



UNIVERSITÉ SULTAN MOULAY SLIMANE

FACULTÉ DES SCIENCES ET TECHNIQUES

CENTRE D'ÉTUDES DOCTORALES : SCIENCES ET TECHNIQUES

FORMATION DOCTORALE : MATHÉMATIQUES ET PHYSIQUES APPLIQUÉES

THÈSE

Présentée en vue d'obtenir le grade de Docteur,

DISCIPLINE : INFORMATIQUE

SPÉCIALITÉ : INFORMATIQUE

PAR :

YOUNES ED-DOUGHMI

CONTRIBUTION AUX DÉVELOPPEMENTS D'APPROCHES BASÉES SUR L'INTELLIGENCE ARTIFICIELLE POUR LA SÉCURITÉ ROUTIÈRE.

THÈSE SOUTENUE LE 23 MAI 2022 DEVANT LE JURY COMPOSÉ DE :

Président du jury:	Pr. Mohamed FAKIR	Faculté des Sciences et Techniques, Béni Mellal.
Directeur de thèse:	Pr. Rachid EL AYACHI	Faculté des Sciences et Techniques, Béni Mellal.
Rapporteurs:	Pr. Youssef ES-SAADY	Faculté Polydisciplinaire, Université Ibn Zohr, Taroudant.
	Pr. Mohammed ERRITALI	Faculté des Sciences et Techniques, Béni Mellal.
	Pr. Mohamed BASLAM	Faculté des Sciences et Techniques, Béni Mellal.
Examineurs:	Pr. Youssef EL MOURABIT	Faculté des Sciences et Techniques, Béni Mellal.

© 2022 - *Younes ED-DOUGHMI*
Tous droits réservés

RÉSUMÉ

Titre: CONTRIBUTION AUX DÉVELOPPEMENTS D'APPROCHES BASÉES SUR L'INTELLIGENCE ARTIFICIELLE POUR LA SÉCURITÉ ROUTIÈRE.

Le réseau routier présente la moyenne la plus fréquentée par la majorité écrasante de la population mondiale, en plus le réseau routier se développe d'une vitesse très rapide, ce qui oblige les personnes à passer plus de temps dans les routes, ainsi choisir la vitesse pour se rendre le plus vite possible d'un point à l'autre, cette nouvelle exigence implique malheureusement l'augmentation des incidents routiers. Il existe plusieurs raisons qui peuvent provoquer des incidents graves, ces incidents peuvent être classifiés en deux catégories, la première catégorie des incidents est liée au comportement humain, cette catégorie d'incident représente la grande majorité d'incident. La seconde catégorie des incidents est liée à l'état de la route ou/et l'état de la voiture.

La distraction au volant est un facteur important dans la conduite. Cependant, il est difficile de lutter contre les phénomènes naturels, qui s'aggravent avec la fatigue et sont compliqués à détecter. Les défenseurs de la prévention font une pause toutes les deux heures, mais il n'y a aucun moyen de vérifier que la règle s'applique. Concrètement, l'envie d'arriver vite prime sur le principe de précaution !

La supervision de la qualité de la route fait aussi partie des principaux axes de recherche dans le domaine de la sécurité routière, en exploitant les différentes techniques d'apprentissage automatique. Notre objectif est de proposer un système efficace et précis, afin de fournir un rapport complet et détaillé sur l'état de la route et d'anticiper la maintenance à effectuer sur la route, ou une aide à la décision si le système est utilisé en temps réel dans le cadre d'une voiture autonome par exemple.

Pour cela nous avons proposé des solutions qui se basent sur des architectures de l'intelligence artificielle, notamment les modèles d'apprentissage en profondeur, ce type de modèles à prouver sa performance sur des données hétérogène et volumineuse, ce qui correspond bien à notre contexte de recherche dans les domaines sécurité routière, tel que nos travaux se basent sur des caméras ou des images qui reflètent le contexte à analyser.

En conséquent, notre premier travail, c'est focaliser sur l'analyse des comportements des conducteurs sur la route, surtout que c'est devenu l'un des principaux sujets de

recherche de ces dernières années, en particulier la somnolence, car elle constitue le facteur le plus élevé d'accidents et la première source de décès sur les routes.

Cet article présente un moyen d'analyser et d'anticiper la somnolence d'un conducteur en appliquant un réseau neuronal récurrent sur une séquence d'images du visage du conducteur. Nous avons utilisé un ensemble de données pour façonner et approuver notre modèle et avons mis en œuvre une architecture de réseau neuronal récurrent basée sur un modèle réseau de neurones récurrents à mémoire court terme et long terme dans un premier travail, et dans un second travail, on a utilisé le modèle de multicouche de réseaux convolutifs 3D pour détecter la somnolence du conducteur. Après une session d'entraînement, nous avons obtenu une précision prometteuse qui approche un taux d'acceptation de 92%, ce qui a permis de développer un système de surveillance des conducteurs en temps réel pour réduire les accidents de la route.

Pour l'analyse de l'état de la route, on a utilisé une architecture de deep learning en deux étapes, la première étape consiste à lancer une Segmentation pour déterminer les côtés de la route et limiter la zone de détection des nids de poule, après la seconde étape est la détection des nids de poule qui sont incluse dans cette zone. Est en seconde partie une autre approche utilisée qui fusionne les deux architectures, cette approche se base sur une architecture Multinet, ce qui réduit le nombre de faux positifs détectés par le modèle, et réduit également les détections de nids de poule qui ne sont pas pertinents pour la détection, y compris les nids de poule sur le côté de la route ou sur le trottoir.

Pour cela, chaque modèle a été entraîné sur des images issues d'une base de données publique, le résultat de notre approche a montré son efficacité en atteignant une précision 93%.

Mots-clés: Sécurité routière, Détection de la Somnolence au volant, Nids-de-poule sur les routes, intelligence artificielle, réseaux neuronaux récurrents, Apprentissage profond, Réseau neuronal convolutif.

ABSTRACT

Title: ARTIFICIAL INTELLIGENCE TO HELP ROAD SECURITY.

The road network presents the most frequented average by the overwhelming majority of the world's population, in addition to the road network is developing at a very fast speed, which forces people to spend more time in the roads, so choose the speed to get as fast as possible from one point to another, this new requirement unfortunately implies the increase in road incidents. There are several reasons that can cause serious incidents, these incidents can be classified into two categories, the first category of incidents is related to human behavior, this category of incidents represents the vast majority of incidents. The second category of incidents is related to the condition of the road and/or the condition of the car.

Distracted driving is an important factor in driving. However, it is difficult to combat natural phenomena, which worsen with fatigue and are complicated to detect. Prevention advocates make a break every two hours, but there is no way to verify that the rule is being followed. In practice, the desire to arrive quickly takes precedence over the precautionary principle!

Supervision of road quality is also one of the main areas of research in the field of road safety, using various machine learning techniques. Our objective is to propose an efficient and accurate system, to provide a complete and detailed report on the state of the road and to anticipate the maintenance to be carried out on the road, or a decision support if the system is used in real time within the framework of an autonomous car for example.

For this, we proposed solutions that are based on artificial intelligence architectures, including deep learning models, this type of model to prove its performance on heterogeneous and large data, which corresponds well to our research context in the areas of road safety, such as our work is based on cameras or images that reflect the context to analyze.

Therefore, our first work is focused on the analysis of drivers' behaviors on the road, especially since it has become one of the main research topics recently, especially drowsiness, as it is the highest factor of accidents and the first source of death on the roads.

This paper presents a way to analyze and anticipate driver drowsiness by applying a recurrent neural network on a sequence of driver face images. We used a dataset to

shape and approve our model and implemented a recurrent neural network architecture based on a short-term and long-term memory recurrent neural network model in a first work, and in a second work, the 3D convolutional network multi-layer model was used to detect driver's drowsiness. After a training session, we achieved a promising accuracy approaching 92% acceptance rate, which allowed us to develop a real-time driver monitoring system to reduce traffic accidents.

For the analysis of the condition of the road, we used a deep learning architecture in two steps, the first step is to launch a Segmentation to determine the sides of the road and limit the area of detection of potholes, after the second step is the detection of potholes that are included in this area. In the second part another approach used that merges the two architectures, this approach is based on a Multinet architecture, which reduces the number of false positives detected by the model, and also reduces the detections of potholes that are not relevant for detection, including potholes on the side of the road or on the sidewalk.

For this, each model was trained on images from a public database, the result of our approach showed its effectiveness by reaching 93% accuracy.

Keywords: Driver fatigue detection, Drowsiness, Potholes on Roads, Road damage detection, Road quality, Computer vision, Artificial intelligence, Recurrent neural networks, Deep learning, Convolutional neural network.

A mes parents ..

C'est avec une joie immense et le cœur ému que je dédie cette thèse à ma chère mère et à mon défunt père, pour leurs affections inépuisables et leurs précieux conseils. Ils n'ont cessé de prier pour moi durant mon cursus scolaire et m'ont encouragé régulièrement.

À ma femme qui m'a soutenue et supporté la charge et la responsabilité de la petite famille, dans les périodes les plus difficiles de la préparation de ma thèse.

REMERCIEMENT

EN se basant sur mon expérience professionnelle comme atout, j'ai entamé une recherche scientifique dans le domaine de l'intelligence artificiel, afin d'apporter des solutions applicables dans le domaine de la sécurité routière, le fait de penser que l'un de mes études peut améliorer des dispositifs dans le but de sauver des vies me donne grande satisfaction.

Avant tout, je voudrais remercier mon directeur de thèse, *Professeur Rachid El Ayachi* pour sa patience et de croire en mon travail. Je le remercie aussi de m'avoir fait confiance et de m'avoir préparé toutes les bonnes conditions pour organiser ma thèse. Je lui suis également reconnaissant pour le temps qu'il m'a accordé, ses qualités pédagogiques et scientifiques, sa franchise et sa sympathie. Pour cela je lui adresse ma gratitude pour tout cela.

Je voudrais remercier le Président de cette thèse *Professeur Mohamed FAKIR* de la faculté des Sciences et Techniques de Béni Mellal, les rapporteurs de cette thèse *Professeur Youssef ES-SAADY* de Faculté Polydisciplinaire d'Université Ibn Zohr à Taroudant, *Professeur Mohammed ERRITALI* de la faculté des Sciences et Techniques de Béni Mellal, *Professeur Mohamed BASLAM* de la faculté des Sciences et Techniques de Béni Mellal et l'examineur de cette thèse *Professeur Youssef EL MOURABIT* de la faculté des Sciences et Techniques de Béni Mellal, pour l'intérêt qu'ils ont porté à mon travail et leur temps consacré à ma thèse.

TABLE DES MATIÈRES

RÉSUMÉ	iii
ABSTRACT	v
REMERCIEMENTS	ix
TABLE DES MATIÈRES	x
LISTE DES FIGURES	xiv
LISTE DES TABLEAUX	xvi
ACRONYMES	xvii
1 INTRODUCTION GÉNÉRALE	1
1.1 ANALYSE DE COMPORTEMENT HUMAIN	2
1.1.1 Approche basée sur l'étude de la somnolence	2
1.1.2 Conduite autonome	3
1.2 ANALYSE DE LA QUALITÉ DE LA ROUTE	3
1.3 OBJECTIF & CONTRIBUTIONS	4
1.4 STRUCTURE DU MANUSCRIT	4
2 ÉTAT DE L'ART	7
2.1 INTRODUCTION	8
2.1.1 Les actions à risque	10
2.1.2 Comment éviter la fatigue	10
2.2 DÉTECTION DE LA SOMNOLENCE	11
2.2.1 Détection de la fatigue du conducteur en fonction de l'état de l'œil	11
2.2.2 Estimation de la position de la tête du conducteur	12
2.2.3 Détection du comportement du conducteur par un système basé sur KINECT	12
2.2.4 Réseaux de neurones à apprentissage profond	12
2.2.5 Détection de la fatigue	12
2.3 ANALYSE DE L'ÉTAT DE LA ROUTE	14
2.3.1 Méthode manuelle (l'enquête visuelle)	15
2.3.2 Détection basée sur des captures	16

2.3.3	Détection par apprentissage profond	16
2.4	RAPPELLE SUR LES RNN	16
2.4.1	Définition	16
2.4.2	Réseaux de neurones	17
2.4.3	Comment les réseaux neuronaux récurrents apprennent	19
2.4.4	Unités de mémoire à long terme	19
2.5	CONCLUSION	20
3	DÉTECTION DE LA FATIGUE DES CONDUCTEURS EN SE BASANT SUR DES RNNs	21
3.1	INTRODUCTION	23
3.2	INFORMATIONS DE BASE	25
3.2.1	Apprentissage profond	25
3.2.2	Réseau de neurones récurrents	26
3.3	ÉTUDE BASÉE SUR L'ARCHITECTURE LSTM	27
3.3.1	Approche proposée	27
3.3.1.1	Réseaux de neurones récurrents (RNN)	27
3.3.1.2	Réseau de neurones récurrents à mémoire court-terme et long terme	28
3.3.1.3	Architectures	28
3.3.1.4	Ensemble de données	29
3.3.1.5	Entraînement	30
3.3.2	Discussion des résultats	31
3.3.3	Proposition de système	31
3.3.4	Conclusion	32
3.4	ÉTUDE BASÉE SUR L'ARCHITECTURE CONVOLUTIFS 3D	33
3.4.1	Approche proposée	33
3.4.1.1	Apprentissage des caractéristiques via des réseaux convolutifs 3D	33
3.4.1.2	Architectures multi-couches	34
3.4.1.3	Max-pooling : pourquoi ?	35
3.4.2	Expérimentations	38
3.4.2.1	L'ensemble de données NTHU-DDD	38
3.4.2.2	Entraînement	39
3.4.3	Analyse des résultats	40
3.4.3.1	Résultats de précision	40
3.4.3.2	Évolution de l'entraînement	41
3.4.3.3	Mesure de la performance	41
3.4.3.4	Comparaison avec d'autres méthodes	42
3.4.4	Système proposé	42

3.4.5 Conclusion	43
4 SYSTÈME DE DÉTECTION DES NIDS-DE-POULE ROUTIERS BASÉ SUR LE RÉSEAU NEURONAL RÉCURRENT	45
4.1 INTRODUCTION	46
4.2 INFORMATIONS DE BASE	47
4.2.1 Méthode Marocain : Le relevé visuel	47
4.2.2 Méthode Française : Le relevé visuel	47
4.2.3 Classification par apprentissage profond	48
4.2.4 Encodeur-décodeur convolutif profond	49
4.3 ARCHITECTURE EN TEMPS DIFFÉRER	49
4.3.1 Approche proposée	49
4.3.2 Expériences	50
4.3.2.1 L'ensemble de données	50
4.3.2.2 Entraînement	50
4.3.3 Analyse du résultat	51
4.3.4 Conclusions	51
4.4 ARCHITECTURE EN TEMPS RÉEL	52
4.4.1 Proposed approach	52
4.4.1.1 Detection	52
4.4.1.2 Segmentation	53
4.4.1.3 Architecture MultiNet	53
4.5 SYSTÈME PROPOSÉ	54
4.5.1 Workflow	54
4.5.1.1 Software architecture	55
4.5.2 Résultats de l'expérience	56
4.5.2.1 Ensemble de donnée	56
4.5.2.2 Entraînement	56
4.5.3 Result analysis	57
4.5.4 Conclusion	58
5 CONCLUSION ET PERSPECTIVES	59
A PUBLICATIONS	61
B TECHNIQUES D'APPRENTISSAGE	63
B.1 VALIDATION CROISÉE	63
B.2 TECHNIQUES D'APPRENTISSAGE	63
B.2.1 AdaBoost	64

B.2.2 Réseaux de neurones	66
BIBLIOGRAPHIE	67

LISTE DES FIGURES

2.1	Action Réalisé pour la sécurité routière.	8
2.2	Développement durable de la sécurité routière	9
2.3	Méthodes de détection des nids de poule.	14
2.4	Réseau de neurones artificiels (ANN).	18
2.5	Structure interne d'un nœud de calcul.	18
3.1	Nombre de décès liés aux accidents de la route dans le monde.	23
3.2	Étude de la de la somnolence au volant sur une population marocaine.	24
3.3	Réseau de neurones récurrents et le déroulement dans le temps.	27
3.4	Long Short-term Memory Cell	28
3.5	LSTM composite model.	29
3.6	Différentes séquences de classe.	30
3.7	Résultat de l'entraînement de modèle LSTM.	32
3.8	Architecture logicielle d'une application mobile et d'un serveur web	33
3.9	Les opérations de convolution en 3D.	34
3.10	Proposition d'architectures multicouches	35
3.11	Groupage avec un fond blanc et un objet noir.	37
3.12	Groupage avec un fond noir et un objet blanc.	37
3.13	Exemples des deux classes, normale et somnolente, de l'ensemble de données NTHU-DDD.	38
3.14	Application des filtres formés de la couche conv4b de notre modèle.	39
3.15	Résultat de l'entraînement de ConvNets 3D.	41
3.16	Architecture logicielle pour une application mobile et un serveur web.	43
4.1	Les conditions d'exploitation en fonction des objectifs de l'étude.	48
4.2	Vue globale de l'architecture cnn.	48
4.3	Encodeur-Décodeur à convolution.	49
4.4	Proposition d'architecture pour la détection des nids de poule	50
4.5	Architecture MultiNet.	52
4.6	Diagramme de flux de travail	54
4.7	Système proposé.	55
4.8	Visualisation du résultat de la sortie.	57

- B.1 A gauche : l'ensemble d'apprentissage S et le sous-ensemble S_1 (points entourés). A droite : l'ensemble S_1 et la droite C_1 apprise sur cet ensemble. . . . 65
- B.2 A gauche : l'ensemble $S - S_1$ et la droite C_1 apprise sur S_1 . A droite : un ensemble S_2 inclus dans $S - S_1$ parmi les plus informatifs pour C_1 (Points entourés) . . 65
- B.3 A gauche : l'ensemble S_2 et la droite séparatrice C_2 apprise sur cet ensemble. Au centre : l'ensemble $S_3 = S - S_1 - S_2$ et la droite séparatrice C_3 apprise sur cet ensemble. A droite : l'ensemble S et la combinaison des 3 droites séparatrices apprises sur cet ensemble 65
- B.4 Réseaux de neurones 66

LISTE DES TABLEAUX

3.1	Résultats expérimentaux sur l'ensemble de données de la somnolence.	31
3.2	Détection de la somnolence au volant sur l'ensemble de données NTHUDDD.	40
3.3	Résultats des expériences F_1 Score.	42
3.4	Comparaison des 3D ConvNets et de la solution de base sur les scénarios de l'ensemble de données sur la détection de la somnolence au volant.	42
4.1	Résultats expérimentaux du calcul.	51
4.2	Résultats des expériences des méthodes appliquées.	57

ACRONYMES

SVM	: Support Vector Machines	Machines à Vecteur Support
RNN	: Recurrent Neural Networks	Réseau de neurones récurrents
SRN	: Simple Recurrent Networks	Réseaux récurrents simples
LSTM	: Long Short-Term Memory cell	Réseau de neurones récurrents à mémoire court-terme et long terme
3D ConvNets	: 3 Dimention Convolutional Networks	3 Dimensions de Neurones Convolutifs
ROI	: 3 Region Of Interest	Région d'intérêt
IOU	: 3 Intersection Over Union	Intersection Sur Union
API	: 3 Application Programming Interface	interface de programmation d'application
ADAS	: Advanced Driver-Assistance Systems Machines	Système d'Aide à la Conduite

SOMMAIRE

1.1 ANALYSE DE COMPORTEMENT HUMAIN	2
1.1.1 Approche basée sur l'étude de la somnolence	2
1.1.2 Conduite autonome	3
1.2 ANALYSE DE LA QUALITÉ DE LA ROUTE	3
1.3 OBJECTIF & CONTRIBUTIONS	4
1.4 STRUCTURE DU MANUSCRIT	4

LA sécurité routière constitue un enjeu de société majeur tant par le nombre des victimes tuées, blessées ou handicapées à vie – treize personnes sont tuées chaque jour sur les routes dans notre pays -, que par le poids du coût économique des accidents de la route pour la société dans son ensemble – des milliards d’euros chaque année.

Selon l’OMS, chaque année, dans le monde, plus d’un million de personnes sont tuées et 54 millions blessés dans un accident de la route. La route était la troisième cause de mortalité au monde est en passe de devenir la seconde avec cette particularité de toucher en priorité les jeunes puisque c’est chez les moins de 25 ans (tranche d’âge 0-24 ans) que se rencontre le plus grand nombre des victimes. Entre quatorze et dix-huit ans, un décès sur trois est dû à un accident de la route. Le cyclomoteur en est la cause principale.

En effet, depuis quelques années, la sécurité routière est entrée dans les préoccupations des Marocains et des efforts importants ont été entrepris pour diminuer le nombre d’accidents sur les routes, sécuriser les véhicules et l’infrastructure, favoriser la prévention et l’éducation routière.

1.1 ANALYSE DE COMPORTEMENT HUMAIN

Depuis une dizaine d'années, l'industrie automobile est en pleine transformation et les véhicules ne sont plus considérés comme des systèmes thermomécaniques contrôlés avec quelques composants électroniques. Les véhicules d'aujourd'hui sont des systèmes complexes dont les fonctions les plus importantes sont contrôlées par des réseaux d'ordinateurs. Plusieurs facteurs ont contribué à ce virage de l'industrie automobile tel que : l'augmentation des coûts d'essence et de la pollution des véhicules, les menaces de sécurité potentielles, la haute densité d'automobiles.

Des caméras, installées dans l'habitacle du véhicule, filment le conducteur pendant que des algorithmes analysent les images à la recherche de signes éventuels qui pourraient conduire à un accident. Une fois les signes correspondants repérés, la voiture prend le contrôle du véhicule ou alerte le conducteur. Cependant, le contrôle du véhicule peut simplement être repris par le conducteur, une fois que les signes de distraction traduisant une fatigue excessive ont cessé.

Le système peut reconnaître quatre sources différentes de distraction : écrire des SMS, appeler, faire un malaise, boire ou manger au volant et enfin, chercher un objet sous le siège passager ou la banquette arrière. De multiples heures de vidéos ainsi que des images ont été nécessaires afin d'entraîner les algorithmes à ces détections. Il a fallu filmer dans le but d'extraire des photos d'un comportement normal d'un automobiliste puis la même chose avec des comportements « anormaux ». Grâce à cela, l'algorithme est devenu de plus en plus capable de faire la différence entre un comportement normal de l'automobiliste et les moments où il est distrait.

1.1.1 Approche basée sur l'étude de la somnolence

La somnolence au volant est devenue la première cause d'accident mortel sur les réseaux autoroutiers. Dangereuse et sournoise, nous n'en prenons en général conscience que trop tard alors que chacun d'entre nous est concerné, quel que soit le lieu et l'instant.

Ainsi, contrairement aux idées reçues, lutter est impossible : ses effets sont imparables, son action inéluctable. Le seul remède efficace, c'est de faire des pauses régulièrement et de prendre le temps de se reposer.

Tout le monde s'accorde sur ses liens à une fatigue générale. Mais, la somnolence peut attaquer par surprise, malgré les précautions. Elle se révèle particulièrement sournoise parce qu'elle arrête le mouvement, elle déconnecte de la relation aux autres, voire à son propre corps, tel que le corps se laisse aller à s'endormir pour répondre à un réel besoin.

1.1.2 Conduite autonome

Les systèmes d'aide à la conduite (ADAS pour Advanced Driver-Assistance Systems) font partie intégrante du monde automobile. Qu'il s'agisse du régulateur automatique de distance, du système d'aide au maintien dans la voie, des systèmes de freinage d'urgence, de la caméra de recul ou de l'aide au stationnement, ces technologies augmentent en même temps le confort de conduite et la sécurité routière. Ce faisant, les systèmes ADAS soutiennent la conduite à différents stades technologiques – de la conduite assistée à la conduite semi-automatique, jusqu'à la conduite autonome. Cependant, le chemin reste pavé d'obstacles avant d'atteindre cet objectif, notamment des défis techniques.

Pour exploiter l'intelligence fournie par les capteurs, les méthodes d'intelligence artificielle (IA) telles que le machine learning (apprentissage machine), l'apprentissage profond et les réseaux neuronaux sont de plus en plus plébiscités. Grâce à des algorithmes avancés, ces technologies sont capables d'analyser de grandes quantités de données et d'en tirer des enseignements. Les systèmes sont entraînés, c'est-à-dire qu'ils apprennent des modèles qui se répètent dans les données des capteurs et en déduisent certaines informations de manière autonome. Intégrées dans les solutions ADAS, ces technologies d'intelligence artificielle peuvent déjà assurer diverses fonctions d'assistance automatisée. Les systèmes de contrôle automatique de la distance ou les assistants au freinage d'urgence sont, par exemple, capables d'identifier de façon fiable une grande variété d'obstacles tels qu'un véhicule devant, des cyclistes ou des piétons sur la route. Cela permet au véhicule de réagir automatiquement et rapidement pour éviter une collision.

1.2 ANALYSE DE LA QUALITÉ DE LA ROUTE

La sécurité routière fait partie des problèmes mondiaux qui tentent de s'améliorer, les investissements, les lois et les contrôles changeant avec l'évolution rapide du transport mondial.

En outre, la détection des parties endommagées et fissurées de l'infrastructure dépendait uniquement de l'expertise de l'équipe d'ingénierie pour l'expérimentation. Cependant, avec la demande croissante d'analyse de l'état des routes, les rapports d'analyse de l'équipe d'ingénieurs consomment plus de temps et de ressources financières. Par conséquent, les systèmes d'analyse automatique des routes apparaissent comme une alternative ou un outil utile pour accélérer le processus d'inspection des routes et fournir des rapports d'analyse plus pertinents en un temps record.

1.3 OBJECTIF & CONTRIBUTIONS

À la lumière de tous ces drames liés à la fatigue ou la somnolence sur la route, nous avons trouvé que la fait de données de temps à ce type de sujet peut sauver des vies, ce que nous a donné une grande motivation pour contribuer au sujet de sécurité routière.

La première contribution, présenter un moyen d'analyser et d'anticiper la somnolence d'un conducteur en appliquant un réseau neuronal récurrent sur une séquence d'images du visage du conducteur.

Après une deuxième contribution est liée à l'analyse de la qualité de route, tel que nous avons proposé une solution qui peut améliorer l'entretien des routes et donner une estimation objective du budget alloué.

1.4 STRUCTURE DU MANUSCRIT

Le plan de cette thèse est structuré autour de six chapitres. Dans le présent chapitre, on va présenter le contexte général du système d'alerte Intelligent, puis on a motivé la nécessité des systèmes intelligents pour le système d'alertes au conducteur en cas de fatigue par la présentation des données concernant l'évolution de taux d'incident au niveau mondial et national.

Le deuxième chapitre est consacré à l'état de l'art sur les méthodes de prédictions et de classification, la détection des actions et la reconnaissance du comportement humaines. Ainsi, l'état de l'art sur les différentes techniques de l'intelligence artificiel et de l'apprentissage profond en mode séquentiel, afin de détecter un comportement de la somnolence du conducteur, on présente notre première contribution avec un article portant sur l'analyse de la somnolence du conducteur.

Le troisième chapitre est la description de notre contribution, tel que, notre travail, c'est focaliser sur l'analyse et la détection des actions et la reconnaissance du comportement humaines. En deuxième temps de notre travail, c'est focaliser sur le temps réel de la détection de la somnolence, est l'anticipation via des alertes, afin d'alerter au bon moment le conducteur, pour cela une nouvelle architecture est proposée dans cette contribution qui nous a permis d'améliorer la performance de taux de détections de la somnolence.

Avec un changement de sujet, le quatrième chapitre est consacré à l'analyse de l'état de la route, spécifiquement la détection des nids de poule, pour cela notre première contribution, est de réaliser avec un article sur la détection des nids de poule en temps déferée, cette contribution nous a permise de proposer une solution qui donnée une estimation globale sur l'état de route. Ainsi, on montre l'application en temps réel de la détection des nids de poule afin de l'associer à des voitures soit pour analyser la route

ou anticiper la présence des nids de poule sur la route afin que le conducteur puisse les éviter.

La conclusion générale du manuscrit fait le bilan de nos contributions, des difficultés rencontrées et propose nos perspectives de travail.

Enfin, dans l'annexe A, présentes les références des travaux publiés pendant la durée de notre formation doctorat. Aussi dans l'annexe B, nous présentons en détail le principe des techniques d'apprentissage utilisées dans nos travaux ainsi que le principe de la validation croisée.

SOMMAIRE

2.1 INTRODUCTION	8
2.1.1 Les actions à risque	10
2.1.2 Comment éviter la fatigue	10
2.2 DÉTECTION DE LA SOMNOLENCE	11
2.2.1 Détection de la fatigue du conducteur en fonction de l'état de l'œil	11
2.2.2 Estimation de la position de la tête du conducteur	12
2.2.3 Détection du comportement du conducteur par un système basé sur KINECT	12
2.2.4 Réseaux de neurones à apprentissage profond	12
2.2.5 Détection de la fatigue	12
2.3 ANALYSE DE L'ÉTAT DE LA ROUTE	14
2.3.1 Méthode manuelle (l'enquête visuelle)	15
2.3.2 Détection basée sur des captures	16
2.3.3 Détection par apprentissage profond	16
2.4 RAPPEL SUR LES RNN	16
2.4.1 Définition	16
2.4.2 Réseaux de neurones	17
2.4.3 Comment les réseaux neuronaux récurrents apprennent	19
2.4.4 Unités de mémoire à long terme	19
2.5 CONCLUSION	20

LE phénomène des accidents de la circulation routière constitue un fléau social qui exige la mobilisation des pouvoirs publics en vue d'en atténuer les conséquences. La lutte contre l'insécurité routière doit revêtir un caractère prioritaire, animé par une détermination sans faille pour espérer la réduction du nombre des accidents et des victimes de la route, une modification des comportements délictueux et l'émergence d'une nouvelle culture en matière de sécurité.

2.1 INTRODUCTION

Chaque année, près de 1,3 million de personnes meurent des suites d'une collision – soit plus de 3 000 morts par jour – et plus de la moitié d'entre elles sont des usagers de la route vulnérables : piétons, cyclistes et motocyclistes. Les collisions font par ailleurs entre 20 et 50 millions de blessés graves et sont une cause importante de handicap dans le monde entier. Quarante-vingt-dix pour cent des décès dus à des collisions se produisent dans les pays à revenu faible ou intermédiaire, où se trouve environ 54% du parc automobile mondial. À moins que l'on agisse efficacement et sans plus attendre, les collisions deviendront la septième cause de décès dans le monde d'ici à 2030 et feront quelque 1,9 million de morts par an.

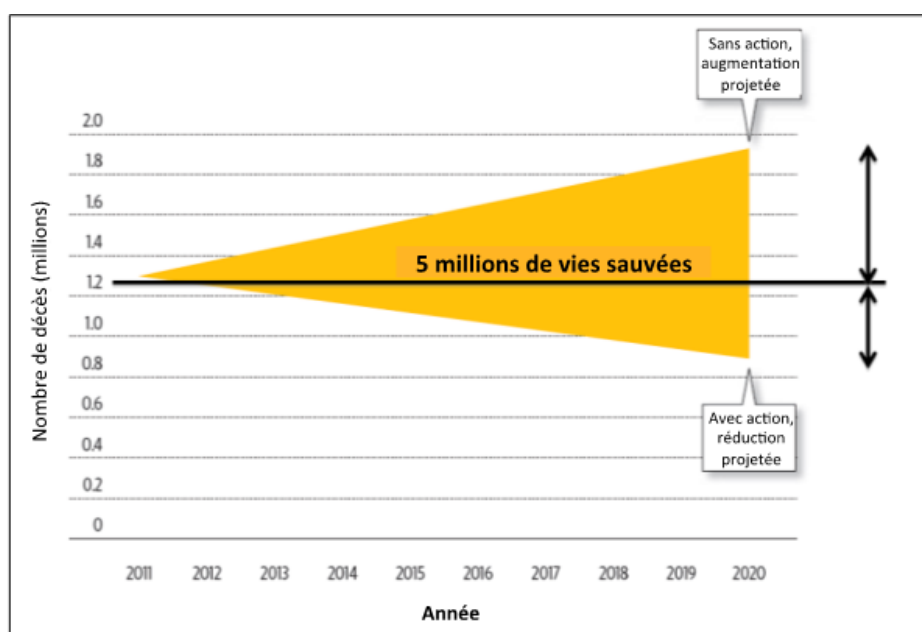


Figure 2.1: Action Réalisé pour la sécurité routière.

Dans sa résolution 64/255, intitulée « Amélioration de la sécurité routière mondiale », l'Assemblée générale des Nations Unies a proclamé la décennie 2011-2020 Décennie d'action pour la sécurité routière, en vue de stabiliser puis de réduire le nombre prévu de décès imputables aux accidents de la route dans le monde en multipliant les activités menées aux niveaux national, régional et mondial.

Dans le droit fil de ces visées ambitieuses, l'ONU a endossé un rôle de chef de file en établissant des normes à l'intention des États Membres, en les sensibilisant au problème et en leur apportant une assistance technique par l'intermédiaire de ses organismes ou commissions régionales ainsi que dans le cadre de ses partenariats internationaux. Afin d'obtenir un engagement politique encore plus soutenu en faveur de la sécurité routière, le Secrétaire général a nommé, en avril 2015, un Envoyé spécial pour la sécurité routière.

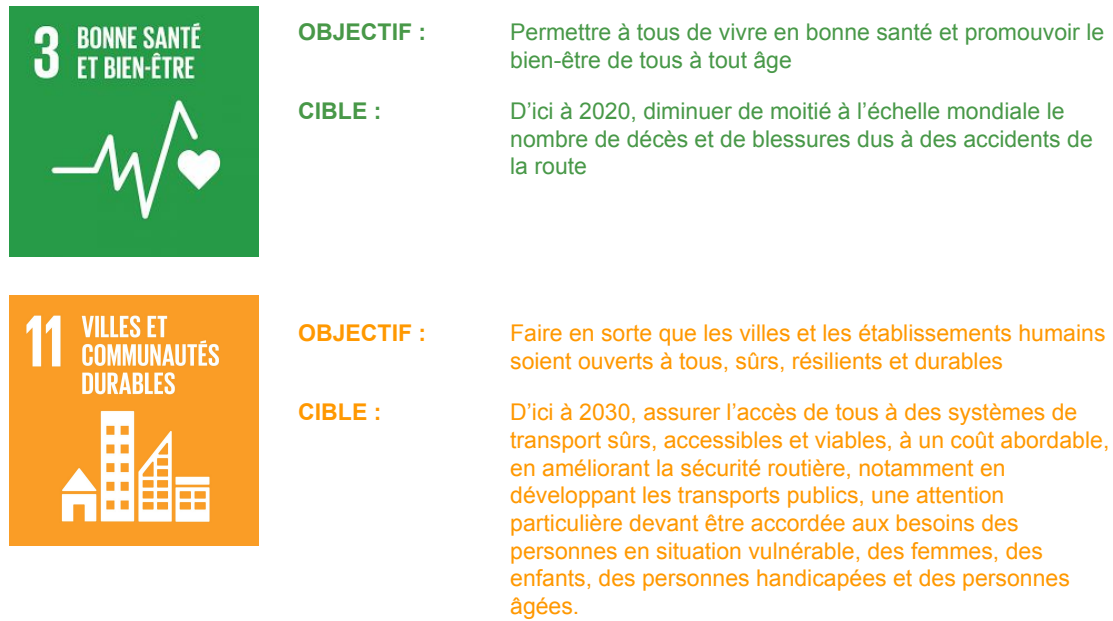


Figure 2.2: Développement durable de la sécurité routière

A l'heure où des millions d'automobilistes prennent la route des vacances et où les sorties nocturnes vont s'intensifier, il est essentiel de rappeler à quel point la fatigue et la somnolence au volant sont des facteurs importants d'insécurité routière.

En effet, des pratiques à risque persistent : 50 % des usagers réduisent leur temps de sommeil habituel au moment des départs en vacances, bien souvent pour avancer l'heure du départ. De plus, 12 % des automobilistes, notamment des jeunes conducteurs, dorment moins de cinq heures la veille d'un départ en vacances ou prennent la route en soirée, sans avoir dormi, ni fait la sieste.

Fatigue et somnolence au volant sont confondues, à tort ! Car chaque état appelle un comportement adapté. Comme le rappelle les spécialistes du sommeil, la fatigue est un « processus cumulatif entraînant des difficultés croissantes à poursuivre la conduite automobile et allant jusqu'à une baisse des performances » - Le traitement consiste alors à s'arrêter. Tandis que la somnolence est la « difficulté à se maintenir éveillé même lors de la conduite automobile » et le traitement consiste alors à dormir.

Pour distinguer fatigue et somnolence, un ensemble de signes précurseurs indique au conducteur qu'il est en train de perdre ses capacités d'attention et qu'il entre dans une phase de fatigue et/ou d'endormissement. La lenteur de réaction, les difficultés à maintenir une vitesse constante, les erreurs de coordination ou l'inattention à la signalisation sont des éléments constitutifs d'une baisse de vigilance.

- Les signes de la fatigue : raideurs dans la nuque, douleurs dans le dos et une certaine fixité du regard doivent alerter le conducteur.

- Les signes de la somnolence : bâillements, paupières lourdes, périodes d'absence ou désir de changer fréquemment de position révèlent les difficultés à se maintenir éveillé. En outre, la somnolence entraîne des périodes de « micro-sommeils » (de 1 à 4 secondes) pouvant être extrêmement dangereuses pour la sécurité de tous.

2.1.1 Les actions à risque

Un accident sur cinq est dû à l'endormissement au volant sur autoroute. Conduire en étant somnolence multiplie par huit le risque d'avoir un accident corporel. Un automobiliste qui prend la route en ayant dormi cinq heures ou moins la veille de son départ a 3 fois plus de risques d'avoir un accident qu'un conducteur reposé. Prendre la route avec un manque de sommeil ou rester volontairement éveillé longtemps afin de parcourir plus de kilomètres, provoquent les mêmes effets négatifs sur les capacités du conducteur que la présence d'alcool dans le sang : 17 heures de veille active équivalente à 0,5 g d'alcool par litre de sang.

2.1.2 Comment éviter la fatigue

Prendre sa voiture pour se rendre sur son lieu de vacances (par exemple), cela se prépare. Et ce qui est vrai du véhicule (freins, pneus...) doit également l'être du conducteur qui avant, comme pendant le trajet, doit respecter un certain nombre de consignes favorisant sa vigilance.

- Privilégier des horaires de conduite où le risque d'accident est plus faible. Il existe des périodes au cours d'une journée plus ou moins influentes sur la perte de vigilance. Ainsi, les études montrent que pour une personne « vierge de maladies », le risque de somnolence augmente sensiblement entre 13 h et 16 h, et encore plus fortement entre 2 h et 5 h. Dans le cas d'un voyage s'effectuant en soirée, il est ainsi recommandé de s'arrêter de conduire avant minuit et de ne pas reprendre le volant avant 6 h du matin. De manière générale, la conduite de nuit favorise la somnolence. L'alternance et la succession des phares de voiture venant à contresens, la réduction du champ de vision, la mauvaise appréciation des distances accélèrent par ailleurs le processus de fatigue.
- Avant le départ
Si le conducteur prend la route le matin, il est important de privilégier un petit-déjeuner consistant accompagné d'une boisson stimulante (thé, café) à dose modérée. En revanche, si le départ est prévu après le déjeuner, le repas doit être léger et sans alcool. Ne pas se fixer d'heure d'arrivée, pour ne pas être tenté d'accélérer. Le voyage en sera d'autant plus serein et agréable. Se renseigner sur les conditions de trafic en consultant le site www.admtrafic.ma.

- Au cours du trajet

Pendant le voyage, il est nécessaire d'adopter les bons réflexes pour écarter tout risque de somnolence :

- effectuer des pauses fréquentes d'une durée de 15 à 20 minutes au minimum toutes les deux heures.
- aérer le plus souvent possible le véhicule, et régler correctement la ventilation et/ou la climatisation.
- boire régulièrement de l'eau ou toute boisson non alcoolisée.
- respecter les règles élémentaires de conduite pour éviter les risques de fatigue et de somnolence : ne pas fixer son regard sur le véhicule qui précède, mais bien regarder dans ses rétroviseurs et faire attention à la signalisation ; respecter les limitations de vitesse et les distances de sécurité.

2.2 DÉTECTION DE LA SOMNOLENCE

Dans cette section, nous présentons un aperçu des études antérieures menées sur la détection de la somnolence. Diverses méthodes ont été proposées pour alerter ou détecter certains signes de somnolence. Ces méthodes sont basées sur le comportement du conducteur, son apparence ou les deux, produisant un résultat précis au temps t pour la plupart de ces méthodes. Cependant, nous ne savons pas à quoi ce comportement est lié dans le cas de la somnolence à t_{-1} par rapport à un cas de comportement normal du conducteur. Pour résoudre ce problème, nous avons utilisé les modèles de calcul RNN, qui intègrent le paramètre temps (t) dans leur complexité de calcul.

2.2.1 Détection de la fatigue du conducteur en fonction de l'état de l'œil

Études menées sur ce type de problème [Kurylyak, Lamonaca, and Mirabelli, 2012; Al-Rahayfeh and Faezipour, 2013; Mandal et al., 2017] présentent en général un système basé sur la vidéosurveillance via une caméra placée devant le conducteur pour détecter et calculer la fréquence des clignements d'yeux et contrôler le niveau de fatigue. Ces études ont utilisé une cascade de classificateurs basée sur Adaboost pour une détection rapide de la zone oculaire.

En résumé, cette technique utilise un descripteur de type Haar [Lienhart and Maydt, 2002] et un algorithme de classification AdaBoost [Freund, Schapire, et al., 1996] pour la détection du visage et des yeux en utilisant le pourcentage de fermeture des yeux (PERCLOS) pour évaluer la fatigue du conducteur. Le PERCLOS évalue la proportion du temps total pendant lequel les paupières des conducteurs sont $\geq 80\%$ fermées et reflète la fermeture lente des paupières plutôt que le clignement.

2.2.2 Estimation de la position de la tête du conducteur

Le système distinct d'estimation de l'orientation de la tête est une technique qui induit l'orientation de la tête par rapport à la vue d'une caméra : [Alioua et al., 2016]. Ainsi, l'objectif est détecté si le conducteur penche la tête en avant, ce qui indique une perte de conscience.

Pour cela, un nouveau descripteur a été utilisé dans cette étude, obtenu en fusionnant quatre descripteurs de l'orientation de la tête les plus pertinents : les filtres orientables, l'histogramme des gradients orientés, les caractéristiques Haar et une version adaptée du descripteur Speeded-Up Robust-Features. Ces caractéristiques ont ensuite été utilisées dans l'algorithme de classification par machine à vecteurs de support : [Burges, 1998].

2.2.3 Détection du comportement du conducteur par un système basé sur KINECT

La solution proposée est un système qui combine la détection de la fatigue et la détection du comportement indésirable du conducteur pendant la conduite [Solomon and Wang, 2015]. Ensuite, le système active un signal en utilisant la boîte à outils du développeur Kinect pour Windows [Zhang, 2012].

Les récentes avancées en matière de caméras 3D de profondeur, comme les capteurs Kinect de Microsoft, ont créé de nombreuses possibilités pour un passionné ou un développeur professionnel de détecter des postures simples, de suivre des gestes, de calculer les positions relatives et les angles entre les articulations d'une personne, et de les convertir dans un format que les développeurs peuvent exploiter pour créer de nouvelles expériences.

2.2.4 Réseaux de neurones à apprentissage profond

L'apprentissage profond est une tendance croissante dans l'analyse générale des données et a été qualifié de l'une des 10 technologies de rupture de 2018 [Rev., 2018].

L'apprentissage profond est une amélioration des réseaux neuronaux artificiels, et consiste en un plus grand nombre de couches qui permettent des niveaux d'abstraction plus élevés et de meilleures prédictions à partir des données [Schmidhuber, 2015]. [Goodfellow et al., 2016]. L'apprentissage profond est encore en train de devenir le principal outil d'apprentissage automatique dans les domaines de l'imagerie générale et de la vision par ordinateur.

2.2.5 Détection de la fatigue

L'objectif des différentes techniques utilisées dans le domaine de la détection de la somnolence du conducteur est de représenter et détecter les signes de somnolence du

conducteur. Cette technique est basée sur le comportement et l'apparence du conducteur ou les deux.

De nombreuses études [Massoz, Verly, and Van Droogenbroeck, 2018; Mandal et al., 2017; Al-Rahayfeh and Faezipour, 2013; Kurylyak, Lamonaca, and Mirabelli, 2012] ont été menées sur l'état de l'œil du conducteur à partir d'un système de vidéosurveillance pour détecter les yeux et calculer leur fréquence de clignement afin de vérifier le niveau de fatigue du conducteur. Certaines études [Kurylyak, Lamonaca, and Mirabelli, 2012], comme celle d'Adaboost [Kurylyak, Lamonaca, and Mirabelli, 2012], ont utilisé une cascade de classificateurs pour la détection rapide de la zone oculaire. Ainsi, dans notre travail, nous utilisons un modèle RNN [Massoz, Verly, and Van Droogenbroeck, 2018], qui vise à analyser la fermeture lente et longue des yeux du conducteur. Cette approche utilise un descripteur de type Haar [Lienhart and Maydt, 2002] et un algorithme de classification AdaBoost [Freund, Schapire, et al., 1996] pour le suivi du visage et des yeux en utilisant le pourcentage de fermeture des yeux (PERCLOS) pour évaluer la fatigue du conducteur. Le PERCLOS évalue la proportion du temps total pendant lequel les paupières d'un conducteur sont $\geq 80\%$ fermées et reflète une fermeture lente des paupières plutôt qu'un clignement standard.

Il existe également des recherches sur l'estimation de la position séparée de la tête du conducteur [Alioua et al., 2016]. Le système d'estimation de la position séparée de la tête est une méthode qui aboutit à l'orientation de la tête par rapport à une vue de caméra. L'objectif est donc de détecter si la tête du conducteur est inclinée vers l'avant, ce qui est révélateur d'une perte de concentration. Pour cela, un descripteur spécifique est créé dans cette recherche en fusionnant quatre descripteurs de l'orientation très pertinente de la tête, et on retrouve les filtres orientables, l'histogramme des gradients orientés, les caractéristiques Haar [Lienhart and Maydt, 2002], et une version adaptée du descripteur Speeded-Up Robust-Features. Ensuite, ces caractéristiques sont appliquées dans l'algorithme de classification de la machine à vecteurs de support [Burges, 1998].

D'autres études ont été menées sur la santé des conducteurs [Dong et al., 2019; Yang et al., 2019; Han et al., 2019; Chaudhuri and Routray, 2019; Wang et al., 2019], afin d'analyser quand le conducteur n'est plus capable de conduire et de détecter les premiers signes de fatigue pour anticiper l'apparition de la somnolence afin d'alerter le conducteur qu'il est nécessaire de faire une pause. Pour ce type de méthode, on utilise l'électroencéphalographie (EEG), qui effectue des mesures à l'aide de captures collées directement sur la peau du conducteur en mesurant le rythme cardiaque, la température corporelle et d'autres mesures en fonction du type de travail. En outre, un autre type de capture est installé dans une voiture directement ; par exemple, il capture les mesures de la température à l'intérieur et à l'extérieur de la voiture. Toutes

ces mesures calculées sont utilisées pour alimenter un modèle vectoriel, qui classe les valeurs mesurées entre les états d'alerte et de fatigue extrême ; ce type de travail donne de bons résultats. Néanmoins, l'utilisation à l'échelle industrielle, notamment pour les conducteurs ordinaires, est coûteuse et peu pratique pour un usage quotidien.

Des travaux sont également menés pour détecter la somnolence [Haggag, n.d.; Liu, 2018; Liu et al., 2019a]. La solution proposée est un système combiné de détection de la fatigue et de détection des comportements indésirables du conducteur pendant la conduite. En utilisant la boîte à outils Kinect for Windows Developer, Kinect a permis de détecter des postures simples en détectant la position des membres du corps du conducteur et en analysant leur mouvement pour détecter les postures les plus proches du comportement de fatigue du conducteur.

Enfin, l'apprentissage en profondeur est une tendance qui se développe dans l'analyse générale des données et qui est identifiée comme l'une des 10 principales technologies de 2018 [Rev., 2018]. Plusieurs travaux utilisent l'apprentissage profond [Liu et al., 2019b; Xiao et al., 2019], comme l'apprentissage profond étant une mise à niveau des réseaux de neurones artificiels. Il consiste à combiner un grand nombre de couches qui permettent d'atteindre un niveau d'abstraction plus élevé et d'améliorer les prédictions à partir de données souvent complexes et hétérogènes.

2.3 ANALYSE DE L'ÉTAT DE LA ROUTE

Dans cette section, nous passons en revue les approches et techniques actuelles utilisées pour la détection des naissances de poulet, c'est-à-dire la détection et la segmentation. Pour cela, nos recherches se concentrent sur les architectures basées sur l'apprentissage profond.

	Vision-based	Vibration-based	Laser-based	Stereo imaging
Device used	Camera	Accelerometer	Laser	Cameras
Technology used	2D imaging	Force and rotation and orientation	3D reconstruction of the image using light reflection	3D reconstruction using multiple cameras
Response time	High	Low	Low	High
Sensing time	While approaching the pothole	While going through a pothole	While approaching the pothole	While approaching the pothole
Processing	Complex image processing algorithms	Readings are directly used	Collection of 3D point cloud with their elevations.	A complex process of 3D image construction by combining image from different camera perceptions
Cost	High because of delicate parts like lens	Low	High	High
Characterization of pothole	Based on size	Based on vibrations	Based on 3D image constructed	Based on 3D image constructed
Detection at night time	Difficult due to poor lighting	Can detect	Can detect	Difficult due to poor lighting
Accuracy	Depends on the algorithm used	High	High	Depends on the alignment of cameras and algorithms used

Figure 2.3: Méthodes de détection des nids de poule.

La détection des nids de poule peut être classée en quatre approches : [Shaghouri, Alkhatib, and Berjaoui, 2021; Kim and Ryu, 2014] : l'approche manuelle établie sur une étude visuelle, l'approche technique par vibration [De Zoysa et al., 2007; Eriksson et al.,

2008; Yu and Yu, 2006], l'approche technique par reconstruction 3D (avec la méthode du scanner laser et la méthode de la vision stéréo et du capteur Kinect) [Chang, Chang, and Liu, 2005; Hou, Wang, and Gong, 2007; Wang, 2004] et l'approche technique de la vision [Ye et al., 2021; Jog et al., 2012; Coenen and Golroo, 2017]. La figure 2.3 résume et compare les différentes approches de détection des nids de poule en fonction de la technologie utilisée, du temps de réponse et de détection, et de la qualité de la détection. La technologie utilisée, le temps de réponse et de détection, le traitement, le coût, la caractérisation des nids de poule et la précision de la détection [Bhatia et al., 2019]. Traditionnellement, un groupe d'employés effectuait la détection des nids de poule en examinant les vidéos numériques enregistrées par des caméras. La détection des nids-de-poule se fait par l'examen de la vidéo numérique enregistrée des chaussées. Cette méthode est coûteuse et prend du temps [Bianchini, Bandini, and Smith, 2010].

2.3.1 Méthode manuelle (l'enquête visuelle)

- Méthode marocaine [Mohamed AFECHKAR, n.d.]:

Le principe de sondage pratiqué pour le réseau routier revêtu est basé sur l'observation à intervalles réguliers de 200m (5 relevés / Km), de l'existence ou non de dégradations observées sur une section transversale de 20 à 30m de longueur, la sévérité a été indiquée par trois niveaux.

- 0 : Aucune dégradation
- 1 : Faible dégradation
- 2 : Dégradation nuisible

In general principle, For each section of road, the operator will have to define a direction of course for the survey. Beginning from the starting point of the section to be surveyed, the operator notes the various degradation observed on the field of the observation zone and gives ratings. Then, he drives his vehicle the distance of 200 m in the direction of the survey without taking into account the existing intermediate degradation, then he carries out a second survey according to the first one and so on until the completion of the section.

- French method [Mohamed AFECHKAR, n.d.]:

L'objectif de cette méthode est de localiser les dégradations existantes, en précisant leur type, leur gravité et l'extension de la chaussée concernée. Ainsi, la section étudiée est parcourue mètre par mètre, en notant l'étendue de chaque dommage.

Les besoins en matière de relevé de dégradations pour les études routières sont couverts par sept modes opératoires définis en fonction de l'objectif, des conditions d'exploitation et du type de chaussée.

2.3.2 Détection basée sur des captures

Des travaux antérieurs réalisés par des chercheurs de [Eriksson et al., 2008; Madli et al., 2015; Dewangan and Sahu, 2020; Mohan, Padmanabhan, and Ramjee, 2008] montrent des détections de rupture de vitesse utilisant des approches basées sur l'accéléromètre/GPS ou avec des vibrations [Gunawan, Soewito, et al., 2015]. Ces détections basées sur l'accéléromètre ou les vibrations ne seront pas en mesure d'alerter le conducteur d'une nouvelle bosse sur la route. Elles ne sont détectées qu'après le passage du véhicule sur la partie défectueuse de la route, sur la base d'un seuil et d'autres techniques de traitement du domaine temporel et fréquentiel appliquées aux données d'accélérométrie et de vibration.

2.3.3 Détection par apprentissage profond

Il y a aussi des travaux basés sur un capteur sensoriel pour détecter les fissures dans les routes basé sur l'architecture du réseau neuronal conventionnel CNN [Fan et al., 2019]. Il y a aussi des travaux basés sur un système de vision stéréo [Fan et al., 2021; Fan et al., 2019], afin de faire des études globales sur la route entière, donc ce type d'études peut aider les systèmes de conduite assistée et les véhicules autonomes à prendre des décisions par rapport au contexte de la route.

D'autre part, la performance des architectures d'apprentissage profond, a permis d'élever les défauts sur des calculs plus complexes, notamment en relation avec notre travail, qui signifie la détection des nids de poule [Maeda et al., 2018; Cao et al., 2020], ce type d'architecture à la capacité de détecter et de classer les fissures et les nids de poule de la route. Aussi, d'autres travaux sont basés sur des architectures de segmentation sémantique, nous trouvons U-NET [Pereira et al., 2019], en plus des architectures basées sur TOLOV4 [Sarmiento, 2021; Jiang et al., 2020], qui s'avère que ces t'un modèle qui répond bien à la nécessité de traitement en temps réel.

2.4 RAPPELLE SUR LES RNN

2.4.1 Définition

Un réseau neuronal récurrent (RNN) est un type de réseau neuronal artificiel couramment utilisé dans la reconnaissance vocale et le traitement du langage naturel (NLP). Les RNN sont conçus pour reconnaître les caractéristiques séquentielles d'une donnée et utiliser des modèles pour prédire le prochain scénario probable.

Les RNN sont utilisés dans l'apprentissage profond et dans le développement de modèles qui simulent l'activité des neurones du cerveau humain. Ils sont particulièrement puissants dans les cas d'utilisation où le contexte est essentiel pour prédire un

résultat et se distinguent des autres types de réseaux neuronaux artificiels, car ils utilisent des boucles de rétroaction pour traiter une séquence de données qui informe la sortie finale, qui peut également être une séquence de données. Ces boucles de rétroaction permettent aux informations de persister ; cet effet est souvent décrit comme une mémoire.

Les cas d'utilisation des RNN ont tendance à être liés à des modèles de langage dans lesquels la connaissance de la lettre suivante d'un mot ou du mot suivant d'une phrase dépend des données qui le précèdent. Une expérience convaincante montre qu'un RNN entraîné avec les œuvres de Shakespeare produit une prose semblable à celle de Shakespeare - avec succès. L'écriture par les RNN est une forme de créativité informatique. Cette simulation de la créativité humaine est rendue possible par la compréhension de l'IA de la grammaire et de la sémantique apprises à partir de son ensemble d'apprentissage.

2.4.2 Réseaux de neurones

Avec l'augmentation rapide des capacités de calcul des ordinateurs et l'évolution des techniques d'apprentissage, l'utilisation des réseaux de neurones connaît un vif essor, en particulier au sein des communautés du traitement d'image, de la traduction automatique et du traitement de la parole. Au cours des dernières années, deux types de réseaux de neurones ont marqué une rupture technologique dans le domaine du traitement de la parole : les réseaux dits "profonds" et les réseaux récurrents. Nous détaillons dans cette partie le fonctionnement de ces réseaux de neurones et les travaux que nous avons réalisés sur les modèles eux-mêmes et sur les méthodes d'apprentissage.

Le réseau de neurones artificiels (Artificial Neural Network - ANN) fut introduit comme un modèle rudimentaire du traitement de l'information dans le cerveau humain. Ainsi, la structure élémentaire d'une ANN est un réseau de petits nœuds de calcul reliés entre eux par des liens dirigés et pondérés (cf. Fig. 2.4) qui est représentations d'un réseau de neurones artificiels (ANN) sous la forme d'un réseau de nœuds de calculs reliés par des liens dirigés et pondérés par les coefficients $w_{i,j}$. L'information se propage en suivant ces liens entre les entrées et les sorties.. Les nœuds représentent les neurones et les liens pondérés représentent la force des connexions synaptiques reliant les neurones entre eux. Dans cette représentation, le neurone peut alors être un sommateur des potentiels des signaux synaptiques qui lui parviennent, et qui transmet à son tour une information basée sur cette somme via une fonction de transfert de préférence non linéaire (cf. Fig. 2.5), cette figure est une structure interne d'un nœud de calcul dans le cas d'un neurone artificiel sommateur. Tout d'abord, la somme pondérée des entrées du nœud est calculée puis on applique à cette somme une fonction de

transfert non-linéaire. La sortie est alors communiquée aux autres nœuds de calculs. Une ANN est activée en injectant des données au niveau de tout ou partie des nœuds puis en propageant l'information en suivant les liens pondérés. Une fois l'information propagée, on peut collecter les niveaux d'activation de tout ou partie des nœuds et les utiliser comme commande d'un système, comme prédiction ou encore comme classification.

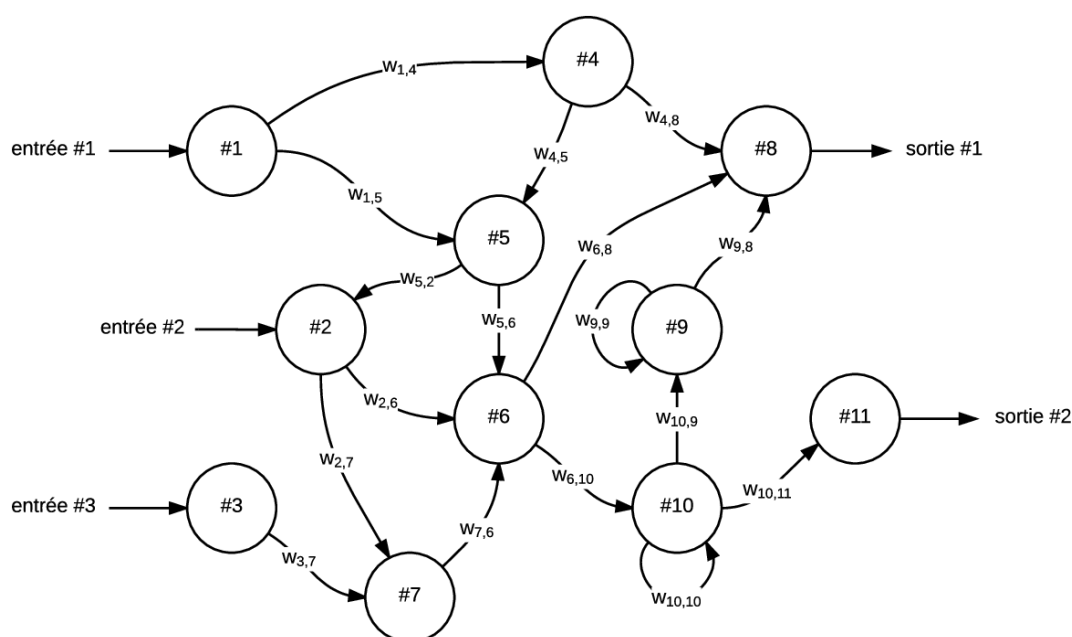


Figure 2.4: Réseau de neurones artificiels (ANN).

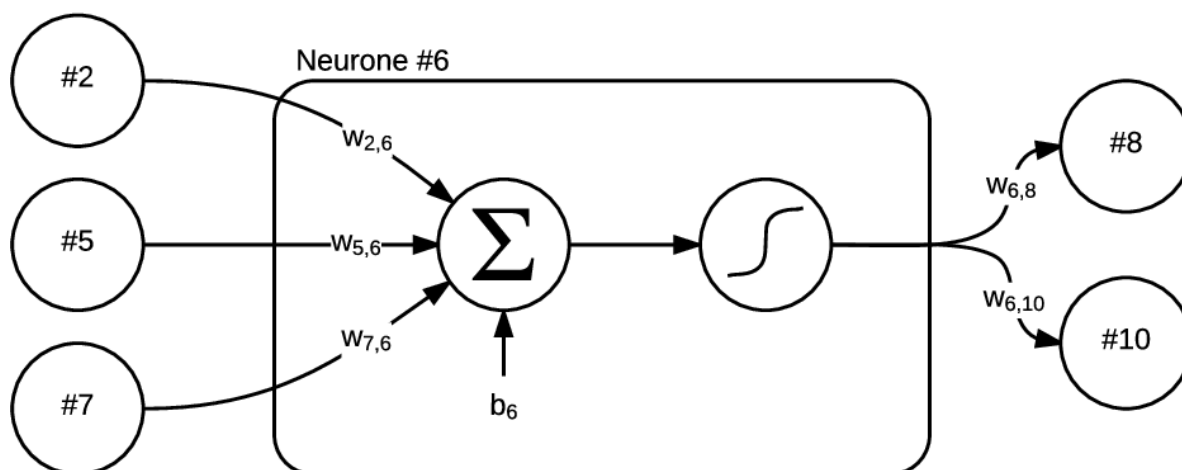


Figure 2.5: Structure interne d'un nœud de calcul.

De nombreux types de réseaux de neurones ayant des propriétés très diverses ont été développés depuis l'apparition des neurones formels dans les années 1940 [McCulloch and Pitts, 1943]. Plus spécifiquement, on distingue deux types de réseaux : ceux dont le graphe de connexion présente au moins un cycle et ceux pour lesquels ce n'est pas le cas.

Les premiers sont dits récurrents et les seconds sont dits acycliques. Parmi les réseaux acycliques, on trouve des réseaux de type perceptron [Rosenblatt, 1957], convolutifs [LeCun et al., 1998], Radial Basis Function [Broomhead and Lowe, 1988] ou les cartes de Kohonen [Kohonen, 1982].

2.4.3 Comment les réseaux neuronaux récurrents apprennent

Les réseaux neuronaux artificiels sont créés avec des composants de traitement de données interconnectés qui sont vaguement conçus pour fonctionner comme le cerveau humain. Ils sont composés de couches de neurones artificiels (nœuds du réseau) capables de traiter les entrées et de transmettre les sorties aux autres nœuds du réseau. Les nœuds sont reliés par des bords ou des poids qui influencent la force d'un signal et la sortie finale du réseau. Dans certains cas, les réseaux neuronaux artificiels traitent les informations dans une seule direction, de l'entrée à la sortie. Ces réseaux neuronaux "feedforward" comprennent les réseaux neuronaux convolutifs qui sous-tendent les systèmes de reconnaissance d'images. Les RNN, en revanche, peuvent être superposés pour traiter l'information dans deux directions.

Comme les réseaux neuronaux "feedforward", les RNN peuvent traiter les données de l'entrée initiale à la sortie finale. À la différence des réseaux neuronaux à anticipation, les RNN utilisent des boucles de rétroaction telles que la rétro-propagation dans le temps (BPTT) tout au long du processus de calcul pour renvoyer les informations dans le réseau. Les entrées sont ainsi reliées entre elles et c'est ce qui permet aux RNN de traiter des données séquentielles et temporelles.

2.4.4 Unités de mémoire à long terme

L'un des inconvénients des RNN standard est le problème du gradient de fuite, qui entraîne une baisse des performances du réseau neuronal en raison de l'impossibilité de le former correctement. Cela se produit avec les réseaux neuronaux à couches profondes, qui sont utilisés pour traiter des données complexes.

Les RNN standard qui utilisent une méthode d'apprentissage basée sur le gradient se dégradent à mesure qu'ils deviennent plus grands et plus complexes. Le réglage efficace des paramètres dans les premières couches prend trop de temps et coûte trop cher en termes de calcul.

Les unités de mémoire à long terme (LSTM), qui ont été inventées par les informaticiens Sepp Hochreiter et Jurgen Schmidhuber en 1997, constituent une solution à ce problème. Les RNN construits avec des unités LSTM classent les données dans des cellules de mémoire à court terme et à long terme. Cela permet aux RNN de déterminer

quelles données sont importantes et doivent être mémorisées et réinjectées dans le réseau et quelles données peuvent être oubliées.

2.5 CONCLUSION

L'intelligence artificielle nous permet l'analyser la complexité de l'environnement qui entoure la voiture et le conducteur, en revanche l'intégration artificielle a besoin plus de donnée pour plus d'efficacité, ce qui il n'est pas aujourd'hui entièrement achevée. À cet égard, l'intégration de nouvelles sources de données semble indispensable.

Avec la multiplication des modes de transport (marche à pied, vélo, scooter, gyropode, ...), tout particulièrement dans les grandes villes, l'enjeu de la sécurité routière du futur est de faire co-exister l'ensemble de ces transports tout en poursuivant les efforts de réduction du nombre d'accidents. L'identification de nouvelles sources de données et l'exploitation de ces données est un défi de taille pour les autorités publiques en charge de la sécurité routière.

Ainsi grâce à ces nouvelles sources de données, le domaine d'intelligence artificiel peut s'améliorer en conteneur, en conséquence de ça, on aura une amélioration sur la sécurité des usagers de routes.

DÉTECTION DE LA FATIGUE DES CONDUCTEURS EN SE BASANT SUR DES RNNs

SOMMAIRE

3.1 INTRODUCTION	23
3.2 INFORMATIONS DE BASE	25
3.2.1 Apprentissage profond	25
3.2.2 Réseau de neurones récurrents	26
3.3 ÉTUDE BASÉE SUR L'ARCHITECTURE LSTM	27
3.3.1 Approche proposée	27
3.3.1.1 Réseaux de neurones récurrents (RNN)	27
3.3.1.2 Réseau de neurones récurrents à mémoire court-terme et long terme	28
3.3.1.3 Architectures	28
3.3.1.4 Ensemble de données	29
3.3.1.5 Entraînement	30
3.3.2 Discussion des résultats	31
3.3.3 Proposition de système	31
3.3.4 Conclusion	32
3.4 ÉTUDE BASÉE SUR L'ARCHITECTURE CONVOLUTIFS 3D	33
3.4.1 Approche proposée	33
3.4.1.1 Apprentissage des caractéristiques via des réseaux convolutifs 3D	33
3.4.1.2 Architectures multi-couches	34
3.4.1.3 Max-pooling : pourquoi ?	35
3.4.2 Expérimentations	38
3.4.2.1 L'ensemble de données NTHU-DDD	38
3.4.2.2 Entraînement	39
3.4.3 Analyse des résultats	40
3.4.3.1 Résultats de précision	40
3.4.3.2 Évolution de l'entraînement	41
3.4.3.3 Mesure de la performance	41

3.4.3.4 Comparaison avec d'autres méthodes	42
3.4.4 Système proposé	42
3.4.5 Conclusion	43

LES accidents de la route deviennent un véritable fléau mondial en raison du nombre élevé de victimes et des conséquences graves qui affectent les usagers de la route ainsi que leurs familles.

Malgré les campagnes de sensibilisation sur la vigilance et la prudence à avoir sur la route, les décès causés par les accidents de la route ne cessent d'augmenter et sont désormais considérés comme un problème majeur de santé publique, plus particulièrement au Maroc où les routes sont parmi les plus meurtrières.

Pour faire face à ce problème, les constructeurs automobiles ont fait des progrès considérables en améliorant l'intelligence et la capacité des véhicules à percevoir et à analyser l'environnement routier pour prévenir les accidents et sécuriser les passagers. Cependant, malgré tous ces efforts, les statistiques d'accidents montrent que dans la plupart des cas, les accidents sont liés à l'inattention des conducteurs et parfois à un comportement irresponsable.

C'est pourquoi un nombre considérable de recherches se sont récemment concentrées sur l'analyse et l'étude des comportements généraux des conducteurs sur la route, en particulier la somnolence, car elle fait partie des facteurs de risque les plus élevés d'accidents et constitue la première cause de décès sur les routes.

Dans notre premier article, nous proposons une approche pour analyser la somnolence des conducteurs en appliquant une architecture de réseau neuronal récurrent aux séquences d'images d'un conducteur. Nous avons utilisé un ensemble de données publiques pour entraîner et valider notre modèle et avons appliqué une architecture de réseau neuronal récurrent appelée "mémoire à long terme" pour détecter la somnolence du conducteur.

Ainsi, on présente un moyen d'analyser et d'anticiper la somnolence d'un conducteur en appliquant un réseau neuronal récurrent sur une séquence d'images du visage du conducteur. Nous avons utilisé un ensemble de données pour façonner et approuver notre modèle et avons mis en œuvre une architecture de réseau neuronal récurrent basée sur un modèle multicouche de réseaux convolutifs 3D pour détecter la somnolence du conducteur. Après une session d'entraînement, nous avons obtenu une précision prometteuse qui approche un taux d'acceptation de 92%, ce qui a permis de développer un système de surveillance des conducteurs en temps réel pour réduire les accidents de la route.

3.1 INTRODUCTION

Selon l'Organisation mondiale de la Santé, les accidents de la circulation constituent un problème majeur de santé publique dans le monde entier. Comme l'indique le Rapport de situation mondial sur la sécurité routière 2015 [Organization, 2015], ils constituent l'une des principales causes de décès et de blessures.

an, près de 1,25 million de personnes meurent et des millions d'autres sont blessées ou handicapées suite à d'accidents de la route. La figure 3.1 montre la tendance à l'augmentation du nombre de décès sur la route, qui passe de 1,1 à 1,25 million entre 2001 et 2013, respectivement.

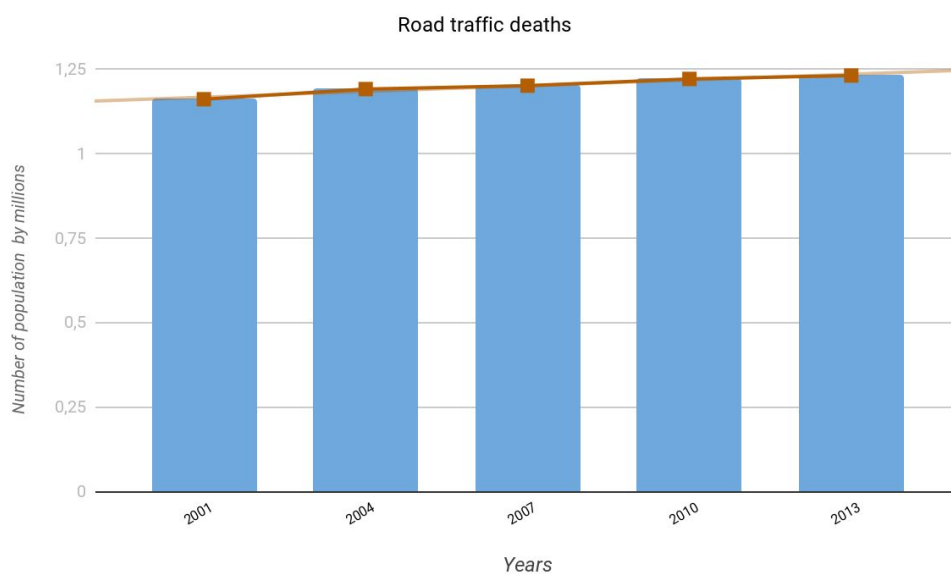


Figure 3.1: Nombre de décès liés aux accidents de la route dans le monde.

En 2013, une étude a été menée au Maroc [El Ftouh et al., 2013], montrant des chiffres alarmants d'accidents liés à la somnolence, responsables chaque année de plus de 4000 décès et de dégâts matériels estimés à 14 milliards de dirhams 3.2. 36,8% des cas se sont déjà sentis endormi pendant quelques secondes et 31,1% des cas ont déjà sortie de la route en raison d'un moment d'inattention ou du somnolence., y compris le quart du mois précédant l'enquête. Par ailleurs, 42,4% des conducteurs interrogés ont reconnu ne pas avoir respecté les recommandations de repos après toutes les 2 h ou 500 km de trajet.

Compte tenu de la gravité du problème, qui cause des dommages sociaux et économiques dans la société, la communauté des chercheurs et l'industrie estiment qu'il est plus que nécessaire de se concentrer sur ce phénomène et de créer un système

Somnolence et manque de vigilance

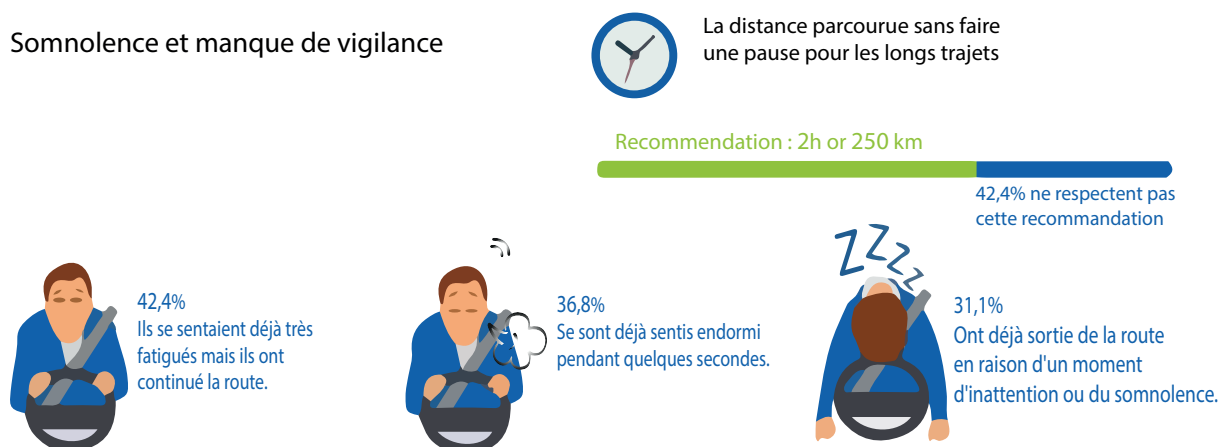


Figure 3.2: Étude de la de la somnolence au volant sur une population marocaine.

efficace pour prévenir l'apparition de la fatigue et réduire le nombre d'accidents afin de diminuer le nombre de blessés et de morts dans les accidents de la route.

Les avancées spectaculaires dans le domaine de l'apprentissage profond, et notamment des réseaux de neurones récurrents (RNN), ont permis d'apporter une solution à ce type de problème. En particulier, les données collectées via les caméras embarquées dans les voitures étant souvent hétérogènes et volumineux, elles correspondent bien aux performances de ces modèles de calcul, puissants et se sont révélés prometteurs dans de nombreuses tâches ardues.

L'idée du RNN est d'utiliser des informations séquentielles et d'exécuter la même tâche pour chaque élément d'une séquence à partir de laquelle le nom est récurrent. Les techniques d'apprentissage des réseaux RNN sont les mêmes que celles des réseaux classiques (Backpropagation [Hecht-Nielsen et al., 1988]), cependant les RNN utilisent la mémoire pour capturer l'état des événements calculés dans le passé. Ainsi, les RNN nous permettent de résoudre le problème de la somnolence du conducteur, et de calculer les séquences du début de la somnolence, permettant ainsi au système d'alerter le conducteur quelques secondes avant la perte de vigilance de manière définitive.

Cependant, les recherches se sont concentrées sur l'état physique, psychologique et physiologique du conducteur, comme l'étude de l'influence de la motivation et de l'émotion du conducteur sur sa capacité à prendre des décisions [Summala, 2007]. De plus, l'environnement qui entoure le conducteur est un facteur qui influence le degré de concentration du conducteur et agit sur sa nervosité [Koiliyas, Mousas, and Rekabdar, 2019]. Par ailleurs, de nombreuses méthodes liées à l'étude de la fatigue du conducteur ont été proposées dans la littérature [Borghini et al., 2014; Mu, Hu, and Min, 2017; Zhang et al., 2017; Chaudhuri and Routray, 2019]. Cependant, ces méthodes analysent l'état du conducteur via des capteurs installés soit sur la voiture, soit directement sur le conducteur. Ces techniques donnent des résultats très optimaux. D'autre part,

notre source de données est différente, car nous effectuons un traitement d'image pour extraire le comportement du conducteur. La combinaison de ces deux solutions peut donner de meilleurs résultats.

Pour cela, nous avons utilisé le jeu de données de [Computer Vision Lab, 2016] pour mettre en œuvre la méthode utilisée dans [Weng, Lai, and Lai, 2016]. Cette méthode vise à améliorer la précision de la détection de la somnolence par classification binaire, car le jeu de données contient plusieurs classes liées à la somnolence, avec une seule couche grise. Cela nous permet d'effectuer des calculs d'apprentissage automatique supervisé sur la partie somnolence pour obtenir un apprentissage très optimal avec un taux de réussite plus élevé. Le découpage des scènes vidéo du jeu de données en courtes durées de 7 s au maximum est nécessaire pour choisir les scènes qui reflètent l'action de la somnolence.

Après la préparation du jeu de données, l'apprentissage a été initié grâce à l'algorithme détaillé dans les sections suivantes. Ensuite, d'autres calculs ont été effectués avec des algorithmes supplémentaires afin de réaliser un tableau statistique permettant de présenter une idée globale sur les résultats des différents algorithmes pouvant être utilisés dans ce type de problème.

En résumé, l'objectif de ce travail est d'améliorer les conditions de sécurité routière afin d'anticiper ce type d'accident mortel. Pour cela, nous proposons des algorithmes RNN qui ont déjà fait leurs preuves dans d'autres domaines semblables à notre problème pour détecter une action à partir d'une séquence d'images. Cela pourrait permettre aux industries automobiles d'utiliser ce type d'étude pour lancer des projets faisables et réalisables.

3.2 INFORMATIONS DE BASE

3.2.1 Apprentissage profond

L'apprentissage profond est une mise à niveau des réseaux de neurones artificiels, constituée d'une structure plus complexe permettant des niveaux d'abstraction plus élevés et de meilleures prédictions à partir des données [Schmidhuber, 2015; Goodfellow et al., 2016]. L'apprentissage profond ne cesse de se développer pour devenir le principal outil d'apprentissage automatique dans les domaines de l'imagerie générale et de la vision par ordinateur.

Le concept d'apprentissage profond fournit plusieurs algorithmes et méthodes qui offrent une multitude de possibilités pour extraire certaines caractéristiques de la somnolence. Pour extraire la représentation intermédiaire, les algorithmes se composent des logiciels propriétaires [Wang and Xu, 2016; Liang et al., 2017; François et al., 2016], ou d'un CNN pré-entraîné [Shih and Hsu, 2016] tel que celui du VGG-16 [Simonyan

and Zisserman, 2014]. Pour déterminer la somnolence, les modèles consistent en une régression logistique [Liang et al., 2017], un réseau neuronal artificiel (ANN) [Wang and Xu, 2016; Ebrahim et al., 2014], une machine à vecteurs de support (SVM) [Cristianini, Shawe-Taylor, et al., 2000], modèle de Markov caché (HMM) [Weng, Lai, and Lai, 2016], Multi-Timescale par CNN [Massoz, Verly, and Van Droogenbroeck, 2018], réseau de mémoire à long et court terme (LSTM) lissé par un CNN temporel [Shih and Hsu, 2016], ou détection 3D-CN de bout en bout [Huynh, Park, and Kim, 2016] sont réalisés.

3.2.2 Réseau de neurones récurrents

Les dernières avancées dans le domaine de l'apprentissage profond ont permis d'éliminer les faits importants des informations brutes, ce qui a permis de faire un grand pas en avant dans le domaine de la vision par ordinateur [Krizhevsky, Sutskever, and Hinton, 2012; Sermanet et al., 2013b; Mnih, 2013].

Le RNN est de manière native quelque part dans le temps car son état caché est un élément de tous les états cachés passés, comme le montre la figure 3.3.

Le réseau récurrent le plus simple comprend une unité répétitive reliant un vecteur d'entrée (x_1, x_2, \dots, x_T) , une couche cachée de vecteurs (h_1, h_2, \dots, h_T) , et un vecteur de sortie (y_1, y_2, \dots, y_T) . Les représentations sont créées par la transformation non linéaire de la séquence d'entrée de $t = 1$ à T . Parmi les réseaux récurrents simples (SRN), on peut citer le réseau d'Elman [Elman, 1990] et le réseau de Jordan [Jordan, 1997].

Cependant, le RNN utilisé dans d'autres problèmes de recherche, tel que [Rekabdar, Mousas, and Gupta, 2019], a été utilisé pour les réseaux adverses génératifs afin de générer un résumé de texte abstrait. Le NNH-CNN [Abdel-Hamid et al., 2012] a appliqué des filtres convolutifs locaux avec max-pooling à la fréquence au lieu. Nous avons également trouvé P-CNN [Chéron, Laptev, and Schmid, 2015] et Auto-Conditioned LSTM [Zhou et al., 2018], qui sont des méthodes pour reconnaître et prédire des actions. Il y a aussi les travaux [Saito et al., 2017; Sela, Richardson, and Kimmel, 2017] sur la prédiction de la forme et de la texture 3D des objets à partir d'une seule image. Dans [Rekabdar and Mousas, 2018], les auteurs ont proposé un réseau neuronal convolutif dilaté pour capturer les dépendances temporelles dans le contexte de la prédiction des manœuvres du conducteur. Toutes ces méthodes montrent la multitude de choix pour traiter un problème séquentiel. Notre choix a été motivé par le résultat publié dans les travaux précédents de la méthode des ConvNets 3D, qui montre une efficacité prédictive très élevée.

Dans la section suivante, nous présentons l'architecture multicouche basée sur un modèle RNN, qui peut être une solution optimale au problème discuté dans ce travail.

3.3 ÉTUDE BASÉE SUR L'ARCHITECTURE LSTM

3.3.1 Approche proposée

Cette section présente l'architecture de réseau proposée. L'architecture de base est d'abord décrite, puis une description détaillée des couches choisies pour les deux solutions est fournie.

3.3.1.1 Réseaux de neurones récurrents (RNN)

Les récentes avancées dans le domaine de l'apprentissage profond ont permis d'extraire des caractéristiques de haut niveau à partir de données sensorielles brutes, ce qui a conduit à des percées dans le domaine de la vision par ordinateur [Sermanet et al., 2013b]. Les RNN sont intrinsèquement profonds dans le temps, car leur état caché est une fonction de tous les états cachés précédents Figure 3.3.

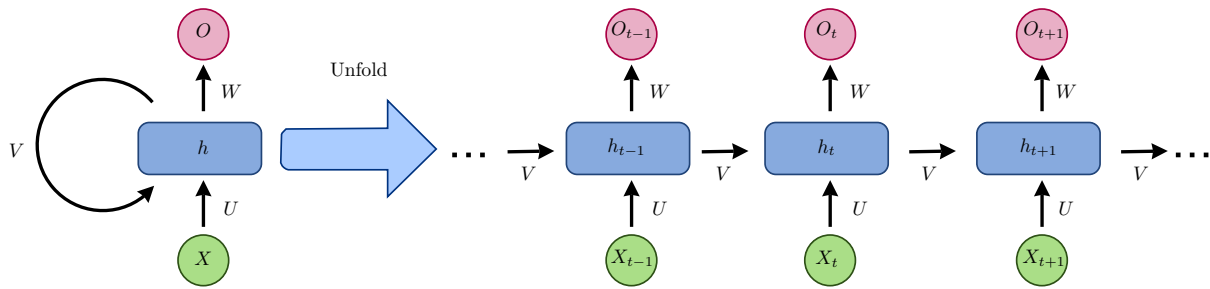


Figure 3.3: Réseau de neurones récurrents et le déroulement dans le temps.

Le réseau récursif le plus simple comprend une unité répétitive reliant un vecteur d'entrée (x_1, x_2, \dots, x_T) , un vecteur de couche cachée des vecteurs (h_1, h_2, \dots, h_T) et un vecteur de sortie (y_1, y_2, \dots, y_T) . Les représentations sont générées par la transformation non linéaire de la séquence d'entrée de $t = 1$ à T . Parmi les réseaux récurrents simples (SRN), on peut citer le réseau d'Elman [Elman, 1990] et le réseau de Jordan [Jordan, 1997], décrits par les équations suivantes :

- Elman network

$$h_x = \sigma_h(Wx_t + Uh_{t-1} + b_h) \quad (3.1)$$

$$y_t = \sigma_o(Wh_t + b_o) \quad (3.2)$$

- Jordan network

$$h_x = \sigma_h(Wx_t + Uy_{t-1} + b_h) \quad (3.3)$$

$$y_t = \sigma_o(Wh_t + b_o) \quad (3.4)$$

dans laquelle W et U sont les matrices de paramètres, b est le vecteur, et σ_h et σ_o sont les fonctions d'activation.

Dans cette section, nous examinons les deux architectures RNN les plus utilisées, qui peuvent constituer une solution optimale au problème abordé dans ce travail.

3.3.1.2 Réseau de neurones récurrents à mémoire court-terme et long terme

Le bloc ou réseau de mémoire à long terme (LSTM) est un RNN simple, qui peut être utilisé comme composant ou bloc de construction (de couches cachées) pour un RNN éventuellement plus grand. Le bloc LSTM est lui-même un réseau récurrent, car il contient des connexions récurrentes comparables aux connexions d'un RNN classique [Sutskever, Vinyals, and Le, 2014].

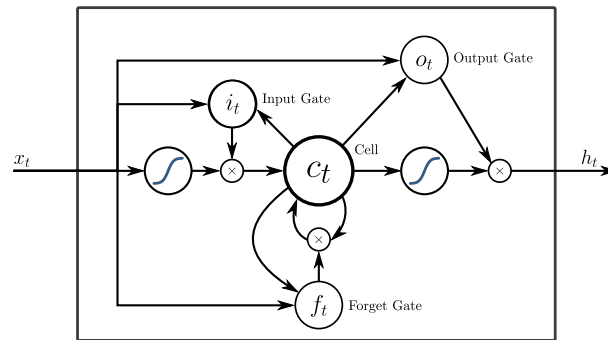


Figure 3.4: Long Short-term Memory Cell

Le LSTM est constitué de trois portes, à savoir la porte d'entrée i , la porte de sortie o et la porte d'oubli f , ainsi que d'une cellule mémoire c . À chaque pas de temps t , le LSTM calcule d'abord les activations des portes i_t , f_t (3.6)(3.5) et actualise la cellule mémoire de c_{t-1} à c_t (3.8). Il calcule ensuite l'activation de la porte de sortie o_t (3.7), et produit enfin la représentation cachée h_t (3.9). Les entrées du LSTM sont les observations x_t et la représentation cachée du pas de temps précédent h_{t-1} . Le LSTM applique l'ensemble suivant d'opérations de mise à jour :

$$f_t = \sigma_g(W_f * x_t + U_f * h_{t-1} + V_f \circ c_{t-1} + b_f) \quad (3.5)$$

$$i_t = \sigma_g(W_i * x_t + U_i * h_{t-1} + V_i \circ c_{t-1} + b_i) \quad (3.6)$$

$$o_t = \sigma_g(W_o * x_t + U_o * h_{t-1} + V_o \circ c_{t-1} + b_o) \quad (3.7)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c * x_t + U_c * h_{t-1} + b_c) \quad (3.8)$$

$$h_t = o_t \circ \sigma_h(c_t) \quad (3.9)$$

3.3.1.3 Architectures

L'architecture globale du modèle de détection de la somnolence proposé est basée sur le modèle LSTM. Nous avons essayé d'établir le meilleur profil de ce modèle de calcul, car il est reconnu comme l'un des modèles les plus précis pour des problèmes

similaires. Les paragraphes suivants décrivent l'architecture et les couches utilisées dans l'apprentissage profond.

Le modèle composite LSTM [Srivastava, Mansimov, and Salakhudinov, 2015] est moins profond et reflète la vitesse de formation ; cependant, il utilise un temps considérable pour converger vers un excellent résultat à partir de plusieurs époques. Par conséquent, pour l'architecture que nous proposons, nous avons utilisé trois couches LSTM avec une augmentation des neurones d'une couche à l'autre, après quoi une couche aplatie est appliquée avant la couche entièrement connectée, suivie d'une couche de sortie softmax qui donne pour la prédiction du training (Figure 3.5).

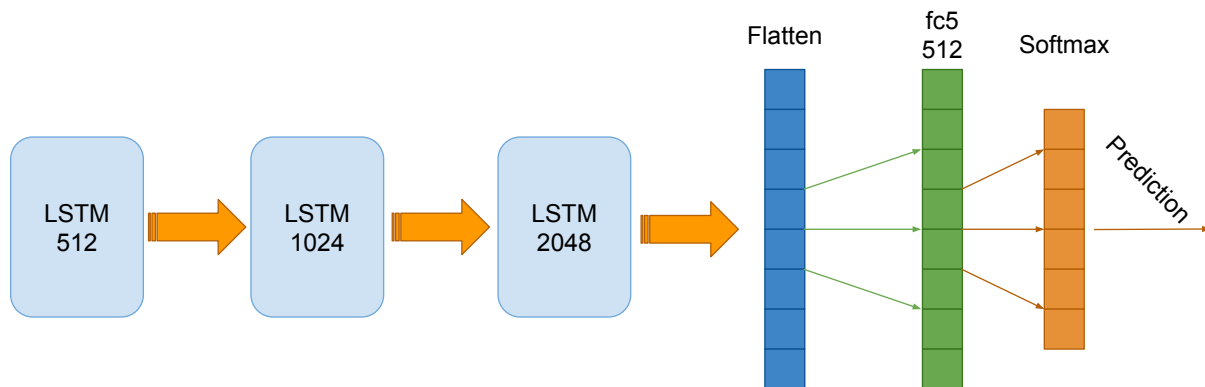


Figure 3.5: LSTM composite model.

La fonction Softmax, qui est la version la plus connue et la plus largement acceptée de la fonction Softmax, est représentée dans les études suivantes [Goodfellow et al., 2016].

3.3.1.4 Ensemble de données

Pour étudier la somnolence des conducteurs, nous avons utilisé un ensemble de données de détection de la somnolence. Cet ensemble de données contient 18 sujets avec 5 scénarios (Visage sans Lunettes, Visage avec Lunettes, Nuit et visage sans lunettes, Nuit et visage avec lunettes, visage avec lunettes de soleil). Les séquences correspondant aux deux scénarios les plus importants, c'est-à-dire la combinaison de symptômes liés à la somnolence (bâillement, hochement de tête, vitesse de clignement lente) et d'actions non liées à la somnolence (parler, rire, regarder des deux côtés), ont été enregistrées pendant 1,5 minute. (voir figure 3.6 – différente séquence de classe utilisée dans l'ordre lors de la détection : conducteur avec lunettes, conducteur sans lunettes de nuit, conducteur avec lunettes la nuit, conducteur sans lunettes de jour et conducteur avec des lunettes de soleil–), ce qui constitue un scénario différent. Cette figure nous permet d'étudier tous les cas possibles de somnolence d'un conducteur.



Figure 3.6: Différentes séquences de classe.

Ainsi, nous avons coupé toutes les vidéos en fonction du comportement souhaité pour obtenir les séquences de la classe de somnolence uniquement. Nous avons extrait deux clips de chaque vidéo, et chacun de ces clips a une durée maximale de 7 s. Ensuite, nous avons préparé un autre sous-ensemble de données qui correspond au comportement normal du conducteur. À la fin de notre découpage, nous avons obtenu 849 clips.

Ce découpage des vidéos de l'ensemble de données a été privilégié pour mieux extraire les séquences qui reflètent la somnolence et éviter les séquences qui peuvent fausser les résultats de l'apprentissage.

3.3.1.5 Entraînement

L'entraînement du jeu de données commence par le découpage des vidéos de 36 sujets, comme décrit dans le paragraphe précédent, afin de créer deux classes d'apprentissages telles que la première classe correspond au comportement normal des conducteurs et la seconde classe correspond au comportement de somnolence des conducteurs. Toutes les vidéos ont été divisées en trois groupes : 60% pour la formation, 30% pour la validation et 10% pour le test.

Ensuite, en fonction de l'architecture choisie, les tailles des images de chaque clip ont été redimensionnées pour optimiser la taille mémoire de la machine.

Pour réaliser cet entraînement, nous avons utilisé un PC avec Alienware R17, Ubuntu 16.04 LTC, 8G GPU, et 16G RAM. Le niveau logiciel est basé sur le cadre Keras [Chollet et al., 2015] avec le script de développement python et est une bibliothèque de niveau modèle, fournissant des blocs de construction de haut niveau pour développer des modèles d'apprentissage profond en utilisant le backend TensorFlow [Abadi et al., 2016].

À la fin de l'apprentissage, une phase de test a été réalisée sur les ensembles de clips du début du découpage pour valider l'efficacité du modèle généré. Le tableau

3.2 montre les résultats des apprentissages réalisés sur plusieurs modèles de calcul compatibles avec l'apprentissage des séquences vidéo.

3.3.2 Discussion des résultats

Table 3.1: Résultats expérimentaux sur l'ensemble de données de la somnolence.

Model	Accuracy	Validate	Test
LSTMs	92.71 %	91.41 %	73.43 %
LSTM-CNN [Donahue et al., 2015; Krizhevsky, Sutskever, and Hinton, 2012]	92.04 %	85.86%	73.18 %
MLP [Kumar, 2018]	71.18 %	72.22 %	60.90 %

Dans cette section, nous présentons les résultats d'apprentissage de chaque architecture sous forme de graphe, le taux d'acceptation convergent vers 92% sur 100 époques, ce qui différencie les autres modèles qui utilisent plusieurs couches en profondeur par rapport à l'architecture LSTM.

Le tableau 3.1 présente le résultat de notre travail sur différents modèles de calcul avec les taux de précision et de validation, ainsi que le taux de test effectué sur des clips non utilisés dans l'étape d'apprentissage.

La figure 3.7 reflète l'évolution de l'apprentissage réalisé par l'architecture LSTM.

3.3.3 Proposition de système

Dans cette section, nous proposons une architecture logicielle pour une application mobile illustrée dans la figure 3.8, dans laquelle l'application mobile utilisera l'appareil photo du téléphone pour capturer une séquence d'images à une fréquence de 5 images par seconde, et à travers le modèle de prédiction de la somnolence, les calculs seront effectués en temps réel, si le modèle prédit la somnolence, un message visuel et audio sera exécuté sur le téléphone de l'utilisateur.

Dans un second temps, et avec l'accord de l'utilisateur, pour chaque prédiction détectée, un message de retour sera demandé à l'utilisateur pour labelliser les séquences d'images utilisées. Ensuite, un transfert est exécuté vers un serveur web, qui reprend les images labellisées, ces images seront sauvegardées et analysées pour le module de prédiction à entraîner. L'objectif de cette étape est de s'assurer que la maintenance du système est évolutive et d'améliorer la précision de la prédiction du modèle.

La plateforme technique proposée pour mettre en œuvre notre architecture logicielle est décrite ci-dessous :

- Application mobile : Kivy et tensorflow (Python3)
- Serveur Web : Flask et Tensorflow (Python3)
- Base de données : Postgresql (sql)



Figure 3.7: Résultat de l'entraînement de modèle LSTM.

- Stockage des données : Système de fichiers ou cloud.

3.3.4 Conclusion

Dans cette étude, nous avons tenté de montrer la possibilité d'utiliser des architectures d'apprentissage profond sur les RNN en utilisant des LSTM entraînés sur des ensembles de données vidéo à grande échelle. Cette architecture répond bien au problème séquentiel. De plus, cette méthode a bien montré son efficacité sur des sujets proches de notre problématique.

De plus, notre étude est basée sur une activité qui dure dans le temps et qui nécessite un changement de posture du conducteur. Cela constitue une limite pour les cas où le conducteur est extrêmement somnolent, mais ne change pas de posture. Ce type de séquence ne peut pas être détecté par la méthode proposée, nous pouvons donc renforcer cette étude en l'associant à des études menées sur la physiologie du conducteur : [Kartsch et al., 2018].

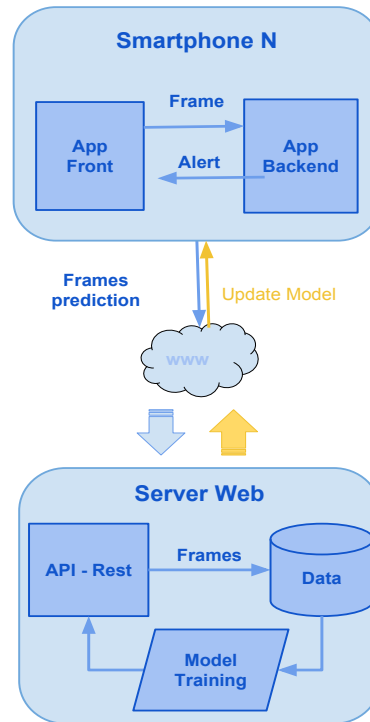


Figure 3.8: Architecture logicielle d'une application mobile et d'un serveur web

3.4 ÉTUDE BASÉE SUR L'ARCHITECTURE CONVOLUTIFS 3D

3.4.1 Approche proposée

Cette section présente l'architecture de réseau multicouche proposée. Le modèle de base a été décrit en premier lieu, puis une description détaillée des couches a été choisie pour notre architecture multicouche.

3.4.1.1 Apprentissage des caractéristiques via des réseaux convolutifs 3D

La méthode proposée de réseaux convolutifs 3D (3D ConvNets) [Tran et al., 2015] apprend des réseaux convolutifs 3D (figure 3.9) sur un support temporel limité à 16 images consécutives avec des noyaux de filtre de taille $3 \times 3 \times 3$, et le choix de la taille du noyau a pour objectif de réduire le poids généré par le modèle pour l'utiliser sur des équipements aux ressources limitées. [Mustafa, Allen, and Appiah, 2019] Il a été démontré que les réseaux neuronaux dynamiques utilisés pour traiter le comportement dépendant du temps s'exécutent beaucoup plus rapidement que des techniques similaires telles que le modèle de Markov caché. Les auteurs ont démontré que les ressources de calcul des HMM ne sont pas pratiques pour les systèmes mobiles et embarqués. De meilleures performances que celles de citekarpathy2014large ont été constatées en permettant à tous les filtres de fonctionner dans l'espace et dans le temps. Cependant, le réseau est considérablement plus profond que celui de [Ji et al., 2013; Karpathy et al.,

2014], avec une structure semblable aux réseaux profonds tels que [Simonyan and Zisserman, 2014]. Une autre façon d'apprendre les relations spatio-temporelles a été proposée dans [Sun et al., 2015], dans laquelle les auteurs ont factorisé une convolution 3D en une convolution spatiale 2D et une convolution temporelle 1D. Plus précisément, leur convolution temporelle a été factorisée en une convolution 2D sur les canaux temporels et les canaux de caractéristiques, et elle n'est utilisée que dans les couches supérieures du réseau.

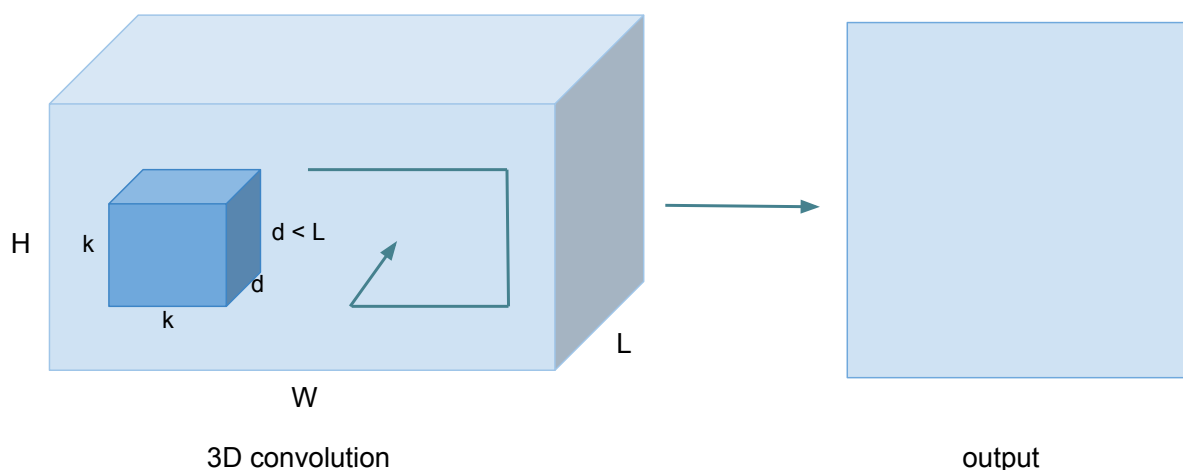


Figure 3.9: Les opérations de convolution en 3D.

Dans la figure 3.9, les trois directions (x, y, z) représentent le calcul d'une convolution. La forme de la sortie représente un volume 3D ; entrée = $[W, H, L]$ et filtre = $[k, k, d]$; ensuite, la sortie = $[W, H, M]$, et surtout, $d < L$ pour obtenir le volume de sortie.

3.4.1.2 Architectures multi-couches

L'architecture multicouche globale de la détection de la somnolence suggérée, basée sur le modèle 3D ConvNets [Tran et al., 2015]. À partir de ce modèle, nous avons établi le meilleur profil, car il est considéré comme l'un des modèles les plus précis pour des problèmes similaires. Les paragraphes suivants expliquent l'architecture et les couches utilisées dans l'apprentissage profond.

D'après les résultats obtenus par les ConvNets 3D, la meilleure option pour un ajustement homogène avec des cœurs de convolution est de $3 \times 3 \times 3$; ces cœurs sont aussi profonds que possible. Un exemple peut être tiré de l'utilisation de données à grande échelle.

En outre, notons que l'architecture ConvNet 3D contient plusieurs couches profondes qui expliquent la durée du calcul, qui est assez longue, mais converge rapidement dès la première époque. Ainsi, dans l'architecture multicouche, il y a six couches pour le modèle ConvNets 3D, quatre couches max-

pooling [Ma, Chang, and Li, 2019; Giusti et al., 2013], une couche flatten [Jin, Dundar, and Culurciello, 2014], et deux couches entièrement connectées, suivies d'une couche de sortie Softmax [Reverdy and Leonard, 2016; Martins and Astudillo, 2016] (Figure 3.10).

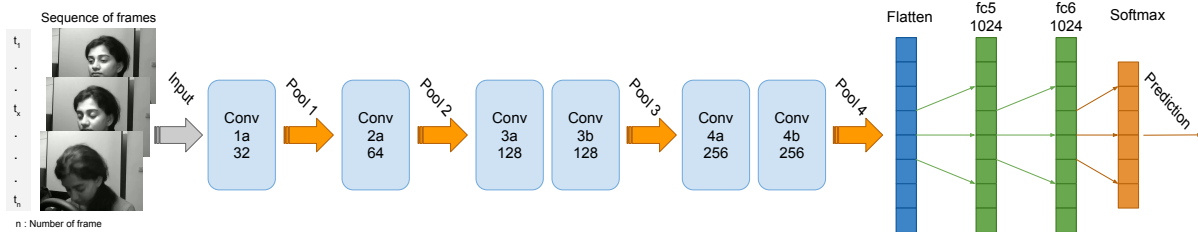


Figure 3.10: Proposition d'architectures multicouches

L'architecture multicouche nous permet de réduire la complexité des données chaque fois que nous passons d'une couche à une autre. Ces couches ConvNets 3D effectuent un filtre qui passe sur la séquence d'images, balayant un petit nombre de pixels à la fois et générant une carte de caractéristiques qui prédit la classe à laquelle chaque caractéristique se rapporte. À l'entrée de la première couche, nous commençons par une dimension de 32×32 pixels. Après lui avoir permis de traiter rapidement la zone sélectionnée de l'image, et de prédire sa classe, chaque passage à une autre couche 3D ConvNets la dimension balayée par le modèle multipliée par 2, ce qui donne 256×256 pixels pour les dernières couches 3d ConvNets ; cette augmentation de la zone filtrée augmente la précision de la prédiction. Ensuite, à chaque couche de pooling, on réduit la taille de l'information pour chaque caractéristique obtenue dans la couche 3D ConvNets tout en conservant les informations essentielles. Après l'étape de filtrage 3D ConvNets et la réduction des données avec la couche de max-pooling, la couche d'aplatissement prend la sortie des couches précédentes, les aplatit, et les transforme en un seul vecteur, qui pourrait être une entrée pour l'étape suivante. Nous avons utilisé le max-pooling comme méthode de réduction en raison de la nature des images qui ont un fond gris. Les couches entièrement connectées continuent d'utiliser l'entrée de l'analyse des caractéristiques et appliquent des poids pour prédire la bonne étiquette. Enfin, nous trouvons la couche de la fonction d'activation Softmax, qui nous donne la classe finale de la séquence d'entrée. Dans notre cas, il s'agit d'une classification binaire.

3.4.1.3 Max-pooling : pourquoi ?

Pooling ou La mise en commun est effectuée dans les réseaux neuronaux pour réduire la variance et la complexité du calcul. Souvent, les débutants utilisent aveuglément une méthode de pooling sans en connaître la raison. Voici une comparaison de trois méthodes de pooling de base largement utilisées.

Les trois types d'opérations de mise en commun sont les suivants :

- Mise en commun maximal : la valeur maximale des pixels du lot est sélectionnée.
- Mise en commun minimal : la valeur minimale du pixel du lot est sélectionnée.
- Mise en commun moyenne : la valeur moyenne de tous les pixels du lot est sélectionnée.

Le lot signifie ici un groupe de pixels de taille égale à la taille du filtre décidée en fonction de la taille de l'image. Dans l'exemple suivant, un filtre de 9x9 est choisi. Le résultat de la méthode de regroupement varie en fonction de la valeur de la taille du filtre.

Nous ne pouvons pas affirmer qu'une méthode de mise en commun particulière est meilleure que d'autres. Le choix de l'opération de mise en commun est fait en fonction des données disponibles. La méthode de mise en commun moyenne lisse l'image et, par conséquent, les caractéristiques nettes peuvent ne pas être identifiées lorsque cette méthode est utilisée.

Le Max pooling sélectionne les pixels les plus clairs de l'image. Il est utile lorsque le fond de l'image est sombre et que nous ne sommes intéressés que par les pixels les plus clairs de l'image. Par exemple : dans le jeu de données MNIST, les chiffres sont représentés en blanc et le fond est noir. On utilise donc le max pooling. De même, le min pooling est utilisé dans l'autre sens.

Les figures suivantes illustrent les effets de la mise en commun sur deux images au contenu différent.

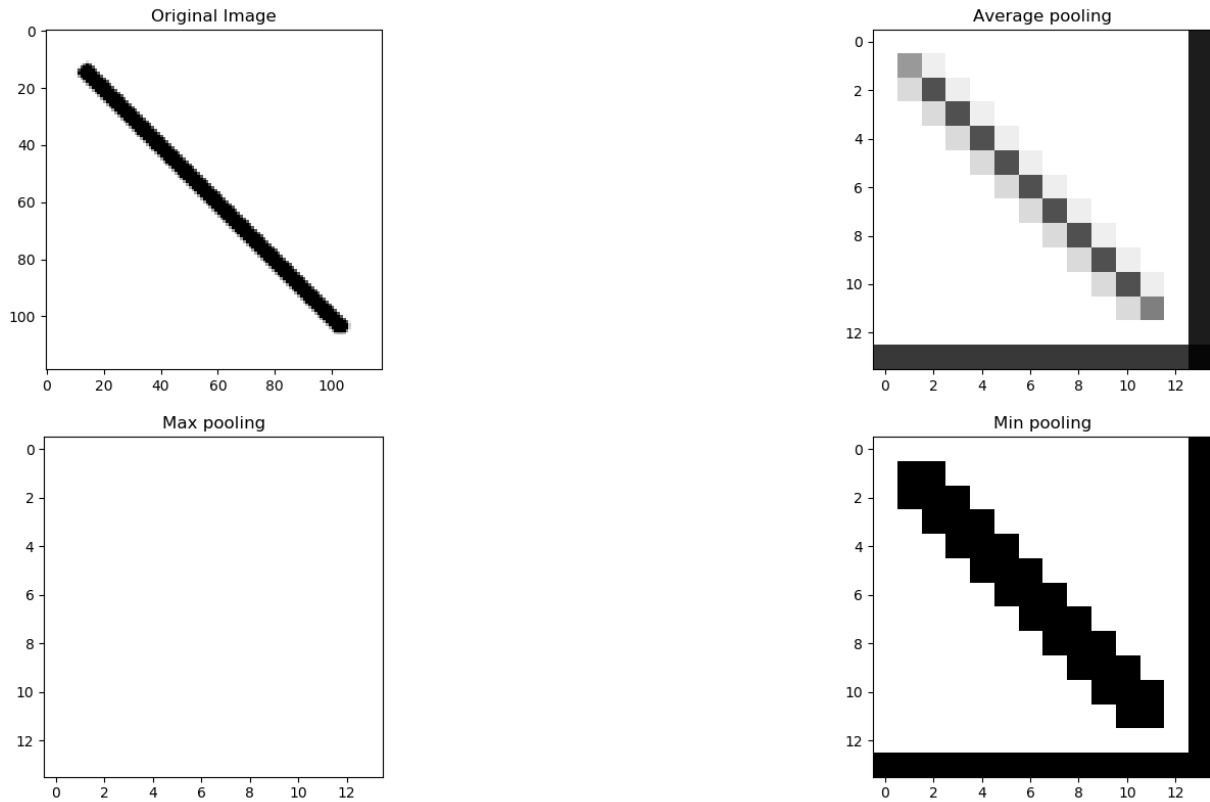


Figure 3.11: Groupage avec un fond blanc et un objet noir.

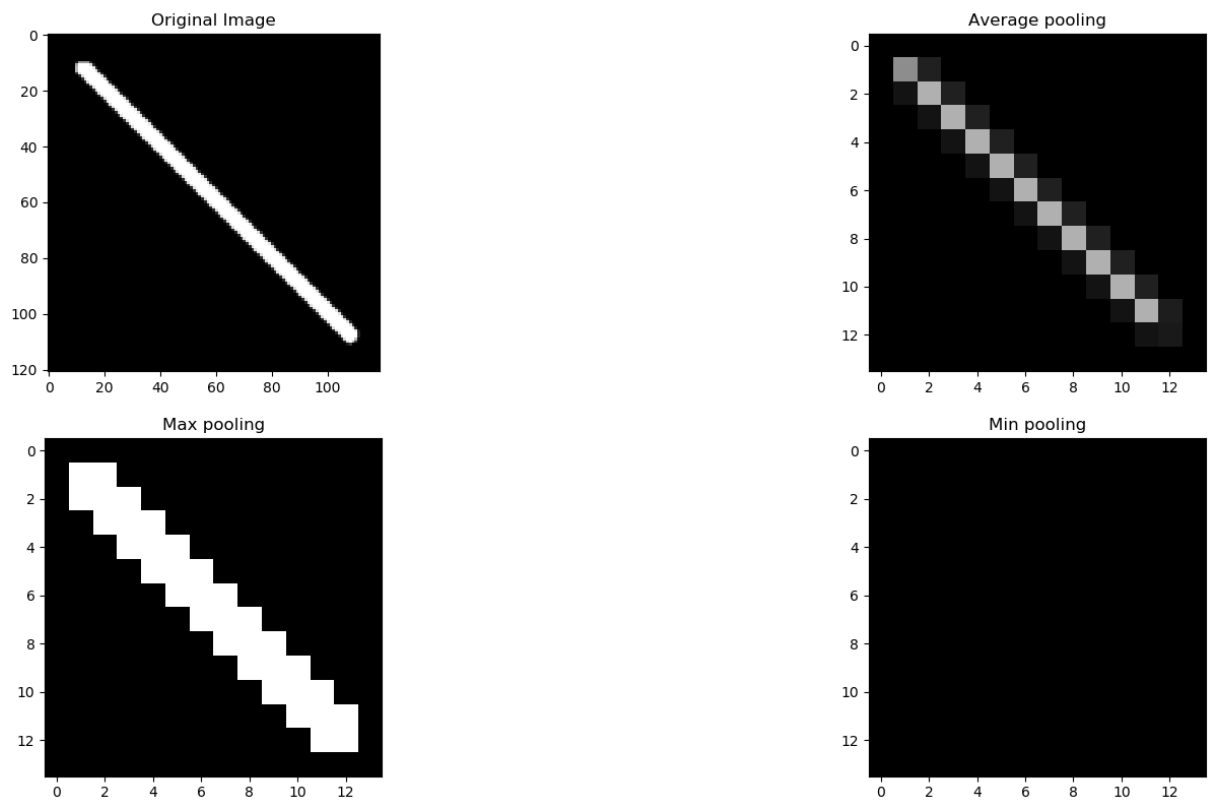


Figure 3.12: Groupage avec un fond noir et un objet blanc.

Lors de la classification du jeu de données des chiffres MNIST à l'aide de CNN, le max pooling est utilisé, car le fond de ces images est rendu noir pour réduire le coût de calcul.

Par conséquent, dans notre calcul, on a utilisé max pooling, car les images plus moins claires, ce qui donne de meilleurs résultats, qui permet d'extraire au maximum les caractéristiques de l'image.

3.4.2 Expérimentations

3.4.2.1 L'ensemble de données NTHU-DDD

Nous avons utilisé le jeu de données NTHU-DDD de détection de la somnolence pour apprendre à connaître la somnolence au volant avec différents sujets (voir figure 3.13). Ce jeu de données met en œuvre cinq scénarios pour chaque étude de classe — En outre, nous avons analysé le cas de la conduite avec des lunettes de soleil dans la journée (voir figure 3.6), qui est un scénario différent— les figures montrent tous les cas possibles de somnolence d'un conducteur.

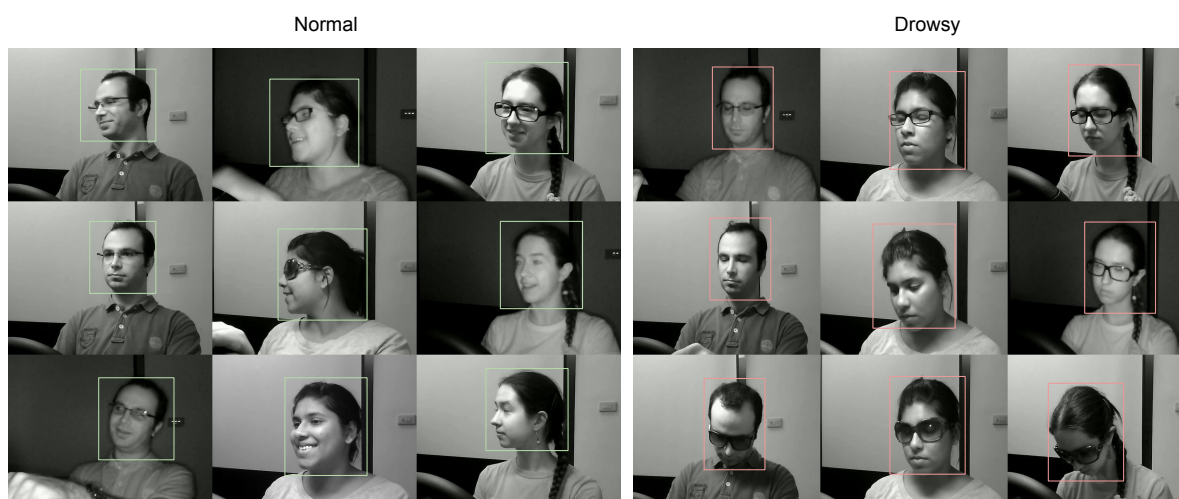


Figure 3.13: Exemples des deux classes, normale et somnolente, de l'ensemble de données NTHU-DDD.

Nous avons donc divisé toutes les vidéos en fonction de l'activité recherchée afin d'obtenir les séquences de la classe de somnolence uniquement. Nous avons sélectionné deux clips dans chaque vidéo et chaque clip à une durée maximale de 7 s. Ensuite, nous avons fourni un autre sous-ensemble de données qui représente le comportement régulier du conducteur. Finalement, nous avons obtenu 849 clips. Ce filtrage des vidéos de l'ensemble de données choisi de préférence pour extraire correctement les séquences qui indiquent la somnolence et pour exclure les séquences qui peuvent nuire aux scores d'apprentissage.

3.4.2.2 Entraînement

La préparation de l'ensemble de données NTHU-DDD commence par le découpage des enregistrements de 36 sujets, comme indiqué dans la sous-section précédente, afin de créer deux classes d'apprentissages, la première classe portant sur le comportement ordinaire des conducteurs et la seconde sur la somnolence des conducteurs. Tous les enregistrements ont été isolés en trois groupes : 60% pour la formation, 30% pour la validation, et 10% pour le test.

Ensuite, en fonction de l'architecture choisie, la taille des images de chaque clip a été ajustée pour optimiser la taille de la mémoire de la machine.

Pour réaliser cet entraînement, nous avons utilisé un PC avec Alienware R17, Ubuntu 16.04 LTC, 16 Go de RAM et 8 Go de GPU. La partie logicielle se compose du framework Keras [Gulli and Pal, 2017] avec le langage de programmation Python, pour que Keras est une bibliothèque fournissant des méthodes prédéfinies pour déployer des modèles d'apprentissage profond avec un backend de formation de bas niveau ; dans notre cas, nous utilisons TensorFlow [Abadi et al., 2016].

Après l'entraînement, une étape de test a été réalisée sur un jeu de données non utilisé lors de la phase d'entraînement afin de valider la précision du modèle produit. Le tableau 3.2 montre les résultats de l'entraînement effectué sur plusieurs modèles de calcul compatibles avec l'entraînement des séquences vidéo.

Pour illustrer le fonctionnement interne des ConvNets 3D, nous avons utilisé la méthode de déconvolution [Zeiler and Fergus, 2014] en utilisant l'outil `tf_cnnvis` [Bhagyesh Vikani, 2017] qui extrait les images de la déconvolution. La figure 3.14 visualise la déconvolution d'un cas de figure de conducteur en état de somnolence, avec utilisation de la méthode de [Zeiler and Fergus, 2014] avec la bibliothèque `tf_cnnvis` [Bhagyesh Vikani, 2017] pour extraire les images de la déconvolution, dans lequel la fonction se concentre sur la personne entière, puis suit l'inclinaison de la tête de la personne dans le reste des images.

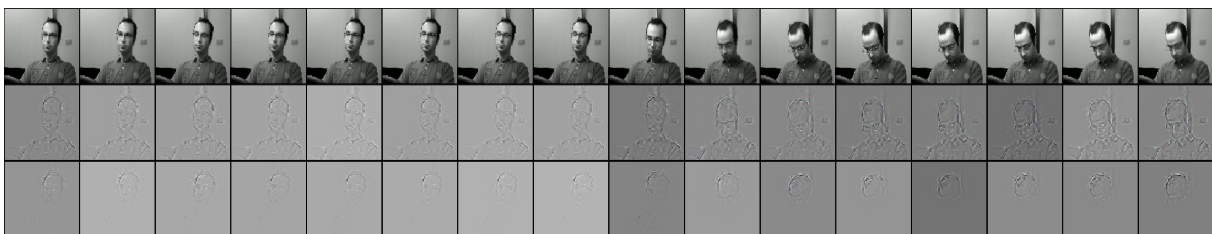


Figure 3.14: Application des filtres formés de la couche conv4b de notre modèle.

3.4.3 Analyse des résultats

Dans cette section, nous présentons les résultats de l'entraînement de notre modèle sous plusieurs mesures et présentation. Ainsi, avec un taux d'acceptation convergeant vers 97% et un taux d'erreur convergeant vers 5.00% sur 100 époques, nous pouvons observer qu'une architecture basée sur le modèle 3D ConvNets donne de parfaits résultats par rapport aux solutions séquentielles. Pour cela, nous avons utilisé d'autres modèles pour enrichir les tableaux de résultats. Nous trouvons les LSTMs [Srivastava, Mansimov, and Salakhudinov, 2015], les réseaux convolutifs récurrents à long terme (LRCNs) [Donahue et al., 2015], le réseau hiérarchique temporel profond (HTDBN) [Weng, Lai, and Lai, 2016], Deep Belief Nets (DBN) [Hinton, Osindero, and Teh, 2006] + SVN [Cristianini, Shawe-Taylor, et al., 2000], et Multilayer Perceptron Neural Network (MLP) [Kumar, 2018].

3.4.3.1 Résultats de précision

Le tableau 3.2 présente le rapport des calculs effectués sur les différents modèles de calcul avec les taux de précision et de validation, ainsi que le taux de test effectué sur les clips non utilisés dans la phase d'entraînement.

Table 3.2: Détection de la somnolence au volant sur l'ensemble de données NTHUDDD.

Modèle	Séquence	Précision %	Validation %	Test %
Méthode proposée	20	97.00	92.19	78.05
	30	97.30	90.19	73.17
	40	97.12	90.40	82.00
LSTMs [Srivastava, Mansimov, and Salakhudinov, 2015]	20	92.51	90.07	80.36
	30	92.58	90.06	78.04
	40	92.71	90.01	80.36
LRCN [Donahue et al., 2015]	20	91.18	82.44	78.04
	30	90.72	81.86	78.04
	40	90.84	80.80	78.04
HTDBN [Weng, Lai, and Lai, 2016]	40	83.04	82.65	80.44
	30	83.26	81.25	78.47
	40	83.04	82.65	80.44
DBN+SVM [Hinton, Osindero, and Teh, 2006] [Cristianini, Shawe-Taylor, et al., 2000]	20	80.65	80.41	76.51
	30	80.01	78.58	75.21
	40	81.12	80.75	76.73
MLP [Kumar, 2018]	20	71.71	73.17	60.97
	30	71.33	73.04	60.97
	40	71.18	72.22	60.97

De plus, les tests dans différentes séquences montrent que nous devons tester avec différents nombres d'images par séquence pour chaque problème et de choisir la meilleure séquence en fonction de l'efficacité et de la vitesse du modèle. Dans notre cas, 40 images

par séquence donnent un meilleur résultat pour notre problème avec le modèle 3D conv3d.

3.4.3.2 Évolution de l'entraînement

La figure 3.15 illustre la convergence du résultat de l'apprentissage de l'architecture ConvNets 3D. Comme on peut le voir, le taux d'acceptation augmente rapidement lors de la première itération et se stabilise à la fin de l'apprentissage. Cependant, comme cette architecture est gourmande en mémoire, il faut plus de temps pour terminer la phase d'apprentissage, le temps d'apprentissage varie en fonction du nombre de couches choisi pour l'architecture 3D ConvNets et de la taille du jeu de données.

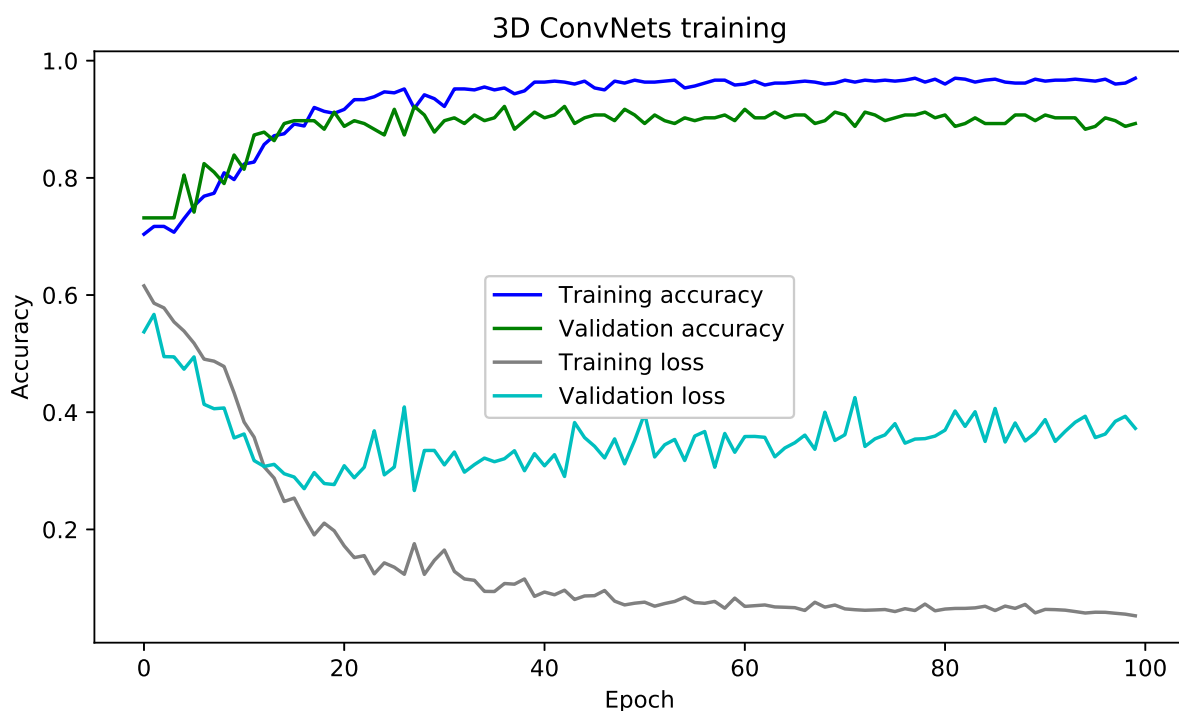


Figure 3.15: Résultat de l'entraînement de ConvNets 3D.

3.4.3.3 Mesure de la performance

Afin de mesurer la performance de notre entraînement, une phase de test a été réalisée avec des données non utilisées dans la phase d'entraînement et de validation, et les résultats obtenus dans la phase de test sont très encourageants.

Le tableau 3.3 présente une analyse du F_1 score [Powers, 2011] de la classification binaire. Le F_1 score est une mesure de la précision d'un test. Il prend aussi en compte simultanément la précision p et le recall r d'un test pour calculer le score. Dans le tableau, nous présentons les différentes mesures de score F_1 réalisées pour le modèle 3D ConvNets, qui montre une précision encourageante qui dépasse 80%.

Table 3.3: Résultats des expériences F_1 Score.

Modèle	Séquence	Précision %	Recall %	F_1 Score %
Proposed method	20	74	100	85
	30	72	92	81
	40	72	92	81
LSTMs [Srivastava, Mansimov, and Salakhudinov, 2015]	20	100	62	77
	30	100	44	61
	40	100	62	77
LRCN [Donahue et al., 2015]	20	75	96	80
	30	80	96	82
	40	75	96	80
HTDBN [Weng, Lai, and Lai, 2016]	20	71	94	79
	30	68	94	77
	40	71	94	79
DBN+SVM [Hinton, Osindero, and Teh, 2006] [Cristianini, Shawe-Taylor, et al., 2000]	20	60	84	64
	30	58	84	64
	40	60	84	61
MLP [Kumar, 2018]	20	61	100	76
	30	61	100	76
	40	61	100	76

3.4.3.4 Comparaison avec d'autres méthodes

Comparé à [Weng, Lai, and Lai, 2016], dans lequel des expériences de somnolence ont été menées sur un ensemble de données. Notre modèle proposé montre un résultat optimal sur le même jeu de données pour la classification de la somnolence.

Le tableau 3.4 montre l'efficacité de l'application de différents modèles de base (SVM, DBN, HMM ou HTDBN) et des nouvelles technologies qui utilisent les réseaux neuronaux profonds.

Table 3.4: Comparaison des 3D ConvNets et de la solution de base sur les scénarios de l'ensemble de données sur la détection de la somnolence au volant.

Method	Accuracy %	F_1 Score %
Driver Alertness Monitoring [Mbouna, Kong, and Chun, 2013]	77.40	43.3
Embedded Smart Cameras [Omidyeganeh et al., 2016]	81.40	43.7
HTDBN [Weng, Lai, and Lai, 2016]	84.82	79.0
Proposed method	92.19	85.0
Multi-timescale CNN [Massoz, Verly, and Van Droogenbroeck, 2018]	94.22	-
HDMS [Chiou et al., 2019]	96.10	81.8

3.4.4 Système proposé

Dans cette section, nous proposons une architecture logicielle pour une application mobile illustrée à la figure 3.16. L'application mobile utilisera l'appareil photo du téléphone pour capturer une séquence d'images à une fréquence de cinq images par seconde. Grâce au modèle de prédiction de la somnolence, les calculs seront effectués

en temps réel ; si le modèle prédit la somnolence, un message visuel et sonore s'affichera sur le téléphone de l'utilisateur.

Dans un deuxième temps, et avec l'accord de l'utilisateur, pour chaque prédiction détectée, un message de retour sera demandé à l'utilisateur pour étiqueter les séquences d'images utilisées. Ensuite, un transfert sera exécuté vers un serveur web, qui récupère les images étiquetées. Ces images seront sauvegardées et analysées pour l'entraînement du module de prédiction. L'objectif de cette étape est de s'assurer que la maintenance du système est évolutive et d'améliorer la précision de la prédiction du modèle.

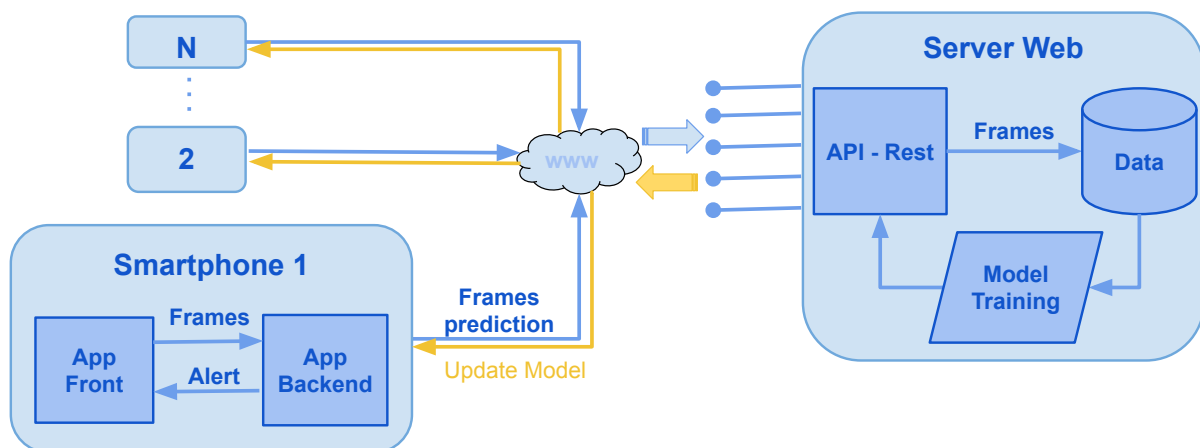


Figure 3.16: Architecture logicielle pour une application mobile et un serveur web.

La plateforme technique proposée pour mettre en œuvre notre architecture logicielle est décrite ci-dessous :

- Application mobile : Kivy et Tensorflow (Python3)
- Serveur Web : Flask et Tensorflow (Python3).
- Base de données : Postgresql (sql).
- Stockage : Système de fichiers ou cloud.

3.4.5 Conclusion

Dans cette étude, nous avons montré comment l'utilisation d'architectures multicouches d'apprentissage profond RNN utilisant des ConvNets 3D, entraînées sur un ensemble de données vidéo à grande échelle, peut aider à résoudre des problèmes séquentiels tels que celui abordé dans cet étude, c'est-à-dire la détection et la prévention de la somnolence.

Notre objectif dans ce travail était de fournir une architecture de réseau neuronal pour un service de détection de la somnolence abordable et portable pour les conducteurs ordinaires. L'architecture ConvNet 3D a atteint une précision proche de 97% sur

toutes les données utilisées. Pour améliorer encore ces résultats, nous devons créer un ensemble de données plus personnalisé et plus approprié au sujet de la somnolence dans un environnement proche de ce que le conducteur peut rencontrer dans un scénario réel.

D'autre part, notre étude s'est basée sur le comportement du conducteur, qui s'est déroulé longtemps avec un changement de posture du conducteur. Cette action est limitée dans le cas où le conducteur s'assoupit sans changer de posture. Notre solution ne peut pas prédire ce cas particulier. Cependant, elle peut s'améliorer avec une association avec les études menées sur la physiologie du conducteur [[Kartsch et al., 2018](#)].

SYSTÈME DE DÉTECTION DES NIDS-DE-POULE ROUTIERS BASÉ SUR LE RÉSEAU NEURONAL RÉCURRENT

SOMMAIRE

4.1 INTRODUCTION	46
4.2 INFORMATIONS DE BASE	47
4.2.1 Méthode Marocain : Le relevé visuel	47
4.2.2 Méthode Française : Le relevé visuel	47
4.2.3 Classification par apprentissage profond	48
4.2.4 Encodeur-décodeur convolutif profond	49
4.3 ARCHITECTURE EN TEMPS DIFFÉRER	49
4.3.1 Approche proposée	49
4.3.2 Expériences	50
4.3.2.1 L'ensemble de données	50
4.3.2.2 Entraînement	50
4.3.3 Analyse du résultat	51
4.3.4 Conclusions	51
4.4 ARCHITECTURE EN TEMPS RÉEL	52
4.4.1 Proposed approach	52
4.4.1.1 Detection	52
4.4.1.2 Segmentation	53
4.4.1.3 Architecture MultiNet	53
4.5 SYSTÈME PROPOSÉ	54
4.5.1 Workflow	54
4.5.1.1 Software architecture	55
4.5.2 Résultats de l'expérience	56
4.5.2.1 Ensemble de donnée	56
4.5.2.2 Entraînement	56
4.5.3 Result analysis	57

LA supervision de la qualité des routes fait partie des principaux axes de recherche en sécurité routière, en exploitant les différentes techniques d'apprentissage automatique. Notre objectif est de proposer un système efficace et précis pour fournir un rapport complet et détaillé sur l'état de la route et anticiper la maintenance à effectuer sur la route ou l'aide à la décision si le système utilise le temps réel dans le contexte d'un exemple de voiture autonome.

Pour cela, nous proposons une architecture d'apprentissage profond en deux étapes. Dans la première étape, nous commençons une segmentation pour déterminer les côtés de la route et limiter la zone de détection des nids de poule. La deuxième étape de détection des nids de poule dans cette zone permet de réduire le nombre de faux positifs détectés par le modèle. Elle réduit les détections de nids de poule qui ne sont pas pertinentes pour cette détection, notamment les nids de poule sur le côté de la route ou sur le trottoir. Pour cela, chaque modèle a été entraîné sur des images issues d'une base de données publique. Le résultat de notre approche a montré son efficacité en atteignant une précision de 93%.

4.1 INTRODUCTION

La sécurité routière fait partie des problèmes mondiaux qui tentent de s'améliorer, les investissements, les lois et les contrôles changeant avec l'évolution rapide du transport mondial.

En outre, la détection des parties endommagées et fissurées de l'infrastructure dépendait uniquement de l'expertise de l'équipe d'ingénierie pour l'expérimentation. Cependant, avec la demande croissante d'analyse de l'état des routes, les rapports d'analyse de l'équipe d'ingénieurs consomment plus de temps et de ressources financières. Par conséquent, les systèmes d'analyse automatique des routes apparaissent comme une alternative ou un outil utile pour accélérer le processus d'inspection des routes et fournir des rapports d'analyse plus pertinents en un temps record.

Pour cela, la communauté des chercheurs s'est penchée sur ce fléau afin de réduire le nombre d'incidents causés, soit par l'erreur humaine, soit par la performance des voitures, soit par la qualité des routes. Dans notre travail, nous nous intéressons à l'analyse de la qualité de la route et aux solutions qui peuvent améliorer l'entretien des routes et donner une estimation objective du budget alloué.

Cette étude présente une approche hybride de la détection des nids de poule basée

sur deux architectures d'apprentissage profond, le réseau de neurones convolutifs (CNN). Une caméra moins chère peut facilement s'adapter à l'intérieur ou à l'extérieur d'une voiture pour récupérer des séquences d'images enregistrées par cette caméra. Par la suite, toutes ces images seront utilisées par notre système, de sorte qu'il commencera à analyser l'état de la route par des techniques d'apprentissage profond. Ce système peut détecter les zones de nids de poule dans diverses conditions de route, telles que des voies sèches, humides et ombragées, ainsi que plusieurs tailles et formes de nids de poule. Pour évaluer la performance, nous comparons le résultat avec la méthode conventionnelle d'apprentissage automatique. Les performances obtenues indiquent que notre système est capable de détecter les nids-de-poule de manière efficace.

4.2 INFORMATIONS DE BASE

4.2.1 Méthode Marocain : Le relevé visuel

Le principe de relevé pratiqué pour le réseau routier revêtu est basé sur le constat à des intervalles réguliers de 200m (5 relevés / Km), de l'existence ou de l'absence de dégradations observées sur la zone d'un profil en travers de 20 à 30m de longueur, on indiquait la sévérité par trois niveaux.

- Cotation 0 : Absence de dégradations
- Cotation 1 : Dégradations peu nuisible
- Cotation 2 : Dégradations nuisibles

Principe général Pour chaque section de route, l'opérateur devra définir un sens de parcours pour le relevé. Partant du point de départ de la section à relever, l'opérateur relève les différentes dégradations observées sur le champ de la zone d'observation et donne des cotations. Ensuite, il parcourt la distance de 200 m avec son véhicule dans le sens de relevé sans tenir compte des dégradations intermédiaires existante, puis il effectue un deuxième relevé conformément au premier et ainsi de suite jusqu'à l'achèvement de la section.

4.2.2 Méthode Française : Le relevé visuel

L'objectif de cette méthode est de localiser les dégradations présentes, en précisant leur type, leur gravité et l'extension de chaussée concernée. Ainsi, la section relevée est parcourue mètre par mètre en notant l'étendue chaque dégradation.

Les besoins de relevé des dégradations pour les études routières sont couverts par sept modes opératoires indiqués dans le tableau 4.1, selon l'objectif, les conditions d'exploitation et le type de chaussée.

Objectifs de l'étude	Conditions d'exploitation et type de chaussée			
	Chaussées souples, semi-rigides, inverses, mixtes, bac			Dalles béton
	Rase campagne et VRU*		Urbain hors VRU	Tous sites
	forts trafics**	faibles trafics **	Tous trafics	
Diagnostic - entretien - renforcement	M1	M1 ou M2	M6	M7
Programmation	M2	M2 ou M3	M6	M7
évaluation - surveillance	M3	M3 ou M4	M6	M7
suivi de sections témoins	M5	M5	-	M5

VRU*: Voies Rapides Urbaines ; trafics**: forts ($\geq T2$), faibles ($< T2$)

Figure 4.1: Les conditions d'exploitation en fonction des objectifs de l'étude.

4.2.3 Classification par apprentissage profond

Les réseaux neuronaux convolutifs [Krizhevsky, Sutskever, and Hinton, 2012] (ConvNets ou CNN) sont une classe de réseaux neuronaux qui se sont révélés très efficaces dans des domaines tels que la reconnaissance et la classification des images. Les ConvNets ont réussi à identifier des visages, des objets et des panneaux de signalisation, en dehors de l'alimentation de la vision dans les robots et les voitures autonomes. Dans notre travail, nous verrons la classification des poules nées en utilisant l'apprentissage profond (CNN).

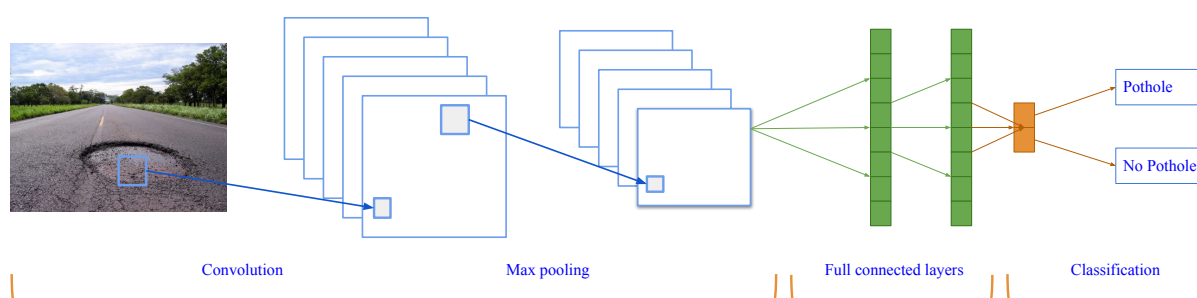


Figure 4.2: Vue globale de l'architecture cnn.

La Fig. 4.2 représente l'architecture de base de ConvNet, qui contient des fonctionnalités de base telles que le Pooling, la couche FC. De nombreuses autres architectures se sont développées à partir de l'architecture de base de la convolution, ce qui a permis des avancées spectaculaires en intelligence artificielle.

4.2.4 Encodeur-décodeur convolutif profond

L'architecture de réseau neuronal encodeur-décodeur se combine avec une nouvelle stratégie pour améliorer la cohérence globale des étiquettes, qu'accompagne un modèle de segmentation d'image amélioré. Les prédictions de distribution d'étiquettes extraites du réseau SegNet [Badrinarayanan, Kendall, and Cipolla, 2017] sont étudiées et utilisées dans le projet d'étiquetage d'images Fig. 4.3. Pour notre travail, ce type de modèle permet d'isoler la zone routière et de cibler les zones à classer et d'augmenter le taux de précision.

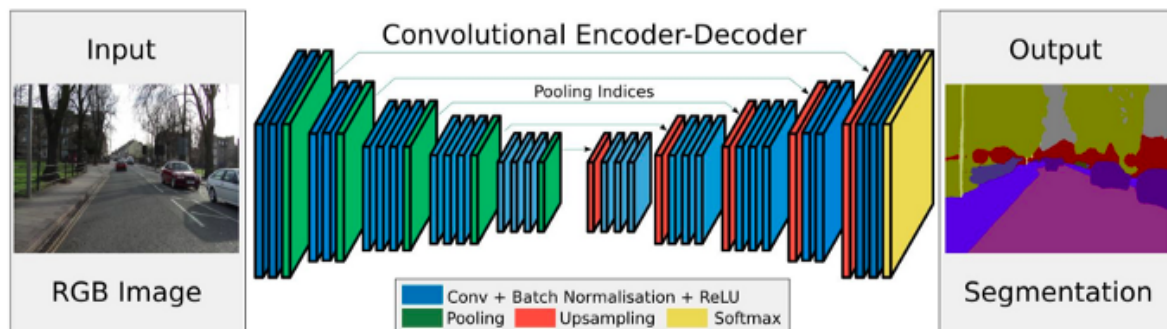


Figure 4.3: Encodeur-Décodeur à convolution.

La figure ci-dessus montre l'architecture de Segnet. Le réseau est uniquement convolutif, car il n'y a pas de couches étroitement connectées. Un décodeur suréchantillonne son entrée en utilisant les indices de pool transférés de son codeur pour produire une ou plusieurs cartes de caractéristiques éparses. Il effectue ensuite une convolution avec un banc de filtres entraînaables pour densifier la carte de caractéristiques, une carte de caractéristiques finale de sortie du décodeur alimentant un classificateur soft-max pour une classification par pixel.

4.3 ARCHITECTURE EN TEMPS DIFFÉRER

4.3.1 Approche proposée

L'architecture proposée utilisée en temps différé Figure 4.4, la combinaison de la segmentation et de la détection augmente le taux de précision. Cependant, l'inconvénient de cette approche est que la gestion du stockage des données sera souvent volumineuse. La conception de l'architecture proposée utilise l'apprentissage automatique fragmenté pour réduire la charge des ressources informatiques massives en ce qui concerne la puissance de traitement des données et la taille de la mémoire.

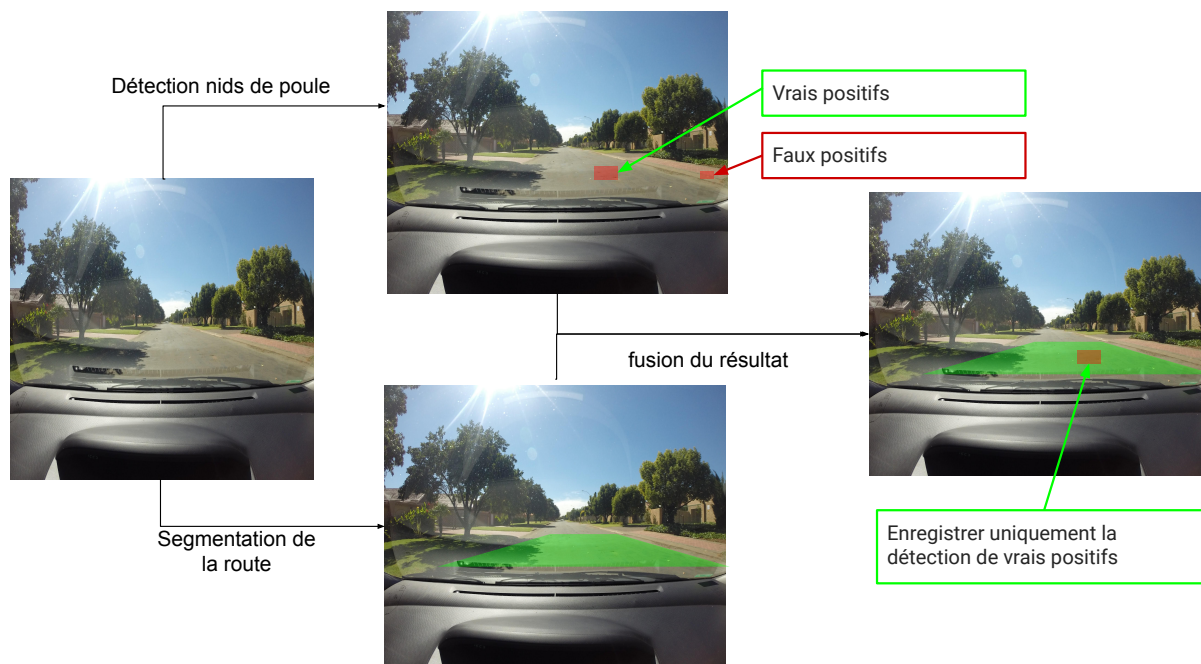


Figure 4.4: Proposition d'architecture pour la détection des nids de poule

4.3.2 Expériences

4.3.2.1 L'ensemble de données

Pour entraîner notre modèle, nous avons utilisé le jeu de données [Nienaber, Booyen, and Kroon, 2015; Nienaber, Kroon, and Booyen, 2015]. Les images du jeu de données utilisées pour cette étude ont été capturées par une caméra GoPro Hero 3+ avec une résolution de 3680×2760 . Le jeu de données est composé de deux ensembles différents, l'un considéré comme simple et l'autre plus complexe. Cette qualité permet de diviser les données en deux groupes, les données utilisées pour l'entraînement qui représente 70 % de l'ensemble de données. Cette division permet au modèle de mieux valider son apprentissage sur la partie du jeu de données sur laquelle il n'a pas fait d'apprentissage, ce qui augmente significativement la précision du modèle.

4.3.2.2 Entraînement

L'entraînement du jeu de données est divisé en deux étapes. La première étape est l'entraînement sur la segmentation de la route, avec une division d'une image en deux classes, comme la première classe qui entoure la route et la deuxième classe est le fond qui représente le reste de l'image, après l'étape d'étiquetage est nécessaire parce que le jeu de données ne fournit pas cet étiquetage.

La deuxième étape de l'apprentissage est la détection des nids de poule. Cette étape de l'apprentissage est basée sur l'étiquetage déjà effectué dans le travail du jeu de données utilisé.

À la fin de l'entraînement, la validation est effectuée en combinant deux modèles de segmentation et de détection pour mesurer la performance de chaque paire.

4.3.3 Analyse du résultat

Dans cette section, nous présentons les résultats de l'apprentissage de plusieurs paires d'architecture étudiées dans le tableau 4.1, avec les taux de précision et de validation obtenus, ainsi que le taux de test effectué sur des images non utilisées dans la phase d'apprentissage. les modèles suivent et utilisé dans notre entraînement afin de comparer l'efficacité de notre solution, tel qu'on a utilisé pour la segmentation les modèles DenseASPP [Yang et al., 2018], Unet[Ronneberger, Fischer, and Brox, 2015], Resnet [He et al., 2016] et Unet-VGG16 [Simonyan and Zisserman, 2014].

Table 4.1: Résultats expérimentaux du calcul.

Segmentation	Method	Accuracy %	Validation %	Test %
DenseASPP	EfficientNet [Tan and Le, 2019]	95.40	94.51	93.3
	YOLOv3 [Redmon and Farhadi, 2018]	92.32	92.9	92.7
	GPipe [Huang et al., 2018]	84.82	79.0	83.0
	SENet [Hu, Shen, and Sun, 2018]	82.7	85.0	83.04
	GoogleNet [Al-Qizwini et al., 2017]	74.8	71.45	73.38
	AlexNet [Wu, Shen, and Van Den Hengel, 2019]	73.55	71.8	70.64
Unet-Resnet	EfficientNet	94.13	91.8	91.3
	YOLOv3	94.57	90.47	90.64
	GPipe	84.6	78.04	80.55
	SENet	82.0	81.79	81.04
	GoogleNet	73.47	71.17	70.15
	AlexNet	70.75	70.64	69.53
Unet-VGG16	EfficientNet	94.06	92.21	90.32
	YOLOv3	94.15	90.88	91.13
	GPipe	85.47	77.51	80.65
	SENet	81.02	80.78	79.13
	GoogleNet	72.5	70.94	68.65
	AlexNet	71.45	70.48	70.25

Notre approche indique que la combinaison des deux modèles augmente la précision et réduit les faux positifs liés à la détection des nids de poule qui ne se trouvent pas dans la zone de la route.

4.3.4 Conclusions

Dans cette contribution, nous avons présenté la détection des nids de poule en combinant deux architectures d'apprentissage profond, la segmentation et la détection. Les deux modèles de calcul utilisés dans notre architecture et la formation sur un ensemble

de données publiques contenant des images recueillies à partir de différentes zones routières urbaines et rurales. Les résultats expérimentaux ont confirmé qu'en utilisant notre approche ; les nids de poule sont bien détectés et atteignent une grande précision.

Ce résultat encourage également les autorités gouvernementales à utiliser notre système pour surveiller et évaluer l'état des routes dans tous les territoires du pays.

Enfin, pour toute solution, il y a toujours une limite. C'est le cas de notre travail, à savoir que notre solution n'est pas optimale pour une utilisation en temps réel. Pour cette raison, dans la suite de notre travail, nous allons essayer de fusionner les deux modèles, la segmentation et la détection, en un seul modèle qui peut faire les deux approches simultanément, de sorte que nous pouvons passer à la détection et à la segmentation en temps réel sans perdre la précision globale.

4.4 ARCHITECTURE EN TEMPS RÉEL

4.4.1 Proposed approach

Dans cette section, nous allons analyser en détail les caractéristiques que l'architecture Multinet [Teichmann et al., 2018] aborde, à savoir la détection et la segmentation sémantique. Pour ce faire, notre étude se concentre sur les architectures basées sur l'apprentissage profond.

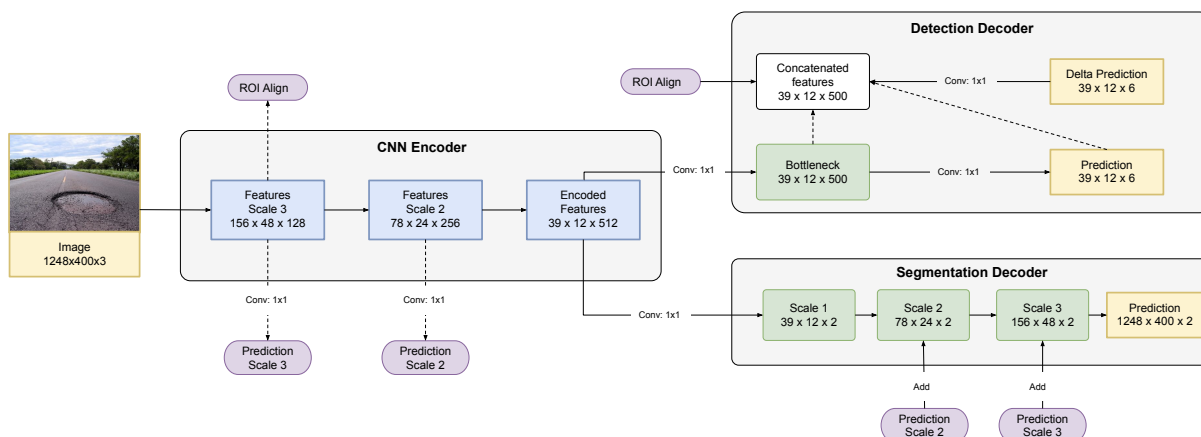


Figure 4.5: Architecture MultiNet.

4.4.1.1 Detection

Les traitements classiques d'apprentissage profond de la détection des objets, il se fait en deux phases, la première phase. C'est le choix des zones de détection [Lampert, Blaschko, and Hofmann, 2008; Hosang, Benenson, and Schiele, 2014; Hosang et al., 2015], ces zones sont d'abord localisées, après via un réseau convolutif [Wang et al., 2015; Gavrilescu et al., 2018] elles sont classées. Dans la continuité, des optimisations sur l'efficacité que l'on peut avoir via un réseau convolutif neuronal (CNN) pour la

partie de la création des suggestions [Szegedy et al., 2014; Gavrilescu et al., 2018], on peut également utiliser trois dimensions [Chen et al., 2016; Chen et al., 2015]. Enfin, diverses techniques proposent un seul réseau profond qui ne peut être entraîné qu'une seule fois et compléter les détections demandées [Carion et al., 2020; Liu et al., 2016; Sermanet et al., 2013a]. Les méthodes basées sur des propositions ont l'avantage d'être très rapides en apprentissage et en inférence. Dans notre travail, nous proposons une architecture qui contient un détecteur facile à entraîner du début à la fin, et qui augmente significativement la performance du modèle généré. Nous soutenons que le principal intérêt des méthodes basées sur les propositions dans notre travail est leur capacité à avoir des tailles de caractéristiques ajustables. Cela a inspiré notre mise en œuvre de la mise en commun des régions d'intérêt (ROI).

4.4.1.2 Segmentation

Inspirés du succès de l'apprentissage profond, les classificateurs basés sur les CNN ont été adaptés à la tâche de segmentation sémantique. Les premières approches ont utilisé l'efficacité inhérente des CNN pour mettre en œuvre une fenêtre glissante implicite. Les réseaux entièrement convolutifs [Long, Shelhamer, and Darrell, 2015] ont été proposés pour modéliser la segmentation sémantique à l'aide d'un pipeline d'apprentissage profond qui peut être formé de bout en bout. Des convolutions transposées [Zeiler et al., 2010; Dumoulin and Visin, 2016; Krizhevsky, Sutskever, and Hinton, 2012] sont utilisées pour surimposer des caractéristiques à faible résolution. Une variété de saveurs de réseaux complètement convolutifs plus profonds a été suggérée depuis [Badrinarayanan, Kendall, and Cipolla, 2017; Noh, Hong, and Han, 2015; Ronneberger, Fischer, and Brox, 2015; Papandreou et al., n.d.]. Avec la combinaison de réseaux entièrement convolutifs, de meilleurs résultats sont obtenus [Chen et al., 2017; Zheng et al., 2015]. [Zheng et al., 2015; Schwing and Urtasun, 2015] ont montré que l'inférence du champ moyen dans les champs aléatoires conditionnels peut être présentée comme un réseau récurrent permettant un apprentissage de bout en bout. Des convolutions deux dilatées ont été introduites dans [Zhang et al., 2014] pour augmenter la taille du champ récepteur sans perdre la résolution. Les techniques ci-dessus, associées aux réseaux résiduels [He et al., 2016], constituent actuellement l'état de l'art.

4.4.1.3 Architecture MultiNet

Dans cette section, l'architecture Multinet est expliquée avec plus de détails sur les couches de l'architecture, pour un raisonnement conjoint sur la segmentation, la détection d'objets et la classification d'images. Dans notre travail, nous proposons une architecture qui a un encodeur commun pour les deux tâches avec deux branches, dès

que chaque branche utilise son propre décodeur pour une tâche donnée. La figure 4.5 illustre l'architecture proposée dans notre travail. En outre, MulitiNet est une architecture qui peut être entraînée du début à la fin, pour que l'inférence sur les sous-modèles de l'architecture Multinet puisse être inférieure à 50 ms.

4.5 SYSTÈME PROPOSÉ

4.5.1 Workflow

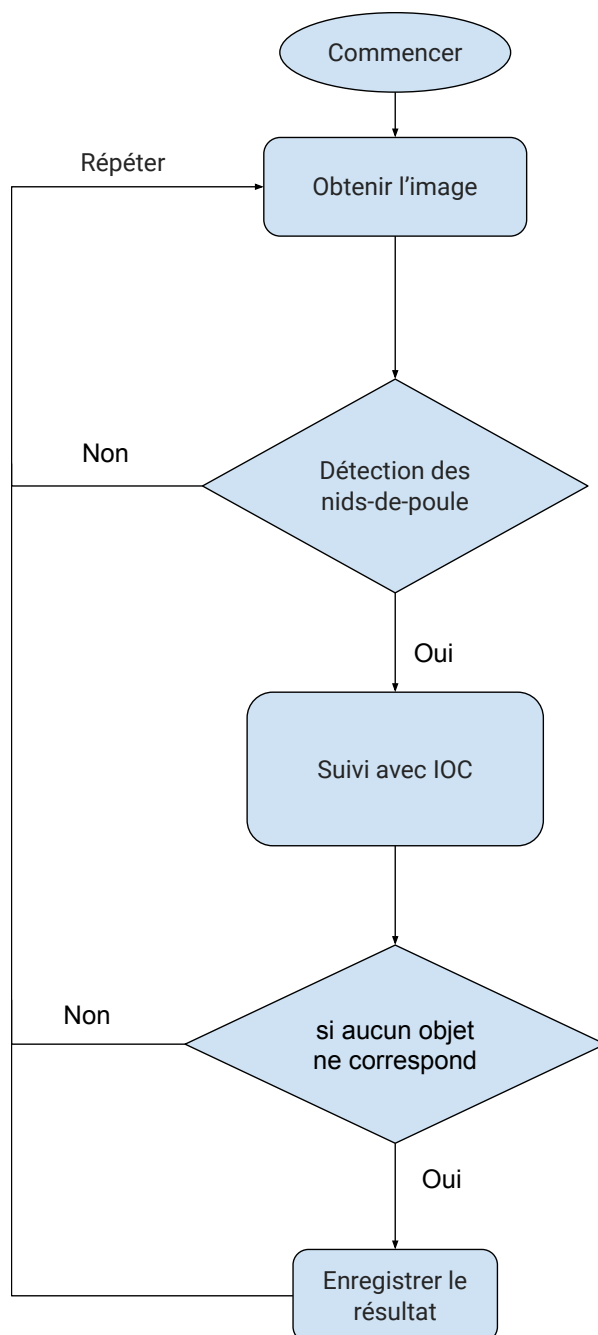


Figure 4.6: Diagramme de flux de travail

Afin de rendre notre système optimal et de répondre aux besoins de l'entreprise, un diagramme de flux de travail Figure 4.6 présente les différentes étapes du traitement de l'information. Par exemple, commencé par retrouver l'image comme point d'entrée, l'étape suivante est la détection des nids de poule comme décrit dans le paragraphe approche, après la détection des nids de poule, l'étape suivante est une étape de principe, car elle évite de détecter plusieurs fois les mêmes nids de poule, comme l'algorithme intersection over union (IOU) [Rezatofighi et al., 2019] permet de tracer le même objet sur une séquence d'images, à la fin les informations de détection sont sauvegardées pour générer un rapport exploitable par l'utilisateur du système.

4.5.1.1 Software archeticture

L'objectif de notre étude est de trouver un compromis entre les performances et les coûts d'installation, pour cette raison toute la complexité liée à l'installation, la configuration et l'optimisation du système sera centralisée dans un serveur distant, permettant de partager le même service avec de nombreux utilisateurs grâce à des API (Application Programming Interface), en particulier que la partie la plus coûteuse est le côté serveur (voir figures 4.7), comme côté serveur, nous chargeons simultanément les modèles de détection et de segmentation, et les services de traitement des données, de même que cette architecture réduit le coût de la maintenance et la supervision du système proposé. D'autre part, du côté utilisateur, il suffit d'un smartphone équipé d'Internet pour envoyer les images enregistrées en temps réel ou en temps différé, et lorsque l'analyse de chaque route est terminée, l'utilisateur peut charger les résultats sous plusieurs formes de données faciles à partager ou à manipuler avec d'autres systèmes, ce qui permet de donner une estimation plus précise de l'état de la route et du budget nécessaire pour effectuer les travaux d'entretien.

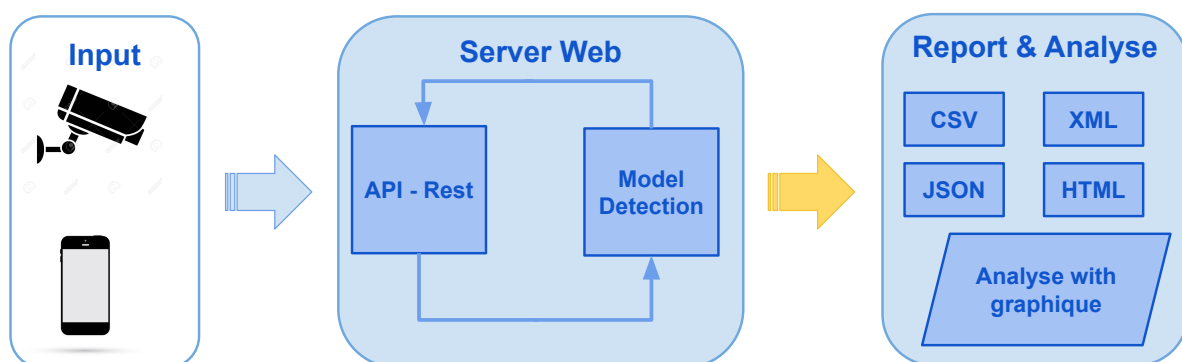


Figure 4.7: Système proposé.

4.5.2 Résultats de l'expérience

4.5.2.1 Ensemble de donnée

Dans notre travail, nous avons utilisé un jeu de données public initié dans ce travail [Nienaber, Booyesen, and Kroon, 2015; Nienaber, Kroon, and Booyesen, 2015]. Les données référencées dans ce jeu de données sont collectées via une caméra Go Pro Hero3+ avec une haute définition qui mesure 3680×2760 , pour augmenter la qualité des images collectées. Le jeu de données est divisé en deux groupes de données, le premier groupe de données qualifiées simples et le second groupe de données est qualifié complexe. Ainsi pour optimiser l'apprentissage et avoir une meilleure précision du modèle, nous avons divisé le jeu de données en deux parties, la première partie représente 80% des données du jeu de données, car cette partie est utilisée dans l'apprentissage du modèle, après la seconde partie des données qui présente 20% du jeu de données, est utilisée pour valider l'apprentissage.

4.5.2.2 Entraînement

Étant donné la capacité de l'architecture MultiNet, il est possible d'entraîner simultanément la segmentation des routes et la détection des nids de poule. Pour cela, l'entraînement est effectué simultanément, en passant les sous-modèles MultiNet, la configuration de l'étiquetage des données relatives à la segmentation de la route et la configuration de l'étiquetage des données relatives à la détection des nids de poule.

Ensuite, la validation du modèle générée à la fin de l'apprentissage est effectuée de telle sorte qu'un prétraitement des données de détection est réalisé afin d'éliminer les faux positifs des nids de poule détectés, de telle sorte que chaque nid de poule détecté en dehors de la segmentation de la route est considéré comme un faux positif, ce prétraitement permet donc d'augmenter le taux de performance de notre apprentissage.



Figure 4.8: Visualisation du résultat de la sortie.

Dans la figure 4.8, nous présentons un exemple du résultat de sortie qui combine les deux tâches : détection et segmentation, tel que, la route est segmentée en bleu et les nids de poule encadrés en rouge, ce qui a permis de distinguer explicitement les nids de poule. Détecter en dehors de la zone de la route, est donc réduit le taux de faux positifs.

4.5.3 Result analysis

Table 4.2: Résultats des expériences des méthodes appliquées.

Method	Précision %	Validation %	Test %	fps	traitement
Faster-RCNN	79.45	75.44	77.62	6	20.5 ms
VGG-pool5	81.91	75.2	80.20	7.2	17.32 ms
ResNet	85.21	89.15	80.15	8.6	12.48 ms
Multinet	91.14	89.98	90.64	11.7	9.26 ms

Nous avons mené une série d'expériences comparant Multinet à d'autres architectures étroitement liées, comme le montre le tableau des résultats 4.2, telles que Fast-RCNN [Ren et al., 2015] et VGG pool5 [Teichmann et al., 2018], qui présentent des architectures utilisées dans le domaine de la détection et de la segmentation.

Actuellement, de l'analyse des résultats des modèles, nous avons observé que la vitesse de calcul est plus rapide lorsque l'inférence est exécutée en parallèle par rapport à une inférence exécutée en mode synchrone. Par ailleurs, il a été constaté que la première couche de calcul est plus coûteuse en termes de traitement. De manière générale, nous avons constaté que l'architecture Multinet offre un excellent rapport entre performance et vitesse de calcul.

4.5.4 Conclusion

Le détecteur de nids de poule utilisant une architecture Multinet a atteint une précision de 91,14%. La vitesse de traitement du SSD était faible et ne peut pas être utilisée pour des applications en temps réel. La vitesse de traitement est d'environ 12 FPS, ce qui peut être considéré comme suffisamment élevé pour notre application en temps réel.

Par conséquent, notre détecteur de nids de poule utilisant une architecture Multinet peut être considéré comme un système robuste en temps réel qui peut être utilisé dans des scénarios réels. Nos futurs travaux comprendront un ensemble de données plus important avec plus d'images de nids-de-poule provenant de différentes routes et présentant plusieurs degrés de gravité ainsi que différentes conditions d'éclairage et météorologiques.

En outre, nous incluons les trous d'homme dans l'entraînement de notre système. Les trous d'homme et les nids de poule ont des caractéristiques communes, ce qui constitue une amélioration importante de notre système actuel pour différencier les trous d'homme des nids de poule. Ainsi, nous avons l'intention de déployer notre système dans de nombreuses voitures pour analyser l'état de la route en direct et en temps réel en ajoutant un GPS pour obtenir les coordonnées des nids de poule pour la maintenance.

LES morts causés par les accidents de la route sont toujours en augmentation et sont actuellement considérés comme un problème médical général important. Les développements les plus récents ont fait progresser les connaissances et les capacités technologiques des véhicules, leur permettant de visualiser et analyser les états de la route afin d'éviter les accidents et de sécuriser les voyageurs. Par conséquent, l'analyse des comportements des conducteurs sur la route est devenue l'un des principaux sujets de recherche de ces dernières années, en particulier la somnolence, car elle constitue le facteur le plus élevé d'accidents et la première source de décès sur les routes.

Dans le chapitre 1 et 2, nous avons montré comment l'utilisation d'architectures multicouches d'apprentissage profond RNN utilisant des ConvNets 3D et LSTM, entraînées sur un ensemble de données vidéo à grande échelle, peut aider à résoudre des problèmes séquentiels tels que celui abordé dans cet article, c'est-à-dire la détection et la prévention de la somnolence.

Notre objectif dans ce travail était de fournir une architecture de réseau neuronal pour un service de détection de la somnolence abordable et portable pour les conducteurs ordinaires. L'architecture ConvNet 3D a atteint une précision proche de 97% sur toutes les données utilisées. Pour améliorer encore ces résultats, il est nécessaire de créer un ensemble de données plus personnalisé et plus approprié au sujet de la somnolence dans un environnement proche de ce que le conducteur peut rencontrer dans un scénario réel.

Dans le chapitre 5 et 6, nous avons présenté solution pour analyser la qualité de la route, tel que notre solution permet la détection des nids de poule en combinant deux architectures d'apprentissage profond, la segmentation et la détection. Les deux modèles de calcul utilisés dans notre architecture et la formation sur un ensemble de données publiques contenant des images recueillies à partir de différentes zones routières urbaines et rurales. Les résultats expérimentaux ont confirmé qu'en utilisant notre approche ; les nids de poule sont bien détectés et atteignent une grande précision.

Ce résultat encourage également les autorités gouvernementales à utiliser notre système pour surveiller et évaluer l'état des routes dans tous les territoires du pays.

Dans le futur, il serait intéressant de pouvoir industrialiser les techniques de

l'intelligence artificielle pour la détection de la somnolence et la détection des nées de poules dans le domaine de l'industrie de l'automobile, notamment les voitures autonomes ou intelligente afin d'améliorer le quotidien des usagers de la route et leurs sécurités.

Aussi, dans les perspectives de notre travail, les villes intelligente ou smart city, tel que les villes sont saisies par la révolution numérique et l'intelligence artificielle, lesquelles transforment l'économie dans son ensemble. Le sujet et les perspectives concernent donc toutes les villes, pour cela les solutions proposer via notre recherche peut être utilisé dans les villes intelligentes, comme la gestion de la congestion afin de réduire la pollution, ou encore l'analyse de la circulation quotidienne afin de réservation des voies pour la mobilité douce comme les vélos et les motos à vitesse réduite.

Journaux

1. Ed-Doughmi Younes and El Ayachi Rachid.
Real time detection of potholes using Multinet.
1. Ed-Doughmi Younes, Najlae Idrissi, and Youssef Hbali.
Real-Time System for Driver Fatigue Detection Based on a Recurrent Neuronal Network.
Journal of Imaging, 2020, vol. 6, no 3, p. 8.

Conférences

1. Ed-doughmi Younes and El Ayachi Rachid. *Detection System Potholes on Roads based-on Recurrent Neuronal Network.* In : **5 Third World Conference on Smart Trends in Systems Security and Sustainability (WorldS4)**. IEEE/Springer, 2021. **London, UK.**
2. Ed-doughmi Younes and Idrissi Najlae.
Driver fatigue detection using recurrent neural networks. In : **Networking, Information Systems Security (NISS19)**. ACM, New York, NY, USA, March 27–29, 2019. p. 1-6. **Rabat, Maroc.**

B TECHNIQUES D'APPRENTISSAGE

SOMMAIRE

B.1 VALIDATION CROISÉE	63
B.2 TECHNIQUES D'APPRENTISSAGE	63
B.2.1 AdaBoost	64
B.2.2 Réseaux de neurones	66

B.1 VALIDATION CROISÉE

La validation croisée [Mitchell et al., 1997] ou la k-fold cross validation est une méthode d'évaluation qui consiste à diviser la collection de test en k sous ensembles de tailles égales (approximativement), d'utiliser $k - 1$ sous ensembles pour l'apprentissage des centres d'intérêt dans notre cas, et le k ème sous ensemble pour le test. On réitère ensuite le processus k fois pour chacun des centres d'intérêt évalué. Dans nos études, nous avons appliqué la procédure K-validation croisée (avec $k = 10$). Suivant cette procédure, nous mesurons la précision de la classification du classificateur considéré.

B.2 TECHNIQUES D'APPRENTISSAGE

La notion d'apprentissage étant importante, nous allons commencer par effectuer un rappel. L'apprentissage par induction permet d'arriver à des conclusions par l'examen d'exemples particuliers. Il se divise en apprentissage supervisé et non supervisé. Les exemples particuliers sont représentés par un ensemble de couples d'entrée/sortie. Le but est d'apprendre une fonction qui correspond aux exemples vus et qui prédit les sorties pour les entrées qui n'ont pas encore été vues. Les entrées peuvent être des descriptions d'objets et les sorties la classe des objets donnés en entrée.

B.2.1 AdaBoost

Le mot *boosting* s'applique à des méthodes générales capables de produire des décisions très précises (au sens d'une fonction de perte) à partir d'un ensemble de règles de décision « faibles », c'est-à-dire dont la seule garantie est qu'elles soient un peu meilleures que le hasard. Ces méthodes s'appliquent aussi bien à l'estimation de densité qu'à la régression ou à la classification. Pour simplifier, nous nous concentrons ici sur la tâche de classification binaire. Dans sa version « par sous-ensembles », cette technique fait produire à l'algorithme trois résultats selon la partie de l'ensemble d'apprentissage sur laquelle il apprend, puis combine les trois apprentissages réalisés pour fournir une règle de classification plus efficace. Examinons d'abord cette technique avant de voir comment la généraliser à l'aide de distributions de probabilités sur les exemples. Schapire [Schapire and Freund, 2013] a développé le premier algorithme de boosting pour répondre à une question de Kearns : est-il possible de rendre aussi bon que l'on veut un algorithme d'apprentissage « faible », c'est-à-dire un peu meilleur que le hasard ? Schapire montra qu'un algorithme faible peut toujours améliorer sa performance en étant entraîné sur trois échantillons d'apprentissage bien choisis. Nous ne nous intéressons ici qu'à des problèmes de classification binaire. L'idée est d'utiliser un algorithme d'apprentissage qui peut être de nature très diverse (un arbre de décision, une règle bayésienne de classification, une décision dépendant d'un hyperplan, etc.) sur trois sous-ensembles d'apprentissage.

1. On obtient d'abord une première hypothèse h_1 sur un sous-échantillon S_1 d'apprentissage de taille $m_1 < m$ (m étant la taille de S l'échantillon d'apprentissage disponible).
2. On apprend alors une deuxième hypothèse h_2 sur un échantillon S_2 de taille m_2 choisi dans $S - S_1$ dont la moitié des exemples sont mal classés par h_1 .
3. On apprend finalement une troisième hypothèse h_3 sur m_3 exemples tirés dans $S - S_1 - S_2$ pour lesquels h_1 et h_2 sont en désaccord.
4. L'hypothèse finale est obtenue par un vote majoritaire des trois hypothèses apprises : $H = \text{vote majoritaire}(h_1, h_2, h_3)$

Le théorème de Schapire sur la « force de l'apprentissage faible » prouve que H a une performance supérieure à celle de l'hypothèse qui aurait été apprise directement sur l'échantillon S . Une illustration géométrique du *boosting* selon cette technique de base est donnée dans les figures B.1, B.2 et B.3.

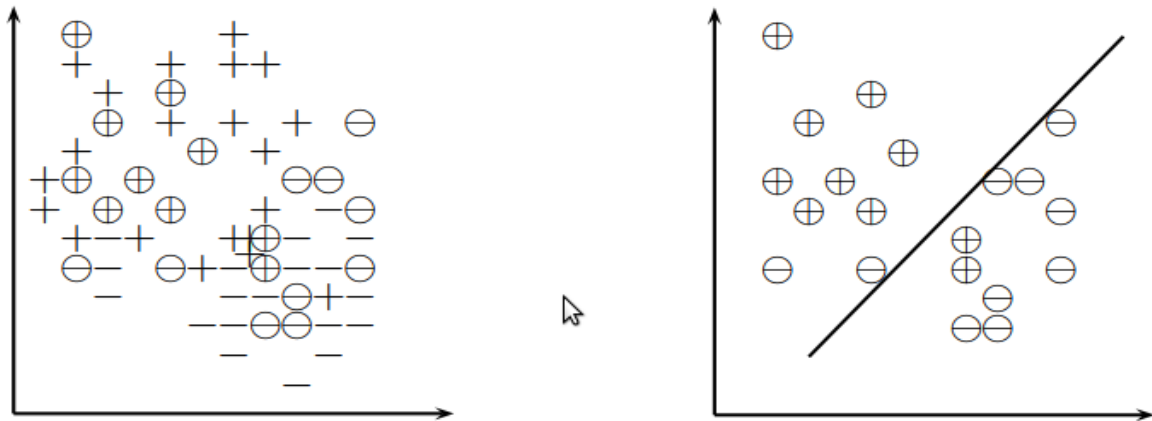


Figure B.1: A gauche : l'ensemble d'apprentissage S et le sous-ensemble S_1 (points entourés). A droite : l'ensemble S_1 et la droite C_1 apprise sur cet ensemble.

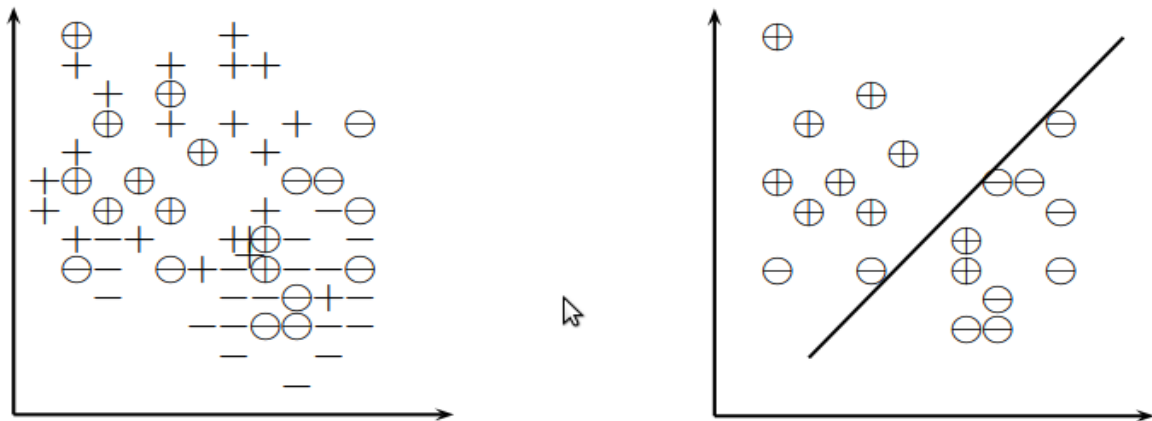


Figure B.2: A gauche : l'ensemble $S - S_1$ et la droite C_1 apprise sur S_1 . A droite : un ensemble S_2 inclus dans $S - S_1$ parmi les plus informatifs pour C_1 (Points entourés)

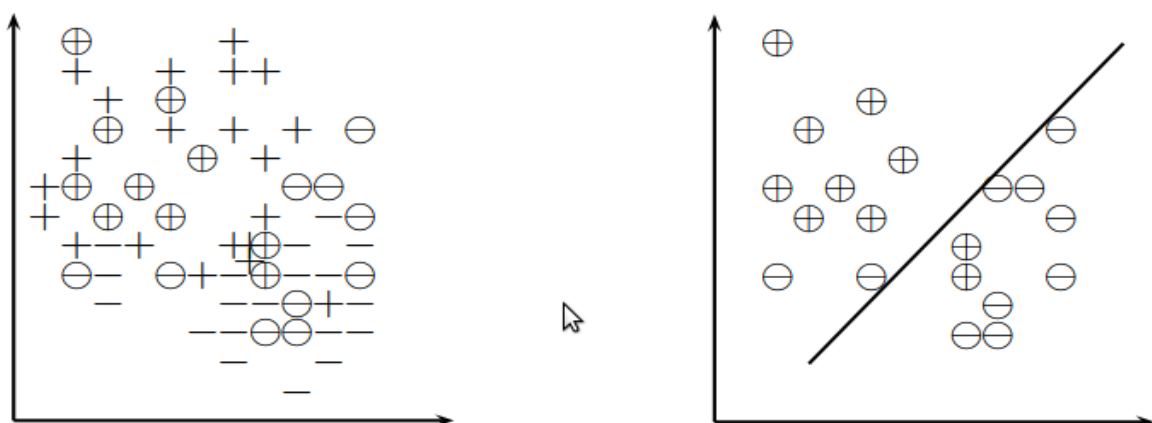


Figure B.3: A gauche : l'ensemble S_2 et la droite séparatrice C_2 apprise sur cet ensemble. Au centre : l'ensemble $S_3 = S - S_1 - S_2$ et la droite séparatrice C_3 apprise sur cet ensemble. A droite : l'ensemble S et la combinaison des 3 droites séparatrices apprise sur cet ensemble ..

B.2.2 Réseaux de neurones

Les réseaux de neurones sont composés d'éléments simples (ou neurones) fonctionnant en parallèle. Ces éléments ont été fortement inspirés par le système nerveux biologique. Comme dans la nature, le fonctionnement du réseau (de neurone) est fortement influencé par la connections des éléments entre eux. On peut entraîner un réseau de neurone pour une tâche spécifique (reconnaissance de caractères par exemple) en ajustant les valeurs des connections (ou poids) entre les éléments (neurone) (voir la figure B.4). En général, l'apprentissage des réseaux de neurones est effectué de sorte que pour une entrée particulière présentée au réseau corresponde une cible spécifique. L'ajustement des poids se fait par comparaison entre la réponse du réseau (ou sortie) et la cible, jusqu'à ce que la sortie corresponde (au mieux ?) à la cible. On utilise pour ce type d'apprentissage dit supervisé un nombre conséquent de pair entrée/sortie. L'apprentissage permet aux réseaux de neurones de réaliser des taches complexes dans différents types d'application (classification, identification, reconnaissance de caractères, de la voix, vision, système de contrôle...). Ces réseaux de neurones peuvent souvent apporter une solution simple à des problèmes encore trop complexes ne pouvant être résolus rapidement par les ordinateurs actuels (puissance de calcul insuffisante) ou par notre manque de connaissances.

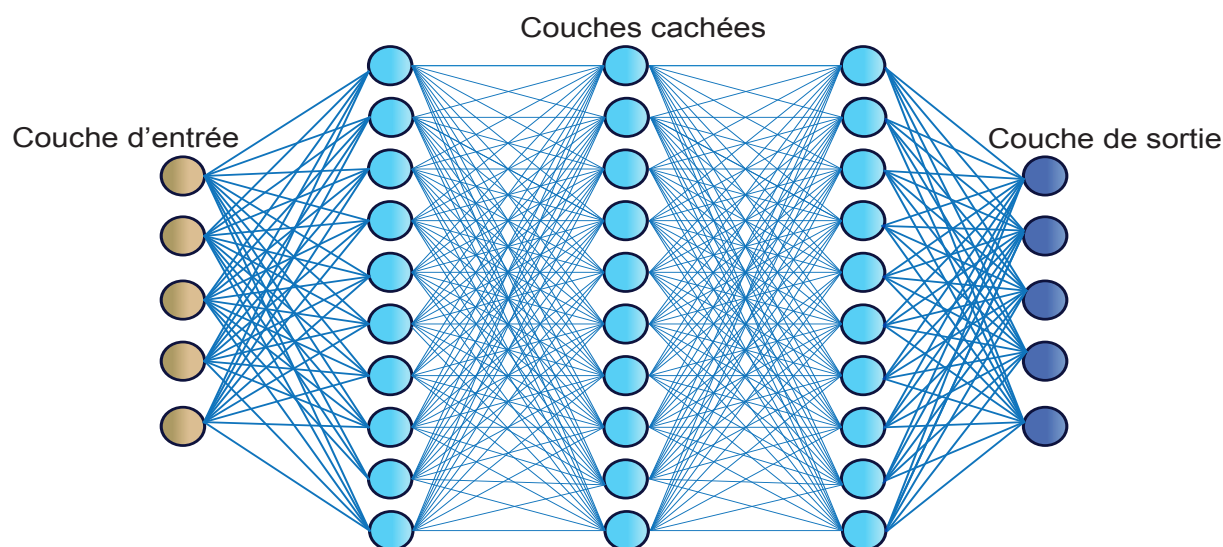


Figure B.4: Réseaux de neurones

BIBLIOGRAPHIE

A

- [Abadi et al., 2016] Abadi, Martín et al. (2016). “TensorFlow: A System for Large-Scale Machine Learning.” In: *OSDI*. Vol. 16, pp. 265–283 (cit. on pp. 30, 39, 67).
- [Abdel-Hamid et al., 2012] Abdel-Hamid, Ossama et al. (2012). “Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition”. In: *2012 IEEE international conference on Acoustics, speech and signal processing (ICASSP)*. IEEE, pp. 4277–4280 (cit. on pp. 26, 67).
- [Al-Qizwini et al., 2017] Al-Qizwini, Mohammed et al. (2017). “Deep learning algorithm for autonomous driving using googlenet”. In: *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, pp. 89–96 (cit. on pp. 51, 67).
- [Al-Rahayfeh and Faezipour, 2013] Al-Rahayfeh, AMER and Miad Faezipour (2013). “Eye tracking and head movement detection: A state-of-art survey”. In: *IEEE journal of translational engineering in health and medicine* 1, pp. 2100212–2100212 (cit. on pp. 11, 13, 67).
- [Alioua et al., 2016] Alioua, Nawal et al. (2016). “Driver head pose estimation using efficient descriptor fusion”. In: *EURASIP Journal on Image and Video Processing* 2016.1, p. 2 (cit. on pp. 12, 13, 67).

B

- [Badrinarayanan, Kendall, and Cipolla, 2017] Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla (2017). “Segnet: A deep convolutional encoder-decoder architecture for image segmentation”. In: *IEEE transactions on pattern analysis and machine intelligence* 39.12, pp. 2481–2495 (cit. on pp. 49, 53, 67).
- [Bhagyesh Vikani, 2017] Bhagyesh Vikani, Falak Shah (2017). *CNN Visualization*. https://github.com/InFoCusp/tf_cnnvis/. commit 2018-04-01 (cit. on pp. 39, 67).
- [Bhatia et al., 2019] Bhatia, Yukti et al. (2019). “Convolutional neural networks based potholes detection using thermal imaging”. In: *Journal of King Saud University-Computer and Information Sciences* (cit. on pp. 15, 67).
- [Bianchini, Bandini, and Smith, 2010] Bianchini, Alessandra, Paola Bandini, and David W Smith (2010). “Interrater reliability of manual pavement distress evaluations”. In: *Journal of Transportation Engineering* 136.2, pp. 165–172 (cit. on pp. 15, 67).
- [Borghini et al., 2014] Borghini, Gianluca et al. (2014). “Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness”. In: *Neuroscience & Biobehavioral Reviews* 44, pp. 58–75 (cit. on pp. 24, 67).
- [Broomhead and Lowe, 1988] Broomhead, DS and D Lowe (1988). “Multivariable functional interpolation and adaptive networks, complex systems, vol. 2”. In: (cit. on pp. 19, 67).
- [Burges, 1998] Burges, Christopher JC (1998). “A tutorial on support vector machines for pattern recognition”. In: *Data mining and knowledge discovery* 2.2, pp. 121–167 (cit. on pp. 12, 13, 67).

C

- [Cao et al., 2020] Cao, Minh-Tu et al. (2020). “Survey on performance of deep learning models for detecting road damages using multiple dashcam image resources”. In: *Advanced Engineering Informatics* 46, p. 101182 (cit. on pp. 16, 68).
- [Carion et al., 2020] Carion, Nicolas et al. (2020). “End-to-end object detection with transformers”. In: *European conference on computer vision*. Springer, pp. 213–229 (cit. on pp. 53, 68).
- [Chang, Chang, and Liu, 2005] Chang, KT, JR Chang, and JK Liu (2005). “Detection of pavement distresses using 3D laser scanning technology”. In: *Computing in civil engineering (2005)*, pp. 1–11 (cit. on pp. 15, 68).
- [Chaudhuri and Routray, 2019] Chaudhuri, Aritra and Aurobinda Routray (2019). “Driver Fatigue Detection Through Chaotic Entropy Analysis of Cortical Sources Obtained From Scalp EEG Signals”. In: *IEEE Transactions on Intelligent Transportation Systems* (cit. on pp. 13, 24, 68).
- [Chen et al., 2017] Chen, Liang-Chieh et al. (2017). “DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs”. In: *IEEE transactions on pattern analysis and machine intelligence* 40.4, pp. 834–848 (cit. on pp. 53, 68).
- [Chen et al., 2015] Chen, Xiaozhi et al. (2015). “3d object proposals for accurate object class detection”. In: *Advances in Neural Information Processing Systems*. Citeseer, pp. 424–432 (cit. on pp. 53, 68).
- [Chen et al., 2016] Chen, Xiaozhi et al. (2016). “Monocular 3d object detection for autonomous driving”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2147–2156 (cit. on pp. 53, 68).
- [Chéron, Laptev, and Schmid, 2015] Chéron, Guilhem, Ivan Laptev, and Cordelia Schmid (2015). “P-cnn: Pose-based cnn features for action recognition”. In: *Proceedings of the IEEE international conference on computer vision*, pp. 3218–3226 (cit. on pp. 26, 68).
- [Chiou et al., 2019] Chiou, Chien-Yu et al. (2019). “Driver monitoring using sparse representation with part-based temporal face descriptors”. In: *IEEE Transactions on Intelligent Transportation Systems* 21.1, pp. 346–361 (cit. on pp. 42, 68).
- [Chollet et al., 2015] Chollet, François et al. (2015). *Keras* (cit. on pp. 30, 68).
- [Coenen and Golroo, 2017] Coenen, Tom BJ and Amir Golroo (2017). “A review on automated pavement distress detection methods”. In: *Cogent Engineering* 4.1, p. 1374822 (cit. on pp. 15, 68).
- [Computer Vision Lab, 2016] Computer Vision Lab, National Tsing Hua University (2016). *Driver Drowsiness Detection Dataset*. <http://cv.cs.nthu.edu.tw/php/callforpaper/datasets/DDD/> (cit. on pp. 25, 68).
- [Cristianini, Shawe-Taylor, et al., 2000] Cristianini, Nello, John Shawe-Taylor, et al. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press (cit. on pp. 26, 40, 42, 68).

D

- [De Zoysa et al., 2007] De Zoysa, Kasun et al. (2007). “A public transport system based sensor network for road surface condition monitoring”. In: *Proceedings of the 2007 workshop on Networked systems for developing regions*, pp. 1–6 (cit. on pp. 14, 68).
- [Dewangan and Sahu, 2020] Dewangan, Deepak Kumar and Satya Prakash Sahu (2020). “Deep learning-based speed bump detection model for intelligent vehicle system using Raspberry Pi”. In: *IEEE sensors journal* 21.3, pp. 3570–3578 (cit. on pp. 16, 68).
- [Donahue et al., 2015] Donahue, Jeffrey et al. (2015). “Long-term recurrent convolutional networks for visual recognition and description”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2625–2634 (cit. on pp. 31, 40, 42, 68).

[Dong et al., 2019] Dong, Na et al. (2019). “A WPCA-Based Method for Detecting Fatigue Driving From EEG-Based Internet of Vehicles System”. In: *IEEE Access* 7, pp. 124702–124711 (cit. on pp. 13, 69).

[Dumoulin and Visin, 2016] Dumoulin, Vincent and Francesco Visin (2016). “A guide to convolution arithmetic for deep learning”. In: *arXiv preprint arXiv:1603.07285* (cit. on pp. 53, 69).

E

[Ebrahim et al., 2014] Ebrahim, Parisa et al. (2014). “Eyelid-based driver state classification under simulated and real driving conditions”. In: *Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on*. IEEE, pp. 3190–3196 (cit. on pp. 26, 69).

[El Ftouh et al., 2013] El Ftouh, M et al. (2013). “Étude de la prévalence et les facteurs de risque de la somnolence au volant dans une population marocaine”. In: *Médecine du Sommeil* 10.4, pp. 141–145 (cit. on pp. 23, 69).

[Elman, 1990] Elman, Jeffrey L (1990). “Finding structure in time”. In: *Cognitive science* 14.2, pp. 179–211 (cit. on pp. 26, 27, 69).

[Eriksson et al., 2008] Eriksson, Jakob et al. (2008). “The pothole patrol: using a mobile sensor network for road surface monitoring”. In: *Proceedings of the 6th international conference on Mobile systems, applications, and services*, pp. 29–39 (cit. on pp. 14, 16, 69).

F

[Fan et al., 2021] Fan, Rui et al. (2021). “Rethinking road surface 3-d reconstruction and pothole detection: From perspective transformation to disparity map segmentation”. In: *IEEE Transactions on Cybernetics* (cit. on pp. 16, 69).

[Fan et al., 2019] Fan, Rui et al. (2019). “Road crack detection using deep convolutional neural network and adaptive thresholding”. In: *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, pp. 474–479 (cit. on pp. 16, 69).

[François et al., 2016] François, Clémentine et al. (2016). “Tests of a new drowsiness characterization and monitoring system based on ocular parameters”. In: *International journal of environmental research and public health* 13.2, p. 174 (cit. on pp. 25, 69).

[Freund, Schapire, et al., 1996] Freund, Yoav, Robert E Schapire, et al. (1996). “Experiments with a new boosting algorithm”. In: *icml*. Vol. 96, pp. 148–156 (cit. on pp. 11, 13, 69).

G

[Gavrilescu et al., 2018] Gavrilescu, Raducu et al. (2018). “Faster R-CNN: an approach to real-time object detection”. In: *2018 International Conference and Exposition on Electrical And Power Engineering (EPE)*. IEEE, pp. 0165–0168 (cit. on pp. 52, 53, 69).

[Giusti et al., 2013] Giusti, Alessandro et al. (2013). “Fast image scanning with deep max-pooling convolutional neural networks”. In: *2013 IEEE International Conference on Image Processing*. IEEE, pp. 4034–4038 (cit. on pp. 35, 69).

[Goodfellow et al., 2016] Goodfellow, Ian et al. (2016). *Deep learning*. Vol. 1. MIT press Cambridge (cit. on pp. 12, 25, 29, 69).

[Gulli and Pal, 2017] Gulli, Antonio and Sujit Pal (2017). *Deep learning with Keras*. Packt Publishing Ltd (cit. on pp. 39, 69).

[Gunawan, Soewito, et al., 2015] Gunawan, Fergyanto E, Benfano Soewito, et al. (2015). “A vibratory-based method for road damage classification”. In: *2015 International Seminar on Intelligent Technology and Its Applications (ISITIA)*. IEEE, pp. 1–4 (cit. on pp. 16, 69).

H

- [Haggag, n.d.] Haggag, Omar (n.d.). "Automated Drowsiness Detection while Driving using Depth Camera". In: *USCCS 2019* (), p. 11 (cit. on pp. 14, 70).
- [Han et al., 2019] Han, Chunxiao et al. (2019). "Brain Complex Network Characteristic Analysis of Fatigue during Simulated Driving Based on Electroencephalogram Signals". In: *Entropy* 21.4, p. 353 (cit. on pp. 13, 70).
- [He et al., 2016] He, Kaiming et al. (2016). "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778 (cit. on pp. 51, 53, 70).
- [Hecht-Nielsen et al., 1988] Hecht-Nielsen, Robert et al. (1988). "Theory of the backpropagation neural network." In: *Neural Networks* 1.Supplement-1, pp. 445–448 (cit. on pp. 24, 70).
- [Hinton, Osindero, and Teh, 2006] Hinton, Geoffrey E, Simon Osindero, and Yee-Whye Teh (2006). "A fast learning algorithm for deep belief nets". In: *Neural computation* 18.7, pp. 1527–1554 (cit. on pp. 40, 42, 70).
- [Hosang, Benenson, and Schiele, 2014] Hosang, Jan, Rodrigo Benenson, and Bernt Schiele (2014). "How good are detection proposals, really?" In: *arXiv preprint arXiv:1406.6962* (cit. on pp. 52, 70).
- [Hosang et al., 2015] Hosang, Jan et al. (2015). "What makes for effective detection proposals?" In: *IEEE transactions on pattern analysis and machine intelligence* 38.4, pp. 814–830 (cit. on pp. 52, 70).
- [Hou, Wang, and Gong, 2007] Hou, Zhiqiong, Kelvin CP Wang, and Weiguo Gong (2007). "Experimentation of 3D pavement imaging through stereovision". In: *International Conference on Transportation Engineering 2007*, pp. 376–381 (cit. on pp. 15, 70).
- [Hu, Shen, and Sun, 2018] Hu, Jie, Li Shen, and Gang Sun (2018). "Squeeze-and-excitation networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141 (cit. on pp. 51, 70).
- [Huang et al., 2018] Huang, Yanping et al. (2018). "Gpipe: Efficient training of giant neural networks using pipeline parallelism". In: *arXiv preprint arXiv:1811.06965* (cit. on pp. 51, 70).
- [Huynh, Park, and Kim, 2016] Huynh, Xuan-Phung, Sang-Min Park, and Yong-Guk Kim (2016). "Detection of driver drowsiness using 3D deep neural network and semi-supervised gradient boosting machine". In: *Asian Conference on Computer Vision*. Springer, pp. 134–145 (cit. on pp. 26, 70).

J

- [Ji et al., 2013] Ji, Shuiwang et al. (2013). "3D convolutional neural networks for human action recognition". In: *IEEE transactions on pattern analysis and machine intelligence* 35.1, pp. 221–231 (cit. on pp. 33, 70).
- [Jiang et al., 2020] Jiang, Zicong et al. (2020). "Real-time object detection method based on improved YOLOv4-tiny". In: *arXiv preprint arXiv:2011.04244* (cit. on pp. 16, 70).
- [Jin, Dundar, and Culurciello, 2014] Jin, Jonghoon, Aysegul Dundar, and Eugenio Culurciello (2014). "Flattened convolutional neural networks for feedforward acceleration". In: *arXiv preprint arXiv:1412.5474* (cit. on pp. 35, 70).
- [Jog et al., 2012] Jog, GM et al. (2012). "Pothole properties measurement through visual 2D recognition and 3D reconstruction". In: *Computing in Civil Engineering (2012)*, pp. 553–560 (cit. on pp. 15, 70).
- [Jordan, 1997] Jordan, Michael I (1997). "Serial order: A parallel distributed processing approach". In: *Advances in psychology*. Vol. 121. Elsevier, pp. 471–495 (cit. on pp. 26, 27, 70).

K

- [[Karpathy et al., 2014](#)] Karpathy, Andrej et al. (2014). "Large-scale video classification with convolutional neural networks". In: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 1725–1732 (cit. on pp. [33](#), [71](#)).
- [[Kartsch et al., 2018](#)] Kartsch, Victor Javier et al. (2018). "A sensor fusion approach for drowsiness detection in wearable ultra-low-power systems". In: *Information Fusion* 43, pp. 66–76 (cit. on pp. [32](#), [44](#), [71](#)).
- [[Kim and Ryu, 2014](#)] Kim, Taehyeong and Seung-Ki Ryu (2014). "A guideline for pothole classification". In: *International Journal of Engineering and Technology* 4.10, pp. 618–622 (cit. on pp. [14](#), [71](#)).
- [[Kohonen, 1982](#)] Kohonen, Teuvo (1982). "Self-organized formation of topologically correct feature maps". In: *Biological cybernetics* 43.1, pp. 59–69 (cit. on pp. [19](#), [71](#)).
- [[Koiliias, Mousas, and Rekadbar, 2019](#)] Koiliias, Alexandros, Christos Mousas, and Banafsheh Rekadbar (2019). "The Effects of Driving Habits on Virtual Reality Car Passenger Anxiety". In: *International Conference on Virtual Reality and Augmented Reality*. Springer, pp. 263–281 (cit. on pp. [24](#), [71](#)).
- [[Krizhevsky, Sutskever, and Hinton, 2012](#)] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (2012). "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*, pp. 1097–1105 (cit. on pp. [26](#), [31](#), [48](#), [53](#), [71](#)).
- [[Kumar, 2018](#)] Kumar, PS Jagadeesh (2018). "Multilayer Perceptron Neural Network Based Immersive VR System for Cognitive Computer Gaming". In: *Progress in Advanced Computing and Intelligent Engineering*. Springer, pp. 91–102 (cit. on pp. [31](#), [40](#), [42](#), [71](#)).
- [[Kurylyak, Lamonaca, and Mirabelli, 2012](#)] Kurylyak, Yuriy, Francesco Lamonaca, and Giovanni Mirabelli (2012). "Detection of the eye blinks for human's fatigue monitoring". In: *Medical Measurements and Applications Proceedings (MeMeA), 2012 IEEE International Symposium on*. IEEE, pp. 1–4 (cit. on pp. [11](#), [13](#), [71](#)).

L

- [[Lampert, Blaschko, and Hofmann, 2008](#)] Lampert, Christoph H, Matthew B Blaschko, and Thomas Hofmann (2008). "Beyond sliding windows: Object localization by efficient subwindow search". In: *2008 IEEE conference on computer vision and pattern recognition*. IEEE, pp. 1–8 (cit. on pp. [52](#), [71](#)).
- [[LeCun et al., 1998](#)] LeCun, Yann et al. (1998). "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11, pp. 2278–2324 (cit. on pp. [19](#), [71](#)).
- [[Liang et al., 2017](#)] Liang, Yulan et al. (2017). "Prediction of drowsiness events in night shift workers during morning driving". In: *Accident Analysis & Prevention* (cit. on pp. [25](#), [26](#), [71](#)).
- [[Lienhart and Maydt, 2002](#)] Lienhart, Rainer and Jochen Maydt (2002). "An extended set of haar-like features for rapid object detection". In: *Image Processing. 2002. Proceedings. 2002 International Conference on*. Vol. 1. IEEE, pp. I–I (cit. on pp. [11](#), [13](#), [71](#)).
- [[Liu et al., 2019a](#)] Liu, Fan et al. (2019a). "A Review of Driver Fatigue Detection: Progress and Prospect". In: *2019 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, pp. 1–6 (cit. on pp. [14](#), [71](#)).
- [[Liu, 2018](#)] Liu, Taijie (2018). "Using Kinect to Capture the Joint Angles of Static Driving Posture". In: *Advances in Physical Ergonomics & Human Factors: Proceedings of the AHFE 2018 International Conference on Physical Ergonomics & Human Factors, July 21-25, 2018, Loews Sapphire Falls Resort at Universal Studios, Orlando, Florida, USA*. Vol. 789. Springer, p. 297 (cit. on pp. [14](#), [71](#)).
- [[Liu et al., 2016](#)] Liu, Wei et al. (2016). "Ssd: Single shot multibox detector". In: *European conference on computer vision*. Springer, pp. 21–37 (cit. on pp. [53](#), [71](#)).
- [[Liu et al., 2019b](#)] Liu, Weihuang et al. (2019b). "Convolutional Two-Stream Network Using Multi-Facial Feature Fusion for Driver Fatigue Detection". In: *Future Internet* 11.5, p. 115 (cit. on pp. [14](#), [71](#)).

[Long, Shelhamer, and Darrell, 2015] Long, Jonathan, Evan Shelhamer, and Trevor Darrell (2015). “Fully convolutional networks for semantic segmentation”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440 (cit. on pp. 53, 72).

M

[Ma, Chang, and Li, 2019] Ma, Zhanyu, Dongliang Chang, and Xiaoxu Li (2019). “Channel Max Pooling Layer for Fine-Grained Vehicle Classification”. In: *arXiv preprint arXiv:1902.11107* (cit. on pp. 35, 72).

[Madli et al., 2015] Madli, Rajeshwari et al. (2015). “Automatic detection and notification of potholes and humps on roads to aid drivers”. In: *IEEE sensors journal* 15.8, pp. 4313–4318 (cit. on pp. 16, 72).

[Maeda et al., 2018] Maeda, Hiroya et al. (2018). “Road damage detection and classification using deep neural networks with smartphone images”. In: *Computer-Aided Civil and Infrastructure Engineering* 33.12, pp. 1127–1141 (cit. on pp. 16, 72).

[Mandal et al., 2017] Mandal, Bappaditya et al. (2017). “Towards detection of bus driver fatigue based on robust visual analysis of eye state”. In: *IEEE Transactions on Intelligent Transportation Systems* 18.3, pp. 545–557 (cit. on pp. 11, 13, 72).

[Martins and Astudillo, 2016] Martins, Andre and Ramon Astudillo (2016). “From softmax to sparsemax: A sparse model of attention and multi-label classification”. In: *International Conference on Machine Learning*, pp. 1614–1623 (cit. on pp. 35, 72).

[Massoz, Verly, and Van Droogenbroeck, 2018] Massoz, Quentin, Jacques Verly, and Marc Van Droogenbroeck (2018). “Multi-Timescale Drowsiness Characterization Based on a Video of a Driver’s Face”. In: *Sensors* 18.9, p. 2801 (cit. on pp. 13, 26, 42, 72).

[Mbouna, Kong, and Chun, 2013] Mbouna, Ralph Oyini, Seong G Kong, and Myung-Geun Chun (2013). “Visual analysis of eye state and head pose for driver alertness monitoring”. In: *IEEE transactions on intelligent transportation systems* 14.3, pp. 1462–1469 (cit. on pp. 42, 72).

[McCulloch and Pitts, 1943] McCulloch, Warren S and Walter Pitts (1943). “A logical calculus of the ideas immanent in nervous activity”. In: *The bulletin of mathematical biophysics* 5.4, pp. 115–133 (cit. on pp. 18, 72).

[Mitchell et al., 1997] Mitchell, Tom M et al. (1997). *Machine learning*. WCB (cit. on pp. 63, 72).

[Mnih, 2013] Mnih, Volodymyr (2013). “Machine learning for aerial image labeling”. PhD thesis. University of Toronto (Canada) (cit. on pp. 26, 72).

[Mohamed AFECHKAR, n.d.] Mohamed AFECHKAR, Abdellah RAHMOUNI (n.d.). *Analysis of the Moroccan method relating to the of pavement degradations*. http://www.ampcr.ma/actes/9eme_congres_national_de_la_route/CONGRE/A4/A4_3.pdf. 9th national road congress (skhirat, june 11, 12 and 13), 2014 (cit. on pp. 15, 72).

[Mohan, Padmanabhan, and Ramjee, 2008] Mohan, Prashanth, Venkata N Padmanabhan, and Ramachandran Ramjee (2008). “Nericell: rich monitoring of road and traffic conditions using mobile smartphones”. In: *Proceedings of the 6th ACM conference on Embedded network sensor systems*, pp. 323–336 (cit. on pp. 16, 72).

[Mu, Hu, and Min, 2017] Mu, Zhendong, Jianfeng Hu, and Jianliang Min (2017). “Driver fatigue detection system using electroencephalography signals based on combined entropy features”. In: *Applied Sciences* 7.2, p. 150 (cit. on pp. 24, 72).

[Mustafa, Allen, and Appiah, 2019] Mustafa, Mohammed Kyari, Tony Allen, and Kofi Appiah (2019). “A comparative review of dynamic neural networks and hidden Markov model methods for mobile on-device speech recognition”. In: *Neural Computing and Applications* 31.2, pp. 891–899 (cit. on pp. 33, 72).

N

- [Nienaber, Booyesen, and Kroon, 2015] Nienaber, S, Marthinus J Booyesen, and RS Kroon (2015). “Detecting potholes using simple image processing techniques and real-world footage”. In: (cit. on pp. 50, 56, 73).
- [Nienaber, Kroon, and Booyesen, 2015] Nienaber, S, RS Kroon, and Marthinus J Booyesen (2015). “A comparison of low-cost monocular vision techniques for pothole distance estimation”. In: *2015 IEEE Symposium Series on Computational Intelligence*. IEEE, pp. 419–426 (cit. on pp. 50, 56, 73).
- [Noh, Hong, and Han, 2015] Noh, Hyeonwoo, Seunghoon Hong, and Bohyung Han (2015). “Learning deconvolution network for semantic segmentation”. In: *Proceedings of the IEEE international conference on computer vision*, pp. 1520–1528 (cit. on pp. 53, 73).

O

- [Omidyeganeh et al., 2016] Omidyeganeh, Mona et al. (2016). “Yawning detection using embedded smart cameras”. In: *IEEE Transactions on Instrumentation and Measurement* 65.3, pp. 570–582 (cit. on pp. 42, 73).
- [Organization, 2015] Organization, World Health (2015). *Global status report on road safety 2015*. World Health Organization (cit. on pp. 23, 73).

P

- [Papandreou et al., n.d.] Papandreou, G et al. (n.d.). “Weakly-and semi-supervised learning of a DCNN for semantic image segmentation. arXiv 2015”. In: *arXiv preprint arXiv:1502.02734* () (cit. on pp. 53, 73).
- [Pereira et al., 2019] Pereira, Vosco et al. (2019). “Semantic segmentation of paved road and pothole image using U-net architecture”. In: *2019 International Conference of Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*. IEEE, pp. 1–4 (cit. on pp. 16, 73).
- [Powers, 2011] Powers, David Martin (2011). “Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation”. In: (cit. on pp. 41, 73).

R

- [Redmon and Farhadi, 2018] Redmon, Joseph and Ali Farhadi (2018). “Yolov3: An incremental improvement”. In: *arXiv preprint arXiv:1804.02767* (cit. on pp. 51, 73).
- [Rekabdar and Mousas, 2018] Rekabdar, Banafsheh and Christos Mousas (2018). “Dilated Convolutional Neural Network for Predicting Driver’s Activity”. In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, pp. 3245–3250 (cit. on pp. 26, 73).
- [Rekabdar, Mousas, and Gupta, 2019] Rekabdar, Banafsheh, Christos Mousas, and Bidyut Gupta (2019). “Generative adversarial network with policy gradient for text summarization”. In: *2019 IEEE 13th international conference on semantic computing (ICSC)*. IEEE, pp. 204–207 (cit. on pp. 26, 73).
- [Ren et al., 2015] Ren, Shaoqing et al. (2015). “Faster r-cnn: Towards real-time object detection with region proposal networks”. In: *Advances in neural information processing systems* 28 (cit. on pp. 57, 73).
- [Rev., 2018] Rev., MIT Technol. (2018). *Dueling Neural Networks*. <https://www.technologyreview.com/lists/technologies/2018/>. [Online; accessed 03-Mars-2018] (cit. on pp. 12, 14, 73).
- [Reverdy and Leonard, 2016] Reverdy, Paul and Naomi Ehrich Leonard (2016). “Parameter estimation in softmax decision-making models with linear objective functions”. In: *IEEE Transactions on Automation Science and Engineering* 13.1, pp. 54–67 (cit. on pp. 35, 73).

- [Rezatofighi et al., 2019] Rezatofighi, Hamid et al. (2019). “Generalized intersection over union: A metric and a loss for bounding box regression”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 658–666 (cit. on pp. 55, 74).
- [Ronneberger, Fischer, and Brox, 2015] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015). “U-net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, pp. 234–241 (cit. on pp. 51, 53, 74).
- [Rosenblatt, 1957] Rosenblatt, Frank (1957). *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory (cit. on pp. 19, 74).

S

- [Saito et al., 2017] Saito, Shunsuke et al. (2017). “Photorealistic facial texture inference using deep neural networks”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5144–5153 (cit. on pp. 26, 74).
- [Sarmiento, 2021] Sarmiento, James-Andrew (2021). “Pavement Distress Detection and Segmentation using YOLOv4 and DeepLabv3 on Pavements in the Philippines”. In: *arXiv preprint arXiv:2103.06467* (cit. on pp. 16, 74).
- [Schapire and Freund, 2013] Schapire, Robert E and Yoav Freund (2013). “Boosting: Foundations and Algorithms”. In: *Kybernetes* 42.1, pp. 164–166 (cit. on pp. 64, 74).
- [Schmidhuber, 2015] Schmidhuber, Jürgen (2015). “Deep learning in neural networks: An overview”. In: *Neural networks* 61, pp. 85–117 (cit. on pp. 12, 25, 74).
- [Schwing and Urtasun, 2015] Schwing, Alexander G and Raquel Urtasun (2015). “Fully connected deep structured networks”. In: *arXiv preprint arXiv:1503.02351* (cit. on pp. 53, 74).
- [Sela, Richardson, and Kimmel, 2017] Sela, Matan, Elad Richardson, and Ron Kimmel (2017). “Unrestricted facial geometry reconstruction using image-to-image translation”. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1576–1585 (cit. on pp. 26, 74).
- [Sermanet et al., 2013a] Sermanet, Pierre et al. (2013a). “Overfeat: Integrated recognition, localization and detection using convolutional networks”. In: *arXiv preprint arXiv:1312.6229* (cit. on pp. 53, 74).
- [Sermanet et al., 2013b] Sermanet, Pierre et al. (2013b). “Pedestrian detection with unsupervised multi-stage feature learning”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3626–3633 (cit. on pp. 26, 27, 74).
- [Shaghouri, Alkhatib, and Berjaoui, 2021] Shaghouri, Anas Al, Rami Alkhatib, and Samir Berjaoui (2021). “Real-Time Pothole Detection Using Deep Learning”. In: *arXiv preprint arXiv:2107.06356* (cit. on pp. 14, 74).
- [Shih and Hsu, 2016] Shih, Tun-Huai and Chiou-Ting Hsu (2016). “MSTN: multistage spatial-temporal network for driver drowsiness detection”. In: *Asian Conference on Computer Vision*. Springer, pp. 146–153 (cit. on pp. 25, 26, 74).
- [Simonyan and Zisserman, 2014] Simonyan, Karen and Andrew Zisserman (2014). “Very deep convolutional networks for large-scale image recognition”. In: *arXiv preprint arXiv:1409.1556* (cit. on pp. 25, 34, 51, 74).
- [Solomon and Wang, 2015] Solomon, Cleshain and Zenghui Wang (2015). “Driver Attention and Behavior Detection with Kinect”. In: *J. Image Graph* 3.2, pp. 84–89 (cit. on pp. 12, 74).
- [Srivastava, Mansimov, and Salakhudinov, 2015] Srivastava, Nitish, Elman Mansimov, and Ruslan Salakhudinov (2015). “Unsupervised learning of video representations using lstms”. In: *International Conference on Machine Learning*, pp. 843–852 (cit. on pp. 29, 40, 42, 74).

- [Summala, 2007] Summala, Heikki (2007). “Towards understanding motivational and emotional factors in driver behaviour: Comfort through satisficing”. In: *Modelling driver behaviour in automotive environments*. Springer, pp. 189–207 (cit. on pp. 24, 75).
- [Sun et al., 2015] Sun, Lin et al. (2015). “Human action recognition using factorized spatio-temporal convolutional networks”. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4597–4605 (cit. on pp. 34, 75).
- [Sutskever, Vinyals, and Le, 2014] Sutskever, Ilya, Oriol Vinyals, and Quoc V Le (2014). “Sequence to sequence learning with neural networks”. In: *Advances in neural information processing systems*, pp. 3104–3112 (cit. on pp. 28, 75).
- [Szegedy et al., 2014] Szegedy, Christian et al. (2014). “Scalable, high-quality object detection”. In: *arXiv preprint arXiv:1412.1441* (cit. on pp. 53, 75).

T

- [Tan and Le, 2019] Tan, Mingxing and Quoc Le (2019). “Efficientnet: Rethinking model scaling for convolutional neural networks”. In: *International Conference on Machine Learning*. PMLR, pp. 6105–6114 (cit. on pp. 51, 75).
- [Teichmann et al., 2018] Teichmann, Marvin et al. (2018). “Multinet: Real-time joint semantic reasoning for autonomous driving”. In: *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, pp. 1013–1020 (cit. on pp. 52, 57, 75).
- [Tran et al., 2015] Tran, Du et al. (2015). “Learning spatiotemporal features with 3d convolutional networks”. In: *Proceedings of the IEEE international conference on computer vision*, pp. 4489–4497 (cit. on pp. 33, 34, 75).

W

- [Wang, 2004] Wang, Kelvin CP (2004). “Challenges and feasibility for comprehensive automated survey of pavement conditions”. In: *Applications of advanced technologies in transportation engineering (2004)*, pp. 531–536 (cit. on pp. 15, 75).
- [Wang et al., 2015] Wang, Naiyan et al. (2015). “Transferring rich feature hierarchies for robust visual tracking”. In: *arXiv preprint arXiv:1501.04587* (cit. on pp. 52, 75).
- [Wang and Xu, 2016] Wang, Xuesong and Chuan Xu (2016). “Driver drowsiness detection based on non-intrusive metrics considering individual specifics”. In: *Accident Analysis & Prevention* 95, pp. 350–357 (cit. on pp. 25, 26, 75).
- [Wang et al., 2019] Wang, Yanjing et al. (2019). “Wavelet Packet Entropy Analysis of Resting State Electroencephalogram in Sleep Deprived Mental Fatigue State”. In: *International Conference on Human-Computer Interaction*. Springer, pp. 484–494 (cit. on pp. 13, 75).
- [Weng, Lai, and Lai, 2016] Weng, Ching-Hua, Ying-Hsiu Lai, and Shang-Hong Lai (2016). “Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network”. In: *Asian Conference on Computer Vision*. Springer, pp. 117–133 (cit. on pp. 25, 26, 40, 42, 75).
- [Wu, Shen, and Van Den Hengel, 2019] Wu, Zifeng, Chunhua Shen, and Anton Van Den Hengel (2019). “Wider or deeper: Revisiting the resnet model for visual recognition”. In: *Pattern Recognition* 90, pp. 119–133 (cit. on pp. 51, 75).

X

[Xiao et al., 2019] Xiao, Zhitao et al. (2019). “Fatigue driving recognition network: fatigue driving recognition via convolutional neural network and long short-term memory units”. In: *IET Intelligent Transport Systems* (cit. on pp. 14, 76).

Y

[Yang et al., 2018] Yang, Maoke et al. (2018). “Denseaspp for semantic segmentation in street scenes”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3684–3692 (cit. on pp. 51, 76).

[Yang et al., 2019] Yang, Yuxuan et al. (2019). “A Complex Network-Based Broad Learning System for Detecting Driver Fatigue From EEG Signals”. In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* (cit. on pp. 13, 76).

[Ye et al., 2021] Ye, Wanli et al. (2021). “Convolutional neural network for pothole detection in asphalt pavement”. In: *Road materials and pavement design* 22.1, pp. 42–58 (cit. on pp. 15, 76).

[Yu and Yu, 2006] Yu, Bill X and Xinbao Yu (2006). “Vibration-based system for pavement condition evaluation”. In: *Applications of Advanced Technology in Transportation*, pp. 183–189 (cit. on pp. 15, 76).

Z

[Zeiler and Fergus, 2014] Zeiler, Matthew D and Rob Fergus (2014). “Visualizing and understanding convolutional networks”. In: *European conference on computer vision*. Springer, pp. 818–833 (cit. on pp. 39, 76).

[Zeiler et al., 2010] Zeiler, Matthew D et al. (2010). “Deconvolutional networks”. In: *2010 IEEE Computer Society Conference on computer vision and pattern recognition*. IEEE, pp. 2528–2535 (cit. on pp. 53, 76).

[Zhang et al., 2017] Zhang, Xiaoliang et al. (2017). “Design of a fatigue detection system for high-speed trains based on driver vigilance using a wireless wearable EEG”. In: *Sensors* 17.3, p. 486 (cit. on pp. 24, 76).

[Zhang et al., 2014] Zhang, Zhanpeng et al. (2014). “Facial landmark detection by deep multi-task learning”. In: *European conference on computer vision*. Springer, pp. 94–108 (cit. on pp. 53, 76).

[Zhang, 2012] Zhang, Zhengyou (2012). “Microsoft kinect sensor and its effect”. In: *IEEE multimedia* 19.2, pp. 4–10 (cit. on pp. 12, 76).

[Zheng et al., 2015] Zheng, Shuai et al. (2015). “Conditional random fields as recurrent neural networks”. In: *Proceedings of the IEEE international conference on computer vision*, pp. 1529–1537 (cit. on pp. 53, 76).

[Zhou et al., 2018] Zhou, Yi et al. (2018). “Auto-conditioned recurrent networks for extended complex human motion synthesis”. In: (cit. on pp. 26, 76).