



# Reconnaissance de l'orientation des piétons par les réseaux de Capsules dans un environnement non structuré pour un ADAS

Safaâ Dafrallah

## ► To cite this version:

Safaâ Dafrallah. Reconnaissance de l'orientation des piétons par les réseaux de Capsules dans un environnement non structuré pour un ADAS. Réseau de neurones [cs.NE]. Normandie Université; Université Ibn Tofail. Faculté des sciences de Kénitra, 2021. Français. NNT : 2021NORMIR29 . tel-03630693

**HAL Id: tel-03630693**

**<https://theses.hal.science/tel-03630693>**

Submitted on 5 Apr 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## THESE

Pour obtenir le diplôme de doctorat

Spécialité Informatique

Préparée au sein de « INSA Rouen Normandie »

ET

« ENSA de Kenitra »

### Reconnaissance de l'orientation des piétons par les réseaux de Capsules dans un environnement non structuré pour un ADAS

Présentée et soutenue par

Safaâ DAFRALLAH

| Thèse soutenue publiquement le (25/11/2021)<br>devant le jury composé de |                                                                           |                        |
|--------------------------------------------------------------------------|---------------------------------------------------------------------------|------------------------|
| M. Mounir Ait Kerroum                                                    | Professeur des Universités, ENCG Kenitra,<br>Université Ibn Tofaïl, Maroc | Président / Rapporteur |
| Mme. Luce MORIN                                                          | Professeur des Universités, INSA Rennes,<br>France                        | Rapporteuse            |
| Mme. Amina RADGUI                                                        | Professeur des Universités, INPT Rabat,<br>Maroc                          | Rapporteuse            |
| M. Fabrice MERIAUDEAU                                                    | Professeur des Universités, Université de<br>Bourgogne, France            | Examineur              |
| Mme. Samia BOUCHAFA                                                      | Professeur des Universités, Université<br>d'Evry, Paris Saclay, France    | Examinatrice           |
| Mme. Aouatif AMINE                                                       | Professeur des Universités, ENSA Kenitra,<br>Université Ibn Tofaïl, Maroc | Directrice de thèse    |
| M. Abdelaziz BENSRAHRA                                                   | Professeur des Universités, INSA Rouen<br>Normandie, France               | Directeur de thèse     |
| M. Stéphane MOUSSET                                                      | Maître de Conférences, Université Rouen<br>Normandie, France              | Co-encadrant           |

Thèse dirigée par Aouatif AMINE (LSI), Abdelaziz BENSRAHRA (LITIS)

*À mes parents.*

Thèse en cotutelle réalisée dans le cadre du projet "SafeRoad : Multi-plateforme pour la Sécurité Routière (MSR)" sous le contrat N° : 24/2017, financé par le Ministère de l'Équipement, du Transport, de la Logistique et de l'Eau (METLE), par le Centre National pour la Recherche Scientifique et Technique (CNRST), ainsi que par le Laboratoire d'Informatique, de Traitement de l'Information et des Systèmes (LITIS) de l'INSA Rouen Normandie.



# Remerciements

Ce travail de recherche s'est déroulé dans le cadre d'une thèse en cotutelle entre l'ENSA de Kénitra à l'université Ibn Tofail au Maroc et l'INSA Rouen Normandie en France.

Je tiens tout d'abord à exprimer ma gratitude à Dieu le tout puissant et miséricordieux pour sa guidance et sa bienveillance et de m'avoir donné la force et la patience d'accomplir cette thèse.

Je tiens à remercier chaleureusement ma directrice de thèse, Mme Aouatif AMINE, de m'avoir accordé l'honneur d'entamer ce projet de recherche, pour son encouragement, son accompagnement et son support moral tout au long de cette thèse. La porte de son bureau était toujours ouverte à chaque fois que j'en avais besoin, ses conseils scientifiques et humains me seront toujours utiles. Qu'elle soit remercié également pour cette opportunité de thèse en cotutelle entre l'ENSA de Kenitra et l'INSA Rouen Normandie, me donnant ainsi la chance de côtoyer deux environnements différents. Cette magnifique expérience m'a permis de forger mes compétences aussi bien sur le plan scientifique que sur le plan personnel et humain.

Mes sincères remerciements vont également à mon second directeur de thèse, M. Abdelaziz BENSRAHAIR, qui m'a accueilli au sein de son équipe de recherche STI au laboratoire LITIS. Je le remercie pour la confiance qu'il m'a accordée et d'avoir facilité mon intégration au sein du LITIS, et pour tous les efforts qu'il a fournis pour que mes séjours se déroulent dans de meilleures conditions. Je tiens à le remercier également, pour ses conseils, son soutien tout au long de ce travail, et pour tout ce temps qu'il m'a consacré malgré son emploi très chargé.

J'exprime mes vifs remerciements à mon encadrant de thèse, M. Stéphane MOUSSET, pour son aide précieuse, ses nombreux encouragements et son assistance. Je lui exprime ma profonde reconnaissance pour le temps et les efforts consacrés pour réviser mon travail tout au long de cette thèse. Il a toujours veillé d'entretenir nos réunions hebdomadaires, me permettant ainsi d'améliorer mon travail.

Un spécial remerciement pour mes directeurs de thèses et mon encadrant pour leur

disponibilité et leur guidance durant cette période de pandémie. Leur suivi permanent pendant cette période et nos réunions hebdomadaires ont permis de mener ce travail à son terme. Je les remercie également pour le temps qu'ils ont consacré à corriger ce manuscrit. Qu'ils trouvent ici l'expression de ma profonde reconnaissance.

J'adresse mes remerciements les plus sincères à Mme Luce MORIN, Mme Amina RADGUI et M. Mounir AIT KERROUM pour l'honneur qu'ils m'ont fait d'être rapporteurs de ce mémoire, je les remercie pour leur temps et leur disponibilité d'évaluer ce travail. J'adresse également mes sincères remerciement à Mme Samia BOUCHAFA et M. Fabrice MERIAUDEAU pour le temps qu'ils m'ont accordé en acceptant d'examiner ma thèse. Je remercie également M. Mounir AIT KERROUM d'avoir accepté de présider le jury de la soutenance.

Je remercie mon comité de thèse, M. Fawzi NASHASHIBI et M. Paul Honeine pour le temps qu'ils m'ont consacré et leurs retours intéressants.

Je remercie Mme. Samia AINOUS, directrice de l'équipe STI, pour les nombreux séminaires et réunions qui nous ont permis de discuter et d'échanger avec les différents chercheurs et permanents de l'équipe.

Mes remerciements vont également à Mme. Brigitte DIARRA et Mme. Sandra HAGUE d'avoir faciliter mes séjours et mes tâches administratives allant de l'hébergement jusqu'au financement. Merci pour votre gentillesse et votre accueil sympathique.

Cette thèse fait partie d'un projet financé par le Ministère de l'Équipement, du Transport, de la Logistique et de l'Eau et par le Centre National de la Recherche Scientifique et Technique (CNRST) au Maroc ainsi que par l'INSA de Rouen pendant mes séjours en France. Toute ma gratitude pour l'équipe de collaboration du Ministère de Transport, l'équipe du CNRST ainsi que l'équipe de l'Agence Nationale de la Sécurité Routière (NARSA).

Je voudrais également exprimer ma gratitude à tous mes collègues trop nombreux pour être nommés. Une mention spéciale pour Fatima Zahra, Fadoua, Mounia, Fatna, Ibtissame, Zakaria, Mariam, Youssef, Bouchra et Mostapha. Une mention spéciale pour mes collègues du labo LITIS avec qui j'ai passé de merveilleux moments et qui

ont laissé leurs traces lors de l'élaboration de cette thèse, particulièrement Djamila, Usman, Imen Beji, Danut, Imen Ben Saleh, Lounis, Gabriel, Soufiane, Qiushi, Franco, Amine et Rachel.

Permettez-moi de dédier cette thèse à mes très chers parents, que je ne saurai remercier assez. C'est grâce à leur soutien sans faille et leur amour inconditionnel que j'ai pu surmonter les obstacles affrontés.

Merci à toi Maman, pour toutes les bonnes manières que tu m'a apprises, pour tes conseils et tes encouragements permanents pour mener cette thèse jusqu'au bout. Merci pour ton aide et ton soutien à chaque fois que j'en avais besoin. Merci d'être une Maman si formidable, unique et très bienveillante.

Merci à toi Papa, pour ta présence, malgré tes préoccupations, dans chaque étape de ma vie, merci pour ton soutien financier et moral, ta générosité, ton suivi régulier et tes encouragements constants tout au long de mon parcours. Tu représentes pour moi le symbole de la persévérance et du travail acharné.

Merci à toute ma famille pour son amour et son encouragement. Enfin merci à tous ceux et celles qui m'ont apporté un jour support et soutien, et que j'ai oublié de mentionner dans ces lignes.

## Résumé

Les accidents de routes représentent la première cause de décès chez les jeunes de moins de 30 ans. Etant l'utilisateur de route le plus vulnérable, le piéton représente 23% de l'ensemble des victimes des accidents mortels au monde. Dans ce contexte, cette thèse s'inscrit dans le cadre des recherches menées sur l'application des méthodes de l'apprentissage profond pour la sécurité des piétons.

Dans le cadre de ce travail, nous proposons un système de détection des orientations des piétons qui pourrait être intégré dans les systèmes avancés d'aide à la conduite (ADAS) afin d'alerter le conducteur de la présence d'un piéton. A cette fin, nous avons créé une nouvelle base de données d'orientation de piéton qu'on a appelé « SAFEROAD Dataset » enregistrée depuis différentes villes Marocaines en utilisant une caméra monoculaire au sein d'un véhicule en mouvement. Cette base contient 8894 images de piétons qui sont manuellement annotées en 4 et 8 directions.

Nous avons ensuite proposé une nouvelle technique de détection de l'orientation du piéton en utilisant les réseaux de Capsules. L'apprentissage et l'évaluation de cette technique sont faits sur notre base SafeRoad, Daimler et sur la base TUD. L'algorithme entraîné sur la base SafeRoad est par la suite appliqué pour la reconnaissance de l'orientation des piétons sur des séquences vidéo prises de la base JAAD.

Et finalement, nous proposons dans cette thèse un système de prévention des accidents piéton-véhicule, qui intègre l'orientation du piéton pour l'évaluation du risque d'un accident. Ce système prend en considération la présence des piétons indisciplinés et des routes non structurées.

**Mots clés :** Réseaux de neurones ; Apprentissage profond ; Réseaux de capsules ; Orientation du piéton ; Système de transport intelligent ; ADAS ; Routes non structurées ; SafeRoad dataset.

## Abstract

Road accidents are the first cause of death for those who are under 30 years old. Represented as the most vulnerable road user, the pedestrian constitutes 23% of all road fatalities. This thesis is part of the research conducted on the application of deep learning methods for pedestrian safety.

In this work, we propose a pedestrian orientation detection system that could be integrated into Advanced Driver Assistance Systems (ADAS) to alert the driver of the presence of a pedestrian. To this end, we created a new pedestrian orientation database called "SAFEROAD Dataset" recorded from different Moroccan cities using a monocular camera in a moving vehicle. This database contains 8894 images of pedestrians that are manually annotated in 4 and 8 directions.

We then proposed a new technique to detect pedestrian orientation using Capsule networks. The training and evaluation of this technique is done on our SafeRoad, Daimler and TUD base. The algorithm trained on the SafeRoad database is then applied for pedestrian orientation recognition on video sequences taken from the JAAD dataset.

Finally, we propose in this thesis a pedestrian-vehicle accident prevention system, which integrates the pedestrian's orientation for the evaluation of the risk of an accident. This system takes into consideration the presence of undisciplined pedestrians and unstructured roads.

**Keywords :** Neural networks; Deep learning; Capsule networks; Pedestrian orientation; Advanced driver assistance system; Unstructured roads; Moroccan dataset.



# Table des matières

|                                                                                 |           |
|---------------------------------------------------------------------------------|-----------|
| Table des figures                                                               | xvi       |
| Liste des tableaux                                                              | xviii     |
| Acronymes                                                                       | xviii     |
| <b>Introduction</b>                                                             | <b>1</b>  |
| <b>1 Etat de l’art</b>                                                          | <b>7</b>  |
| 1.1 Introduction . . . . .                                                      | 7         |
| 1.2 Systèmes Avancés d’Aide à la Conduite . . . . .                             | 9         |
| 1.2.1 Evolution des ADAS . . . . .                                              | 11        |
| 1.3 Méthodes de détection de piétons . . . . .                                  | 14        |
| 1.3.1 Classification par descripteurs . . . . .                                 | 14        |
| 1.3.2 Classifieurs . . . . .                                                    | 18        |
| 1.3.3 Approches classiques de détections de piétons . . . . .                   | 21        |
| 1.3.4 Méthodes basées sur l’apprentissage automatique . . . . .                 | 23        |
| 1.4 Conclusion . . . . .                                                        | 31        |
| <b>2 Proposition d’un système de prévention des accidents Piétons–Véhicules</b> | <b>33</b> |
| 2.1 Introduction . . . . .                                                      | 33        |
| 2.2 Conflit Piéton-Conducteur . . . . .                                         | 37        |
| 2.2.1 Comportement du conducteur . . . . .                                      | 37        |
| 2.2.2 Comportement du piéton . . . . .                                          | 40        |

|          |                                                                                           |            |
|----------|-------------------------------------------------------------------------------------------|------------|
| 2.2.3    | Interaction piéton-conducteur . . . . .                                                   | 42         |
| 2.3      | Système de prévention de collision avec piétons . . . . .                                 | 47         |
| 2.4      | Proposition d'un système de prévention des accidents avec piétons . .                     | 54         |
| 2.4.1    | Temps d'arrêt du véhicule . . . . .                                                       | 59         |
| 2.4.2    | Estimation du risque de collision . . . . .                                               | 62         |
| 2.5      | Conclusion . . . . .                                                                      | 66         |
| <b>3</b> | <b>Détection de l'Orienta-tion du Piéton</b>                                              | <b>69</b>  |
| 3.1      | Introduction . . . . .                                                                    | 69         |
| 3.2      | Etude de l'Existant . . . . .                                                             | 71         |
| 3.2.1    | Méthodes Classiques . . . . .                                                             | 71         |
| 3.2.2    | Méthodes utilisant l'apprentissage automatique . . . . .                                  | 73         |
| 3.3      | Les réseaux de Capsule . . . . .                                                          | 77         |
| 3.4      | Bases de données . . . . .                                                                | 80         |
| 3.4.1    | Base SafeRoad . . . . .                                                                   | 80         |
| 3.4.2    | Base Daimler . . . . .                                                                    | 84         |
| 3.4.3    | Base TUD Multiview Pedestrians . . . . .                                                  | 86         |
| 3.5      | Reconnaissance de l'Orienta-tion du piéton (4 Orientations) . . . . .                     | 87         |
| 3.5.1    | L'architecture proposée . . . . .                                                         | 87         |
| 3.5.2    | Expérimentations et Résultats . . . . .                                                   | 89         |
| 3.5.3    | Comparaison avec différents algorithmes d'optimisation pour la<br>base SafeRoad . . . . . | 96         |
| 3.6      | Reconnaissance de l'orientation du piéton (8 orientations) . . . . .                      | 99         |
| 3.6.1    | Architecture . . . . .                                                                    | 99         |
| 3.6.2    | Evaluation sur la base SafeRoad . . . . .                                                 | 99         |
| 3.6.3    | Evaluation sur la base TUD . . . . .                                                      | 100        |
| 3.7      | Conclusion . . . . .                                                                      | 105        |
| <b>4</b> | <b>Reconnaissance de l'orientation du piéton dans une séquence vidéo</b>                  | <b>107</b> |
| 4.1      | Introduction . . . . .                                                                    | 107        |
| 4.2      | Suivi de piétons . . . . .                                                                | 109        |



|                            |                                                        |            |
|----------------------------|--------------------------------------------------------|------------|
| 4.2.1                      | Suivi d'un seul objet . . . . .                        | 109        |
| 4.2.2                      | Suivi Multi-objets . . . . .                           | 109        |
| 4.2.3                      | Méthodes traditionnelles . . . . .                     | 110        |
| 4.2.4                      | Méthode du flux optique . . . . .                      | 111        |
| 4.2.5                      | Approches basées sur l'apprentissage profond . . . . . | 111        |
| 4.3                        | La base JAAD . . . . .                                 | 116        |
| 4.4                        | Expérimentations et Résultats . . . . .                | 117        |
| 4.5                        | Conclusion . . . . .                                   | 127        |
| <b>Conclusion Générale</b> |                                                        | <b>135</b> |



# Table des figures

|      |                                                                                                                                                                                               |    |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1-1  | Evolution des ADAS depuis 1980 [14] . . . . .                                                                                                                                                 | 12 |
| 1-2  | (a) Description d'un visage avec les ondelettes de Haar. (b) Exemples<br>des ondelettes de Haar [103] . . . . .                                                                               | 16 |
| 1-3  | Etapes de calcul d'un HOG : a) Image initiale, b) norme du gradient.<br>c) orientation du gradient. d) division de l'image en plusieurs cellules.<br>e) calcul du histogramme. [17] . . . . . | 17 |
| 1-4  | Le choix de l'hyperplan optimal [96] . . . . .                                                                                                                                                | 20 |
| 1-5  | L'influence des vecteurs de support sur la marge de l'hyperplan [96] .                                                                                                                        | 21 |
| 1-6  | Représentation des étapes du détecteur ACF [36] . . . . .                                                                                                                                     | 23 |
| 1-7  | Architecture du réseau de neurone convolutif basé sur les régions [54]                                                                                                                        | 24 |
| 1-8  | Régulation des cadres englobants [53] . . . . .                                                                                                                                               | 25 |
| 1-9  | Architecture du Fast-RCNN [52] . . . . .                                                                                                                                                      | 26 |
| 1-10 | Comparaison entre les algorithmes de détection en terme de vitesse [43]                                                                                                                       | 27 |
| 1-11 | Architecture de l'approche Faster RCNN [114] . . . . .                                                                                                                                        | 28 |
| 1-12 | Vitesse de Faster-RCNN par rapport à ces prédécesseurs dans la phase<br>de test [43] . . . . .                                                                                                | 28 |
| 1-13 | Les étapes du système de détection YOLO [113] . . . . .                                                                                                                                       | 29 |
| 1-14 | Le modèle YOLO [113] . . . . .                                                                                                                                                                | 30 |
| 1-15 | L'architecture de YOLO [113] . . . . .                                                                                                                                                        | 30 |
| 1-16 | L'architecture de YOLOv3 [42] . . . . .                                                                                                                                                       | 31 |

|      |                                                                                                                                                                          |    |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2-1  | Vue aérienne montrant le lieu de l'accident, le chemin du piéton et du véhicule ainsi que la vitesse du véhicule à partir de 5.6 seconde avant l'accident [97] . . . . . | 35 |
| 2-2  | Personnes tuées par classe d'âge en 2018 en France [98] . . . . .                                                                                                        | 38 |
| 2-3  | Personnes blessées par classe d'âge en 2018 en France [98] . . . . .                                                                                                     | 38 |
| 2-4  | Conducteurs tués par classe d'âge et par sexe en France [98] . . . . .                                                                                                   | 39 |
| 2-5  | Histogramme des piétons blessés et tués au Maroc par genre et par âge [90] . . . . .                                                                                     | 42 |
| 2-6  | Vitesse moyenne des conducteurs par rapport au passage piéton avec la présence d'un ADAS [12] . . . . .                                                                  | 46 |
| 2-7  | Vitesse moyenne des conducteurs par rapport au passage piéton sans la présence d'un ADAS [12] . . . . .                                                                  | 46 |
| 2-8  | Répartition des décès par type d'usager de la route, par région (OMS) [140] . . . . .                                                                                    | 48 |
| 2-9  | Taux de décès sur le trafic routier en 2013 et 2016 par population de 100.000 pour chaque région [140] . . . . .                                                         | 49 |
| 2-10 | Interface de Mobileye permettant d'alerter le conducteur en cas d'un éventuel accident [33] . . . . .                                                                    | 49 |
| 2-11 | Chronologie des mesures de protection des piétons. Le temps de réaction d'une certaine distance diminue avec l'augmentation de la vitesse [50] . . . . .                 | 50 |
| 2-12 | Diagramme de flux de données montrant la distillation des informations à partir des données brutes jusqu'à l'action appropriée [50] . . . . .                            | 51 |
| 2-13 | Les quatre zones de danger entre le véhicule et le piéton [134] . . . . .                                                                                                | 52 |
| 2-14 | Exemple de route structurée dans la capitale Rabat . . . . .                                                                                                             | 55 |
| 2-15 | Exemple de piéton traversant de façon non appropriée dans une zone structurée . . . . .                                                                                  | 55 |
| 2-16 | Exemple de piéton indiscipliné dans une zone mal structurée . . . . .                                                                                                    | 56 |
| 2-17 | Des piétons changeant leurs directions dans un temps court . . . . .                                                                                                     | 57 |
| 2-18 | Les quatre zones de risques entre le véhicule et le piéton . . . . .                                                                                                     | 58 |

|      |                                                                                                                                                                           |    |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2-19 | Architecture du système de prévention des accidents proposé . . . . .                                                                                                     | 61 |
| 3-1  | Détection d'un piéton en mouvement et la détermination de sa position<br>sur l'image [58] . . . . .                                                                       | 72 |
| 3-2  | Extraction de l'orientation du piéton à partir des positions communes<br>[25] . . . . .                                                                                   | 74 |
| 3-3  | Calcul du vecteur d'orientation pour le corps humain [71] . . . . .                                                                                                       | 75 |
| 3-4  | Image représentant l'installation de la caméra utilisée à l'intérieur du<br>véhicule pour l'acquisition de la base SafeRoad . . . . .                                     | 81 |
| 3-5  | Annotation des images de piétons de la base SafeRoad en 4 directions                                                                                                      | 81 |
| 3-6  | Annotation des images de piétons de la base SafeRoad en 8 directions                                                                                                      | 82 |
| 3-7  | Les différentes conditions d'acquisition de la base SafeRoad . . . . .                                                                                                    | 83 |
| 3-8  | Image représentant des piétons disciplinés prise de la base SafeRoad .                                                                                                    | 83 |
| 3-9  | Image représentant des piétons indisciplinés prise de la base SafeRoad                                                                                                    | 83 |
| 3-10 | Image représentant une route structurée de la base SafeRoad . . . . .                                                                                                     | 84 |
| 3-11 | Image représentant une route non structurée de la base SafeRoad . .                                                                                                       | 84 |
| 3-12 | Représentation des différents comportements des piétons acquissent<br>dans la base SafeRoad . . . . .                                                                     | 85 |
| 3-13 | Images de piétons de la base Daimler . . . . .                                                                                                                            | 86 |
| 3-14 | Images de piétons de la base TUD . . . . .                                                                                                                                | 87 |
| 3-15 | Architecture CapsNet utilisé pour détecter l'orientation du piéton . .                                                                                                    | 88 |
| 3-16 | Représentation de la perte de marge d'une valeur de 0.004, la perte de<br>reconstruction d'une valeur de 0.1 ainsi que la perte totale d'une valeur<br>de 0.014 . . . . . | 90 |
| 3-17 | Exemple d'images reconstruites des réseaux de Capsules durant la<br>phase d'apprentissage . . . . .                                                                       | 91 |
| 3-18 | Matrice de Confusion des réseaux CapsNet, AlexNet et ResNet sur la<br>base Daimler . . . . .                                                                              | 93 |
| 3-19 | ROC curve of Capsule Network on Daimler dataset . . . . .                                                                                                                 | 94 |
| 3-20 | Exemple d'images représentant une mauvaise classification . . . . .                                                                                                       | 95 |

|      |                                                                                                       |     |
|------|-------------------------------------------------------------------------------------------------------|-----|
| 3-21 | Pedestrian orientation recognition using Capsule Networks on Safe-Road dataset . . . . .              | 95  |
| 3-22 | Matrices de Confusion Matrices des réseaux CapsNet, AlexNet et ResNet pour la base SafeRoad . . . . . | 97  |
| 3-23 | CapsNet architecture pour la détection de 8 orientations . . . . .                                    | 99  |
| 3-24 | Les matrices de confusion des réseaux CapsNet, AlexNet et ResNet sur la base SafeRoad . . . . .       | 101 |
| 3-25 | Courbe ROC du réseau de Capsule sur la base SafeRoad . . . . .                                        | 102 |
| 3-26 | Matrices de confusion de CapsNet, AlexNet et ResNet sur la base TUD                                   | 104 |
| 3-27 | Courbe ROC curve du réseau de Capsule sur la base TUD . . . . .                                       | 105 |
| 4-1  | Schéma explicatif du réseau de régression profonde [59] . . . . .                                     | 112 |
| 4-2  | schéma explicatif de l'algorithme de suivi ROLO [95] . . . . .                                        | 113 |
| 4-3  | Architecture du DeepSort [142] . . . . .                                                              | 115 |
| 4-4  | Reconnaissance des orientations différentes de plusieurs piétons sur la même image . . . . .          | 118 |
| 4-5  | Détection de la direction avant le changement . . . . .                                               | 119 |
| 4-6  | Détection de la direction après son changement rapide . . . . .                                       | 120 |
| 4-7  | Détection de l'orientation des piétons occultés qui courent . . . . .                                 | 120 |
| 4-8  | Détection de l'orientation d'un enfant courant . . . . .                                              | 121 |
| 4-9  | Exemple de piétons ayant une orientation oblique . . . . .                                            | 122 |
| 4-10 | Exemple d'une mauvaise classification de l'orientation d'un piéton dû à la distance . . . . .         | 123 |
| 4-11 | Exemple de piétons non détectés dans l'obscurité . . . . .                                            | 123 |
| 4-12 | Exemple d'une mauvaise classification de l'orientation des piétons dans l'obscurité . . . . .         | 124 |
| 4-13 | Matrice de confusion pour N=1 . . . . .                                                               | 125 |
| 4-14 | Matrice de confusion pour N=5 . . . . .                                                               | 125 |
| 4-15 | Matrice de confusion pour N=10 . . . . .                                                              | 126 |

# Liste des tableaux

|      |                                                                                                            |     |
|------|------------------------------------------------------------------------------------------------------------|-----|
| 3.1  | Comparaison des résultats des trois méthodes sur l'orientation du piéton . . . . .                         | 73  |
| 3.2  | Performance du réseau CNN par rapport aux méthodes traditionnelles sur la base Human 1.6M . . . . .        | 74  |
| 3.3  | Performance du réseau CNN par rapport aux méthodes traditionnelles sur la base Human 1.6M . . . . .        | 76  |
| 3.4  | Classification de l'orientation du piéton basé sur différentes architectures sur la base Daimler . . . . . | 90  |
| 3.5  | Comparaison de l'exactitude entre les réseaux AlexNet, ResNet et CapsNet pour la base Daimler . . . . .    | 91  |
| 3.6  | Comparaison des métriques de classification pour chaque classe de la base Daimler . . . . .                | 92  |
| 3.7  | Comparaison des résultats entre les réseaux AlexNet, ResNet et CapsNet pour la base SafeRoad . . . . .     | 94  |
| 3.8  | Comparaison des métriques de classification pour chaque classe de la base SafeRoad . . . . .               | 96  |
| 3.9  | Comparaison avec différents algorithmes d'optimisation pour la base SafeRoad . . . . .                     | 98  |
| 3.10 | Comparaison de la performance de différentes architectures sur la base SafeRoad . . . . .                  | 100 |
| 3.11 | Comparaison des métriques de classification metrics pour chaque classe de la base SafeRoad . . . . .       | 102 |

|      |                                                                                                    |     |
|------|----------------------------------------------------------------------------------------------------|-----|
| 3.12 | Comparaison de la performance de classification sur la base TUD pour<br>huit orientations. . . . . | 103 |
| 3.13 | Comparaison des métriques de classification pour chaque classe de la<br>base TUD . . . . .         | 103 |



# Acronymes

**AASHTO** American Association of State Highway and Transportation Officials

**ABS** Anti-lock Braking System

**ACC** Adaptive Cruise Control

**ACF** Aggregated Channel Features

**AdaBoost** Adaptive Boosting

**ADAS** Advanced Driver Assistance System

**AEB** Automatic Emergency Braking

**APROSYS** Advanced Protection Systems

**BSD** Blind Spot Detection

**CapsNet** Capsule Network

**CAS** Collision Avoidance System

**CNN** Convolutional Neural Network

**DTS** Distance To Stop

**ESC** Electronic Stability Control

**FHWA** Federal Highway Administration

**GNSS** Global Navigation Satellite System

**HOG** Histograms of Oriented Gradients

**ICF** Integral Channel Features

**IHC** Intelligent Headlight Control

**IMM** Interacting Multiple Model

**ITS** Intelligent Transportation System

**JAAD** Joint Attention in Autonomous Driving

**KF** Kalman Filter

**LBP** Local Binary Patterns

**LDW** Lane Departure Warning

**LSTM** Long Short-Term Memory

**METLE** Ministère de l'Équipement, du Transport, de la Logistique et de l'Eau

**MNIST** Modified National Institute of Standards and Technology database

**MOT** Multiple Object Tracking

**OMS** Organisation Mondiale de la Santé

**PCAS** Pedestrian Collision Avoidance System

**RCNN** Region Based Convolutional Neural Networks

**RCTA** Rear Cross Traffic Alert

**ROC** Receiver Operating Characteristic

**RoI** Region of Interest

**ROLO** Recurrent You Only Look Once

**SCCR** Short Circuit Current Rating

**SIFT** scale-Invariant Feature Transform

**SORT** Simple Online Real-Time Tracking

**SURF** Speeded Up Robust Features

**SVM** Support Vector Machine

**TCCR** Temporal Crossing Compliance Rate

**TCS** Traction Control System

**TSA** Traffic-Sign Assist

**TTS** Time To Stop

**TTZ** Time to Zebra

**VJ** Viola and Jones

**V2I** Vehicle to Infrastructure

**V2V** Vehicle to Vehicle

**YOLO** You Only Look Once

# Introduction

Les accidents de la route représentent la première cause de décès chez les jeunes de moins de 30 ans. Environ 1,35 million de personnes meurent chaque année dans le monde suites d'un accident de la route, dont 23% des victimes sont des piétons. Selon l'Organisation Mondiale de la Santé (OMS), 93% des décès sur la route surviennent dans les pays à revenu faible et intermédiaire en raison de leur population élevée [141].

Au Maroc, la sécurité routière correspond à un enjeu majeur de santé publique et de protection des personnes. Selon les statistiques du ministère de l'Équipement, du Transport, de la Logistique et de l'Eau (METLE), le nombre d'accidents corporels a atteint 89.375 pour l'année de 2017, où chaque jour une moyenne de 10 personnes décèdent et 361 sont blessées suite à des accidents de la circulation selon les statistiques livrées par le ministère (METLE). La sécurité routière est donc placée au cœur des politiques publiques du pays [90].

Dans ce contexte, METLE a lancé des appels à projets de recherche en sécurité routière pour remédier aux problèmes d'accidents de la circulation au Maroc. Notre projet intitulé "SAFEROAD : une Méta-plateforme pour la sécurité routière", est l'un des 6 projets acceptés parmi 130 projets soumis à travers le Maroc.

## Description du projet SafeRoad

SafeRoad est un projet qui relève d'une étude scientifique bien réfléchiée en sécurité routière, il traite plusieurs volets phares de cette problématique ; à savoir : la détection des piétons, l'analyse comportementale des usagers de la voie publique, la surveillance du conducteur, mesure des indicateurs de la sécurité routière, scénarios accidentogènes

et évaluation prévisionnelle.

L’objectif de ce projet est la réalisation, le test et l’évaluation de prototypes de systèmes d’aide à la conduite et la réalisation d’un système de transport intelligent. Ce projet doit contribuer à la mise en place de nouvelles stratégies de développement de solutions de ce type.

Il est divisé en trois aspects fondamentaux (Système routier avec le véhicule comme acteur, système routier avec l’environnement comme acteur et système routier avec l’Homme comme acteur).

## **Système routier avec le véhicule comme acteur**

L’objectif de cette partie est de développer un système qui avertit le conducteur de la possibilité d’une collision, en utilisant la détection d’entités autour du véhicule, la communication entre les véhicules (V2V) et entre les véhicules et l’infrastructure (V2I) [119, 120, 121].

## **Système routier avec l’environnement comme acteur**

Cette partie se concentre principalement sur trois éléments essentiels :

- Conception de l’infrastructure : Dans cet élément, la priorité doit être donnée en premier lieu aux points suivants : les alignements horizontaux et verticaux, la largeur de la chaussée, le marquage au sol, la conception des intersections, les espaces dédiés aux piétons, les aires de stationnement, la distance de visibilité, etc.
- Détection de panneaux de signalisation : Il s’agit ici de détecter et de reconnaître, à partir de l’analyse d’images, certaines formes caractéristiques des panneaux de signalisation (carré, triangle, rond, octogone). Les méthodes utilisées dans cette partie sont basées sur la segmentation et la classification en traitement d’images.
- Gestion du trafic routier : Nous proposons d’installer des caméras intelligentes dans un environnement contrôlé, à des endroits spécifiques mettant en évi-

dence un ou plusieurs facteurs d'accident. En plus de la sécurité que ces caméras peuvent apporter, elles permettront de gérer le trafic via le comptage des véhicules.

## **Système routier avec l'Homme comme acteur**

Cette partie représente l'un des principaux aspects du projet. Dans le cadre de cette thèse, nous avons contribué dans cette partie en accentuant nos recherches sur la sécurité de l'utilisateur le plus vulnérable de la route qui est le piéton. Ce dernier représente 27.70% des accidents mortels au Maroc selon les statistiques de l'année 2017. Nous allons présenter une contribution pour améliorer la sécurité du piéton. Dans ce contexte, la section suivante décrit nos principales contributions.

## **Contributions**

L'objectif de cette thèse est de protéger le piéton comme étant l'utilisateur le plus vulnérable de la route, plus précisément lorsque l'infrastructure routière possède un manque dans les mesures de sécurité pour les piétons, ce qui les conduisent à traverser la route de façon non convenable.

Afin de réduire les accidents mortels impliquant les piétons, de nombreuses recherches sont mises en oeuvre sur les Systèmes Avancés d'Aide à la Conduite (ADAS) notamment pour l'atténuation des accidents de piétons «Systèmes de préventions des collisions avec piétons» (PCAS) qui sont utilisés chez plusieurs constructeurs automobiles haut de gamme.

À cette fin, les recherches dans ces systèmes sont davantage axées sur la détection des piétons que sur la reconnaissance de leur direction de marche. Par conséquent, rares sont ceux qui ont l'intention d'inclure l'orientation des piétons dans ces systèmes.

Dans le cadre de ce travail, nous proposons un système de détection des orientations des piétons qui pourrait être intégré dans les systèmes avancés d'aide à la conduite (ADAS) afin d'alerter le conducteur de la présence d'un piéton.

Dans ce cadre, nos principales contributions dans ce travail incluent :

- La proposition d’un nouveau système PCAS prenant en considération la présence des piétons indisciplinés et des routes les moins structurées.
- Création de la première base de données marocaine sur la direction des piétons que nous avons appelée SafeRoad. Cette dernière est collectée à partir de scènes naturelles depuis différentes villes marocaines, en utilisant un enregistrement vidéo d’une caméra intégrée au bord du véhicule en mouvement. Après l’analyse de cette base, nous constatons que le code de la route est peu respecté par les conducteurs et les piétons. Ces derniers partagent la chaussée avec les véhicules, en particulier dans les zones mal structurées. En raison de ce comportement, l’incidence des accidents de piétons au Maroc est très élevé.
- Proposition d’une nouvelle technique de détection de l’orientation du piéton en utilisant les réseaux de Capsules. Cette technique peut être intégrée dans un ADAS afin de détecter les piétons indisciplinés et alerter le conducteur de leur présence. L’apprentissage et l’évaluation de cette technique sont faits sur notre base SafeRoad et sur la base Daimler.
- Application de la reconnaissance de l’orientation du piéton sur des scènes vidéos prises de la base JAAD.

## Organisation de la thèse

Ce rapport contient 4 chapitres qui sont organisés comme suit :

Une introduction générale qui décrit le projet SafeRoad et cite les principales contributions de cette thèse. Ensuite nous présentons dans le premier chapitre un état de l’art sur l’évolution des ADAS depuis leur apparition. Il présente également les différentes techniques de vision par ordinateur utilisées pour la détection des objets notamment les piétons, allant des méthodes de détection par descripteur jusqu’à l’apprentissage profond. Ensuite, nous proposons un nouveau système de prévention des accidents piéton-véhicule dans le deuxième chapitre. Ce dernier aborde en première partie le conflit piéton-véhicule en décrivant le comportement des conducteurs et des piétons en fonction de leurs âges et leurs genres. Le deuxième chapitre décrit par la

suite l'interaction et la communication entre les piétons et les véhicules et présente ensuite une étude sur les systèmes PCWS existants avant d'introduire notre système proposé qui implique l'intégration de l'orientation du piéton pour évaluer le risque d'un accident.

La détection de l'orientation du piéton est entamée par la suite dans le troisième chapitre en utilisant les réseaux de Capsules. Ce chapitre représente un des noyaux importants de cette thèse. Il introduit la création de la première base de données marocaine pour la direction des piétons. Cette base est labellisée manuellement en 4 et 8 orientations. Nous avons proposé par la suite une nouvelle approche de la détection de l'orientation du piéton basée sur les réseaux de Capsule pour 4 et 8 directions, entraîné sur la base marocaine SafeRoad, et les bases publiques Daimler et TUD Multiview Pedestrians. La dernière partie du chapitre contient les expérimentations et la comparaison des résultats avec les réseaux AlexNet et ResNet.

Le réseau est entraîné sur la base SafeRoad est par la suite appliqué pour la reconnaissance de l'orientation des piétons sur des séquences vidéo prises de la base JAAD. Ce dernier chapitre traite trois points essentiels pour la reconnaissance de l'orientation sur vidéo. Nous avons dans un premier temps détecter les piétons sur chaque séquence vidéo en utilisant l'algorithme issu d'un réseau de neurones "You Only Look Once" (YOLO). Ensuite, chaque piéton détecté est suivi en lui attribuant un identifiant unique jusqu'à sa disparition en utilisant l'algorithme DeepSort. Et enfin, le réseau de Capsules entraîné sur la base SafeRoad est utilisé pour la reconnaissance de l'orientation de chaque piéton détecté et suivi.

Finalement nous allons présenter les conclusions et discuter les perspectives de ce travail de recherche.





# Chapitre 1

## Etat de l'art

### 1.1 Introduction

Les Systèmes Avancés d'Aide à la Conduite (ADAS) sont devenus indispensables dans les véhicules d'aujourd'hui. En raison du besoin croissant de mobilité, le trafic est devenu beaucoup plus complexe, c'est donc un défi de plus en plus grand pour tous les usagers de la route. Les ADAS sont essentiels pour éviter les accidents et les blessures concomitantes ou les décès possibles. Par ailleurs, ils apportent des solutions pour une conduite confortable, économique et intelligente. Ces dernières années, de plus en plus de types d'unités de commande complexes ont été développés et intégrés dans les véhicules. Ces systèmes diffèrent dans leurs principes de fonctionnement et leurs domaines d'application. Les ADAS utilisent des capteurs environnants tels que radar, infrarouge, vidéo, pour surveiller et analyser l'environnement d'un véhicule.

La majorité des ADAS utilisent la vision par ordinateur pour alerter le conducteur de la présence d'un accident imminent, en utilisant des techniques de détection et de classification pour reconnaître la nature de l'objet autour du véhicule (véhicule, piéton, etc). Les méthodes de détection d'objets, ont connu une amélioration importante en terme de performance ces dernières années, notamment avec l'apparition des techniques de l'apprentissage profond tel que les Réseaux de Neurones Convolutifs (CNN).

Dans ce chapitre, nous allons présenter un état de l'art sur les différentes applica-

tions des ADAS depuis leur apparition ainsi que leur évolution dans le temps. Nous allons faire le point également sur les méthodes de détection d'objets, allant des méthodes classiques en utilisant les descripteurs et les classifieurs, jusqu'aux méthodes récentes utilisant les CNN pour une détection plus précise et rapide permettant une utilisation en temps réel.

## 1.2 Systèmes Avancés d'Aide à la Conduite

L'objectif des Systèmes Avancés d'Aide à la Conduite (ADAS) est de prévenir les accidents de la route, ou de réduire les conséquences au cas où l'accident ne peut pas être évité. Le développement des systèmes d'assistance à la conduite a commencé avec les systèmes de freinage (ABS) introduits dans une production en série à la fin des années 70.

Les principales étapes du développement des systèmes d'aide à la conduite peuvent être classées comme suit :

1. L'utilisation des capteurs proprioceptifs (permettant de mesurer l'état interne du véhicule tels que la vitesse, position des roues, etc)
2. L'intégration des capteurs extéroceptifs (permettant de réagir aux stimuli externes du véhicule)
3. L'utilisation des réseaux de capteurs

Les capteurs proprioceptifs sont capables de détecter et de réagir à une situation de danger en analysant le comportement du véhicule (vitesse ou position des roues, etc). Par contre les extéroceptifs tels que les ultrasons, radar, lidar, infrarouge et de vision sont capables de réagir rapidement et de prévoir les dangers possibles en analysant l'environnement extérieur du véhicule. Des améliorations supplémentaires sont attendues par l'application de plates-formes multisensorielles et de réseaux de capteurs de trafic. Dans cette section, nous allons faire le point sur les capteurs extéroceptifs.

Les ADAS n'agissent pas de manière autonome mais fournissent des informations supplémentaires sur la situation du trafic pour accompagner un conducteur et l'assister dans la mise en œuvre des actions critiques. La synchronisation des actions du conducteur et des informations de l'environnement ainsi que la reconnaissance des éventuelles manœuvres du véhicule sont essentielles pour la performance efficace des diverses applications d'ADAS. Quelques exemples d'application ADAS sont décrits ci-dessous.

La détection des angles morts (BSD) surveille la zone à côté d'un véhicule. La fonction d'un système de détection des angles morts est d'avertir le conducteur lors-

qu'il y a des objets dans les angles morts, avec un signal visuel tel qu'un voyant qui s'allume sur le rétroviseur ou avec un alerte sonore. Le but de ce système est d'éviter les accidents potentiels, en particulier lors des manœuvres de changement de voie vers un trafic dense [123, 124, 60, 110, 145, 45, 62].

L'alerte de circulation transversale arrière (RCTA) permet d'éviter les accidents lors de la marche arrière spécialement dans un parking, ce qui peut souvent entraîner de graves accidents avec des piétons ou cyclistes qui impliquent des blessures corporelles. Pour cette fonction, l'environnement derrière le véhicule est surveillé et vérifié. Dans le cas où un objet est détecté dans le sens de marche arrière, le conducteur reçoit un avertissement sonore et visible [55, 72, 21, 147, 148, 68].

La commande intelligente des phares (IHC) régule les feux d'un véhicule selon les conditions environnementales. Cette application optimise les changements entre les feux de route et les feux de croisement pendant les trajets de nuit. Conduire la nuit ou à travers les tunnels est ainsi plus confortable et plus sécurisé. De plus, les conducteurs venant en sens inverse des véhicules ne sont plus aveuglés par les feux des véhicules [111, 112, 85, 78, 82, 41, 125].

Une autre application de sécurité est l'assistance aux panneaux de signalisation (TSA). Ce système reconnaît automatiquement les panneaux de signalisation (également les panneaux de différents pays) et peut traiter les informations qu'ils contiennent. Par conséquent, un conducteur est en mesure de recevoir des informations importantes telles que les limites de vitesse légales ou les règles de priorité réelles. Lorsque ces informations sont fournies, l'assistance aux panneaux de signalisation permet une conduite plus détendue et plus sûre [51, 92, 118, 10, 18, 130, 129, 86, 153].

L'avertissement de sortie de voie (LDW) scanne les côtés de la route et détecte quand un véhicule quitte involontairement sa voie ou une sortie de route. En contrôlant le mouvement de direction, le système est en mesure d'évaluer si le changement de voie est intentionnel. Le système avertit le conducteur que les voies ont été changées par inadvertance avec un avertissement visuel ou haptique tel que les vibrations du volant. Ce système permet de réduire les accidents de circulation causés par des véhicules quittant la route [66, 15, 144, 94, 77].

L'assistance au freinage d'urgence (AEB) améliore la sécurité de conduite grâce au freinage actif permettant de freiner automatiquement dans les situations dangereuses. Les collisions par arrière peuvent donc être évitées entièrement. De plus, les conséquences des accidents sont réduites en raison de la réduction de la vitesse d'impact et de l'énergie d'impact. Le frein d'urgence est également une interface possible pour les applications pré-collision et les systèmes de protection des piétons [20, 126, 56].

Le régulateur de vitesse adaptatif avec fonctions Stop & Go (ACC + S&G) contrôle la distance par rapport au véhicule qui précède, même en cas d'arrêt et de départ. Et ce, soit en avertissant le conducteur soit en ralentissant activement la vitesse du véhicule lorsque la distance relative devient plus courte. Cette application est bénéfique en particulier dans les situations de trafic congestionné, permettant au conducteur une conduite plus confortable et plus décontractée avec la fluidité du trafic. La sécurité est améliorée grâce à une distance prédéfinie, en cas d'urgence un freinage d'urgence est nécessaire [88, 132, 135, 91].

De nombreuses nouvelles applications mentionnées dans [14] sont en cours de développement et optimisées en permanence pour améliorer la sécurité des passagers, des piétons ou des animaux. Elles offrent un véritable confort de conduite et réduisent la fatigue sur la route.

### 1.2.1 Evolution des ADAS

L'évolution des systèmes d'assistance du conducteur d'un point de vue technologique est illustrée sur la Figure 1-1, montrant son développement depuis 1980 avec une évolution future potentielle après les années 2030.

L'un des premiers systèmes d'assistance actifs basés sur les capteurs proprioceptifs est le système de freinage antiblocage (ABS), avec une production en série à partir de 1978. Le système de contrôle de Traction (TCS) a par la suite complété l'évolution des ABS. Des années plus tard en 1995, l'introduction des systèmes de contrôle dynamique, tels que le contrôle électronique de la stabilité (ESC), a marqué une étape importante dans le développement des ADAS. En terme de sécurité routière, des études ont montré que les systèmes de contrôle dynamique sont les plus efficaces

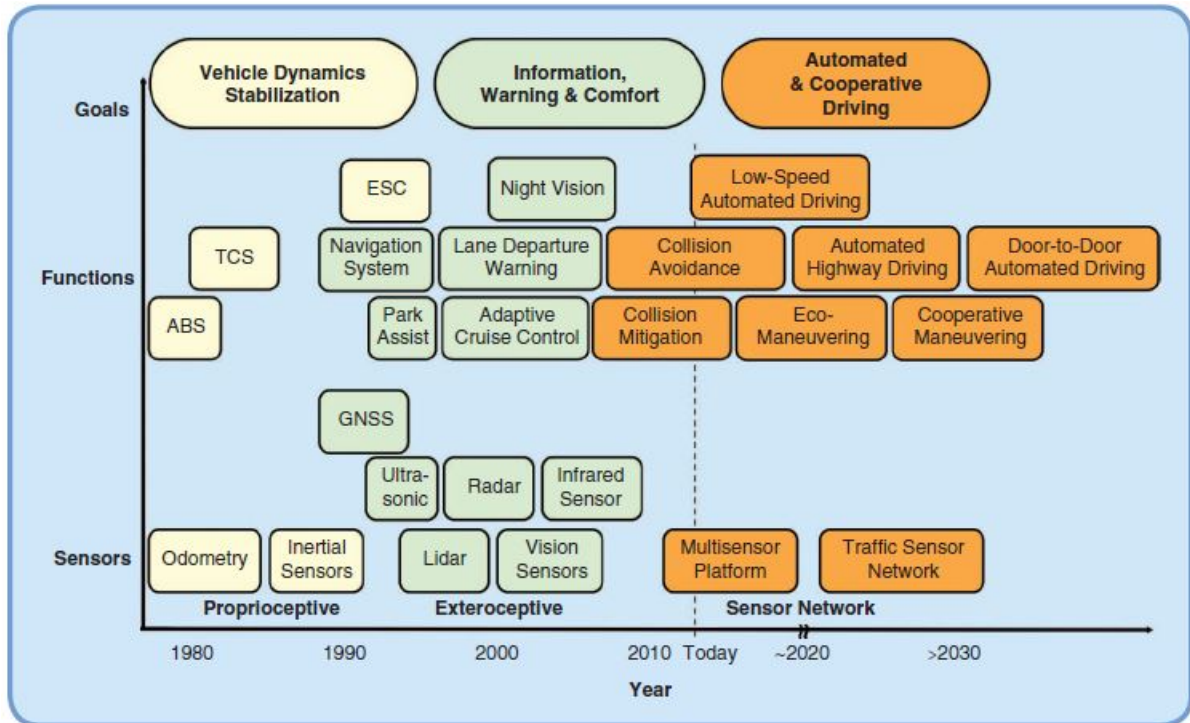


FIGURE 1-1 – Evolution des ADAS depuis 1980 [14]

pour la sécurité des passagers, après la ceinture du siège [2]. Avec la reconnaissance publique de la sécurité potentielle des systèmes de contrôle dynamique, la fréquence de la mise en œuvre de ces systèmes a augmenté de manière significative, et ils ont, en conséquence, sauvés plusieurs milliers de vies. À partir de novembre 2014, ESC (en plus des systèmes d’assistance de freinage développés en parallèle) sont devenus même une obligation légale pour chaque nouvelle voiture dans l’Union Européenne.

Les capteurs extéroceptifs (ultrasons, radar, lidar ou caméras, ...), quant à eux, acquièrent des informations du côté extérieur du véhicule. Ces capteurs fournissent des informations à propos de la route à parcourir, sur la présence des autres utilisateurs de la route, ainsi que sur la position du véhicule en utilisant les systèmes de positionnement par satellite (GNSS).

La deuxième génération des Systèmes d’Aide à la Conduite, introduite pour la première fois vers 1990, est basée sur les capteurs extéroceptifs. Elle se concentre sur la fourniture d’informations au conducteur, sur son avertissement en cas de besoin

ainsi que sur l'amélioration de son confort de conduite. Grâce à la réduction des coûts de navigation sur les appareils mobiles, la technologie utilisant le GNSS est devenue répandue dans les véhicules de nos jours. Permettant ainsi de réduire le risque des accidents dues à l'inattention, en aidant le conducteur dans son orientation et en réduisant ainsi la charge de travail du conducteur, en consacrant son attention à sa tâche primaire qui est la conduite.

Les Systèmes de Prévention de Collision (CAS) sont des systèmes proactifs de pré-collision permettant de prévenir un accident imminent en alertant le conducteur. En utilisant les niveaux d'alertes croissants, le conducteur est averti d'une collision imminente. Si ce dernier ne réagit pas à l'alerte, le véhicule freine activement pour atténuer la gravité de l'accident une fois qu'il n'est plus évitable [89]. Parmi ces systèmes, nous trouvons les systèmes anti-collision piétons-engins qui visent à réduire les accidents avec l'utilisateur de route le plus vulnérable qui est le piéton. Ces derniers systèmes utilisent des techniques de détection de piéton pour alerter le conducteur de sa présence mais n'incluent aucune information sur sa direction de marche. Notre intérêt dans cette thèse est d'améliorer les systèmes de prévention de collisions, notamment sur les zones peu structurées. Dans ce cadre, nous proposons un nouveau système d'alerte prenant en considération les piétons ayant un comportement à risque en se basant sur leur direction de marche.

Le secteur ADAS est actuellement dominé par les technologies basées sur les radars et les caméras. L'objectif à court terme est d'automatiser la conduite pour certaines situations. A titre d'exemple, les systèmes d'assistance aux embouteillages basés sur le radar et les caméras stéréo ont été récemment introduits. Ces systèmes sont conçus pour automatiser la conduite à basse vitesse sur les autoroutes encombrées, grâce à l'intégration du contrôle latéral et longitudinal. La vitesse de déplacement maximale est encore faible (30 km / h), limitée à des situations d'arrêt et de départ, mais cette fonction peut éventuellement émerger vers la conduite automatisée sur autoroute.

D'autres extensions des ADAS actuels sont à venir, tels que l'assistance pour la détection du trafic venant en sens inverse dans des conditions de vision défavorables. Certains de ces systèmes nécessitent l'échange de données entre les acteurs de la route



ou avec l’infrastructure.

Dans le cadre de cette thèse, nous proposons un nouvel ADAS pour la protection des piétons ayant un comportement à risque, notamment sur les environnements non structurés, en se basant sur la direction de marche du piéton.

## 1.3 Méthodes de détection de piétons

Les techniques de détection des piétons sont basées soit sur des descripteurs, soit sur l’apprentissage profond. Au cours des dernières années, les approches d’apprentissage profond ont maintenu leur supériorité sur les approches classiques. Compte tenu de cela, nous nous concentrons dans la suite de cette thèse, sur les approches basées sur l’apprentissage profond. Dans ce chapitre, nous mettons particulièrement l’accent sur la création d’une taxonomie des techniques de détection des piétons.

Nous commençons cette section en notant que presque toutes les techniques contemporaines de détection des piétons basées sur l’apprentissage profond sont dérivées de détecteurs d’objets classiques. Par conséquent, les mécanismes suivis dans différents détecteurs d’objets classiques ont un impact sur la conception et le fonctionnement de divers détecteurs de piétons. Partant de ce point de vue, nous classons d’abord les détecteurs d’objets classiques, puis nous les décrivons. Nous décrivons ensuite divers détecteurs de piétons comme des extensions et des modifications de détecteurs d’objets classiques.

### 1.3.1 Classification par descripteurs

Avant le succès des réseaux de neurones convolutifs dans le domaine de la vision par ordinateur, une variété de descripteurs de caractéristiques traditionnels, y compris SIFT [93], LBP [143], SURF [9], HOG [35] et Haar [138], ont été étudiés dans le contexte de la détection d’objets en général et dans la détection des piétons en particulier. Ces caractéristiques classiques extraient généralement de l’image des informations sur la couleur, la texture ou les bords.

Il existe deux types de descripteurs : les descripteurs locaux et les descripteurs globaux. Les premiers décrivent une simple portion de l'image. Généralement, il s'agit du voisinage d'un point précis de l'image qui contient des informations intéressantes pour décrire la scène ou l'objet présent dans celle-ci. Ces points appelés points d'intérêt sont généralement choisis via un détecteur. Une fois les points détectés, ils sont décrits par un descripteur local qui utilise le voisinage de ces points. Les descripteurs globaux font, quant à eux, une description de l'ensemble de l'image. Généralement, l'image est découpée et traitée en sous-blocs. Une description de chaque partie est calculée et au final une description de l'image entière est obtenue.

Nous allons présenter dans cette section les descripteurs les plus utilisés pour la détection du piéton.

### **Ondelettes de Haar**

L'analyse par ondelettes a été introduite dans les années 1980 pour étudier des signaux. Cette représentation donne simultanément des informations temporelles et fréquentielles, facilitant ainsi l'identification des caractéristiques physiques de la source du signal. D'abord utilisées en traitement du signal, leur champ d'applications s'est étendu à bien d'autres domaines, notamment au traitement d'images. Aujourd'hui elles représentent un descripteur largement répandu en reconnaissance des formes.

Papageorgiou et al. [103] ont adapté ces ondelettes afin de les utiliser en tant que descripteur d'images pour la reconnaissance d'objets. L'idée est d'encoder les différences d'intensité entre des zones particulières d'une image. Prenons l'exemple d'une image de visage en niveaux de gris (voir figure 1-2a) : il apparaît aisément que la bouche, le nez et les yeux sont des zones de changements d'intensité. Les ondelettes de Haar permettent alors de capturer ces caractéristiques en les adaptant en position, en taille et en orientation. La figure 1-2b schématise les différentes orientations usuellement exploitées en reconnaissance d'objets.

Pour caractériser au mieux un objet, chaque ondelette est adaptée en position et taille. Par exemple, pour un visage, une ondelette d'orientation horizontale est très souvent utilisée pour la bouche avec une taille adéquate, et une ondelette verticale

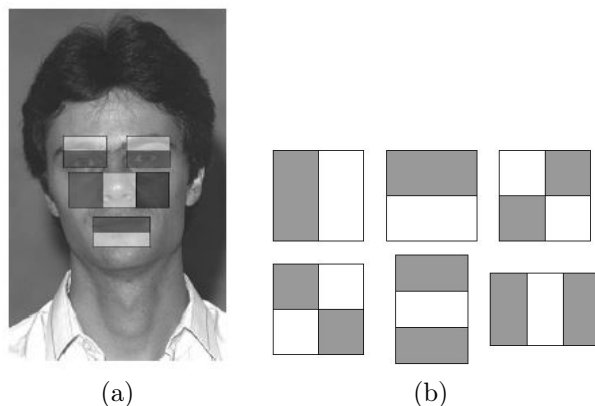


FIGURE 1-2 – (a) Description d'un visage avec les ondelettes de Haar. (b) Exemples des ondelettes de Haar [103]

pour caractériser le nez comme sur la figure 1-2a. Mais pour améliorer le modèle, les ondelettes sont déplacées autour de la position initiale afin d'obtenir un ensemble complet pour décrire l'image.

### Histogramme de Gradient Orienté (HOG)

Une des méthodes les plus populaires et les plus utilisées pour la détection des piétons est l'histogramme du gradient (HOG) [35, 48, 133, 70, 8]. La technique se base sur le calcul de gradient qui peut se faire en tout point d'une image. De façon générique, le gradient permet de calculer les variations d'une fonction par rapport aux variations de ses différents paramètres. En ce qui concerne le calcul de gradients dans les images de luminance, il s'agit de calculer la variation de l'intensité des pixels dans différentes directions. Cela revient à un calcul de gradient 1D dans les directions intéressantes (classiquement en horizontal, vertical ou diagonal). Un gradient 1D horizontal ou vertical est tout simplement un calcul de la dérivée partielle de la fonction image  $I(x, y)$ .

Les histogrammes de gradients orientés, permettent de calculer les occurrences des orientations du gradient dans une portion localisée de l'image. Ils ont été introduits par Dalal et Triggs [35] pour faire la reconnaissance des formes sur des piétons dans des images fixes en trois étapes, comme illustré sur la figure 1-3.

— Première étape : calcul du gradient. Il s'agit de calculer le gradient pour tous

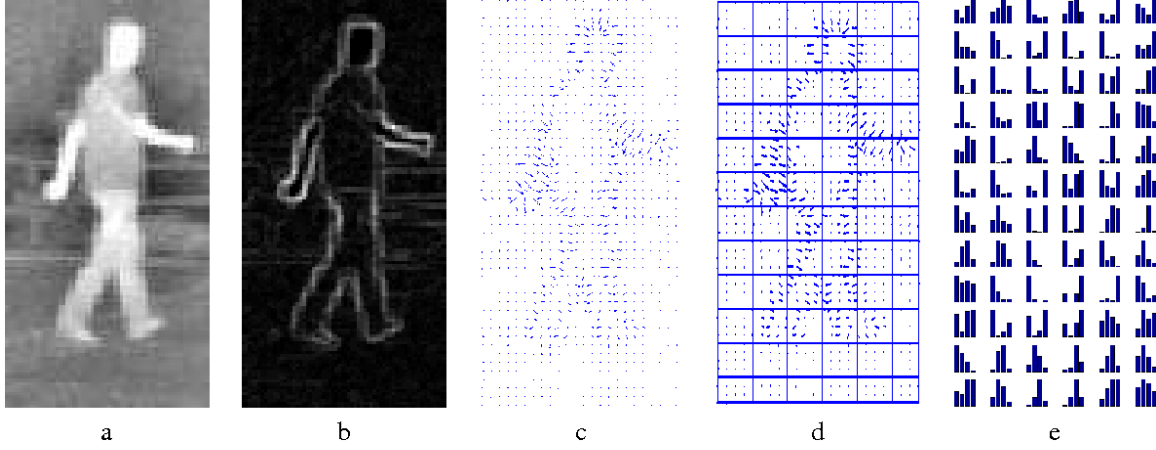


FIGURE 1-3 – Etapes de calcul d’un HOG : a) Image initiale, b) norme du gradient. c) orientation du gradient. d) division de l’image en plusieurs cellules. e) calcul du histogramme. [17]

les points de l’image ; usuellement deux masques de dérivation sont appliqués sur l’image, un horizontal- (eq.1.1) - et un vertical - (eq.1.2)

$$M_1 = [-1, 0, 1] \quad (1.1)$$

$$M_2 = [-1, 0, 1]^T \quad (1.2)$$

- Deuxième étape : calcul de l’orientation. Pour chaque point de l’image, une approximation de la composante horizontale du gradient notée  $G_x$  et de la composante verticale notée  $G_y$  est ainsi obtenue. La plupart du temps la valeur absolue du gradient est prise car c’est le contraste entre deux régions qui importe le plus : un objet noir sur un fond blanc aura donc la même réponse qu’un objet blanc sur un fond noir.

Une fois que le calcul du gradient a été effectué pour tous les pixels de l’image, il faut en calculer l’orientation ; celle-ci peut être définie par l’équation 1.3

$$\theta = \arctan\left(\frac{G_y}{G_x}\right) \quad (1.3)$$

- Troisième étape : Construction de l’histogramme. L’image est divisée en plu-

sieurs cellules. Pour chacune d'entre elles, un histogramme de gradients orientés est construit en comptabilisant les occurrences du gradient dans une barre correspondant à un intervalle d'orientation spécifique.

Une barre d'histogramme correspond au nombre d'occurrences de l'orientation du gradient pour un intervalle donné. Le choix du nombre de barres influe directement sur la précision des informations : un petit nombre de barres réduit l'information disponible, mais trop de barres risquent de ne pas faire ressortir une information globale intéressante.

### 1.3.2 Classifieurs

Le descripteur permet d'extraire les caractéristiques d'une image, puis le classifieur est entraîné sur ces caractéristiques. Le rôle du classifieur est de déterminer les caractéristiques communes aux exemples d'une même classe afin de pouvoir ultérieurement reconnaître à quelle classe appartient un nouvel exemple inconnu.

Dans cette section, nous présentons les méthodes de classification classiques les plus connues telles que SVM et AdaBoost.

#### Adaptive Boosting (AdaBoost)

L'algorithme de Boosting [46, 149, 115, 61] permet de construire un classifieur fort en apprenant des erreurs de plusieurs algorithmes faibles. L'algorithme commence par créer un modèle à partir des données d'apprentissage. Ensuite, un second modèle est créé, ce dernier essaie de corriger les erreurs présentes dans le premier modèle. Les modèles sont créés séquentiellement, chacun corrigeant les erreurs du modèle qui le précède, jusqu'à ce que les données d'apprentissage soient correctement prédites, ou bien à ce que le nombre maximal des modèles soit atteint. De cette façon, il est facile d'intégrer un grand nombre de classifieurs faibles dans la création du classifieur final.

La technique de Boosting la plus populaire est l'Adaptive Boosting (AdaBoost), introduit en 1996 par Freund et Schapire [46]. Basée sur le principe fondamental du Boosting, l'AdaBoost associe plusieurs classifieurs faibles afin de créer un classifieur

fort.

Un seul classifieur peut ne pas être précis pour prédire la classe d'un objet, par contre, un modèle fort peut être construit en groupant multiples classifieurs faibles entre eux, où chacun d'eux apprend progressivement des mauvaises classifications des autres. Les classifieurs mentionnés peuvent être l'arbre de décision, régression logistique, etc.

La nouveauté introduite par l'AdaBoost est de proposer une distribution de probabilités a priori sur tout l'ensemble d'apprentissage en fonction de la réponse de l'algorithme à l'itération précédente.

Notons  $S$  l'ensemble d'apprentissage composé de  $N$  exemples tel que :  $S = (x_i, y_i)_{i=1, \dots, N}$  où  $x_i$  représente le vecteur de caractéristiques d'un exemple et  $y_i$  représente un scalaire indiquant la classe de l'objet. Dans le cas présent d'une classification en deux classes,  $y_i = 1$  si l'objet appartient à la classe d'objet recherché (exemple positif), et  $y_i = -1$  sinon (exemple négatif). Notons  $p_t$  un vecteur poids dont chaque composante est le poids associé à chaque exemple ; initialement  $p_t = p_1$  et tous les éléments ont la même valeur. Un échantillon, noté  $S_1$  et comprenant  $n_1$  exemples (avec  $n_1 < N$ ), est ensuite sélectionné parmi l'ensemble d'apprentissage. Un classifieur  $C_1$  est entraîné sur cet échantillon, qui fournit une hypothèse  $h_1$  correspondant directement à la classe de l'objet. Cette règle de décision est ensuite appliquée pour tous les exemples compris dans  $S$  afin de calculer l'erreur  $\epsilon_1$  du classifieur  $C_1$  sur l'ensemble  $S$ . Le poids correspondant à chaque exemple est ensuite mis à jour en fonction du résultat de la classification. Si un exemple est bien classé, son poids diminue et s'il est mal classé son poids augmente. Pour cela, un coefficient  $\alpha_1$  est calculé tel que :

$$\alpha_1 = \frac{1}{2} \ln\left(\frac{1 - \epsilon_1}{\epsilon_1}\right) \quad (1.4)$$

Où  $\epsilon_1$  représente l'erreur du classifieur  $C_1$ .

## Machine à Vecteurs de Support (SVM)

Développés dans les années 1990, les Machines à Vecteurs de Support (SVM) [96, 24, 5, 80] sont un algorithme d'apprentissage automatique supervisé avec des séparateurs à vaste marge. Il vise à classer les données en les séparant par une frontière, de façon à ce que la distance (marge) entre la frontière et les classes soit maximale, où les vecteurs de supports sont les données les plus proches de la marge. Leur objectif est de trouver un hyperplan dans un espace de  $N$  dimensions qui classe les points de données. Un hyperplan dans un espace de deux dimensions est une ligne qui sépare les deux classes. Pour une distribution de trois dimensions, l'hyperplan sera une surface de deux dimensions et non une ligne. Il existe plusieurs possibilités pour choisir un hyperplan pour séparer les classes des données, l'objectif est de choisir le plan ayant la marge (distance) maximale entre les points de données des deux classes, comme représenté sur la figure 1-4.

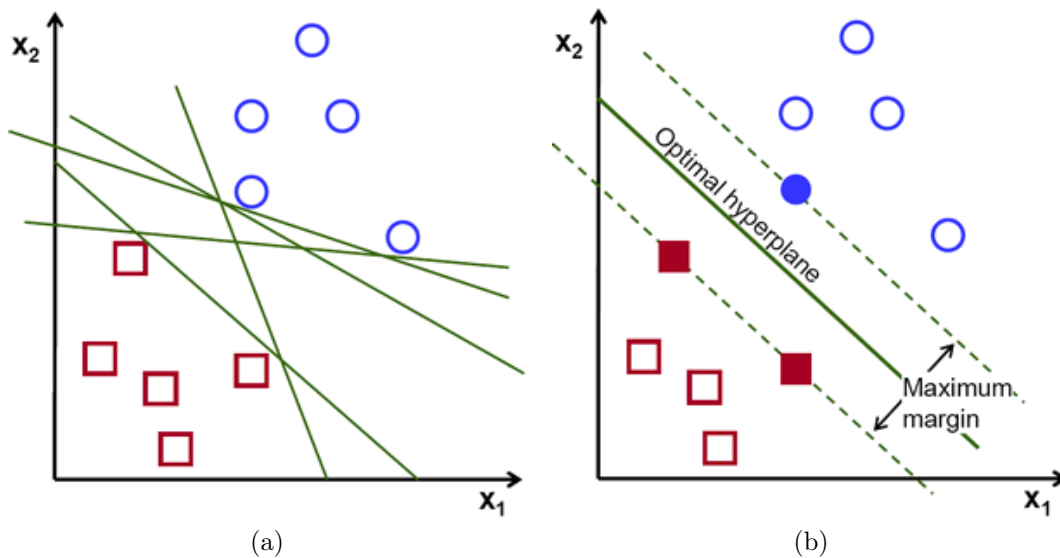


FIGURE 1-4 – Le choix de l'hyperplan optimal [96]

Les vecteurs de support sont les points de données qui sont proches de l'hyperplan et influencent la position et l'orientation de l'hyperplan. L'utilisation de ces vecteurs de support maximise la marge du classifieur et leur suppression changera la position de l'hyperplan. La figure 1-5 illustre la position des vecteurs de support.

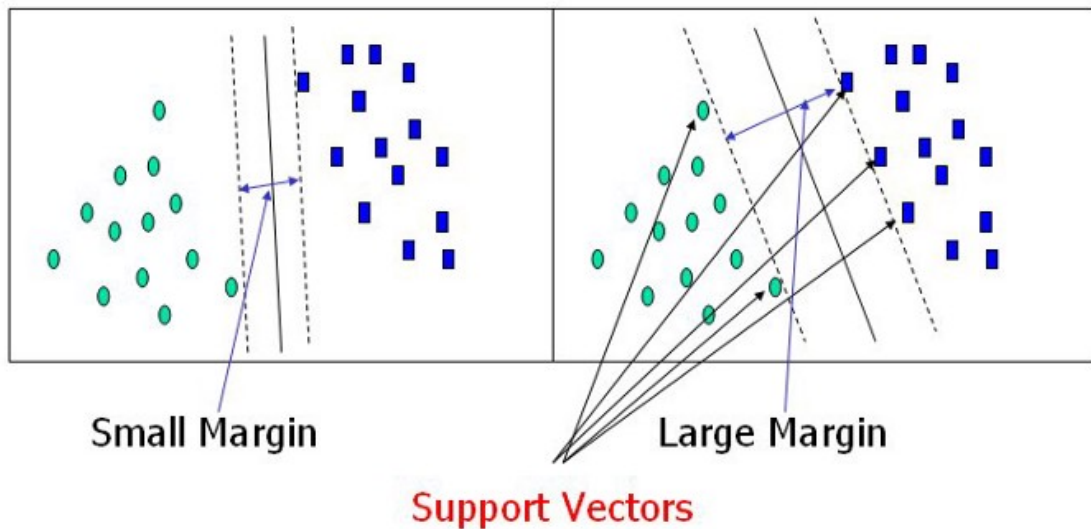


FIGURE 1-5 – L'influence des vecteurs de support sur la marge de l'hyperplan [96]

### 1.3.3 Approches classiques de détections de piétons

L'apprentissage se fait en utilisant des échantillons positifs et négatifs, les caractéristiques de ses échantillons sont tout d'abord extraites par les méthodes classiques (ex : HOG). Une fois que les caractéristiques extraites, les classifieurs (ex : SVM, Boosting,...) sont alors utilisés pour apprendre au détecteur à distinguer entre le piéton (classe positive) et son arrière-plan (classe négative). Généralement, les échantillons positifs et les échantillons négatifs sont générés par la méthode des fenêtres glissantes d'une taille fixe (par exemple,  $128 \times 64$  pixels) sur les images d'apprentissage. Le détecteur de piétons est alors utilisé pour classer ou détecter les piétons dans des images de tests non observées pendant l'apprentissage.

La phase de test est similaire à celle de l'apprentissage, les caractéristiques sont d'abord extraites des échantillons, en faisant glisser une fenêtre d'un pas fixe sur les images de test. Une fois que les caractéristiques des échantillons extraites, elles sont envoyées dans un détecteur pour la classification.

Parmi les méthodes les plus populaires de classification par descripteurs nous trouvons la méthode Viola et Jones (VJ).



La méthode VJ extrait d'abord les caractéristiques Haar pour chaque fenêtre de détection (échantillon), puis utilise la cascade AdaBoost pour l'apprentissage. Initialement, la méthode VJ a été utilisée pour effectuer la détection de visage [3, 28, 81, 7]. Étant donné que les caractéristiques de Haar sont basées sur les opérations de somme locale et de différence, la technique de l'image intégrale peut être facilement incorporée pour accélérer l'extraction des caractéristiques traitées. Basée sur la structure en cascade et l'image intégrale, la méthode VJ est le premier détecteur de visage en temps réel capable d'être déployé sur un périphérique CPU. Pourtant, par rapport à la détection de visage, la détection des piétons est plus difficile, et la structure initiale de VJ s'est avérée moins efficace sur cette tâche.

Les caractéristiques du canal intégrales (ICF) est une méthode de détection d'objets utilisée par Dollar et al. [37] pour la détection des piétons. Cette méthode utilise les images intégrales pour extraire des caractéristiques de l'image à partir de plusieurs canaux de l'image (LUV, magnitude du gradient, histogramme).

La méthode permet de détecter les objets en suivant ces différentes étapes :

1. La méthode calcule en premier les canaux de l'image d'entrée : trois types de canaux sont utilisés (canaux de couleurs - histogrammes du gradient - magnitude du gradient).
2. La deuxième étape est l'extraction des caractéristiques à partir de chaque canal.
3. La troisième étape est l'apprentissage. Dollar et al. ont utilisé le classifieur AdaBoost pour l'apprentissage.
4. Finalement, le modèle entraîné est utilisé pour la détection des objets.

Par la suite, plusieurs variantes de la méthode ICF ont été proposées, la plus connue d'entre elles est celle de Dollár et al. [36] appelée "Les caractéristiques du canal agrégées (ACF)". L'ACF est un détecteur de piéton rapide, qui contient 10 canaux dont le premier canal est la valeur du gradient, 6 canaux des orientations HOG où chaque orientation représente un canal, ainsi que les 3 canaux de couleur LUV. Ensuite l'apprentissage se fait par le détecteur AdaBoost, comme illustré sur

la figure 1-6.

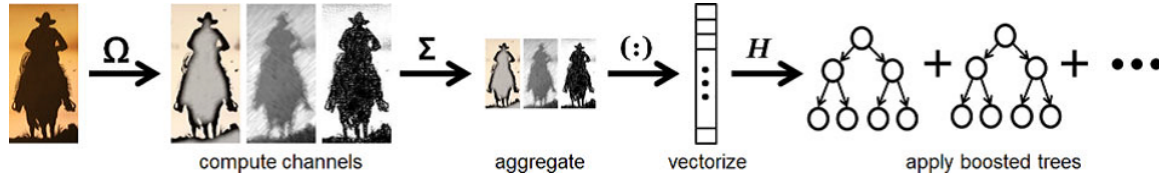


FIGURE 1-6 – Représentation des étapes du détecteur ACF [36]

Avant le calcul des 10 canaux, l'image est lissée en utilisant un filtre  $[1 \ 2 \ 1]/4$ , ensuite les canaux sont divisés en des blocs de  $4 \times 4$  et les pixels de chaque bloc sont agrégés. Le modèle contient 4 étapes d'apprentissage, la cascade d'Adaboost est utilisée. La première étape contient 32 classifieurs ensuite la 2ème contient 128 puis 512 et enfin 2048 classifieurs. La méthode de fenêtre glissante de taille  $[64 \ 32]$  est utilisée pour détecter la présence des piétons avec un pas de 4.

### 1.3.4 Méthodes basées sur l'apprentissage automatique

Ces dernières années, les réseaux de neurones convolutifs profonds (CNN) ont obtenu un grand succès sur la vision par ordinateur [150, 104, 105, 106, 83, 34]. Avec le succès de l'apprentissage profond en détection d'objets, plusieurs tentatives ont été faites pour appliquer des caractéristiques CNN approfondies à la détection des piétons. Les techniques d'apprentissage profond ont permis d'améliorer la précision de détection par rapport aux méthodes classiques. Plusieurs méthodes ont prouvé leur efficacité pour la détection de piéton en terme de précision et de vitesse de détection, notamment les Réseaux de Neurones Convolutifs basés sur les Régions (RCNN) [54].

#### Réseaux de Neurones Convolutifs basés sur les Régions (RCNN)

La différence entre les algorithmes de détection d'objets et les algorithmes de classification, est que dans les algorithmes de détection nous avons besoin de dessiner un cadre englobant pour chaque objet afin de le localiser sur l'image. Une des difficultés de la détection est que le nombre de cadres que l'algorithme doit dessiner peut varier en fonction du nombre d'objets dans l'image qui est inconnu au préalable.

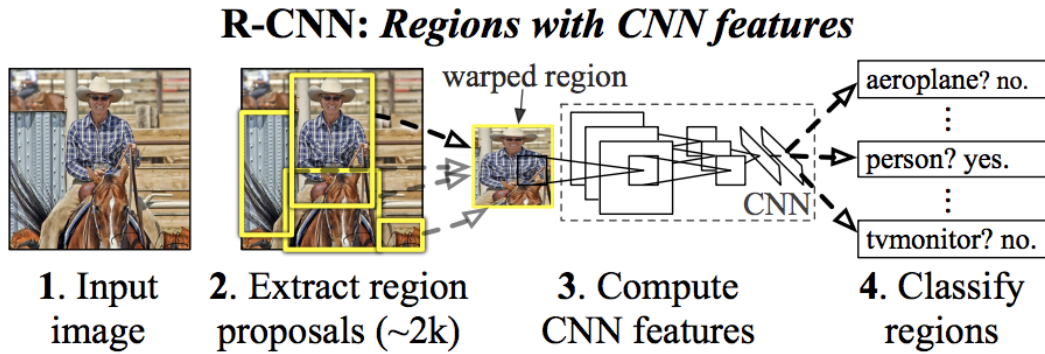


FIGURE 1-7 – Architecture du réseau de neurone convolutif basé sur les régions [54]

Une des solutions proposées est de prendre aléatoirement différentes régions d'intérêt (RoI) d'une image et d'entraîner par la suite le réseau CNN pour détecter la présence des objets dans ces régions. Pour procéder avec cette solution, un grand nombre de régions est sélectionné dans une image, afin de traiter tous les emplacements qui peuvent contenir un objet, ce qui peut être coûteux en terme de calcul. De ce fait, des algorithmes tels que RCNN, YOLO ont pu résoudre ce problème.

Ross et al. [54] ont résolu ce problème en proposant la méthode RCNN utilisant un algorithme de recherche sélective pour extraire 2000 régions de l'image appelées les régions proposées.

Ces 2000 régions proposées sont par la suite introduites dans un réseau de neurones convolutif (CNN) qui produit un vecteur de caractéristiques d'une dimension de 4096 comme illustré sur la Figure 1-7. Le CNN représente un extracteur de caractéristiques où la couche dense se compose des caractéristiques extraites de l'image qui sont introduites dans un SVM pour détecter la présence d'un objet dans une région proposée, comme illustré sur la Figure 1-8. L'algorithme prédit aussi quatre valeurs qui sont des valeurs de décalage pour augmenter la précision des cadres englobants. A titre d'exemple, si un piéton est détecté dans une région proposée et que son visage n'apparaît pas en entier, ces valeurs aident à ajuster le cadre de la région proposée.

Malgré d'avoir borné le nombre de régions proposées dans 2000, le temps de traitement par image reste important. De même le RCNN ne peut être implémenté en

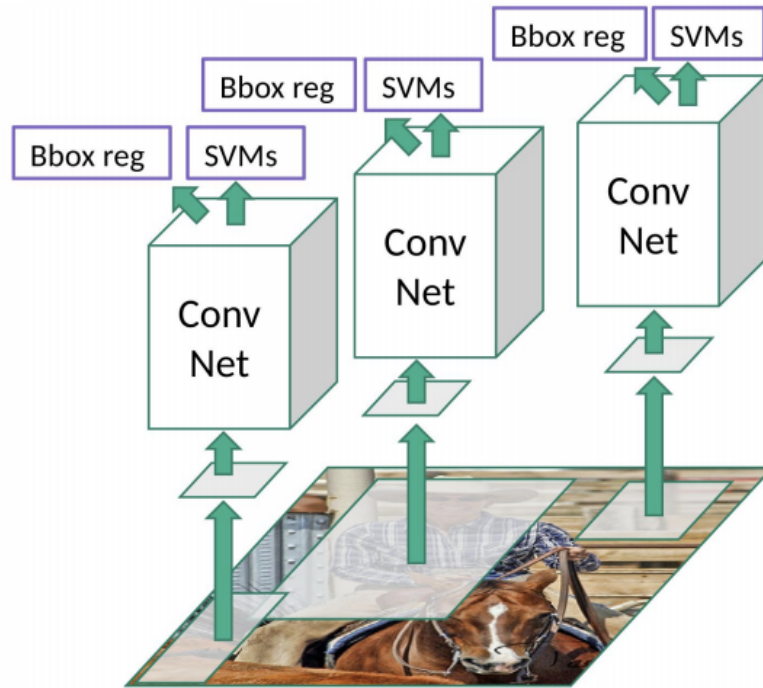


FIGURE 1-8 – Régulation des cadres englobants [53]

temps réel vu que chaque image de test prend environ 47 secondes de traitement. Comme solution le Fast-RCNN a été proposé.

## Fast-RCNN

Proposé par le même auteur de RCNN, le Fast-RCNN [52] résout les inconvénients du RCNN, permettant ainsi de construire un détecteur d'objet plus rapide. L'approche est similaire à l'algorithme RCNN, la différence se situe au niveau d'entrée du réseau CNN. Contrairement à RCNN, cette nouvelle approche introduit en entrée du CNN l'image d'entrée au lieu des régions proposées. Par la suite, le réseau CNN génère une carte de caractéristiques à partir de l'image d'entrée. A partir de cette carte, les régions proposées sont identifiées et redimensionnées à une taille fixe pour pouvoir l'introduire dans la couche dense en utilisant les régions d'intérêt. A partir du RoI, la couche Softmax est utilisée pour classer la région proposée et les valeurs du décalage du cadre englobant.

L'architecture du réseau Fast-RCNN est illustrée dans la Figure 1-9.

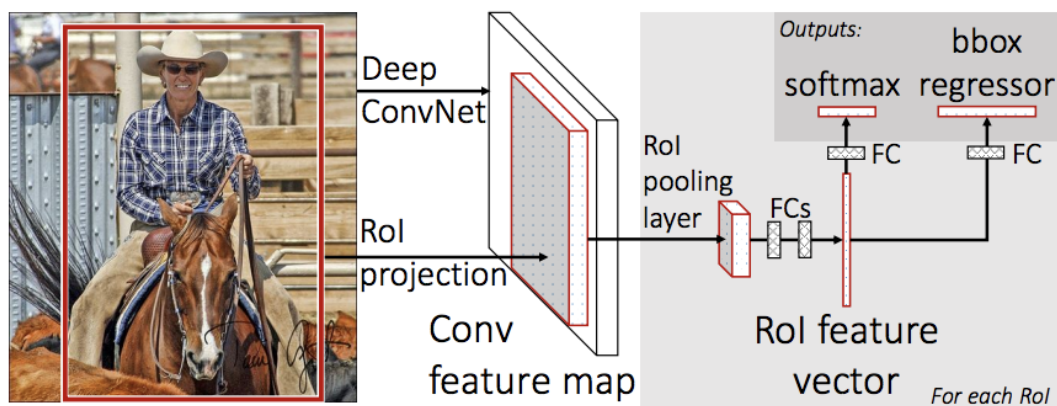


FIGURE 1-9 – Architecture du Fast-RCNN [52]

La raison pour laquelle le Fast R-CNN est plus rapide que le RCNN est que les 2000 régions ne vont pas être introduites pour chaque image en entrée du CNN. L'opération de convolution est faite seulement une fois par image et la carte des caractéristiques à extraire est générée à partir de cette opération. La Figure 1-10 représente une comparaison entre les algorithmes de détection en terme de vitesse de détection.

Nous pouvons conclure à partir de ces graphes que le Fast-RCNN est significativement plus rapide que le RCNN pour les phases d'apprentissage et de test. La vitesse de détection durant le test en utilisant les régions proposées est inférieure à celle sans régions proposées. De ce fait, l'utilisation des régions proposées affecte la performance du réseau Fast-RCNN.

### Faster R-CNN

Les deux architectures RCNN et Fast RCNN utilisent la technique de recherche sélective pour chercher les régions proposées. La recherche sélective est une technique très lente, ce qui affecte négativement le temps de calcul du réseau. Faster RCNN est un algorithme proposé par Shaoqing Ren et al. [114]. Ce dernier permet au réseau d'apprendre les régions seul au lieu d'utiliser la recherche sélective. De façon similaire au réseau RCNN, l'image d'entrée est fournie en entrée au réseau de convolution

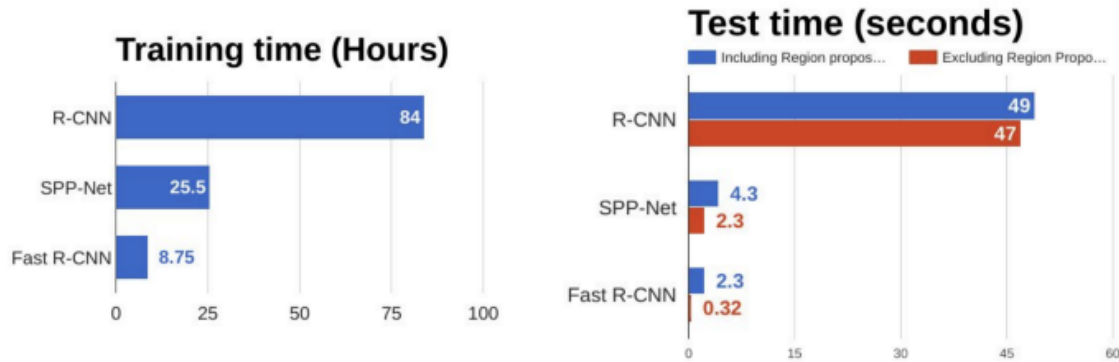


FIGURE 1-10 – Comparaison entre les algorithmes de détection en terme de vitesse [43]

qui produit une carte de caractéristiques. Au lieu d'utiliser la recherche sélective sur la carte des caractéristiques pour identifier les régions proposées, le Faster-RCNN utilise un réseau séparé pour préparer les régions proposées. La région prédite est par la suite redimensionnée en utilisant le RoI. Faster-RCNN a été largement utilisé pour la détection des piétons [152, 155, 151, 79, 69].

La figure 1-11 illustre l'architecture du Faster-RCNN.

Le graphe 1-12 illustre la vitesse de Faster-RCNN par rapport à ses prédécesseurs. De ce fait, il peut être utilisé pour les applications en temps réel.

## You Only Look Once (YOLO)

Les algorithmes de détection d'objet que nous venons de citer précédemment, utilisent les régions pour localiser un objet dans une image. Par conséquent, le réseau ne traite pas l'image entière, mais seulement la partie de l'image qui possède une forte probabilité de contenir un objet. L'algorithme de détection You Only Look Once (YOLO) proposé par Redmon et al. [113] est différent des algorithmes basés sur les régions. Dans YOLO un seul réseau convolutif prédit les cadres délimitant ainsi que la probabilité de classe pour ces cadres. La Figure 1-13 représente le système de détection YOLO en trois étapes, la première consiste à redimensionner l'image d'entrée en 448x448, cette image est par la suite fournie en entrée du réseau convolutif,

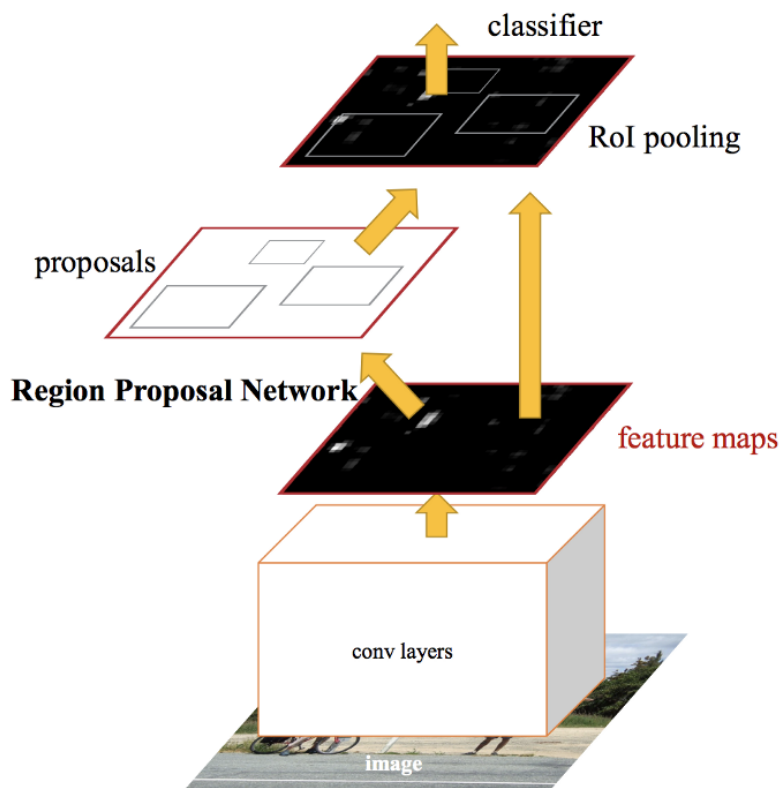


FIGURE 1-11 – Architecture de l'approche Faster RCNN [114]

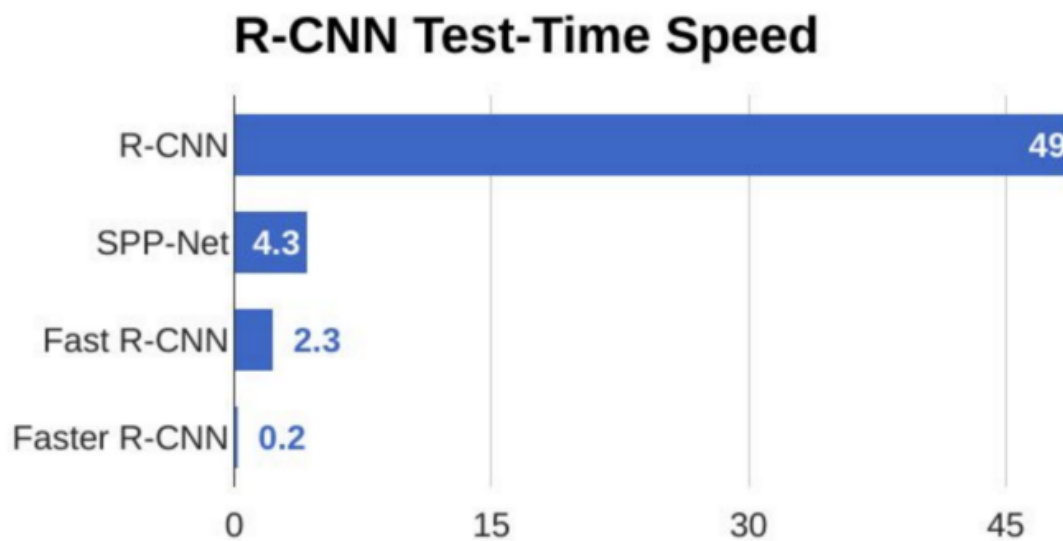


FIGURE 1-12 – Vitesse de Faster-RCNN par rapport à ces prédécesseurs dans la phase de test [43]

et finalement la Non Max Suppression (NMS) est utilisée pour ne garder qu'un seul cadre par détection.

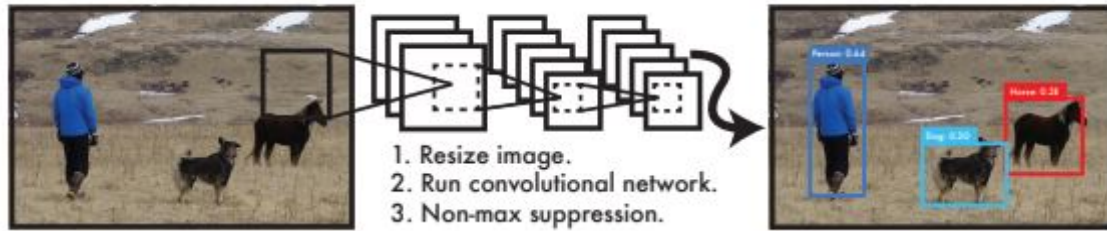


FIGURE 1-13 – Les étapes du système de détection YOLO [113]

L'algorithme YOLO découpe l'image d'entrée en grilles de dimension  $S \times S$ , où chaque grille contient  $m$  cadre englobants. Pour chacun des cadres englobants, le réseau donne en sortie une probabilité de la classe de son appartenance. Les cadres englobants ayant la probabilité d'une classe supérieure à un certain seuil sont sélectionnés et utilisés pour localiser l'objet dans l'image. Ce processus est illustré sur la figure 1-14

Le réseau de détection YOLO contient 24 couches de convolution suivi par 2 couches denses. Les couches de convolution réduisent la dimension des caractéristiques des couches précédentes. Ces couches de convolution sont pré-entraînées sur la base ImageNet avec une résolution de  $224 \times 224$  pour l'image d'entrée, ensuite la dimension est doublée pour la phase de détection. L'architecture est représentée par la Figure 1-15.

YOLO a connu par la suite plusieurs variantes notamment la version YOLOv3 [42]. Cette dernière est une amélioration de la version originale de YOLO, elle contient 53 couches de convolution pour la classification (figure 1-16). YOLOv3 est aussi performant que d'autres détecteurs de l'état de l'art tel que RetinaNet, tout en étant considérablement plus rapide.

YOLO est l'algorithme de détection d'objet le plus rapide parmi ceux que nous venons de citer traitant 45 images par seconde, ce qui le rend adéquat pour les applications en temps réel. Pour cette raison, nous avons choisi dans le cadre de cette



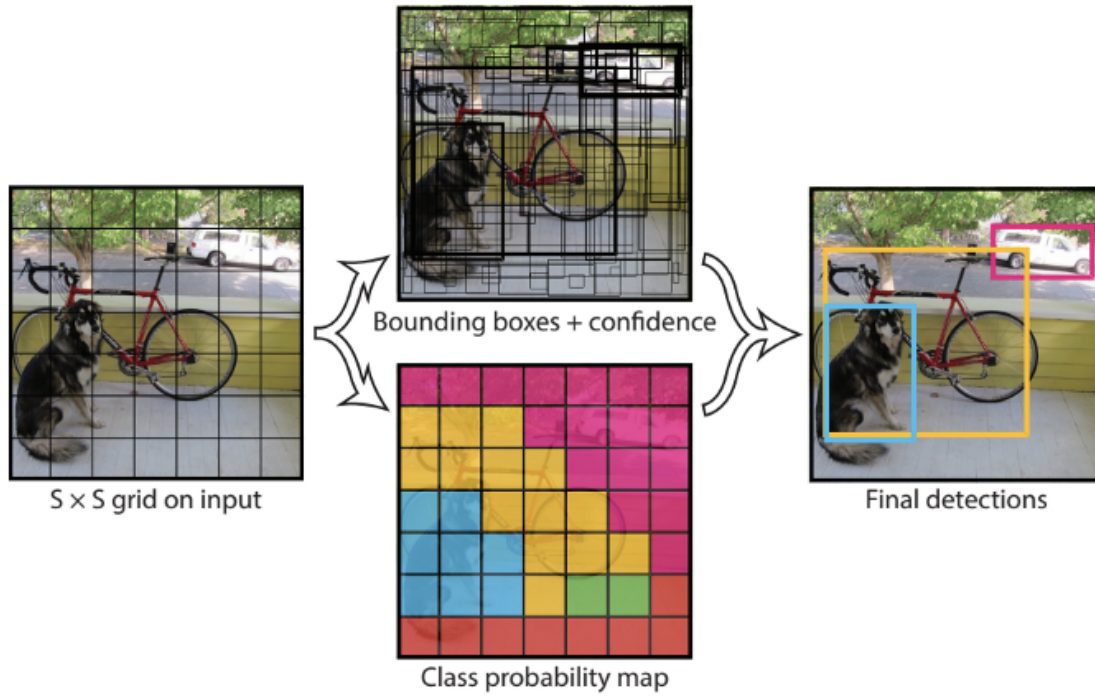


FIGURE 1-14 – Le modèle YOLO [113]

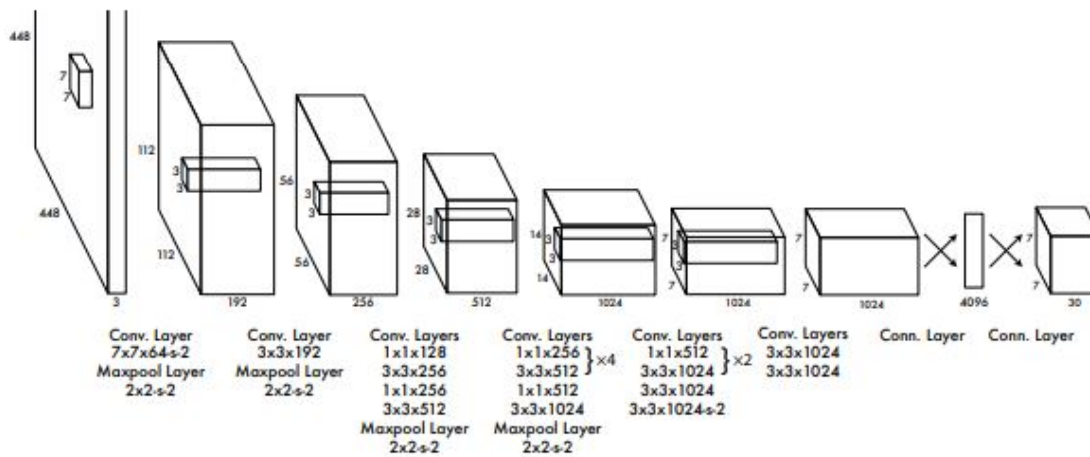


FIGURE 1-15 – L'architecture de YOLO [113]

thèse de travailler en utilisant YOLOv3 pour la détection des piétons [75, 154, 64, 109, 146, 4, 22].





# Chapitre 2

## Proposition d'un système de prévention des accidents Piétons—Véhicules

### 2.1 Introduction

Dans ce chapitre, nous allons présenter une étude conçue pour améliorer le comportement des conducteurs et assurer la sécurité des piétons. Afin de réduire les accidents mortels impliquant des piétons, de nombreuses recherches sont mises en oeuvre sur les systèmes d'aide à la conduite avancés (ADAS) destinés à la protection des piétons "Systèmes de préventions des collisions avec piétons (PCAS)" [49, 116, 50, 23, 67, 11]. À cette fin, les recherches dans ces systèmes sont davantage axées sur la détection des piétons que sur la prédiction de leur direction de marche. Par conséquent, rares sont ceux qui ont l'intention d'inclure l'orientation des piétons dans ces systèmes. De plus, les systèmes existants sont conçus pour des zones bien structurées contenant des panneaux de signalisation et des marquages au sol, mais cela n'est pas toujours le cas sur nos routes.

Dans ce travail, nous avons sélectionné le Maroc comme étude de cas, où presque 28% des victimes d'accidents mortels sont des piétons, comme indiqué par le Ministère

Marocain des Transports (METLE) en 2017 [90]. Dans ce contexte, nous avons collecté une nouvelle base de données contenant des piétons marocains prise dans différentes villes Marocaines. Après analyse de ces données, nous constatons que le code de la route est non respecté par les conducteurs et les piétons. Ces derniers partagent la chaussée avec les véhicules, en particulier dans les zones mal structurées. A cause des mauvais comportements, un taux élevé d'accidents de piétons est remarqué au Maroc.

Notre objectif est de protéger l'utilisateur le plus vulnérable de la route qui est le piéton, principalement lorsque l'infrastructure ne dispose pas des passages protégés, menant ainsi les piétons à traverser d'une façon non convenable.

Bien que les ADAS connaissent un succès considérable au niveau de la prévention des accidents, ils n'ont pu réduire le taux d'accidents mortels liés aux piétons, notamment dans les pays en développement. En Mars 2018, un véhicule autonome Uber a tué un piéton en Arizona aux Etats Unis, devenant ainsi le premier véhicule autonome à tuer un piéton. L'investigation du National Transportation Safety Board résumé dans le rapport [97] a révélé que le véhicule n'a pas réussi à détecter la victime correctement, car il n'était pas programmé pour prédire le mouvement des personnes qui traversent en dehors des clous.

La Figure 2-1 représente une image satellitaire prise par Google Earth quelques secondes avant l'accident, nous pouvons voir l'emplacement du véhicule et du piéton à partir de 4.2 secondes avant l'accident jusqu'à la position de l'accident. On peut remarquer aussi d'après la figure que le piéton traversait de façon diagonale en dehors du passage clouté.

D'après les résultats de l'investigation sur l'accident on peut conclure que les ADAS et les véhicules autonomes sont conçus en première partie pour les piétons bien disciplinés traversant des routes bien structurées disposant des marquages et des signalisations, chose qui n'est pas toujours valide.

Dans ce chapitre, nous proposons un nouveau système de prévention intégrant l'orientation du piéton et permettant de détecter un accident imminent indépendamment des installations piétonnes.

Dans ce qui suit, nous allons étudier de près le conflit piéton-conducteur afin de

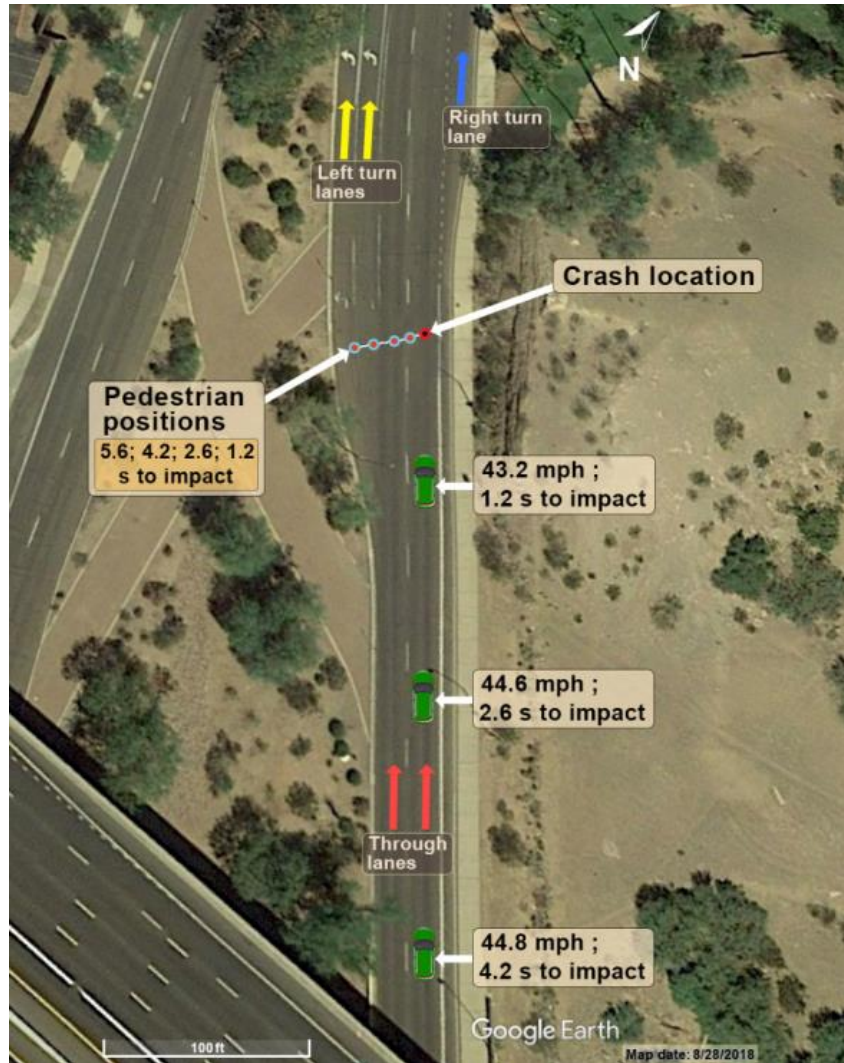


FIGURE 2-1 – Vue aérienne montrant le lieu de l'accident, le chemin du piéton et du véhicule ainsi que la vitesse du véhicule à partir de 5.6 seconde avant l'accident [97]

mieux comprendre les causes menant à une collision impliquant ces deux usagers de routes. Pour cela, nous allons aborder tout d'abord les études réalisées sur le comportement du conducteur [30] ainsi que celui du piéton et leurs différences selon l'âge et le genre, avant d'adresser l'interaction et la communication entre le piéton et le conducteur.

## 2.2 Conflit Piéton-Conducteur

### 2.2.1 Comportement du conducteur

#### Comportement du conducteur (selon l'âge et le genre)

Au Maroc, selon le recueil des statistiques des accidents corporels de la circulation routière pour l'année de 2017 [90], le sexe masculin dominait le pourcentage de victimes d'accidents corporels avec 78.52% contre seulement 20.84% pour le sexe féminin. Côté âge, plus de 46.11% des tués ont moins de 34 ans en particulier celle entre 25 et 34 ans représente la tranche d'âge la plus touchée par la mortalité routière avec plus de 20.61% de l'ensemble des tués.

En France, selon le bilan de l'accidentalité de l'année 2018 [98], les personnes âgées entre 20-24 ans représentent le plus grand nombre de tués et blessés, suivi par les 15-19 ans et les 25-29 ans comme représentent les Figures 2-2 et 2-3. En terme de pourcentage, les jeunes entre 15 et 24 ans représentent 19.2% de la mortalité routière pour 11.8% de la population. Ce qui rend cette catégorie la plus touchée par les accidents routiers et la plus vulnérable notamment les 18-24 ans qui décèdent majoritairement avec un pourcentage de 62% lors d'un accident routier en véhicule de tourisme.

Côté genre, la proportion d'hommes tués sur la route est plus élevée de trois fois que les femmes selon les statistiques de 2018, notamment pour les classes de 15-29 ans, 30-44 ans et 45-59 ans comportant 83% de mortalité masculine dont 79% sont des conducteurs, représenté par la figure 2-4 par classe d'âge et de sexe. Les accidents chez les conducteurs masculins sont majoritairement dû à l'excès de vitesse (29%), l'alcool (20%) et le non-respect des priorités (10%), représentant ainsi 85% des conducteurs tués. Tandis que les conductrices représentent 15% des tués avec pour causes principales la vitesse excessive avec 19%, le non-respect des priorités avec 14% et l'inattention avec 13%. Le bilan a également cité que les piétons représentent 25% des tués dans les territoires urbains.

Selon l'Organisation Mondiale de la Santé (OMS) [140], les accidents de la route



### Personnes tuées par classe d'âge en 2018

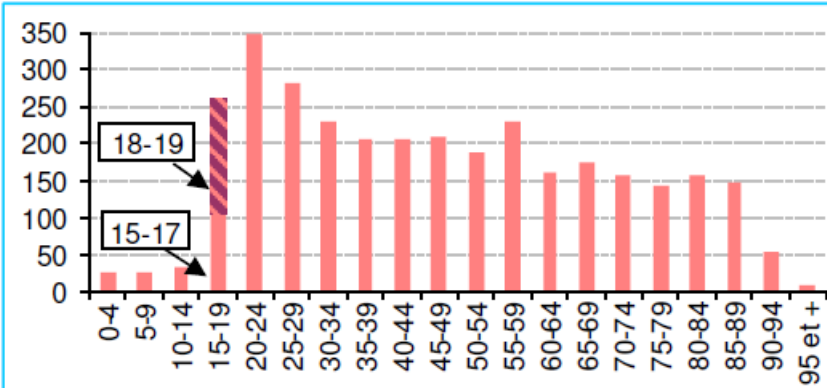


FIGURE 2-2 – Personnes tuées par classe d'âge en 2018 en France [98]

### Personnes blessées par classe d'âge en 2018

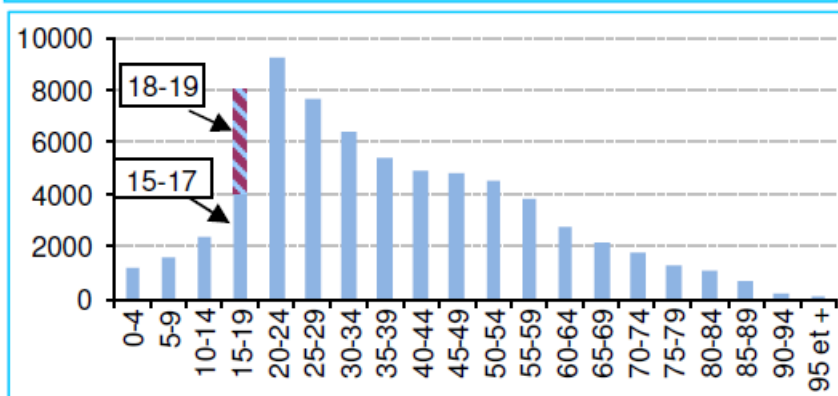


FIGURE 2-3 – Personnes blessées par classe d'âge en 2018 en France [98]

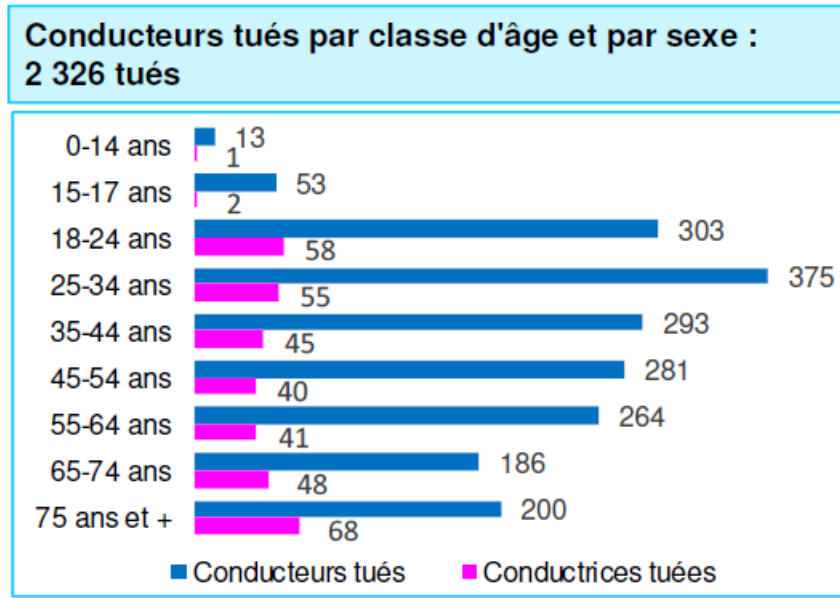


FIGURE 2-4 – Conducteurs tués par classe d'âge et par sexe en France [98]

sont la cause principale des décès pour les jeunes de moins de 30 ans. Ce qui prouve que les jeunes conducteurs sont les plus impliqués dans les accidents les plus dangereux. Les recherches de Lancaster [76] ont montré que les jeunes conducteurs étaient en outre les plus impliqués dans les accidents liés à l'ivresse, l'utilisation de drogues ainsi que le non-port de la ceinture de sécurité. Ceci peut s'expliquer par leur tendance à surestimer leur capacité de conduire associée à leur manque d'expérience de conduite. Le taux d'accidents est également élevé pour les conducteurs les plus âgés bien qu'il soit inférieur à celui des jeunes, ceci est majoritairement dû à leur faiblesse visuelle.

Il a été montré que les conducteurs masculins ont tendance à avoir des attitudes moins sûres que les conductrices, selon l'OMS les hommes sont impliqués 3 fois plus que les femmes dans les accidents routiers, ce qui peut être expliqué par la prise de risque dans le trafic qui est beaucoup plus courant chez les hommes que les femmes. Les recherches de Harris et Harrison [128] ont démontré que parmi les jeunes conducteurs, les hommes possèdent un comportement de conduite bien plus risqué que celui des femmes, avec des taux plus élevés de vitesse et de conduite en état d'ivresse. Cette différence de comportement est expliquée sur [128] par la différence de stéréotypes sexuels entre la féminité et la masculinité, cette dernière est généralement associée à

la prise de risque.

Dans ce même principe, Mahalik et al. [87] affirment qu'un niveau élevé de masculinité est associé à un niveau élevé de prise de risque, d'agressivité ainsi qu'un niveau élevé d'accidents. Alors, qu'un niveau élevé de féminité est associé à moins de violation et d'accidents.

### 2.2.2 Comportement du piéton

Malgré son importance, le comportement du piéton reçoit moins d'attention dans la littérature par rapport à celui du conducteur. Toutefois, la compréhension du comportement du piéton et sa manière de traverser peuvent s'avérer importantes pour améliorer la planification des routes, et par conséquent l'amélioration de la sécurité de ce dernier.

Parmi les chercheurs qui se sont intéressés à l'analyse en profondeur du comportement du piéton, nous trouvons Oxley et al. [100] qui ont examiné la traversée des piétons âgés sur les passages à niveau, en mesurant plusieurs indicateurs tels que le temps de traversée, le temps d'arrivée, la marge de sécurité minimale, etc. Une étude relative faite par Bernhoft Carstensen [16], a montré que les piétons âgés apprécient les passages piétons et les installations piétonnes beaucoup plus que les jeunes piétons. Rosenbloom et al. [117], ont utilisé une méthode similaire pour étudier le comportement des enfants et ont trouvé que le comportement le plus dangereux et le plus courant chez les enfants est de traverser sans regarder la route, suivi par le fait de ne pas s'arrêter avant de traverser. Les résultats ont également montré que les enfants accompagnés par un adulte se comporte plus dangereusement, surtout quand l'adulte accompagnant ne leur tient pas la main. Sisiopiku Akin [127] ont défini la manière de traverser des piétons par rapport à leur emplacement (SCCR), et à la présence des feux de circulation (TCCR). Les résultats ont prouvé que l'emplacement du passage piéton par rapport au point de départ et de destination du piéton est le facteur le plus influent sur sa façon de traverser.

Les recherches de Papadimitriou et al. [102], sur les facteurs humains du comportement du piéton et de sa traversée, ont démontré selon leur sondage que 20 à

35% des piétons préfèrent traverser diagonalement et qu'ils sont prêts à saisir toute opportunité pour traverser rapidement y compris les manœuvres dangereuses. Cependant, la plupart des piétons réalisent que traverser en dehors de l'espace désigné est une attitude risquée, en reconnaissant qu'elle leur permet de gagner du temps. Selon un questionnaire d'auto-évaluation, la majorité des piétons affirment qu'ils se sentent plus prudents et plus rapides que les autres piétons et qu'ils sont les moins susceptibles d'être impliqués dans un accident de route par rapport aux autres. Ce qui laisse entendre que chaque piéton a la conviction de traverser en vigilance malgré que sa manière soit dangereuse, et que les accidents ne surviennent qu'aux piétons les moins prudents. Seulement 43% des piétons ont signalé n'avoir jamais traversé au milieu de la route, en revanche la majorité confirme le faire entre les véhicules arrêtés en cas de congestion routière ou si le trafic est vide. Il est important de signaler aussi que le comportement des piétons change avec le changement des conditions de trafic et du type de route : seulement 10% qui ont reconnu avoir traversé au milieu d'une route urbaine, alors que 64% le font sur les petites ruelles et les routes résidentielles. Finalement les piétons participant aux sondages ont tendance à s'accorder à avoir des avis négatifs sur les conducteurs : seulement 13% contredisent que les conducteurs ne respectent pas les piétons et seulement 7% admettent que les conducteurs ne sont pas agressifs et insoucians.

Au Maroc, les personnes âgées de plus de 65 ans représentent le plus grand pourcentage de décès spécialement pour le sexe masculin comme le démontre la figure 2-5, suivi par les 35-44 ans et les 15-24 ans. Par contre, on peut remarquer d'après cette figure que les jeunes entre 15-24 ans représentent la plus grande partie des blessés au niveau des piétons suivie par les enfants entre 5 et 9 ans. Cependant, pour les deux histogrammes on peut voir la domination du genre masculin au niveau des tués et blessés et ce pour toutes les tranches d'âge.

L'étude effectuée par Liu et al. [84] a montré que la vitesse de marche des piétons âgés de plus de 65 ans est moins rapide comparée à celle des jeunes entre 21-30 ans, pour 1.17m/s et 1.49m/s respectivement. Néanmoins, les sujets âgés n'ont pas réalisé que leur capacité de marche a diminué et jugent subjectivement que leur temps pour

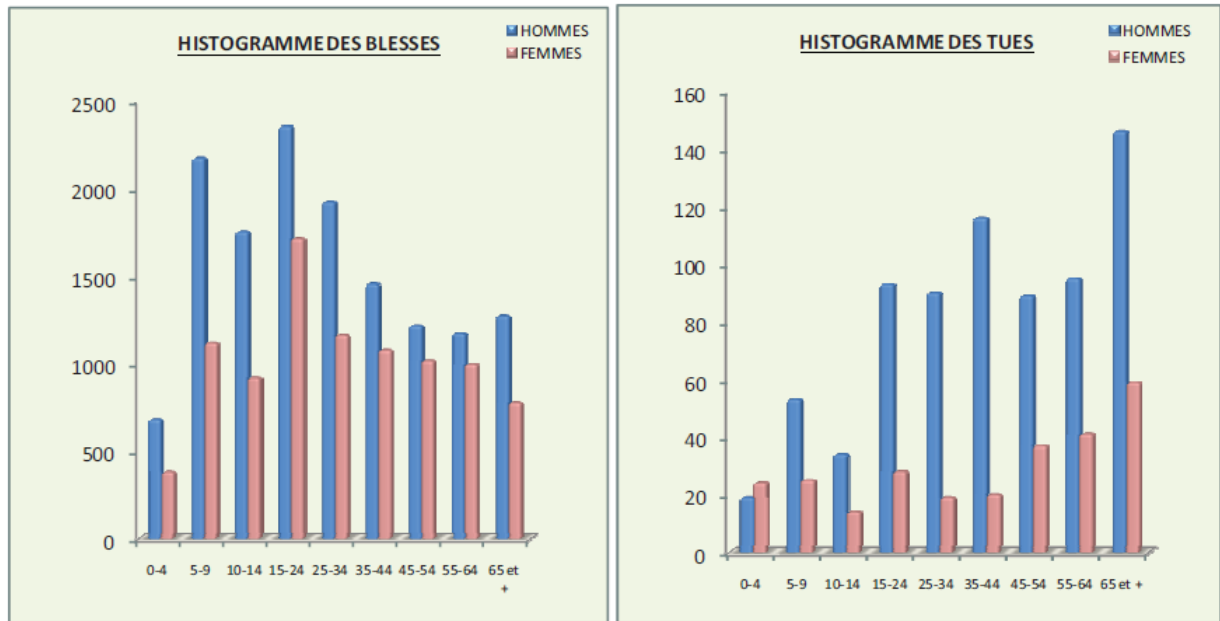


FIGURE 2-5 – Histogramme des piétons blessés et tués au Maroc par genre et par âge [90]

traverser la route est toujours le même que ceux des sujets plus jeunes, ce qui peut expliquer le haut pourcentage d'accidents pour les piétons âgés.

Ferrenchak dans son étude [44] a comparé le comportement de traversée des piétons selon le genre en Inde. Comme pour le cas du conducteur, le genre féminin a fait preuve de prudence pour les piétons aussi. Son étude a montré que le genre féminin possède un temps d'attente 88% plus long que celui du genre masculin, avec un temps d'attente de 34.4s pour les femmes et de 18.3s pour les hommes. En terme d'usage du passage piéton, 53.2% du genre féminin l'utilisent pour 39.8% du genre masculin.

### 2.2.3 Interaction piéton-conducteur

Fuller [47] a décrit l'interaction dynamique entre le piéton et le véhicule en Europe par un modèle d'évitement des menaces. Selon ce modèle, le conducteur est concentré sur l'évitement des stimuli indésirables dans le trafic. La présence d'un piéton est un des stimuli indésirables que le conducteur essaye d'éviter. Par conséquent, il peut soit décider de ralentir le véhicule pour laisser passer le piéton ou bien choisir de

garder la même vitesse communiquant ainsi au piéton qu’il ne va pas lui céder la priorité. Dans le premier cas, le conducteur est sanctionné par la perte de temps, alors que dans le second le conducteur est récompensé car il n’a pas subi de retard. Néanmoins, dans ce dernier cas, le piéton peut assumer un « comportement compétitif » en voulant traverser la route malgré la réaction du conducteur. Dans cette situation, le conducteur doit faire face à une réponse retardée de freinage rapide pour éviter la collision.

Parmi les facteurs influençant le conducteur à céder le passage au piéton, nous trouvons la densité du trafic, le nombre de piétons voulant traverser ainsi que la distance du piéton du bord du trottoir. Le conducteur cède dans la majorité des cas le passage si le nombre des piétons voulant traverser est supérieur à 3, alors qu’il ne donne pas la priorité au piéton quand la densité du trafic est élevée, et quand le piéton s’arrête à distance de plus d’un demi mètre du bord du trottoir.

D’après cette étude, nous pouvons constater qu’il y a un rapport de force dans la façon de communiquer entre le piéton et le conducteur, ce qui conduit à une incertitude sur le fait que le conducteur va céder le passage au piéton ou non.

L’interaction piéton-véhicule est aussi souvent décrite en utilisant le paramètre Time-To-Zebra ( $TTZ_{arr}$ ) proposé par Varhelyi [136], ce paramètre est défini comme étant le temps restant au conducteur pour arriver au passage piéton au moment où ce dernier arrive au bord du trottoir.

Elle est obtenue en calculant le rapport entre la distance du véhicule vis-à-vis du passage piéton, et la vitesse du véhicule quand le piéton arrive au bord du trottoir. Varhelyi, observe que la vitesse du conducteur en approchant le passage du piéton dépend de l’emplacement du piéton par rapport au bord du trottoir relativement au temps prévu pour que le conducteur arrive au passage.

Bella et Silvestri [13] ont utilisé une approche similaire à celle de Varhelyi, en classifiant l’interaction piéton-conducteur en terme de  $TTZ_{arr}$ . Selon cette classification, les faibles valeurs de  $TTZ_{arr}$  correspondent à un comportement agressif, et en contrepartie les valeurs élevées correspondent à un comportement plus prudent.

Dans leur étude, Bella et al. [12] ont étudié le comportement du conducteur face au

passage piéton sur trois différents types de routes (urbaine, suburbaine et rurale) avec la présence et l'absence d'un ADAS en utilisant un simulateur de conduite. Sachant que la vitesse limite sur les zones urbaines et suburbaines s'élève à 50km/h et 60km/h respectivement alors que la sur la zone rurale s'élève à 90km/h, trois scénarios ont été présentés lors de l'étude : un passage piéton sans la présence d'un piéton, un passage avec la présence d'un piéton sans ADAS et un passage avec la présence d'un piéton et d'un ADAS sur le véhicule.

Le piéton marchant à une vitesse de 1.4m/s a été programmé pour commencer à traverser le passage du côté droit du conducteur, quand le véhicule est à une distance de 55.6m du passage sur la zone urbaine, 66.6m sur la zone suburbaine et 88.8m sur la zone rurale, pour une vitesse du conducteur de 50km/h, 60km/h et 80km/h pour respectivement les zones urbaines, suburbaines et rurales. Ces distances représentent une valeur  $TTZ_{arr}$  de 4 secondes. Les scénarios contenant un système ADAS produisent une alarme visuelle et auditive lors de la présence d'un piéton. Le déclenchement du système ADAS se produit lorsque le temps de collision est de 3.5 secondes donnant théoriquement une distance de 49m, 59m et 78m du passage piéton pour les trois zones, si la vitesse de rapprochement du véhicule est de 50, 60 et 80km/h respectivement. L'étude a intégré 41 participants (21 hommes et 20 femmes). La vitesse des conducteurs est analysée à partir de 150m du passage piéton, afin d'obtenir les variables permettant de décrire le comportement de vitesse du conducteur en approchant le passage piéton et sa volonté de céder le passage, et ce en calculant le  $TTZ_{arr}$  (eq. 2.1).

Les paramètres permettant de la calculer sont :

- $V_i$  : Vitesse initiale du conducteur quand le conducteur perçoit le passage piéton et diminue sa vitesse en relâchant la pédale de l'accélérateur.
- $L_i$  : Distance entre le véhicule et le passage, au moment où  $V_i$  est enregistrée.

$$TTZ_{arr} = \frac{L_i}{V_i} \quad (2.1)$$

Le temps au passage piéton représenté par  $TTZ_{arr}$  donne une information im-

portante sur le comportement du conducteur, une faible valeur  $TTZ_{arr}$  indique un conducteur caractérisé comme étant agressif avec une faible volonté de céder le passage, alors qu'une valeur élevée du  $TTZ_{arr}$  indique un comportement plus vigilant de la part du conducteur.

Trois intervalles sont considérés :  $TTZ_{arr} < 4s$ ,  $4s < TTZ_{arr} < 6s$  et  $TTZ_{arr} > 6s$ .

La première analyse, se concentre sur l'effet de l'ADAS sur la vitesse du conducteur pour les différents intervalles de  $TTZ_{arr}$  peu importe le type de routes. Pour chaque intervalle de  $TTZ_{arr}$  la vitesse moyenne des conducteurs est illustrée en présence d'un ADAS sur la figure 2-6 et en son absence sur la figure 2-7.

- Pour  $TTZ_{arr} < 4s$  indique une conduite avec une vitesse élevée par rapport aux valeurs plus élevées de  $TTZ_{arr}$ . En l'absence d'un ADAS, les conducteurs commencent par ralentir aux environs de 55m avant le passage piéton significativement inférieure à la distance où les conducteurs commencent à ralentir avec la présence d'un ADAS, qui est de 70m avant le passage piéton. D'autant plus, lorsque le message d'alerte est fourni, les conducteurs atteignent la vitesse minimale plus rapidement que lorsque l'alerte n'est pas fournie, avec une distance de 12m sans ADAS et 18m avec ADAS.
- Pour  $4s < TTZ_{arr} < 6s$ , la vitesse des conducteurs était moins élevée que celle de  $TTZ_{arr} < 4s$ , le conducteur commence à diminuer la vitesse du véhicule approximativement 100m avant le passage piéton avec et sans ADAS, et atteint une vitesse minimale de 32km/h aux environs de 25m avant le passage. Dans ce cas, la présence d'un ADAS n'a eu aucun effet sur le comportement du conducteur en le comparant avec le premier scénario.
- Pour  $TTZ_{arr} > 6s$ , la vitesse du conducteur est la plus faible par rapport aux deux autres scénarios. Le conducteur atteint une vitesse minimale de 25km/h pour approximativement 25m avant le passage piéton. En conséquence, la présence d'un ADAS n'avait aucun effet sur le comportement du conducteur pour ce scénario.

Finalement, l'effet des ADAS a été étudié aussi pour chaque type de route (urbaine, suburbaine et rurale). Sur la zone urbaine et avec la présence d'un ADAS, le



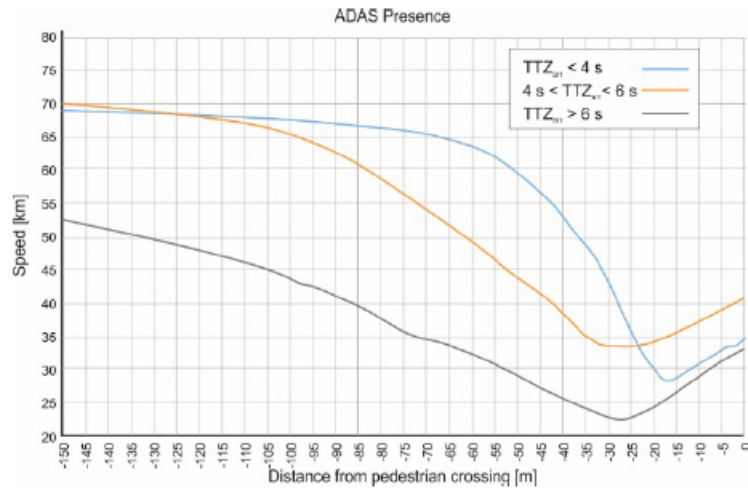


FIGURE 2-6 – Vitesse moyenne des conducteurs par rapport au passage piéton avec la présence d'un ADAS [12]

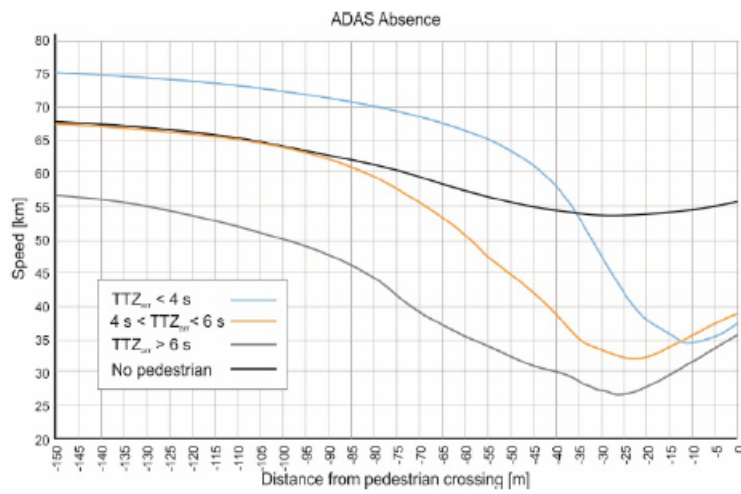


FIGURE 2-7 – Vitesse moyenne des conducteurs par rapport au passage piéton sans la présence d'un ADAS [12]

conducteur commence à ralentir directement après la réception de l’alerte (entre 50 et 55m), alors que dans le cas d’absence d’un ADAS le ralentissement commence après un retard de 10m, c.-à-d. approximativement 40m avant le passage piéton. Le même effet a été remarqué pour les deux autres zones mais en raison de la vitesse plus élevée des conducteurs sur ses deux zones, la distance d’alerte avant le passage piéton était plus grande, dans les environs de 66.6m pour la zone suburbaine et 78m pour la zone rurale.

De ce fait, le comportement du conducteur peut être affecté par la présence et l’absence d’un ADAS, spécialement pour le conducteur agressif ayant un  $TTZ_{arr} < 4s$ , alors que le système n’est pas si impactant pour les conducteurs prudents ayant un  $4s < TTZ_{arr} < 6s$  et  $TTZ_{arr} > 6s$ .

## 2.3 Système de prévention de collision avec piétons

Le système de prévention de collision avec les piétons, connu sous le nom de Pedestrian Collision Avoidance System (PCAS) ou Pedestrian Collision Avoidance Mitigation (PCAM), est un type particulier d’ADAS permettant de protéger l’utilisateur le plus vulnérable de la route qui est le piéton. Il prédit les accidents piétons-véhicules et alerte ensuite le conducteur pour freiner le véhicule. Dans le cas où l’accident ne peut pas être évité, ce système intégré au sein du véhicule peut être utilisé pour réduire l’impact de l’accident sur le piéton.

L’Organisation mondiale de la Santé (OMS) dans son rapport [140], a décrit que plus de la moitié des décès au niveau mondial des accidents de routes sont des piétons, cyclistes et motocyclistes où les piétons représentent à eux seuls 23% comme exposé sur la Figure 2-8.

Selon le même rapport de l’OMS, l’Afrique contient la plus grande proportion de tués des piétons et cyclistes avec 44% de décès au niveau mondial, ceci est majoritairement dû à la vaste présence de piétons et de cyclistes qui partagent la route avec les véhicules. L’Afrique représente aussi le taux le plus élevé d’accidents au monde d’une valeur de 26.6 décès pour 100.000 habitants en 2016, comme le montre la figure 2-9.

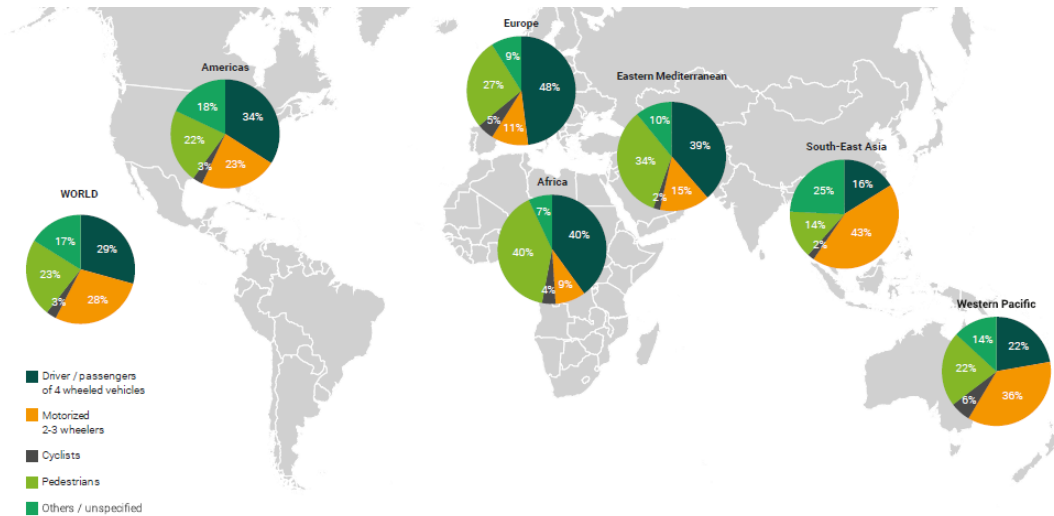


FIGURE 2-8 – Répartition des décès par type d'utilisateur de la route, par région (OMS) [140]

De ce fait, la sécurité des piétons est une motivation majeure tant pour la communauté scientifique que pour l'industrie automobile. Plusieurs organismes et projets scientifiques au niveau global ont été établis pour participer à la résolution du problème du piéton sur la route, notamment le programme de recherche sur la sécurité des piétons et des cyclistes élaboré par l'Administration Fédérale des Routes (FHWA) aux Etats-Unis. En Europe, la commission européenne a lancé des projets tels que Watch-over [108] et APROSYS [107] visant à protéger l'utilisateur le plus vulnérable des routes qui est le piéton en avisant le conducteur de sa présence et ce en utilisant les dernières technologies. Au Maroc également, le Ministère des Transports a lancé des projets visant à améliorer la sécurité routière au royaume y compris la sécurité des piétons.

Côté industrie, Mobileye [33] utilise un système permettant de détecter les piétons en utilisant une mono-caméra. Il alerte ensuite le conducteur en cas d'un éventuel accident, en utilisant une interface sur le tableau de bord comme le montre la figure 2-10.

La figure 2-11 représente les différentes étapes pour améliorer la sécurité des piétons, allant des mesures de renforcement de la sécurité à long terme telles que la conception des infrastructures et des véhicules, jusqu'à l'utilisation des systèmes per-

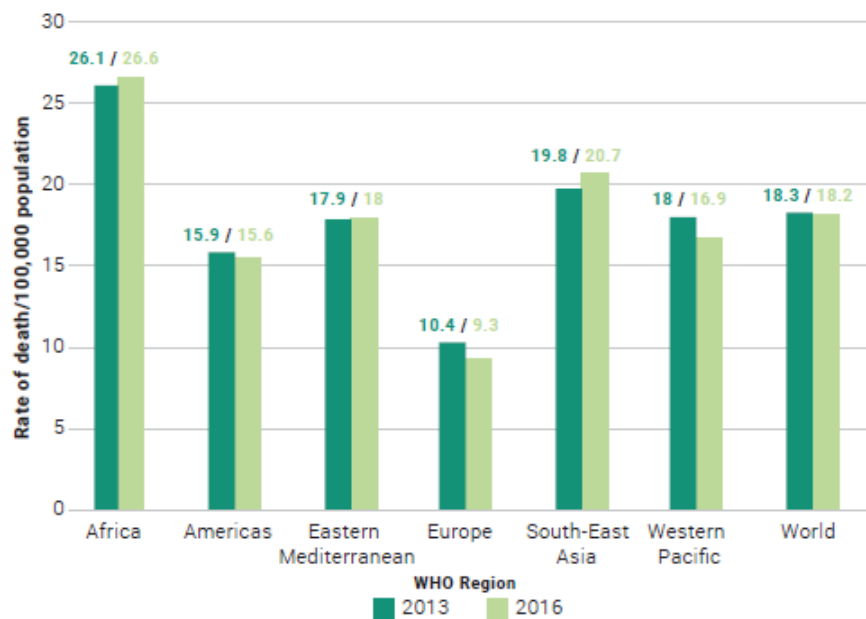


FIGURE 2-9 – Taux de décès sur le trafic routier en 2013 et 2016 par population de 100.000 pour chaque région [140]

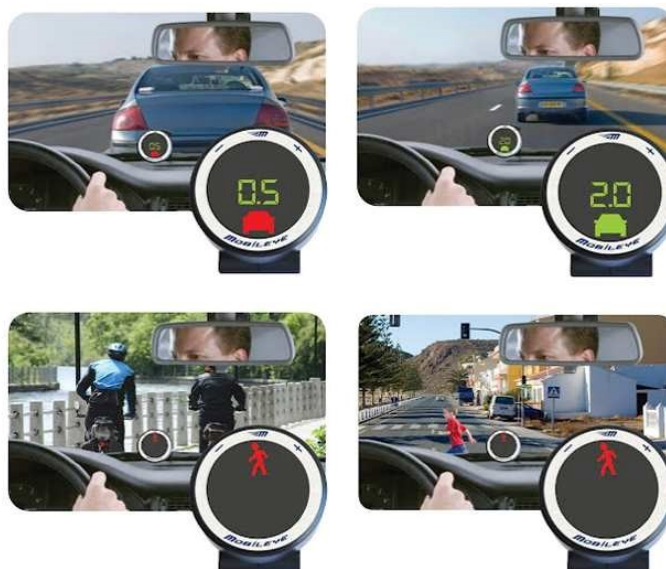


FIGURE 2-10 – Interface de Mobileye permettant d’alerter le conducteur en cas d’un éventuel accident [33]

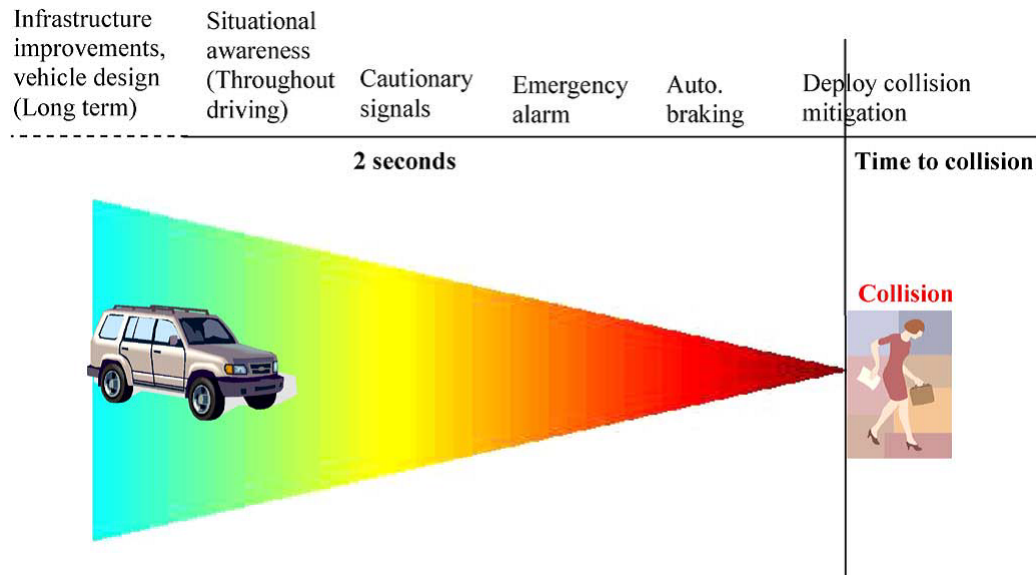


FIGURE 2-11 – Chronologie des mesures de protection des piétons. Le temps de réaction d’une certaine distance diminue avec l’augmentation de la vitesse [50]

mettant de protéger le piéton par une alerte au conducteur, ou par un déclenchement du freinage automatique.

Gandhi et Trivedi ont présenté sur [50] un système à différents niveaux allant de l’acquisition des données brutes aux actions prises par le système destiné au conducteur, représenté sur la Figure 2-12. Le processus est représenté sous forme de pyramide d’une base contenant une large quantité de données brutes. Le plus grand niveau de la pyramide reçoit l’information de sortie du niveau inférieur en combinaison avec les modèles à gauche de la pyramide. Au fur et à mesure que l’on gravit la pyramide, les informations utiles sont distillées par étapes successives, jusqu’à ce que finalement, on agisse sur la base d’une décision oui / non.

Les niveaux 1, 2, 3, 4 de la figure 2-12 consistent à faire la détection et le suivi du piéton. D’autres approches peuvent combiner les étapes de détection, de reconnaissance et de classification du piéton en une seule. L’étape de prétraitement traite directement les données brutes pour pouvoir identifier un potentiel candidat piéton en utilisant des algorithmes rapides. L’enjeu principal de cette étape est d’avoir un taux de détection élevé, même au détriment de l’autorisation des fausses alarmes.

La 3ème et la 4ème étape qui sont la classification et le suivi, appliquent des

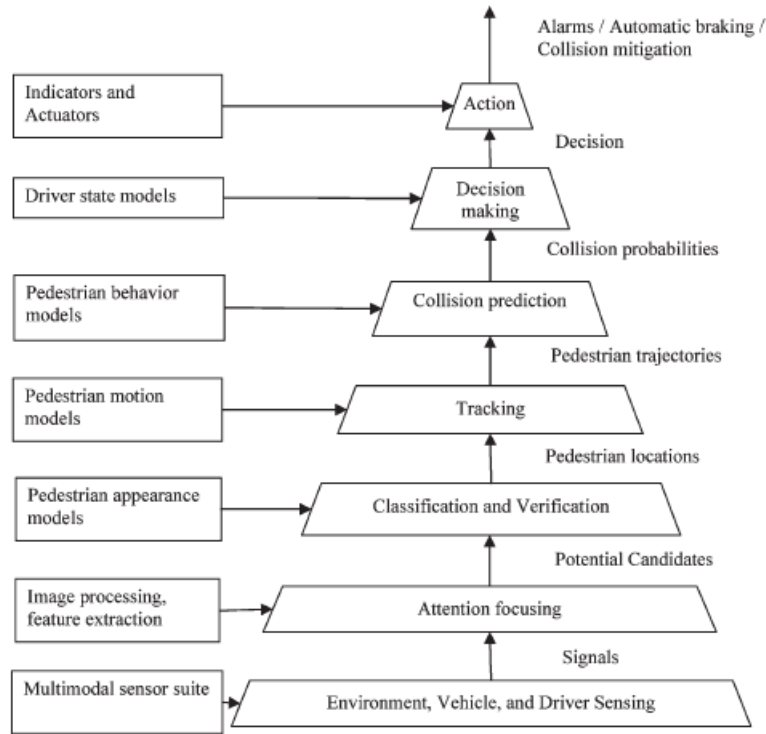


FIGURE 2-12 – Diagramme de flux de données montrant la distillation des informations à partir des données brutes jusqu'à l'action appropriée [50]

algorithmes plus complexes aux candidats pour pouvoir séparer les piétons réels des fausses alarmes. Les piétons détectés sont par la suite, suivis dans le temps pour avoir leur trajectoire. Cette dernière est envoyée ensuite au module de prédiction de collision qui va prédire la probabilité de collision entre le piéton et le véhicule. Dans le cas d'une haute probabilité, le conducteur reçoit l'alerte appropriée. Si l'accident est imminent, le système peut déclencher la décélération automatique du véhicule et diminuer l'impact de l'accident sur le piéton. Le même schéma peut être utilisé pour les capteurs montés sur les infrastructures qui peuvent détecter les véhicules et les piétons, calculer la probabilité de collision et envoyer les signaux d'alerte au conducteur dans le cas d'un danger en utilisant la communication sans fil.

Pour déterminer le niveau de danger entre le piéton et le véhicule, Thamasbi et al. [134] ont défini quatre zones de danger représentées sur la figure 2-13 en fonction de la vitesse du véhicule ainsi que de la distance entre le véhicule et le piéton.

La première zone représentée en rouge sur la figure 2-13, est une zone où l'accident

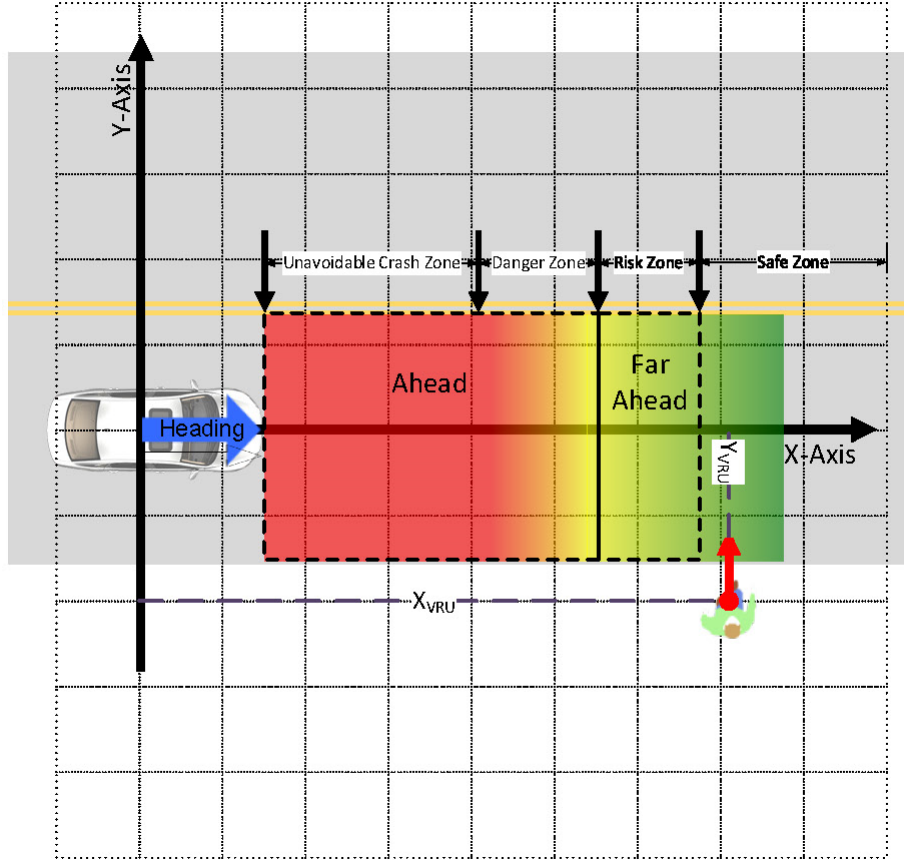


FIGURE 2-13 – Les quatre zones de danger entre le véhicule et le piéton [134]

est inévitable, autrement dit si un piéton est détecté sur cette zone, le système ADAS ne peut éviter la collision.

Pour déterminer la longueur de cette zone nous calculons le temps d'arrêt minimum ( $TTS_{min}$ ) du véhicule qui est égal au temps nécessaire pour que le véhicule s'arrête, en utilisant la décélération maximale pour freiner le véhicule ( $d_{max}$ ) basé sur la vitesse actuelle ( $v_{cur}$ ), l'accélération actuelle ( $a_{cur}$ ), et le temps de réaction du conducteur ( $T_{DRD}$ ) représenté par (eq 2.2).

$$TTS_{min} = -\frac{v_{brk}}{d_{max}} + T_{DRD} \quad (2.2)$$

Où ( $v_{brk}$ ) est la vitesse du véhicule après la réaction du conducteur qui égale à (eq 2.3) :

$$v_{brk} = a_{cur} * T_{DRD} + v_{cur} \quad (2.3)$$

La distance traversée durant ce  $TT S_{min}$  représente la distance minimale  $DT S_{min}$  pour que le véhicule s'arrête. La longueur de la zone est ainsi définie par (eq 2.4) :

$$DT S_{min} = D_{DRD} + \left( -\frac{v_{brk}^2}{2 * a_{max}} \right) \quad (2.4)$$

Où  $D_{DRD}$  est la distance traversée après la réaction du conducteur (eq :2.5) :

$$D_{DRD} = \frac{1}{2} * a_{cur} * T_{DRD}^2 + v_{cur} * T_{DRD} \quad (2.5)$$

La deuxième zone est la zone de danger, dans cette zone certains accidents peuvent être évités si le conducteur est alerté au bon moment. La longueur de cette zone est définie par la distance traversée durant le temps  $T_{dz}$  qui est égale au temps d'arrêt minimal  $TT S_{min}$  plus le temps nécessaire pour alerter le conducteur  $T_{alert}$  (eq :2.6) :

$$T_{dz} = TT S_{min} + T_{alert} \quad (2.6)$$

La troisième zone représente la zone de risque où le danger n'est pas imminent mais qui nécessite tout de même une alerte au conducteur si un piéton apparaît. La longueur de cette zone dépend du temps que le véhicule peut prendre pour un freinage souple  $d_{mod}$ , cette durée est représentée par (eq :2.7) :

$$T_{risque} = T_{dz} + T_{mod} \quad (2.7)$$

La dernière est une zone de sûreté où aucune alerte n'est nécessaire.

Nous nous sommes inspirés de ces quatre zones de sévérité entre un véhicule et un piéton, pour concevoir un système permettant d'alerter le conducteur lors de la présence d'un piéton indiscipliné, et ceci en utilisant son orientation. Un piéton indiscipliné traversant entre les véhicules d'une manière inadéquate dans la zone de sûreté peut augmenter le risque d'un accident. D'où la nécessité d'alerter le conducteur



de sa présence.

Nous proposons dans ce cadre, un nouveau système de prévention des accidents avec piétons. Ce système prend en considération la présence des piétons indisciplinés et les routes mal structurées.

## 2.4 Proposition d'un système de prévention des accidents avec piétons

Notre objectif est de proposer un système permettant de prévenir les accidents entre les piétons et les véhicules surtout pour les zones les moins structurées. Ces zones sont généralement définies par l'absence de signalisation des routes y compris les passages réservés aux piétons, ce qui augmente la présence de piétons traversant d'une manière indisciplinée, mettant ainsi leur vie en danger.

Bien que la présence des piétons indisciplinés soit surtout fréquente dans les zones représentant un manque d'installations piétonnes, traverser au milieu de la chaussée entre les véhicules reste préférable pour quelques piétons malgré la présence des passages cloutés.

Durant nos expérimentations réelles au Maroc, nous avons remarqué ce type de piétons dans différentes villes marocaines, y compris les villes les plus structurées du pays notamment la capitale Rabat comme nous pouvons voir sur la Figure 2-14.

En effet, lors de nos observations, des piétons insistent pour traverser au milieu de la chaussée pour différentes raisons. A titre d'exemple, la Figure 2-15 représente une dame qui traverse obliquement entre les véhicules pour rejoindre un taxi. Elle s'arrête ensuite au milieu de la route pour demander le trajet du taxi, et enfin retourne vers son emplacement initial de la même manière.

Cela n'empêche que l'existence des routes moins structurées au royaume encourage les piétons à adopter une traversée inappropriée comme le représente la figure 2-16, où nous remarquons l'absence du passage clouté et la présence de plusieurs piétons marchant dans la direction du trafic sur l'espace réservé aux véhicules, risquant ainsi



FIGURE 2-14 – Exemple de route structurée dans la capitale Rabat



(a)



(b)



(c)

FIGURE 2-15 – Exemple de piéton traversant de façon non appropriée dans une zone structurée



FIGURE 2-16 – Exemple de piéton indiscipliné dans une zone mal structurée

leur vie sans avoir la possibilité de voir les véhicules venant du trafic.

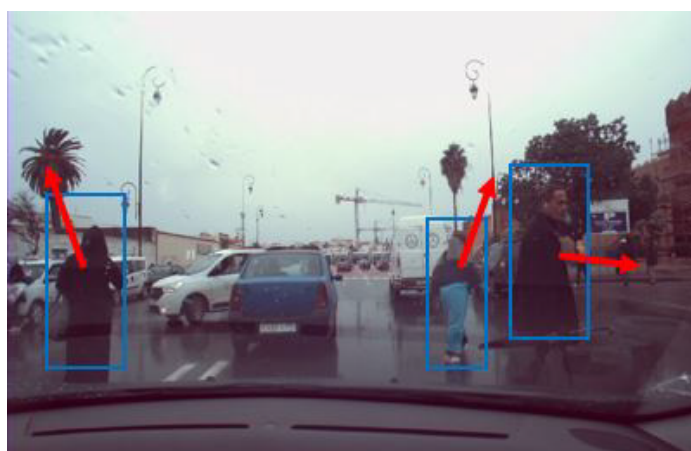
D'après nos observations, nous nous apercevons que le piéton peut changer sa direction rapidement comme illustré sur la figure 2-17. D'où la nécessité d'étudier la direction du piéton et de l'inclure dans un ADAS.

Pour l'ensemble des raisons citées plus haut, nous avons eu l'idée de proposer un nouveau système permettant de détecter la présence d'un piéton indiscipliné traversant entre les véhicules et ce en se basant sur son orientation, contenant un système d'alerte pour avertir le conducteur de sa présence ainsi que le niveau de risque de l'accident. Nous définissons trois situations possibles lors d'un conflit piéton-véhicule :

- La première situation représentée par l'équation (2.8) est lorsque le véhicule traverse avant le piéton, cette situation se produit quand le temps nécessaire pour que le véhicule traverse le point de collision est inférieur à celui dont le piéton a besoin pour atteindre le point de collision.  $D_v$  et  $D_p$  représentent respectivement la distance du véhicule et du piéton à la zone de collision,  $V_v$  et  $V_p$  représentent respectivement la vitesse du véhicule et du piéton, et  $W_p$  représente la largeur du véhicule. On suppose que la caméra est au milieu du véhicule,  $\frac{W_v}{2}$  est la distance entre le piéton et la caméra du véhicule.



(a)



(b)

FIGURE 2-17 – Des piétons changeant leurs directions dans un temps court

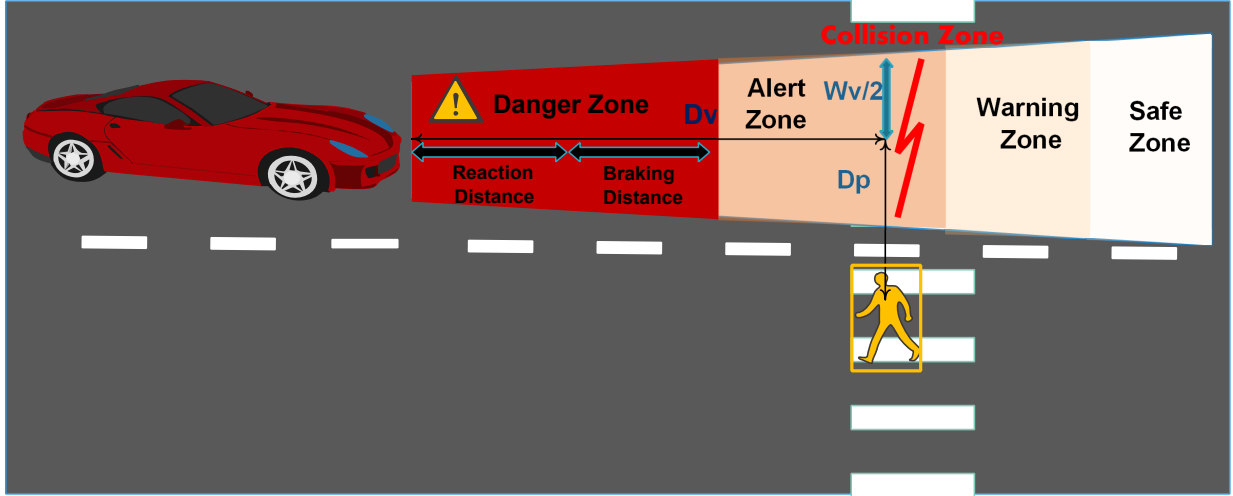


FIGURE 2-18 – Les quatre zones de risques entre le véhicule et le piéton

$$\frac{D_v}{V_v} < \frac{D_p - \frac{W_v}{2}}{V_p} \quad (2.8)$$

- La seconde situation (eq.2.9) est lorsque le piéton traverse avant le véhicule, ce cas peut être risqué pour le piéton si le conducteur accélère.

$$\frac{D_v}{V_v} > \frac{D_p + \frac{W_v}{2}}{V_p} \quad (2.9)$$

- La troisième situation est quand le piéton et le véhicule traversent le point de collision en même temps, représentée par (eq.2.10).

$$\frac{D_p - \frac{W_v}{2}}{V_p} < \frac{D_v}{V_v} < \frac{D_p + \frac{W_v}{2}}{V_p} \quad (2.10)$$

Pour ce troisième cas la probabilité d'une collision dépend du temps d'arrêt du véhicule. Nous définissons la gravité de la situation dans ce cas est définie selon quatre zones de risques : La zone de danger, la zone d'alerte, la zone d'avertissement et la zone de sûreté comme représentées sur la Figure 2-18.

### 2.4.1 Temps d'arrêt du véhicule

La longueur de chaque zone dépend du temps d'arrêt du véhicule (TTS), ce dernier est lié au temps de réaction du conducteur et à la distance de freinage du véhicule.

#### Distance de freinage

La distance de freinage représente la distance parcouru par le véhicule à partir du moment où le conducteur appuie sur le frein. Elle dépend de la vitesse du véhicule, les conditions de la route et le taux de décélération, l' « American Association of State Highway and Transportation Officials (AASHTO) [1] » recommande d'utiliser la valeur de  $3.4\text{m/s}^2$ . La distance d'arrêt se calcule en utilisant l'équation 2.11 où  $v$  représente la vitesse du véhicule,  $f$  représente le coefficient de friction entre les pneus et la route (0.7 sur une route sèche et entre 0.3 et 0.4 sur une route mouillée),  $G$  représente la pente de la route (positive pour une montée et négative pour une descente).

$$D_b = \frac{v^2}{(254 * (f + G))} \quad (2.11)$$

#### Temps de réaction du conducteur

Le temps de réaction du conducteur est le temps dont le conducteur a besoin pour prendre conscience qu'une réaction est nécessaire et qu'il commence à appuyer sur le frein. Le temps de réaction dépend de la vigilance du conducteur au moment de l'événement. Selon une étude faite par Johansson et al. [65] le temps de réaction moyen d'un conducteur pour un événement attendu est de 0.6 seconde. Ce temps peut augmenter si l'événement est inattendu où si le conducteur est distrait. Pour mesurer la distraction du conducteur, Choudhary et al. [26] ont observé dans leur papier que le temps de réaction du conducteur dans différentes conditions d'utilisation du téléphone et ont trouvé que le temps de réaction moyen pour un conducteur envoyant des messages est de 2.72 secondes comparé à 0.99 seconde lorsque le conducteur n'utilise pas de téléphone. Le temps de réaction augmente aussi si l'événement est

inattendu : Johansson et al. [65] dans leur étude ont montré que le temps de réaction augmente de 35% pour les événements inattendus ce qui mène à une réaction de 2.7 secondes pour quelques conducteurs. L'AASHTO [1] adopte un temps de réaction de 2.5 secondes qui inclu l'aptitude de tous les conducteurs y compris les sujets âgés.

Selon la norme d'AASHTO, la distance d'arrêt est représenté par l'équation 2.12, où  $t$  représente le temps de réaction du conducteur, et  $v$  représente la vitesse du véhicule.

$$StoppingDistance = (0.278 * t * v) + D_b \quad (2.12)$$

Le temps d'arrêt du véhicule peut être extrait de l'équation 2.12 en divisant la distance d'arrêt par le taux de décélération du véhicule.

Afin d'alerter le conducteur sur un possible accident ont doit prendre en considération ces indications pour avoir un système convenable pour chaque situation. Pour cela nous avons adopté pour notre système un temps d'arrêt minimum ( $TTS_{min}$ ) et un temps d'arrêt de distraction ( $TTS_{dist}$ ). Le temps d'arrêt minimum ( $TTS_{min}$ ) représente le temps d'arrêt du véhicule avec le temps de réaction minimum du conducteur qui est de 0.6 seconde. Alors que le temps d'arrêt de distraction ( $TTS_{dist}$ ) représente le temps d'arrêt du conducteur avec un temps de réaction de 2.5 secondes.

La Figure 2-19 représente un aperçu du système proposé. Le système prend en entrée des images prises d'une caméra monoculaire intégrée dans un véhicule en mouvement. En utilisant des algorithmes de l'apprentissage profond comme représenté sur l'algorithme 1, le système détecte la présence des piétons et reconnaît leurs orientations. Ensuite, l'imagette de chaque piéton détecté est coupée en utilisant les cadres englobants obtenus par l'algorithme YOLOv3. Afin d'obtenir une taille fixe pour toutes les images, nous avons redimensionné les imagettes en 48x48 pixels, avant de les utiliser comme entrée au réseau de Capsules pour avoir l'orientation du piéton.

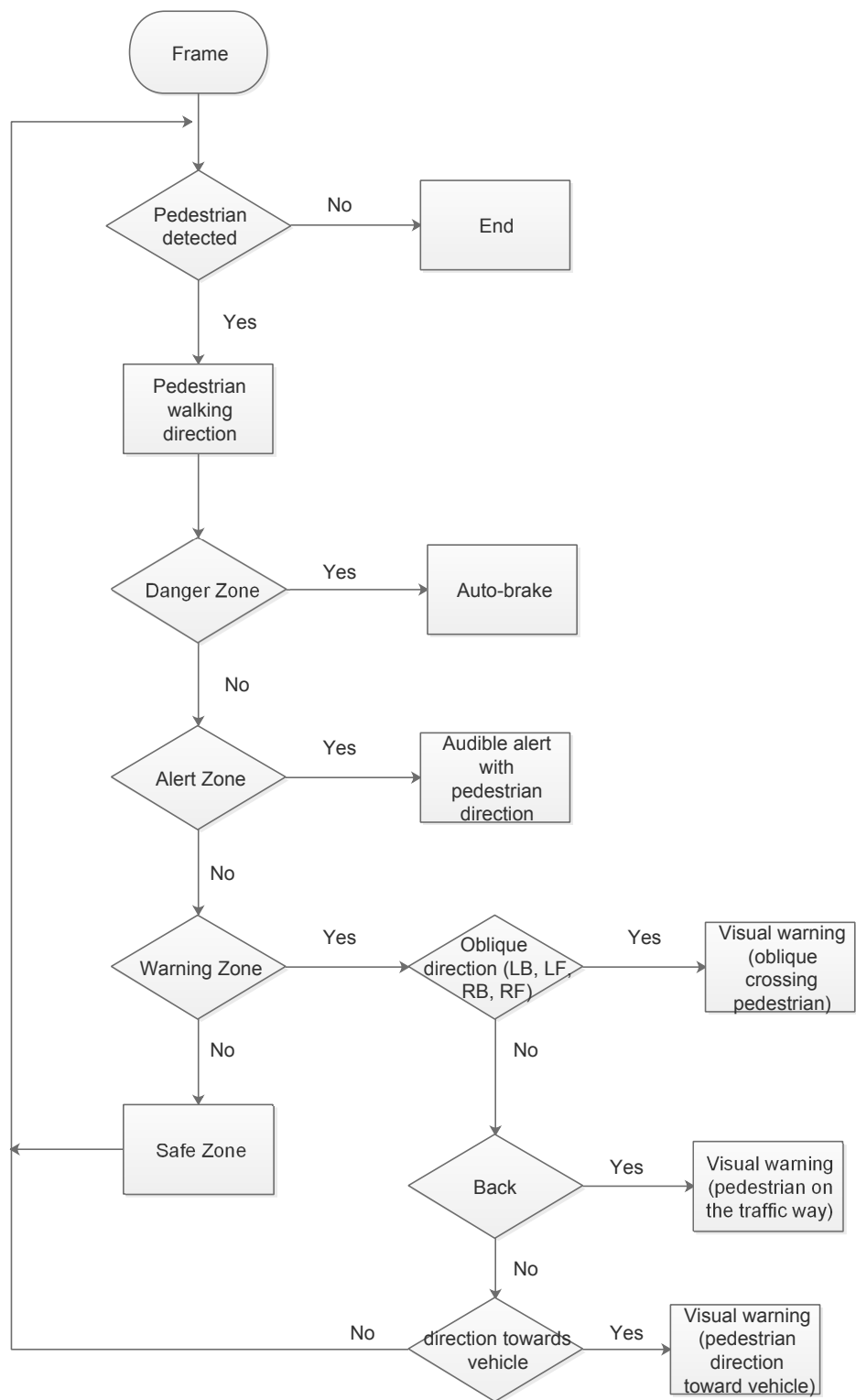


FIGURE 2-19 – Architecture du système de prévention des accidents proposé



---

**Algorithm 1:** Reconnaissance de l'orientation du piéton

---

**Input:** Image à partir d'une caméra monoculaire

**Output:** Orientation du piéton

- 1: Détection des piétons en utilisant YOLOv3.
  - 2: Couper les imagerie des piétons détectés en utilisant les cadres englobants.
  - 3: Redimensionner les imagerie en 48x48 pixels.
  - 4: Convertir les imagerie en niveau de gris.
  - 5: Utiliser les imagerie en gris comme entrée au réseau de Capsule pour avoir l'orientation du piéton.
- 

### 2.4.2 Estimation du risque de collision

Le système proposé permet d'alerter le conducteur si un accident est sur le point de se produire avec un piéton, suivant le schéma ci-dessus. La gravité de la situation qui est représentée par les zones de risque dépend de la vitesse du véhicule, de la distance entre le piéton et le véhicule ainsi que de la direction de marche du piéton. Le risque de l'accident augmentera si le piéton marche sur la chaussée avec des véhicules roulant à grande vitesse. Dans ce cas, déterminer l'orientation de marche du piéton est nécessaire pour évaluer si le piéton est en situation de danger ou non, et choisir le mode d'alerte qu'il faut activer (Visuel – Audio).

#### La zone de danger

La zone de danger indique la zone où l'accident est inévitable. Nous supposons qu'aucun accident peut être évité si le temps d'arrêt du véhicule minimal ( $TTS_{min}$ ) est supérieur que le temps requis par le piéton pour arriver à l'autre côté du véhicule. Ce temps est représenté par l'équation (2.13).

$$TTS_{min} \geq \frac{D_p - \frac{W_v}{2}}{V_p} \quad (2.13)$$

#### La zone d'alerte

La zone d'alerte est une zone où le risque d'accident est toujours élevé mais qui peut être évité si le conducteur réagit rapidement. La longueur de cette zone est définie en considérant les conditions qui peuvent augmenter le temps d'arrêt du véhicule

tel que l'inattention du conducteur. Nous considérons qu'un piéton est dans la zone d'alerte si le temps nécessaire pour atteindre l'autre bord du véhicule est entre  $TTS_{min}$  et  $TTS_{dist}$  comme représenté sur l'équation (2.14).

$$TTS_{min} < \frac{D_p + \frac{W_v}{2}}{V_p} < TTS_{dist} \quad (2.14)$$

La raison derrière le choix de la longueur de cette zone, est de s'assurer que le conducteur distrait est pris en considération pour qu'il puisse éviter l'accident s'il est alerté au bon moment. Pour cette zone le système active une alarme sonore pour alerter le conducteur comme représenté sur la Figure 2-19.

### La zone d'avertissement

Le risque d'un accident peut augmenter si le piéton traverse d'une manière inappropriée, ce qui est très remarqué sur les routes mal structurées. Dans ce cas, l'orientation du piéton peut avoir un impact significatif sur la survenue de l'accident, spécialement pour les zones les moins structurées où la conception de l'infrastructure routière ne permet pas de séparer le piéton du véhicule en raison de l'absence de l'installation piétonne, telle que les trottoirs et les passages cloutés, ce qui incite le piéton à partager la route avec des véhicules roulant à grande vitesse, et traverser ainsi la chaussée d'une manière inappropriée. Les accidents de piétons sont deux fois plus susceptibles de se produire sur les routes sans passages piétons, tel que rapporté par l'Organisation Mondiale de la Santé (OMS) [141].

Un piéton est considéré indiscipliné si :

- Il traverse la route obliquement ce qui diminue la distance entre le piéton et le véhicule et augmente le risque de collision. La traversée oblique est représentée par les directions de piétons suivantes : Gauche-arrière, Droite-arrière, Gauche-devant, Droite-devant.
- Quand le piéton partage la route avec les véhicules ce qui est représenté par les direction suivantes : Arrière, Devant. Le risque d'un accident augmente 2 fois de plus comme cité par l'Organisation Mondiale de la Santé (OMS) [99]

- surtout si le piéton marche vers la direction du trafic qui est représenté par [101, 131] comme étant plus risqué que de marcher contre la direction du trafic, vu que les véhicules ne sont pas visibles pour le piéton dans ce cas.
- Lorsque le piéton traverse perpendiculairement vers la direction du véhicule ce qui est représenté par : Droite, Gauche.

De ce fait, le risque de collision dépend de l'orientation du piéton relativement à celle du véhicule. Une alerte visuelle est activée quand le piéton traverse d'une manière dangereuse. Nous recommandons pour cette zone une marge de deux secondes comme représenté par l'équation (2.15). L'orientation du piéton est traitée en utilisant les réseaux de Capsules.

$$ST_{min} < \frac{D_p + \frac{W_v}{2}}{V_p} < ST_{dist} + 2sec \quad (2.15)$$

## La zone de sûreté

Dans cette zone, le système n'envoie aucun avertissement ou alerte au conducteur car aucun danger n'est signalé.

L'ensemble du processus d'estimation du risque de collision est détaillé dans l'al-

gorithme 2.

---

**Algorithm 2:** Estimation du risque de collision

---

**Input:** *Paramètres environnementaux* (vitesse, distance, météo, pente de la route)

*Images* : caméra monoculaire

**Output:** Estimation du risque de collision

**foreach** *piéton* **do**

1: Calculer  $\frac{D_p - \frac{W_v}{2}}{V_p}$  and  $\frac{D_p + \frac{W_v}{2}}{V_p}$

2: Calculer  $\frac{D_v}{V_v}$

**if**  $\frac{D_p - \frac{W_v}{2}}{V_p} < \frac{D_v}{V_v} < \frac{D_p + \frac{W_v}{2}}{V_p}$  **then**

Calculer  $TTS_{min}$  and  $TTS_{dist}$  ;

**else if**  $TTS_{min} \geq \frac{D_p - \frac{W_v}{2}}{V_p}$  **then**

$Risk\ zone \leftarrow Danger\ zone$  ;

**else if**  $TTS_{min} < \frac{D_p + \frac{W_v}{2}}{V_p} < TTS_{dist}$  **then**

$Risk\ zone \leftarrow Alert\ zone$  ;

**else if**  $TTS_{min} < \frac{D_p + \frac{W_v}{2}}{V_p} < TTS_{dist} + 2sec$  **then**

$Risk\ zone \leftarrow Warning\ zone$  ;

**else**

$Risk\ zone \leftarrow Safe\ zone$  ;

**end**

---

## 2.5 Conclusion

Le piéton est l'utilisateur le plus vulnérable de la route, selon les statistiques de l'OMS, le piéton reste l'utilisateur de route le plus impliqué par les accidents et par le plus grand taux de décès en Afrique. La diminution du nombre de victimes au niveau des piétons, passe par la compréhension des causes de ces accidents en étudiant le comportement du piéton et de son antagoniste le conducteur, leur interaction ainsi que leur communication.

La majorité des études faites sur ce volet, s'entendent à conclure que les jeunes piétons et conducteurs partagent un comportement plus ou moins aventurier surtout pour le genre masculin, contrairement aux femmes qui représentent un comportement généralement prudent. L'étude de l'interaction entre le piéton et le conducteur reste indispensable pour comprendre l'augmentation du nombre d'accidents impliquant un piéton. Les recherches réalisées montrent que cette interaction est majoritairement conditionnée à une faible communication entre les deux usagers de routes en raison du manque de moyen de communication. Ceci engendre un malentendu sur qui va traverser en premier, le conducteur de sa part refuse de céder le passage en jugeant que le piéton renoncera à traverser, alors que ce dernier persiste à traverser en pensant que le conducteur s'arrêtera.

Une des solutions pour éviter le conflit piéton-conducteur, est d'intégrer sur les véhicules un système d'alerte de collisions avec les piétons visant à alerter le conducteur pour diminuer sa vitesse lors de la présence d'un piéton afin d'éviter un éventuel accident, ou pour diminuer les dégâts si l'accident est inéluctable.

Nous avons proposé dans ce travail, un nouveau système permettant d'alerter le conducteur lors d'un risque d'accident avec un piéton. Le type d'alerte dépend de la gravité du risque selon la zone qui a été déterminée : la zone de danger, la zone d'alerte, la zone d'avertissement et la zone de sûreté. Ce système prend en considération la présence des piétons indisciplinés et des routes non structurées ne possédant pas d'installations piétonnes. Afin de détecter un piéton indiscipliné le système étudie la direction du piéton selon la situation, ce qui représente la première contribution

majeure de cette thèse.



# Chapitre 3

## Détection de l’Orientation du Piéton

### 3.1 Introduction

Le piéton est l’utilisateur de la route le plus vulnérable avec 28% de la mortalité routière au monde. Il se caractérise par sa mobilité physique lui permettant de changer rapidement de direction, d’où la nécessité de l’intégrer aux systèmes de protection. La majorité des recherches impliquant la sécurité des piétons s’intéresse autant à sa détection qu’à prédire sa direction de mouvement. Toutefois, la direction du piéton reste un aspect primordial pour les systèmes de protection. Ceci peut alors améliorer considérablement la sécurité des usagers de la route.

Les erreurs humaines dues à l’excès de vitesse, l’inattention, la conduite en état de fatigue et la distraction (utilisation de téléphone, ...) accompagnées parfois d’un comportement à risque de la part du piéton peuvent générer des accidents importants.

Ces accidents peuvent être largement diminués si les Systèmes Avancés d’Assistance du Conducteur (ADAS) intègrent non seulement la détection du piéton mais aussi la reconnaissance de sa direction de marche. Ce qui permettra d’améliorer la sécurité des piétons notamment sur les environnements non structurés caractérisés par le manque d’installations obligeant parfois le piéton à adopter des comportements à risque.

Comme cité dans le chapitre précédent, les différents comportements et interactions des piétons et des conducteurs varient selon plusieurs facteurs, notamment l’âge



et le genre du sujet concerné. Le comportement des conducteurs tend à être plus risqué et inapproprié chez les jeunes du genre masculin. Statistiquement parlant, les jeunes hommes de moins de 30 ans représentent la proportion la plus élevée des accidents mortels et qui est trois fois plus importante que celle des conductrices. Ce fait peut être expliqué par les excès de vitesse chez les jeunes conducteurs masculins et un taux élevé de la conduite en état d'ivresse.

Les jeunes piétons du genre masculin possèdent également un comportement à risque qui se manifeste par une traversée inappropriée loin de l'emplacement réservé aux piétons. Contrairement aux conducteurs qui peuvent être protégés par les véhicules, les piétons subissent des attaques directes et deviennent donc les usagers les plus vulnérables de la route.

D'autres facteurs peuvent influencer le comportement du piéton, notamment l'infrastructure routière. En d'autres termes, les routes non structurées ne disposant d'aucun aménagement pour les piétons, poussent ces derniers à adopter un comportement dangereux, comme par exemple la traversée en diagonale des piétons.

Nous avons proposé dans le chapitre précédent un système d'alerte de collision entre piétons et véhicules incluant l'orientation du piéton. Dans ce chapitre, nous traitons la reconnaissance de l'orientation du piéton en proposant une nouvelle approche basée sur les réseaux de Capsules. La majorité des bases de données existantes traitent de la détection des piétons mais pas leur direction de mouvement. À notre connaissance, seulement trois bases de données traitent l'orientation du piéton. Les bases ROVIT (3 directions), TUD (8 directions) et JAAD (4 directions). Parmi ces trois bases, JAAD est la seule qui peut être définie comme réaliste, en filmant à partir d'un véhicule en mouvement des piétons traversant en zone urbaine. Elle sera utilisée ultérieurement dans le chapitre suivant pour la reconnaissance de l'orientation des piétons sur des séquences vidéos.

Pour cette raison, nous avons créé une nouvelle base de données comportant l'orientation des piétons, que nous avons nommée SafeRoad. La base SafeRoad, est une base de données de zone urbaine marocaine. Elle restitue des scènes routières de trafic urbain depuis un véhicule en mouvement comportant des piétons se déplaçant

dans leur espace. Une des difficultés que l'on rencontre avec une base réaliste est l'impossibilité d'utiliser les techniques classiques de classification telle que la soustraction de l'arrière-plan « Background subtraction » à cause de l'environnement en mouvement.

La base SafeRoad est manuellement annotée pour 4 ou 8 directions du piétons.

Dans ce chapitre nous détaillons les contributions suivantes :

- Proposition d'une nouvelle base de données réaliste de l'orientation du piéton.
- Proposition d'une nouvelle méthode de reconnaissance de l'orientation du piéton basée sur les réseaux de Capsule.
- Comparaison du résultat avec les réseaux CNN (AlexNet et ResNet).

## 3.2 Etude de l'Existant

### 3.2.1 Méthodes Classiques

Parmi les auteurs traitant l'orientation du piéton pour les véhicules intelligents nous trouvons Enzweiler et Gravila [40]. Ils présentent dans une seule approche la classification du piéton et l'estimation de son orientation à partir d'une seule image. L'approche est basée sur un framework de probabilité en utilisant des modèles discriminatifs. La classification du piéton implique une décision à posteriori entre la classe "piéton" et la classe "non-piéton". L'orientation du piéton est intégrée en analysant la probabilité de densité à l'orientation du piéton, comme représenté sur la Figure 3-1 . Sauf que leur travail consiste à traiter seulement les orientations Avant/Arrière et Droite/Gauche, alors que dans un environnement réel, les orientations obliques sont souvent remarquées et représentent des situations dangereuses pour les collisions entre piétons et véhicules.

Hariyono et al. [58] présentent dans leur papier une méthode pour estimer la direction de marche du piéton et pour prédire son chemin à partir d'un ensemble d'images prises à l'aide d'une caméra monoculaire intégrée dans un véhicule en mouvement. Le système proposé adresse dans un premier temps l'orientation du piéton en utilisant

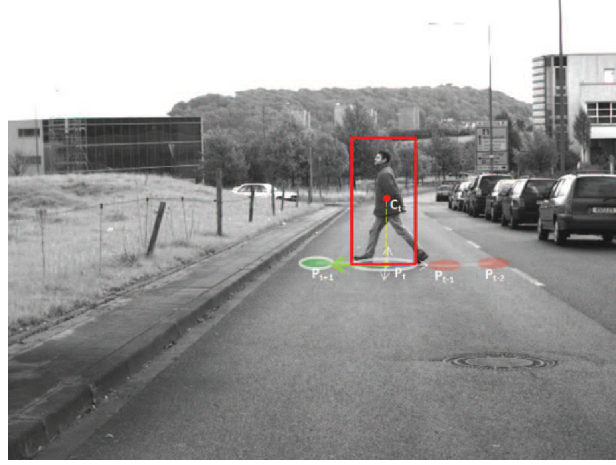


FIGURE 3-1 – Détection d'un piéton en mouvement et la détermination de sa position sur l'image [58]

une seule image prise à partir d'une scène réelle, en prenant en considération la localisation du piéton sur l'image et sa distance avec le véhicule. La distance du piéton par rapport au véhicule est calculée en projetant le centre de la bounding box du piéton détectée sur le plan terrain pour avoir ses coordonnées globales et sa position dans la scène (Fig. 3-1). La direction de marche du piéton est par la suite estimée en utilisant la concaténation de plusieurs images consécutives en se basant sur le résultat de l'orientation estimée à partir d'une seule image. Ensuite, le papier prédit le chemin du piéton en évaluant ce dernier avec le Filtre Kalman (KF), les Modèles Multiples Interactives (IMM) et la correspondance de trajectoire hiérarchique probabiliste.

Le papier a défini trois méthodes pour estimer la direction du piéton sur plusieurs images, la première méthode estime que l'orientation du piéton est déterminée à partir de la direction qui est détectée fréquemment. Elle est calculée par l'équation suivante (3.1) :

$$D_f = \operatorname{argmax}_i(f_1, f_2, \dots, f_n) \quad (3.1)$$

Où  $D_f$  représente la direction détectée la plus fréquente et  $f_n$  représente la fréquence d'apparence à l'orientation  $i$ . La deuxième méthode consiste à calculer la moyenne arithmétique des directions de plusieurs images consécutives, pour trouver ainsi la direction moyenne  $D_a$ .

La troisième méthode est la méthode hybride, elle consiste à rassembler les deux méthodes introduites précédemment. La combinaison des deux méthodes résout le problème de changement imprévu de direction. La méthode est représentée par l'équation (3.2) :

$$D = \operatorname{argmax}_i(D_i, D_f, D_a) \quad (3.2)$$

Où  $D_i$  représente la direction dans la dernière image appartenant à l'ensemble des images prises durant une seconde. La méthode hybride a donné de meilleurs résultats par rapport à d'autres méthodes avec une exactitude (eq.3.3) de 97.50% sur la base Daimler. Le tableau 3.1 représente une comparaison des résultats entre les trois méthodes.

$$\text{Exactitude} = \frac{TP + TN}{TP + FP + TN + FN} \quad (3.3)$$

TABLE 3.1 – Comparaison des résultats des trois méthodes sur l'orientation du piéton

| Méthode   | Exactitude    |
|-----------|---------------|
| Fréquence | 92.50%        |
| Moyenne   | 81.80%        |
| Hybride   | <b>97.50%</b> |

### 3.2.2 Méthodes utilisant l'apprentissage automatique

Choi et al. [25] ont présenté dans leur papier une estimation de l'orientation du piéton en utilisant les réseaux de neurones de convolutions (CNN). Le modèle prend en entrée une image de piéton de résolution 32x32 pixels et donne en sortie son orientation. Le modèle CNN proposé contient deux couches de convolutions et deux couches denses. Contrairement aux réseaux CNN, le modèle proposé par [25] ne contient pas de couche Pooling, ceci est dû à l'importance de l'information spatiale pour estimer l'orientation du piéton. L'entrée est sous la forme d'une image de piéton de résolution 32x32 pixels, l'algorithme fournit en sortie l'orientation du piéton. Pour l'évaluation du modèle, les auteurs ont utilisé la base Human 3.6M [63] contenant des vidéos de 11

personnes (6 hommes et 5 femmes) filmées par de multiples caméras. La base fournit la position commune 3D des sujets, par contre elle ne contient pas une annotation de l'orientation du corps. Pour annoter la base, les auteurs ont pu extraire l'orientation du piéton en utilisant les positions communes fournies par la base pour calculer le produit vectoriel du vecteur reliant le cou à la jambe droite (vecteur en rouge sur la Figure. 3-2) et du vecteur reliant le cou à la jambe gauche (vecteur en vert sur la Figure. 3-2). Ensuite, le vecteur calculé (en bleu) est projeté sur un plan XY, afin de calculer l'angle d'orientation du corps (vecteur orange) en utilisant le vecteur projeté. La base est labellisée en 8 orientations ayant un angle de  $45^\circ$  chacune. La base contient 74.862 images dont 64.862 sont utilisées pour l'apprentissage et 10.000 pour la validation.



FIGURE 3-2 – Extraction de l'orientation du piéton à partir des positions communes [25]

La performance du modèle a été comparée par les méthodes traditionnelles d'extraction des caractéristiques HOG et SVM. Comme représenté sur le Tableau 3.2, le modèle proposé par [25] a surpassé les méthodes traditionnelles avec une performance de 81.58% sur l'ensemble de validation.

TABLE 3.2 – Performance du réseau CNN par rapport aux méthodes traditionnelles sur la base Human 1.6M

| Modèle              | Exactitude    |
|---------------------|---------------|
| CNN                 | <b>81.58%</b> |
| HOG+SVM(Linéaire)   | 57.26%        |
| HOG+SVM(Gaussienne) | 57.12%        |

L'orientation du corps humain a aussi été traitée aussi par Kohari et al. [71] pour la robotique, en utilisant un modèle basé sur les CNN. Les auteurs ont utilisé un réseau de 13 couches de convolutions et de 3 couches denses. L'entrée est une image, en niveau de gris de 100x100 pixels, d'une personne. Pour entraîner le modèle, la base SURREAL [137] a été utilisée. C'est une base générée en combinant d'images synthétiques des humains avec un arrière-plan réel. La base contient un ensemble d'images RGB, d'images de profondeur, ainsi que des locations communes 2D et 3D, par contre elle ne contient pas d'annotation sur l'orientation du corps. Les auteurs ont calculé le vecteur d'orientation en utilisant le produit vectoriel 3D entre le vecteur de l'épaule droite et celui de l'épaule gauche (comme représente la Figure 3-3). Le vecteur 3D est par la suite projeté sur le plan XY. Le nombre d'images sur la base utilisée est de 2.336.851. Pour tester le modèle, les auteurs ont créé leur propre base de piétons en utilisant un système avec une table de rotation. Pour estimer l'orientation, trois capteurs ont été utilisés, une caméra RGB, un lidar 2D et un lidar 3D. La base de test contient 15.142 images. Les orientations utilisées sont de  $0^\circ$  à  $360^\circ$ .

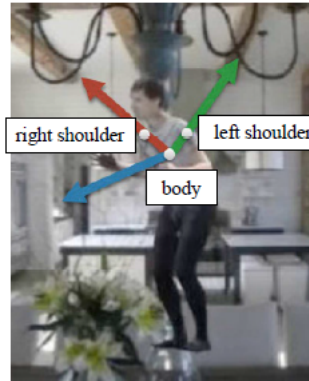


FIGURE 3-3 – Calcul du vecteur d'orientation pour le corps humain [71]

Les auteurs ont évalué la performance du modèle avec un intervalle de zéro, dix et vingt degrés, donnant respectivement une exactitude de 47.7%, 89.7%, 97.5%.

Hara et al. [57], ont utilisé les réseaux de neurones de convolutions profonds (DCNN) pour une estimation de l'orientation continue des objets, pour une prédiction de  $0^\circ$  à  $360^\circ$  degrés. Les auteurs comparent trois approches pour prédire l'orientation continue de l'objet. Les deux premières approches représentent l'orientation comme

un point sur un cercle afin de minimiser par la suite soit la fonction de perte L2 soit la différence angulaire. Alors que la troisième approche convertit dans un premier temps l'estimation de l'orientation continue en un ensemble de tâches d'estimation discrète de l'orientation. Par la suite, les sorties de ses estimations discrètes sont converties vers une orientation continue en utilisant l'algorithme « mean-shift ». La base TUD Multiview Pedestrian Dataset [6] a été utilisée pour une évaluation sur les piétons, elle contient 5228 images de piétons annotées en 8 orientations, avec 4732 images pour l'apprentissage, 290 pour la validation et 309 pour le test. Les images de la base TUD ont été converties en images couleurs pour une utilisation sur le réseau DCNN proposé. L'exactitude représentée par l'eq.3.3 des prédictions de l'orientation avec un intervalle de 22.5° et 45° degrés de la vérité terrain, a été calculée pour les trois approches. Selon les résultats, la troisième méthode utilisant la discrétisation a reçu la meilleure performance avec une exactitude de 63.1% et 82.8% pour respectivement 22.5° et 45° degrés. Le tableau 3.3 représente une comparaison entre l'exactitude des trois méthodes pour 22.5 et 45 degrés.

TABLE 3.3 – Performance du réseau CNN par rapport aux méthodes traditionnelles sur la base Human 1.6M

| Modèle | Exactitude -22.5° | Exactitude - 45° |
|--------|-------------------|------------------|
| 1      | 46.9%             | 75.7%            |
| 2      | 44%               | 70.9%            |
| 3      | 63.1%             | 82.8%            |

Appliquée au Système des Transports Intelligents (ITS), Sanchez et al. [38], proposent une variante du CNN pour la reconnaissance de la direction de mouvement du piéton. Le papier propose une nouvelle base de données nommée ROVIT, traitant de l'orientation du piéton. Elle contient 7416 images pour l'apprentissage, 1752 pour la validation et 1632 pour le test, annotées manuellement en 3 directions (Droite, Gauche et Devant), avec 4035 images labellisées en Droite, 4227 en Gauche et 2538 en Devant. Pour l'apprentissage de la base, les auteurs ont utilisé les réseaux AlexNet, GoogLeNet et ResNet. Pour comparer les approches, un algorithme de flux optique et le gradient de l'historique du mouvement de l'image (MHI) ont été utilisés. La direc-

tion globale du flux optique est calculée pour fournir des orientations entre 0 et 360 degrés. Ces valeurs ont été discrétisées en trois intervalles de 120 degrés chacun pour les directions suivantes : Gauche, Droite et Avant. Les résultats obtenus montrent que cette approche a une performance faible sur la base de tests avec une exactitude de 51%, 39% et 40% pour chaque direction. Entre les trois approches CNN, ResNet a reçu la meilleure performance avec 79% sur la base de test, alors que GoogLeNet et AlexNet ont reçus respectivement une exactitude de 77% et 71%.

### 3.3 Les réseaux de Capsule

Le réseau de Capsule (CapsNet) est une nouvelle architecture de réseaux de neurones proposée par Geoffrey Hinton et Sara Sabour [122]. Ce nouveau réseau permet de dépasser les limites des réseaux convolutifs pour la reconnaissance d'image.

Les réseaux de neurones convolutifs se distinguent par la couche de pooling. Cette dernière, permet de compresser l'information et de réduire la taille des images dans le réseau. Ce qui en résulte une perte de la relation spatiale entre les caractéristiques extraites par la couche de convolution. De ce fait, les réseaux convolutifs permettent de classer correctement un visage dont des éléments caractéristiques, comme par exemple la bouche ou les yeux sont positionnés de façon aléatoire.

Les CapsNets viennent répondre à ce débat connu comme le "problème de Picasso", en prenant en compte la position relative des caractéristiques dans l'image. L'idée est d'ajouter des structures appelées «capsules» à un réseau de neurones à convolution (CNN) et de réutiliser les sorties de plusieurs de ces capsules pour former des représentations plus stables (par rapport à diverses perturbations) de capsules d'ordre supérieur. La sortie est un vecteur représentant la probabilité d'une observation et d'une pose.

Au sein de chaque capsule, les neurones représentent les propriétés d'une caractéristique de l'image pouvant inclure différents paramètres d'instanciation, notamment la « pose » (position, taille, orientation), luminosité, etc. Ce qui explique le choix de l'utilisation de ce nouveau réseau pour la reconnaissance de l'orientation du piéton.



Les réseaux de Capsules contiennent une partie d'encodage et une autre de décodage. La partie d'encodage permet de classer l'image d'entrée, qui est dans notre cas un piéton. La partie de décodage quant à elle, permet de reconstruire l'image d'entrée à partir du résultat de classification de l'encodage. La phase d'encodage contient deux couches de convolutions et une couche entièrement connectée. Le principal rôle de cette partie étant la classification, alors que celle du décodage contient trois couches entièrement connectées qui servent à reconstruire l'image d'entrée.

La première couche de convolution appartenant à la partie d'encodage permet d'extraire la carte de caractéristiques de l'image d'entrée, le résultat de la première couche constitue l'entrée de la deuxième couche de convolution appelée couche primaire des capsules. Cette couche contient 32 capsules de 8 dimensions. Les vecteurs de cette couche doivent avoir une valeur entre 0 et 1 puisqu'ils représentent la probabilité d'existence de l'entité, pour cela une fonction appelée "squash" est appliquée pour normaliser les vecteurs entre 0 et 1. Alors que le rôle principal de la troisième couche est la classification, elle permet d'attribuer à l'image d'entrée la classe ayant la plus grande probabilité, le nombre de capsules dans cette couche égale au nombre de classes possible où chaque capsule représente une classe.

Les capsules de la deuxième couche (couche primaire des capsules) prédisent les vecteurs de sortie  $\hat{u}_{j/i}$  (eq.3.4) de la troisième couche en utilisant leur propre vecteur de sortie  $u_i$  multiplié par la matrice de transformation  $W_{ij}$  :

$$\hat{u}_{j/i} = W_{ij}u_i \quad (3.4)$$

La matrice de transformation  $W_{ij}$  est entraînée par le réseau progressivement en utilisant la rétropropagation durant le processus d'apprentissage de la couche primaire des capsules. Par la suite un agrément  $a_{ij}$  entre la valeur prédite  $\hat{u}_{j/i}$  par la capsule  $i$  de la deuxième couche et la valeur réelle par la capsule  $j$  de la troisième couche  $v_j$ , se calcule en utilisant le produit scalaire suivant (eq.3.5) :

$$a_{ij} = \hat{u}_{j/i} \cdot v_j \quad (3.5)$$

Pour chaque vecteur prédit un poids de routage appelé  $b_{ij}$  est utilisé. Il est initialisé à zéro pour toutes les capsules sur les deux couches. Ensuite, une fonction softmax  $c_{ij}$  est appliquée à ce poids de routage pour chaque capsule de la deuxième couche. La somme pondérée  $s_j$  de tous les vecteurs de prédiction est par la suite calculée pour chaque capsule appartenant à la 3ème couche (eq.3.6) :

$$s_j = \sum_i c_{ij} \hat{u}_{j/i} \quad (3.6)$$

Ensuite, la fonction d'activation de squash est appliquée à cette somme pondérée afin que la norme du vecteur de sortie  $v_j$  soit comprises entre 0 et 1 (eq.3.7). Cette longueur représente la probabilité de l'existence de la caractéristique.

$$v_j = \frac{||s_j||^2}{1 + ||s_j||^2} \frac{s_j}{||s_j||} \quad (3.7)$$

Et finalement, le poids de routage  $b_{ij}$  est mis à jour en lui ajoutant l'agrément entre le vecteur réel et le vecteur prédit (eq.3.8).

$$b_{ij} = b_{ij} + a_{ij} \quad (3.8)$$

Tout le processus de la deuxième couche représente une itération de l'algorithme de routage. Dans le cas d'une bonne prédiction, le poids de routage  $b_{ij}$  augmente ce qui augmente la norme du vecteur de sortie pour la prochaine itération ainsi que la probabilité d'existence de l'entité représentée par le vecteur. Ensuite cette norme du vecteur est utilisée pour calculer la probabilité que l'entité existe en calculant la perte de marge  $L_k$ . Une perte de marge est calculée pour chaque classe  $k$  :

$$L_k = T_k \cdot \max(0, m^+ - ||v_k||)^2 + \lambda(1 - T_k) \max(0, ||v_k|| - m^-)^2 \quad (3.9)$$

Où :

- $T_k = 1$  si l'entité est présente
- $||v_k||$  représente la norme du vecteur de sortie
- $m^+$  et  $m^-$  sont des hyper-paramètres qui égalent respectivement à 0.9 et 0.1

—  $\lambda = 0.5$ .

Sur l'équation ci-dessus,  $T_k$  représente la présence ou l'absence de l'entité de la classe en question,  $m^+$  et  $m^-$  sont mis en place pour borner la longueur du vecteur de la fonction de perte. Le coefficient  $\lambda$  est mis à 0.5 afin de contrôler l'effet de la baisse du poids des classes absentes sur les décisions du modèle.

## 3.4 Bases de données

### 3.4.1 Base SafeRoad

Dans cette partie nous introduisons une nouvelle base de données nommée SafeRoad [29, 32]. C'est la première base de données marocaine permet l'analyse de l'orientation du piéton. Elle contient des images de piétons collectées durant 3 heures d'acquisitions depuis plusieurs villes marocaines et contenant des scènes de routes réelles. Elles sont prises avec deux caméras monoculaires positionnées à l'intérieur d'un véhicule en mouvement. La première est une caméra industrielle de référence DFK33UX174 d'une résolution de 2.3MP avec un capteur CMOS avec une fréquence pouvant aller jusqu'à 60 images par seconde illustré sur la figure 3-4. Néanmoins pour une bonne qualité nous avons choisi de limiter la fréquence à 30 images par seconde. La deuxième est une caméra GoPro Hero 5 avec une résolution vidéo de 8.3 MP et une fréquence de 30 fps.

#### Annotation de la base

Notre base de données est manuellement annotée pour traiter la reconnaissance de l'orientation du piéton, deux versions de la base sont disponibles. La première version contient 5160 piétons labellisés sous 4 orientations (avant, arrière, gauche, droit) comme représenté sur la figure 3-5. 4160 images ont été utilisées pour l'apprentissage et 1000 images pour le test.

La deuxième version contient 8894 images de piéton annotées en 8 directions représentées sur la Figure 3-6 (avant, arrière, gauche, droit, avant-droit, avant-gauche,



FIGURE 3-4 – Image représentant l’installation de la caméra utilisée à l’intérieur du véhicule pour l’acquisition de la base SafeRoad



FIGURE 3-5 – Annotation des images de piétons de la base SafeRoad en 4 directions (avant, arrière, gauche, droite). Nous avons divisé la base suivant un ratio de 80/10/10 pour l’apprentissage, la validation et le test.

La majorité de la base a été collectée pendant la journée, seulement quelques vidéos ont été enregistrées pendant le soir. Les conditions météorologiques de nos acquisitions sont variables, allant d’un soleil éblouissant, passant par un ciel clair, jusqu’à un temps nuageux ou même pluvieux comme illustré sur la figure 3-7. Ces différents points font que la détection du piéton ainsi que la classification de sa direction une tâche difficile.

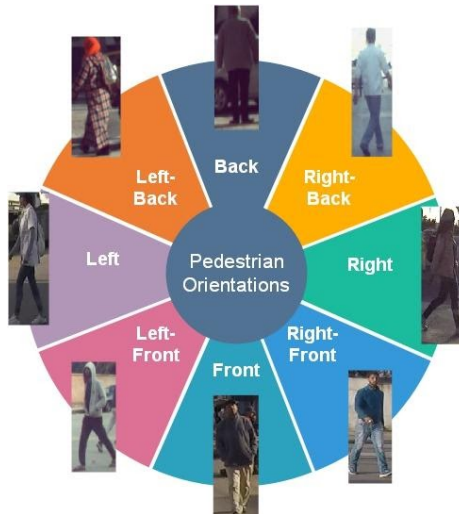


FIGURE 3-6 – Annotation des images de piétons de la base SafeRoad en 8 directions

Il faut noter que la majorité des piétons de notre base sont capturés en train de traverser la route. La variation d'échelle de la hauteur des piétons, en pixels, dépend de la distance entre la caméra et le piéton, autrement dit, elle dépend de la position du véhicule par rapport au piéton. Pour normaliser la taille des imagerie de piétons, nous avons choisi de les redimensionner en 48x48 pixels. L'originalité dans notre base de données est associée à la présence d'une variation du comportements du piéton. On trouve des piétons disciplinés respectant leur passage comme le montre la figure 3-8, et d'autres partageant la route avec les véhicules et traversant sur la chaussée mettant ainsi leur vie en danger malgré la présence d'installations piétonnes représentées sur la figure 3-9.

Les variations existent aussi au niveau de l'aménagement des routes, certaines sont bien structurées (Figure 3-10), alors que d'autres le sont moins, manquant de services de sécurité de base nécessaire pour protéger les piétons contribuant ainsi à l'augmentation des accidents routiers impliquant des piétons, comme illustré sur la figure 3-11.

Dans l'ensemble nous pouvons conclure que le comportement de l'utilisateur le plus vulnérable qu'est le piéton peut être défini par risqué et aléatoire, s'il traverse dans tous les sens en partageant la chaussée avec des véhicules, négligeant la présence de conducteurs imprudent et mettant ainsi leur vie en danger, comme cela est le montre



(a) Temps pluvieux le soir.



(b) Ciel clair.



(c) Temps nuageux.



(d) Soleil éblouissant.

FIGURE 3-7 – Les différentes conditions d’acquisition de la base SafeRoad



FIGURE 3-8 – Image représentant des piétons disciplinés prise de la base SafeRoad



FIGURE 3-9 – Image représentant des piétons indisciplinés prise de la base SafeRoad

la Figure 3-12. D’après nos connaissances aucune des bases de données existantes ne



FIGURE 3-10 – Image représentant une route structurée de la base SafeRoad



FIGURE 3-11 – Image représentant une route non structurée de la base SafeRoad

contient une variété des comportements des piétons comme cela est le cas pour la base SafeRoad, avec un apport non négligeable de piétons indisciplinés et de routes non structurées, ce qui rend la base SafeRoad beaucoup plus réaliste que les autres bases existantes.

### 3.4.2 Base Daimler

Nous avons évalué également notre réseau en utilisant la base publique Daimler [39]. Elle contient 13.000 images de piétons de dimension 48x96px, que nous avons redimensionné à 48x48px pour des raisons de ressources de calcul. Nous avons utilisé 12.000 images pour l'apprentissage et 1000 pour le test. La base Daimler ne contient aucune annotation sur l'orientation du piéton, donc nous avons procédé à une annotation manuelle de la base pour 4 directions. L'état d'acquisition de la base Daimler ressemble à ceux de notre base SafeRoad, les images sont prises à partir d'un véhicule en mouvement dans un trafic urbain, ce qui rend la base réaliste (Figure 3-13). Sauf





(a)



(b)



(c)



(d)

FIGURE 3-12 – Représentation des différents comportements des piétons acquisent dans la base SafeRoad



que, les conditions de routes et le comportement des piétons ne sont pas les mêmes que dans la base SafeRoad. Contrairement à la base SafeRoad, dans la base Daimler le comportement des piétons est ordonné, ce qui simplifie la détection de la direction de marche.



FIGURE 3-13 – Images de piétons de la base Daimler

### 3.4.3 Base TUD Multiview Pedestrians

La base TUD [6] est une base de données publique traitant l'estimation de pose chez le piéton, contenant 5228 images pour l'apprentissage annotées en 8 directions (droite, gauche, avant, arrière, gauche-arrière, droite-arrière, gauche-devant, droite-devant) avec une différence de 45 degré entre chaque direction, comme représenté sur la figure 3-14. TUD est largement utilisée dans les études de recherche notamment pour l'estimation de l'orientation du piéton. Une des raisons de l'utilisation de cette base est sa conformité avec le nombre d'orientations utilisé dans notre base de données SafeRoad.

Le modèle proposé est évalué en utilisant les trois bases citées précédemment. Les bases TUD et Daimler représentent un modèle parfait des piétons, par rapport à la base SafeRoad où les comportements à risque sont prédominants, avec des piétons traversant de façon aléatoire.



FIGURE 3-14 – Images de piétons de la base TUD

## 3.5 Reconnaissance de l’Orientation du piéton (4 Orientations)

### 3.5.1 L’architecture proposée

Notre système de classification de l’orientation du piéton [29] permet de classer l’orientation du piéton vers une des quatre directions suivantes (Droite, Gauche, Devant, Arrière). Chaque direction représente une classe pour notre système basé sur les réseaux de capsules (CapsNet). Comme nous l’avons déjà cité, les réseaux de Capsules contiennent deux parties qui sont l’encodage et le décodage comme définies dans la section 1.3. Nous avons entraîné cinq architectures de réseaux de capsules représentés ci-dessous avant de choisir le modèle final.

- Les trois premières architectures contiennent une seule couche de convolution, une couche de capsule primaire et une couche dense, la première couche de convolution possède 256 filtres et 32 capsules sur la couche de capsule primaire, la deuxième architecture contient 128 filtres et 16 capsules, alors que la troisième possède 64 filtres et 8 capsules.
- La quatrième architecture que nous avons testée contient deux couches de convolution de 256 et 128 filtres respectivement et 16 capsules dans la couche de capsule primaire.
- La cinquième et dernière architecture contient deux couches de convolution de 64 et 128 filtres respectivement et 16 capsules. Cette dernière architecture a obtenu le plus grand taux de précision, et est de ce fait le modèle que nous avons choisi.

Nous avons tout d’abord testé l’architecture proposée par [122] comme première

architecture. Contrairement aux résultats obtenu par [122] pour classer la base MNIST, cette architecture a obtenu le taux d'exactitude le plus faible pour notre étude. Ceci peut être expliqué par la différence des objets à classer, où la classification des piétons nécessite une extraction de caractéristiques plus importantes. A cette fin, nous avons proposé l'utilisation de deux couches de convolution pour l'extraction des caractéristiques comme représenté sur la cinquième architecture permettant ainsi d'avoir un meilleur résultat.

Ci-après nous allons détailler la partie encodage de notre modèle choisi (cinquième architecture) comme représenté sur la Figure 3-15.

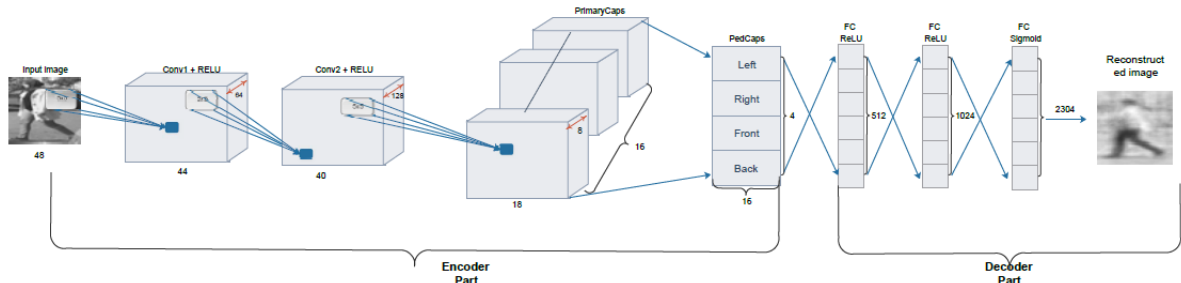


FIGURE 3-15 – Architecture CapsNet utilisé pour détecter l'orientation du piéton

- La première couche est une couche de convolution de 64x5x5 filtres et d'un pas de 1. Elle prend en entrée une image en niveau de gris d'une dimension de 48x48 pixels, et donne en sortie un tenseur de taille 44x44x64. Cette première couche permet d'extraire les caractéristiques de l'image d'entrée pour les envoyer à la deuxième couche.
- La deuxième couche est une couche de convolution de 128x5x5 filtres et d'un pas de 1. Elle donne en sortie un tenseur de taille 40x40x128 qui sera utilisé comme entrée pour la couche suivante.
- La troisième couche représente la couche primaire des capsules, elle contient 16 capsules de 8 dimensions. Chaque capsule appartenant à cette couche reçoit en entrée les caractéristiques extraites par les deux couches de convolution, soit une dimension totale de 40x40x128x16.
- La dernière couche que nous avons nommée PedCaps, contient 4 capsules de

16 dimensions. Chaque capsule fait référence à une classe parmi les 4 classes d'orientations. Cette couche permet de classer l'image d'entrée et de l'attribuer à une des 4 classes susmentionnées.

En ce qui concerne la partie de décodage, elle contient 3 couches denses de 512, 1024 et 2304 filtres. Elle sert à reconstruire l'image d'entrée en utilisant le résultat de classification de la couche PedCaps. La partie décodage permet de calculer la perte de reconstruction en calculant la distance euclidienne entre l'image de reconstruction et l'image d'entrée.

### 3.5.2 Expérimentations et Résultats

Le processus d'apprentissage pour tous les modèles a été réalisé en utilisant le framework d'apprentissage profond Tensorflow 1.15. Pour entraîner le réseau, nous avons utilisé un apprentissage de 50 itérations avec un lot de taille 64. L'exactitude représente le pourcentage des vrais résultats par rapport au nombre total des cas examinés.

Pour évaluer la performance de chaque modèle, l'exactitude est calculée en divisant les vrais positifs sur l'ensemble des prédictions de la base du test (eq.3.3).

#### Evaluation sur la base Daimler

L'approche proposée a été évaluée en utilisant la base SafeRoad et la base Daimler. Le tableau 3.4 représente les résultats du test pour chaque modèle des cinq architectures cités plus haut sur la base Daimler.

Selon le Tableau 3.4, une architecture avec deux couches de convolution donne un meilleur résultat que celles avec une seule couche de convolution. La réduction du nombre des filtres et du nombre de capsules utilisées améliore également le résultat final. L'exactitude a augmenté de plus de 6% en réduisant le nombre de filtres de 256 à 128, et le nombre des capsules de 32 à 16. Cependant, l'architecture donnant le meilleur résultat est la cinquième architecture avec une exactitude de 97.60% et une perte de 0.014. La perte totale représentée par (eq :3.10) est calculée en rassemblant la

TABLE 3.4 – Classification de l’orientation du piéton basé sur différentes architectures sur la base Daimler

| Architecture | Nb de couches de convolution | Nombre de filtres        | Nombre de capsules primaire | Perte totale | Exactitude    |
|--------------|------------------------------|--------------------------|-----------------------------|--------------|---------------|
| A1           | 1                            | 256                      | 32                          | 0.07         | 90,62%        |
| A2           | 1                            | 128                      | 16                          | 0.016        | 96,87%        |
| A3           | 1                            | 64                       | 8                           | 0.02         | 96,66%        |
| A4           | 2                            | Conv1 :256<br>Conv2 :128 | 16                          | 0.06         | 95.20%        |
| A5           | 2                            | Conv1 :64<br>Conv2 :128  | 16                          | 0.014        | <b>97.60%</b> |

perte de marge (eq :3.9) et la perte de reconstruction. Cette dernière est obtenue par le calcul de la distance euclidienne entre l’image de reconstruction et l’image d’entrée.

$$Perte\ totale = Perte\ de\ marge + \lambda(Perte\ de\ reconstruction) \quad (3.10)$$

Avec  $\lambda = 0.0005$  représente l’échelle de régularisation.

La figure 3-16 représente la perte de marge, la perte de reconstruction ainsi que la perte totale, obtenues lors de l’apprentissage en utilisant la base Daimler. L’abscisse représente le nombre d’époques durant le processus d’apprentissage, alors que l’ordonnée représente le taux de perte.

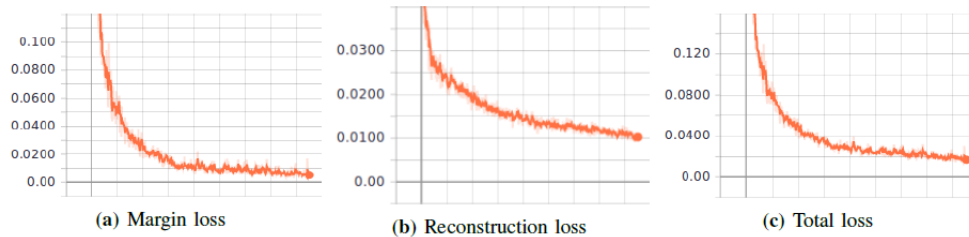


FIGURE 3-16 – Représentation de la perte de marge d’une valeur de 0.004, la perte de reconstruction d’une valeur de 0.1 ainsi que la perte totale d’une valeur de 0.014

La figure 3-17 représente un échantillon des images reconstruites prises de la base Daimler. Ces images montrent le rôle des réseaux de Capsule à préserver les détails de l’image d’entrée y compris l’orientation des piétons pour notre cas.

Afin de tester la performance des réseaux de Capsules (CapsNet) par rapport aux autres réseaux, nous avons comparé notre approche à certaines architectures CNN

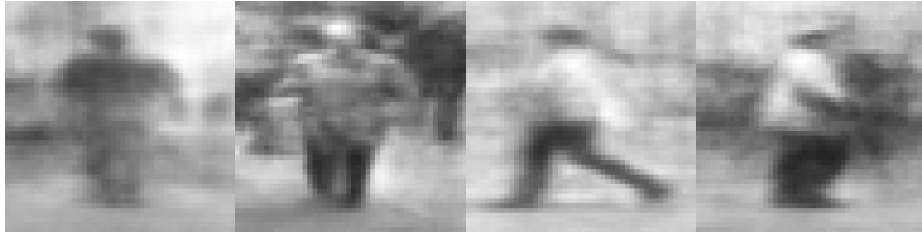


FIGURE 3-17 – Exemple d’images reconstruites des réseaux de Capsules durant la phase d’apprentissage

largement utilisées pour classer les images. Le tableau 3.5 montre les résultats obtenus pour les architectures AlexNet et ResNet sur la base Daimler avec une exactitude de 95.52% et 96.45% respectivement alors que l’exactitude obtenue par les réseaux de Capsules a dépassé ces deux résultats obtenues avec une précision de 97.60%. Le modèle est évalué en calculant le pourcentage des vraies positives, ce qui représente l’exactitude obtenue par l’eq.3.3. La vérité terrain dont nous disposons est sous forme d’images de piétons labellisées par leurs orientations. Si la prédiction du modèle sur une image de piéton convient à l’orientation dont nous disposons avec la vérité terrain, elle constitue une vraie positive.

TABLE 3.5 – Comparaison de l’exactitude entre les réseaux AlexNet, ResNet et CapsNet pour la base Daimler

| Réseau  | Exactitude    |
|---------|---------------|
| AlexNet | 95.52%        |
| ResNet  | 96.45%        |
| CapsNet | <b>97.60%</b> |

L’utilisation de la matrice de confusion permet d’indiquer le nombre de prédictions correctes et incorrectes pour chaque classe, et ce, en calculant les vraies positives, les fausses positives, les vraies négatives et les fausses négatives pour chaque classe.

- Vraie Positive : Indication correcte d’une prédiction positive
- Vraie Négative : Indication correcte d’une prédiction négative
- Fausse Positive : Indication incorrecte d’une prédiction positive
- Fausse Négative : Indication incorrecte d’une prédiction négative

La matrice de confusion nous permet ainsi d'analyser la relation entre chaque classe prédite et les autres classes. La Figure 3-18 représente les matrices de confusion pour chaque réseau (CapsNet, AlexNet et ResNet) pour la base Daimler. Nous pouvons remarquer que les réseaux Capsules sont légèrement plus performants que les réseaux AlexNet et ResNet. D'après les matrices de confusion, les fausses prédictions sur la base Daimler sont légèrement présentes, sauf pour la classe « Back (Arrière) » qui a une fausse prédiction comme étant la classe « Front (Avant) » avec un pourcentage de 4%, 8% et 7% respectivement pour les réseaux CapsNet, AlexNet et ResNet. Ceci peut expliquer la difficulté du réseau à différencier correctement entre une personne se dirigeant vers l'avant et une autre vers l'arrière.

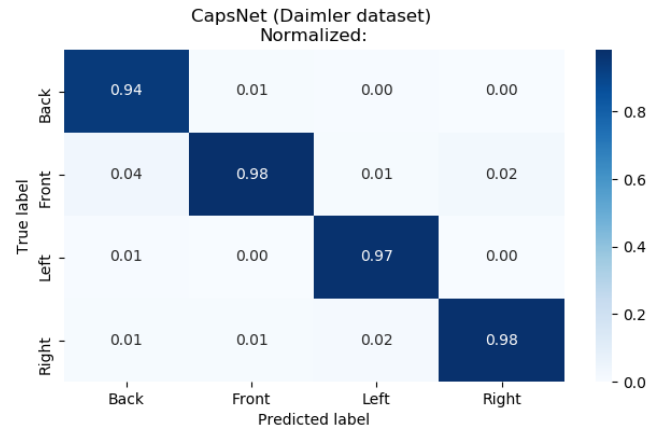
Le tableau 3.6 représente une comparaison des métriques d'évaluation de performance entre les réseaux CapsNet, AlexNet et ResNet pour la base Daimler. Les métriques utilisées sont la précision, la spécificité, le F1-Score et le recall. La précision représente le pourcentage des vrais positifs par rapport à toutes les prédictions positives pour chaque classe. Elle est calculée selon l'équation suivante (eq. 3.11) :

$$Precision = \frac{TP}{TP + FP} \quad (3.11)$$

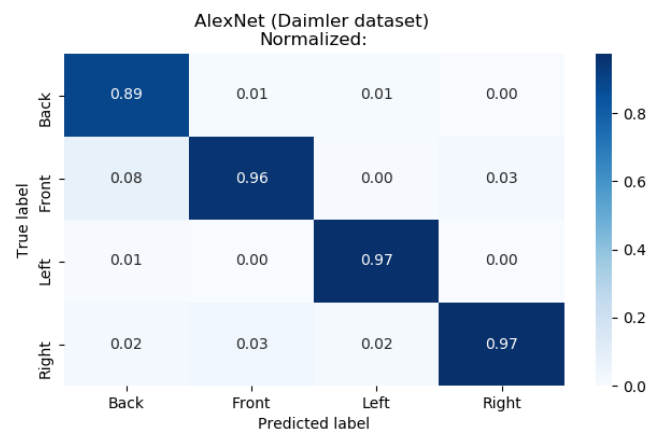
TABLE 3.6 – Comparaison des métriques de classification pour chaque classe de la base Daimler

|         | Classes<br>Métriques | Arrière     | Avant       | Gauche      | Droite      |
|---------|----------------------|-------------|-------------|-------------|-------------|
| CapsNet | Précision            | <b>0.94</b> | <b>0.98</b> | <b>0.97</b> | 0.98        |
|         | Recall               | <b>0.99</b> | <b>0.93</b> | <b>0.99</b> | <b>0.96</b> |
|         | Spécificité          | <b>0.98</b> | <b>0.99</b> | <b>0.99</b> | <b>0.99</b> |
|         | F1-Score             | <b>0.96</b> | <b>0.95</b> | <b>0.98</b> | <b>0.97</b> |
| AlexNet | Précision            | 0.89        | 0.96        | <b>0.97</b> | 0.97        |
|         | Recall               | 0.98        | 0.88        | <b>0.99</b> | 0.93        |
|         | Specificity          | 0.96        | <b>0.99</b> | <b>0.99</b> | <b>0.99</b> |
|         | F1-Score             | 0.94        | 0.92        | <b>0.98</b> | 0.95        |
| ResNet  | Précision            | 0.84        | <b>0.98</b> | 0.96        | <b>0.99</b> |
|         | Recall               | 0.98        | 0.89        | <b>0.96</b> | 0.91        |
|         | Specificity          | 0.94        | <b>0.99</b> | 0.98        | <b>0.99</b> |
|         | F1-Score             | 0.91        | 0.93        | 0.96        | 0.95        |

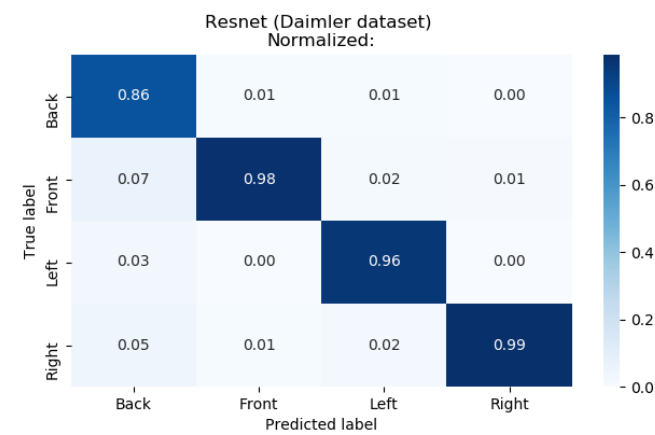
D'après le tableau 3.6, les réseaux de Capsules dépassent les autres réseaux pour



(a) Matrice de Confusion de CapsNet sur la base Daimler



(b) Matrice de Confusion de AlexNet sur la base Daimler



(c) Matrice de Confusion de ResNet sur la base Daimler

FIGURE 3-18 – Matrice de Confusion des réseaux CapsNet, AlexNet et ResNet sur la base Daimler

presque toutes les métriques sauf au niveau de la précision où le réseau ResNet possède une augmentation de 0.01 par rapport au Capsule sur la classe "Droite".



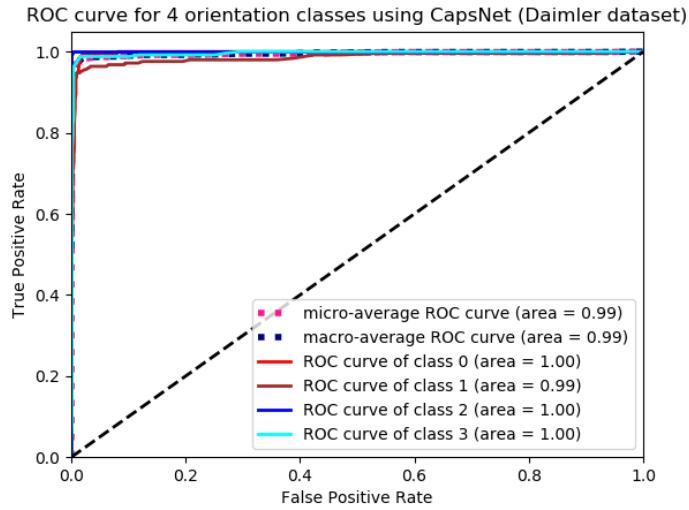


FIGURE 3-19 – ROC curve of Capsule Network on Daimler dataset

### Evaluation sur la base SafeRoad

Comme sur la base Daimler, les réseaux de Capsules évalués sur la base SafeRoad ont obtenu la meilleure exactitude par rapport aux réseaux ResNet et AlexNet, avec respectivement une précision de 73.64%, 73.12% et 70.10% comme représenté sur le Tableau 3.7.

TABLE 3.7 – Comparaison des résultats entre les réseaux AlexNet, ResNet et CapsNet pour la base SafeRoad

| Réseau  | Exactitude    |
|---------|---------------|
| AlexNet | 70.10%        |
| ResNet  | 73.12%        |
| CapsNet | <b>73.64%</b> |

Nous pouvons également observer que la base Daimler dépasse largement en terme d'exactitude la base SafeRoad pour les trois approches. Ceci peut être expliqué par la différence qui existe entre les deux bases dans l'environnement d'acquisition, ce qui engendre une différence sur le comportement du piéton et sa façon de traverser. Les images de la base Daimler représentent des piétons traversant généralement de façon horizontale ce qui fait que la direction est bien claire. Par contre les images de notre base de données SafeRoad contiennent des scènes prises depuis des environnements

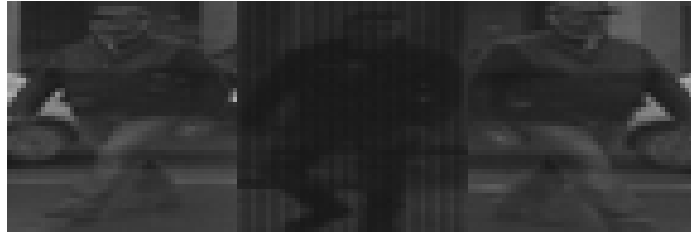


FIGURE 3-20 – Exemple d’images représentant une mauvaise classification



FIGURE 3-21 – Pedestrian orientation recognition using Capsule Networks on Safe-Road dataset

non structurés dans le milieu urbain contenant des piétons avec un comportement à risque traversant d’une manière oblique qui ne peut être représenté par les quatre orientations citées plus haut (Gauche, Droite, Avant, Arrière) produisant ainsi des erreurs de classification. Ces mauvaises classifications sont surtout remarquées entre deux directions adjacentes. La Figure 3-20 illustre des exemples de mauvaise classification où un piéton se dirigeant en Arrière est classifié à Droite, et deux directions de Gauche représentées comme étant Avant.

La figure 3-21 illustre le résultat de la reconnaissance de l’orientation du piéton par les réseaux de Capsules sur la base SafeRoad. La figure montre que la reconnaissance de l’orientation est correcte pour plusieurs piétons à la fois, chaque piéton est traité indépendamment en utilisant un système de suivi.

Les matrices de confusion des réseaux CapsNet, AlexNet et ResNet illustrées sur la Figure 3-22 démontrent la performance du réseau de Capsules à classer correctement

l'orientation du piéton pour la base SafeRoad. Comme avec la base Daimler, la classe "Arrière (Back)" est la classe ayant le plus grand pourcentage de fausses estimations dans CapsNet arrivant à 34%. Nous pouvons également remarquer que le réseau CapsNet surpasse au niveau des prédictions correctes les autres réseaux mis à part la classe "Back" où le réseau AlexNet performe mieux avec un pourcentage de 72% de classification correcte contre 66% pour CapsNet. Le tableau 3.8 représente d'autres métriques de classification pour la base SafeRoad telles que le Recall, Spécificité et F1-score.

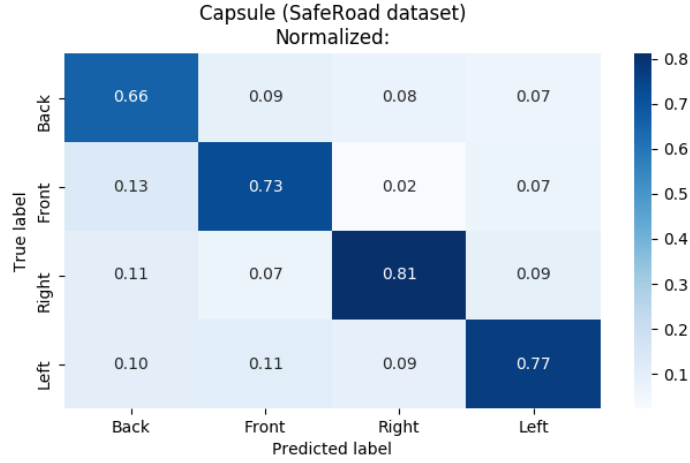
TABLE 3.8 – Comparaison des métriques de classification pour chaque classe de la base SafeRoad

|         | Classes<br>Métriques | Arrière     | Avant       | Droite      | Gauche      |
|---------|----------------------|-------------|-------------|-------------|-------------|
| CapsNet | Précision            | 0.66        | 0.73        | 0.71        | <b>0.77</b> |
|         | Recall               | <b>0.78</b> | <b>0.76</b> | 0.72        | 0.68        |
|         | Specificity          | 0.87        | 0.90        | <b>0.94</b> | <b>0.93</b> |
|         | F1-Score             | <b>0.71</b> | <b>0.74</b> | 0.76        | 0.72        |
| AlexNet | Précision            | <b>0.72</b> | 0.69        | 0.69        | 0.71        |
|         | Recall               | 0.64        | 0.75        | 0.74        | 0.67        |
|         | Specificity          | <b>0.91</b> | 0.89        | 0.89        | 0.91        |
|         | F1-Score             | 0.68        | 0.72        | <b>0.71</b> | 0.69        |
| ResNet  | Précision            | 0.66        | <b>0.75</b> | <b>0.80</b> | 0.73        |
|         | Recall               | 0.68        | 0.71        | <b>0.77</b> | <b>0.77</b> |
|         | Specificity          | 0.89        | <b>0.92</b> | 0.93        | 0.90        |
|         | F1-Score             | 0.67        | 0.73        | <b>0.78</b> | <b>0.75</b> |

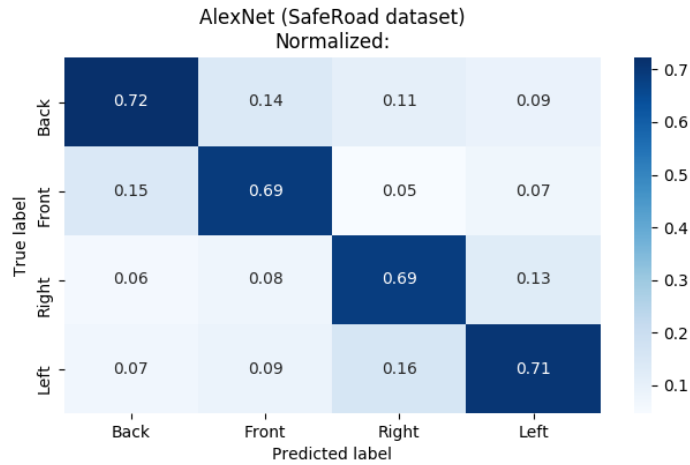
### 3.5.3 Comparaison avec différents algorithmes d'optimisation pour la base SafeRoad

#### Les algorithmes d'Optimisation

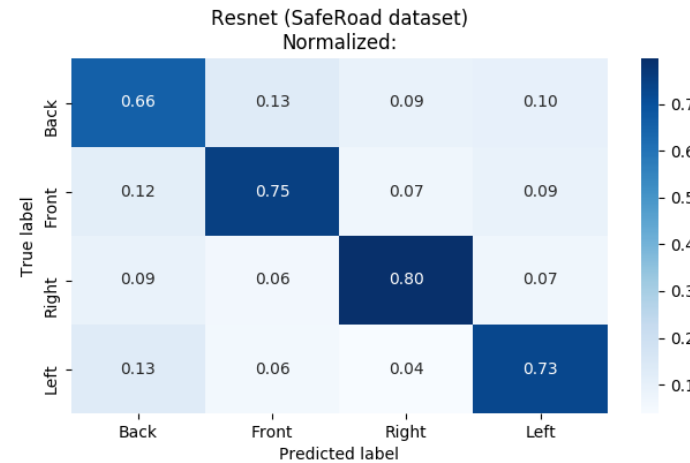
Les algorithmes d'optimisation ont un effet significatif sur les performances du réseau neuronal ; leur objectif est de trouver une solution optimale pour minimiser la perte du réseau pendant le processus d'apprentissage. Parmi les algorithmes existants, Gradient Descent est le plus populaire, le gradient est calculé à chaque itération de l'apprentissage. Par la suite, les poids et les biais sont mis à jour afin de minimiser la



(a) Matrice de Confusion de CapsNet sur la base SafeRoad



(b) Matrice de Confusion de AlexNet sur la base SafeRoad



(c) Matrice de Confusion de ResNet sur la base SafeRoad

FIGURE 3-22 – Matrices de Confusion Matrices des réseaux CapsNet, AlexNet et ResNet pour la base SafeRoad

TABLE 3.9 – Comparaison avec différents algorithmes d’optimisation pour la base SafeRoad

| Optimiseurs<br>Architectures | AdamOptimizer | Gradient<br>Descent | Adadelata | Adagrad | RMSProp       |
|------------------------------|---------------|---------------------|-----------|---------|---------------|
| CapsNet                      | 73.64%        | 75.41%              | 66.87%    | 76.87%  | <b>78.95%</b> |
| AlexNet                      | 70.10%        | 77.60%              | 61.35%    | 74.37%  | 77.60%        |
| ResNet                       | 73.12%        | 78.12%              | 56.66%    | 75.41%  | 78.54%        |

fonction de perte et atteindre la valeur minimale globale. Le RMSProp est un autre optimiseur qui aide à trouver le minimum global en ajustant le taux d’apprentissage pour chaque étape d’itération et restreint l’oscillation dans une direction afin de converger plus rapidement. L’optimiseur Adagrad adapte également le taux d’apprentissage en fonction des paramètres en se référant aux gradients passés et en utilisant la moyenne de désintégration de tous les gradients carrés passés, tandis que son extension Adadelata limite les gradients passés à une taille fixe. L’estimation du moment adaptatif connue par AdamOptimizer calcule le taux d’apprentissage adaptatif pour chaque paramètre et stocke une moyenne de décroissance exponentielle des gradients passés. Les algorithmes d’optimisation précités sont ceux utilisés dans ce travail pour améliorer les performances du réseau.

## Résultat

Bien qu’il possède le temps d’apprentissage le plus long, le réseau de Capsules donne le meilleur résultat par rapport aux réseaux CNN (AlexNet et ResNet).

Afin d’améliorer le résultat pour la base SafeRoad, nous avons entraîné le réseau en utilisant différents algorithmes d’optimisation tels que Gradient Descent, Adadelata, Adagrad et RMSProp que nous avons comparé avec les résultats déjà obtenus en utilisant l’algorithme d’optimisation AdamOptimizer. Le tableau 3.9 représente une comparaison [32] de ces algorithmes pour chaque approche (CapsNet, AlexNet and ResNet) où l’exactitude a augmenté de 5% pour les réseaux de Capsules en utilisant l’algorithme RMSProp, allant de 73.64% pour AdamOptimizer arrivant à 78.95% pour RMSProp avec un taux d’apprentissage de 0.01.

## 3.6 Reconnaissance de l'orientation du piéton (8 orientations)

### 3.6.1 Architecture

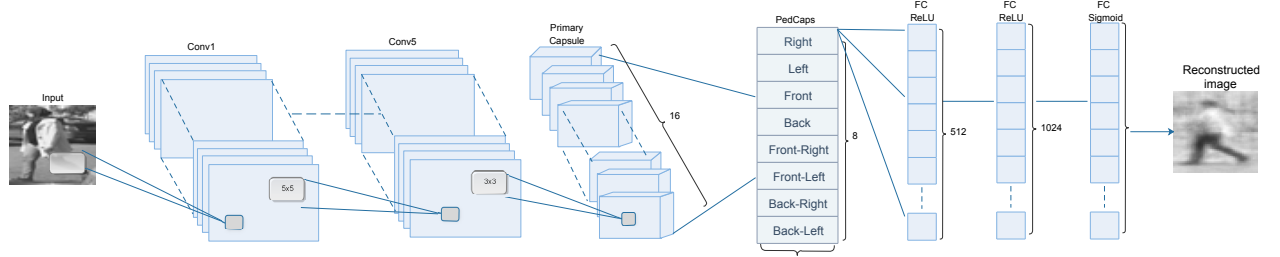


FIGURE 3-23 – CapsNet architecture pour la détection de 8 orientations

Après plusieurs évaluations nous avons proposé pour le cas de classification de 8 orientations, une architecture du réseau des capsules légèrement différente de celle proposée pour la classification de 4 orientations [31]. La différence majeure se situe sur l'utilisation de cinq couches de convolution au lieu de deux pour l'architecture antérieure. La première couche de convolution est de 64x5x5 filtres alors que les quatre autres couches sont de 128x3x3.

### 3.6.2 Evaluation sur la base SafeRoad

De même pour cette architecture, les réseaux de Capsule ont obtenu la meilleure précision par rapport aux autres réseaux CNN, AlexNet et ResNet. Le tableau 3.10 représente la précision des réseaux de capsule appliquée sur la base SafeRoad pour les 8 orientations. En terme de temps, l'apprentissage des réseaux de capsule est le plus long avec 15h d'apprentissage pour la base SafeRoad, alors que ResNet et AlexNet ont pris respectivement 7h et 24 min, exécutés sur une machine Nvidia GeForce 1510 GTX avec une mémoire de 4 GB, 6 coeurs et un lot d'apprentissage de 16. L'évaluation des performances sur la base SafeRoad, a montré que les réseaux de Capsules ont obtenu une précision de 70.22% sur les directions prédites qui ont la même direction

que la vérité terrain, et une précision de 83.5% pour les directions prédites qui ont la même direction ou une direction adjacente de la vérité terrain, comme représenté sur le Tableau 3.10.

TABLE 3.10 – Comparaison de la performance de différentes architectures sur la base SafeRoad

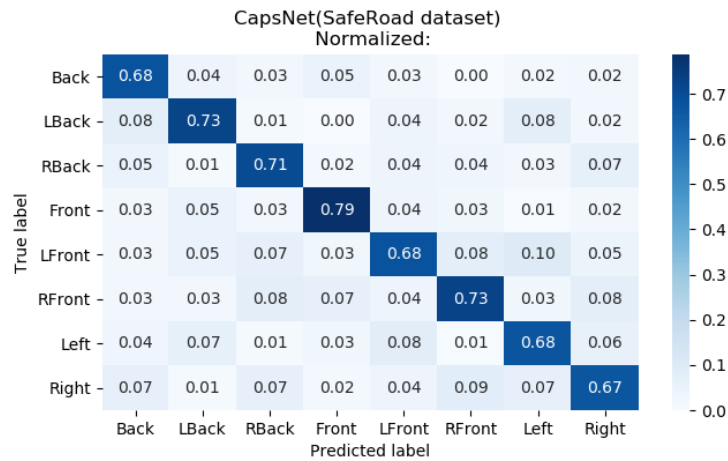
| Architectures | Exactitude pour les directions prédites ayant la même direction que la vérité terrain | Précision pour les directions prédites ayant la même direction ou une direction adjacente que la vérité terrain | Temps d'apprentissage |
|---------------|---------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|-----------------------|
| CapsNet       | <b>70.22%</b>                                                                         | <b>83.5%</b>                                                                                                    | 15h                   |
| AlexNet       | 65.34%                                                                                | 82.12%                                                                                                          | 7h                    |
| ResNet        | 54.43%                                                                                | 73.75%                                                                                                          | 24min                 |

La Figure 3-24 et le tableau 3.11 représentent respectivement la matrice de confusion et les différentes métriques d'évaluation pour les réseaux CapsNet, AlexNet et ResNet sur la base SafeRoad. La performance des réseaux de capsules dépassent clairement celle des autres réseaux pour presque toutes les classes utilisées mis à part la classe "Droite (Right)" où AlexNet performe légèrement mieux avec une précision de 72% contre 67% pour CapsNet. Également pour les autres métriques, les réseaux de Capsules ont une meilleure performance par rapport aux réseaux AlexNet et ResNet. La figure 3-25 illustre la courbe ROC des réseaux de Capsules sur la base SafeRoad, où les classes 0-7 représentent les huit orientations (Arrière, Gauche-Arrière, Droite-Arrière, Devant, Gauche-Devant, Droite-Devant, Gauche et Droite).

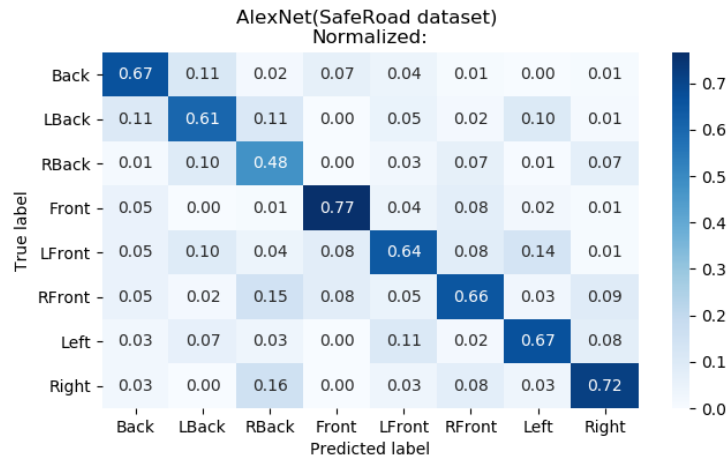
Nous considérons la base SafeRoad comme une base complexe pour la tâche de classification pour distinguer les directions de piétons en raison de la proximité de certaines orientations adjacentes. Ceci explique le taux d'erreurs de classification.

### 3.6.3 Evaluation sur la base TUD

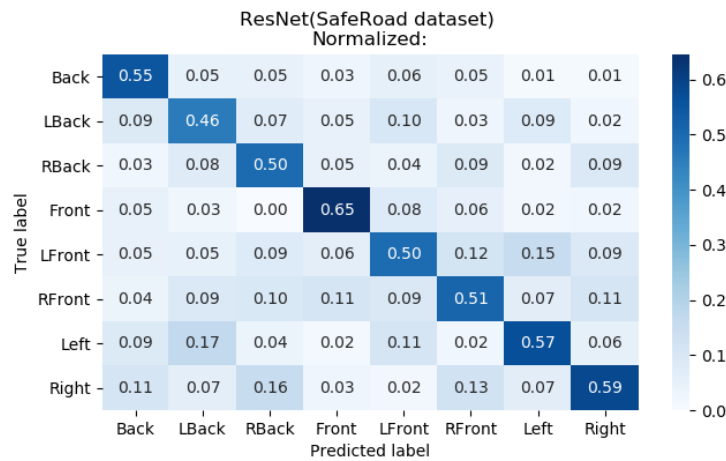
Pour s'assurer de la performance du réseau de Capsules sur la reconnaissance de l'orientation des piétons pour diverses bases de données, nous avons effectué l'apprentissage et le test pour une détection de huit orientations sur la base TUD. Iden-



(a) Confusion Matrice du réseau de Capsule sur la base SafeRoad



(b) Confusion Matrice de AlexNet sur la base SafeRoad



(c) Confusion Matrice de ResNet sur la base SafeRoad

FIGURE 3-24 – Les matrices de confusion des réseaux CapsNet, AlexNet et ResNet sur la base SafeRoad



ROC curve for the 8 orientations classes using CapsNet (SafeRoad dataset)

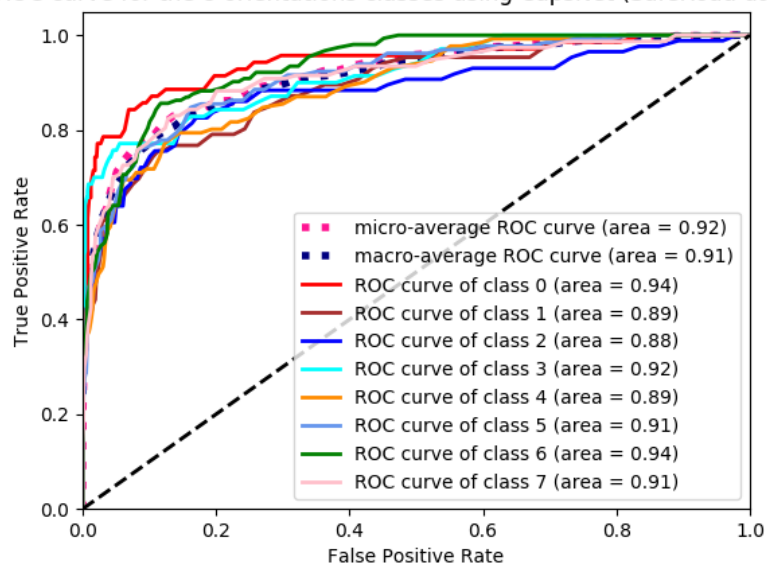


FIGURE 3-25 – Courbe ROC du réseau de Capsule sur la base SafeRoad

|         | Classes<br>Métriques | Arrière     | Gauche-Arrière | Droite-Arrière | Avant       | Gauche-Avant | Droite-Avant | Gauche      | Droite      |
|---------|----------------------|-------------|----------------|----------------|-------------|--------------|--------------|-------------|-------------|
| CapsNet | Précision            | <b>0.68</b> | <b>0.73</b>    | <b>0.71</b>    | <b>0.79</b> | <b>0.68</b>  | <b>0.73</b>  | <b>0.68</b> | 0.67        |
|         | Recall               | <b>0.74</b> | <b>0.62</b>    | 0.62           | <b>0.69</b> | <b>0.63</b>  | <b>0.70</b>  | <b>0.79</b> | 0.74        |
|         | Spécificité          | <b>0.97</b> | <b>0.97</b>    | <b>0.97</b>    | <b>0.98</b> | <b>0.94</b>  | <b>0.95</b>  | 0.91        | 0.92        |
|         | F1-Score             | <b>0.71</b> | <b>0.67</b>    | <b>0.66</b>    | <b>0.73</b> | <b>0.65</b>  | <b>0.72</b>  | <b>0.73</b> | <b>0.71</b> |
| AlexNet | Précision            | 0.67        | 0.61           | 0.48           | 0.77        | 0.64         | 0.66         | 0.67        | <b>0.72</b> |
|         | Recall               | 0.71        | 0.43           | <b>0.64</b>    | 0.66        | 0.58         | 0.60         | 0.75        | <b>0.75</b> |
|         | Spécificité          | <b>0.97</b> | <b>0.97</b>    | 0.92           | <b>0.98</b> | <b>0.94</b>  | 0.94         | <b>0.92</b> | <b>0.94</b> |
|         | F1-Score             | 0.69        | 0.50           | 0.55           | 0.71        | 0.61         | 0.62         | 0.71        | 0.73        |
| ResNet  | Précision            | 0.55        | 0.46           | 0.50           | 0.65        | 0.50         | 0.51         | 0.57        | 0.59        |
|         | Recall               | 0.63        | 0.41           | 0.48           | 0.57        | 0.40         | 0.50         | 0.69        | 0.61        |
|         | Spécificité          | 0.96        | 0.95           | 0.95           | 0.97        | 0.93         | 0.92         | 0.89        | 0.90        |
|         | F1-Score             | 0.59        | 0.43           | 0.49           | 0.61        | 0.45         | 0.50         | 0.62        | 0.60        |

TABLE 3.11 – Comparaison des métriques de classification metrics pour chaque classe de la base SafeRoad

TABLE 3.12 – Comparaison de la performance de classification sur la base TUD pour huit orientations.

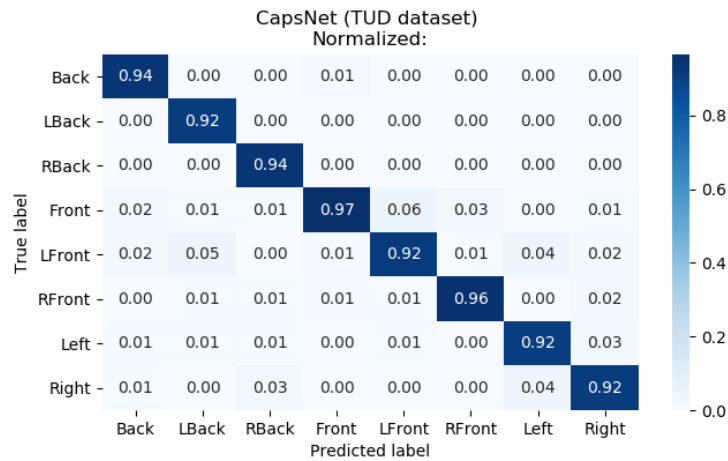
| Architectures | Exactitude    |
|---------------|---------------|
| CapsNet       | <b>93.48%</b> |
| AlexNet       | 91.79%        |
| ResNet        | 92.70%        |

|         | <del>Classes<br/>Métriques</del> | Arrière     | Gauche-<br>Arrière | Droite-<br>Arrière | Avant       | Gauche-<br>Avant | Droite-<br>Avant | Gauche      | Droite      |
|---------|----------------------------------|-------------|--------------------|--------------------|-------------|------------------|------------------|-------------|-------------|
| CapsNet | Précision                        | <b>0.94</b> | <b>0.92</b>        | 0.94               | <b>0.97</b> | 0.92             | <b>0.96</b>      | 0.92        | <b>0.92</b> |
|         | Recall                           | 0.99        | <b>1</b>           | <b>1</b>           | <b>0.87</b> | <b>0.85</b>      | <b>0.94</b>      | 0.93        | 0.92        |
|         | Spécificité                      | <b>0.99</b> | <b>0.99</b>        | <b>0.99</b>        | <b>1</b>    | <b>0.99</b>      | <b>0.99</b>      | <b>0.99</b> | <b>0.99</b> |
|         | F1-Score                         | <b>0.96</b> | <b>0.96</b>        | <b>0.97</b>        | <b>0.91</b> | <b>0.88</b>      | <b>0.95</b>      | <b>0.92</b> | <b>0.92</b> |
| AlexNet | Précision                        | 0.90        | 0.88               | 0.91               | <b>0.97</b> | 0.91             | 0.93             | 0.94        | <b>0.92</b> |
|         | Recall                           | <b>1</b>    | 0.99               | 0.99               | 0.80        | 0.82             | <b>0.94</b>      | 0.87        | 0.93        |
|         | Spécificité                      | 0.98        | 0.98               | <b>0.99</b>        | <b>1</b>    | <b>0.99</b>      | <b>0.99</b>      | <b>0.99</b> | <b>0.99</b> |
|         | F1-Score                         | 0.95        | 0.93               | 0.95               | 0.88        | 0.86             | 0.93             | 0.91        | <b>0.92</b> |
| ResNet  | Précision                        | 0.92        | 0.88               | <b>0.95</b>        | 0.94        | <b>0.96</b>      | 0.93             | 0.93        | 0.91        |
|         | Recall                           | 0.99        | <b>1</b>           | 0.97               | 0.82        | 0.84             | 0.92             | <b>0.94</b> | <b>0.95</b> |
|         | Spécificité                      | 0.99        | 0.98               | 0.99               | 0.99        | <b>1</b>         | 0.99             | 0.99        | 0.98        |
|         | F1-Score                         | 0.96        | 0.94               | 0.96               | 0.88        | 0.90             | 0.92             | 0.93        | 0.93        |

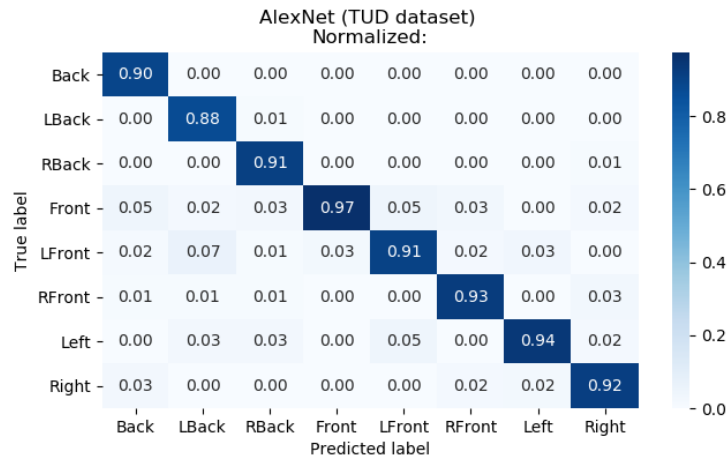
TABLE 3.13 – Comparaison des métriques de classification pour chaque classe de la base TUD

tiquement aux autres bases, le réseau de Capsules se distingue par sa performance au niveau des classifications correctes. Il surpasse les réseaux AlexNet et ResNet pour la majorité des classes utilisées comme le démontre la Figure 3-26 et le Tableau 3.13. La précision globale des réseaux de Capsule est de 93.48% contre 91.79% et 92.70% pour respectivement AlexNet et ResNet, comme représenté sur le tableau 3.12.

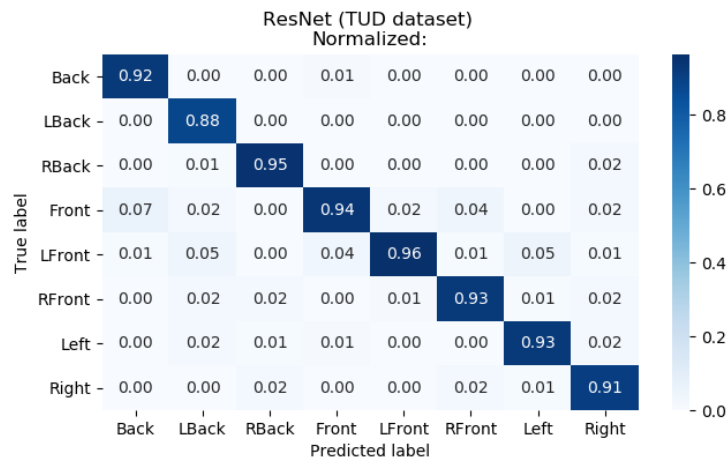
Le modèle donne une meilleure précision pour les bases TUD et Daimler, vu que les images de la base contiennent des piétons ayant des directions bien claires, faisant d'elle une base moins réaliste et facilitant ainsi la classification de l'orientation par le réseau. Cependant, malgré le taux de précision élevé lors de l'évaluation, un système entraîné sur des images de piétons idéales, serait incompatible avec le flux de traversée des piétons observé sur les routes réelles, et plus particulièrement les moins structurées où le comportement indiscipliné des piétons est dominant.



(a) Matrice de confusion du réseau de Capsule sur la base TUD



(b) Matrice de confusion de AlexNet sur la base TUD



(c) Matrice de confusion de ResNet sur la base TUD

FIGURE 3-26 – Matrices de confusion de CapsNet, AlexNet et ResNet sur la base TUD

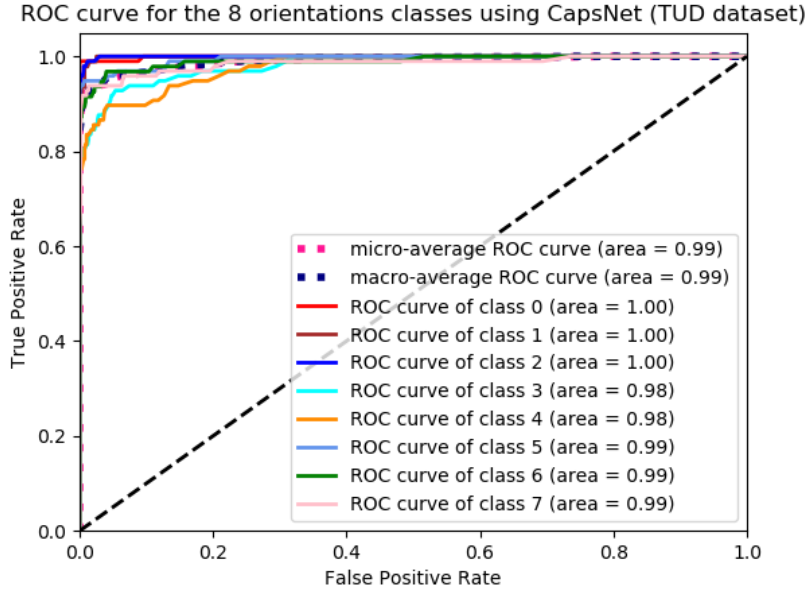


FIGURE 3-27 – Courbe ROC curve du réseau de Capsule sur la base TUD

### 3.7 Conclusion

Dans ce chapitre nous avons mis le point sur le conflit piéton-véhicule présent d'une façon exorbitante sur les routes souffrant d'un manque de structures pour la sécurité des piétons.

Afin d'étudier de près le comportement du piéton marocain nous avons créé la première base de données marocaine associée à l'orientation du piéton, qui représente la première contribution majeure de cette thèse. Elle a été collectée sous deux versions à partir des villes marocaines en utilisant une caméra à bord d'un véhicule en mouvement. La première version de la base contient 5160 images de piétons annotés en 4 orientations prises par une caméra industrielle de résolution 2.3MP, la deuxième version a été prise par une GoPro Hero 5 de résolution 8.3MP, contenant 8894 images de piétons annotées en 8 directions. Cette base a été utilisée pour l'apprentissage et le test du système de détection de l'orientation du piéton et comparée avec les bases publiques Daimler et TUD. Le système est entraîné par le réseau de Capsules, ce qui représente notre deuxième contribution.

L'approche vise à classer la direction du piéton en quatre orientations (Avant,

Droite, Gauche, Arrière) et en huit orientations (Avant, Arrière, Droite, Gauche, Gauche-Arrière, Droite-Arrière, Gauche-Avant, Droite-Avant). Le réseau de Capsules a obtenu les meilleurs résultats en terme de précision avec 97.60% sur la base Daimler pour une détection de 4 orientations et 78.95% sur notre base SafeRoad, par rapport aux réseaux AlexNet et ResNet sur la base Daimler dont les résultats sont respectivement 95.52% et 96.45%. Le réseau de Capsules a également surpassé les autres réseaux par la reconnaissance des 8 orientations avec une précision de 70.22% et 93.48% pour respectivement la base SafeRoad et TUD. Nous pouvons remarquer d'après les résultats que la base SafeRoad a une précision plus faible relativement à celle de Daimler et TUD. Cette différence des résultats est probablement due à la différence dans l'environnement de collection entre les bases.

La performance des réseaux de Capsules vis-à-vis des réseaux de Convolutions, peut être expliqué d'une part par la propriété de CapsNet qui lie les positions des caractéristiques de l'image, les neurones de chaque capsule représentent une propriété caractéristique de l'image notamment son orientation pour notre cas. D'autre part, la couche de pooling des réseaux de neurones Convolutifs compresse la taille des images, il en résulte une perte d'information contenant les propriétés de ces images.

Dans ce chapitre, nous avons traité l'orientation du piéton à partir d'une seule image ce qui ne représente pas le cas réel où le système capture une série d'images. De ce fait, nous allons représenter dans le chapitre suivant la reconnaissance de la direction de mouvement de chaque piéton apparaissant sur une séquence vidéo.

# Chapitre 4

## Reconnaissance de l'orientation du piéton dans une séquence vidéo

### 4.1 Introduction

L'étude de la reconnaissance de la direction du mouvement du piéton sur des séquences vidéos est une étape nécessaire avant son intégration sur les Systèmes Avancés d'Aide à la Conduite.

Dans les premiers chapitres, nous avons présenté les trois objectifs principaux de cette thèse, en proposant dans le deuxième chapitre un nouveau système d'alerte de collision entre piétons et véhicules impliquant l'orientation du piéton. Ensuite, nous avons introduit dans le chapitre précédent une approche de détection de l'orientation des piétons basée sur les réseaux de Capsules, entraînée sur les images de notre base de données SafeRoad qui représente notre 3ème proposition.

Dans ce chapitre, nous présentons la dernière contribution de cette thèse traitant la reconnaissance de la direction de marche du piéton sur des séquences vidéos, en utilisant la base publique "Join Attention for Autonomous Driving (JAAD) [73]" pour les expérimentations. Un des défis du traitement des scènes vidéos est l'attribution de la direction de marche pour chaque piéton indépendamment des autres piétons apparaissant dans la séquence, ce qui nécessite trois étapes de traitement dû à sa complexité par rapport à celle dans une image fixe, détaillées comme suit :

- La première étape est la détection du piéton et sa localisation dans l'image, pour cette étape nous avons choisi le détecteur d'objet temps réel YOLO [113].
- Chaque vidéo contient un ensemble de piétons en mouvement qui peuvent apparaître et disparaître successivement au sein de la même séquence, d'où la nécessité d'effectuer le suivi. Pour tout piéton détecté, est attribué un identifiant unique permettant d'effectuer son suivi pour les prochaines images de la séquence vidéo. A cette fin, l'algorithme DeepSort [142] est utilisé pour le suivi jusqu'à ce que le piéton quitte la séquence.
- Chaque piéton identifié est utilisé comme entrée au réseau de Capsules précédemment entraîné sur les images de la base SafeRoad et testé sur les vidéos de la base JAAD.

Ce chapitre est organisé comme suit : Section 2 introduit quelques approches existantes sur le suivi des objets et plus précisément les piétons dans des séquences vidéos. section 3 et 4 décrivent respectivement la base JAAD et les résultats des expérimentations faites sur la base JAAD. Section 5 représente notre conclusion.

## 4.2 Suivi de piétons

Les algorithmes de suivi multi-objets (MOT), permettent de trouver les objets en mouvement dans une séquence d'images, et de les identifier en utilisant un identifiant unique (id) pour chaque objet détecté et suivi. Ces objets peuvent être des piétons, véhicules, etc. Pour cette thèse nous nous intéressons au suivi des piétons.

Parmi les algorithmes les plus connus du suivi de mouvement on trouve l'algorithme SORT [19].

### 4.2.1 Suivi d'un seul objet

L'objectif est de se focaliser simplement sur un seul objet sur une image et de le suivre. Ce type de suivi est le plus simple vu que le plus grand défi des algorithmes de suivi est la distinction des objets.

Le suivi d'un seul objet ne peut être appliqué pour les systèmes destinés aux voitures intelligentes. De ce fait, nous allons nous concentrer dans ce chapitre sur le suivi d'objets multiples.

### 4.2.2 Suivi Multi-objets

L'objectif principal des algorithmes de suivi multi-objets, est de localiser les objets en mouvement dans une séquence d'images, et d'identifier ces objets dans différentes images.

Un algorithme de suivi doit être combiné avec un détecteur pour extraire les informations concernant les bounding box des piétons pour chaque frame. Dans cette section, nous allons faire le point sur les différents algorithmes de suivi et leurs étapes. Une fois le piéton détecté, un identifiant unique (ID) lui est attribué jusqu'à ce qu'il quitte la séquence. Pour ceci deux techniques sont utilisées :

nous concentrer sur les étapes de suivi et nous supposons que les piétons sont détectés par un algorithme de détection et que nous avons les bounding box. Compte tenu des informations du bounding box pour un ID dans la première image de la



séquence, l’ID est attribué aux images subséquentes en utilisant une des deux techniques :

- Attribution d’un ID basé sur le centroïde : Le centroïde de chaque bounding box est calculé sur la première image. Par la suite, les centroïdes de la deuxième image de la séquence sont calculés et comparés avec ceux de la première image. Les IDs sont ensuite attribués basés sur la distance relative entre les deux centroïdes. L’hypothèse de base est que les centroïdes se déplacent légèrement entre deux images consécutives, la distance de déplacement des centroïdes entre deux images consécutives ne doit pas être assez large. Selon cette hypothèse, la distance entre les centroïdes ne doit pas être très importante. L’inconvénient de cette approche, est la possibilité de commutation entre les IDs dans le cas d’une occultation.
- Filtre de Kalman (KF) : Le Filtre de Kalman permet de modéliser le suivi en se basant sur la position, la vitesse et la prédiction de la position de l’objet. Il modélise la position future ainsi que la vitesse en utilisant une distribution gaussienne. Les nouvelles mesures sont ajoutées à l’ancienne prédiction calculée pour prédire la nouvelle position. Le Filtre de Kalman a de meilleurs résultats par rapport à la technique du centroïde dû à l’utilisation de la position et la vitesse de mouvement.

### 4.2.3 Méthodes traditionnelles

#### Méthode Mean-Shift

Mean-Shift [27] est une méthode populaire utilisée prioritairement pour les problèmes de classifications non supervisés et de regroupement. Cette méthode se caractérise par le calcul du centre des groupes par l’utilisation de la moyenne pondérée qui donne de l’importance aux points les plus proches de la moyenne. Le but de cet algorithme est de trouver tous les noeuds dans une distribution de données.

L’application de l’algorithme Mean-Shift pour le suivi d’objets permet de fournir des informations sur le changement de la distribution des caractéristiques de l’objet

(couleur, texture, histogramme, etc.) sur une séquence d'images, et par conséquent sur le changement du mouvement de l'objet et donc sur son suivi.

#### 4.2.4 Méthode du flux optique

Cette méthode diffère de la précédente, vu qu'elle ne nécessite pas l'extraction des caractéristiques des objets détectés. Comme alternative, l'objet est suivi en utilisant les variations de lumières spatio-temporel dans une séquence d'images au niveau des pixels. Le suivi des objets en utilisant le flux optique dépend sur trois hypothèses importantes :

- Cohérence de la luminosité : La luminosité autour d'une petite région doit rester presque constante même si l'emplacement de la région change.
- Cohérence spatiale : Les points appartenant à la même surface possèdent typiquement des mouvements similaires.
- Mouvement limité : Le mouvement des points ne doit pas être si loin où d'une façon aléatoire.

#### 4.2.5 Approches basées sur l'apprentissage profond

##### Deep Regression Network

Le réseau de régression profonde [59] est une des premières méthodes utilisant l'apprentissage profond, pour le suivi d'un seul objet. Le modèle est entraîné sur un ensemble de données constitué de vidéos avec des images cibles étiquetées. L'objectif du modèle est de suivre simplement un objet donné à partir d'une coupe d'image donnée.

Pour y parvenir, le réseau se constitue d'une architecture CNN à deux trames utilisant à la fois la trame actuelle et la trame précédente pour régresser avec précision sur l'objet.

Comme le montre la figure 4-1 , nous prenons le recadrage de l'image précédente basée sur les prédictions et définissons une "région de recherche" dans l'image actuelle basée sur ce recadrage. Le réseau est ensuite entraîné à régresser pour trouver l'objet

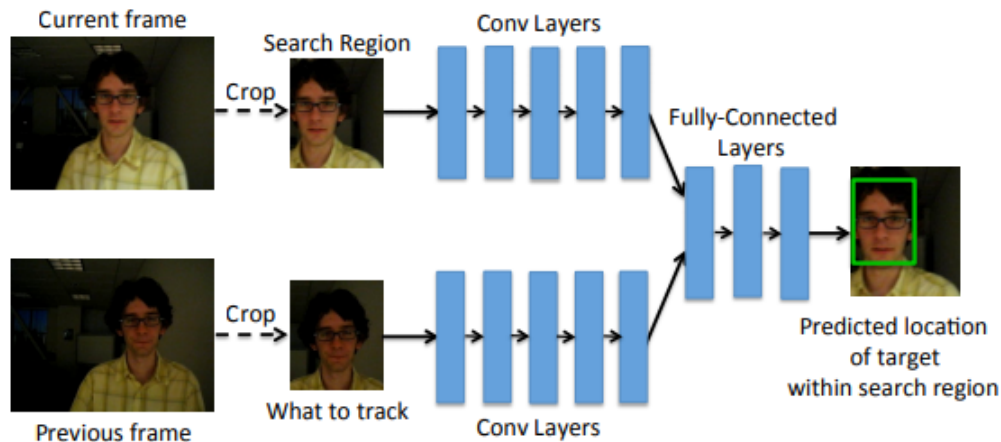


FIGURE 4-1 – Schéma explicatif du réseau de régression profonde [59]

dans cette région de recherche. L'architecture du réseau est simple avec des CNN suivis de couches entièrement connectées, qui nous fournissent directement les coordonnées du bounding box.

### Recurrent You Only Look Once (ROLO)

ROLO [95] est une méthode pour suivre des objets en utilisant l'apprentissage profond. Cette méthode contient de légères modifications par rapport au détecteur YOLO. Elles sont concentrées sur l'ajout d'une unité LSTM récurrente à la fin du réseau, aidant à suivre les objets en capturant les caractéristiques spatio-temporelles. Comme illustré par la figure 4-2, YOLO est utilisé pour extraire les caractéristiques des objets et d'effectuer la détection, ce qui constitue les informations spatiales de l'objet. Ces informations sont ensuite utilisées comme entrée pour la couche LSTM qui permet de calculer les informations temporelles. La concaténation de ces deux informations spatiales et temporelles permet de prédire la position du bounding box de l'objet en question et de faire ainsi son suivi.

### Algorithme SORT

Simple Online and Realtime Tracking (SORT), est un algorithme de suivi en ligne et en temps réel proposé par Bewley et al. [19]. Cet algorithme permet de résoudre le problème de suivi d'objets multiples. Cette méthode de suivi d'objets contient

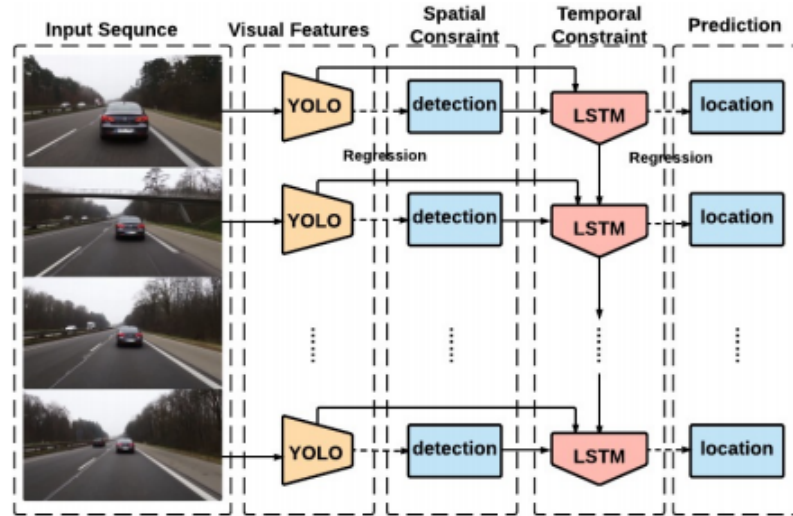


FIGURE 4-2 – schéma explicatif de l’algorithme de suivi ROLO [95]

4 étapes clés qui sont la détection, estimation de l’état des objets dans les images futures, association des détections actuelles avec les objets existants, et finalement création et suppression de l’identité (id) des objets suivis. L’objet dans notre cas est un piéton.

- Détection : La qualité de détection possède un impact significatif sur la performance du suivi, une meilleure détection permet d’avoir un meilleur suivi. Bewley et al. ont choisi les réseaux de convolutions pour la détection des objets en utilisant l’algorithme Faster Region CNN (FrRCNN) [152]. FrRCNN est un algorithme de détection d’objets basé sur deux étapes, la première consiste à extraire les caractéristiques et à proposer les zones d’intérêts pour la deuxième étape, qui classe les objets appartenant aux zones proposées. Pour ce travail, nous sommes concernés uniquement les piétons, et nous avons choisi d’utiliser YOLOv3 [42] comme détecteur pour des raisons décrites sur le chapitre 1.
- Estimation : Une fois le piéton détecté, l’algorithme propage l’identité (Id) des piétons détectés à l’image suivante. Le modèle utilisé estime le déplacement approximatif entre deux images consécutives pour chaque piéton, en utilisant un modèle linéaire de vitesse constante. Quand une détection est associée à une cible, le bounding box détecté est utilisé pour actualiser l’état de la cible où

les composantes de vitesse sont résolues en utilisant le Filtre de Kalman [139]. Si aucune détection n'est associée à la cible, l'état est prédit sans correction en utilisant le modèle linéaire de vitesse.

- Association des données : Lors de l'attribution des détections à des cibles existantes, chaque bounding box d'une cible est estimée en prédisant son nouvel emplacement dans l'image courante. La matrice d'attribution est calculée en utilisant la distance de Jaccard connu par Intersection over Union (IoU), entre chaque détection et toutes les bounding box prédits de la cible existante. L'attribution est résolue de manière optimale en utilisant l'algorithme Hungarian [74]. Une distance IoU minimale (IoUmin) est également imposée pour rejeter les attributions où le chevauchement entre la détection et la cible est inférieur à un certain IoUmin.

La distance IoU des bounding box gère les occlusions de court terme causées par les cibles passées. Puisque la distance IoU favorise les détections avec une échelle similaire, lorsqu'un piéton cible est couvert par un autre piéton, le piéton obturateur est alors détecté. Ceci permet de corriger la cible occlus avec la détection, alors que la cible occlus n'est pas affectée puisque aucune attribution n'est faite.

- Création et suppression des identifiants de suivis : Lorsque des piétons entrent et sortent de l'image, des identifiants uniques doivent être créés ou détruits en conséquence. Pour la création des traqueurs, l'algorithme considère que toute détection avec une distance de chevauchement moins de IoUmin signifie l'existence d'un piéton non suivi et donc de la création d'un nouvel identifiant. Le traqueur est initialisé en utilisant la géométrie des bounding box avec une vitesse mise à zéro. Par la suite, la cible suivie a besoin d'être associée à une détection pour la continuité de l'identifiant. La cible n'est plus suivie si elle n'est associée à aucune détection pour un certain nombre d'images (TLost). L'identifiant de cette cible est détruit en conséquence. De ce fait, si un piéton disparaît dans un nombre d'images TLost et réapparaît par la suite, le suivi de ce piéton reprend sous un nouvel identifiant.

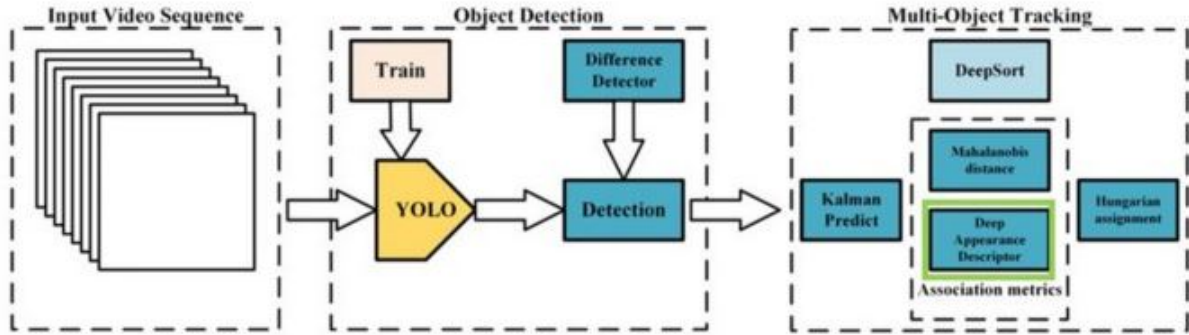


FIGURE 4-3 – Architecture du DeepSort [142]

Malgré sa performance en terme de précision de suivi et son efficacité, l'algorithme SORT retourne un grand nombre d'identifications commutées lors des occlusions. Afin d'améliorer ce point, l'algorithme Deep SORT introduit une nouvelle métrique de distance basée sur l'apparence de l'objet suivi, appelée « métrique d'association profonde ».

## Deep SORT

L'algorithme Deep SORT proposé par Wojke et al. [142], est un des algorithmes les plus utilisés pour le suivi d'objets multiples. Il améliore la performance de SORT en permettant le suivi des objets pour une plus longue période en proposant une nouvelle métrique de distance basée sur l'apparence de l'objet suivi (Fig. 4-3) permettant de réduire le nombre d'identifiants commutés durant les occlusions.

Pour chaque objet suivi, l'algorithme calcule le nombre d'images depuis la dernière association de métrique. Les objets suivis qui ont dépassé un certain nombre d'images appelé « âge maximal » sont considérés avoir quitté la scène et sont supprimés. Les objets apparaissant pour la première fois sont considérés comme des tentatives pendant les trois premières images. Tout objet non associé à une mesure pendant ses trois premières images sera alors supprimé. Toute nouvelle détection doit être associée à un identifiant, pour cela l'algorithme propose une métrique de distance ainsi que la méthode Hongroise pour associer les données.

— Métrique de distance : Les auteurs ont décidés d'utiliser la distance d'inter-

point carrée généralisée connue par la distance de Mahalanobis. Le seuillage de cette distance peut nous donner une très bonne idée des associations réelles. Cette métrique est plus précise que la distance euclidienne, car nous mesurons effectivement la distance entre deux distributions.

- Algorithme Hongrois : Est un algorithme d'optimisation combinatoire permettant de résoudre les problèmes d'affectation en temps polynomial. Il est utilisé par l'algorithme DeepSort afin d'associer les identifiants de chaque piéton en définissant si le piéton détecté dans l'image actuelle est le même que celui de l'image précédente.

## 4.3 La base JAAD

Joint Attention in Autonomous Driving (JAAD) [73], est une base de données pour étudier l'attention des utilisateurs de la route dans le cadre d'une conduite autonome. La base se concentre sur le comportement des piétons, des conducteurs et sur les facteurs qui influencent leurs manières de traverser les croisements. JAAD contient 346 vidéos d'une longueur entre 5 et 10 secondes pour chacune. Ces vidéos sont extraites pendant plus de 240 heures de conduite en Amérique du Nord et l'Europe de l'Est dans diverses conditions météorologiques.

La base contient plusieurs types d'annotations dans chaque vidéos. Parmi les annotations fournies nous trouvons des étiquettes sur les conditions météorologiques, l'emplacement de la vidéo, d'autres représentant le comportement horodatées des piétons nécessitant l'attention des conducteurs (ex : arrêté, marchant, regardant, etc.). La base contient également des attributs fournis pour chaque piéton (ex : genre, direction de marche, etc.), ainsi qu'une liste des signes de trafic visible sur la route.

Pour ce chapitre, nous nous intéressons à l'orientation du piéton représenté par le label "direction de marche"

## 4.4 Expérimentations et Résultats

L’objectif de cette partie est d’appliquer le réseau de Capsules proposé dans le chapitre précédent sur des séquences vidéos.

Cette section expose le résultat des expérimentations faites en utilisant le réseau des Capsules entraîné sur les images de la base SafeRoad et testé sur les vidéos de la base JAAD.

Pour chaque image de la séquence vidéo, les piétons sont tout d’abord détectés par le détecteur YOLO. Ensuite, l’algorithme Deep Sort est utilisé pour faire le suivi des piétons sur la vidéo, en identifiant chaque piéton détecté avec un identifiant unique (ID). Le cadre du piéton détecté est par la suite transformé en niveau de gris et redimensionné en 48x48 pixels (Taille d’entrée des images au réseau). Dans le but de réduire les variations d’intensité, ces images sont ensuite normalisées en soustrayant la moyenne et en divisant par l’écart type. L’image normalisée est utilisée comme entrée au réseau de Capsules entraîné auparavant sur la base SafeRoad, pour avoir en sortie l’orientation de marche pour chaque piéton identifié.

L’avantage de ce système réside dans la possibilité de son utilisation pour plusieurs piétons ayant différentes directions de marche sur la même image, comme illustré sur les figures 4-4.

Plusieurs méthodes ont été testées pour la reconnaissance de l’orientation du piéton au niveau d’une séquence d’images :

Dans le chapitre précédent, nous avons estimé l’orientation du piéton à partir d’une seule image. De ce fait, nous proposons dans ce chapitre l’estimation de la direction de marche du piéton en utilisant le résultat de l’algorithme présenté dans le chapitre 3.

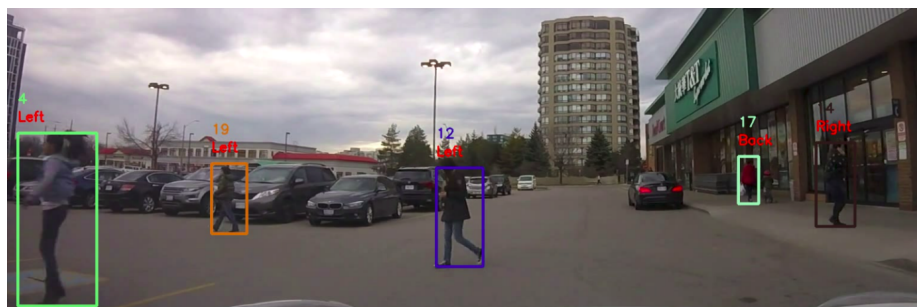
La méthode proposée pour estimer la direction de marche du piéton consiste à utiliser l’orientation du piéton la plus fréquente pour une durée T comme direction de marche, représenté par l’eq. 4.1.

$$Direction = argmax_i(f_{o1}, f_{o2}, f_{o3}, f_{o4}) \quad (4.1)$$





(a)



(b)

FIGURE 4-4 – Reconnaissance des orientations différentes de plusieurs piétons sur la même image

Où  $f_{oi}$  représente l'occurrence de la fréquence de l'orientation  $i$ . Cette méthode peut mener à des erreurs si l'orientation la plus fréquente n'est pas la même que la vérité terrain. Le résultat de l'estimation de l'orientation la plus fréquente dépend de la durée utilisée (nombre d'images consécutives prises pour calculer la fréquence).

Pour ceci, nous avons choisi de tester 3 durées :

- Prendre en considération le résultat de l'orientation de chaque image de la séquence vidéo.
- Calculer la fréquence de l'orientation pour chaque 5 images.
- Calculer la fréquence de l'orientation pour chaque 10 images.

Pour évaluer les modèles l'exactitude (eq.3.3) est calculée en divisant les vrais positifs et les vrais négatifs sur l'ensemble des prédictions. D'après les expérimentations, l'exactitude de la direction de marche augmente avec l'augmentation des nombres d'images utilisées pour calculer la fréquence avec une valeur de 63%, 74.53% et 80.5% pour respectivement 1, 5 et 10 images consécutives. Nous nous sommes arrêtés sur 10 images vu que la fréquence des vidéos est de 30 images par seconde, afin d'avoir au minimum dans une seconde 3 estimations de direction de marche.

Une des caractéristiques du piéton est sa capacité à changer rapidement sa direction de marche d'où la nécessité d'un système de détection rapide et efficace. Durant nos expérimentations, nous avons rencontré des piétons changeant leur direction dans un laps de temps très court, comme le montre les figures 4-5 et 4-6 prises de la base JAAD où le système proposé a pu détecter le changement de direction du piéton de la gauche vers l'avant.



FIGURE 4-5 – Détection de la direction avant le changement



FIGURE 4-6 – Détection de la direction après son changement rapide

Le système proposé permet également la reconnaissance de l'orientation du piéton même en cas de mouvement rapide pour un piéton qui court, comme le montre les Figures 4-7 et 4-8.



FIGURE 4-7 – Détection de l'orientation des piétons occultés qui courent

Durant les expérimentations, nous avons rencontrés des cas où le piéton possède une orientation diagonale oblique, ce qui rend la détection de sa direction compliquée comme représenté sur la Figure 4-9a où notre algorithme estime que l'orientation de la dame est tantôt devant tantôt à gauche, alors que la vérité terrain la considère comme étant gauche. Un autre cas représenté sur la Figure 4-9b où les piétons ayant l'indice 1 et 2 possèdent une orientation oblique entre les directions devant et gauche. Le système proposé dans ce cas détecte les deux directions possibles alternativement comme nous pouvons le remarquer sur les figures 4-8 et 4-9b où sur la première le piéton avec l'indice 1 avait comme direction "gauche" alors que sur la deuxième sa

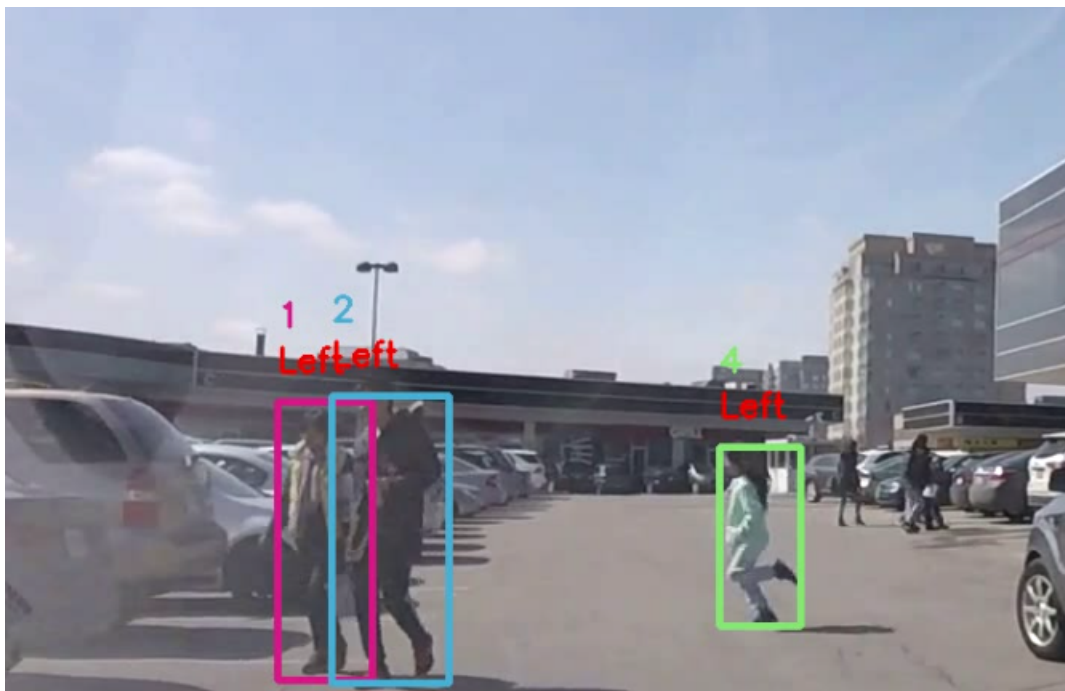


FIGURE 4-8 – Détection de l'orientation d'un enfant courant

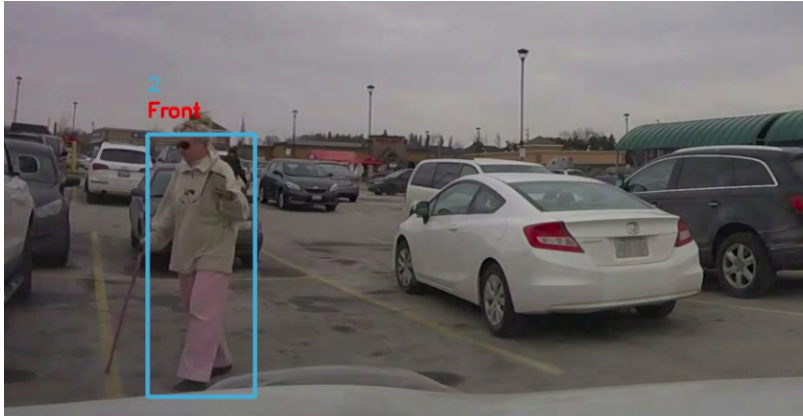
direction a changé vers "avant". Dans ce cas, l'estimation de notre algorithme et celle de la vérité terrain sont toutes deux correctes. Alors qu'en terme de précision (eq.3.11), la direction prédite est considérée comme incorrecte vu qu'elle ne convient pas avec celle de la vérité terrain.

Un autre cas de fausses estimations est lorsque le piéton détecté est assez loin du véhicule au point que l'algorithme n'estime pas correctement son orientation, comme représenté sur la figure 4-10, où le système a classé l'orientation du piéton allant vers la droite comme étant gauche.

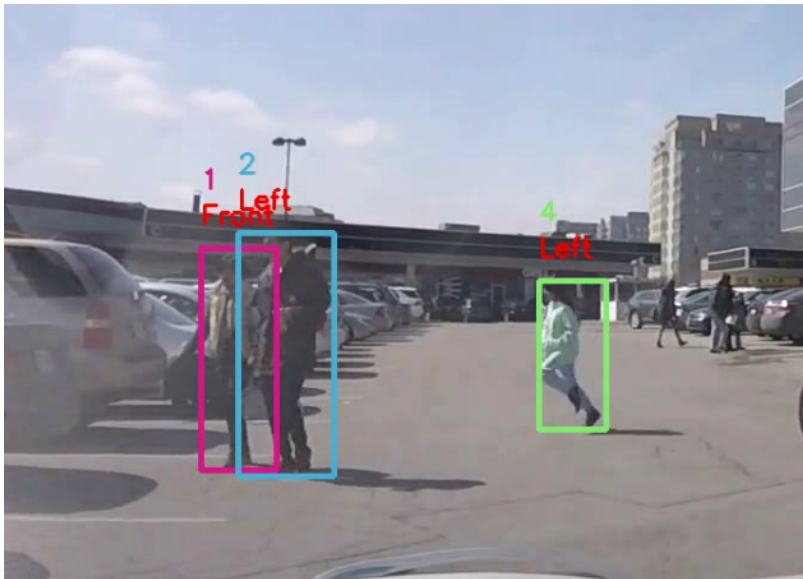
L'estimation de l'orientation est incorrecte également lorsque les conditions d'éclairages ne permettent pas la reconnaissance de l'orientation du piéton ni sa détection pour certains cas. Les figures 4-11a et 4-11b contiennent des piétons non détectés par le système, prises dans des conditions pluvieuses et d'obscurités.

La figure 4-12 représente un des cas où l'éclairage de l'environnement est trop faible entraînant une mauvaise classification de l'orientation des trois piétons, qui sont à peine détectables.

Dans la suite, nous allons présenter le résultat des expérimentations en terme de



(a)



(b)

FIGURE 4-9 – Exemple de piétons ayant une orientation oblique





FIGURE 4-10 – Exemple d’une mauvaise classification de l’orientation d’un piéton dû à la distance



(a)



(b)

FIGURE 4-11 – Exemple de piétons non détectés dans l’obscurité



FIGURE 4-12 – Exemple d’une mauvaise classification de l’orientation des piétons dans l’obscurité

précision. Ces résultats sont issus des tests réalisés sur 10 séquences vidéos prises de la base JAAD.

La figure 4-13 représente la matrice de confusion de la première expérimentation prenant en considération le résultat de l’orientation de chaque image de la séquence vidéo ( $N=1$  où  $N$  représente le nombre d’images consécutives dans une séquence). Selon la matrice, nous pouvons remarquer que les orientations "Gauche" et "Droite" possèdent le plus grand pourcentage de vrais positifs avec respectivement 72% et 71%. Alors que le plus grand taux d’erreurs appartient à l’orientation "Avant" avec 65% de fausses détections où 30% des orientations "Avant" ont été classés comme "Droite" et 31% comme "Gauche". Les vrais positives de la classe "Arrière" peuvent aussi être considérées faible avec une précision de 52% seulement.

L’utilisation de l’orientation la plus fréquente pour  $N=5$  a augmenté le taux des vrais positives pour chaque classe, comme le montre la matrice sur la figure 4-14. La direction "Avant" reste toujours celle ayant le taux le plus élevés des mauvaises classifications avec seulement 49% des vrais positives.

L’utilisation de 10 images augmente la précision pour toutes les classes arrivant à 83% et 82% pour respectivement la classe "Droite" et "Gauche" et une précision de

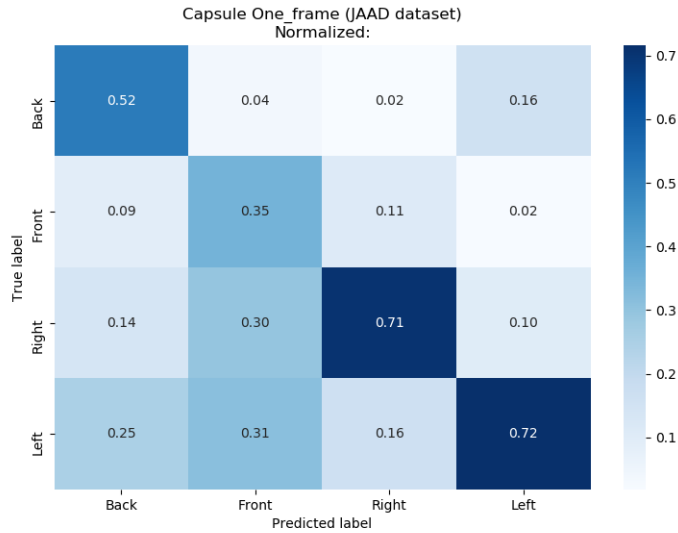


FIGURE 4-13 – Matrice de confusion pour N=1

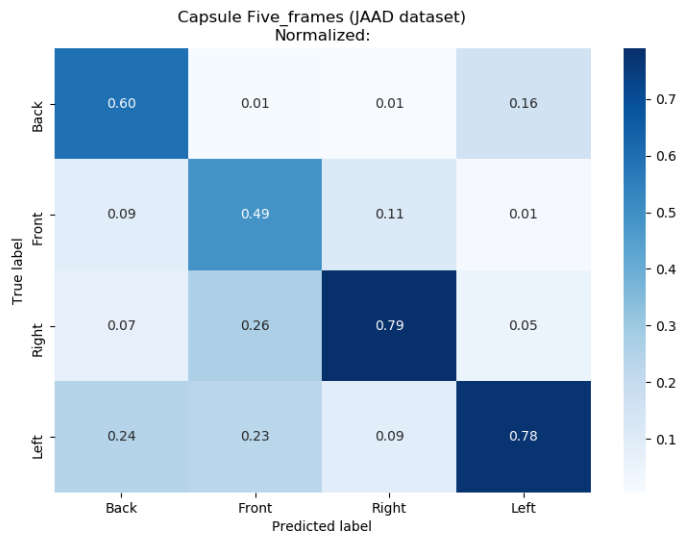


FIGURE 4-14 – Matrice de confusion pour N=5

68% et 60% pour respectivement les classes "Arrière" et "Avant". Comme représenté sur la Figure 4-15 les mauvaises classifications sur la classe "Gauche" ont été surtout marquées pour la classe "Arrière" avec un taux arrivant à 14%. De même, la classe "Gauche" représente 17% des mauvaises classifications de la classe "Arrière", alors que la classe "Avant" représente 13% des mauvaises classifications.

Selon les trois matrices de confusion nous pouvons remarquer que les classes "Ar-



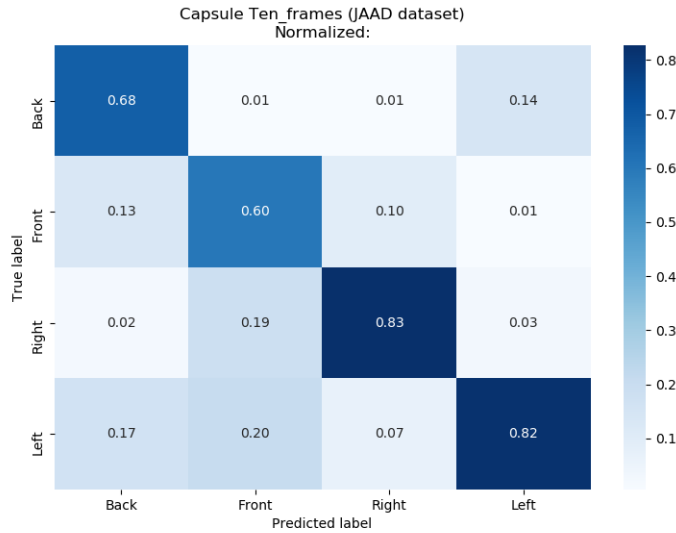


FIGURE 4-15 – Matrice de confusion pour N=10

rière" et "Avant" ont le plus grand taux d'erreurs. Les mauvaises classifications pour l'orientation "Avant" sont souvent classées comme étant "Gauche" ou "Droite", alors que ceux de l'orientation "Arrière" sont mal classés en tant que "Gauche" ou "Avant".

Ceci peut être expliqué par le problème de l'orientation oblique que nous avons cité plus haut. La spécification de l'orientation oblique réside dans la difficulté de correctement préciser la direction du piéton qui se situe entre deux orientations (Avant/droite - Avant/gauche - Arrière/droite - Arrière/gauche). Le système proposé pour ce cas détecte les deux orientations en alternance, ce qui ne convient pas avec la vérité terrain, qui considère ces orientations comme étant "Arrière" ou "Avant". D'où le taux d'erreurs augmentés pour ces deux classes.

D'après les résultats, l'utilisation de l'orientation la plus fréquente pour chaque 10 images consécutives, permet d'avoir un résultat plus exact sur la direction de marche du piéton. Ceci peut être expliqué par la non intégration des mauvaises classification faites par l'algorithme qui sont moins fréquentes par rapport aux bonnes classifications. Cette méthode permet également de respecter la contrainte temporelle pour avoir une détection rapide et en temps réel.

## 4.5 Conclusion

Dans ce chapitre, nous avons évalué la reconnaissance de l'orientation du piéton dans des séquences vidéos en utilisant la base JAAD.

Notre approche est constitué dans un premier temps, par la détection du piéton et sa localisation dans l'image en utilisant YOLO, ensuite le suivi de tous les piétons détectés en utilisant l'algorithme DeepSort, et finalement la reconnaissance de l'orientation de chaque piéton suivi en utilisant le réseau de Capsules entraîné sur les images de la base SafeRoad.

Nous avons évalué la reconnaissance de la direction de marche des piétons, en utilisant la méthode de l'orientation du piéton la plus fréquente. Cette technique vise à considérer l'orientation la plus répandue pour un certain nombre d'images consécutives ( $N$ ) comme étant la direction du piéton. Nous avons testé tout d'abord en prenant en considération toutes les images d'une séquence vidéo et en utilisant  $N=1$ ,  $N=5$  et  $N=10$ .

D'après les expérimentations, l'exactitude de la direction de marche augmente avec l'augmentation du nombre d'images utilisées pour calculer la fréquence de l'orientation allant de 62.98% pour  $N=1$  jusqu'à une précision de 73.53% et 80.49% pour respectivement 5 et 10 images. Les expérimentations ont été réalisées sur des vidéos montrant différentes conditions météorologiques et pendant différents moments de la journée représentant des piétons traversant de différentes manières (horizontale, diagonale).

En analysant les résultats, nous pouvons constater l'importance de la méthode de la fréquence à éliminer les fausses classifications pour ne garder que les classifications dominantes qui sont les plus susceptibles d'être correctes. D'après les matrices de confusion présentées dans ce chapitre, nous pouvons constater aussi la difficulté de reconnaître l'orientation d'un piéton traversant de manière diagonale. Pour résoudre ce problème, nous envisageons dans le futur d'évaluer l'approche proposée en utilisant huit orientations au lieu de quatre, comme nous considérons la création d'un système de bout en bout et son intégration au sein d'un système ADAS.



# Conclusion Générale

Au cours de la dernière décennie, les systèmes de transport intelligents ont connu un énorme progrès, principalement au niveau des systèmes d'aide à la conduite avancés (ADAS). Cependant, les accidents mortels sont toujours en hausse, conservant ainsi la sécurité routière comme un enjeu important pour la recherche scientifique.

Dans le cadre de cette étude, nous avons fait le point sur le conflit piéton-véhicule présent d'une façon exorbitante dans les routes les moins structurées, souffrantes d'un manque de mesures pour la sécurité piétonnes.

Nous avons proposé un nouveau système de prévention des accidents piétons-véhicules bien conçu pour les zones avec manque d'installations piétonnes. Ce système pourrait également être efficace pour les conducteurs imprudents et les piétons ayant un comportement à risque. Il alerte le conducteur en fonction de la gravité du risque d'accident. Cette dernière dépend de la position du piéton par rapport au véhicule, et du temps nécessaire au véhicule pour s'arrêter complètement afin d'éviter l'accident.

Dans un premier temps, nous avons effectué un état de l'art sur les principaux concepts des ADAS et leurs différentes fonctions, ainsi que leur évolution technologique au cours du temps. Cette dernière a connu l'apparition des capteurs extéroceptifs (caméras, lidars, etc.) nécessitant l'application des techniques de vision et du traitement d'images. Dans ce contexte, nous avons apporté une vue d'ensemble des algorithmes de vision existants pour la détection des piétons, allant des descripteurs classiques aux méthodes d'apprentissage automatique.

Notre recherche a abouti à la définition d'un système d'alerte pour la prévention des accidents piétons-véhicules. L'originalité de ce travail, consiste dans la proposition d'un nouveau prototype tenant en compte de l'orientation des piétons. Cette méthode

calcule le risque d'un éventuel accident selon trois cas de figure possible lors d'un conflit piéton-véhicule :

- Le véhicule traverse le point de collision avant le piéton.
- Le piéton traverse le point de collision avant le véhicule.
- Le piéton et le véhicule traversent le point de collision simultanément.

Le risque de collision dépend de la distance entre le véhicule et le piéton, du temps d'arrêt du véhicule ainsi que la direction du piéton. Cette dernière peut intensifier le risque d'une collision si le piéton se dirige vers le sens d'arrivée du véhicule, alors que le risque diminue pour le cas inverse.

Le système de vision utilisé pour l'orientation du piéton représente un modèle basé sur les réseaux de Capsules. Ce dernier utilise respectivement deux architectures pour classer l'orientation des piétons en quatre directions (Avant, Droite, Gauche, Arrière) et 8 directions (devant, arrière, gauche, droit, devant-droit, devant-gauche, arrière-droit, arrière-gauche).

Afin d'entraîner le réseau susmentionné, nous avons créé une nouvelle base de données des directions de piétons prise à partir de scènes réelles, que nous avons nommée SafeRoad. Cette base a été collectée depuis plusieurs villes Marocaines en utilisant une caméra intégrée dans un véhicule en mouvement. Elle contient 4160 images de piétons manuellement annotées en 4 directions et 8894 images annotées en 8 directions.

Le modèle de classification de l'orientation des piétons est entraîné sur notre nouvelle base SafeRoad et sur les bases publiques Daimler et TUD Multiview Pedestrians. Les réseaux de Capsules ont obtenus les meilleurs résultats en terme de précision avec 97.60% sur la base Daimler et 78.95% sur notre base SafeRoad pour les 4 directions, par rapport à 95.52% et 96.45% respectivement pour les réseaux AlexNet et ResNet sur la base Daimler et 77.60% et 78.54% sur la base SafeRoad. Les réseaux de Capsules ont obtenus également les meilleurs résultats pour la classification à 8 directions avec une précision de 70.22% et 93.48% respectivement pour les bases SafeRoad et TUD, comparé à 65.34% et 54.43% pour les réseaux AlexNet et ResNet sur la base SafeRoad et 91.79% et 92.70% pour la base TUD.

Nous pouvons constater d’après les résultats que la base SafeRoad a une précision plus faible relativement à celle de Daimler et TUD. Cette différence des résultats est probablement due à la différence dans l’environnement de collection des bases, où Daimler et TUD ont été prises dans des villes européennes respectant différentes mesures de sécurité pour les piétons. En revanche, la base SafeRoad contient davantage d’orientations adjacentes qui sont compliquées à classer, ce qui semble être plus réaliste.

Finalement, nous avons proposé dans le dernier chapitre la reconnaissance de l’orientation du piéton sur des séquences vidéos prises de la base JAAD, en utilisant les réseaux de Capsules entraîné sur la base SafeRoad. La reconnaissance sur des séquences vidéos nécessite plusieurs étapes énumérées comme suit :

- La détection : Effectuer par l’algorithme de détection en temps réel YOLO.
- Le suivi de chaque piéton détecté en lui affectant un identifiant unique jusqu’à sa disparition. L’algorithme de suivi utilisé est DeepSort.
- La reconnaissance de chaque piéton suivi en utilisant les réseaux de Capsules.

Le travail présenté dans ce rapport propose un prototype de système d’alerte pour éviter les collisions piétons-véhicules. Cette étude peut-être approfondie en la développant sous un environnement RTMaps. La mise en place du système sous un environnement de développement permet de tester son fonctionnement réel. Ceci nous permettra d’estimer le niveau d’efficacité du système et de connaître les différents scénarios possibles en cas de simulation.

Cette étape, nous permettra ensuite de franchir le pas vers une implémentation réelle sur des véhicules de test, ce qui nous donnera la possibilité de déterminer les limites du système proposé dans un environnement réel, et de le valider par la suite.

Afin d’améliorer la précision des bonnes classifications et de diminuer les mauvaises, nous tenons comme perspective d’augmenter la taille de notre base et d’améliorer la qualité de ses images.



# Productions scientifiques en lien avec la thèse

## Conférences internationales avec comité de lecture

- **Safaâ Dafrallah**, Aouatif Amine, Stéphane Mousset, and Abdelaziz Bensrhair. Will capsule networks overcome convolutional neural networks on pedestrian walking direction ? In 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pages 3702–3707. IEEE, 2019
- **Safaâ Dafrallah**, Aouatif Amine, Stéphane Mousset, and Abdelaziz Bensrhair. Driver behavior assistance in road intersections. In Embedded Systems and Artificial Intelligence, Proceedings of ESAI 2019, Fez, Morocco 1076 (2020).
- Zakaria Sabir, **Safaâ Dafrallah**, and Aouatif Amine. A novel solution to prevent accidents using v2i in moroccan smart cities. In 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), pages 621–625. IEEE, 2019.
- **Safaâ Dafrallah**, Zakaria SABIR, Aouatif AMINE, Stéphane MOUSSET, and Abdelaziz BENSRRHAIR. Pedestrian walking direction classification formoroccan road safety. In the proceedings of the International Conference on Industrial Engineering and Operations Management (IEOM), August, 2020.

## Chapitres dans des ouvrages collectifs

- **Safaâ Dafrallah**, Aouatif Amine, Stéphane Mousset, and Abdelaziz Bensrhair. Driver behavior assistance in road intersections. In Embedded Systems



- and Artificial Intelligence, pages 53–60. Springer, Singapore 2020.
- **Safaâ Dafrallah**, Stéphane Mousset, Aouatif Amine, Abdelaziz Bensrhair. The Integration of Pedestrian Orientation for the Benefit of ADAS, based on the Moroccan Case. From AI to Autonomous and Connected Vehicles. ISTE Ltd., London, and John Wiley and Sons, New York, August 2021.

## Revue avec comité de lecture

- **Safaâ Dafrallah**, Aouatif Amine, Stéphane Mousset, and Abdelaziz Bensrhair. Monocular pedestrian orientation detection based on Capsule Network for a Novel Collision Warning System. IEEE access, 2021.

# Bibliographie

- [1] A Aashto. Policy on geometric design of highways and streets. *American Association of State Highway and Transportation Officials, Washington, DC*, 1 :1048, 2018.
- [2] Masami Aga and Akio Okada. Analysis of vehicle stability control (vsc)’s effectiveness from accident data. In *Proceedings : International Technical Conference on the Enhanced Safety of Vehicles*, volume 2003, pages 7–p. National Highway Traffic Safety Administration, 2003.
- [3] Samiksha Agrawal and Pallavi Khatri. Facial expression detection techniques : based on viola and jones algorithm and principal component analysis. In *2015 Fifth International Conference on Advanced Computing & Communication Technologies*, pages 108–112. IEEE, 2015.
- [4] Faizan Ahmad, Li Ning, and Mustafa Tahir. An improved d-cnn based on yolov3 for pedestrian detection. In *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*, pages 405–409. IEEE, 2019.
- [5] Ignacio Parra Alonso, David Fernández Llorca, Miguel Ángel Sotelo, Luis M Bergasa, Pedro Revenga de Toro, Jess Nuevo, Manuel Ocaña, and Miguel Ángel García Garrido. Combination of feature extraction methods for svm pedestrian detection. *IEEE Transactions on Intelligent Transportation Systems*, 8(2) :292–307, 2007.
- [6] Mykhaylo Andriluka, Stefan Roth, and Bernt Schiele. Monocular 3d pose estimation and tracking by detection. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 623–630. IEEE, 2010.
- [7] J Anitha, G Mani, and K Venkata Rao. Driver drowsiness detection using viola jones algorithm. In *Smart Intelligent Computing and Applications*, pages 583–592. Springer, 2020.
- [8] Tudor Barbu. Pedestrian detection and tracking using temporal differencing and hog features. *Computers & Electrical Engineering*, 40(4) :1072–1079, 2014.
- [9] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf : Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer, 2006.

- [10] Rachid Belaroussi and Jean Philippe Tarel. Détection des panneaux de signalisation routière par accumulation bivariée. *Traitement du signal*, 27(3) :pp-265, 2010.
- [11] Francesco Bella, Chiara Ferrante, Manuel Silvestri, and Maria Rosaria De Blasiis. Effects on driver’s yielding behavior of a pedestrian collision warning system in different road environments. In *International Conference on Applied Human Factors and Ergonomics*, pages 307–314. Springer, 2020.
- [12] Francesco Bella, Valentina Natale, and Manuel Silvestri. Driver-pedestrian interaction under different road environments. *Transportation research procedia*, 27 :148–155, 2017.
- [13] Francesco Bella and Manuel Silvestri. Effects of safety measures on driver’s speed behavior at pedestrian crossings. *Accident Analysis & Prevention*, 83 :111–124, 2015.
- [14] Klaus Bengler, Klaus Dietmayer, Berthold Farber, Markus Maurer, Christoph Stiller, and Hermann Winner. Three decades of driver assistance systems : Review and future perspectives. *IEEE Intelligent transportation systems magazine*, 6(4) :6–22, 2014.
- [15] Brian Bennie, Eric L Reed, and Kristopher Hook. System and method for lane departure warning, November 25 2010. US Patent App. 12/468,491.
- [16] Inger Marie Bernhoft and Gitte Carstensen. Preferences and behaviour of pedestrians and cyclists by age and gender. *Transportation Research Part F : Traffic Psychology and Behaviour*, 11(2) :83–95, 2008.
- [17] Massimo Bertozzi, Alberto Broggi, Mike Del Rose, Mirko Felisa, Alain Rakotomamonjy, and Frédéric Suard. A pedestrian detector using histograms of oriented gradients and a support vector machine classifier. In *2007 IEEE Intelligent Transportation Systems Conference*, pages 143–148. IEEE, 2007.
- [18] B Besserer, S Estable, B Ulmer, and D Reichardt. Shape classification for traffic sign recognition. *IFAC Proceedings Volumes*, 26(1) :487–492, 1993.
- [19] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. Simple online and realtime tracking. In *2016 IEEE international conference on image processing (ICIP)*, pages 3464–3468. IEEE, 2016.
- [20] John Vincent Bond III, Gerald H Engelman, Jonas Ekmark, Jonas LZ Jansson, M Nabeel Tarabishy, and Levasseur Tellis. Autonomous emergency braking system, February 25 2003. US Patent 6,523,912.
- [21] Maria Eugenia Garcia Bordes. Combined lane change assist and rear, cross-traffic alert functionality, February 16 2012. US Patent App. 12/855,238.

- [22] Wen Boyuan and Wu Muqing. Study on pedestrian detection based on an improved yolov4 algorithm. In *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, pages 1198–1202. IEEE, 2020.
- [23] Lorenzo Caminiti, Jeffrey Clark Lovell, James Joseph Richardson, and Christopher Thomas Higgins. Communication based vehicle-pedestrian collision warning system, December 2 2014. US Patent 8,903,640.
- [24] Yunqiang Chen, Xiang Sean Zhou, and Thomas S Huang. One-class svm for learning in image retrieval. In *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, volume 1, pages 34–37. IEEE, 2001.
- [25] Jinyoung Choi, Beom-Jin Lee, and Byoung-Tak Zhang. Human body orientation estimation using convolutional neural network. *arXiv preprint arXiv :1609.01984*, 2016.
- [26] Pushpa Choudhary and Nagendra R Velaga. Modelling driver distraction effects due to mobile phone use on reaction time. *Transportation Research Part C : Emerging Technologies*, 77 :351–365, 2017.
- [27] Dorin Comaniciu and Peter Meer. Mean shift analysis and applications. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pages 1197–1203. IEEE, 1999.
- [28] Mehul K Dabhi and Bhavna K Pancholi. Face detection system based on viola-jones algorithm. *International Journal of Science and Research (IJSR)*, 5(4) :62–64, 2016.
- [29] Safaa Dafrallah, Aouatif Amine, Stéphane Mousset, and Abdelaziz Bensrhair. Will capsule networks overcome convolutional neural networks on pedestrian walking direction ? In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 3702–3707. IEEE, 2019.
- [30] Safaa Dafrallah, Aouatif Amine, Stéphane Mousset, and Abdelaziz Bensrhair. Driver behavior assistance in road intersections. In *Embedded Systems and Artificial Intelligence*, pages 53–60. Springer, 2020.
- [31] Safaa Dafrallah, Aouatif Amine, Stéphane Mousset, and Abdelaziz Bensrhair. Monocular pedestrian orientation recognition based on capsule network for a novel collision warning system. *IEEE Access*, 2021.
- [32] Safaa DAFRALLAH, Zakaria SABIR, Aouatif AMINE, Stéphane MOUSSET, and Abdelaziz BENSRAIR. Pedestrian walking direction classification for moroccan road safety. In *International Conference on Industrial Engineering and Operations Management*. IEOM, 2020.
- [33] Erez Dagan, Ofer Mano, Gideon P Stein, and Amnon Shashua. Forward collision warning with a single camera. In *IEEE Intelligent Vehicles Symposium, 2004*, pages 37–42. IEEE, 2004.

- [34] Xiaobiao Dai, Junping Hu, Hongmei Zhang, Abubakar Shitu, Chunlei Luo, Ahmad Osman, Stefano Sfarra, and Yuxia Duan. Multi-task faster r-cnn for night-time pedestrian detection and distance estimation. *Infrared Physics & Technology*, 115 :103694, 2021.
- [35] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. Ieee, 2005.
- [36] Piotr Dollár, Ron Appel, Serge Belongie, and Pietro Perona. Fast feature pyramids for object detection. *IEEE transactions on pattern analysis and machine intelligence*, 36(8) :1532–1545, 2014.
- [37] Piotr Dollár, Zhuowen Tu, Pietro Perona, and Serge Belongie. Integral channel features. 2009.
- [38] Alex Dominguez-Sanchez, Miguel Cazorla, and Sergio Orts-Escolano. Pedestrian movement direction recognition using convolutional neural networks. *IEEE transactions on intelligent transportation systems*, 18(12) :3540–3548, 2017.
- [39] M Enzweiler and M Gavrila. Daimler mono pedestrian detection bechmark dataset. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pages 2179–2195, 2009.
- [40] Markus Enzweiler and Darius M Gavrila. Integrated pedestrian classification and orientation estimation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 982–989. IEEE, 2010.
- [41] Sungmin Eum and Ho Gi Jung. Enhancing light blob detection for intelligent headlight control using lane detection. *IEEE transactions on intelligent transportation systems*, 14(2) :1003–1011, 2013.
- [42] Ali Farhadi and Joseph Redmon. Yolov3 : An incremental improvement. *Computer Vision and Pattern Recognition, cite as*, 2018.
- [43] Justin Johnson Fei-Fei Li and Serena Yeung. *R-CNN, Fast R-CNN, Faster R-CNN, YOLO - Object Detection Algorithms*. Stanford University, 2017.
- [44] Nicholas N Ferencsik. Pedestrian age and gender in relation to crossing behavior at midblock crossings in india. *Journal of Traffic and Transportation Engineering (English Edition)*, 3(4) :345–351, 2016.
- [45] Nicholas Frazier, Jeremy Lerner, Taylor HAWLEY, and Joseph Gifford. Blind spot detection and alert, April 1 2021. US Patent App. 16/588,376.
- [46] Yoav Freund, Robert E Schapire, et al. Experiments with a new boosting algorithm. In *icml*, volume 96, pages 148–156. Citeseer, 1996.

- [47] Ray Fuller. A conceptualization of driving behaviour as threat avoidance. *Ergonomics*, 27(11) :1139–1155, 1984.
- [48] Guolong Gan and Jian Cheng. Pedestrian detection based on hog-lbp feature. In *2011 Seventh International Conference on Computational Intelligence and Security*, pages 1184–1187. IEEE, 2011.
- [49] Tarak Gandhi and Mohan M Trivedi. Pedestrian collision avoidance systems : A survey of computer vision based recent studies. In *2006 IEEE Intelligent Transportation Systems Conference*, pages 976–981. IEEE, 2006.
- [50] Tarak Gandhi and Mohan Manubhai Trivedi. Pedestrian protection systems : Issues, survey, and challenges. *IEEE Transactions on intelligent Transportation systems*, 8(3) :413–430, 2007.
- [51] Dariu M Gavrila. Traffic sign recognition revisited. In *Mustererkennung 1999*, pages 86–93. Springer, 1999.
- [52] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.
- [53] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.
- [54] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 38(1) :142–158, 2015.
- [55] Charles A Green and Uzmaa H Balbale. Cross traffic alert system for a vehicle, and related alert display method, August 12 2010. US Patent App. 12/370,239.
- [56] Umar Zakir Abdul Hamid, Fakhurul Razi Ahmad Zakuan, Khairul Akmal Zulkepli, Muhammad Zulfaqar Azmi, Hairi Zamzuri, Mohd Azizi Abdul Rahman, and Muhammad Aizzat Zakaria. Autonomous emergency braking system with potential field risk assessment for frontal collision mitigation. In *2017 IEEE conference on systems, process and control (icspc)*, pages 71–76. IEEE, 2017.
- [57] Kota Hara, Raviteja Vemulapalli, and Rama Chellappa. Designing deep convolutional neural networks for continuous object orientation estimation. *arXiv preprint arXiv :1702.01499*, 2017.
- [58] Joko Hariyono, Ajmal Shahbaz, and Kang-Hyun Jo. Estimation of walking direction for pedestrian path prediction from moving vehicle. In *2015 IEEE/SICE International Symposium on System Integration (SII)*, pages 750–753. IEEE, 2015.

- [59] David Held, Sebastian Thrun, and Silvio Savarese. Learning to track at 100 fps with deep regression networks. In *European conference on computer vision*, pages 749–765. Springer, 2016.
- [60] Christopher Hinnerger, Makoto Murata, Masahito Shingyoji, Fumie Nakamura, and Yasuhiko Fujita. Vehicle blind spot detection and indicator system, November 13 2012. US Patent 8,310,353.
- [61] Ya-Wen Hsu, Yue-Sheng Ciou, and Jau-Woei Perng. Object recognition system design in regions of interest based on adaboost algorithm. In *2017 20th International Conference on Information Fusion (Fusion)*, pages 1–5. IEEE, 2017.
- [62] Tsung-Ming Huang. Vehicle blind spot detection system, March 4 2021.
- [63] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3. 6m : Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE transactions on pattern analysis and machine intelligence*, 36(7) :1325–1339, 2013.
- [64] Marina Ivasic-Kos, Mate Kristo, and Miran Pobar. Person detection in thermal videos using yolo. In *Proceedings of SAI Intelligent Systems Conference*, pages 254–267. Springer, 2019.
- [65] Gunnar Johansson and Kåre Rumar. Drivers’ brake reaction times. *Human factors*, 13(1) :23–27, 1971.
- [66] Claudio Rosito Jung and Christian Roberto Kelber. A lane departure warning system based on a linear-parabolic lane model. In *IEEE Intelligent Vehicles Symposium, 2004*, pages 891–895. IEEE, 2004.
- [67] Hyungil Kim, Joseph L Gabbard, Alexandre Miranda Anon, and Teruhisa Misu. Driver behavior and performance with augmented reality pedestrian collision warning : An outdoor user study. *IEEE transactions on visualization and computer graphics*, 24(4) :1515–1524, 2018.
- [68] Jae Suk Kim and Tae Bong Noh. Apparatus and method for controlling a rear cross traffic alert, June 18 2020. US Patent App. 16/710,530.
- [69] Jong Hyun Kim, Ganbayar Batchuluun, and Kang Ryoung Park. Pedestrian detection based on faster r-cnn in nighttime by fusing deep convolutional features of successive images. *Expert Systems with Applications*, 114 :15–33, 2018.
- [70] Takuya Kobayashi, Akinori Hidaka, and Takio Kurita. Selection of histograms of oriented gradients features for pedestrian detection. In *International conference on neural information processing*, pages 598–607. Springer, 2007.

- [71] Yoshiki Kohari, Jun Miura, and Shuji Oishi. Cnn-based human body orientation estimation for robotic attendant. In *IAS-15 Workshop on Robot Perception of Humans*, 2018.
- [72] Tetsuya Komoguchi, Tomoya Kawasaki, and Yoshihiko Takahashi. Rear cross traffic alert device, June 23 2015. US Patent 9,064,408.
- [73] Iuliia Kotseruba, Amir Rasouli, and John K Tsotsos. Joint attention in autonomous driving (jaad). *arXiv preprint arXiv :1609.04741*, 2016.
- [74] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2) :83–97, 1955.
- [75] Wenbo Lan, Jianwu Dang, Yangping Wang, and Song Wang. Pedestrian detection based on yolo network model. In *2018 IEEE international conference on mechatronics and automation (ICMA)*, pages 1547–1551. IEEE, 2018.
- [76] John Lancaster. *Engineering Catastrophes Causes and Effects of Major Accidents*. CRC Press, 2005.
- [77] Sukhan Lee, Woong Kwon, and Jae-Won Lee. A vision based lane departure warning system. In *Proceedings 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human and Environment Friendly Robots with High Intelligence and Emotional Quotients (Cat. No. 99CH36289)*, volume 1, pages 160–165. IEEE, 1999.
- [78] Joël Leleve, Abdelaziz Bensrhair, and Julien Rebut. Method for the early detection of the arrival of a motor vehicle in a dark sector, November 22 2011. US Patent 8,063,935.
- [79] Chengyang Li, Dan Song, Ruofeng Tong, and Min Tang. Illumination-aware faster r-cnn for robust multispectral pedestrian detection. *Pattern Recognition*, 85 :161–171, 2019.
- [80] Dawei Li, Lihong Xu, Erik D Goodman, Yuan Xu, and Yang Wu. Integrating a statistical background-foreground extraction algorithm and svm classifier for pedestrian detection and tracking. *Integrated Computer-Aided Engineering*, 20(3) :201–216, 2013.
- [81] Qian Li, Usman Niaz, and Bernard Merialdo. An improved algorithm on violajones object detector. In *2012 10th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pages 1–6. IEEE, 2012.
- [82] Ying Li and Sharathchandra Pankanti. Intelligent headlight control using camera sensors, July 23 2013. US Patent 8,493,446.
- [83] Tianshan Liu, Kin-Man Lam, Rui Zhao, and Guoping Qiu. Deep cross-modal representation learning and distillation for illumination-invariant pedestrian detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.



- [84] Yung-Ching Liu and Ying-Chan Tung. Risk analysis of pedestrians' road-crossing decisions : Effects of age, time gap, time of day, and vehicle speed. *Safety science*, 63 :77–82, 2014.
- [85] Antonio López, Jörg Hilgenstock, Andreas Busse, Ramón Baldrich, Felipe Lumbreras, and Joan Serrat. Nighttime vehicle detection for intelligent headlight control. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 113–124. Springer, 2008.
- [86] Gareth Loy and Nick Barnes. Fast shape-based road sign detection for a driver assistance system. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, volume 1, pages 70–75. IEEE, 2004.
- [87] James R Mahalik, Glenn E Good, and Matt Englar-Carlson. Masculinity scripts, presenting concerns, and help seeking : Implications for practice and training. *Professional Psychology : Research and Practice*, 34(2) :123, 2003.
- [88] John-Jairo Martinez and Carlos Canudas-de Wit. A safe longitudinal control for adaptive cruise control and stop-and-go scenarios. *IEEE Transactions on control systems technology*, 15(2) :246–258, 2007.
- [89] Markus Maurer and A Eskandarian. Forward collision warning and avoidance. *Handbook of intelligent vehicles*, pages 657–687, 2012.
- [90] METLE. *Statistics report of road accidents for the year 2017*. Ministry of Environment, Transport, Logistics and Water, 2018.
- [91] Harald Michi, Michael Scherl, and Werner Uhler. Speed control with stop function, October 3 2006. US Patent 7,117,077.
- [92] Andreas Mogelmose, Mohan Manubhai Trivedi, and Thomas B Moeslund. Vision-based traffic sign detection and analysis for intelligent driver assistance systems : Perspectives and survey. *IEEE Transactions on Intelligent Transportation Systems*, 13(4) :1484–1497, 2012.
- [93] Eric N Mortensen, Hongli Deng, and Linda Shapiro. A sift descriptor with global context. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 184–190. IEEE, 2005.
- [94] Sandipann P Narote, Pradnya N Bhujbal, Abbhilasha S Narote, and Dhiraj M Dhane. A review of recent advances in lane detection and departure warning system. *Pattern Recognition*, 73 :216–234, 2018.
- [95] Guanghan Ning, Zhi Zhang, Chen Huang, Xiaobo Ren, Haohong Wang, Canhui Cai, and Zhihai He. Spatially supervised recurrent convolutional neural networks for visual object tracking. In *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–4. IEEE, 2017.

- [96] William S Noble. What is a support vector machine? *Nature biotechnology*, 24(12) :1565–1567, 2006.
- [97] NTSB. *Vehicle Automation Report*. National Transportation Safety Board, 2018.
- [98] ONISR. *Bilan 2018 de la sécurité routière*. Observatoire National Interministériel de la Sécurité Routière, 2018.
- [99] World Health Organisation. *Pedestrian Safety : A Road Safety Manual For Decision-Makers and Practitioners*. Geneva, Switzerland, 3rd edition, 2013.
- [100] Jennifer A Oxley, Elfriede Ihlen, Brian N Fildes, Judith L Charlton, and Ross H Day. Crossing roads safely : an experimental study of age differences in gap selection by pedestrians. *Accident Analysis & Prevention*, 37(5) :962–971, 2005.
- [101] Chih-Wei Pai, Ping-Ling Chen, Shiao-Tzu Ma, Shan-Hong Wu, Václav Linkov, and Hon-Ping Ma. Walking against or with traffic? evaluating pedestrian fatalities and head injuries in taiwan. *BMC public health*, 19(1) :1280, 2019.
- [102] Eleonora Papadimitriou, Sylvain Lassarre, and George Yannis. Human factors of pedestrian walking and crossing behaviour. *Transportation research procedia*, 25 :2002–2015, 2017.
- [103] Constantine Papageorgiou and Tomaso Poggio. A trainable system for object detection. *International journal of computer vision*, 38(1) :15–33, 2000.
- [104] Dănuț Ovidiu Pop, Alexandrina Rogozan, Clement Chatelain, Fawzi Nashashibi, and Abdelaziz Bensrhair. Multi-task deep learning for pedestrian detection, action recognition and time to cross prediction. *IEEE Access*, 7 :149318–149327, 2019.
- [105] Dănuț Ovidiu Pop, Alexandrina Rogozan, Fawzi Nashashibi, and Abdelaziz Bensrhair. Incremental cross-modality deep learning for pedestrian recognition. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pages 523–528. IEEE, 2017.
- [106] Danut Ovidiu Pop, Alexandrina Rogozan, Fawzi Nashashibi, and Abdelaziz Bensrhair. Pedestrian recognition using cross-modality learning in convolutional neural networks. *IEEE Intelligent Transportation Systems Magazine*, 13(1) :210–224, 2020.
- [107] APROSYS Project. Advanced protection systems (aprosys), 2009. .
- [108] WATCH-OVER Project. Vehicle-to-vulnerable road user cooperative communication and sensing technologies to improve transport safety, 2009. .
- [109] Hongquan Qu, Tongyang Yuan, Zhiyong Sheng, and Yuan Zhang. A pedestrian detection method based on yolov3 model and image enhanced by retinex. In *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–5. IEEE, 2018.

- [110] Manoharprasad K Rao, Mark Cuddihy, and Eric L Reed. System and method for combined blind spot detection and rear crossing path collision warning, October 8 2013. US Patent 8,552,848.
- [111] Julien Rebut, Abdelaziz Bensrhair, and Joël Leleve. Method and apparatus for determining the distance of visibility of the driver of a vehicle, November 22 2004. European Patent EP1422663.
- [112] Julien Rebut, Abdelaziz Bensrhair, and Joël Leleve. Techniques for early detection of arrival at dark area, November 22 2004. Japanese Patent JP2006298362.
- [113] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once : Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.
- [114] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn : Towards real-time object detection with region proposal networks. *arXiv preprint arXiv :1506.01497*, 2015.
- [115] John-Paul Renno, Dimitrios Makris, and Graeme A Jones. Object classification in visual surveillance using adaboost. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.
- [116] Dan Rosenbaum, Amiad Gurman, Yonatan Samet, Gideon P Stein, and David Aloni. Pedestrian collision warning system, January 12 2016. US Patent 9,233,659.
- [117] Tova Rosenbloom, Adar Ben-Eliyahu, and Dan Nemrodov. Children’s crossing behavior with an accompanying adult. *Safety Science*, 46(8) :1248–1254, 2008.
- [118] Yassmina Saadna and Ali Behloul. An overview of traffic sign detection and classification methods. *International journal of multimedia information retrieval*, 6(3) :193–210, 2017.
- [119] Zakaria Sabir and Aouatif Amine. Connected vehicles using ndn for intelligent transportation systems. In *Proceedings of the international conference on industrial engineering and operations management*, volume 2018, pages 2433–2441, 2018.
- [120] Zakaria Sabir and Aouatif Amine. Ndn vs tcp/ip : Which one is the best suitable for connected vehicles ? In *Recent Advances in Mathematics and Technology*, pages 151–159. Springer, 2020.
- [121] Zakaria Sabir, Safaâ Dafrallah, and Aouatif Amine. A novel solution to prevent accidents using v2i in moroccan smart cities. In *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, pages 621–625. IEEE, 2019.

- [122] Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. Dynamic routing between capsules. In *Advances in Neural Information Processing Systems*, pages 3856–3866, 2017.
- [123] Kenneth Schofield and Niall R Lynam. Vehicle blind spot detection display system, July 28 1998. US Patent 5,786,772.
- [124] Kenneth Schofield and Niall R Lynam. Vehicle blind spot detection display system, July 27 1999. US Patent 5,929,786.
- [125] Pushkar Sevekar and SB Dhonde. Nighttime vehicle detection for intelligent headlight control : A review. In *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, pages 188–190. IEEE, 2016.
- [126] EA Sirotkin, AS Martyanov, EV Solomin, and SV Kozlov. Emergency braking system for the wind turbine. In *2016 2nd International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*, pages 1–4. IEEE, 2016.
- [127] Virginia P Sisiopiku and D Akin. Pedestrian behaviors at and perceptions towards various pedestrian facilities : an examination based on observation and survey data. *Transportation Research Part F : Traffic Psychology and Behaviour*, 6(4) :249–274, 2003.
- [128] Diana Smart, Suzanne Vassallo, Ann Sanson, Samantha Cockfield, Anne Harris, and Warren Harrison. In the driver’s seat : Understanding young adults’ driving behaviour. 2005.
- [129] Bahman Soheilian, Aurore Arlicot, and Nicolas Paparoditis. Extraction de panneaux de signalisation routière dans des images couleurs. In *Reconnaissance des Formes et Intelligence Artificielle*, pages 1–8, 2010.
- [130] Bahman Soheilian, Nicolas Paparoditis, and Bruno Vallet. Detection and 3d reconstruction of traffic signs from multiple view color images. *ISPRS journal of photogrammetry and remote sensing*, 77 :1–20, 2013.
- [131] Lisa K Spainhour, Isaac A Wootton, John O Sobanjo, and Patrick A Brady. Causative factors and trends in florida pedestrian crashes. *Transportation research record*, 1982(1) :90–98, 2006.
- [132] Raphael E Stern, Shumo Cui, Maria Laura Delle Monache, Rahul Bhadani, Matt Bunting, Miles Churchill, Nathaniel Hamilton, Hannah Pohlmann, Fangyu Wu, Benedetto Piccoli, et al. Dissipation of stop-and-go waves via control of autonomous vehicles : Field experiments. *Transportation Research Part C : Emerging Technologies*, 89 :205–221, 2018.

- [133] Frédéric Suard, Alain Rakotomamonjy, Abdelaziz Bensrhair, and Alberto Broggi. Pedestrian detection using infrared images and histograms of oriented gradients. In *2006 IEEE Intelligent Vehicles Symposium*, pages 206–212. IEEE, 2006.
- [134] Amin Tahmasbi-Sarvestani, Hossein Nourkhiz Mahjoub, Yaser P Fallah, Ehsan Moradi-Pari, and Oubada Abuchaar. Implementation and evaluation of a cooperative vehicle-to-pedestrian safety application. *IEEE Intelligent Transportation Systems Magazine*, 9(4) :62–75, 2017.
- [135] Levasseur Tellis, Gerald H Engelman, John Vincent Bond III, John Patrick Joyce, Stephen Samuel, Jeffrey Dan Rupp, Gyu Myeong Joh, M Nabeel Tarabishy, Nigel John Clarke, Timothy Zwicky, et al. Stop and go adaptive cruise control system, March 16 2004. US Patent 6,708,099.
- [136] Andras Varhelyi. Drivers’ speed behaviour at a zebra crossing : a case study. *Accident Analysis & Prevention*, 30(6) :731–743, 1998.
- [137] Gul Varol, Javier Romero, Xavier Martin, Naureen Mahmood, Michael J Black, Ivan Laptev, and Cordelia Schmid. Learning from synthetic humans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 109–117, 2017.
- [138] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, volume 1, pages I–I. IEEE, 2001.
- [139] Greg Welch, Gary Bishop, et al. An introduction to the kalman filter. 1995.
- [140] WHO. *Global status report on road safety 2018*. World Health Organisation, 2018.
- [141] WHO. *Road Traffic Injuries*. World Health Organisation, 2018.
- [142] Nicolai Wojke, Alex Bewley, and Dietrich Paulus. Simple online and realtime tracking with a deep association metric. In *2017 IEEE international conference on image processing (ICIP)*, pages 3645–3649. IEEE, 2017.
- [143] Lior Wolf, Tal Hassner, and Yaniv Taigman. Descriptor based methods in the wild. In *Workshop on faces in’real-life’images : Detection, alignment, and recognition*, 2008.
- [144] Bing Yu, Weigong Zhang, and Yingfeng Cai. A lane departure warning system based on machine vision. In *2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, volume 1, pages 197–201. IEEE, 2008.

- [145] CHIU Yu-Hao. Blind spot detection device for vehicle, method thereof, and steering wheel device, April 8 2021. US Patent App. 17/009,687.
- [146] Eduard Zadobrischi and Mihai Negru. Pedestrian detection based on tensorflow yolov3 embedded in a portable system adaptable to vehicles. In *2020 International Conference on Development and Application Systems (DAS)*, pages 21–26. IEEE, 2020.
- [147] Shuqing Zeng. Cross traffic collision alert system, July 31 2012. US Patent 8,232,872.
- [148] Shuqing Zeng, Xiang Gong, and Bakhtiar B Litkouhi. Method and apparatus for rear cross traffic avoidance, May 16 2017. US Patent 9,650,026.
- [149] Nabil Zerrouki, Fouzi Harrou, Ying Sun, and Amrane Houacine. Vision-based human action classification using adaptive boosting algorithm. *IEEE Sensors Journal*, 18(12) :5115–5121, 2018.
- [150] Haodi Zhang, Dănuț Ovidiu Pop, Alexandrina Rogozan, and Abdelaziz Bensrhair. Accelerate high resolution image pedestrian detection with non-pedestrian area estimation. *IEEE Access*, 9 :8625–8636, 2021.
- [151] Hui Zhang, Yu Du, Shurong Ning, Yonghua Zhang, Shuo Yang, and Chen Du. Pedestrian detection method based on faster r-cnn. In *2017 13th International Conference on Computational Intelligence and Security (CIS)*, pages 427–430. IEEE, 2017.
- [152] Liliang Zhang, Liang Lin, Xiaodan Liang, and Kaiming He. Is faster r-cnn doing well for pedestrian detection? In *European conference on computer vision*, pages 443–457. Springer, 2016.
- [153] Tao Zhang, Jie Zou, and Wenjing Jia. Fast and robust road sign detection in driver assistance systems. *Applied Intelligence*, 48(11) :4113–4127, 2018.
- [154] Congcong Zhao and Bin Chen. Real-time pedestrian detection based on improved yolo model. In *2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, volume 2, pages 25–28. IEEE, 2019.
- [155] Xiaotong Zhao, Wei Li, Yifang Zhang, T Aaron Gulliver, Shuo Chang, and Zhiyong Feng. A faster rcnn-based pedestrian detection system. In *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, pages 1–5. IEEE, 2016.