

Entraîneur sportif fondé sur l'intelligence artificielle

Rapport final

Projet industriel

Younes BELKADA, Jérôme BONACCHI,
Abdallah DAABOUL, Najwa MOURS LI & Arthur ZUCKER

POLYTECH SORBONNE
Spécialité
mathématiques appliquées et informatique numérique
Année 4
2019 – 2020



L'ensemble de l'équipe remercie
Hacène OUZIA pour son rôle de coordinateur,
Catherine ACHARD pour son encadrement et son aide précieuse,
Mégane MILLAN pour ses conseils,
Tarik LARJA et Marc DURAND pour leurs enseignements sur le management et la gestion de projet.

Table des matières

Table des figures	v
Liste des tableaux	vii
1 Introduction	1
1.1 L'importance de l'intelligence artificielle	1
1.2 Brève introduction au <i>machine learning</i>	1
1.3 Définitions, sigles et acronymes	2
2 Présentation du sujet	3
2.1 Contexte	3
2.2 Projet initial	3
2.2.1 Personnes impliquées	3
2.2.1.1 Membre de l'équipe	4
2.2.1.2 Représentant de Polytech Sorbonne	4
2.2.1.3 Représentant académique	4
2.2.2 Besoins	5
2.2.3 Objectifs du projet	5
2.2.4 Contraintes	5
2.3 Projet remodelé	6
2.3.1 Équipement	6
2.3.1.1 Langages de programmation	6
2.3.1.2 Modules et bibliothèques logicielles	7
2.3.1.3 Environnements de travail	7
2.3.1.4 Outils	7
2.3.1.5 Ressources	7
2.3.2 Études déjà effectuées	7
2.3.3 Plan de réalisation	8
3 Réalisation	11
3.1 Tâches réalisées en amont	11
3.1.1 Essais d'OpenPose	11
3.1.2 Création de notre base de données	11
3.1.3 Traitement des vidéos de la base de données	11
3.1.4 Découpage et annotation des vidéos	13
3.1.5 Conversion vidéo – tenseur	13

3.1.6	Création des réseaux de neurones	13
3.2	Réalisation du <i>CNN</i>	14
3.2.1	Première approche	15
3.2.2	Seconde approche	15
3.2.3	Architecture finale	15
3.3	Réalisation du <i>LSTM</i>	16
3.4	Réalisation de la combinaison d'un <i>CNN</i> et d'un <i>LSTM</i>	18
3.5	Autre solution : le <i>SVR</i>	18
3.5.1	<i>SVR</i>	18
3.5.2	Présentation du code	19
4	Résultats	21
4.1	Résultats du <i>CNN</i>	21
4.1.1	Première approche	21
4.1.2	Seconde approche en utilisant l'architecture finale	21
4.2	Résultats du <i>LSTM</i>	23
4.3	Résultats de la combinaison d'un <i>CNN</i> et d'un <i>LSTM</i>	23
4.4	Résultats du <i>SVR</i>	24
5	Conclusion	25
5.1	Évaluation des exigences	25
5.2	Perspectives d'évolution	25
5.3	Obstacles et difficultés rencontrés	27
5.4	Bilans personnels	29
5.4.1	Bilan de Younes BELKADA	29
5.4.2	Bilan de Jérôme BONACCHI	29
5.4.3	Bilan d'Abdallah DAABOUL	30
5.4.4	Bilan de Najwa MOURSЛИ	30
5.4.5	Bilan d'Arthur ZUCKER	31
A	Extraits du cahier des charges	33
A.1	Expression fonctionnelle du besoin	33
A.1.1	Fonctions et contraintes	33
A.1.2	Évaluation des exigences	34
A.2	Analyse des risques	34
	Glossaire	37
	Sigles et acronymes	39
	Bibliographie	41

Table des figures

2.1	Organigramme	4
2.2	Architecture utilisée par LI, HUANG, CAI et al. [17]	8
2.3	Architecture utilisée par LI, HUANG, CAI et al. [17] (suite)	8
2.4	<i>Work Breakdown Structure</i> (succinct)	9
2.5	<i>Work Breakdown Structure</i> (détailé)	9
3.1	Articulations d'OpenPose	12
3.2	Les gestes élémentaires qui constituent un exercice	12
3.3	Histogramme du nombre de <i>frames</i> par exercice	13
3.4	Schéma explicatif d'une couche Conv1D	14
3.5	Schéma explicatif d'une couche Conv2D	14
3.6	Structure des tenseurs utilisés	15
3.7	Modèle final retenu pour l'apprentissage	17
3.8	Schéma du <i>support vector machine (SVM)</i>	19
3.9	schéma du <i>support vector regression (SVR)</i>	20
4.1	Résultat du sans les <i>squats</i>	22
4.2	Résultat du <i>CNN</i> avec les <i>squats</i>	22
4.3	Exemple de résultat du <i>LSTM</i>	23
4.4	Résultat de la combinaison d'un <i>CNN</i> et d'un <i>LSTM</i>	24
4.5	Mauvaises détections du <i>SVR</i>	24
5.1	Invariants utiles à la normalisation de la morphologie humaine	26
A.1	Bête à corne	33
A.2	Diagramme pieuvre	34

Liste des tableaux

5.1	Registre des risques rencontrés	27
5.2	Niveaux de criticité des risques	28
A.1	Registre des risques du cahier des charges	35
A.2	Matrice de probabilité d'impact	35

1. Introduction

1.1 L'importance de l'intelligence artificielle

Depuis l'émergence de la robotique et de l'informatique, l'Homme essaye d'injecter ses notions d'intelligence dans des machines. Il tente également, et de plus en plus, d'introduire en elles ses propres capacités, ses propres mécanismes. Étant conçue et fabriquée par l'homme, on qualifie cette forme d'intelligence comme l'intelligence artificielle. Ce domaine de connaissances est sûrement l'un des plus importants dans l'histoire moderne de l'humanité et il va continuer de révolutionner notre façon de vivre.

L'intelligence artificielle (I.A.) est souvent définie comme la science de créer et de programmer des systèmes capables d'accomplir des tâches qui nécessitent de l'intelligence lorsqu'elles doivent être faites par un être humain : que ce soit faire des mathématiques, jouer aux échecs ou même parler. Depuis les débuts de l'I.A. dans les années 1950, ce domaine de recherche s'est enrichi avec des nombre incalculables de recherches et d'expériences. Aujourd'hui, il existe des ordinateurs capables de résoudre des équations et des problèmes en quelques dixièmes de secondes auxquels l'Homme était incapable de résoudre il y a 100 ans. Un exemple d'application de l'I.A. pour résoudre un problème considéré comme extrêmement complexe est l'utilisation d'une variante de la théorie de la fonctionnelle de la densité complétée par du *deep learning* pour reproduire des propriétés thermodynamique de l'eau afin de résoudre à posteriori l'équation de SCHRÖDINGER. D'après l'article de SACCO [1], « il a été possible de retrouver certaines des propriétés de l'eau uniquement à partir de l'application de l'équation de SCHRÖDINGER à des molécules d'eau ». De plus, il y est mentionné qu'il existe une possibilité d'améliorer cette application pour élargir le champs d'étude.

« Le professeur Krzysztof SZALEWICZ pense que leur modèle ne s'applique pas seulement à l'eau sous forme liquide, mais devrait permettre aussi de mieux comprendre celle-ci dans différentes phases, comme par exemple sous ses multiples formes de glace. Plus généralement, les techniques qu'ils ont mises au point devraient faire progresser les simulations sur ordinateurs d'autres liquides et systèmes moléculaires, les protéines ou l'A.D.N. en biologie par exemple. » [1]

De nos jours, l'I.A. s'est grandement développée et il existe des ordinateurs capables de mener une discussion avec un humain, capables de répondre à des questions comme un humain ; il y a même des ordinateurs capables de reconnaître son propriétaire juste par la voix. L'une des conséquences sur notre vie quotidienne est la pratique d'activités telles que le sport, dans lequel l'I.A. s'immisce afin d'améliorer notre confort, nos aptitudes, notre attention, etc. Ce projet illustre cette application de intelligence artificielle au service de l'Homme.

1.2 Brève introduction au *machine learning*

Pour comprendre pourquoi nous utilisons le *machine learning*, il nous paraît nécessaire d'expliquer brièvement ce qu'il représente. Le lecteur ayant déjà connaissance de ce sujet peut passer directement à la partie 2.

Le *machine learning* ou apprentissage automatique est une branche de l'intelligence artificielle cherchant à donner à une machine la capacité d'apprendre à partir de données, d'apprendre à résoudre des tâches sans être

initialement programmé pour. Afin d'apprendre, la machine nécessite un très grand nombre de données que l'on appelle les données d'entraînement (ou *training set* en anglais). C'est pourquoi le *machine learning* fait aussi appel au *big data*. Concrètement, cela permet d'effectuer des prédictions sur une nouvelle donnée qui serait fournie à la machine.

Au sein même du *machine learning*, il existe trois branches distinctes [2] :

- Le *supervised machine learning* ou apprentissage supervisé (par l'Homme) est utilisé lorsque l'on connaît les données de sortie de notre prédiction. On entraîne alors la machine à partir d'entrées et de données connues, qui vont, à l'aide d'algorithmes de classification ou de régression, créer un modèle de prédiction capable de prédire la sortie correspondant à une nouvelle donnée.
- Le *unsupervised machine learning* ou apprentissage non-supervisé utilise principalement des méthodes de *clustering* qui regroupent des données, n'ayant *a priori* aucun liens, en différents *clusters*, ou groupement de données. Cela permet de découvrir un schéma caché reliant les données.
- L'apprentissage par renforcement consiste à étudier l'environnement dans lequel se trouve un agent et apprendre à ce dernier comment choisir les meilleures décisions en fonction du contexte. La base du modèle d'apprentissage par renforcement consiste en :
 - un ensemble d'états de l'agent dans l'environnement;
 - un ensemble d'actions que l'agent peut effectuer;
 - un ensemble de récompenses que l'agent peut obtenir.

Le but de cet apprentissage est alors de maximiser les récompenses.

1.3 Définitions, sigles et acronymes

Avant même de commencer l'explication de ce projet, il est important d'introduire la définition des termes que nous allons utilisés ainsi que la signification des sigles et des acronymes que nous employons. Nous utilisons assez souvent des termes anglophones car ceux-ci sont plus usités que leur équivalent français, c'est pourquoi nous sommes plus familiers avec ceux-ci. Toutefois, nous avons essayé de faire figurer dans le glossaire chacun de ces termes avec leur traduction. Cette liste de définitions, de sigles et d'acronymes est présente à la fin de ce document (cf. page 37).

2. Présentation du sujet

2.1 Contexte

De nos jours, les gymnases sont de plus en plus populaires et l'industrie du *fitness* est en plein essor. En 2018, le chiffre d'affaires de l'industrie américaine des salles de sport a atteint 32,3 milliards de dollars, contre 30 milliards en 2017, soit une croissance de 7,8 % [3]. Il en va de même pour les pays en développement comme le Vietnam : l'industrie du *fitness* devrait croître annuellement de 19,9 % pour atteindre une valeur de 112,7 millions de dollars en 2020 [4]. Par ailleurs, le printemps de l'industrie du *fitness* au Vietnam est assez récent : le marché présente donc beaucoup d'opportunités.

Une salle de sport offre à la fois un lieu et des outils pour s'entraîner, ainsi qu'un service : le *coaching* et la formation des clients. Si l'on exclut le coût des équipements requis dans un gymnase, la principale source de dépenses dans les centres de *fitness* provient des employés. En effet, le secteur sportif privé (clubs de sport, gestion d'installation sportives, enseignements du sport, etc.) employait 124 286 salariés en 2015 selon le Ministère de l'Économie et des Finances français. Les entraîneurs et les formateurs sont difficiles à recruter et leur salaire est donc élevé. Ainsi, leur faible disponibilité les rend précieux pour tout club de sport. Afin de réduire les dépenses et améliorer le service à la clientèle, les salles de sport pourraient utiliser un entraîneur artificiel qui serait disponible en tout temps. Il corrigera automatiquement et aiderait tout apprenant qui l'utilisera sans avoir besoin d'un contact direct. Cela permettrait d'améliorer le service à la clientèle en augmentant la disponibilité des entraîneurs sportifs.

Les domaines d'application de l'intelligence artificielle se sont grandement étendus et le sport n'est pas une exception. Une intelligence artificielle (I.A.) précise jouant le rôle d'un entraîneur permettrait de meilleurs retours sur tous types de mouvements, faciliterait la correction des positions, préciseraient les défauts et le tout à moindre coût. Les juges de compétitions comme ceux des jeux olympiques (J.O.) pourraient être assistés permettant ainsi une notation dont l'impartialité serait garantie. Aujourd'hui, les entraîneurs sportifs artificiels se fondant sur des données vidéos ne sont utilisés que dans des domaines de recherche. Toutefois, bien que les versions grand public ne soient encore que des ébauches par rapport à la demande, le marché est prêt à les accueillir comme le montre cet entraîneur artificiel auditif qui a soulevé près de 1,6 millions de dollars [5].

2.2 Projet initial

2.2.1 Personnes impliquées

Ce projet fait parti de l'enseignement nommé « Projet industriel » donné aux élèves d'année 4 de la spécialité mathématiques appliquées et informatique numérique à l'école polytechnique universitaire Polytech Sorbonne¹. Ce projet est supervisé par Polytech Sorbonne et l'institut des systèmes intelligents et de robotique (ISIR)² par

1. cf. www.polytech.sorbonne-universite.fr

2. cf. www.isir.upmc.fr

un représentant de chaque partie. Il l'était également initialement par un représentant de Chainos. L'ISIR est un laboratoire de recherche multidisciplinaire qui appartient à Sorbonne Université³ et au centre national de la recherche scientifique (C.N.R.S.)⁴.

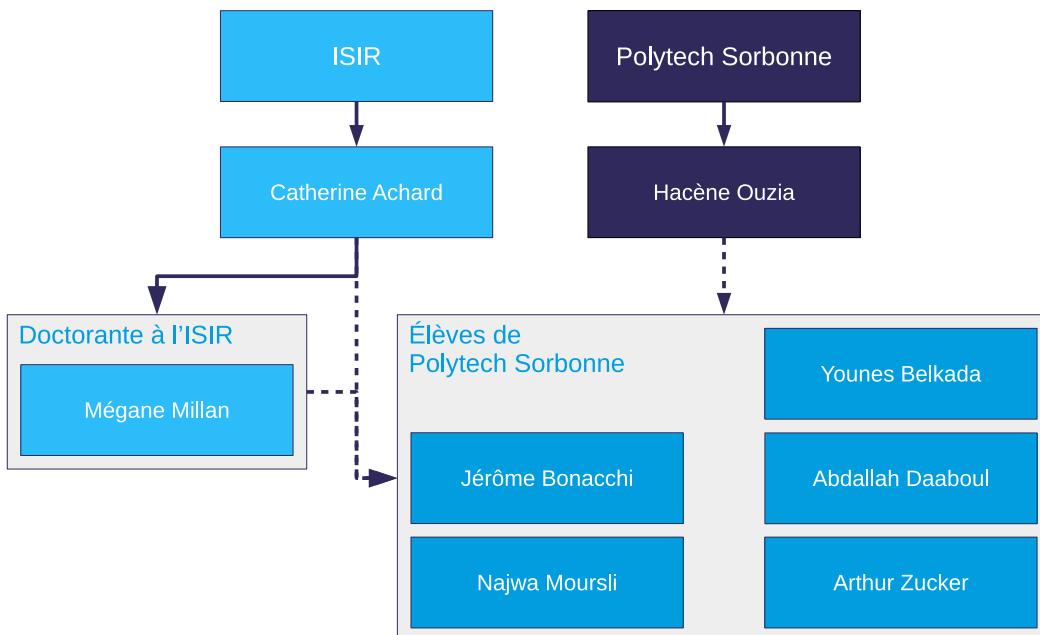


FIGURE 2.1 – Organigramme

2.2.1.1 Membre de l'équipe

- Younes BELKADA
- Jérôme BONACCHI
- Abdallah DAABOUL
- Najwa MOURS LI
- Arthur ZUCKER

Arthur ZUCKER et Younes BELKADA ont un accompli un stage à Chainos, ont déjà travaillé dans le domaine de l'intelligence artificielle et ont un bonne maîtrise de Python. Jérôme BONACCHI et Abdallah DAABOUL ont également tous les deux une bonne connaissance de Python et ont déjà suivis des cours sur Internet à propos du *deep learning*. Najwa MOURS LI a des connaissances médicales qui seront utiles afin de prendre en compte la variabilité du corps humain dans les algorithmes ; elle suit également des cours de *deep learning*.

2.2.1.2 Représentant de Polytech Sorbonne

Le docteur Hacène OUZIA est le coordinateur du projet. Ses compétences en mathématiques et en informatique pourraient nous être utiles si nous rencontrions des difficultés.

2.2.1.3 Représentant académique

Le docteur Catherine ACHARD qui travaille à l'ISIR à Sorbonne Université. Elle est notre mentor du fait du fait de sa formation et de ses connaissances en intelligence artificielle dans les domaines de *computer vision* et d'*image processing*.

3. cf. www.sorbonne-universite.fr

4. cf. www.cnrs.fr

2.2.2 Besoins

Les exercices de *fitness* sont très bénéfiques pour la santé et la forme physique des personnes. Cependant, ils peuvent également être inefficaces, potentiellement dangereux et peuvent causer de graves blessures aux muscles et aux ligaments s'ils sont mal exécutés. Des erreurs de mouvement durant des exercices surviennent lorsque un individu n'effectue pas correctement le bon geste. Cela pourrait être dû à un manque de formation. En outre, les gymnases peuvent ne pas disposer d'un nombre suffisant d'entraîneurs personnels pour satisfaire la demande en raison de leur coût élevé.

Comme expliqué précédemment, le marché du *fitness* est en plein essor, particulièrement au Vietnam. C'est pour l'ensemble de ces raisons que l'entreprise Chainos, spécialisée en cryptomonnaie ainsi que d'autres services informatiques, par le biais de Nam NGUYEN VAN, a proposé à notre équipe de leur fournir après sept mois de travail un système d'entraînement basé sur l'I.A. composé des éléments suivants.

1. **Un système de détection et de suivi des mouvements** basé sur des mécanismes qui font correspondre les informations de profondeur avec les couleurs et qui prennent en compte les informations spatiales et temporelles collectées par une caméra temps de vol. Par la suite, ce système sera formé et testé sur notre base de données, ce qui nous amènera à sélectionner un cadre d'apprentissage approfondi de l'I.A. et un réseau préentraîné approprié pour la détection et le suivi. Ensuite, ce réseau devra être validé en l'entraînant sur notre ensemble de données.
2. **Un système de notation** approprié qui doit être simple, efficace et le moins faussé possible grâce à la conception d'un algorithme de système de notation testé et validé sur des environnements de simulation dans un premier temps.
3. **Un système de feedback** basé sur notre système de notation afin de donner les meilleures recommandations aux apprenants dans le but d'améliorer leurs mouvements.
4. **Une application hybride** contenant l'ensemble des trois systèmes.

Ce système permettant à l'entreprise d'obtenir une ébauche gratuite d'une application qu'elle voudrait mettre sur le marché dans un futur proche. Le produit final qui devra être livré aux différents partenaires est un programme fonctionnel documenté capable de : détecter un humain réalisant un mouvement sportif, reconnaître sa pose, l'évaluer et la commenter. Cette note devra se fonder sur les performances d'un athlète ou d'un entraîneur et le commentaire devra servir à l'amélioration du mouvement.

2.2.3 Objectifs du projet

L'objectif de ce projet est d'aider les athlètes à perfectionner leur entraînement et de réduire les coûts des propriétaires de salles de sport grâce aux salaires des entraîneurs personnels. Nous cherchons à aider les gens à adopter la bonne posture pour les exercices, afin qu'ils ne nuisent pas à leur santé, en concevant une application qui détecte et note les poses faites par les utilisateurs et enfin retourne des commentaires. Cette application utilisera les dernières avancées en matière de vision par ordinateur (*computer vision*), de reconnaissance des formes (*pattern recognition*), et d'apprentissage profond (*deep learning*).

2.2.4 Contraintes

À travers ce projet, nous avons réalisé que de nombreuses contraintes étaient liés au domaine d'étude : l'intelligence artificielle. En effet, la moitié des projets menés dans le domaine de l'I.A. sont voués à l'échec selon le journal scientifique [6]. Là, est l'un des principaux enseignements tiré d'une étude réalisée début 2019 par *International Data Corporation* [7] auprès de quelque 2 500 organisations à travers le monde. Les décideurs consultés, Andrea MOGINI, *data scientist* senior au sein du cabinet de conseil Keyrus et Didier GAULTIER, consultant de Business & Decision, évoquent le plus souvent des attentes irréalistes et un manque de main-d'œuvre parmi les causes de ces revers. Les autres raisons invoquées sont le coût des solutions d'I.A. et les potentiels biais présents dans les jeux de données d'entraînement.

Voici, ci-dessous les contraintes auxquelles ce projet doit faire face.

« Pas de data, pas d'I.A. ». Voici la première et la plus grande contrainte de notre projet si la base de données est trop petite ou potentiellement inexiste, il est impossible de commencer le projet. Il faut que le jeu de donné soit labellisé, sinon cela prendra des mois et entraînera un retard considérable dans l'avancement de notre projet.

Disposer de données recouvrant l'ensemble des variables explicatives nécessaires. Si l'ensemble de nos données ne nous permet pas d'avoir une vision dynamique de la personne effectuant des mouvements (vitesse, morphologie, précision, etc.), les résultats de la modélisation seront forcément incorrects.

Disposer d'exemples d'apprentissage fondés sur l'expérience. Dans le même temps, ces exemples sont indispensables et doivent être présents en suffisamment grand nombre pour parvenir à une I.A. généralisable et applicable sur le terrain.

Garantir la confidentialité des données. Sachant que les données privées seront utilisées à des fins de formation, de test et d'évaluation, nous devons respecter le réglementation générale sur la protection des données (R.G.P.D.)⁵ sur les politiques en matière de données et assurer la sécurité de notre programme.

Surmonter des difficultés pratiques. La détection de la pose humaine en trois dimensions, qui est la technique la plus fine pour ce projet, est difficile à mettre en œuvre car presque toutes les recherches dans ce domaine utilisent la détection et le suivi bi-dimensionnels.

Prendre en compte la variabilité du corps humain. Notre objectif serait de traiter différents types de corpulence en adaptant notre système de détection afin d'éviter l'erreur de détection de certains mouvements aléatoires. Par exemple, notre application attribuerait un score plus bas que prévu à l'apprenant, considérant qu'il a fait une erreur de comparaison avec les mouvements du formateur, alors que la seule différence est leur hauteur mais que les mouvements sont exactement les mêmes.

Faire face a des questions éthiques. Finalement, ce projet est au centre des débats sur l'I.A. en soulignant des questions éthiques. Ce projet est destiné à remplacer le travail d'un formateur qui pourrait être interprété comme l'incarnation des excès de l'I.A.

Afin de pallier les différentes contraintes, nous avons décidé d'utiliser la théorie des contraintes ou approche de chaîne critique. La chaîne critique propose de traiter les dérives comportementales les plus fréquentes. À cet effet, elle fournit des réponses pratiques à la surévaluation des marges de sécurité des temps attribuées à chaque tâche, au démarrage des travaux au dernier moment (ou syndrome de l'étudiant), ou encore à l'utilisation de toute la durée prévue d'une tâche même si le temps effectif de réalisation en est bien inférieur.

2.3 Projet remodelé

Le projet a été remodelé suite à la séparation avec Chainos qui n'a pas tenu son engagement de fournir la base de données. Les objectifs initiaux et les contraintes restent inchangés.

2.3.1 Équipement

Au cours de ce projet, nous utilisons différents outils et ressources à notre disposition.

2.3.1.1 Langages de programmation

Au cours de ce projet, nous avons utilisés ces langages de programmation :

- Matlab [8];
- Python.

5. cf. eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679

2.3.1.2 Modules et bibliothèques logicielles

Nous sommes amenés à utiliser les modules suivants :

- TensorFlow [9];
- Keras [10].

2.3.1.3 Environnements de travail

Nous utilisons ces environnements de travail :

- Google Drive⁶;
- Google Colaboratory⁷;
- GitHub⁸;
- Overleaf⁹;
- WhatsApp¹⁰.

2.3.1.4 Outils

Nous avons la capacité d'utiliser les outils suivants :

- les *graphics processing unit (GPU)* de Google;
- les caméras de l'ISIR;
- les ordinateurs personnels des membres de l'équipe ainsi que ceux de Polytech Sorbonne.

2.3.1.5 Ressources

Nous avons accès aux ressources suivantes :

- des bases de données libres de droits (par exemple, celle des J.O.);
- des *massive open online course (MOOC)* sur le *deep learning*;
- des articles scientifiques décrivant les méthodes de l'état de l'art.

2.3.2 Études déjà effectuées

L'ensemble de l'équipe chercha des articles scientifiques [11]-[16] en rapport avec notre projet. Madame Catherine ACHARD nous conseilla plusieurs articles [17]-[20] à propos de méthodes innovantes.

En regardant dans la littérature scientifique, nous avons remarqué que l'architecture utilisée par LI, HUANG, CAI et al. [17] (cf. figures 2.2 et 2.3) était intéressante. En effet, elle permettrait de donner une note globale et une note à chaque instant. Ainsi, nous décidâmes de nous en inspirer pour la suite de notre projet.

Son architecture se compose d'un module d'attention spatial (cf. partie bleue sur la figure 2.2 et figure 2.3) qui focalise le réseau de neurones sur une partie de l'image seulement. L'architecture contient ensuite un module d'agrégation temporelle censé modéliser la transition temporelle des actions. Dans l'article d'origine, le premier module prend en entrée un tenseur d'ordre 3 X_t à chaque instant. Nous souhaiterions à la place donner un tenseur contenant les positions des articulations de la personne à l'instant t . Pour cela, nous utiliserons OpenPose [11]-[14] qui, à partir d'une vidéo, détecte les positions des articulations de personnes à chaque instant. Après une longue recherche dans l'état de l'art, cette application s'avère être la plus adaptée à notre projet. Celle-ci semble avoir de très bons résultats sur le traitements de vidéos. Nous avons également envisagé l'utilisation d'AlphaPose

6. cf. www.google.com/drive/

7. cf. colab.research.google.com/

8. cf. github.com

9. cf. www.overleaf.com

10. cf. www.whatsapp.com

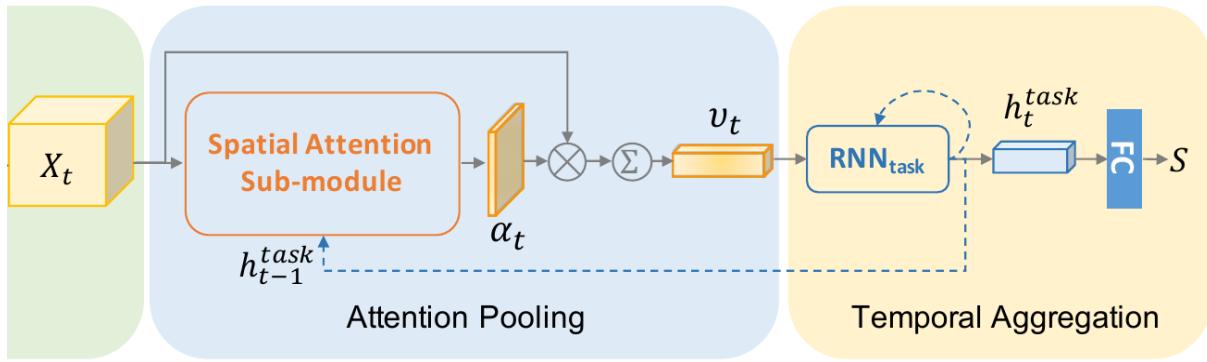


FIGURE 2.2 – Architecture utilisée par LI, HUANG, CAI et al. [17]

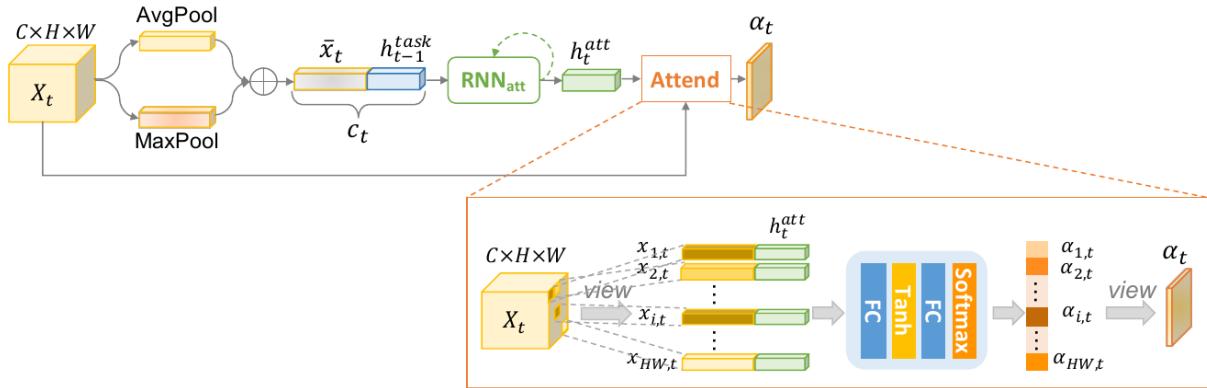


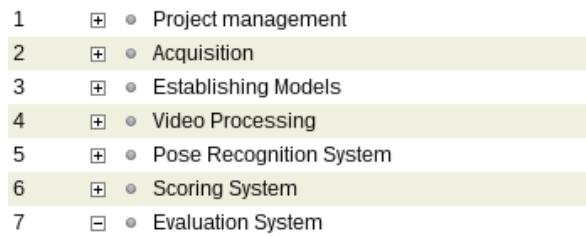
FIGURE 2.3 – Architecture utilisée par LI, HUANG, CAI et al. [17] (suite)

FANG, XIE, TAI et al. [21] (créé par Facebook) s'avérant cependant moins précis et non adapté à notre domaine d'application.

De plus, nous nous sommes intéressés au code utilisé par PIRSIAVASH, VONDRIK et TORRALBA [18] car il permettrait de donner une note globale et de connaître l'importance de chaque instant dans cette note. Par ailleurs, le code montré dans l'article est capable de donner un *feedback*. Une autre solution envisagée a été, toujours à partir d'OpenPose, d'utiliser une autre architecture de réseau de neurones artificiels inspirée de celle constituée par MILLAN et ACHARD [22]. en codant un *long short-term memory (LSTM)* et un *convolutional neural network (CNN)*.

2.3.3 Plan de réalisation

Nous avons fixé dès le début du projet et pendant plusieurs mois les étapes importantes à effectuer sous la forme d'un *WBS* présent sur les figures 2.4 et 2.5. Toutefois, nos avancées n'étant pas aussi ordonnées, nous avons changé notre méthode de travail. Notre façon de travailler s'est alors organisée de manière *agile*. Lors des réunions avec notre encadrante, nous définissions les tâches à faire dans la semaine et nous discutions sur les projections à moyen terme. Nous travaillions donc de manière itérative, avec des itérations qui durèrent entre 1 à 2 semaines.

FIGURE 2.4 – *Work Breakdown Structure* (succinct)FIGURE 2.5 – *Work Breakdown Structure* (détailé)

3. Réalisation

3.1 Tâches réalisées en amont

3.1.1 Essais d'OpenPose

Arthur ZUCKER installa OpenPose sur son ordinateur et l'essaya en utilisant premièrement des vidéos de test, puis sa *webcam*. Younes BELKADA utilisa OpenPose via *Google Colaboratory* sur des vidéos de test. Nous avons exécuté nos vidéos sur OpenPose par le biais de Google Colaboratory dû à une puissance de nos ordinateurs trop faible et à une installation des dépendances nécessaires difficile. En effet, Abdallah DAABOUL et Jérôme BONACCHI essayèrent d'installer OpenPose sur leurs ordinateurs respectifs sans succès. L'un des avantages majeurs de l'utilisation de Google Colaboratory est la puissance de calcul mise à notre disposition qui est bien supérieure à celle que nous aurions avec tous nos ordinateurs personnels réunis.

OpenPose renvoie pour chaque *frame*, un fichier *JSON* contenant pour chaque articulation ses coordonnées (x, y) ainsi que c le niveau de confiance associé donné par le réseau OpenPose vis à vis des coordonnées des articulations détectées. En vérité, comme montré sur la figure 3.1, les points renvoyés par OpenPose ne correspondent pas tous à des articulations. Cependant, par simplification, nous nous permettrons par la suite d'appeler ces points des articulations.

3.1.2 Création de notre base de données

Le but est de créer une base de données annotée arbitrairement par nous-même contenant des vidéos du même exercice composé de 9 gestes élémentaires. Dans un premier temps, Abdallah DAABOUL filma Najwa MOURS LI en train de réaliser les mouvements souhaités (cf. figure 3.2). Puis, Arthur ZUCKER monta la vidéo à partir de l'application Imovie de manière à ce qu'elle soit didactique . Jérôme BONACCHI utilisa OpenPose et AlphaPose [21] sur la vidéo « modèle » pour tester la précision des différentes applications à notre disposition. Avec l'aide de notre encadrante académique Madame Catherine ACHARD et de la doctorante qu'elle supervise Mégane MILLAN, nous réussîmes à réunir une vingtaine de volontaires. Deux caméras prêtées par l'ISIR furent utilisées, une de face et l'autre tournée de 45°. Le même exercice a été réalisé par 21 personnes différentes à 5 vitesses d'exécution différentes, et deux fois pour chaque vitesse, afin de créer de la variabilité. Nous obtenons finalement 210 vidéos dont la répartition du nombre de *frames* est présentée sur la figure 3.3.

3.1.3 Traitement des vidéos de la base de données

Younes BELKADA réalisa un programme Python sur Google Colaboratory afin de récupérer les sorties d'OpenPose sur les vidéos de la base de données *UNLV Olympic Scoring Dataset*, puis sur la notre par la suite. Nous souhaitons enregistrer les tenseurs utilisés dans notre base dans des fichiers *numpy*. Le principe de l'obtention de ces fichiers est le suivant. Nous récupérons les sorties d'OpenPose au format *JavaScript object notation (JSON)*,

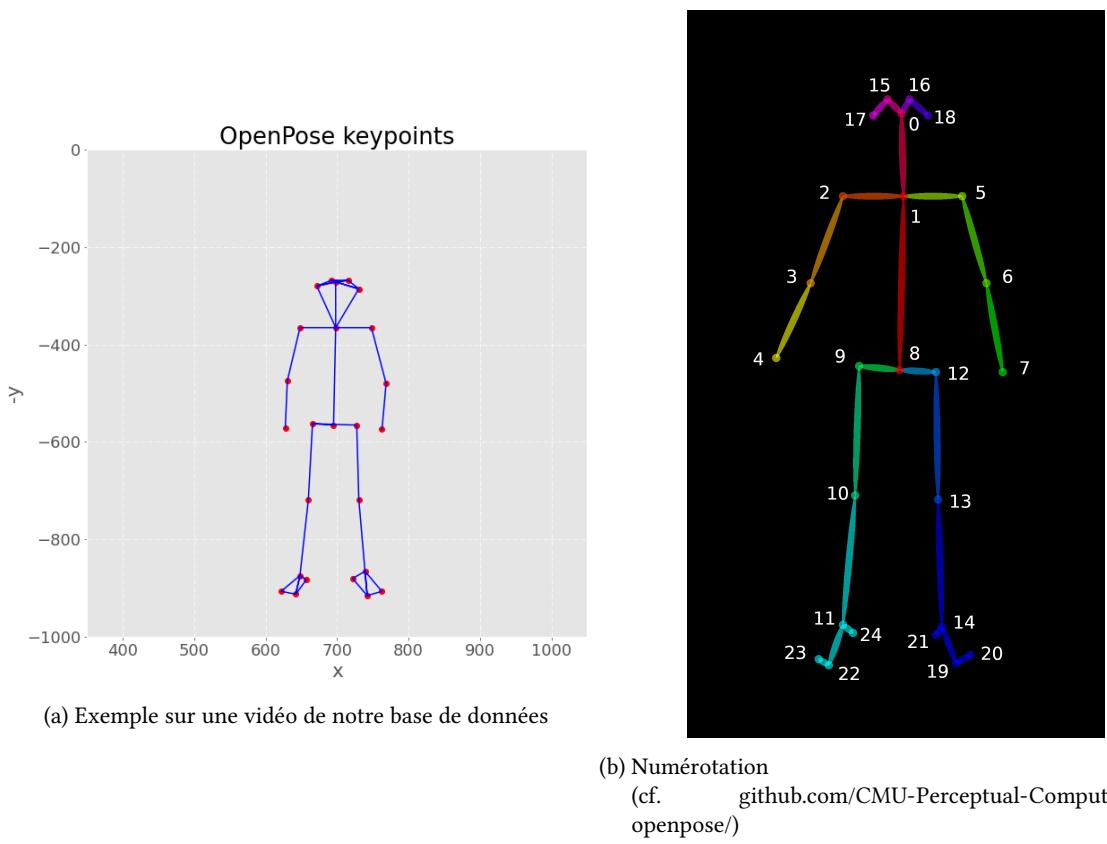


FIGURE 3.1 – Articulations d'OpenPose

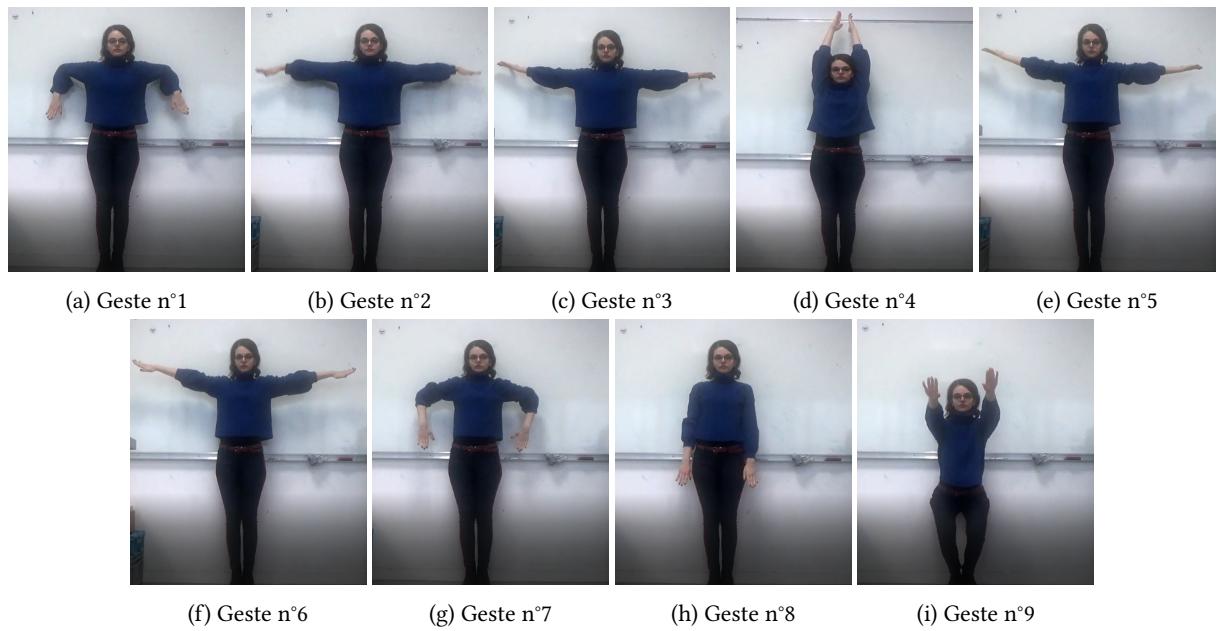


FIGURE 3.2 – Les gestes élémentaires qui constituent un exercice

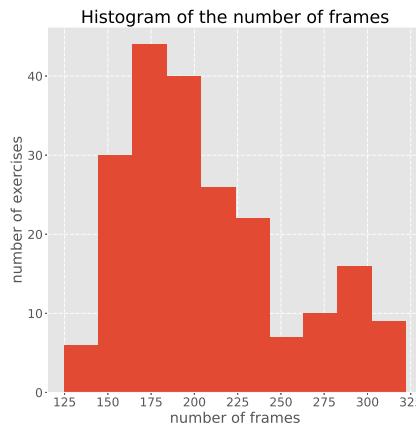


FIGURE 3.3 – Histogramme du nombre de *frames* par exercice

puis, grâce à un script Python, nous traitons ces fichiers un par un et nous regroupons toutes ces coordonnées dans un tenseur d'ordre 3. Ces tenseurs sont de taille $2 \times 25 \times n_{frame}$, avec n_{frame} le nombre de *frame* de la vidéo, et contiennent les coordonnées x et y des 25 points retournés par OpenPose pour chaque image de la vidéo.

3.1.4 Découpage et annotation des vidéos

Suite à la demande de Madame Catherine ACHARD, Jérôme BONACCHI se chargea de la création d'un programme Matlab [8] permettant la découpe et l'annotation des vidéos. Il a pour but d'être utilisé lors du traitement des vidéos de notre base de données. Un script a été créé prenant en entrée une vidéo d'un exercice, une matrice de 2 colonnes, pour noter le début et la fin de chaque geste, et N lignes, N est le nombre de gestes élémentaires qui composent un exercice. Il crée également un vecteur de taille N contenant les notes. Nous avons constaté que la lecture de vidéos sur Matlab est extrêmement lente comparée à d'autre lecteurs vidéos ce qui nous a amené à délaisser Matlab [8] et à utiliser Openshot.

Younes BELKADA créa un fichier *comma-separated values* (CSV) contenant les *frames* représentant le début et la fin de chaque geste. L'équipe utilisa ce patron pour noter le début et la fin de chaque geste sur toutes les vidéos. L'ensemble du groupe se mis d'accord sur la façon d'annoter les vidéos et un fichier CSV pour les annotations fut créé. Nous avons ainsi pu noter les 9 gestes élémentaires que compose un exercice avec des notes comprises entre 0 et 5.

3.1.5 Conversion vidéo – tenseur

Afin de pouvoir entraîner nos réseaux de neurones, nous avons utilisé des tenseurs correspondant à un exercice d'une personne (ou une vidéo). Il a donc fallu convertir chaque vidéo en tenseur. Pour chaque personne, un fichier CSV lui est associé contenant sur chaque ligne le numéro du geste, la *frame* de début $frame_{debut}$, la *frame* de fin $frame_{fin}$ et la note attribuée au geste. Il est donc possible de pouvoir accéder à des gestes spécifiques dans chaque tenseur, grâce à ce fichier.

3.1.6 Crédit des réseaux de neurones

La création des différentes architectures de réseaux de neurones artificiels après l'obtention de tenseurs correspondants aux sorties d'OpenPose s'est faite à partir de l'architecture d'un *CNN* et d'un *LSTM*. Ces deux archi-

1D CONVOLUTIONAL - EXAMPLE

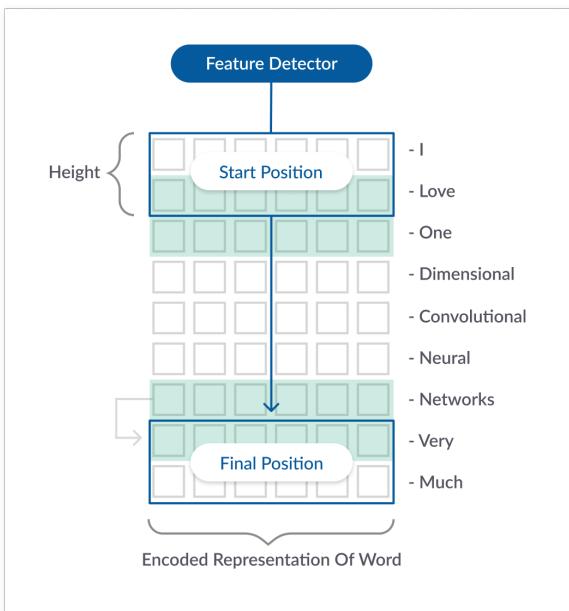


FIGURE 3.4 – Schéma explicatif d'une couche Conv1D

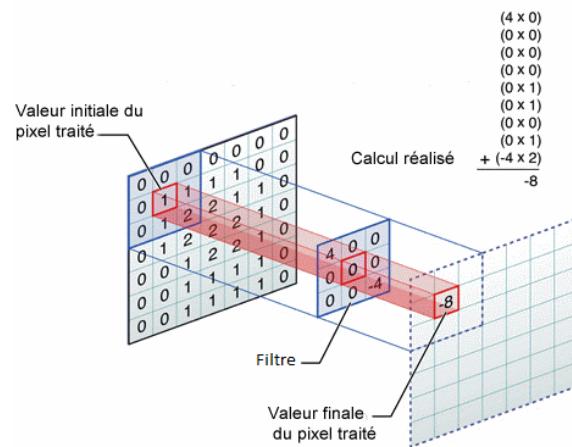


FIGURE 3.5 – Schéma explicatif d'une couche Conv2D

tectures dont nous avons essayé de nous inspirer pour la réalisation de nos propres réseaux de neurones artificiels sont issues du travail de MILLAN et ACHARD [22]. Ces réseaux permettent l'étude de mouvements dans le cadre de la chirurgie : ce sont des mouvements plus minutieux que ceux que nous tentons d'analyser. Néanmoins, de telles architectures étant difficiles à mettre œuvre, nous avons décidé de nous en inspirer brièvement et de construire des architectures très basiques dans un premier temps afin de pouvoir effectuer des tests et avoir des résultats probants sur notre base de données. Jérôme BONACCHI et Najwa MOURS LI étaient en charge du codage d'un *long short-term memory* (*LSTM*), Younès BELKADA et Abdallah DAABOUL de celui d'un *convolutional neural network* (*CNN*) et Arthur ZUCKER du *support vector regression* (*SVR*) utilisé par PIRSIAVASH, VONDRIK et TORRALBA [18]. Les deux binômes ont travaillé sur Google Colaboratory afin de programmé en Python chaque architecture et le *SVR* a été codé via Matlab [8].

3.2 Réalisation du CNN

Le but ici est de chercher à savoir si une architecture fondée sur un *convolutional neural network* est adaptée à notre problème, c'est à dire, si nous arrivons à obtenir un score raisonnable sur l'ensemble de notre jeu de données. Il faut savoir qu'il existe plusieurs types de *CNN*, plus ou moins adaptés à des tâches différentes. En effet, la taille et le nombre de filtres de convolution vont déterminer la précision de notre réseau.

Il est très commun d'utiliser les couches Conv2D (cf. figure 3.5) pour des applications au traitement d'images (classification, segmentation, etc.) et il existe plusieurs architectures notables comme l'*AlexNet* ou le *VGG16*. Lorsqu'on s'apprête à aborder une tâche dans le traitement d'images le plus commun est de s'inspirer de ces architectures ou de faire tout simplement du *transfert learning*¹ sur ces architectures. Notons que la couche Conv2D renvoie en sortie un tenseur dont la taille varie selon le nombre de filtres et la taille de ces derniers.

Cependant, nous allons dans notre cas essayer un autre type de *CNN*, il s'agit des couches Conv1D (cf. figure 3.4). Ces dernières sont utilisées pour du traitement de texte, mais surtout pour des données qui dépendent du temps, ce qui est adapté à notre cas. Le fonctionnement du réseau diffère dans le comportement du filtre, au lieu que ce dernier se déplace dans les 2 directions (*x* et *y*), il se déplacera dans une direction (celle du temps).

1. Ré-entrainer un modèle déjà entraîné en utilisant une nouvelle base de données, en général plus petite.

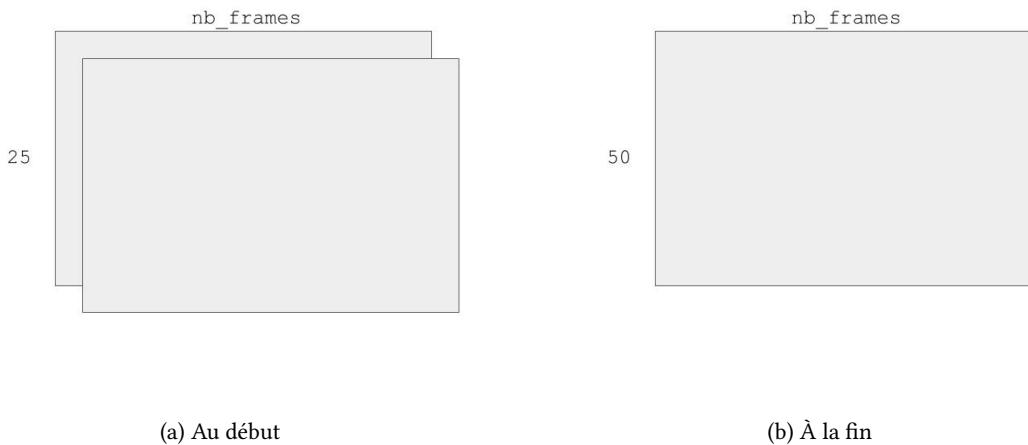


FIGURE 3.6 – Structure des tenseurs utilisés

Afin de faciliter l'apprentissage, une certaine transformation des données s'impose. Nous devons faire en sorte que les valeurs à l'intérieur de nos tenseurs varient entre -1 et 1 , sinon, les coordonnées avec une grande variance vont dominer celles avec une petite variance, et par conséquent nuire au modèle. Il faut savoir que nous avons deux tenseurs, un pour les x et un pour les y . C'est pour cela que nous soustrayons, puis divisions l'ensemble des valeurs (x, y) de nos tenseurs respectivement par la moyenne et par l'écart-type des x et des y des *keypoints* correspondants. D'autres transformations s'imposent également, le réseau de neurones doit pouvoir récupérer en entrée des tenseurs de taille constante. C'est pour cela que nous faisons du *padding*² sur les tenseurs pour qu'ils aient tous une taille identique. Après avoir récupéré le nombre maximal de *frames* n_{\max} , nous avons fait en sorte que les tenseurs soient tous de taille $2 \times 25 \times n_{\max}$.

3.2.1 Première approche

Nous avons réalisé plusieurs essais sur différentes architectures en utilisant des couches Conv2D. Le premier essai était donc de récupérer les vidéos d'exercice sous forme de tenseurs de dimension $2 \times 25 \times n_{\text{frames}}$. L'approche intuitive serait de faire passer ces tenseurs par des *CNN* avec des filtres 2D verticaux c'est à dire de taille $1 \times n$ avec n un paramètre à ajuster au fur et à mesure. Le filtre appliquera l'opérateur de convolution sur les deux tenseurs donnés en entrée. Cela permettrait au réseau de mettre en avant les corrélations temporelles de nos mouvements. Il s'est avéré que cette approche était incorrecte puisque le réseau apprenait à donner en sortie un score constant. L'erreur semble liée au fait que les filtres de convolution additionnent les valeurs x et y . Cela fausserait l'apprentissage, il faut donc certainement redéfinir une nouvelle structure pour le tenseur.

3.2.2 Seconde approche

Nous avons décidé de changer la structure des tenseurs, en concaténant la matrice des x et la matrice des y afin d'essayer de mettre en place des couches Conv1D. Les filtres ici sont de taille n et sont de dimension 1. Nous nous retrouvons avec un tenseur de dimension $50 \times n_{\text{frames}}$, ce qui permet de ne pas biaiser le *CNN*.

3.2.3 Architecture finale

Après plusieurs essais nous avons décidé de travailler avec l'architecture présentée sur la figure 3.7, avec les hyper-paramètres suivants :

2. Traduit par bourrage.

```

1   — nb_epochs = 200
2   — batch_size = 32
3 model = Sequential()
4
5 model.add(Convolution1D(filters = 16, filter_size = 3, activation='relu', input_shape
6     =(322,50)))
7 model.add(MaxPooling1D(3))
8 model.add(SpatialDropout1D(0.2))
9
10 model.add(Convolution1D(filters = 32, filter_size = 3, activation='relu'))
11 model.add(MaxPooling1D(3))
12 model.add(SpatialDropout1D(0.2))
13
14
15 model.add(Convolution1D(filters = 64, filter_size = 3, activation='relu'))
16 model.add(MaxPooling1D(3))
17 model.add(SpatialDropout1D(0.2))
18
19 model.add(Flatten())
20 model.add(Dense(50, activation='relu'))
21 model.add(Dropout(0.5))
22
23 model.add(Dense(1, activation='linear'))
24

```

Listing 3.1 – Code utilisé pour coder le modèle final

Notre architecture se compose des couches suivantes :

- Des couches dites Conv1D. Comme expliqué dans les parties précédentes, les couches prennent en paramètre le nombre de filtres, ainsi que de la largeur de ce dernier. Le filtre occupe la totalité du tenseur, et se déplace dans une direction (direction du temps) contrairement à la fonction Conv2D qui renvoie une matrice en sortie et où les filtres se déplacent dans les deux directions.
- Des couches MaxPooling1D qui prennent en argument un paramètre correspondant à la taille t du filtre. Elle réalisera une opération de MaxPool sur une surface de taille t du vecteur donné en argument.
- Des couches SpatialDropout qui réalisent une opération *dropout* qui consiste en l'annulation aléatoire de coefficients d'apprentissages (selon un paramètre) afin d'éviter le sur-apprentissage.

Ces blocs sont répétés 4 fois, avec des paramètres différents avant de passer à une succession de couches denses. Les détails sont explicités sur la figure 3.7.

3.3 Réalisation du *LSTM*

Nous utilisons un réseau *long short-term memory (LSTM)* qui est une architecture de réseau de neurones artificiels fondée sur les principes des *recurrent neural networks (RNNs)*, pouvant mémoriser de l'information, et constituerait ainsi une première approche nécessaire à l'élaboration d'une architecture plus complexe.

Nous récupérons un tenseur d'ordre 3 par exercice et par personne de taille $(2, n_{frames}, 25)$. Nous chargeons également les notes associées à chaque geste de chaque exercice de chaque personne. Afin d'aider le réseau à apprendre plus facilement quels sont les paramètres qui nous intéressent, nous ne lui donnons que les articulations du haut du corps. Ces articulations sont les points $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 12\}$ de la numérotation OpenPose (cf. figure 3.1b). Pour faire cela, nous construisons une copie du tenseur précédent où nous nous contentons d'ajouter seulement les coordonnées des articulations qui nous intéressent. Par ailleurs, il est nécessaire de changer la forme du tenseur d'ordre 3 en ordre 2 car le *LSTM* ne prend en entrée que des tenseurs d'ordre 2. Ainsi, nous avons décidé de concaténer l'ensemble des ordonnées des points après l'ensemble des abscisses des points. En outre, si une articulation n'est pas détectée, au lieu d'avoir une coordonnée à 0, nous copions la valeur de la position précédente. De plus, tous les tenseurs obtenus ne font pas la même taille car les vidéos dont ils proviennent étaient bien évidemment de durées différentes, et donc d'un nombre de *frames* différent. Afin de donner au *LSTM*

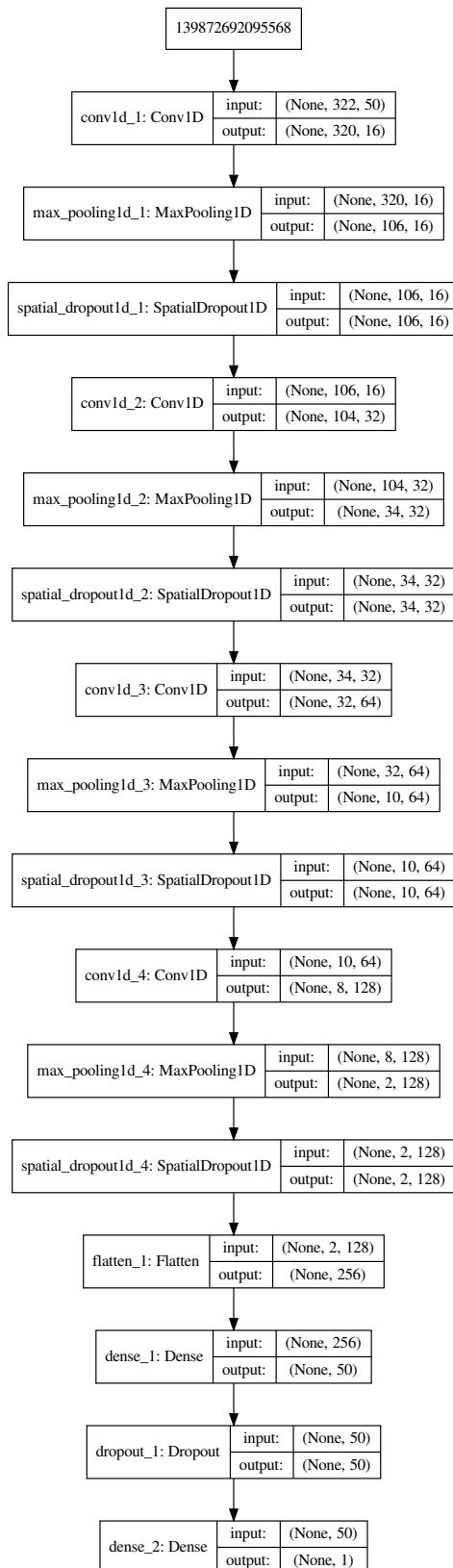


FIGURE 3.7 – Modèle final retenu pour l'apprentissage

des tenseurs de même taille, nous avons décidé de remplir les tenseurs par des 0 jusqu'à ce qu'ils fassent tous la même taille (correspondant à la vidéo la plus longue). Ensuite, afin que le réseau apprenne efficacement, il convient de lui donner des entrées standardisées. Pour cela, suite au conseil de Madame Catherine ACHARD, nous centrons toutes les coordonnées par rapport à un point médian du corps (correspondant au point 8 d'OpenPose), puis nous divisons ces nouvelles distances par la distance entre les points 1 et 8. Concernant les notes, afin d'avoir une note par exercice comprise entre 0 et 1, nous sommes toutes les notes des gestes d'un exercice que nous divisons par la note maximale théorique.

Le réseau quant à lui est conçu avec une architecture assez simple composée d'une couche d'entrée *LSTM*, une couche cachée dense qui combine les informations provenant de la couche précédente, une couche Dropout consistant à annuler aléatoirement certains coefficients d'apprentissages afin d'éviter le sur-apprentissage, et une sortie scalaire³. Nous lançons le code sur plusieurs réseaux différents avec la même structure mais où les dimensions des couches diffèrent afin de connaître quel modèle donne les meilleurs résultats. Nous essayons des tailles qui sont des multiples de 2. Nous séparons les données en ensemble de test et en ensemble d'apprentissage avec un ratio 20/80. Et, nous étudions l'erreur quadratique moyenne (ou *MSE*). De plus, nous utilisons l'optimiseur Adam et nous fixons le *learning rate* initialement à 0,001, puis nous le divisons par 2 toutes les 10 *epochs*. Cela devrait permettre de converger vers un minimum. Il est à noter que nous avons essayé plein de réglages différents. Par exemple, nous avons testé plusieurs réglages d'hyper-paramètres comme une régularisation L¹ ou L² avec des coefficients différents, une initialisation des poids du réseau selon une loi normale centrée d'écart-type 0,01, divers taux de *dropout* d'au moins 20%.

3.4 Réalisation de la combinaison d'un *CNN* et d'un *LSTM*

Afin de tirer partie des qualités du *CNN* pour agréger les informations et celles du *LSTM* pour les mémoriser, nous nous inspirons de l'architecture finale du *CNN* (cf. figure 3.7) où nous remplaçons la couche Flatten permettant de passer d'une couche de convolution à une couche dense par une couche *LSTM*. Nous utilisons l'ensemble des gestes et toutes les articulations nous traitons de la même manière que pour le *LSTM* mis à part la normalisation qui est celle utilisée par le *CNN*. Nous exécutons le modèle avec 150 *epochs*, des *batchs* de taille 32, une couche Dropout (l'avant-dernière couche) avec un taux de 80%, l'optimiseur Adam avec un *learning rate* initial de 0,001 multiplié par 0,85 toutes les 10 *epochs*.

3.5 Autre solution : le *SVR*

Bien que l'utilisation d'un *convolutional neural network* semble très adapté et populaire, il a été porté à notre attention qu'un article scientifique utilisant un algorithme de *support vector regression* (*SVR*) (décrit ci-après) ainsi qu'un détecteur de pose différent d'OpenPose existait. Cet article de PIRSIAVASH, VONDRIK et TORRALBA [18] présente une *pipeline* (une série de programmes) permettant l'annotation automatique de mouvements issus des J.O. et de leurs épreuves de plongée. Notre projet ayant une visée similaire, nous avons donc décidé de nous pencher sur cet article et d'étudier le code adjoint à celui-ci afin de reproduire les résultats présentés, qui semblaient très concluants.

3.5.1 *SVM* et *SVR*

Afin de pouvoir présenter le mécanisme de fonctionnement du *SVR*, nous devons tout d'abord présenter l'algorithme plus simple sur lequel il se fonde : le *support vector machine* (*SVM*).

Ce modèle d'apprentissage supervisé (appartenant au domaine de l'intelligence artificielle), repose sur des propriétés géométriques pour séparer, au mieux, deux classes ou modalités présentes dans les données d'appren-

3. Détail sur l'implémentation : nous utilisons un CuDNNLSTM qui est une version optimisée de *LSTM* de la bibliothèque Keras [10].

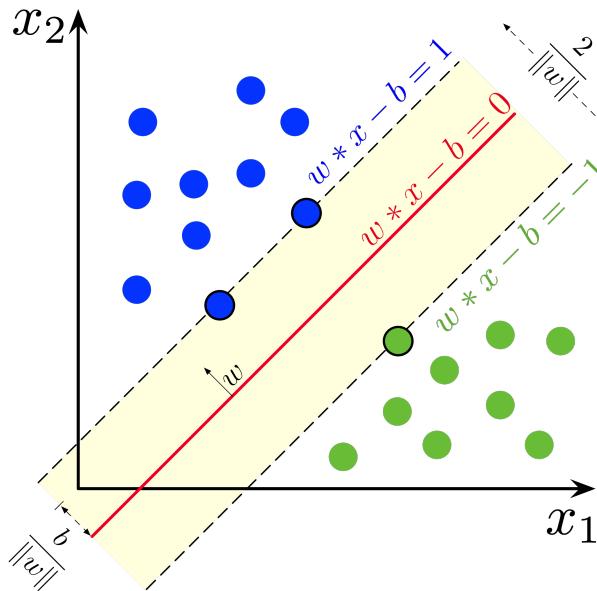


FIGURE 3.8 – Schéma du *support vector machine* (SVM) (cf. en.wikipedia.org/wiki/Support-vector_machine)

tissage. En notant le vecteur x_i le vecteur réel correspondant aux paramètres étudiés sur les données, et y_i la classe des données, l'ordinateur va chercher un réel w et un réel b optimaux tels que $w \times x - b = 0$ définit l'hyperplan maximisant la distance entre les deux classes. Ensuite, il va étudier le signe de $w \times x_{\text{new}} - b$ et, selon son signe (± 1) prédire s'il appartient à la classe $y = 1$ ou $y = -1$ qui peuvent correspondre ensuite à des noms de classes. Les données d'entraînement permettent de trouver w^*, b^* optimaux, permettant le bon fonctionnement du modèle. Finalement, afin d'obtenir un modèle résistant à des petites perturbations, il faut minimiser $\|w\|$, pour maximiser la distance entre deux classes. Le schéma 3.8 représente l'hyperplan calculé par le SVM en deux dimensions.

Le SVR utilise les mêmes propriétés que le SVM, tout en permettant non pas de prédire une classe, mais de calculer un coefficient réel correspondant à la valeur prédite de y , qui cette fois-ci peut prendre une infinité de valeurs. Sachant que la solution ne peut pas être exacte, une marge de tolérance est définie (ε) à partir du modèle SVM lié au problème. Le but est toujours de minimiser l'erreur de prédiction. De plus, dans le cas où les données sont corrélées de manière non-linéaire, le SVR transforme les données à l'aide d'une fonction noyau afin de rendre la prédiction linéaire. Un exemple de transformation est présenté dans 3.9.

3.5.2 Présentation du code

Après avoir récupéré le code implémentant le SVR, Arthur a essayé de le comprendre et de l'entraîner avec notre base de données. Cependant, le code n'ayant pas été mis à jour depuis longtemps, il y avait de nombreuses erreurs liées à la fois à la version de Matlab et au code en lui-même qui n'a pas été écrit de façon à être facilement utilisé pour une autre base de données. De plus, des problèmes liés à l'ordinateur utilisé par Arthur (étant un Mac) sont apparus. Le code permet de couper automatiquement les vidéos en *frames* à partir des annotations. Le format des annotations étant différent, Arthur a écrit un simple script permettant de passer de notre format d'annotation à celui adapté au code. La détection est ensuite effectuée sur l'ensemble des *frames*, puis l'ensemble des données récupérées sur les articulations sont concaténées en un seul fichier matriciel. Finalement, le SVR est entraîné à l'aide des positions des articulations et des annotations associées à chaque *frame*.

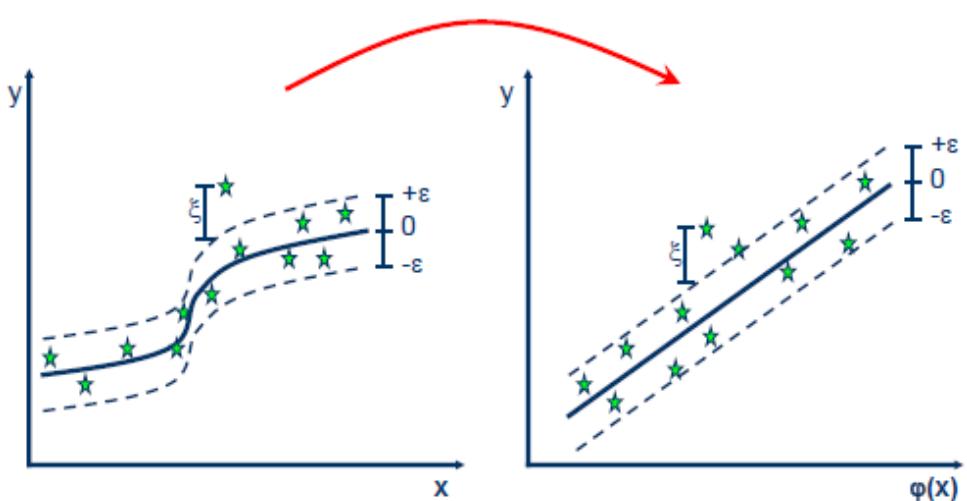


FIGURE 3.9 – schéma du *support vector regression* (SVR)

4. Résultats

4.1 Résultats du CNN

4.1.1 Première approche

En traitant les données en utilisant la première approche, c'est-à-dire en donnant en entrée des tenseurs de taille $2 \times 25 \times n_{frames}$, le réseau renvoie un score constant quelle que soit la matrice donnée en entrée. Nous pouvons conclure que la façon dont nous avons utilisé les couches Conv2D ainsi que nos tenseurs ne sont pas adaptés afin d'obtenir des scores exactes. L'apprentissage n'a pas abouti à quelque chose d'exploitable pour les raisons suivantes. Premièrement, le fait d'avoir deux tenseurs, x et y biaise l'opération de convolution qui va additionner les couples (x, y) . Il n'existe certainement pas de corrélation à proprement parler entre les points x et y , d'où la nécessité de les traiter séparément. Deuxièmement, la couche *Conv2D* permet au filtre de se déplacer dans les deux directions du tenseur. Cette approche est cohérente dans les problèmes de traitement d'images, mais pas adaptée à notre problème puisque les tenseurs sont organisés de manière particulière (il s'agit de tenseurs organisés en coordonnées spatiales \times temps, ce qui est totalement différent des images qui sont des pixels \times pixels).

4.1.2 Seconde approche en utilisant l'architecture finale

Ici, nous avons utilisé l'architecture finale (en utilisant les couches Conv1D), et en appliquant la transformation sur les tenseurs. Rappelons que chaque exercice, et donc chaque tenseur, est constitué de 9 gestes élémentaires dont le geste final est le *squat* (donc des gestes impliquant des membres inférieurs principalement). Pour avoir de meilleurs résultats, nous avons entraîné le modèle geste par geste au lieu de lui donner l'ensemble du tenseur directement. Ceci signifie que nous avons créé des *batchs*¹ triés selon les gestes. Dans un premier temps, nous avons entraîné le modèle sans les *squats* et sans les articulations du bas corps et de la tête. En entraînant le réseau, nous obtenons une *root mean squared error (RMSE)* de 0,006, une corrélation de 0,91 entre les notes prédictes et les notes réelles (cf. figure 4.1). Nous pouvons voir qu'en changeant la structure des données ainsi qu'avec une meilleure visualisation du problème, nous pouvons aboutir à de meilleurs résultats. Nous discuterons des perspectives d'évolution dans la prochaine partie.

Nous avons entraîné dans un second temps le modèle 3.7 avec les squats et toutes les articulations. En entraînant le réseau, nous obtenons une *root mean squared error* de 0,005, une corrélation entre les notes prédictes et les notes réelles de 0,94 (cf. figure 4.2).

1. groupement de tenseurs

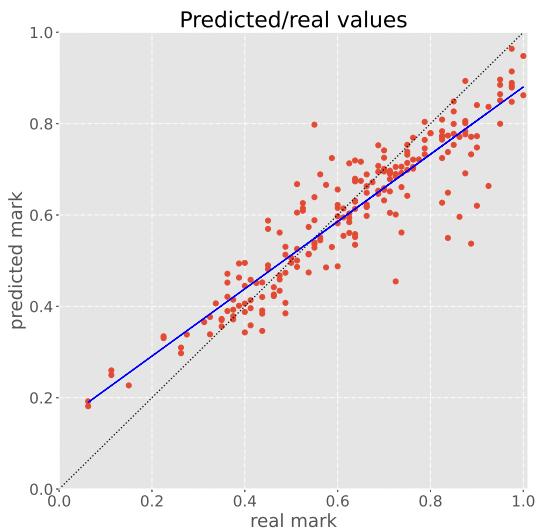


FIGURE 4.1 – Résultat du CNN sans les *squats*. Graphique des notes prédites en fonction des notes réelles. En abscisses, le score réel et en ordonnées le score prédit.

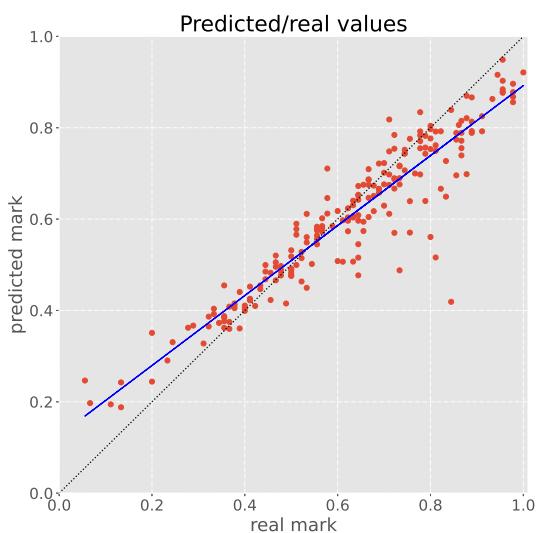
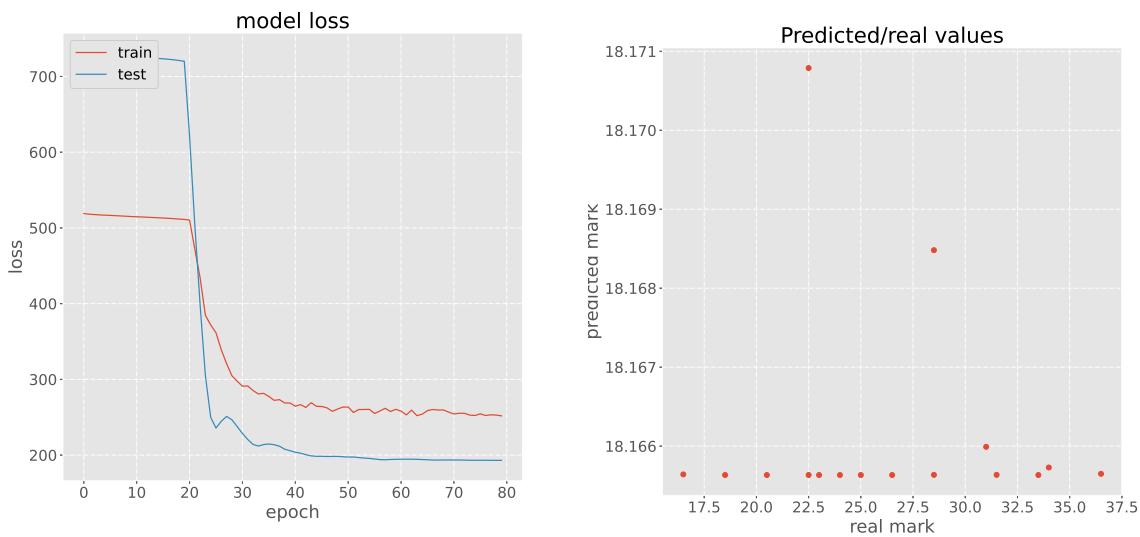


FIGURE 4.2 – Résultat du CNN avec les *squats*. Graphique des notes prédites en fonction des notes réelles. En abscisses, le score réel et en ordonnées le score prédict.



(a) Évolution de la fonction de perte en fonction du nombre d'itérations (b) Graphique des notes prédites en fonction des notes réelles

FIGURE 4.3 – Exemple de résultat du *LSTM*

4.2 Résultats du *LSTM*

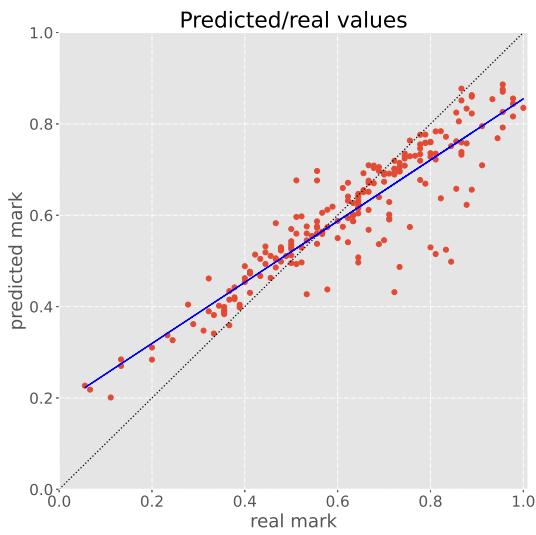
Les résultats obtenus avec le réseau *long short-term memory* sont constants la plus part du temps, qu’importe les dimensions des couches. Cela dénote que le réseau n’est pas capable d’apprendre les caractéristiques souhaitées. De plus, il est à noter qu’il y a un caractère aléatoire qui a permis d’obtenir une fois ou deux une courbe reliant les valeurs prédites aux valeurs réelles brièvement linéaires. Cependant, la corrélation entre les deux grandeurs n’a jamais dépassé la valeur de 0,68. Il est à noter que nous avons essayé plein de réglages différents en vain. Par exemple, nous avons testé plusieurs réglages d’hyper-paramètres comme la régularisation, l’initialisation des poids du réseau, le taux de *dropout*. Par ailleurs, même en ne donnant que les réalisations les plus lentes des exercices afin que les données se ressemblent plus et que le *padding*² de 0 soit moins important, le réseau ne donna pas de bons résultats (cf. figure 4.3).

4.3 Résultats de la combinaison d’un *CNN* et d’un *LSTM*

Cette architecture permet d’obtenir une corrélation de 0,91 entre les notes prédites sur l’ensemble de test et les notes réelles ainsi qu’une *RMSE* de 0,006. La figure 4.4 montre les notes prédites en fonction des notes réelles. Le *CNN* a donné des résultats un peu meilleurs, mais ceux-ci n’ont été réalisés qu’avec une partie des articulations et sans les *squats* contrairement au résultat obtenu ici.

La combinaison de *CNN* et de *LSTM* utilisée a donné de bons résultats car elle permet de mixer les qualités des deux types de réseaux. En effet, traiter directement les données d’entrées par un *LSTM* est beaucoup plus lent que de passer d’abord par des couches de convolution car les données sont alors agrégées intelligemment, et donc moins nombreuses. De plus, nos résultats nous poussent à croire que l’utilisation d’un *LSTM* seul est largement insuffisant. Toutefois, l’architecture que nous avons construite n’est pas encore à la hauteur du problème que nous cherchons à résoudre : produire des notes et des commentaires digne d’experts.

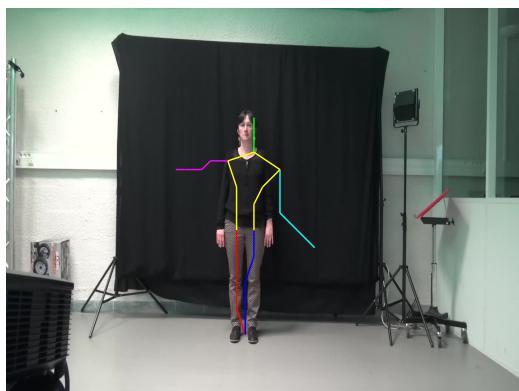
2. Traduit par bourrage.


 FIGURE 4.4 – Résultat de la combinaison d'un *CNN* et d'un *LSTM*

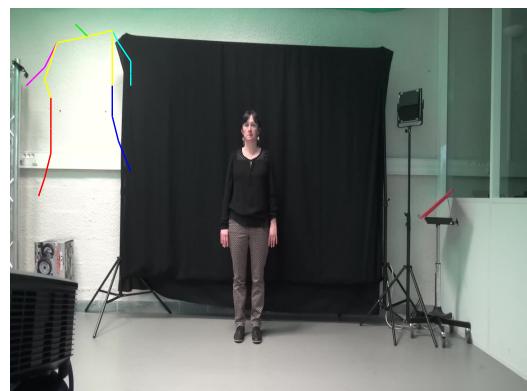
4.4 Résultats du *SVR*

Avant de pouvoir entraîner le modèle *SVR*, la détection de pose doit être lancée en utilisant la méthode des *N-best* positions tirée de DENNIS PARK [23]. Celle-ci se fonde sur l'estimation de plusieurs positions pour chaque *frame* à partir d'une méthode décrite par YANG et RAMANAN [24] pour ensuite calculer parmi N estimations les meilleures positions possibles pour chaque articulation.

Le temps d'exécution par *frame* était de l'ordre de 2 minutes, ce qui rendait impossible l'application à l'ensemble de la base de données (contenant plus de 40 000 *frames*). De plus, bien que la forme globale de la personne soit bien détectée, les positions des articulations, principalement celles des bras, sont chaotiques (cf. figure 4.5), ne permettant pas l'utilisation des squelettes récupérées.



(a) Mauvaise détection des bras



(b) Mauvaise position de la détection

 FIGURE 4.5 – Mauvaises détections du *SVR*

5. Conclusion

5.1 Évaluation des exigences

Nous pouvons maintenant étudier les résultats que nous avons obtenus par rapport aux exigences fixées initialement dans le cahier des charges (cf. en annexe page 34).

Nous avons utilisé OpenPose pour la détection et le suivi des mouvements car cela est l'un des systèmes les plus adaptés et les plus précis. Par conséquent, nous ne pouvons pas contrôler la précision de ce système, mais il nous semble que cela est l'un des meilleurs aujourd'hui.

De plus, n'ayant pas utilisé une base de données contenant des gestes d'athlètes notés par des entraîneurs ou des experts, nous pouvons difficilement estimer la qualité des prédictions renvoyées par notre système basé sur une notation tout à fait arbitraire. Toutefois, au vu des 91% de corrélation obtenu avec l'architecture combinant le *CNN* et le *LSTM*, nous estimons avoir validé le niveau que nous nous étions fixé qui était de 75%.

5.2 Perspectives d'évolution

Une future étape est de tester notre système sur une base de données avec des gestes de professionnels et des notes données par des entraîneurs ou des experts (par exemple, *UNLV Olympic Scoring Dataset*). En outre, il serait intéressant d'entraîner le réseau sur d'autres exercices et d'autres sports. De plus, il serait très intéressant de prédire également la note de chaque geste élémentaire afin de savoir si la note totale est ainsi mieux estimée qu'en apprenant directement cette note et si nous retrouvons les notes les plus basses, et donc les instants où il y a des erreurs. Nous espérerions ainsi avoir un réseau avec les mêmes avantages que celui de PIRSIAVASH, VONDRIK et TORRALBA [18] sans ses défauts.

Concernant la méthode *SVR*, nous pourrions utiliser comme système de détection OpenPose qui est beaucoup plus précis que celui présent originellement afin d'entraîner le *SVR*. Il faudrait alors transformer les fichiers obtenus et adapter le code Matlab.

Actuellement, nos architectures fonctionnelles normalisent les coordonnées des articulations (soustraction de la moyenne, puis division par l'écart-type). Nous pourrions également étudier si notre normalisation est la plus optimale ou si une autre, fondée sur des invariants de l'anatomie humaine, permettrait à notre réseau de mieux s'habituer à la diversité des corps. D'un point de vue anatomique, les seuls invariants du squelette de chaque être humain sont ce que l'on appelle le centre de gravité S3 ou troisième vertèbre sacrale en passant par le sommet des deux hanches (le toit de l'acétabulum) ainsi que l'appui proximal inter-articulaire qui correspond à l'espace formé par l'ensemble d'articulations de l'épaule (reliant le tronc aux membres supérieurs). Nous pensons que normaliser toutes les morphologies rencontrées dans notre base de données par la distance entre la vertèbre S3 et cet espace au niveau des deux épaules améliorera largement la précision de nos résultats. La figure 5.1 illustre cette distance.

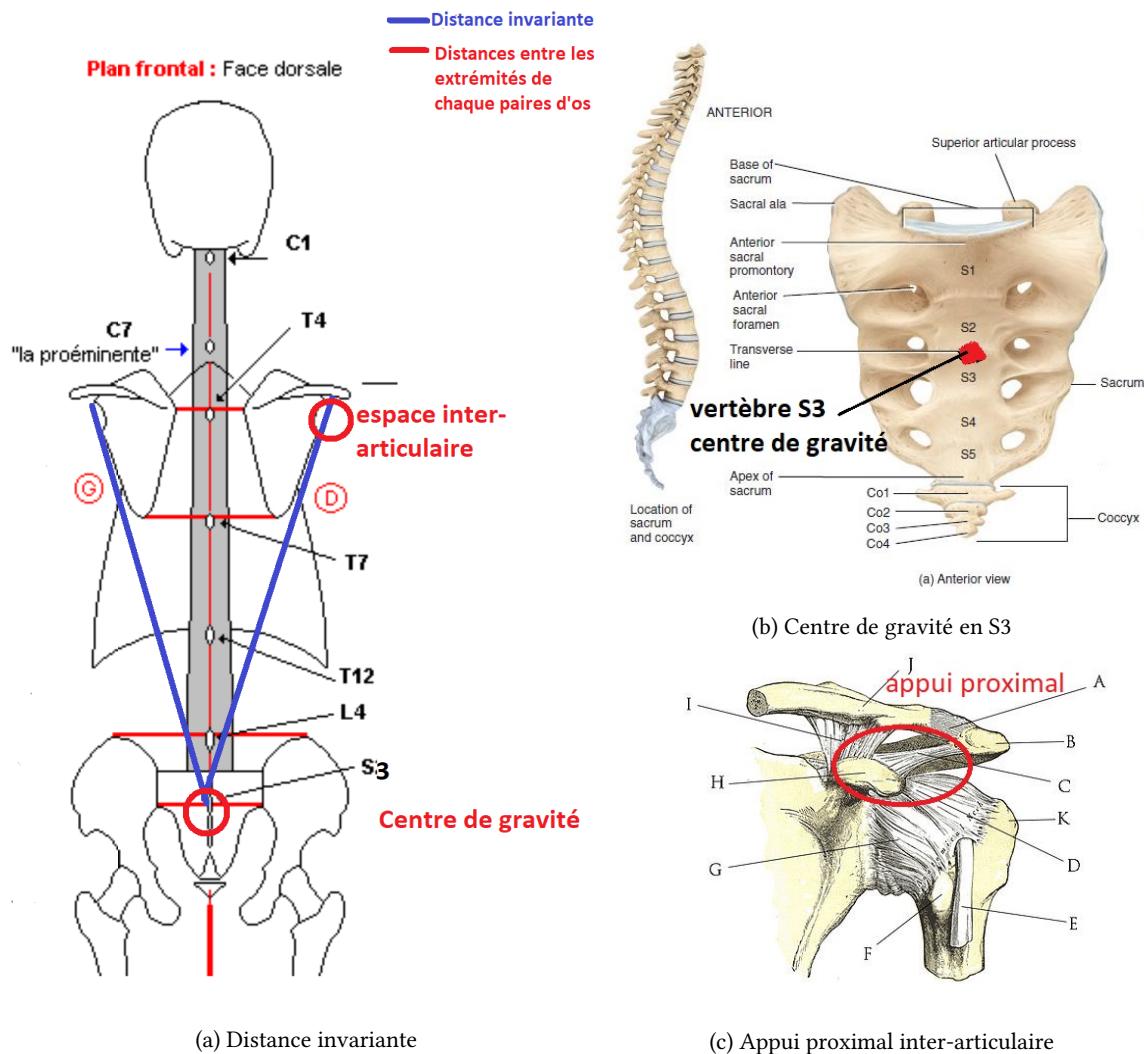


FIGURE 5.1 – Invariants utiles à la normalisation de la morphologie humaine

Par ailleurs, bien qu'aucune interface graphique n'ait été demandée par l'ISIR, il est à noter que ce projet devrait en avoir une à terme car, étant destiné à des néophytes de l'informatique, ce système doit être facilement accessible et utilisable. De plus, cette application devra avoir de bonnes performances et une faible consommation afin qu'elle fonctionne sur des serveurs ou sur des appareils portables. Finalement, afin d'affiner la précision de notre modèle, nous pourrions étudier l'intérêt de diverses propositions comme :

- explorer la piste, sans doute prometteuse, des *convolutional LSTM* ;
- utiliser un réseau siamois ;
- utiliser le niveau de confiance renvoyer par OpenPose ;
- utiliser, s'il existe, un système de détection encore plus précis.

Le plus important à faire lorsque nous abordons des problèmes liés à l'intelligence artificielle s'agit de trouver la bonne approche pour y faire face. À travers nos expérimentations nous avons bien pu voir que notre approche, constituée d'une série d'étapes assez longue est très largement exploitable, grâce aux résultats que nous avons pu obtenir (taux de corrélation élevée). Avec plus de temps, nous aurions pu développer une base de données plus précise, avec des mouvements plus en rapport à la musculation.

5.3 Obstacles et difficultés rencontrés

Le tableau 5.1 est le registre décrivant l'ensemble des risques qui sont devenus de réels problèmes au cours de l'avancement du projet, bien que certains aient été jugés comme faiblement probables. Il est à noter que ce registre n'est qu'une partie de celui initialement rédigé dans le cahier des charges présent en annexe à la page 34. Chaque risque est associé à des stratégies pour y pallier et à une criticité dont les niveaux sont détaillés dans le tableau 5.2. Cependant, la réalité a été toute autre, certains obstacles ont été difficiles à surmonter et d'autres inattendus sont survenus au cours du projet.

Id.	Description du risque	Criticité	Responsable	Prévention	Réparation
A	Création d'une base de données par nos soins	Moyen – Haut	Arthur ZUCKER et Younes BELKADA	Recherche de bases de données adaptées	Automatisation de la création de la base de données
B	Organisation du travail sous-optimale	Moyen	Jérôme BONACCHI	Utilisation des outils de gestion de projet	Réduction de la portée du projet
C	Le <i>feedback</i> est inappropriate	Moyen – Haut	Younes BELKADA, Jérôme BONACCHI et Abdallah DAABOUL	Demander à un vrai formateur un commentaire correct de chaque mouvement	Supprimer le système de <i>feedback</i>
D	Un membre de l'équipe est indisponible	Bas	Tous les membres	Travailler toujours en binômes	Répartir les tâches entre le reste du groupe
E	Manque de communication	Moyen	Tous les membres	Utiliser constamment les TIC	Faire le point chaque semaine sur l'avancement de chacun

Tableau 5.1 – Registre des risques rencontrés

Voici, ci-dessous, la description des obstacles et des difficultés rencontrés identifiés de la même manière que dans le registre des risques rencontrés (cf. tableau 5.1).

A. **La création d'une nouvelle base de donnée** a été notre plus gros frein et a fait survenir des risques

Bas	Les risques sont presque improbables et ont un faible impact.
Bas – Moyen	Les risques ont un faible impact, mais ils sont plus fréquent.
Moyen	Les risques sont dangereux, mais pas si fréquent.
Moyen – Haut	Les risques dangereux qui pourraient arriver.
Haut	Comme Moyen – Haut, mais les conséquences sont pires.

Tableau 5.2 – Niveaux de criticité des risques

non prévus tels que la démotivation de l'ensemble de l'équipe et la perte de trois mois de travail. Il a été extrêmement compliqué pour chacun de recommencer de zéros sans aucune base de données à notre disposition comme il avait été prévu avec l'entreprise Chainos qui nous a fait faux bond. Par conséquent, cela a grandement impacté la progression de l'équipe. Pouvoir travailler avec l'ISIR sur le même sujet, nous a redonné confiance, ayant une nouvelle encadrante proche, disponible et attentive à notre travail. Nous avons pu être guidé pas à pas pour la création de la base de données qui a été effectivement laborieuse mais enrichissante à la fois. Tout le monde a participé à la réparation de ce problème et tout particulièrement Arthur et Younès qui avaient travaillé avec Chainos auparavant et qui s'y connaissaient un peu en matière de *deep learning*.

- B. **La cadence de travail et les conditions** n'ont pas été optimales dû en partie à la raison citée ci-dessus, mais également à d'autres facteurs dont les risques D et E. Nous avons discuté bien évidemment de la réduction des objectifs de ce projet. Ceci a été effectué par Jérôme principalement avec l'aide de l'ensemble de l'équipe.
- C. **La réalisation du système de feedback** a été malheureusement impossible car après toutes ces péripéties, le temps restant était trop court pour réaliser cette partie du projet.
- D. **L'indisponibilité d'un ou plusieurs membres de l'équipe** a été le deuxième obstacle le plus difficile à surmonter car sa fréquence d'apparition, jugée faible au premier abord, s'est révélée très élevée dans notre groupe de part les obligations personnelles ou maladie d'une partie de l'équipe. Le reste de l'équipe qui pouvait travailler au moment présent s'est réparti les différentes tâches que les membres absents ne pouvait exécuter. Younès, Abdallah et Jérôme se sont occupés de la répartition équitables des tâches.
- E. **Le manque de communication** a été inhérent à notre groupe. Chaque membre du groupe avait tendance à apprécier le travail en autonomie complète ce qui a entraîné des coupures de communication de notre travail personnel pendant plusieurs semaines. Puisque nous avions commencé à travailler chacun de notre côté, aucun membre de l'équipe n'était au courant de l'ensemble de l'avancée du projet. À la moitié du projet, après avoir constitué notre base de données, nous avons décidé de se réunir chaque mercredi pour quatre heures de travail minimum par binôme nouvellement formé : Jérôme – Najwa, Abdallah – Younès et Arthur travaillant sur une mission précise mais aidant les binômes quand il était nécessaire. La communication entre nos encadrants s'est faite de façon plus efficace par Younès et le problème de communication interne a été soulevé par Najwa qui a proposé des solutions comme le partage permanent du flux de travail. La mise en binôme de travail s'est faite par accord commun.

Il est évident que nous avons rencontré d'autres obstacles inhérents à tout projet tels que les *deadlines* à respecter, le manque de moyens matériels ou humains périodiquement, le temps consacrer à la réalisation du projet, etc.

De plus, dans le registre des risques du cahier des charges (cf. tableau A.1 en annexe) nous avions cité le risque de ne pas avoir en possession des processeur graphique. En utilisant OpenPose sur Google Colaboratory, nous avons fait face à un inconveniant assez conséquent qui est l'obligation d'installer OpenPose à chaque fois que nous souhaitions utiliser notre code (car ce n'est pas un module Python), ce qui est une perte de temps récurrente.

5.4 Bilans personnels

5.4.1 Bilan de Younes BELKADA

Le projet semblait commencer sur de bonnes bases lors de l'été 2019 où l'entreprise dans laquelle j'étais stagiaire nous présenta le projet. Mis à part le fait que nous sommes tombés de haut avec l'inexistence de la base de données promise par Chainos Solution, je pense que ce projet m'a beaucoup apporté personnellement. J'ai toujours voulu approfondir mes connaissances dans le *deep learning* : c'est surtout pour cela que j'ai décidé de m'investir dans ce projet. J'ai aussi voulu mettre en pratique ce que j'ai pu apprendre à travers les *MOOC* que j'ai réalisé avec Coursera sur le *deep learning*. Dans le fond, j'ai pu éclaircir beaucoup de points et d'interrogations à travers les discussions avec Madame Catherine ACHARD. J'ai également appris comment les problèmes liés à l'intelligence artificielle peuvent être traités, comment avoir les bonnes approches et intuitions dans ces cas de figure. Je ne me considère pas comme quelqu'un qui maîtrise parfaitement l'I.A. mais plutôt comme quelqu'un qui voit très bien comment des outils de l'intelligence artificielle peuvent être utilisés dans certains problèmes. J'estime que malgré notre projet j'ai encore beaucoup à apprendre, et que je n'ai encore compris qu'une infime partie de ce domaine très vaste.

Travailler avec mes camarades dans ce domaine fut une très belle expérience, je regrette néanmoins qu'on ne soit pas formés davantage en *deep learning*. J'ai également apprécié travailler avec l'ISIR, à travers cette expérience j'ai pu voir concrètement quelles sont les exigences et le rythme de travail dans un institut de recherche. Je regrette que notre équipe n'ai pas pu fonctionner à la fin, mais je ne blâme personne car j'estime que nous avons chacun une part de responsabilité.

5.4.2 Bilan de Jérôme BONACCHI

Initialement, j'ai choisi ce projet car il aborde un domaine que je pense très important, intéressant et dans lequel j'ai envie de me former qui est l'intelligence artificielle. À cela s'ajoutait l'importance de travailler en collaboration avec Chainos : une entreprise présente dans tous les secteurs des nouvelles technologies, où Arthur et Younes y avaient fait un stage. De plus, ce qui m'intéressait était l'application de nos connaissances au profit de la santé des gens : lutter contre les douleurs musculaires et osseuses et réaliser une application qui puisse encourager les gens à pratiquer une activité physique et sportive efficace et sécurisée au quotidien. En outre, il m'importait de travailler dans une équipe composée de personnes compétentes disposant de connaissances diverses et complémentaires.

Toutefois, nous avons dû faire face à l'absence de notre encadrant de Chainos qui s'est, par la suite, concrétisée en séparation pure et simple. Par conséquent, nous avons dû réaliser laborieusement une base de données que nous n'avions pas concrètement prévue. Au fur et à mesure que le projet continuait, je me suis rendu compte que le sujet n'était peut-être pas une application des mathématiques aussi pertinente que je le pensais par le passé. De surcroît, je considère que le développement de mes connaissances n'a pas été aussi important que je le croyais. Je comptais sur ce projet pour qu'un encadrant professionnel de type ingénieur soit présent à nos côtés hebdomadairement pour nous partager ses conseils et son savoir qui sont différents de ceux d'un encadrant académique.

L'un des plus gros points négatifs est l'absence de cohésion de notre groupe la plupart du temps. J'estime que cette collaboration est infructueuse dans le sens où le travail du groupe n'a pas été assez efficace et je constate que le travail réalisé par chacun est inférieur à ce qu'il aurait dû être au vu d'un projet d'une telle ampleur et d'une telle importance dans notre cursus. En conséquence de la mauvaise gestion de projet et des impacts des facteurs extérieurs au projet, de nombreux objectifs initialement prévus avec Chainos n'ont pas été réalisés malgré l'aide apporté par Catherine ACHARD et Mégane MILLAN. Je suis donc à la fois déçu de nos résultats et du comportement de l'équipe — une incompatibilité des comportements et des caractères des membres du groupe que je n'avais pas considérée originellement.

5.4.3 Bilan d'Abdallah DAABOUL

Ma curiosité dans le domaine de l'intelligence artificielle m'a naturellement poussé à choisir ce projet. J'ai trouvé que ce projet était une opportunité pour approfondir mes connaissances dans le domaine du *data science*, et surtout, découvrir un domaine autour duquel je n'avais pas beaucoup de connaissances : le *deep learning*. Par ailleurs, le projet était encadré en partie par l'entreprise vietnamienne Chainos dans laquelle deux de mes collègues ont passé leur stage et étaient très satisfaits des compétences acquises suite à celui-ci. Par conséquent, j'étais motivé pour apprendre de mes camarades, découvrir de nouvelles bibliothèques Python, gagner en professionnalisme en communiquant à distance, pratiquer mon anglais dans le monde professionnel et collaborer avec une entreprise intéressée par les technologies modernes. Le but du projet était innovant : automatiser un entraîneur de sport. C'est un problème concret, pratique, et dans lequel nous avons une grande chance d'être confronté en tant qu'ingénieur dans une entreprise. Cela a été un véritable challenge, révélateur dans le domaine dans lequel je veux exercer.

En revanche, les choses ne se sont pas déroulées tout à fait comme prévues. Nous avons perdu notre confiance en Chainos au bout de quelques semaines suite à la non présence de la base de données. Cela a causé beaucoup de déception au sein de l'équipe et a mené à un retard d'un mois complet par rapport au projet. Madame Catherine ACHARD nous a donné la chance de travailler sur un projet similaire, dans le même esprit que celui de Chainos mais cette fois en parallèle avec l'ISIR, au lieu de Chainos.

Je me suis notamment impliqué dans les deux parties suivantes. La première est la création de la base de données, une étape essentielle pour un projet en *data science*. Cela m'a appris plusieurs choses : D'une part, se poser des questions avant de commencer le travail. À titre d'exemple, que sera l'entrée du modèle que je vais créer ? Lequel est plus bénéficiaire pour le résultat ? Classer les vidéos par gestes, par personne ou bien par vitesses ? D'autre part, l'importance de l'annotation et de la labélisation de la base de données. Durant le projet, nous avons eu besoin de faire des changements dans la base à plusieurs reprises. Cela n'a pas causé de problème car les changements étaient faciles à mettre en œuvre grâce au fait que nous pouvions créer des routines Python pouvant accéder facilement à ces vidéos avec de simples boucles for. La deuxième partie importante consistait à créer le modèle du CNN, une partie que j'ai travaillé en binôme avec Younes avec qui je travaille pour la première fois. Ayant plus de connaissances dans ce domaine, Younes m'a aidé à entrer plus rapidement dans le sujet, bien que nous ayons rencontré plusieurs difficultés, et au début nous avions de mauvais résultats. Cependant, avec l'aide de Madame ACHARD, nous avons pu surpasser ces problèmes et nous sommes arrivés aux résultats attendus par ce modèle. Cette partie m'a permis de découvrir la force des réseaux neuronaux convolutifs concernant le traitement d'image. En outre, j'ai découvert Google Colaboratory, OpenPose et la bibliothèque Python Keras [10].

Finalement, ce projet m'a exposé à plusieurs problèmes auxquels je pourrais faire face dans des projets de ce niveau. Le changement de collaborateur, la pandémie, différents caractères et méthodologie de travail au sein de l'équipe, etc. Tout cela m'a appris comment réagir devant de telles situations. Je remercie Madame ACHARD pour ses encouragements infinis et mes collègues avec lesquels nous avons pu passer toutes les difficultés après huit mois de travail en équipe.

5.4.4 Bilan de Najwa MOURS LI

J'ai choisi de prendre part à la réalisation de ce projet afin d'élargir mes compétences et connaissances sur le domaine de l'intelligence artificielle et plus particulièrement l'architecture des réseaux de neurones artificiels car cela m'aidera à poursuivre les recherches que je souhaite faire dans le domaine des neurosciences, et notamment modéliser des réseaux de neurones normaux et présentant des pathologies dans le but d'expliquer les mécanismes de neurophysiologie et ceux de certaines pathologies neurologiques encore inconnues comme le mécanisme de la migraine ou l'estimation de la dégénérescence neuronale.

Au cours de ce projet, j'ai fait face à certaines désillusions. En effet, les trois quarts de notre projet se sont concentrés sur la création d'une base de données qui devait nous être fournie, par conséquent l'étude et la réalisation d'ANN n'a été que survolé, malheureusement. De plus, le projet s'est révélé très centré sur la programmation

informatique, ce n'est pas ce qui me plaisait le plus dans un tel projet mais plutôt la résolution du problème de la variabilité morphologique des individus, qui apparemment n'était pas une priorité parmi nos objectifs.

Malgré tous ces points négatifs, j'en ressors grandi. La divergence d'opinion et la diversité des profils de notre équipe avaient l'air d'être des obstacles mais avec le temps nos différences ont fait notre force principale dans la réalisation du projet. Cela nous a permis de jouer sur les points forts de chacun (dans la répartition des tâches) et les points faibles de chacun (entraide et explications à ceux qui n'étaient pas aux mêmes niveaux de connaissance). Nous avons pu travailler en groupe après certains désaccord sans aucunes difficultés, ce qui est un point fort pour nous tous étant donné qu'en entreprise un tel professionnalisme n'est pas forcément appliqué. En outre, étant la personne la moins à l'aise en programmation du groupe, les autres membres m'ont beaucoup apporté en termes de connaissances en *deep learning* et en programmation avec Python. De mon côté, j'ai apporté aux autres membres du groupes mes connaissances médicales sur la variabilité des morphologies ainsi que mes compétences relationnelles et de médiation.

Je suis personnellement satisfaite des objectifs que nous avons atteints même si notre projet rendu est très différent de la proposition de projet initiale. D'un point de vu technique, la réalisation d'une base de donnée est un long processus que nous avons réussi ensemble et nous sommes arrivés à créer un CNN qui fonctionne. D'un point de vu managérial, nous avons appliquer des méthodes d'organisations et de travail agiles afin de booster notre productivité. Finalement, d'un point de vu humain, je sais me positionner et faire entendre mes idées dans un groupe, j'ai pu améliorer ma capacité de communication avec mes camarades, comprendre et accepter des idées opposées aux miennes.

Ce que je retiens de l'ensemble de ce projet est que les obstacles sont nombreux et parfois imprévisibles surtout dans ce domaine d'étude, cependant avec de la rigueur et des compromis, les applications de l'I.A. permettent de résoudre en mathématique, informatique ou médecine ce qui nous paraît encore impossible.

Ce projet de deuxième année du cycle d'ingénieur est le plus proche de notre réalité future dans le monde du travail. Voulant poursuivre ma carrière dans un laboratoire à l'étranger, je suis prête à effectuer de nombreux projets de recherche comme celui-ci, mais, par cette expérience, je saurais m'affranchir de certaines difficultés rencontrées durant ce projet.

Finalement, je remercie mes quatre camarades ayant travaillé avec moi, pour avoir mené à bien ce projet et le partage de nos connaissances et nos difficultés qui ont apporté à chacun ce qui lui manquait pour être un futur ingénieur accompli.

5.4.5 Bilan d'Arthur ZUCKER

Lorsque l'entreprise Chainos me présenta ce projet, il y a de cela presque un an, j'étais stagiaire à temps plein dans leur locaux. J'avais passé deux mois à travailler dans leur équipe sur des projets similaires, utilisant des bases de données afin d'entraîner une intelligence artificielle à résoudre une certaine tâche. Le bon déroulement de mon stage et de ces projets m'a donné confiance en cette entreprise et c'est pourquoi avec mon collègue Younes BELKADA nous avons décidé de proposer la réalisation de ce projet. J'avais envie, en guise de remerciement, de m'investir pour la réussite de ce projet. Les cours et l'expérience que j'ai reçus en travaillant dans l'équipe de Chainos me donnèrent beaucoup de confiance en moi. C'est pourquoi les premières semaines du projets se déroulèrent pour moi de manière naturelle et sans encombre.

Cependant, lorsque la communication avec l'entreprise est devenue de plus en plus faible, et après avoir appris l'inexistence de la base de donnée qui nous avait été promise, j'ai perdu toute motivation. J'aurais aimé démarrer un nouveau projet pour faire « page blanche » et repartir d'un bon pieds. Cependant, nous avons eu la chance de recevoir de l'aide de la part de l'ISIR en la personne de Catherine ACHARD. Elle nous a redonné un nouvel objectif et guidé à la réalisation de celui-ci. Ma contribution au projet s'est alors focalisée sur la création de la base de données et l'étude du code Matlab implémentant un *SVR* et utilisant la méthode des *N-best* pour l'estimation des poses. La pandémie de coronavirus s'ajoutant à tout cela m'a complètement empêché d'aller plus loin dans la

réalisation en réduisant le peu de motivation qu'il me restait. J'ai stoppé une grande partie des contacts avec mon équipe pendant quelques semaines, à la fois par culpabilité et par manque de volonté.

J'aurais grandement aimé recevoir plus de cours et plus de connaissances quant au fonctionnement et à la réalisation d'un réseau de neurones artificiels adapté à une tâche spécifique, mais j'aurais surtout aimé avoir une base de données prête et bien fournie. J'ai appris que ma naïveté pouvait me porter atteinte, et que malgré un manque d'investissement et de l'égoïsme, mes collègues ont su faire preuve d'indulgence envers moi, ce que j'ai beaucoup apprécié. Certes, la réalisation du projet a été mise en échec, mais je ne doute pas que l'ensemble des membres ayant participé saura trouver une nouvelle opportunité de travail en I.A. qui cette fois-ci sera mieux préparée par l'expérience acquise durant ce projet.

A. Extraits du cahier des charges

A.1 Expression fonctionnelle du besoin

Cette partie est un extrait du cahier des charges abordant l'expression fonctionnelle du besoin du projet remodelé. Pour plus d'informations concernant cette partie, nous conseillons le lecteur de se reporter au cahier des charges. La figure A.1 représente la raison d'être de l'objet d'étude du projet sous la forme d'un diagramme bête à corne. La figure A.2 illustre sous forme d'un diagramme pieuvre les fonctions et fonctions contraintes de l'objet.

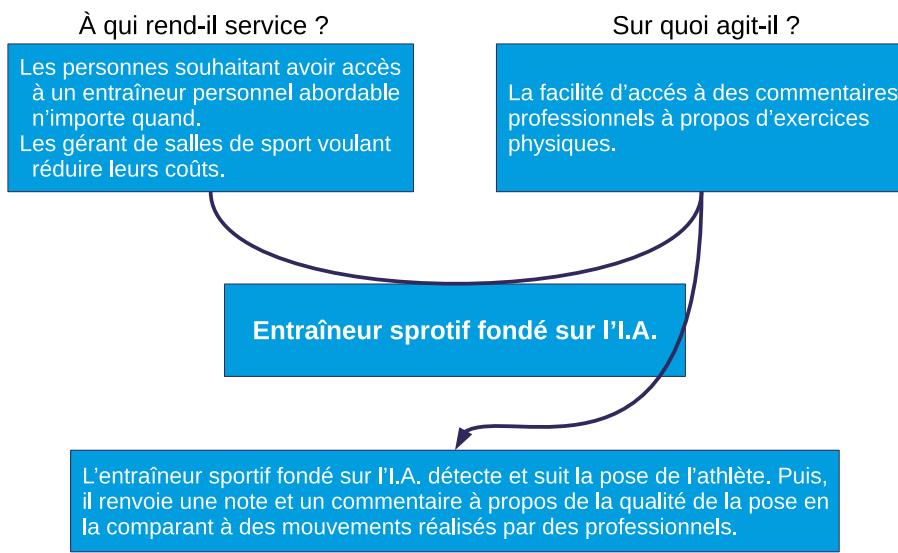


FIGURE A.1 – Bête à corne

A.1.1 Fonctions et contraintes

F1	Détecter, suivre, évaluer et classer correctement les mouvements de l'utilisateur selon un exercice.
F2	Retourner une note des mouvements réalisés par l'utilisateur.
F3	Donner un commentaire expliquant les erreurs commises par l'utilisateur dans le but qu'il améliore ses futurs mouvements.
C1	Détecter et analyser les mouvements rapidement et efficacement.
C2	Capturer avec précision les articulations de l'utilisateur uniquement.
C4	Gérer la confidentialité des vidéos enregistrées.

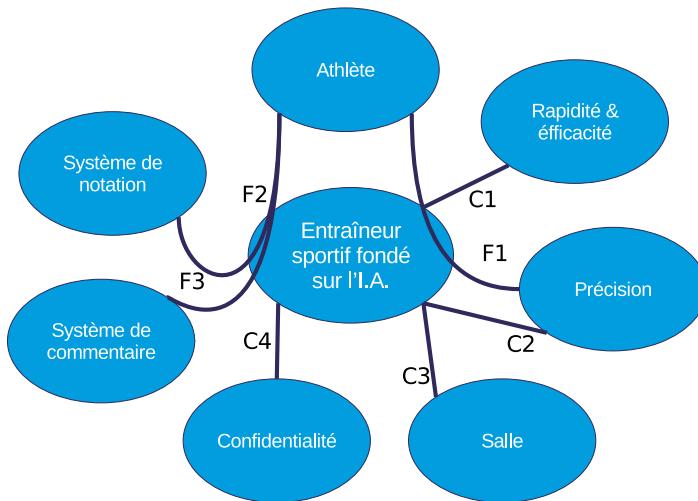


FIGURE A.2 – Diagramme pieuvre

A.1.2 Évaluation des exigences

Fonction	Niveau requis	Niveau désiré mais révisable	Mesure d'évaluation
F1	85%	99 %	Précision
F2	75%	95%	Corrélation des notes du système avec des notes de véritables entraîneurs sportifs
F3	50%	90%	Taux de pertinence du commentaire
C1	10 FPS	60 FPS	Nombre de FPS traités par notre programme
C2	95%	100%	Nombre de fois où toutes les articulations de l'utilisateur ne sont pas reconnues
C4	aucun problème	problèmes présents	Avec les lois de confidentialité (R.G.P.D., etc.)

A.2 Analyse des risques

Cette partie est un extrait de l'analyse des risques du projet initial. Le tableau A.1 est la copie du registre des risques initialement présent dans le cahier des charges. Il est suivi du tableau A.2 qui associe à chaque risque sa probabilité et son impact.

Id.	Description du risque	Criticité	Responsable	Prévention	Réparation
A	Accès non contrôlé à la base de données donc accessible à un utilisateur non autorisé	Haut	Abdallah DAABOUL	Contrôler l'accès à notre base de données avant de l'utiliser	Utiliser une base de données publique
B	Perte de performance du système due à un mauvais choix de puissance de calcul	Haut	Jérôme BONACCHI	Déterminer la puissance de calcul requise	Modifier l'algorithme pour qu'il consomme moins d'énergie
C	Clients non réactifs à ce produit	Haut	Chainos	Enquête réalisée par l'entreprise auprès de ses clients	Campagne marketing intensive

D	Impossible d'obtenir des fonds pour les caméras	Moyen – Haut	Arthur ZUCKER	Demande de fonds à plusieurs départements	Achat de caméras avec notre propre argent
E	Impossible d'obtenir des fonds pour les GPU	Moyen – Haut	Arthur ZUCKER	Demander des fonds à plusieurs départements	Utiliser des ordinateurs du <i>cloud</i>
F	Création d'une base de données par nos soins	Moyen – Haut	Arthur ZUCKER et Younes BELKADA	Recherche de bases de données adaptées	Automatisation de la création de la base de données
G	Le système de notation ne fonctionne pas	Haut	Abdallah DAABOUL et Najwa MOURSLI	Apprendre les méthodes de pointe sur ce sujet	Supprimer le système de notation
H	Organisation du travail sous-optimale	Moyen	Jérôme BONACCHI	Utilisation des outils de gestion de projet	Réduction de la portée du projet
I	Ne pas tenir compte complètement de la variabilité du corps humain	Moyen – Haut	Najwa MOURSLI	Normaliser les informations extraites des vidéos	Entraîner à nouveau notre algorithme
J	Notre jeu de données n'est pas assez large, ainsi la diversité des mouvements ne serait pas assez étendue, donc l'application ne reconnaîtrait pas certaines poses	Moyen – Haut	Younes BELKADA et Najwa MOURSLI	Utiliser un jeu de données très large	Indication de l'étendue du jeu de données
K	Le <i>feedback</i> est inappropriate	Moyen – Haut	Younes BELKADA, Jérôme BONACCHI et Abdallah DAABOUL	Demander à un vrai formateur un commentaire de chaque mouvement	Supprimer le système de <i>feedback</i>
L	Un membre de l'équipe est indisponible	Bas	Tous les membres	Travailler toujours en binômes	Répartir les tâches entre le reste du groupe
M	Le travail d'un membre de l'équipe n'est pas compréhensible	Moyen	Tous les membres	Toujours travailler en binômes et expliquer notre travail	Réviser l'ensemble du travail

Tableau A.1 – Registre des risques du cahier des charges

Probabilité/Conséquences	Négligeables	Mineures	Modérées	Majeures	Sévères
Très probable			D		A, B
Probable				K, J	C
Possible		M	H	F, J	E
Presque improbable		L			
Improbable					G

Tableau A.2 – Matrice de probabilité d'impact

Glossaire

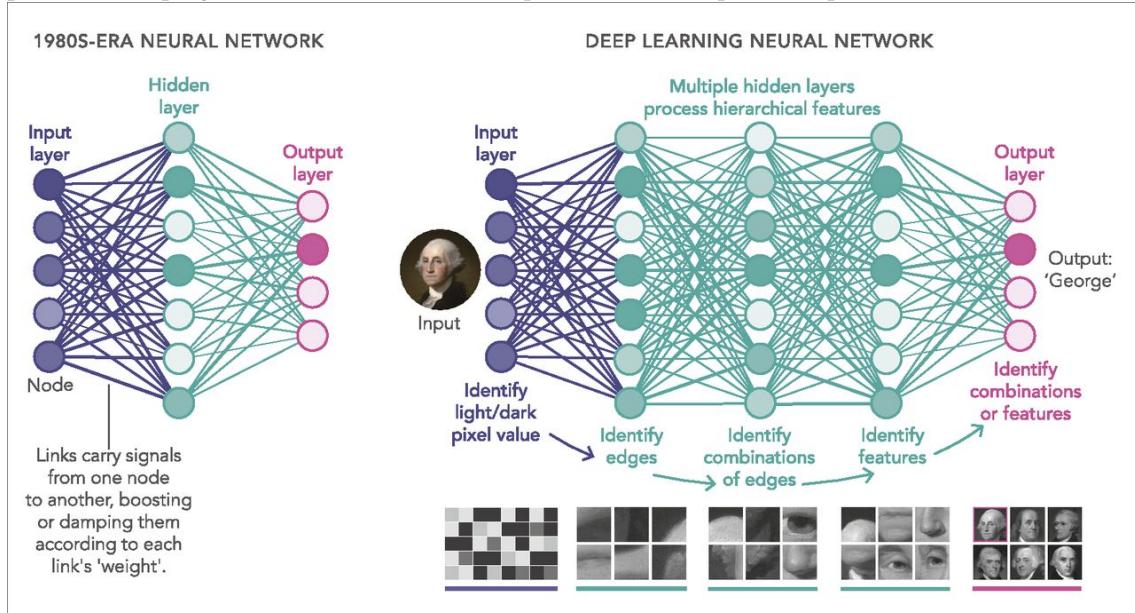
Application hybride Application s'appuyant sur des logiciels informatiques mis en œuvre sur de multiples plates-formes informatiques.

☞ Page 5.

Apprentissage automatique L'étude scientifique des algorithmes et des modèles statistiques que les systèmes informatiques utilisent pour effectuer une tâche spécifique sans utiliser d'instructions explicites, en se basant plutôt sur des modèles et des inférences. Cette discipline est considérée comme un sous-ensemble de l'intelligence artificielle. Les algorithmes d'apprentissage automatique construisent un modèle mathématique fondé sur des données d'échantillon, appelées « données d'entraînement », afin de faire des prédictions ou de prendre des décisions sans être explicitement programmés pour effectuer la tâche [25].

☞ Page 1.

Apprentissage profond Une classe d'algorithmes d'apprentissage automatique utilisant plusieurs couches pour extraire progressivement des caractéristiques de niveau supérieur à partir de d'une entrée brute.



Par exemple, en traitement d'images, les couches inférieures peuvent identifier les bords, tandis que les couches supérieures peuvent identifier les concepts pertinents pour un humain tels que les chiffres ou les lettres ou les visages. Les couches d'unité de traitement non linéaire sont utilisés pour l'extraction et la transformation des caractéristiques ; chaque couche prenant en entrée la sortie de la précédente.

☞ Page 5.

Caméra temps de vol Une caméra utilisant la vitesse de la lumière pour mesurer des distances en comptant efficacement le temps qu'il faut à un faisceau de lumière réfléchi pour revenir au capteur de la caméra.

☞ Page 5.

Convolutional neural network Traduit par réseau neuronal convolutif.

☞ Voir réseau neuronal convolutif.

☞ Pages 14, 18.

Deep learning Traduit par apprentissage profond.

☞ Voir apprentissage profond.

☞ Pages 1, 4, 5, 7, 28–31.

Epoch Traduit par itération du réseau de neurones artificiels. Cela correspond à une passe entière du jeu de données d'entraînement dans le réseau.

☞ Page 18.

Intelligence artificielle L'intelligence artificielle est l'étude des systèmes capables d'interpréter correctement des données externes, à tirer des enseignements de ces données et à utiliser ces enseignements pour atteindre des objectifs et des tâches spécifiques.

☞ Pages 1, 3–5, 18, 27, 29, 30.

Interface graphique L'interface graphique est un dispositif de dialogue entre l'humain et la machine. Cette interface se concrétise par une présentation d'un programme aisément compréhensible par les utilisateurs néophytes du système.

☞ Page 27.

Learning rate Traduit par coefficient d'apprentissage. C'est le coefficient utilisé pour la méthode de la descente de gradient du réseau de neurones artificiels.

☞ Page 18.

Long short-term memory Une variante améliorée et plus complexe de RNN utilisée en apprentissage profond.

☞ Page 23.

Machine learning Traduit par apprentissage automatique.

☞ Voir apprentissage automatique.

☞ Pages 1, 2.

Processeur graphique Une unité de traitement graphique est un circuit électronique spécialisé conçu pour manipuler et modifier rapidement la mémoire grâce à sa structure hautement parallèle.

☞ Page 28.

Python Un langage de programmation haut niveau couramment utilisé dans des projets en rapport avec l'intelligence artificielle (cf. www.python.org).

☞ Pages 4, 6, 11, 13, 14, 28, 30, 31.

Root mean squared error L'erreur quadratique moyenne est un estimateur statistique quantifiant la différence entre les valeurs expérimentales et les valeurs théorique de la quantité étudiée par la distance entre les différents points obtenus sur le graphique modélisant le phénomène considéré.

☞ Page 21.

Réseau de neurones artificiels Le réseau neuronal artificiel ou réseau neuronal simulé est un réseau interconnecté de neurones artificiels (ou naturels en neurosciences) qui utilise un modèle mathématique ou informatique pour le traitement de l'information basé sur une approche connectiviste du calcul.

☞ Pages 8, 13, 14, 16, 30, 32.

Réseau neuronal convolutif Une variante de réseau de neurones artificiels acyclique majoritairement utilisée pour le traitement d'images. La différence avec les réseaux de neurones artificiels basiques est l'ajout de couches convolutives qui transforment leurs propres entrées, d'une manière ou d'une autre, avant de les passer aux couches prochaines ; cela dans le but de simplifier l'image avant le traitement..

☞ Voir réseau de neurones artificiels.

☞ Page 30.

Sigles et acronymes

ANN *Artificial neural network.*

- ☞ Voir réseau de neurones artificiels.
- ☞ Page 30.

CNN *Convolutional neural network.*

- ☞ Voir *convolutional neural network.*
- ☞ Pages 8, 13–15, 18, 23–25, 30, 31.

CSV Correspond à un format de fichier pratique pour la structuration de données. Les entrées sont séparés d'un caractère (généralement la virgule) afin d'indiquer qu'on passe à la colonne suivante.

- ☞ Page 13.

FPS *Frames per Secons.*

- ☞ Voir *frames per second.*
- ☞ Page 34.

GPU *Graphics processing unit.*

- ☞ Voir *graphics processing unit.*
- ☞ Pages 7, 35.

I.A. Intelligence artificielle.

- ☞ Voir intelligence artificielle.
- ☞ Pages 1, 3, 5, 6, 29, 31, 32.

ISIR Institut des systèmes intelligents et de robotique.

- ☞ Pages 3, 4, 7, 11, 27–31.

J.O. Jeux olympiques.

- ☞ Pages 3, 7, 18.

JSON Correspond à un format de fichier qui permet de représenter de l'information structurée et de la stocker.

- ☞ Page 11.

LSTM *Long short-term memory.*

- ☞ Voir *long short-term memory.*
- ☞ Pages 8, 13, 14, 16, 18, 23–25.

MOOC Traduit par formation en ligne ouverte à tous.

- ☞ Pages 7, 29.

MSE Traduit par erreur quadratique moyenne.

- ☞ Page 18.

R.G.P.D. Réglementation générale sur la protection des données.

- ☞ Pages 6, 34.

RMSE Root mean squared error.

☞ Voir root mean squared error.

☞ Pages 21, 23.

RNN Recurrent neural network.

☞ Voir réseau de neurones artificiels.

☞ Page 16.

SVM Support vector machine.

☞ Voir support-vector machine.

☞ Pages 18, 19.

SVR Support vector regression.

☞ Pages 14, 18–20, 24, 25, 31.

Bibliographie

- [1] L. SACCO. (4 jan. 2019). L'IA s'attaque aux secrets de l'eau avec l'équation de Schrödinger, adresse : www.futura-sciences.com/sciences/actualites/physique-ia-attaque-secrets-eau-equation-schrodinger-10478/ (visité le 01/05/2020).
- [2] MATHWORKS. (2019). What Is Machine Learning? | How It Works, Techniques, adresse : <https://www.mathworks.com/discovery/machine-learning.html> (visité le 26/05/2019).
- [3] M. RODRIGUEZ. (28 mar. 2019). Latest IHRSA Data : Over 6B Visits to 39,570 Gyms in 2018, adresse : www.ihsra.org/about/media-center/press-releases/latest-ihsra-data-over-6b-visits-to-39-570-gyms-in-2018/ (visité le 13/10/2019).
- [4] M. SON. (19 juin 2019). Fitness Industry Grows as Vietnam Becomes Body Conscious, adresse : e.vnexpress.net/news/business/industries/fitness-industry-grows-as-vietnam-becomes-body-conscious-3940603.html (visité le 13/10/2019).
- [5] LIFEBEAM. (2017). Vi. The First True Artificial Intelligence Personal Trainer, adresse : www.kickstarter.com/projects/1050572498/vi-the-first-true-artificial-intelligence-personal/ (visité le 12/11/2019).
- [6] A. CROCHET-DAMAI. (15 oct. 2019). Pourquoi la moitié des projets IA échouent, adresse : www.journaldunet.com/solutions/dsi/1459042-ia-pourquoi-la-moitie-des-projets-echouent/ (visité le 30/04/2020).
- [7] K. WIGGERS. (8 juil. 2019). IDC : For 1 in 4 companies, half of all AI projects fail, adresse : <https://venturebeat.com/2019/07/08/idc-for-1-in-4-companies-half-of-all-ai-projects-fail/> (visité le 30/04/2020).
- [8] MATLAB version (R2019b), The Mathworks, Inc., Natick, Massachusetts, 2019.
- [9] M. ABADI, A. AGARWAL, P. BARHAM, E. BREVDO, Z. CHEN, C. CITRO, G. S. CORRADO, A. DAVIS, J. DEAN, M. DEVIN, S. GHEMAWAT, I. GOODFELLOW, A. HARP, G. IRVING, M. ISARD, Y. JIA, R. JOZEFOWICZ, L. KAISER, M. KUDLUR, J. LEVENBERG, D. MANÉ, R. MONGA, S. MOORE, D. MURRAY, C. OLAH, M. SCHUSTER, J. SHLENS, B. STEINER, I. SUTSKEVER, K. TALWAR, P. TUCKER, V. VANHOUCKE, V. VASUDEVAN, F. VIÉGAS, O. VINYALS, P. WARDEN, M. WATTENBERG, M. WICKE, Y. YU et X. ZHENG, *TensorFlow : Large-Scale Machine Learning on Heterogeneous Systems*, Software available from tensorflow.org, 2015. adresse : tensorflow.org.
- [10] F. CHOLLET et al., *Keras*, keras.io, 2015.
- [11] Z. CAO, G. HIDALGO, T. SIMON, S.-E. WEI et Y. SHEIKH, « OpenPose : realtime multi-person 2D pose estimation using Part Affinity Fields », *CoRR*, t. abs/1812.08008, 2018. arXiv : 1812.08008v2.
- [12] T. SIMON, H. JOO, I. MATTHEWS et Y. SHEIKH, « Hand Keypoint Detection in Single Images using Multiview Bootstrapping », *CoRR*, t. abs/1704.07809, 2017. arXiv : 1704.07809. adresse : arxiv.org/abs/1704.07809.
- [13] Z. CAO, T. SIMON, S.-E. WEI et Y. SHEIKH, « Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields », *CoRR*, t. abs/1611.08050, 2016. arXiv : 1611.08050v2. adresse : arxiv.org/abs/1611.08050v2.
- [14] S.-E. WEI, V. RAMAKRISHNA, T. KANADE et Y. SHEIKH, « Convolutional pose machines », *CoRR*, t. abs/1602.00134, 2016. arXiv : 1602.00134v4. adresse : arxiv.org/abs/1602.00134v4.
- [15] S. CHEN et R. YANG, « Pose Trainer : Correcting Exercise Posture using Pose Estimation », mar. 2018. doi : 10.13140/RG.2.2.29224.47367.
- [16] J. BÜTEPAGE, M. BLACK, D. KRAGIC et H. KJELLSTRÖM, « Deep representation learning for human motion prediction and classification », *CoRR*, t. abs/1702.07486, 2017. arXiv : 1702.07486v2. adresse : arxiv.org/abs/1702.07486v2.

- [17] Z. LI, Y. HUANG, M. CAI et Y. SATO, « Manipulation-skill Assessment from Videos with Spatial Attention Network », *CoRR*, t. abs/1901.02579, 2019. arXiv : 1901.02579v2. adresse : arxiv.org/abs/1901.02579v2.
- [18] H. PIRSIAVASH, C. VONDICK et A. TORRALBA, « Assessing the Quality of Actions », in *Computer Vision – ECCV 2014*, D. FLEET, T. PAJDLA, B. SCHIELE et T. TUYTELAARS, éd., Cham : Springer International Publishing, 2014, p. 556-571, ISBN : 978-3-319-10599-4.
- [19] P. PARMAR et B. T. MORRIS, « Action Quality Assessment Across Multiple Actions », *CoRR*, t. abs/1812.06367, 2018. arXiv : 1812.06367v2. adresse : arxiv.org/abs/1812.06367v2.
- [20] H. ISMAIL FAWAZ, G. FORESTIER, J. WEBER, L. IDOUMGHAR et P.-A. MULLER, « Evaluating Surgical Skills from Kinematic Data Using Convolutional Neural Networks », *Lecture Notes in Computer Science*, p. 214-221, 2018, ISSN : 1611-3349. DOI : 10.1007/978-3-030-00937-3_25. adresse : dx.doi.org/10.1007/978-3-030-00937-3_25.
- [21] H.-S. FANG, S. XIE, Y.-W. TAI et C. LU, « RMPE : Regional Multi-person Pose Estimation », in *ICCV*, t. abs/1612.00137, 2016. arXiv : 1612.00137v5. adresse : arxiv.org/abs/1612.00137v5.
- [22] M. MILLAN et C. ACHARD, « Segmenting Surgical Tasks using Temporal Convolutional Neural Network », in *Computer Assisted Radiology and Surgery*, 2019.
- [23] D. R. DENNIS PARK, « N-best maximal decoders for part models », p. 2627-2634, mar. 2011.
- [24] Y. YANG et D. RAMANAN, « Articulated pose estimation with flexible mixtures-of-parts », in *CVPR 2011*, 2011, p. 1385-1392.
- [25] (29 mai 2019). Apprentissage automatique – Wikipedia – Apprendre une langue étrangère, adresse : www.euro-langues.org/apprentissage-automatique-wikipedia-apprendre-une-langue-etrangere/ (visité le 01/05/2020).