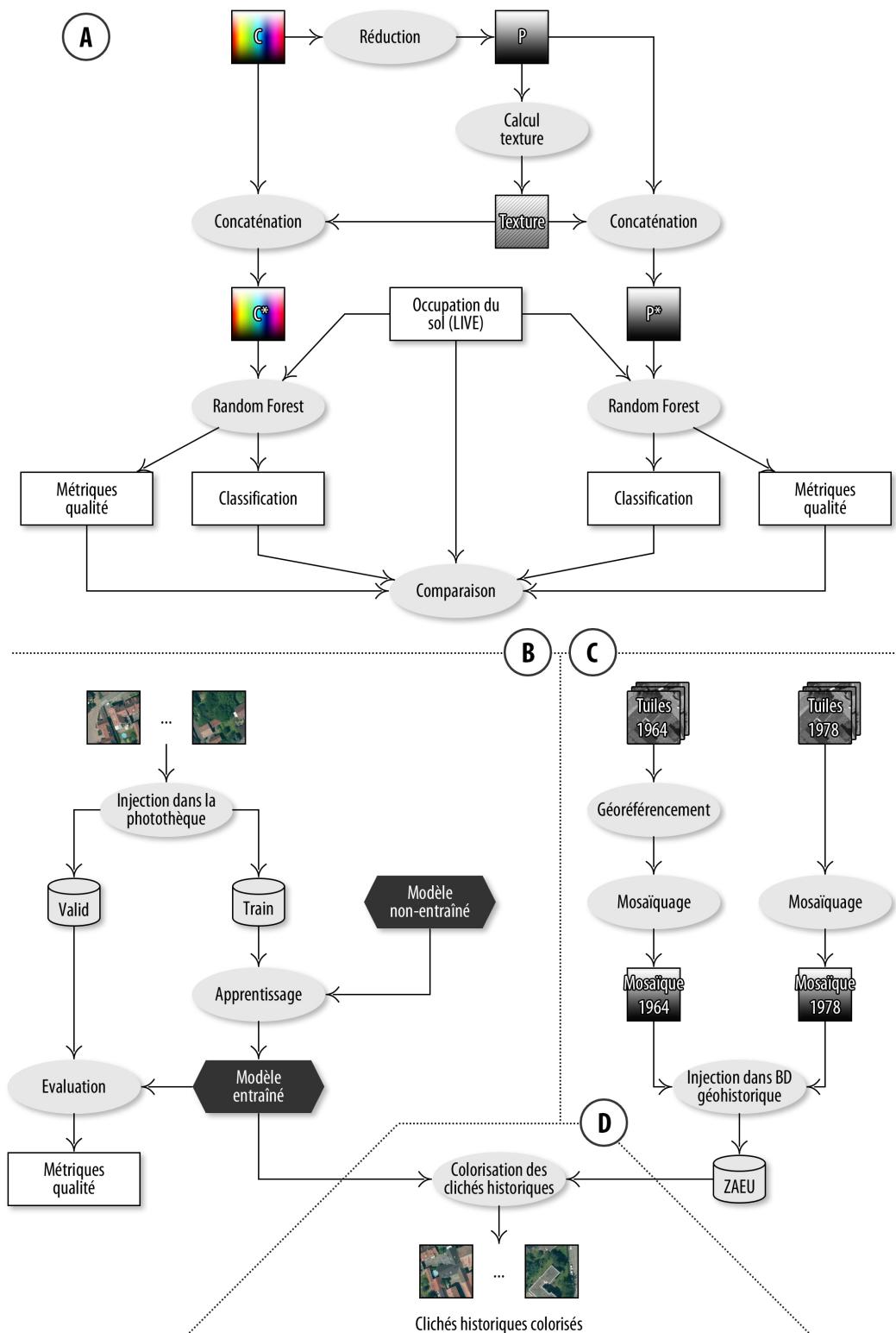


FIGURE A.2 – Description générale des étapes du travail de recherche.

B Préparation des mosaïques

La ZAEU dispose d'une base de données géohistorique constituée d'une série temporelle de photographies aériennes orthoréctifiées, capturées entre 1932 et 2013. Afin d'être colorisées, certaines nécessitent un ensemble de pré-traitements qui constituent à les caler et/ou mosaïquer, notamment les clichés de 1964 et 1978. En effet, un travail préliminaire avait été réalisé sur ces produits, mais une inspection visuelle révèle des erreurs dans le géoréférencement des tuiles qui les constituent (1964 et 1978), ou des erreurs de mosaïquage (1978).

Dans le cadre de ce projet de colorisation de photographies historiques, une correction du calage des tuiles ainsi qu'un mosaïquage de celles-ci ont été réalisés pour 1964 et 1978.

Ne disposant pas d'un logiciel adapté pour cette tâche au départ, les orthophotographies de 1978 ont été assemblées à l'aide de Photoshop. Pour des raisons techniques également, le traitement a été partagé en un ensemble de huit chantiers, permettant une meilleure gestion des ressources mémoire. Plus tard, chacun des huits produits a été géoréférencé, en prenant comme références les orthophotographies de 2013 et 2007. Certaines portions du produit initialement fourni pour l'année 1978 ont également été utilisées comme références, uniquement lorsque celles-ci ne présentaient *a priori* aucun problème de calage ou de mosaïquage. Le tout a ensuite été assemblé en une mosaïque finale, dont la résolution est de 30cm.

Pour l'année 1964, une inspection des tuiles a d'abord été réalisée afin de vérifier l'exactitude du géoréférencement. En cas de mauvaise qualité, ce qui a malheureusement souvent été le cas, les produits concernés ont été recalés en prenant comme références les orthophotographies de 2013, 2007, 1978 et 1956. Les problèmes de calage semblent être avant tout le fait de l'orthorectification, réalisée en amont par un-e autre opérateur-trice, avec un modèle numérique de terrain d'une résolution de 25m. Afin d'obtenir des produits d'une qualité acceptable, une méthode spline a été utilisée lors du géoréférencement, nécessitant donc de saisir au moins 10 points d'amer. Compte-tenu des déformations, nous avons un maximum essayé d'obtenir des clichés correctement coregistrés au niveau des zones de recouvrement. Cette étape terminée, nous avons effectué le mosaïquage des tuiles à l'aide du logiciel ArcGIS cette fois-ci. L'ensemble de cette procédure en particulier est décrit sur la Figure B.1.

Annexe B. Préparation des mosaïques

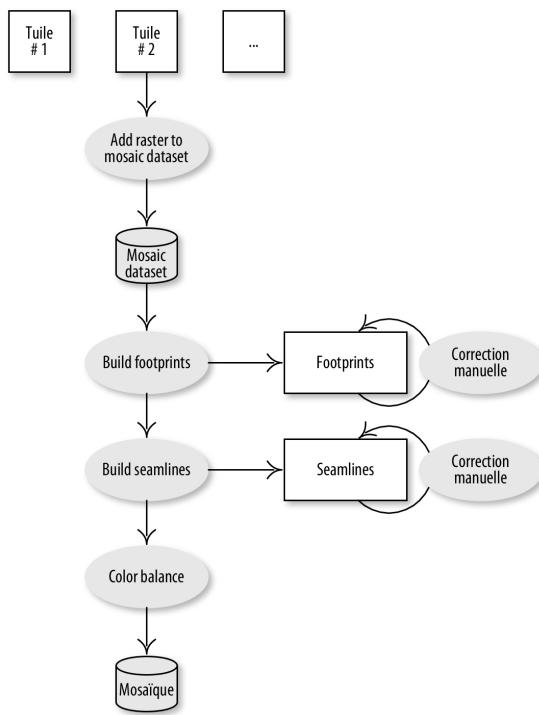


FIGURE B.1 – Description générale des étapes du mosaïquage pour les photographies prises dans l'année 1964.

Pour plus de détails sur la méthodologie employée pour l'année 1964, nous avons développé, à l'aide de solutions ESRI, une base de données géographique dans laquelle a été créé un jeu mosaïqué. L'ensemble des tuiles précédemment géoréférencées ont été ajoutées à ce dernier. Plusieurs étapes ont ensuite été nécessaires pour aboutir à un résultat acceptable. Le premier consiste à créer les emprises de chaque tuile, à l'aide de l'outil "Build footprints". Différents paramètres sont indispensables pour espérer générer des contours débarrassant les clichés de leur cartouche, et sont les suivants : BuildFootprints{Méthode : Radiométrique ; Minimum : 10 ; Maximum : 255 ; Nombre de vertices : 80 ; Shrink : 0 ; Maintain sheet edges : NO_MAINTAIN_EDGES ; Update boundary : UPDATE_BOUNDARY ; Request Size : 2 000 ; Minimum Region Size : 25 000 ; Simplification Method : None ; Maximum Sliver Size : 20 ; Minimum Thinness Ratio : 0,05}. Les bordures ont ensuite été reprises manuellement afin d'éliminer les artefacts restants. Une fois cette étape terminée, nous avons généré les lignes de jointure entre les différentes tuiles, avec les paramètres suivants : BuildSeamlines{Méthode : Radiométrique ; Blend Type : Both ; Request Size : 10 ; Type : PIXELS}. Elles ont aussi été corrigées afin d'obtenir une sortie correcte. A noter que la correction des bordures et lignes de jointure requiert d'utiliser un mode d'édition topologique pour conserver la cohérence des entités décrites par le jeu de données mosaïqué. Enfin, les niveaux de gris ont été corrigés pour chaque tuile afin que le résultat soit le plus uniforme possible, avec la méthode suivante : ColorBalance{Méthode : Dodging ; Color Surface Type : Single Color}. Le résultat a ensuite été exporté à une résolution spatiale de 20cm.

C Mise en place de l'environnement de programmation

Les méthodes d'apprentissage profond requièrent de disposer de capacités de calcul importantes, auxquelles répondent les GPUs, capables de paralléliser les opérations linéaires sur plusieurs centaines, voire milliers de cœurs. En effet, l'entraînement des modèles classiquement utilisés pour la colorisation nécessitent autrement plusieurs jours, voire semaines, sur un CPU classique. La mise en œuvre de ce travail a donc contraint à la location d'un serveur de calcul disposant de GPUs NVIDIA, *leader* mondial dans la conception de processeurs graphiques.

Nous avons ainsi souscrit à Google Cloud, qui offre 300\$ de crédits pour la location d'un serveur GPU, soit l'équivalent de 258€ au moment de la rédaction de ce mémoire. La plateforme donne accès à trois modèles de GPUs NVIDIA — Tesla K80, P100 et V100 — dont les prix sont compris entre 0,40€ et 2,50€ de l'heure. Afin de disposer de suffisamment de crédits pour la mise en œuvre de tests et l'entraînement d'un modèle, nous avons opté pour l'option la moins onéreuse, soit le NVIDIA K80. Outre le GPU, des choix sont à faire vis-à-vis de l'architecture souhaitée, notamment en termes de capacités de stockage. L'ensemble de l'opération, avec mise en place de l'instance, est décrit dans le document */ARCHIVE_CD/ENVIRONNEMENT/TutorielGoogleCloud.odt* qui est disponible dans l'archive fournie avec ce mémoire. A noter que cette procédure n'est pas à suivre si vous disposez déjà d'un GPU suffisamment puissant.

Une fois le serveur créé et le GPU réservé, il est possible de concevoir un environnement de programmation adapté aux besoins de l'étude, comme cela est décrit dans le fichier */ARCHIVE_CD/ENVIRONNEMENT/InstallEnvironment.odt*. Afin de disposer de l'ensemble des bibliothèques utilisées pour la manipulation des produits et le développement des modèles, nous avons utilisé la base système proposée par Howard *et al.* (2018) dans le cadre du projet *fast.ia*. A noter que Jeremy Howard et Rachel Thomas sont également les auteurs du MOOC "*Practical Deep Learning For Coders*". Python 3.6 correspond à la version logicielle utilisée, les modules suivants ayant été mobilisés pour aboutir aux résultats présentés dans ce mémoire : collections, cv2, gdal, math, matplotlib, numpy, ogr, os, osr, pandas, random, skimage, sklearn, subprocess, torch et torchvision. L'ensemble est mis à disposition sous la forme

Annexe C. Mise en place de l'environnement de programmation

d'un environnement Conda, au chemin suivant dans l'archive fournie avec ce mémoire : */ARCHIVE_CD/ENVIRONNEMENT/env_colorize.txt*. Différents scripts et notebooks utilisant ces bibliothèques, il est conseillé de suivre la procédure d'installation pour pouvoir réaliser les traitements.

Le programme permettant de réaliser les colorisations a été écrit avec un GPU à l'esprit, mais fonctionne aussi sur un CPU contre un temps de calcul plus long. En effet, avec une carte NVIDIA, le traitement d'un cliché de 1024×1024 pixels se fait en moins de 1s, contre 9s pour un CPU 8 cœurs d'une fréquence de 3.1GHz. Dans les deux cas, un système avec environ 10Go de RAM (CPU) ou VRAM (GPU) est nécessaire pour pouvoir faire tourner le programme sur cette même taille d'image. Il est possible de traiter des clichés de dimensions inférieures à 1024×1024 pixels, mais cela se fait généralement au prix des performances de colorisation.

A noter enfin que n'importe quel GPU devrait être en mesure de faire fonctionner le programme, à condition qu'il dispose de suffisamment de mémoire vidéo, à nouveau 10Go ou plus. Les temps de calcul varient cependant largement d'un modèle à l'autre, compte-tenu du nombre de cœurs CUDA embarqués sur la carte. En effet, ce sont les unités de calcul qui permettent de paralléliser le traitement. Plus elles sont nombreuses, mieux c'est. A moyen et long termes, la location d'un serveur est cependant déconseillée, dans la mesure où la plupart des GPUs proposés sont finalement peu performants par rapport à d'autres disponibles sur le marché. A titre d'exemple, environ 300€ de crédits (en majeure partie fournis par Google Cloud) ont été dépensés pour ce travail. Pour environ 500€, il est possible d'acheter une carte NVIDIA GTX 1080 Ti, qui ressort donc comme un investissement plus intéressant, à la fois en termes de performance et de coût monétaire. Il reste cependant nécessaire de prendre garde au prix des GPUs, qui varie très rapidement du fait du minage des cryptomonnaies principalement.

D Présentation des tests de colorisation réalisés

Avant de retenir les cGAN, BEGAN et DRAGAN présentés dans le cadre de ce travail, une série de tests a été menée avec une petite base d'images. L'objectif était ici d'évaluer la capacité de chaque méthode à produire des résultats corrects, mais aussi à tenir compte de la nature multi-modale du problème de colorisation (Annexe E).

La Table D.1 rend compte des expérimentations effectuées. La majeure partie des combinaisons possibles a été réalisée, mais pas la totalité. En effet, certains tests n'étaient tout simplement pas concluants et n'auraient pas donné de meilleurs résultats autrement.

Variables	Valeurs testées
Espace colorimétrique	RGB, Lab
Espace des valeurs	Continu, discret
Orientation	Classification {CNN, PixelCNN, ColorNet} Régression {CNN} Génération {cGAN, BEGAN, DRAGAN, VAE-GAN, VAE, AE, Pix2Pix, WGAN, autres} Transfert {Classique, Apprentissage profond} Scribble {Classique}
Régularisation*	Label smoothing {Oui, Non} Fonction objectif supplémentaire {Oui {L1, L2, MSE, Gram, SSIM, MSSIM}, Non} Normalisation {Batch {Oui, Non}, Spectrale {Oui, Non}} Dropout {Oui, Non} Bruit {Oui, Non} Weight decay {Oui, Non}
Optimisation*	Générateur {Adam, SGD} Discriminateur{Adam, SGD}
Architecture*	Plus ou moins de couches Plus ou moins d'attributs appris Méthodes de rééchantillonnage {Oui {Paramétrique, Non paramétrique}, Non} Différentes fonctions d'activation intermédiaires {ReLU, PRELU, ELU, autres} Type {U-Net, ResNet, autres}

TABLE D.1 – Description des différents tests réalisés dans le cadre du travail de recherche. Les variables suivies d'une étoile * sont valables uniquement pour les GANs et leurs dérivés.

Quelques exemples de visualisations sont également disponibles sur la Figure D.1.

Annexe D. Présentation des tests de colorisation réalisés

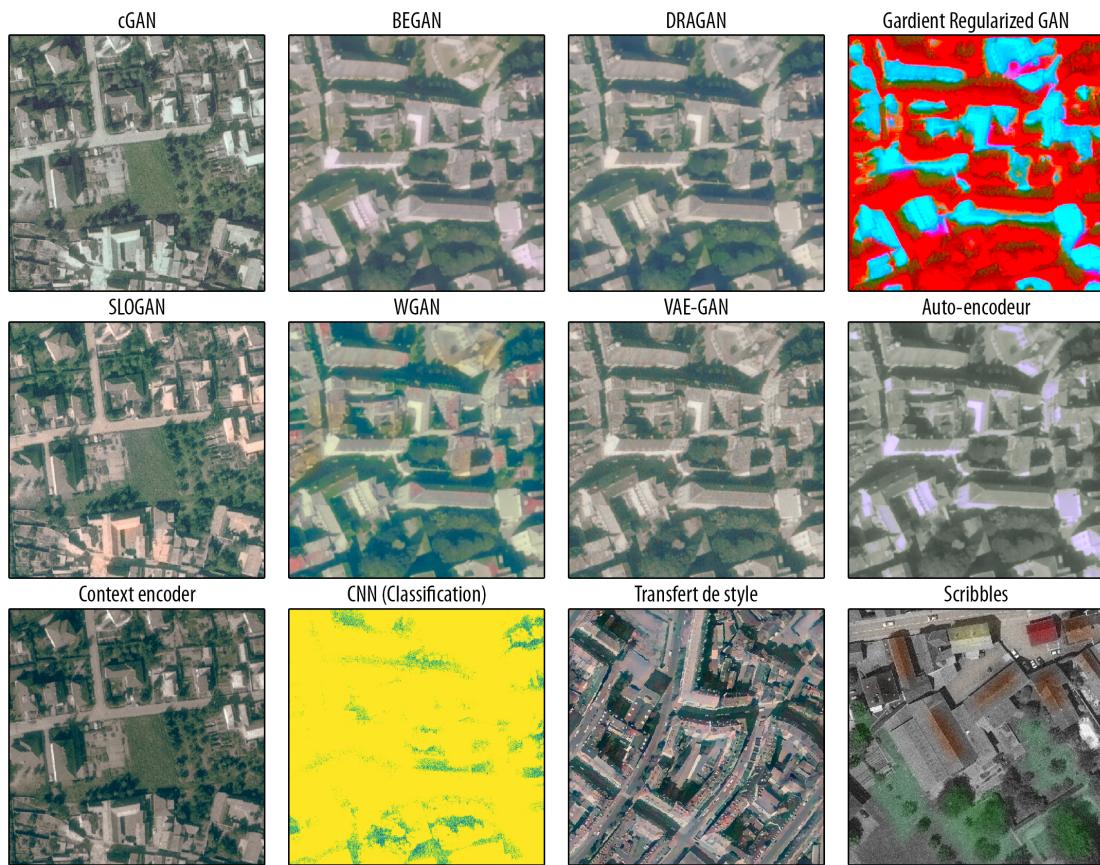


FIGURE D.1 – Visualisation des résultats de colorisation obtenus pour différents paradigmes, algorithmes, modèles et architectures. Ces représentations constituent un extrait des différents tests qui ont été réalisés, pour un apprentissage dont le nombre d’itérations est compris en 25 et 50, et sur une base d’entraînement modeste. Seuls les *scribbles* correspondent ici à une méthode de colorisation classique, n’impliquant pas d’apprentissage ou de GPU.

D'une façon générale, les différentes catégories de modèles et paradigmes ont apporté des colorisations très hétérogènes. Les CNNs donnent de très bons résultats sur le jeu d'entraînement, mais parviennent difficilement à généraliser, en particulier lorsqu'une fonction objectif de classification est utilisée. Il serait possible d'obtenir de meilleures colorisations qu'avec la méthode finalement retenue, à condition de disposer d'une photothèque exhaustive, sur laquelle l'ensemble du territoire de l'EMS est représenté par exemple. Les modèles génératifs donnent tous des résultats assez proches les uns des autres au début de l'entraînement, avec un apprentissage très rapide des sémantiques les plus discriminantes. Ils se distinguent cependant après suffisamment d'itérations, avec certains algorithmes qui parviendront, ou non, à apprendre les modes de la distribution. D'autres, comme le SLOGAN et le WGAN, sont efficaces au départ, puis finissent par donner des produits complètement incohérents après quelques dizaines de passes. En dernier lieu, des méthodes plus classiques ont aussi été testées. La première, basée sur le transfert, a fourni des résultats intéressants d'un point de vue colorimétrique, mais peu pertinents car incohérents sur le plan spatial. Les méthodes basées

sur les échantillons ne parviennent quant à elles pas à travailler sur les orthophotographies historiques. Des tests réalisés sur des images (non géographiques) ont pourtant fonctionné, avec un succès modéré cependant. Nous pouvons ainsi supposer que le bruit présent sur les clichés historiques limite leur utilisation.

La transition entre les espaces colorimétriques RGB et Lab améliore quant à elle très largement la qualité des produits en sortie. En effet, le modèle ne s'occupe de prédire qu'un champ de chrominance pour les canaux a et b . La concaténation d'une bande (pseudo-)panchromatique P ou L , les deux notations étant ici interchangeables, permet ensuite d'affiner les produits. En effet, elle donne l'information nécessaire pour restituer correctement les contours et la forme des objets. Cette même bande renseigne aussi sur la luminosité des surfaces, et permet d'aboutir à une information colorimétrique pertinente pour l'œil humain.

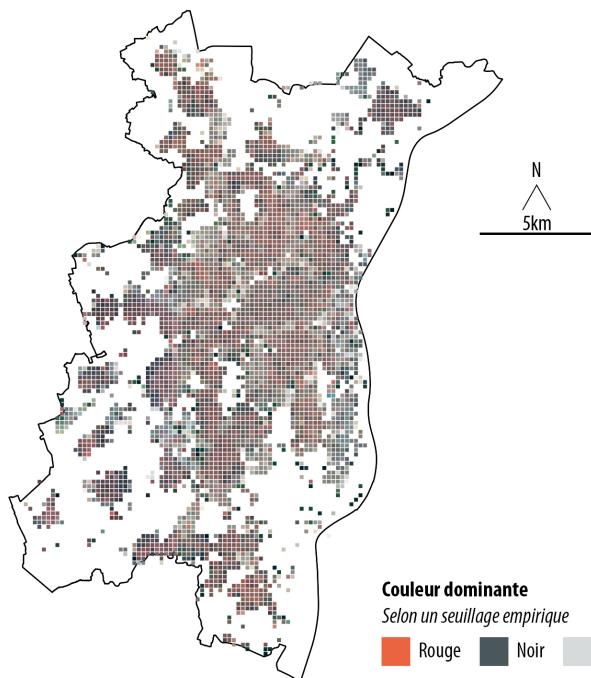
Les modifications apportées aux autres variables de la Table D.1 n'ont pas été particulièrement remarquables. Le fait de travailler avec un espace continu ou discret, outre les modèles pour lesquels l'un ou l'autre est indispensable, n'a pas permis une meilleure prédiction des valeurs de chrominance. Certaines méthodes de régularisation, comme le bruit ou le *dropout*, ont eu tendance à détériorer la qualité des résultats en sortie, bien qu'un temps d'apprentissage plus long aurait certainement permis d'annuler voire d'inverser cet effet. Un point qui pourrait permettre d'optimiser le modèle concerne son architecture, en particulier le nombre de couches utilisées. En effet, selon les tests menés, réduire de moitié leur quantité n'engendre pas une dégradation notable des sorties. Ainsi, revoir l'architecture pourrait au moins accélérer l'apprentissage, dans la mesure où il y aurait moins de poids à apprendre.

A noter que l'ensemble de ces tests a été rendu possible grâce aux individus et groupes de chercheurs-euses qui acceptent de distribuer leurs algorithmes sur des plateformes, telles GitHub ou GitLab. Ces travaux ont servi de référence et/ou d'inspiration pour la rédaction d'une partie importante du code auquel ce mémoire est adossé. Une liste exhaustive des dépôts et billets consultés est disponible dans l'archive distribuée avec le présent document, au chemin suivant : [/ARCHIVE_CD/LIENS/References.html](#).

E La colorisation comme problème multi-modal

La raison pour laquelle la colorisation est un problème difficile résulte de sa nature multi-modale, le fait qu'un même objet puisse revêtir différentes couleurs. C'est notamment le cas en milieu urbain, où les toitures possèdent des tuiles oranges, rouges, bleues ou noires par exemple, compte-tenu des matériaux qui les composent. En d'autres termes, une même sémantique peut être associée à différentes valeurs de chrominance dans un espace colorimétrique donné.

Composition colorée des valeurs médianes de réflectance
Selon l'ortophotographie de 2013, dans une grille de 200m de côté



Couleurs de toitures dominantes selon la réflectance
Selon l'ortophotographie de 2013, dans une grille de 200m de côté

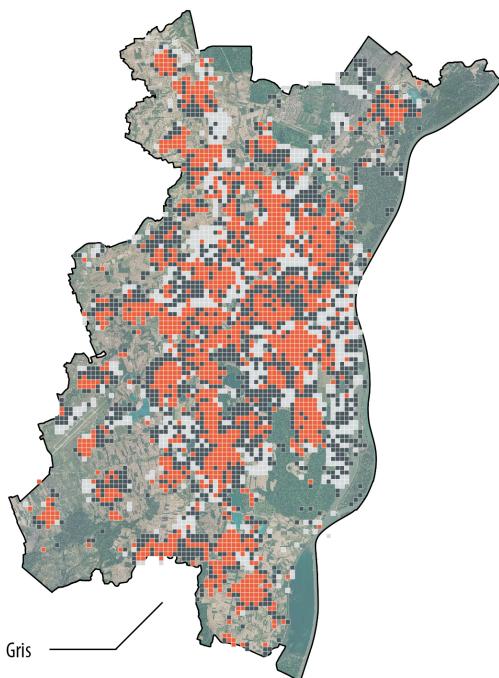


FIGURE E.1 – Couleurs médianes et dominantes des toitures dans l'agglomération strasbourgeoise. Le contraste de la composition colorée a été augmenté de sorte à pouvoir mieux discriminer les différentes couleurs.

Annexe E. La colorisation comme problème multi-modal

A titre de visualisation seulement, la Figure E.1 met en évidence les trois couleurs de toitures principalement rencontrées dans l'agglomération. La question des toits était particulièrement importante pour nous, car il existe une relation entre la couleur de leurs tuiles et leur localisation. Le centre-ville et les centre-bourgs, en tous cas pour Strasbourg, présentent presque exclusivement des toitures orangées. Les espaces marqués par une artificialisation récente, datant de la deuxième moitié du XX^e siècle, sont quant à eux caractérisés par un mélange de noir, de gris et d'orange principalement.

Ayant choisi d'utiliser les GANs, l'effet de *mode collapse* souvent rencontré apparaissait ainsi comme un frein à la mise en œuvre de colorisations plausibles. En effet, l'objectif était de disposer de couples sémantique — chrominance cohérents, faisant donc intervenir cette dimension spatiale sous-jacente. Les résultats obtenus avec le DRAGAN global sont certes prometteurs. Cependant, les approches basées sur un CNN, et utilisant une fonction objectif de classification, parvenaient à restituer des colorisations de qualité supérieure, mais seulement pour le jeu d'entraînement. Avec une base d'images plus exhaustive, il serait donc possible d'obtenir de meilleurs résultats, grâce à un corpus de sémantiques plus large, qui profiterait donc aux GANs, mais aussi aux CNNs qui seraient alors en mesure de mieux généraliser leur apprentissage.

RÉSUMÉ — La bancarisation des données géographiques constitue aujourd’hui un enjeu majeur pour les collectivités territoriales. Dans ce cadre, la Zone Atelier Environnementale Urbaine a initié le développement d’un SIG historique contenant une série temporelle de photographies aériennes anciennes, et décrivant le territoire de l’Eurométropole de Strasbourg pour huit dates entre 1932 et 2013. Pour les plus anciennes, seuls les clichés panchromatiques sont disponibles, limitant donc leur usage. De ce fait, il a été proposé de les coloriser afin de disposer de nouveaux attributs radiométriques, plus facilement mobilisables dans les chaînes de traitements géographiques classiques. Après avoir réalisé un catalogue des méthodes de colorisation actuellement disponibles, les réseaux génératifs antagonistes (GANs), développés dans le domaine de l’apprentissage profond, ont été retenus pour leur flexibilité et capacité d’automatisation. A l’issue d’une comparaison de différents algorithmes, le DRAGAN conditionnel, un dérivé du GAN classique, a été sélectionné pour entraîner un modèle à partir d’un prototype de photothèque géographique. Celle-ci a été constituée à partir de couples de produits panchromatiques et couleurs provenant de différentes sources, aériennes principalement. Les résultats obtenus grâce à cette technique sont corrects, malgré des erreurs de colorisation qui apparaissent localement. Après une validation qualitative et quantitative, il en ressort cependant qu’il est nécessaire de compléter la photothèque avec de nouveaux clichés, et d’entraîner le modèle plus longuement, afin d’obtenir de meilleurs produits. L’absence de métriques réellement adaptées à l’évaluation des résultats apparaît également comme une limite, à laquelle nous proposons différentes solutions.

Mots-clés : colorisation, photographie aérienne, apprentissage profond, DRAGAN.

ABSTRACT — Warehousing of geographic data has become a major challenge for land management. The *Zone Atelier Environnementale Urbaine* has created its own database in order to store a time series of aerial photographs. Taken between 1932 and 2013, some of these products were captured using a panchromatic imaging system, thus limiting their use. This thesis presents a method created for colorizing such images, allowing the community to access more radiometric attributes, leveraged in traditional geographic workflows. After assessing the relevance of available techniques, Generative Adversarial Networks (GANs) appeared to be efficient enough for this task. Having compared various implementations of this algorithm, we investigated conditional DRAGAN more specifically, and trained a model using a handmade geographic picture library. This structure is made of pairs of grayscale and colour images, captured by various platforms, aerial for the most part. Even though the results are plausible, the reconstructed color images show unrealistic chroma locally. Both qualitative and quantitative evaluations indicated the model would probably deliver better reconstructions after a longer training time, and using a more extensive photo library. The lack of proper metrics for assessing the quality of generated colorizations also appeared as a cap, that the scientific community could overcome by following various instructions proposed in this thesis.

Keywords : colorization, aerial photograph, deep learning, DRAGAN.