



جامعة العربي التبسي - تبسة
Université Larbi Tébessi - Tébessa

République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique
Université Echahid Chiekh Larbi Tébessi - Tébessa
Faculté des Sciences Exactes et des Sciences de la Nature et
de la Vie



كلية العلوم الدقيقة وعلوم الطبيعة والبيئة
FACULTÉ DES SCIENCES EXACTES
ET DES SCIENCES DE LA NATURE ET DE LA VIE

Département : Mathématiques et Informatique

Mémoire de fin d'études
Pour l'obtention du diplôme de MASTER
Domaine : Mathématiques et Informatiques
Filière : Informatique
Option : Système d'information.
Thème :

**Conception d'un système de détection automatique des
panneaux routiers approche basée « deep learning »**

Présenté Par :
Bouhaha Malek

Devant le jury :

Mr : Bendjenna Hakim	Pr	Université Larbi Tébessi	Président
Mme : Ghrieb Nawel	MCB	Université Larbi Tébessi	Examineur
Mr : Bennour Akram	MCB	Université Larbi Tébessi	Encadrant

Date de soutenance : 06/06/2023.

REMERCIEMENT

Tout d'abord, je tiens à exprimer ma gratitude à Allah, le tout-puissant, qui m'a accordé la force, l'intelligence et la patience nécessaires pour mener à bien ce modeste travail.

Je souhaite également exprimer mes sincères remerciements à mon encadrant, Dr. Bennour Akram, pour sa disponibilité, son aide précieuse et ses conseils tout au long de la préparation de ce travail.

Mes remerciements vont également à tous les membres du jury pour leur disponibilité et leur acceptation d'examiner et d'évaluer mon travail.

Je tiens également à remercier tous les enseignants du département de Mathématiques et Informatique.

Enfin, je voudrais exprimer ma gratitude envers toutes les personnes qui m'ont apporté leur aide, que ce soit directement ou indirectement. Je suis sincèrement reconnaissant envers eux.

DEDICACE

À mes chers parents, ma femme et mes enfants, vous êtes mon monde.

Votre amour et votre soutien sont la force motrice derrière mes réalisations

الملخص

توفر أنظمة الكشف عن لافتات الطريق والتعرف عليها مستوى إضافي لمساعدة السائق، مما يعزز سلامة الركاب والعملاء بطرق تعزز سلامة المركبات. كجزء من أنظمة مساعدة السائق المتقدمة، يمكن استخدام التعرف على إشارات الطريق للسائقين (خاصة أولئك ذوي الاحتياجات الخاصة) عن طريق تنبيههم إلى وجود إشارات الطريق للحد من مخاطر التشتيت. التعب وسوء الرؤية والظروف الجوية. على الرغم من أن العديد من أنظمة التعرف على إشارات الطريق قد تم اقتراحها، إلا أن تصميم خوارزمية قوية لا يزال مشكلة بحثية مفتوحة. وقد أظهر استخدام نماذج التعلم العميق، مثل الشبكات العصبية التلافيفية (CNN) نتائج واعدة في هذا المجال.

في هذا العمل، تم تطوير نموذج التعلم العميق لتصنيف إشارات الطرق باستخدام استراتيجيتين: الشبكة العصبية التلافيفية ونقل التعلم. حيث تم تدريب النموذج على مجموعة بيانات مرجعية للتعرف على إشارات المرور الألمانية (GTSRB) نظرًا لأن مجموعة البيانات قد تكون غير متوازنة، فقد تم إجراء زيادة في البيانات لتحسين أداء النموذج وتم تقييم الأداء باستخدام مقاييس مختلفة مثل مصفوفة الارتباك والدقة.

النتائج التي تم الحصول عليها مشجعة للغاية، مع دقة تصل إلى 99% حققها نموذج CNN المقترح. ما يعني أن النموذج قادر على تصنيف 99% من إشارات المرور بشكل صحيح في 43 فئة محددة مستخدمة في الدراسة.

ABSTRACT

It is interesting to note that computerized recognition of road signs is considered an essential element of modern traffic systems. The use of deep learning models, such as convolutional neural networks (CNNs), has shown promising results in this field.

In this work, a deep learning model was developed to classify road signs using two strategies: a convolutional neural network and transfer learning. The model was trained on a reference dataset of German traffic sign recognition (GTSRB). Since the dataset may exhibit imbalance, data augmentation was performed to enhance the model's performance. The model's performance was evaluated using different measures such as confusion matrix and accuracy.

The results obtained are very encouraging, with a 99% accuracy achieved by the proposed CNN model. This means that the model is capable of correctly classifying 99% of the traffic signs across the 43 specific classes used in the study.

RESUME

Il est intéressant de constater que la reconnaissance informatisée des panneaux routiers est considérée comme un élément essentiel des systèmes de circulation modernes. L'utilisation de modèles d'apprentissage profond, tels que les réseaux de neurones convolutifs (CNN), a montré des résultats prometteurs dans ce domaine.

Dans ce travail, un modèle d'apprentissage profond a été développé pour classer les panneaux routiers en utilisant deux stratégies : un réseau de neurones convolutif et l'apprentissage par transfert. Le modèle a été entraîné sur un ensemble de données de référence de reconnaissance des panneaux de signalisation allemands (GTSRB). Étant donné que l'ensemble de données peut présenter un déséquilibre, une augmentation des données a été effectuée pour améliorer les performances du modèle. Les performances ont été évaluées à l'aide de différentes mesures telles que la matrice de confusion et la précision.

Les résultats obtenus sont très encourageants, avec une précision de 99 % atteinte par le modèle CNN proposé. Cela signifie que le modèle est capable de classer correctement 99 % des panneaux de signalisation dans les 43 classes spécifiques utilisées dans l'étude.

TABLE DE MATIERES

Introduction générale	1
 Chapitre 01 : Système de reconnaissance des panneaux routières	
Introduction	3
1. Les Systèmes Avancés d'Aide à la Conduite.....	3
2. Les types de capteurs utilisés pour les ADAS.....	5
3. Les ADAS basés sur la vision	6
4. Système de reconnaissance des panneaux routières (Traffic Sign Recognition – TSR)	13
5. Processus de reconnaissance des panneaux routières.....	16
6. Défis majeurs des TSR.....	18
Conclusion.....	18
 Chapitre 02 : Deep Learning et la reconnaissance des images	
Introduction	19
1. L'apprentissage profond : (Deep Learning)	19
2. La Détection des objets et la classification des images	20
3. Fonctionnement de Deep Learning	22
4. Le réseau de neurones convolutionnel (CNN)	23
5. Transfer Learning : (l'apprentissage par transfert).....	25
6. Les modèles de classification basés sur CNN	26
7. Les modèles de détections des objets basée sur CNN	29
Conclusion.....	33
 Chapitre 03 : Conception et implémentation du modèle CNN de classification	
Introduction	34
1. Classification des panneaux routières dans la littérature	34
2. La solution proposée	37
3. Environnement de travail	43
4. Expérimentation et résultats	44
Conclusion.....	51
Conclusion générale	52
Bibliographie	53

LISTE DES FIGURES

Figure 1-1. Les types d'un ADAS [2].	4
Figure 1-2. Image RVB d'une caméra monoculaire [5].	7
Figure 1-3. Une image de caméra stéréoscopique [2].	8
Figure 1-4. Une image capturée à partir d'une vidéo de caméra thermique passive [6].	9
Figure 1-5. La caméra thermique vs la caméra active et caméra monoculaire [6].	9
Figure 1-6. Exemple de détection des piétons [3].	10
Figure 1-7. Guide de stationnement sur l'ordinateur [7].	11
Figure 1-8. Exemple de la reconnaissance des panneaux routiers [8].	11
Figure 1-9. Alerte de somnolence après le quatrième dépassement du seuil Détections [9].	12
Figure 1-10. Schéma général d'un système TRS [10].	13
Figure 1-11. Les panneaux routiers de danger [11].	14
Figure 1-12. Les panneaux routiers de priorité [11].	14
Figure 1-13. Les panneaux routiers d'interdiction [11].	15
Figure 1-14. Les panneaux routiers d'obligation [11].	15
Figure 1-15. Panneaux de prescriptions particulières [11].	16
Figure 1-16. Le flux de travail de TSR [12].	16
Figure 2- 1. Classification d'images par apprentissage profond [15].	20
Figure 2- 2. La différence entre la classification [16].	22
Figure 2- 3. Parcours de filtre sur une image [17].	24
Figure 2- 4. Max/Average Pooling avec un filtre 2x2 et un pas de 2 [18].	24
Figure 2- 5. Couche de correction Relu [19].	25
Figure 2- 6. Illustration du concept de transfer learning [20].	26
Figure 2- 7. L'architecture du réseau AlexNet [23].	28
Figure 2- 8. L'architecture de VGGNet-16 [24].	29
Figure 2- 9. L'apprentissage résiduel [27].	29
Figure 2- 10. Architecture de modèle Faster R-CNN [29].	31
Figure 2- 11. Modèle YOLO [30].	32
Figure 3- 1. Solution proposé.	37
Figure 3- 2. Échantillons d'images de dataset.	38
Figure 3- 3. Liste des classes des panneaux utilisés.	39
Figure 3- 4. Architecture du modèle CNN proposé.	42
Figure 3- 5. Architecture du modèle VGG16 [45].	43
Figure 3- 6. Graphe de précision et de perte du modèle CNN proposé.	45
Figure 3- 7. Matrice de confusion du modèle CNN.	46
Figure 3- 8. Graphe de précision et de perte du méthode 2.	47
Figure 3- 9. Matrice de confusion du méthode 2.	48
Figure 3- 10. Graphe de précision et de perte du méthode 2.	49
Figure 3- 11. Matrice de confusion du méthode 3.	49

LISTE DES TABLEAUX

Tableau 1. Résultat d'évaluation du model CNN pour différentes valeurs de batch size.....	45
Tableau 2. Résultat d'évaluation du méthode 1 de transfer learning pour différentes valeurs de batch size.....	46
Tableau 3. Résultat d'évaluation du méthode 2 de transfer learning pour différentes valeurs de batch size.....	48
Tableau 4. Meilleur résultat de chaque méthode.....	50
Tableau 5. Comparaison du résultats	50

INTRODUCTION GENERALE

Ces dernières années, le domaine des transports a connu des avancées significatives avec l'émergence des systèmes avancés d'aide à la conduite (ADAS). Ces systèmes visent à renforcer la sécurité de conduite, à améliorer la fluidité du trafic et à réduire les accidents grâce à l'intégration de technologies de pointe. Parmi les composants cruciaux de l'ADAS, la reconnaissance des panneaux de signalisation (TSR) joue un rôle central en permettant aux véhicules intelligents de percevoir et de comprendre l'environnement complexe de la circulation [2].

La reconnaissance des panneaux routiers implique la détection, la classification et l'interprétation de divers panneaux de routiers, y compris les limites de vitesse, les panneaux d'arrêt, les panneaux de cession et les panneaux d'avertissement. Il fournit des informations importantes au conducteur et aux systèmes de contrôle autonomes du véhicule, facilitant une prise de décision éclairée et une navigation efficace. Traditionnellement, les algorithmes TSR s'appuyaient sur des techniques de vision par ordinateur conventionnelles, telles que l'extraction de caractéristiques et la correspondance de modèles. Cependant, avec l'avènement de l'apprentissage en profondeur, une nouvelle ère s'est ouverte dans le domaine de la TSR, révolutionnant ses performances et ses capacités.

Les techniques d'apprentissage en profondeur, en particulier les réseaux de neurones convolutifs (CNN), ont connu un succès sans précédent dans diverses tâches de vision par ordinateur, notamment la classification d'images, la détection d'objets et la segmentation. Leur capacité à apprendre et à extraire automatiquement des caractéristiques complexes à partir de données brutes les a propulsés à l'avant-garde de la recherche TSR.

Notre but dans ce projet est de classifier des images des panneaux routiers, donnant naissance à des systèmes capables de renforcer la nature irréprochable de l'ADAS (Advanced Driver Assistance System). Dans ce travail nous allons utiliser les techniques de l'apprentissage profond sur une large ensemble des images de panneaux routières pour une classification multi class (43 classes). Ce mémoire est organisé en trois chapitres :

- Le premier chapitre est consacré à la présentation du domaine d'application, pour expliquer en générale le ADAS, le TRS et ses étapes principales.
- Le deuxième chapitre a été consacré à la description de fonctionnement du Deep Learning et les réseaux de neurones convolutifs.

- Dans le troisième chapitre, nous allons montrer la partie expérimentale de notre travail. Nous avons détaillé l'architecture et les paramètres choisis et discuter les différents résultats obtenus.

CHAPITRE 01 : Système de reconnaissance des panneaux routières

1. Introduction

Les accidents de la route constituent un problème majeur à l'échelle mondiale, avec plus de 1,2 million de décès et plus de 50 millions de blessés chaque année, selon l'Organisation Mondiale de la Santé [1]. En Algérie, en 2018, 23 024 accidents de la route ont entraîné 3 310 décès et 23 570 blessés, avec le facteur humain comme principale cause de ces accidents, suivi de l'état des véhicules et des routes et de l'environnement. Les principales raisons de ces accidents sont l'excès de vitesse, le dépassement dangereux, l'imprudence des piétons, les manœuvres dangereuses et le non-respect de la signalisation. Pour tenter de résoudre ce problème, les ADAS (Advanced Driver Assistance Systems) ont été développés pour aider les conducteurs à éviter les accidents et à conduire plus prudemment. Les exemples d'ADAS incluent les systèmes d'alerte de collision, les systèmes de maintien de voie et les systèmes de reconnaissance de signalisation routière.

Dans ce premier chapitre, nous présentons les ADAS et les systèmes de reconnaissance des plaques routières, en décrivant les technologies utilisées, notamment celles basées sur la vision par ordinateur, ainsi que les défis à relever lors de leur mise en pratique.

1. Les Systèmes Avancés d'Aide à la Conduite

Le système avancé d'assistant de conduite (ADAS) est un système intégré intelligent conçu pour être intégré aux véhicules. Ces dernières années, il a accompli de grands progrès, rendant la conduite facile et plus efficace. L'ADAS utilise les informations fournies par les capteurs installés sur le véhicule pour percevoir l'environnement autour de celui-ci. Ces systèmes peuvent être classifiés selon les types de capteurs utilisés, la plupart d'entre eux étant principalement basés sur la vision (caméra), la détection de la lumière et la télémétrie (lidar), ainsi que la détection et la télémétrie radio (radar). D'autres technologies telles que les Photonic Mixer Device (PMD), les Global Positioning Systems (GPS) et les Inertial Measurement Units (IMU) sont également utilisées pour améliorer les fonctionnalités de ces technologies [2]. La figure 1.1 montre les différentes technologies utilisées dans les applications des systèmes avancés d'aides à la conduite.

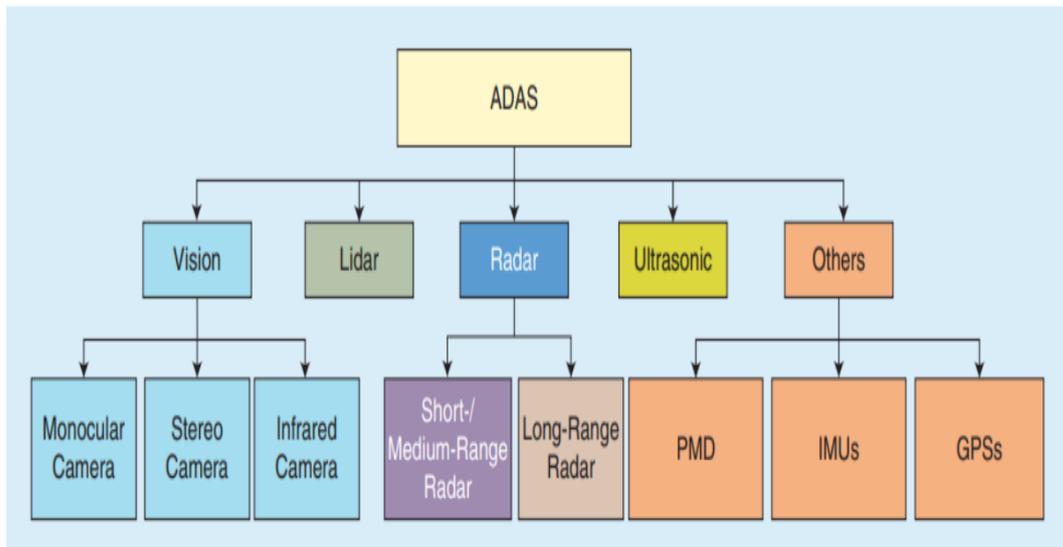


Figure 1-1. Les types d'un ADAS [2].

La plupart des programmes de systèmes avancés d'aide à la conduite (ADAS) effectuent des opérations comparables destinées à aider les conducteurs. Initialement, ADAS recueille des données sur l'environnement du véhicule à l'aide des capteurs susmentionnés. La nature des données acquises est ensuite déterminée, y compris toute imagerie visuelle, signaux lumineux et autres informations pertinentes. Après ce point, les informations sont affinées par ADAS en utilisant des techniques de prétraitement et de traitement en conjonction avec des modèles d'intelligence artificielle comme l'apprentissage automatique et l'apprentissage en profondeur. Ces méthodes facilitent la compréhension et l'analyse de l'environnement du véhicule [2]. Enfin, ce système effectue un choix après avoir analysé les informations qu'il a recueillies et détecté des situations importantes comme une collision imminente avec une autre voiture ou une sortie de voie du conducteur. Pour éviter un accident, ce processus décisionnel consiste à prendre le contrôle du véhicule (comme l'application des freins d'urgence) ou à avertir le conducteur. La majorité des ADAS incluent de telles caractéristiques de sécurité active comme éléments essentiels. Un certain nombre de composants connectés liés au véhicule, au conducteur humain et à l'environnement sont inclus dans les systèmes de sécurité active. Ces systèmes espèrent atténuer l'impact des accidents ou les éviter complètement en étant présents ou en fonctionnant. Les systèmes de sécurité active réagissent de manière proactive avant qu'un accident ne se produise, se différenciant des systèmes de sécurité passive qui fonctionnent après un incident [3].

2. Les types de capteurs utilisés pour les ADAS

Les ADAS sont catégorisés en fonction des types de capteurs qu'ils utilisent, lesquels sont essentiels pour fournir une perception précise de l'environnement du véhicule. Parmi les capteurs couramment utilisés figurent les capteurs de vision, le LIDAR, le RADAR et les capteurs à ultrasons. Néanmoins, des capteurs supplémentaires sont également incorporés afin de renforcer les fonctionnalités du système. Par exemple, le système de positionnement global (GPS) est exploité pour déterminer l'emplacement spatial précis du véhicule. L'unité de mesure inertielle (IMU) fournit des informations de position et une estimation de mouvement fiables, facilitant ainsi les ADAS en termes de stabilisation et de navigation. De plus, l'utilisation du Photonic Mixer Device (PMD) permet une identification optique rapide et un décodage synchrone des signaux lumineux incohérents [2].

a- Le capteur de vision

La caméra est le principal capteur visuel des systèmes avancés d'aide à la conduite (ADAS) basés sur la vision. Son travail principal consiste à enregistrer des photos qui fournissent des données visuelles au système, lui permettant d'évaluer et d'analyser l'environnement du véhicule. La caméra, en particulier, est essentielle à la détection et à l'identification d'objets. Il peut également être utilisé pour surveiller l'intérieur d'un véhicule. L'intégration de la caméra dans ADAS présente des avantages majeurs, notamment une intégration facile avec le véhicule, la rentabilité et, surtout, la capacité de fournir au système une quantité importante d'informations environnementales précises par rapport à d'autres capteurs.

b- Le capteur de LIDAR

Le LIDAR, acronyme de Light Detection and Ranging, est une technique de télédétection à base de laser utilisée pour détecter l'environnement du véhicule. Les systèmes LIDAR peuvent générer des images 3D haute résolution et construire une représentation 3D complète à 360 degrés de l'environnement, avec des informations exactes sur la profondeur, en produisant un faisceau laser et en mesurant le temps nécessaire au faisceau pour revenir au capteur après réflexion sur les objets. Les avantages significatifs du LIDAR sont ses capacités d'identification d'objets et sa portée opérationnelle exceptionnelle, qui peut atteindre des distances allant jusqu'à 60 mètres. Cependant, il est crucial de noter que les capteurs LIDAR se distinguent par leur grande taille, leur poids et leur coût, et leur précision peut être entravée par de mauvaises conditions météorologiques telles que la pluie ou le brouillard [2].

c- Le capteur de RADAR

Le capteur de détection et de télémétrie radio RADAR, est un composant clé des systèmes d'estimation de distance sécurisés, étant donné sa précision de mesure. La fonction du RADAR consiste à émettre des micro-ondes qui, lorsqu'elles entrent en contact avec un objet, provoquent un décalage de fréquence (acoustique ou électromagnétique) entre l'émission et la réception du signal, ce qui permet de mesurer la distance entre l'émetteur et le récepteur en temps réel. Par rapport au LIDAR, le RADAR présente l'avantage de pouvoir détecter des objets à de plus longues distances et d'être moins sensible aux conditions météorologiques telles que la pluie ou le brouillard. En outre, il est relativement peu coûteux. Selon leur portée, les RADAR peuvent être classés en trois catégories : courte portée (0,2 à 30 mètres), moyenne portée (30 à 80 mètres) et longue portée (80 à 200 mètres) [2].

d- Les capteurs ultrasoniques

Les capteurs ultrasoniques sont des dispositifs qui se basent sur l'utilisation d'ondes sonores pour mesurer la distance entre un objet et le capteur. Ils sont couramment utilisés dans les ADAS pour détecter des objets situés à proximité immédiate du véhicule [2].

3. Les ADAS basés sur la vision

Les constructeurs automobiles travaillent dur pour développer des systèmes ADAS basés sur la vision et dotés de meilleures fonctionnalités de sécurité. Tesla a dévoilé le système de pilote automatique, un programme ADAS exploitant les technologies de vision par ordinateur, en 2015. Cette technologie intègre des fonctions telles que le régulateur de vitesse adaptatif, le stationnement automatisé et d'autres fonctionnalités en plus de permettre la reconnaissance des voies sur la route. Il est important de souligner que malgré les capacités du système, le conducteur reste responsable et doit faire preuve d'une vigilance vigilante [4].

A. Les types des caméras utilisées

Les caméras monoculaires, les caméras stéréo et les caméras thermiques sont les trois types de caméras les plus fréquemment utilisées dans les ADAS basés sur la vision. Ces types sont décrits ci-dessous.

- **Les caméras monoculaires**

Le système de caméra monoculaire se caractérise par sa configuration à objectif unique, ce qui entraîne la fourniture d'une image à la fois et nécessite un traitement d'image moins

complexe par rapport aux autres types de caméra. Les caméras monoculaires trouvent une utilité dans une gamme d'applications, y compris la détection d'obstacles, la reconnaissance des piétons (comme illustré à la figure 1.2), la détection des panneaux, ainsi que la surveillance du conducteur dans le véhicule, comme les systèmes de détection de la fatigue. Néanmoins, il convient de noter que les images capturées par une caméra monoculaire ne possèdent pas d'informations de profondeur adéquates, ce qui les rend inadaptées à une estimation fiable de la distance.



Figure 1-2. Image RVB d'une caméra monoculaire) [5]

- **Les caméras stéréos**

Grâce à l'analyse d'appariements d'images 2D et à l'utilisation d'une carte de disparité pour déterminer la profondeur relative à l'intérieur d'une scène, ces caméras sont excellentes pour extraire des informations tridimensionnelles [2]. Les caméras stéréo sont plus performantes que les caméras monoculaires dans une variété d'applications, notamment la reconnaissance des panneaux, la détection des voies, l'identification des piétons et le calcul précis de la distance (Figure 1.3). La majorité des systèmes ADAS utilisent des caméras stéréo, qui sont souvent montées à l'intérieur des voitures derrière le rétroviseur et inclinées vers le bas pour se concentrer sur l'environnement routier.

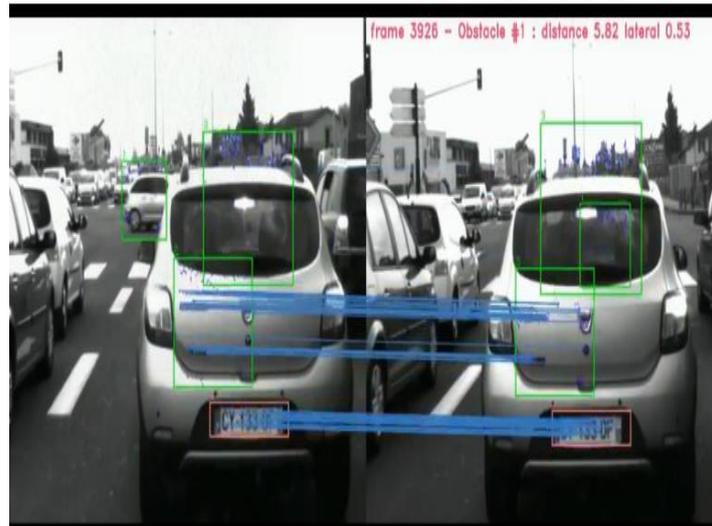


Figure 1-3. Une image de caméra stéréoscopique [2].

- **Les caméras thermiques (infrarouges)**

Les caméras thermiques peuvent être divisées en deux types : les caméras thermiques actives qui utilisent une source de lumière infrarouge proche (avec une longueur d'onde de 750 nm à 1400 nm) pour illuminer la scène, et les caméras thermiques passives qui utilisent des capteurs infrarouges pour capturer le rayonnement thermique émis par toute matière. La première peut être directement intégrée dans les véhicules et nécessite une illumination spéciale pour la scène, tandis que la seconde ne nécessite aucune illumination particulière et est largement utilisée dans les applications ADAS (comme le montre la figure 1.4) [6]. La figure 1.5 illustre la diversité entre les images capturées par une caméra numérique standard et une caméra thermique active dans des conditions de faible luminosité. Les caméras thermiques sont une solution utile pour améliorer la vision nocturne et aider les conducteurs à mieux voir dans des conditions de faible luminosité.



Figure 1-4. Une image capturée à partir d'une vidéo de caméra thermique passive [6].



Figure 1-5. La caméra thermique vs la caméra active et caméra monoculaire [6].

B. Le fonctionnement

Le système nécessite une collecte de données visuelles à partir de caméras afin de prendre des décisions en fonction de l'environnement d'un véhicule. Tout d'abord, une image (ou une vidéo) est capturée, puis sa qualité est améliorée. Ensuite, des algorithmes sont utilisés pour reconnaître l'environnement environnant du véhicule, tels que la détection des voies ou des véhicules. Selon les résultats de ces algorithmes, des décisions ou des actions sont prises, telles que des alertes, un freinage d'urgence ou un signal sonore, par exemple.

C. Exemples

- **Le système de protection des piétons**

Les accidents impliquant des piétons sont responsables d'un grand nombre de décès sur les routes à travers le monde. Ainsi, les industries et les chercheurs s'efforcent de développer des

ADAS basés sur la vision pour réduire le nombre de ces accidents. Par exemple, Mercedes-Benz a développé un système d'alerte pour les piétons et Toyota a conçu un système d'aide à l'évitement des piétons [3]. Ces systèmes de protection des piétons fonctionnent en détectant les piétons et en alertant le conducteur de leur présence grâce à des séquences d'images prises par une caméra montée sur le véhicule. (Figure 1-6).



Figure 1-6. Exemple de détection des piétons [3].

- **Le système assistant de stationnement**

C'est un système basé sur la vision qui aide les conducteurs à décider où garer leur véhicule. Ce système peut prendre des photos à 360 degrés à l'aide d'une caméra grand angle ou de quatre caméras fish-eye, et il peut évaluer les données d'image de l'avant, de l'arrière et des côtés pour déterminer les dimensions des places de stationnement disponibles. Avec la vidéo de la caméra avant en haut à gauche, la vidéo de la caméra latérale droite en haut à droite, le mouvement de l'automobile détecté par le système en bas à gauche et les instructions que le système envoie au conducteur en bas à droite, Figure 1.7 illustre l'affichage du guidage de stationnement sur un ordinateur.



Figure 1-7. Guide de stationnement sur l'ordinateur [7].

- **La détection et la reconnaissance des panneaux de routiers**

La détection et la reconnaissance des panneaux de signalisation sont des éléments clés des ADAS basés sur la vision, étant donné que ces panneaux régulent la circulation routière, signalent l'état de la route et sont omniprésents. Les panneaux se distinguent par leur forme (triangulaire, circulaire ou octogonale), leur couleur (rouge, bleu) et leur contenu. (Figure 1.8).



Figure 1-8. Exemple de la reconnaissance des panneaux routiers [8].

- **La détection de fatigue**

Il existe des risques de collision assez importants pour les conducteurs qui conduisent lorsqu'ils sont fatigués. L'étude de mouvements faciaux spécifiques est à la base des techniques de détection de la fatigue du conducteur. Pour identifier les indicateurs d'épuisement, certains systèmes évaluent les traits buccaux et oculaires. Les mesures peuvent être séparées en deux catégories : fatiguées et normales en utilisant des méthodes d'apprentissage automatique comme la méthode Viola & Jones ou l'algorithme CNN. Ces valeurs permettent de séparer les autres émotions faciales, comme le sourire et le bavardage, des symptômes de lassitude, comme le bâillement et les yeux endormis. Lorsque la fatigue d'un conducteur est reconnue, le système l'avertit de prendre des mesures, comme l'arrêt de la voiture. Un résultat d'un système de détection de fatigue est présenté dans la figure ci-dessous.

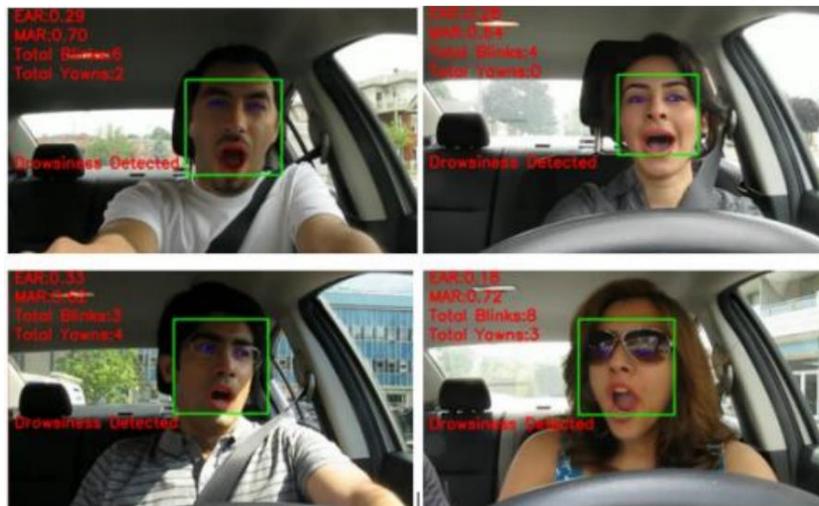


Figure 1-9. Alerte de somnolence après le quatrième dépassement du seuil Détections [9].

- **Le système de la détection des feux de circulation**

La reconnaissance des feux de circulation est un élément clé des systèmes d'assistance à la conduite basés sur la vision. Elle permet de fournir des informations cruciales au conducteur concernant les intersections et les passages piétons, et peut contribuer à réduire les accidents dus à une distraction face aux feux. Le système repère les feux de circulation dans un contexte urbain et en détermine l'état pour avertir le conducteur.

Dans ce mémoire, nous étudions les systèmes de reconnaissance des panneaux afin de créer un classifieur des panneaux routiers.

4. Système de reconnaissance des panneaux routières (Traffic Sign Recognition – TSR)

Un système de reconnaissance de panneaux routiers (TSR) est souvent équipé d'une caméra installée à l'avant du véhicule. Le principe de fonctionnement du TSR est de traiter le flux vidéo capturé par la caméra pour identifier les panneaux de signalisation présents dans la scène. Cette opération se déroule en deux étapes principales (Figure 1.10) : la détection des panneaux routiers et la reconnaissance et classification de ces panneaux. Au cours de la première étape, les attributs visuels tels que la couleur et la forme sont exploités pour localiser les signes dans les images capturées. Ensuite, dans l'étape de reconnaissance, le type de chaque signe détecté est déterminé en analysant son pictogramme. En pratique, les signes identifiés sont soit communiqués au conducteur, soit partagés avec d'autres systèmes ADAS.[10]



Figure 1-10. Schéma général d'un système TRS [10].

Les dispositifs de circulation tels que les panneaux routiers et les feux de signalisation sont essentiels pour informer, guider et avertir les usagers de la route. Ils sont cruciaux pour garantir un trafic routier de qualité et pour améliorer la sécurité en permettant un mouvement ordonné et prévisible des véhicules et des piétons. Les panneaux routiers sont conçus de manière à être facilement identifiables par les conducteurs, en utilisant des formes et des couleurs qui se démarquent clairement de leur environnement. Cette distinction visuelle permet de classer les panneaux en différentes catégories selon leur forme et leur couleur.

Les panneaux sont classifiés en cinq catégories [11] :

- **Les panneaux de danger** : ont pour fonction d'alerter les usagers de la route de situations dangereuses. Ils se reconnaissent aisément grâce à leur forme triangulaire, avec un bord rouge épais et un fond blanc ou jaune. Pour préciser le type de danger, un pictogramme est souvent représenté à l'intérieur du triangle. La figure 1.11 présente plusieurs exemples de panneaux routiers de danger.



Figure 1-11. Les panneaux routiers de danger [11].

- **Les panneaux de priorité** : permet d'informer les usagers de la route des règles spéciales de priorité à respecter aux intersections, comme montré dans la figure 1.12. Leur objectif est d'assurer une circulation ordonnée et sûre des véhicules et des piétons en réglant leur ordre de passage.

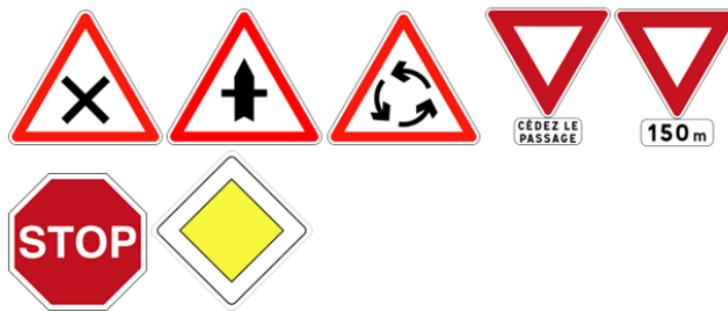


Figure 1-12. Les panneaux routiers de priorité [11].

- **Les panneaux d'interdiction (restriction)** : sont des signaux routiers qui indiquent aux usagers de la route qu'une interdiction doit être respectée. Ils ont une forme circulaire et sont dotés d'un fond blanc, jaune ou bleu, bordé d'une couronne rouge (Figure 1.13).



Figure 1-15. Panneaux de prescriptions particulières [11].

5. Processus de reconnaissance des panneaux routières

Dans la figure 1-16, nous présentons le flux de travail de TSR, qui comprend principalement trois parties. La première partie concerne le prétraitement de l'image, qui encapsule généralement l'amélioration de l'image, le redimensionnement de l'image et d'autres opérations. La deuxième partie est la détection des panneaux de signalisation routière, qui comprend deux étapes importantes : (1) l'extraction des régions candidates ; (2) la deuxième classification des panneaux de signalisation routière consiste à déterminer s'ils sont des panneaux routières ou non ; la troisième partie est la classification détaillée des panneaux routière, à savoir la catégorie spécifique de panneaux routière.

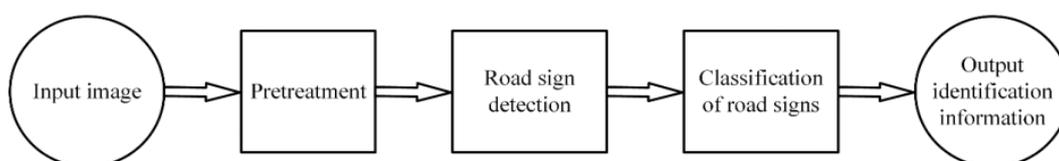


Figure 1-16. Le flux de travail de TSR [12]

a- Prétraitements des images

TSR est différente de la reconnaissance d'objets générale. L'influence irrésistible de facteurs externes tels que la brume, la pluie et la neige, ainsi que toutes sortes de conditions d'éclairage affecteront l'image de panneaux de signalisation que nous obtenons, ce qui rendra la luminosité de l'image trop faible ou avec des bruits, ce qui ne sera pas propice à la TSR. Le prétraitement de l'image améliore efficacement la qualité de l'image, réduit ou élimine même les facteurs de perturbation qui ne sont pas propices à la reconnaissance d'objets et améliore l'efficacité et la précision de la détection des panneaux de signalisation routière. Par

conséquent, pour améliorer la précision de reconnaissance des panneaux de signalisation routière, l'image doit être prétraitée avant la reconnaissance des panneaux.

b- La détection des panneaux

La deuxième partie de TSR est la détection de panneaux de signalisation routière, qui comprend principalement deux processus : (1) l'extraction de cadres candidats, (2) la classification. L'extraction de la boîte candidate (candidate box) consiste à segmenter la partie contenant les panneaux de signalisation routière dans l'image originale en fonction des caractéristiques pertinentes des panneaux de signalisation routière. La deuxième classification consiste à extraire les caractéristiques des panneaux de signalisation routière et à déterminer si les panneaux de signalisation routière appartiennent aux zones correspondantes. Les méthodes de détection de panneaux routière comprennent la détection de forme, la segmentation de couleur, l'extraction de caractéristiques combinée à un classifieur et l'algorithme de détection d'apprentissage profond. Cependant, en situation réelle, la détection n'est jamais parfaite. Outre les vrais positifs (VP) qui correspondent aux panneaux correctement détectés, il peut y avoir des fausses détections (FP). Cela est dû à la complexité de la scène routière, avec des objets qui peuvent masquer les panneaux ou qui sont très proches d'eux, ce qui complique la tâche de détection et génère des False Positives. La variété des catégories de panneaux, ainsi que les changements d'éclairage et le faible contraste qui rendent difficile la distinction entre les signes et l'arrière-plan, ne sont que quelques-uns des éléments que la détection doit prendre en compte.

c- La classification des panneaux (la reconnaissance)

La troisième partie est la classification des panneaux de signalisation routière, qui sert à déterminer la catégorie spécifique de la zone de panneaux de signalisation routière segmentée. Cette étape est réalisée à l'aide d'un classifieur préalablement entraîné sur les différentes classes de panneaux, lui permettant d'attribuer la classe appropriée à chaque panneau détecté. Les panneaux ainsi identifiés peuvent être signalés au conducteur ou transmis à d'autres systèmes d'aide à la conduite (ADAS) [6]. Les méthodes de classification traditionnelles encapsulent la combinaison d'un vecteur de caractéristiques et d'un classifieur, ainsi que la correspondance de modèles. Avec le développement continu de l'intelligence artificielle, l'application de l'apprentissage en profondeur dans TSR a donné de meilleurs résultats [13]. Par rapport aux méthodes traditionnelles, la méthode de classification d'apprentissage en profondeur est capable d'apprendre automatiquement les caractéristiques sans intervention

humaine. Plus il y a de données d'entraînement, meilleures sont les performances de classification. En particulier, l'apparition du réseau de neurone convolutif (CNN) améliore la capacité d'extraction de caractéristiques du réseau neuronal. Il est devenu l'un des réseaux les plus représentatifs de l'apprentissage en profondeur, donc il est largement appliqué à la classification des objets.

6. Défis majeurs des TSR

La reconnaissance des panneaux de signalisation est difficile car elle nécessite de trouver et de reconnaître des plaques dans des environnements compliqués, ce qui peut créer un certain nombre de problèmes, tels que :

- Changements d'éclairage : les conditions d'éclairage peuvent varier en fonction de facteurs tels que l'heure de la journée, la saison, la couverture nuageuse et les conditions météorologiques, affectant la visibilité des panneaux de signalisation.
- Décoloration : une exposition prolongée au soleil et à la pluie peut entraîner une décoloration des signes avec le temps, ce qui réduit la clarté.
- Variations d'échelle : La taille d'un panneau de signalisation apparaît différemment selon sa distance à la caméra, ce qui oblige à tenir compte des changements d'échelle.
- Rotation et translation : les panneaux de signalisation peuvent être capturés sous différents angles ou orientations, ce qui entraîne des variations dans leur rotation et leur translation dans la scène.
- Présence d'objets similaires : la présence d'autres objets dans la scène qui partagent des couleurs ou des formes similaires peut présenter des difficultés pour distinguer les panneaux de signalisation des autres éléments.

Ces facteurs contribuent à la complexité de la reconnaissance des panneaux routiers.

7. Conclusion

Ce chapitre a introduit les principes fondamentaux des systèmes avancés d'aide à la conduite (ADAS) et des systèmes de reconnaissance des panneaux routiers (TSR). Nous avons examiné en détail le processus de développement des systèmes de reconnaissance de panneaux routiers basés sur la vision, qui sont conçus pour améliorer la sécurité routière en réduisant le nombre d'accidents.

CHAPITRE 02 : Deep Learning et la reconnaissance des images

Introduction

De nombreuses inventions ont été inspirées par la nature, ce qui justifie l'idée d'utiliser l'architecture du cerveau comme modèle pour construire une machine intelligente. C'est précisément cette idée qui a conduit au développement des réseaux de neurones artificiels, qui jouent un rôle central dans l'apprentissage profond. Ces réseaux sont à la fois puissants et évolutifs, ce qui les rend particulièrement adaptés aux tâches d'apprentissage automatique complexes et volumineuses, telles que la classification de milliards d'images, la reconnaissance vocale et la recommandation de vidéos.

Dans ce chapitre, nous allons d'abord présenter le fonctionnement de l'apprentissage profond, ainsi que les différentes architectures utilisées. Nous aborderons également les techniques spécifiquement dédiées à la classification d'images.

1. L'apprentissage profond : (Deep Learning)

L'apprentissage profond est une branche de l'apprentissage automatique qui vise à former des machines à exécuter des tâches humaines telles que la reconnaissance d'objets visuels et l'identification de sons et d'images. Contrairement à l'approche traditionnelle qui consiste à organiser les données, l'apprentissage profond définit ses propres paramètres de base, ce qui permet à la machine d'apprendre de manière autonome.[14]

Le deep learning (apprentissage profond) est un sous-domaine de l'intelligence artificielle et plus spécifiquement de l'apprentissage machine. Il utilise des réseaux de neurones artificiels spécifiques et partage certaines caractéristiques avec ces techniques, notamment la capacité à apprendre de façon autonome. Pour cela, il existe trois approches d'apprentissage en fonction du contexte d'utilisation :

- **L'apprentissage supervisé** : Cette approche consiste à entraîner un programme informatique en lui fournissant des données préétiquetées, puis à utiliser cet apprentissage pour classer de nouvelles observations.
- **L'apprentissage non supervisé** : Ici, l'objectif est de créer un modèle qui structure l'information sans étiquettes. Les données ne sont pas préalablement classifiées et l'objectif est de découvrir les classes ou structures sous-jacentes.
- **L'apprentissage par renforcement** : Cette approche implique l'apprentissage d'actions à prendre dans un environnement en évolution afin de maximiser une récompense. La machine apprend à travers l'expérience et à travers les retours qu'elle reçoit sur ses actions.

2. La Détection des objets et la classification des images

a- Classification

La classification est une méthode utilisée pour déterminer et assigner une catégorie à de nouvelles observations en se basant sur des données préexistantes (données d'entraînement). Il existe deux types de classification :

- **La classification binaire** : Dans ce cas, les données entrantes sont classées dans l'une des deux catégories possibles.
- **La classification multi-classe** : Ce type de classification divise les données en plusieurs catégories possibles, supérieures à deux. Ce type de classification est particulièrement utile lorsqu'il y a un grand volume de données à traiter.

La classification des images consiste à attribuer une ou plusieurs étiquettes à une image, qui est l'un des problèmes les plus fondamentaux de la vision par ordinateur et de la reconnaissance des formes, et a un large éventail d'applications, par exemple, la récupération d'images et de vidéos, surveillance vidéo, et la biométrie...

En général, le processus de classification d'image consiste à extraire des caractéristiques d'image puis à les classer. Les méthodes de classification traditionnelles utilisent des fonctionnalités de bas niveau ou de niveau intermédiaire pour représenter une image. Les fonctionnalités de bas niveau sont généralement basées sur des informations de densité, de couleur, de texture, de forme et de position en niveaux de gris, qui sont définies par l'homme (également appelées fonctionnalités fabriquées à la main, « hand-crafted features »).

Les fonctionnalités de niveau intermédiaire, ainsi que les fonctionnalités basées sur l'apprentissage, sont généralement distillées par des algorithmes de sac de mots visuels, qui sont efficaces et populaires dans le cadre de la classification ou de la récupération d'images au cours des dernières années. En vision par ordinateur, après l'extraction des caractéristiques, un classificateur (par exemple SVM, forêt aléatoire, etc.) est généralement utilisé pour attribuer l'étiquette aux différents types d'objets.

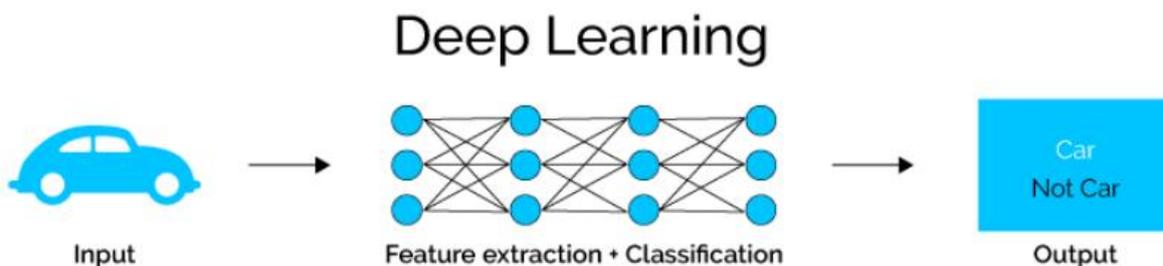


Figure 2- 1. Classification d'images par apprentissage profond [15]

Différente de la méthode de classification d'image traditionnelle, la méthode d'apprentissage profond combine le processus d'extraction et de classification des caractéristiques de l'image sur un même réseau. Le processus de classification en apprentissage profond est illustré à la figure 2.1.

La représentation des fonctionnalités de haut niveau du deep learning s'est avérée supérieure aux fonctionnalités de bas niveau et aux fonctionnalités de niveau intermédiaire réalisées à la main et a obtenu de bons résultats en termes de reconnaissance et de classification d'images. Ce concept est à la base du modèle d'apprentissage profond (réseau), qui est composé de nombreuses couches (telles que les couches convolutions et les couches entièrement connectées) qui transforme les données d'entrée (par exemple, les images) en sorties (par exemple, le résultat de la classification) tout en apprenant de plus en plus haut de niveau supérieur. [14]

b- La détection des objets

La détection d'objets est une tâche fondamentale en vision par ordinateur qui joue un rôle crucial dans diverses applications, telles que la conduite autonome, les systèmes de surveillance et la reconnaissance d'objets. Elle implique la localisation et la classification simultanées d'objets au sein d'une image. Le premier aspect de la détection d'objets est la localisation, qui fait référence à la détermination de l'étendue spatiale ou de la position des objets dans l'image. Ceci est réalisé en prédisant les coordonnées d'une boîte englobante qui enferme étroitement chaque objet. La boîte englobante fournit des informations sur l'emplacement et l'échelle de l'objet, permettant une analyse et une compréhension plus approfondies de son contexte dans l'image. Le deuxième aspect de la détection d'objets est la classification, qui consiste à attribuer une étiquette ou une classe à chaque objet identifié. L'objectif est de reconnaître et de catégoriser avec précision les objets dans des classes ou catégories prédéfinies. Ceci est généralement réalisé à l'aide de techniques d'apprentissage automatique, telles que l'apprentissage en profondeur, où les modèles sont formés sur de grands ensembles de données annotées pour apprendre des caractéristiques discriminantes et faire des prédictions précises.

La détection d'objets est une tâche difficile en raison de plusieurs facteurs. Une localisation précise nécessite une estimation précise de la position, de l'échelle et de l'orientation de l'objet, qui devient de plus en plus complexe en présence d'occlusions, d'arrière-plans encombrés et de variations dans l'apparence de l'objet. De plus, une classification précise est essentielle pour attribuer correctement l'objet à la catégorie appropriée, compte tenu de la présence potentielle d'objets similaires ou d'instances de la même classe avec des apparences différentes. Pour relever ces défis, les algorithmes de détection d'objets s'appuient sur des techniques avancées telles que les réseaux de neurones convolutifs (CNN), les méthodes de proposition de région et les mécanismes de fusion de caractéristiques. Ces approches permettent d'extraire des caractéristiques discriminantes de l'image et permettent la détection et la classification d'objets avec une grande précision. Le développement de

méthodes de détection d'objets robustes et efficaces a considérablement contribué aux progrès des applications de vision par ordinateur, permettant la reconnaissance, le suivi et la compréhension d'objets en temps réel. Les recherches en cours dans ce domaine visent à améliorer les performances des modèles de détection d'objets, à renforcer leur capacité à gérer divers scénarios et à relever des défis spécifiques, tels que la détection de petits objets ou la gestion des variations d'échelle. [16].

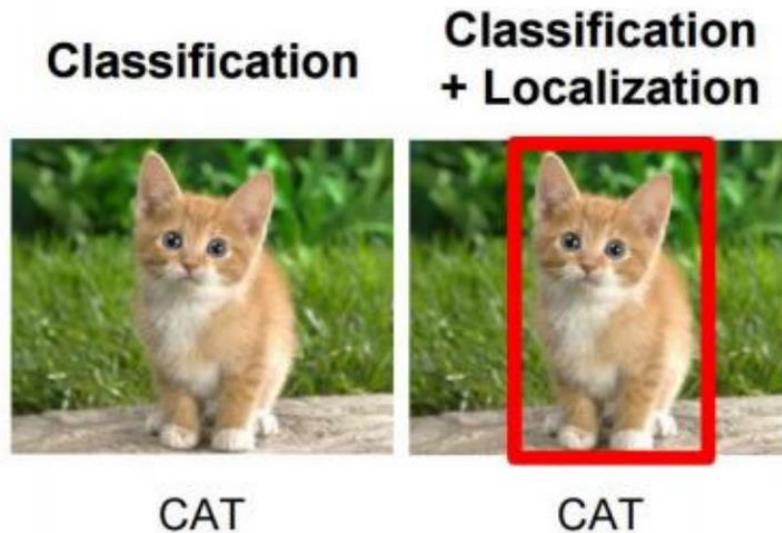


Figure 2- 2 La différence entre la classification [16]

3. Fonctionnement de Deep Learning

L'apprentissage profond est basé sur un réseau de neurones artificiels inspiré du cerveau humain. Ce réseau est composé de plusieurs « couches » de neurones, les données sont transmises à une couche d'entrée et sont ensuite transférées vers une ou plusieurs couches de traitement - appelées couches cachées - jusqu'à la couche de sortie. Chaque couche est composée de neurones interconnectés entre eux. Un neurone peut ou non être connecté à tous les neurones des couches adjacentes. Chaque connexion entre deux neurones est associée à un « poids ». Lorsque les données sont introduites dans le modèle, elles traversent chaque couche et sont transformées en représentations de plus en plus abstraites. Chaque couche utilise les informations de la couche précédente pour apprendre à détecter des caractéristiques spécifiques. Le modèle associera ensuite des caractéristiques à un objet spécifique en ajustant les poids des connexions neuronales associées à ces caractéristiques. [14]

Par exemple, si le modèle essaie de détecter une balle dans une image, les poids des neurones associés aux cercles seront favorisés. Cet ajustement est effectué à l'aide d'une méthode appelée « rétropropagation ». Il fonctionne en calculant l'erreur entre la sortie du modèle et l'entrée. Les poids sont ensuite modifiés pour minimiser cette erreur, en commençant par la couche de sortie et en revenant à la couche d'entrée. En fournissant suffisamment d'exemples au modèle, il peut ajuster les poids jusqu'à ce qu'il puisse détecter avec précision les objets souhaités. Ces exemples n'ont pas besoin

de dire au modèle où se trouve un objet. Ils ont juste besoin de lui dire qu'un objet est là, et il apprendra à le détecter par lui-même, tout comme un humain apprendrait à reconnaître un objet s'il est montré plusieurs fois.

4. Le réseau de neurones convolutionnel (CNN)

Il existe plusieurs types d'apprentissage profond, parmi eux le réseau de neurones convolutionnel (CNN). CNN est une amélioration des réseaux de neurones artificiels (ANN) traditionnel, qui comprend généralement des couches convolutions, des couches de pool et des couches entièrement connectées. Le CNN peut être divisé en deux parties :

- Une partie d'extraction d'entités (couches de convolution et couches de regroupement)
- Une partie de classification (couches entièrement connectées).

L'image est d'abord passée à travers une série de convolution, regroupant des couches pour l'extraction d'entités, puis passer à travers des couches entièrement connectées pour la classification.

a- Couche convolution (Convolution layer)

la couche convolution est un composant fondamental des réseaux CNN, jouant un rôle crucial dans l'extraction de caractéristiques à partir des images d'entrée. Son objectif principal est d'identifier et de capturer les caractéristiques pertinentes dans une image. Ceci est réalisé grâce au processus de filtrage convolutif. La couche convolutive fonctionne en faisant glisser une fenêtre de filtre sur l'image, balayant systématiquement chaque partie de l'image. A chaque position, le filtre et la région correspondante de l'image sont multipliés élément par élément, et les valeurs résultantes sont additionnées pour obtenir une seule valeur de sortie. Cette opération de convolution est effectuée pour chaque position dans l'image, résultant en une carte de caractéristiques de sortie. Les filtres utilisés dans la couche convolutive sont conçus pour capturer des caractéristiques ou des motifs d'intérêt spécifiques dans l'image.

Chaque filtre met en évidence un aspect différent, comme les bords, les textures ou les formes. En appliquant divers filtres, la couche convolutive peut détecter efficacement un large éventail de caractéristiques dans l'image d'entrée. Pour chaque paire d'image d'entrée et de filtre, la couche convolutive produit une carte d'activation de caractéristiques (comme la figure 2-3). Cette carte indique les emplacements dans l'image où la caractéristique correspondante est détectée. Les valeurs d'activation dans la carte représentent le degré de similitude entre le filtre et la région de l'image. Des valeurs d'activation plus élevées indiquent des correspondances plus fortes entre les positions du filtre et de l'image qui possèdent la caractéristique souhaitée.

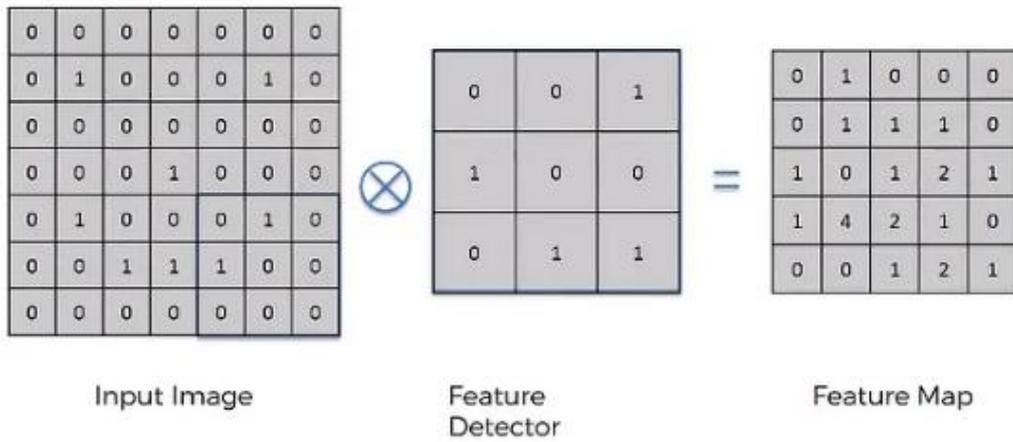


Figure 2- 3. Parcours de filtre sur une image[17]

b- **la couche de pooling (pooling layer)** : elle reçoit plusieurs « feature map » en entrée et applique une opération de pooling (*subsampling*) à chaque feature map. Elle permet la réduction de la taille de l'image en tenant compte de ses caractéristiques importantes. Le principe est de couper l'image en cellules régulières, puis nous conservons la valeur maximale dans chaque cellule par rapport au filtre utilisé (2 x 2 pixels, 3x3...). Donc des petites cellules carrées sont souvent utilisées afin de ne pas perdre trop d'informations. Nous obtenons le même nombre de « features map » que l'entrée, mais elles sont beaucoup plus petites. Cela permet d'accélérer non seulement les calculs, mais d'éviter également le problème du sur-ajustement.

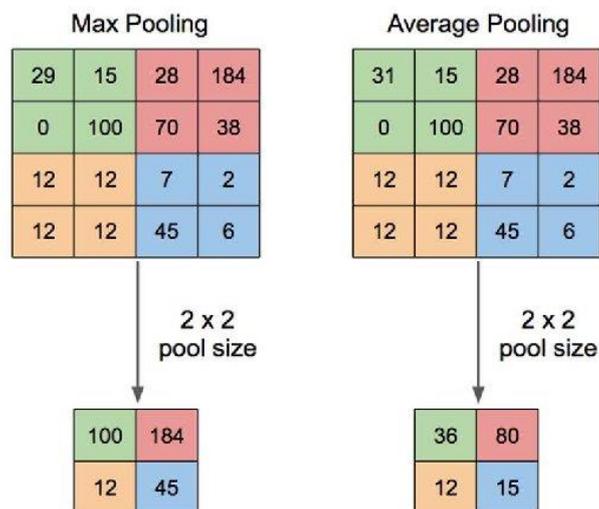


Figure 2- 4. Max/Average Pooling avec un filtre 2x2 et un pas de 2[18]

Il existe trois principaux types de pooling :

- **Le pooling maximal** : est basé sur la détection de la valeur maximale dans la région sélectionnée (voir la figure 2-4) ;
- **Le pooling minimal** : sur la détection de la valeur minimale dans la région sélectionnée

- **Le pooling moyenne (average pooling)** : basé sur la détection de la valeur moyenne dans la région sélectionnée (voir la figure 2-4).
- c- **La couche de correction ReLU** : (Rectified Linear Units) représente la fonction réelle non linéaire qui permet de remplacer toutes les valeurs d'entrée négatives par des zéros (Figure 2-5). Il remplit la fonction d'activation. $\text{ReLU}(x) = \max(0, x)$.

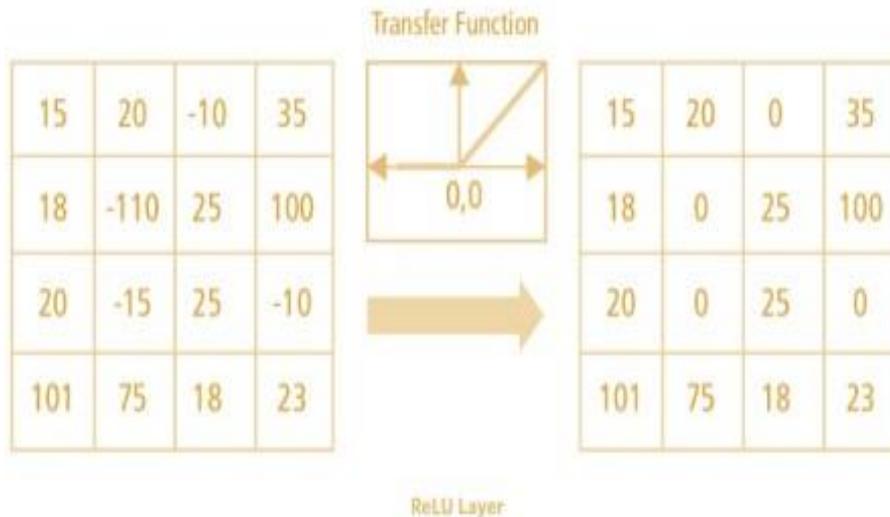


Figure 2- 5. Couche de correction Relu [19].

- d- **La couche entièrement connectée (fully-connected)** : Elle permet de classifier l'image en entrée dans une étiquette. Elle reçoit un vecteur en entrée (représente le résultat d'aplatir de notre matrice) et produit un nouveau vecteur en sortie. Chaque élément du vecteur représentant une probabilité qu'une certaine caractéristique appartienne à une étiquette.

5. Transfer Learning : (l'apprentissage par transfert)

L'apprentissage par transfert est une stratégie dans laquelle un modèle de réseau de neurones est initialement formé sur un problème similaire avant d'être appliqué au problème actuel. En tirant parti des connaissances acquises lors de la formation initiale, une ou plusieurs couches du modèle pré-entraîné sont utilisées dans un nouveau modèle spécialement formé pour le problème cible. Cette approche a démontré son efficacité, en particulier dans des scénarios avec des données d'entraînement limitées.

Les CNN utilisés pour la classification des images se composent de deux parties : les couches de convolution, qui sont utilisées pour l'extraction des caractéristiques, et les couches de connexion complète, qui sont utilisées pour la classification. Par conséquent, nous pouvons utiliser directement le réseau de neurone pré-entraîné pour extraire des caractéristiques de l'image, puis le vecteur de caractéristique extrait est utilisé comme entrée pour former une nouvelle couche entièrement connectée afin de résoudre d'autres problèmes de classification. [20] (Figure 2-6).

La couche de sortie du modèle d'origine doit généralement être remplacée car elle n'est probablement pas du tout utile pour la nouvelle tâche, et elle peut même ne pas avoir le bon nombre de sorties pour la nouvelle tâche.

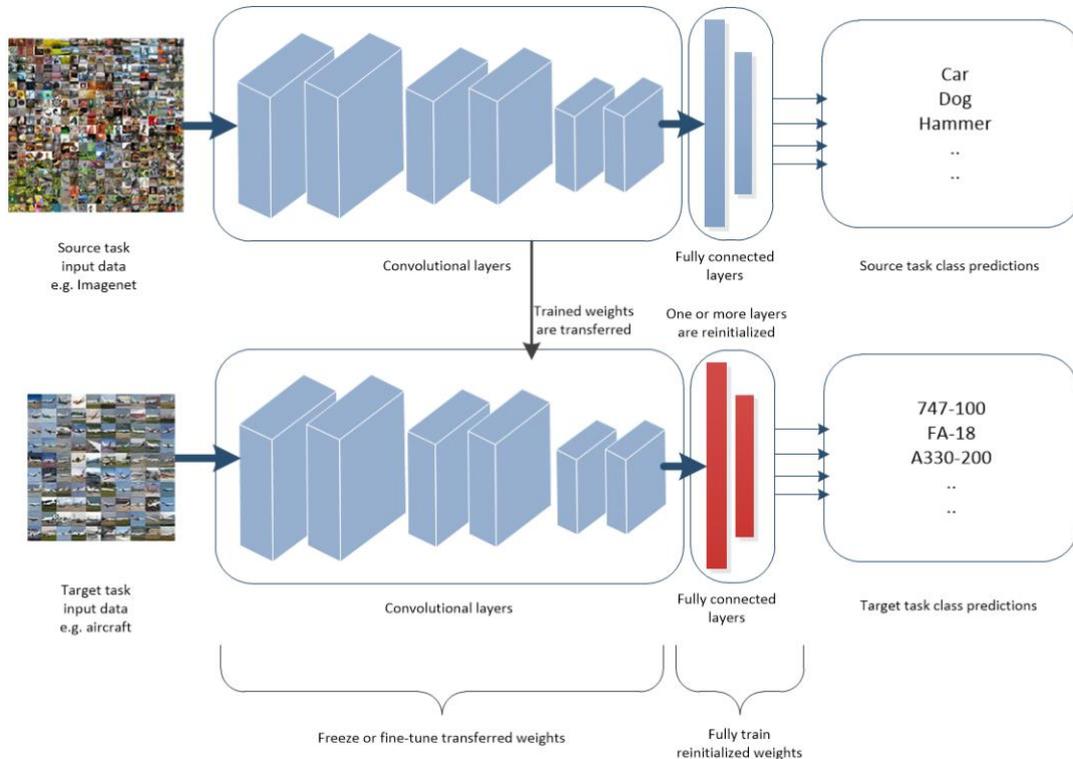


Figure 2- 6. Illustration du concept de transfer learning [20].

Le composant souhaité du modèle ou du modèle pré-entraîné peut être immédiatement inclus dans un nouveau modèle de réseau neuronal. Dans cette application, les poids du modèle précédemment entraîné peuvent être gelés pour éviter qu'ils ne soient modifiés pendant l'entraînement du nouveau modèle, empêchant ainsi toute modification due à la descente de gradient. Ensuite, on peut essayer de dégeler une ou deux des couches cachées supérieures afin que la rétropropagation puisse les modifier. Cela peut être fait pour voir si les performances s'améliorent. de plus en plus de couches peuvent être dégelées si les données de formation deviennent de plus en plus abondantes. Comme alternative, les poids peuvent être initialisés à l'aide du modèle pré-entraîné lors de l'apprentissage du nouveau modèle et mis à jour ultérieurement, peut-être avec un taux d'apprentissage réduit [20].

6. Les modèles de classification basés sur CNN

Les architectures CNN empilent quelques couches convolutionnelles (chacune généralement suivie d'une couche ReLU), puis une couche de regroupement, puis quelques autres couches convolutionnelles (+ ReLU), puis une autre couche de regroupement, etc. L'image devient de plus en plus petite au fur et à mesure qu'elle progresse à travers le réseau, mais elle devient généralement de plus en plus profonde (c'est-à-dire avec plus de cartes de caractéristiques) grâce aux couches de

convolutions. Au sommet de la pile, un réseau de neurones feedforward régulier est ajouté, composé de quelques couches entièrement connectées (+ ReLU), et la couche finale génère la prédiction (par exemple, une couche softmax qui génère des probabilités de classe estimées). Au fil du temps, des variantes de cette architecture fondamentale ont été développées, conduisant à des avancées étonnantes dans le domaine.

Avant de détailler quelques architectures de CNN, nous présenterons d'abord l'un des ensembles de données les plus importants dans la classification d'images ImageNet [21]. Le jeu de données ImageNet est un jeu de données de référence largement utilisé dans le domaine de la vision par ordinateur. Il s'agit de l'un des ensembles de données les plus vastes et les plus complets disponibles pour les tâches de reconnaissance et de classification d'objets. L'ensemble de données contient des millions d'images étiquetées couvrant un large éventail de catégories d'objets.

ImageNet a été créé dans le cadre du projet ImageNet, qui visait à développer une base de données visuelle à grande échelle pour la formation et l'évaluation des algorithmes de vision par ordinateur. L'ensemble de données est organisé hiérarchiquement selon la hiérarchie WordNet, chaque image étant associée à un synset spécifique (un ensemble de mots synonymes représentant un concept ou une catégorie d'objet). L'ensemble de données a d'abord gagné en popularité grâce au ImageNet Large Scale Visual Recognition Challenge (ILSVRC), un concours annuel où les participants ont développé des modèles pour classer et détecter des objets dans l'ensemble de données. L'ILSVRC a joué un rôle important dans l'avancement du domaine de l'apprentissage en profondeur, en particulier avec l'introduction des réseaux de neurones convolutifs (CNN).

L'ensemble de données ImageNet couvre un large éventail de catégories d'objets, y compris les animaux, les plantes, les véhicules, les articles ménagers et bien d'autres. Il contient des images diverses et complexes capturées à partir de diverses sources, telles qu'Internet et des photographes professionnels, garantissant son applicabilité à des scénarios réels. En raison de sa taille et de sa diversité, le jeu de données ImageNet est devenu une référence standard pour évaluer les performances des modèles de vision par ordinateur. De nombreux modèles de pointe ont été formés sur l'ensemble de données ImageNet, ce qui a conduit à des avancées significatives dans la reconnaissance d'objets, la classification d'images et les tâches connexes. La disponibilité de l'ensemble de données ImageNet a permis aux chercheurs et aux praticiens de développer et d'évaluer des modèles avec une base solide dans la compréhension visuelle générale. Son impact sur le développement et les progrès des algorithmes de vision par ordinateur ne peut être surestimé, et il continue d'être une ressource précieuse pour faire progresser le domaine.

- **AlexNet** : L'architecture AlexNet CNN [22] a remporté le défi ImageNet ILSVRC 2012 par une large marge : elle a atteint 17% de taux d'erreur dans le top 5 tandis que la deuxième meilleure n'a atteint que 26%! Il a été développé par Alex Krizhevsky (d'où son nom), Ilya Sutskever et Geoffrey Hinton. Il a été le premier à empiler des couches convolutionnelles directement les unes sur les autres, au lieu d'empiler une couche de regroupement au-dessus de chaque couche convolutionnelles.

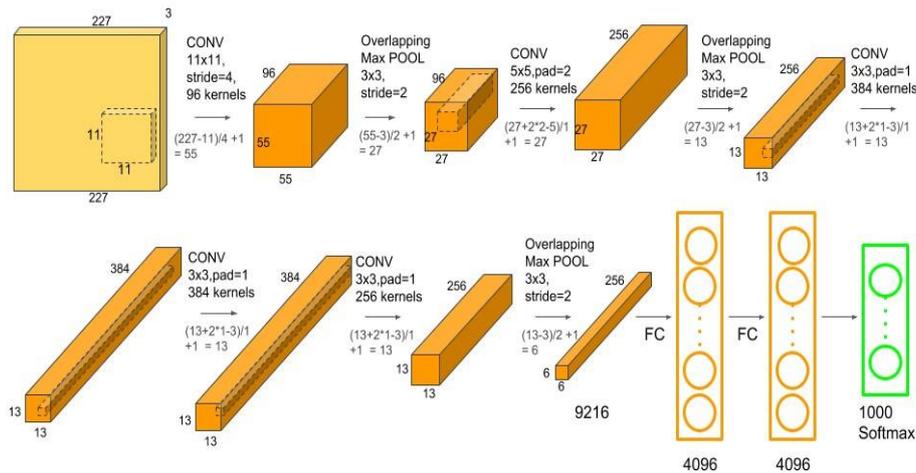


Figure 2- 7. L'architecture du réseau AlexNet [23]

- **GoogLeNet** : L'architecture GoogLeNet a été développée par Christian Szegedy et al de Google Research [24] et il a remporté le défi ILSVRC 2014 en poussant le taux d'erreur du top 5 en dessous de 7%. Cette excellente performance provient en grande partie du fait que le réseau était beaucoup plus profond que les CNN précédents. Cela a été rendu possible par des sous-réseaux appelés modules de création, qui permet à GoogLeNet d'utiliser les paramètres beaucoup plus efficacement que les architectures précédentes : GoogLeNet a en fait 10 fois moins de paramètres qu'AlexNet (environ 6 millions au lieu de 60 millions).
- **VGGNet** : Le finaliste du défi ILSVRC 2014 était VGGNet [25] développé par K. Simonyan et A. Zisserman. Il avait une architecture très simple et classique, avec 2 ou 3 couches convolutionnelles, une couche de regroupement, puis à nouveau 2 ou 3 couches convolutionnelles, une couche de regroupement, etc. (avec un total de seulement 16 couches convolutionnelles), plus une densité finale réseau avec 2 couches cachées et la couche de sortie. Il n'utilisait que des filtres 3×3 , mais de nombreux filtres.

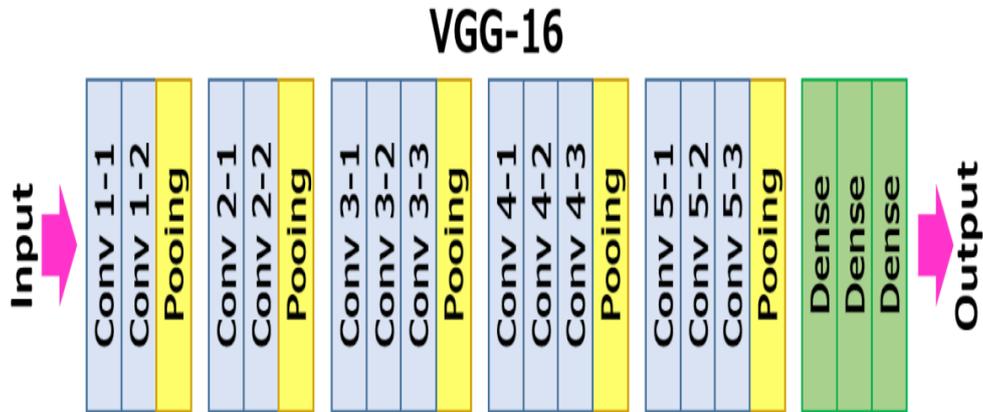


Figure 2- 8.L'architecture de VGGNet-16 [24]

- ResNet** : Le défi ILSVRC 2015 a été remporté en utilisant un réseau résiduel (ou ResNet), développé par Kaiming He et al [26] qui a réalisé un taux d'erreur top-5 incroyable sous 3,6%, en utilisant un CNN extrêmement profond composé de 152 couches. Cela confirme la tendance générale : les modèles deviennent de plus en plus profonds, avec de moins en moins de paramètres. La clé pour pouvoir former un réseau aussi profond est d'utiliser des connexions de saut (également appelées connexions de raccourci) : le signal entrant dans une couche est également ajouté à la sortie d'une couche située un peu plus haut dans la pile. Voyons pourquoi cela est utile. Lors de l'entraînement d'un réseau de neurone, l'objectif est de lui faire modéliser une fonction cible $h(x)$. Si vous ajoutez l'entrée x à la sortie du réseau (c'est-à-dire que vous ajoutez une connexion de saut), le réseau sera forcé de modéliser $f(x) = h(x) - x$ plutôt que $h(x)$. C'est ce qu'on appelle l'apprentissage résiduel (voir la figure 2-8).

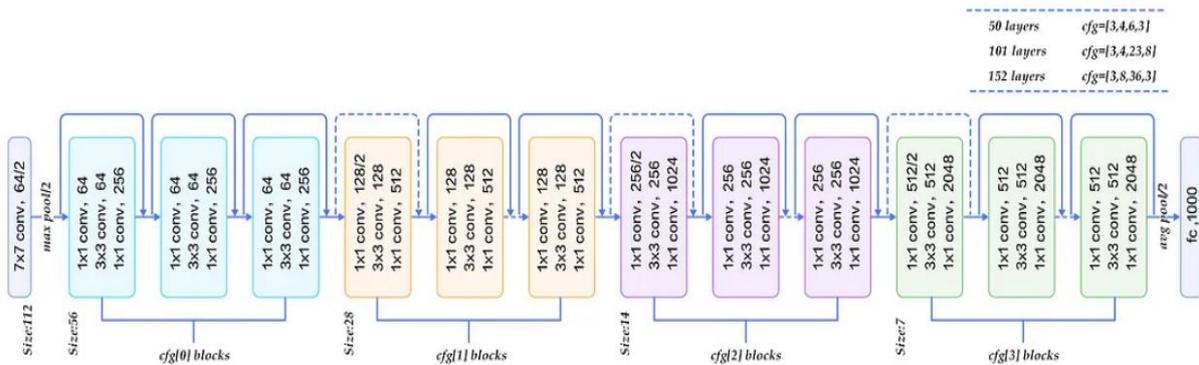


Figure 2- 9. L'apprentissage résiduel [27]

7. Les modèles de détections des objets basée sur CNN

□ Le détecteur à deux étages

Implique généralement une approche en deux étapes. Dans la première étape, un réseau de proposition de région (RPN) génère des propositions d'objets potentiels en balayant

l'image entière et en identifiant les régions d'intérêt. Ces propositions servent de boîtes englobantes candidates pour les objets. Dans la deuxième étape, les régions sélectionnées sont introduites dans un réseau de classificateurs, qui classe chaque région proposée et affine les coordonnées de la boîte englobante. Cette étape consiste à affiner le réseau pour classer et localiser avec précision les objets dans les régions proposées.

L'approche en deux étapes permet une détection efficace des objets en réduisant le nombre de régions à traiter. Le RPN agit comme un pré-filtre, identifiant les régions prometteuses pour une analyse plus approfondie, ce qui aide à concentrer les ressources de calcul sur les zones pertinentes. Cette approche s'est avérée efficace pour obtenir une précision et une vitesse élevées dans les tâches de détection d'objets, ce qui la rend largement adoptée dans diverses applications telles que la conduite autonome, la surveillance et l'analyse d'images.[28]. Exemple : Faster R-CNN.

- **Réseau convolutif régional plus rapide (Faster R-CNN)**

Faster R-CNN est un modèle de détection d'objets populaire qui s'appuie sur l'approche en deux étapes de la proposition et de la classification des régions. Il améliore les méthodes précédentes en introduisant un réseau de proposition de région (RPN) qui partage des couches convolutionnelles avec le réseau de détection d'objets ultérieur. Dans Faster R-CNN, la première étape implique le RPN, qui analyse l'image d'entrée et génère un ensemble de propositions de région. Ces propositions sont des emplacements potentiels de boîte englobante qui pourraient contenir des objets d'intérêt. Le RPN y parvient en faisant glisser une petite fenêtre (appelée ancre) sur l'image et en prédisant la probabilité de présence d'objet et en ajustant les coordonnées de la boîte englobante.

Dans la deuxième étape, les régions proposées sont introduites dans un réseau de classificateurs, généralement un Fast R-CNN, qui classe les objets et affine davantage les coordonnées de la boîte englobante. Le réseau de classification prend les régions proposées en entrée et extrait les caractéristiques de ces régions à l'aide de la mise en commun RoI ou d'une technique similaire. Ces caractéristiques sont ensuite utilisées pour la classification et la régression de la boîte englobante. (voir figure 2-10).

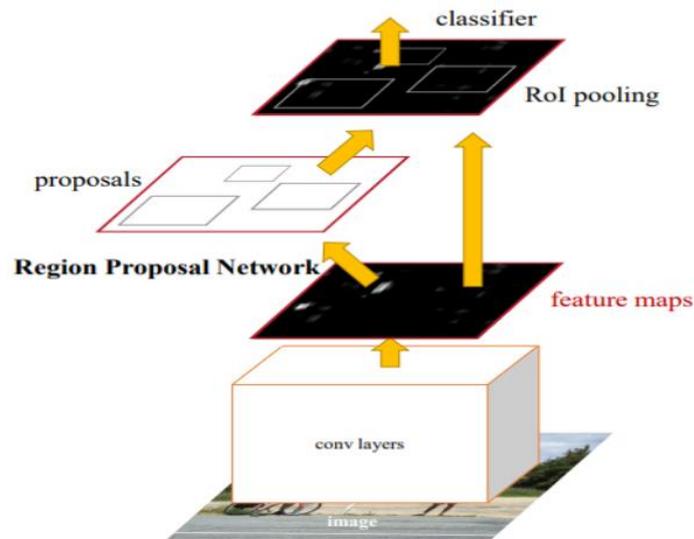


Figure 2- 10. Architecture de model Faster R-CNN [29]

□ Le détecteur à un étage

Contrairement aux approches en deux étapes, les détecteurs d'objets en une étape prédisent directement les coordonnées de la boîte englobante et les probabilités de classe en un seul passage sur l'image d'entrée. Ces modèles divisent l'image en une grille de cellules ou de boîtes d'ancrage et font des prédictions pour chaque cellule ou boîte. Ils utilisent généralement un ensemble de boîtes d'ancrage prédéfinies avec différentes échelles et rapports d'aspect pour capturer des objets de différentes tailles.

Les détecteurs à un étage utilisent des réseaux de neurones convolutionnels (CNN) pour extraire les caractéristiques de l'image d'entrée. Ces caractéristiques sont ensuite traitées par des couches convolutionnelles supplémentaires pour prédire les probabilités de classe et les coordonnées de la boîte englobante pour chaque boîte d'ancrage ou cellule de grille. Les prédictions sont affinées à l'aide de techniques telles que la suppression non maximale pour éliminer les détections en double et améliorer la précision de la localisation.

L'avantage des détecteurs à un étage est leur simplicité et leur efficacité, car ils éliminent le besoin d'une étape de proposition de région explicite. Cela conduit à des temps d'inférence plus rapides, ce qui les rend adaptés aux applications en temps réel. Cependant, ils peuvent avoir du mal à détecter de petits objets et à gérer des scènes denses en raison de la conception grossière de la grille ou de la boîte d'ancrage.[28]. Exemple : Yolo model.

- **You Only Look Once (YOLO)**

YOLO (You Only Look Once) [30] est un algorithme populaire de détection d'objets en temps réel qui a révolutionné le domaine de la vision par ordinateur. Il a introduit une approche innovante de la détection d'objets en combinant la localisation et la classification d'objets dans un seul modèle unifié. L'idée principale derrière YOLO est de diviser l'image d'entrée en une grille et de prédire les boîtes englobantes et les probabilités de classe directement dans chaque cellule de la grille. Cette approche de cellule de grille permet à YOLO de faire des prédictions pour plusieurs objets simultanément, ce qui se traduit par une vitesse d'inférence impressionnante.

YOLO utilise une architecture de réseau neuronal à convolution profonde (CNN), généralement basée sur l'architecture Darknet, pour extraire les caractéristiques de l'image d'entrée. Le réseau divise l'image en une grille $S \times S$, où chaque cellule de grille prédit B boîtes englobantes et leurs probabilités de classe correspondantes. Ces boîtes englobantes sont paramétrées par leurs coordonnées par rapport à la cellule de la grille, ainsi que les prédictions de largeur et de hauteur. Les probabilités de classe représentent les scores de confiance pour chaque classe d'objets.

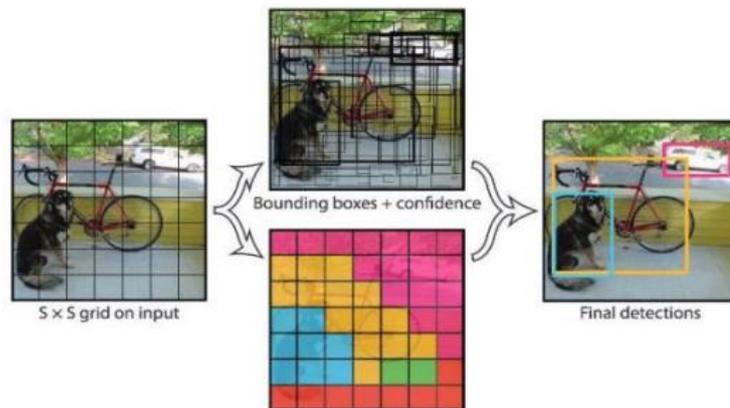


Figure 2- 11. Modèle YOLO [30]

Pendant la formation, YOLO utilise un ensemble prédéfini de boîtes d'ancrage avec différents rapports d'aspect pour correspondre aux objets de vérité terrain. Le modèle est entraîné à l'aide d'une combinaison de perte de localisation (mesure de la précision des prédictions de la boîte englobante) et de perte de classification (mesure de la précision des prédictions de classe). La fonction de perte est optimisée à l'aide de techniques telles que la rétropropagation et la descente de gradient. L'un des principaux avantages de YOLO est sa vitesse remarquable, permettant la détection d'objets en temps réel sur des appareils à ressources limitées. Cependant, l'approche en un seul passage de YOLO peut entraîner des difficultés à détecter avec précision les petits objets, et elle peut avoir des difficultés avec

des objets d'échelles et de rapports d'aspect variables dans la même cellule de grille. Au fil des ans, plusieurs versions de YOLO ont été développées, notamment YOLOv2, YOLOv3 et YOLOv4, chacune introduisant des améliorations en termes de précision et de performances. Ces versions intègrent des techniques telles que le regroupement de boîtes d'ancrage, des réseaux pyramidaux et des modifications architecturales avancées pour améliorer la précision de la détection et répondre aux limites du YOLO d'origine.

Conclusion

Dans ce chapitre on a expliqué la notion du Deep Learning et les réseaux de neurones convolutionnels et ces différents modèles. On a parlé de l'apprentissage par transfert et la classification des images, ainsi que quelques modèles CNN utilisés pour la classification et la détection des objets. Dans le chapitre suivant, on va présenter la partie expérimentale de notre travail : l'architecture et les paramètres choisis ainsi que la discussion des différents résultats obtenus.

CHAPITRE 03 : Conception et implémentation du modèle CNN de classification

Introduction :

Nous allons décrire dans ce chapitre, l'approche proposée qui consiste à élaborer des algorithmes basés Deep Learning pour la classification des images de panneaux routières en 43 classes. Pour cela nous avons proposé deux algorithmes, le premier consiste à proposer un réseau profond de type CNN élaboré frome Scratch. Quant à la deuxième solution, elle consiste à faire recours à un type particulier de réseau profond basé sur le concept de transfert Learning. Cette technique utilise une variante du réseau profond VGG celui de VGG16. Ce chapitre décrit aussi la base de données utilisée pour entrainer, tester et valider les modèles profonds utilisés dans ce travail.

1. Classification des panneaux routières dans la littérature :

Les méthodes de classification et de reconnaissance des panneaux de signalisation peuvent être divisées en méthodes basées sur l'apprentissage automatique et en méthodes basées sur l'apprentissage profond.

Méthodes d'apprentissage automatique

Le concept d'un modèle d'apprentissage automatique consiste à appliquer tout d'abord des techniques manuelles telles que l'histogramme du gradient orienté (HOG) et le modèle binaire localisé (LBP) afin d'extraire les couleurs ou les caractéristiques de forme d'une image d'entrée, puis d'utiliser des classificateurs pour la classification des images.

Liu et al [31] ont proposé une méthode de détection des panneaux d'interdiction de circuler basée sur HOG et SVM (Support Vector Machine). Le prétraitement de l'image, l'extraction des caractéristiques HOG et la classification de l'image basée sur le modèle SVM sont les trois étapes de la méthode proposée. Le grisage de l'image couleur, l'extraction des caractéristiques HOG, l'ajustement de la taille de l'image du panneau de signalisation et la correction Gamma de l'image grise sont les techniques de prétraitement de l'image appliquées à l'image d'entrée. La classification des panneaux de signalisation d'interdiction comprend l'apprentissage d'un classificateur binaire unique et la reconnaissance des panneaux de signalisation basée sur le classificateur binaire SVM.

Li et al [32] ont proposé une reconnaissance des panneaux de signalisation basée sur les caractéristiques locales et les réseaux neuronaux ELM (Extreme Learning Machine). La méthode combine le réseau neuronal ELM et une variété de caractéristiques locales. Les

caractéristiques des panneaux sont extraites à l'aide des caractéristiques HOG et LBP, puis introduites dans le réseau neuronal ELM à des fins d'apprentissage. Après le processus de formation, ces réseaux neuronaux formés sont utilisés pour la reconnaissance des panneaux de signalisation. Namyang et al [33] ont proposé un système de classification et de reconnaissance des panneaux de signalisation thaïlandais basé sur les descripteurs HOG et Color Layout et sur le coefficient de corrélation normalisé. La précision de la méthode proposée était de 93,98 %. La forêt aléatoire et le SVM sont deux classificateurs utilisés dans le processus de classification.

Sugiharto et al [34] ont proposé une détection des panneaux de signalisation basée sur HOG et PHOG à l'aide de SVM binaires et de K-NN. Les trois étapes de la méthode proposée sont la conversion de l'espace colorimétrique RVB en HSI (teinte, saturation, intensité), le redimensionnement de l'image suivi de l'extraction des caractéristiques à l'aide de HOG et PHOG, puis la classification. Rahmad et al [35] ont proposé la détection et la reconnaissance des panneaux de signalisation indonésiens en utilisant l'extraction des caractéristiques de texture et de couleur et le classificateur SVM. Coctovanu et al [36] ont proposé une détection des panneaux de signalisation basée sur le SVM utilisant les caractéristiques HOG. La méthode proposée comprend l'acquisition d'images, le prétraitement HSV (Hue Saturation Value), la segmentation des couleurs et la classification SVM. Zhang et al [37] ont proposé la détection et la reconnaissance des panneaux de signalisation dans une scène complexe. La détection des panneaux de signalisation utilise le MSER (Maximally Stable Extremal Region) et la reconnaissance des panneaux de signalisation utilise le modèle Lenet-5.

Zheng et al. [47] ont présenté une méthode de détection de fenêtre glissante, qui recherche les panneaux de signalisation à différentes échelles avec le classificateur de caractéristiques de canal intégré. Ciblant les panneaux d'interdiction et d'obligation dans les repères de signalisation routière allemands (GTSDDB), Gim et al. [48] ont développé un système contenant deux modules de filtres grossiers : le premier module est basé sur le HOG et l'analyse discrète linéaire (LDA) ; la seconde repose sur une petite fenêtre coulissante ; les deux modules impliquent une grande fenêtre et un classificateur SVM.

Bien que ces efforts soient efficaces pour reconnaître les panneaux de signalisation basés sur des méthodes graphiques, ces solutions ne fonctionnent pas bien dans des scénarios complexes (par exemple, faible luminosité, panneaux partiellement obscurcis, etc.) et sont particulièrement inefficaces pour reconnaître les panneaux avec des orientations ou des points de vue différents.

Méthodes d'apprentissage profond

Récemment, les méthodes d'apprentissage profond ont été largement utilisées dans le domaine de la classification des panneaux de signalisation. Par rapport aux méthodes basées sur l'apprentissage automatique, les caractéristiques peuvent être extraites automatiquement et une grande précision peut être atteinte.

Serna et al [38] ont proposé la reconnaissance et la classification de panneaux de signalisation à la fois textuels et symboliques. Le système proposé se compose de trois modules : détection des panneaux de signalisation (Mask RCNN), affinement (amélioration de la localisation et filtrage) et classification. Le processus d'affinement permet d'extraire les régions d'intérêt ROI à partir des sorties de masque précédentes. Dans la classification, les CNN sont utilisés sur la base de cinq blocs pour identifier une classe particulière, et trois blocs sont utilisés pour identifier une catégorie de panneaux routiers. Les catégories Danger, Interdiction, Spécial, Priorité, Obligatoire, Direction, Panneaux supplémentaires et Autres sont huit catégories de panneaux. L'architecture Class CNN a atteint une précision de 99,61 % sur les ensembles de données GTSRB.

Loukmane et al. [39] ont proposé une nouvelle architecture de réseau basée sur le VGG au lieu d'utiliser le CNN pour la classification des images. L'architecture se compose de 3 couches maxpool, de 6 couches convolutives, d'une couche entièrement connectée et d'une couche softmax. L'algorithme CLASHE (Contrast Limited Adaptive Histogram Equalization) et la normalisation sont utilisés pour l'amélioration de l'image. Des transformations géométriques et de retournement sont effectuées sur les images des panneaux de signalisation pour résoudre certains problèmes tels que les occlusions, les changements d'éclairage et les conditions météorologiques déséquilibrées. La matrice de confusion, la précision et le rappel sont les mesures utilisées ici. La précision globale de la méthode proposée est de 99,33 %.

Vincent et al [40] ont proposé un modèle CNN pour la classification des panneaux de signalisation à l'aide de l'ensemble de données GTSRB. La rotation, la mise en miroir et la mise à l'échelle sont des techniques d'augmentation des données qui augmentent le nombre d'images dans un ensemble de données. Le modèle utilisé dans cet article est le modèle Keras séquentiel. Le système proposé a utilisé un ensemble de données GTSRB composé d'images de panneaux de signalisation. Ce modèle a atteint une précision supérieure de 98,44 % dans l'ensemble de données de test.

Sanjeevani et al [41] ont proposé une reconnaissance et une détection automatiques des panneaux de limitation de vitesse. Il contient deux classes. La première comprend une segmentation pixel par pixel pour toutes les limitations de vitesse et la seconde est un nouveau classificateur avec une classe de fausses alarmes. Le système proposé développe une nouvelle technologie qui peut améliorer la précision du programme AusRAP (Australian Road Assessment Program) et réduire les erreurs de classification des panneaux de limitation de vitesse.

Sun et al [42] ont proposé une reconnaissance des panneaux de signalisation basée sur le CNN. Il se concentre principalement sur la détection et la reconnaissance des panneaux de signalisation circulaires. Sarku et al [43] ont proposé une approche de classification des panneaux de signalisation basé sur l'apprentissage par transfert pour économiser les ressources de développement comme le temps, l'effort et l'argent. L'apprentissage par transfert basé sur ResNet a été identifié comme une solution efficace pour le processus de reconnaissance des panneaux de signalisation.

Shao et al. [46] ont proposé une méthode de reconnaissance des panneaux de signalisation pour un travail complet sur le réseau de trafic : le CNN basé sur la région (R-CNN) a été étendu par la méthode de proposition d'objet d'EdgeBox.

2. La solution proposée

Dans ce travail, et afin de classifier les images en 43 classes nous avons suivi l'approche décrite dans la figure ci-dessous. L'approche utilise deux réseaux profonds, le premier est un CNN from scratch, le deuxième est basé sur le transfert learning en utilisant VGG16

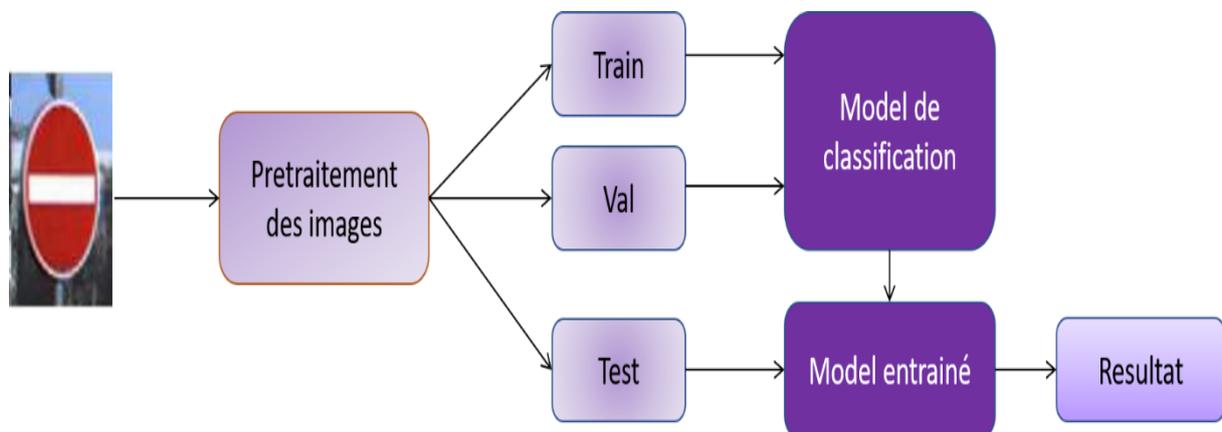


Figure 3- 1. Solution proposé

A- Dataset

L'ensemble de données utilisé dans cette étude est le GTSRB (German Traffic Sign Recognition Benchmark), qui est fréquemment utilisé par les universitaires travaillant sur la détection et la catégorisation des panneaux de signalisation [44]. Plus de 50 000 photos de panneaux de signalisation allemands sont incluses dans l'ensemble de données de référence GTSRB, qui provient d'Allemagne. Ces images sont divisées en 43 classifications. Ces images ont été prises pendant la journée en Allemagne à partir de dix heures de séquences vidéo de divers types d'autoroutes.

Un total de 51 839 images des panneaux compose l'ensemble de données, qui est ensuite divisé en deux sous-ensembles : un ensemble d'apprentissage de 39 209 photos et un ensemble de test distinct de 12 630 images, qui représentent ensemble 43 classes de signes différentes.



Figure 3- 2. Échantillons d'images de dataset

La liste suivante représente les différentes classes et le code correspond de chacun des panneaux de routiers utilisés :

```

classes = {0:'Speed limit (20km/h)',
          1:'Speed limit (30km/h)',
          2:'Speed limit (50km/h)',
          3:'Speed limit (60km/h)',
          4:'Speed limit (70km/h)',
          5:'Speed limit (80km/h)',
          6:'End of speed limit (80km/h)',
          7:'Speed limit (100km/h)',
          8:'Speed limit (120km/h)',
          9:'No passing',
          10:'No passing veh over 3.5 tons',
          11:'Right-of-way at intersection',
          12:'Priority Road',
          13:'Yield',
          14:'Stop',
          15:'No vehicles',
          16:'Veh > 3.5 tons prohibited',
          17:'No entry',
          18:'General caution',
          19:'Dangerous curve left',
          20:'Dangerous curve right',
          21:'Double curve',
          22:'Bumpy Road',
          23:'Slippery Road',
          24:'Road narrows on the right',
          25:'Road work',
          26:'Traffic signals',
          27:'Pedestrians',
          28:'Children crossing',
          29:'Bicycles crossing',
          30:'Beware of ice/snow',
          31:'Wild animals crossing',
          32:'End speed + passing limits',
          33:'Turn right ahead',
          34:'Turn left ahead',
          35:'Ahead only',
          36:'Go straight or right',
          37:'Go straight or left',
          38:'Keep right',
          39:'Keep left',
          40:'Roundabout mandatory',
          41:'End of no passing',
          42:'End no passing veh > 3.5 tons'}

```

Figure 3- 3. Liste des classes des panneaux utilisés.

B- Prétraitement de dataset

Nous avons redimensionné les images de l'apprentissage en taille de 30*30 pixels, et faire une augmentation de données pour enrichir notre base de données et améliorer le résultat. Autrement dit, effectués des transformations aléatoires sur les images d'entrée. Cela permet d'empêcher le réseau de mémoriser la position et l'orientation exactes des objets, ou encore en générant des échantillons artificiels pour une ou plusieurs classes du jeu de données en rotation, zoom, mise en miroir, flou sur l'ensemble original d'images.

L'augmentation de données est faite grâce à la classe prédéfinie de Keras « ImageDataGenerator » qui permet d'appliquer des transformations géométriques, des effets de

couleur, de luminosité, de résolution... etc. Il suffit de les parcourir et de choisir les plus pertinentes pour l'étude. Dans notre travail nous avons appliqué les transformations comme suis :

- *Fill_mode="nearest"*
- *Rotation_range=10*
- *Width_shift_range=0.1*
- *Shear_range=0.15,*
- *Zoom_range=0.15*
- *Height_shift_range=0.1*

C- Model de classification

Stratégie 1 : model CNN from scratch

Les réseaux profonds de type CNN est largement utilisé et recommande pour le problème de classification, pour cette raison nous avons proposé un réseau convolutionnel (CNN) comporte 14 couches : Une couche d'entrée, 4 couches de convolution, 2 couches de *MaxPooling*, 3 couches de normalisation par lot (*batch normalization*), une couche *Flatten* et 1 couches de sortie Fully-Connected. (Figure 3.4)

a) Les couches de convolution :

Notre architecture comprend 5 couches de convolution laquelle :

- **La première couche : (conv2D)** est destiné à l'extraction d'identité (feature extraction).il effectue une convolution avec 16 filtres de taille 3x3 pour filtrer une image d'entrée de taille 32*32*3. Cette couche est effectuée avec une stride de (1,1) avec un *padding same*, la fonction d'activation utilisée est RELU.
- **La deuxième couche :** occupe la deuxième position dans le modèle. Il présente de nombreuses similitudes avec la première couche. Cette couche prend en entrée la sortie de la première couche convolution avec 32 filtres de taille 3x3.
- **La troisième couche** occupe la cinquième position du modèle avec 64 filtres de taille 3x3 avec l'application d'une fonction d'activation RELU.
- **La quatrième couche** occupe la septième position est contient les mêmes paramètres que la couche de convolution précédente sauf le nombre de filtre qui est 128.

b) Les couches pooling

Notre architecture se compose de 2 couches de pooling (maxPooling) pour réduire la dimension spatiale (feature map) des couches convolutionnelles

Toutes les deux couches de pooling ont la même taille de filtre qui est 2x2, ils occupent les positions : troisième et septième couche.

c) La couche batch normalisation

Notre architecture comprend 3 couches de normalisation par lot utilisé pour entraîner efficacement le modèle en accélérant l'apprentissage, en améliorant la stabilité et en réduisant le surapprentissage.

d) Les couches fully connected (dense)

Notre modèle se termine par une couche *flatten* et 2 couches entièrement connectées (FC) avec respectivement 512 et 43 unités. En choisissant ce nombre des unités pour minimiser le nombre des paramètres et accélérer le calcul (le taux de calcul). La première couche fully connected utilise une fonction d'activation RELU suivi par une couche dropout pour réduire le problème d'*overfitting*.

La couche de sortie constitue de 43 unités qui ressemble à nos classes de sortie (catégories des panneaux) avec une fonction Softmax.

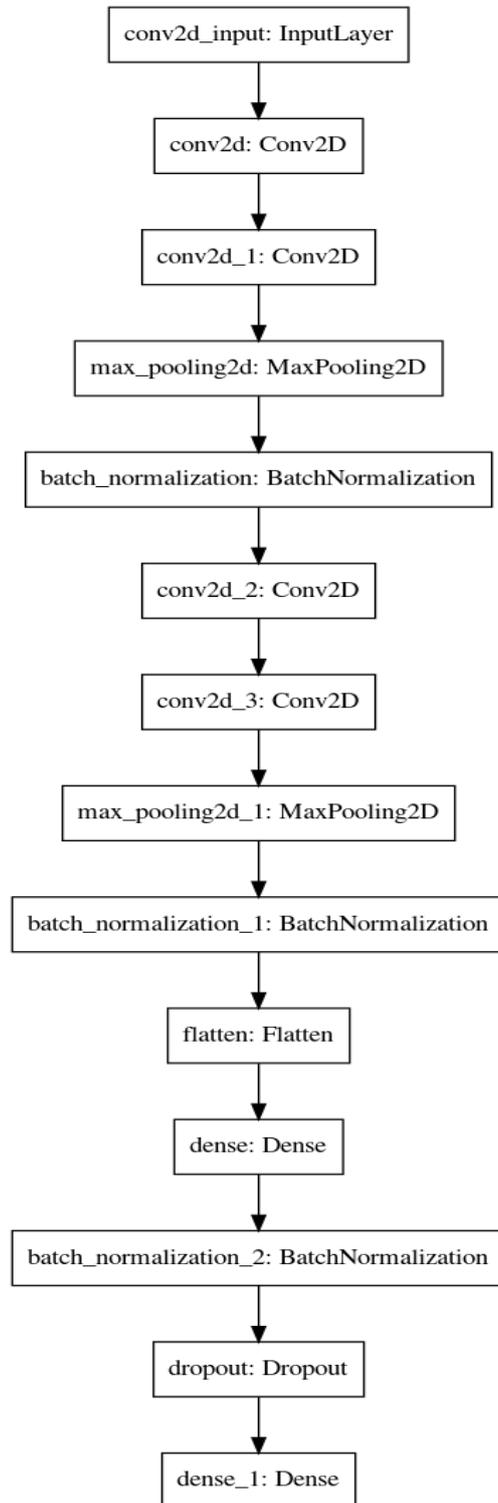


Figure 3- 4. Architecture du model CNN proposé

Stratégie 2 : Transfer learning

Dans cette méthode, nous avons utilisé une approche d'entraînement basé sur le Transfer learning, où on utilise un model existant pré-entraîné sur une large dataset.

Pour cela nous avons choisi le modèle VGG pré-entraîné sur les données d'ImageNet. Ce modèle comporte 16 couches avec un nombre de paramètres de 138 millions paramètres. (Figure 3.4)

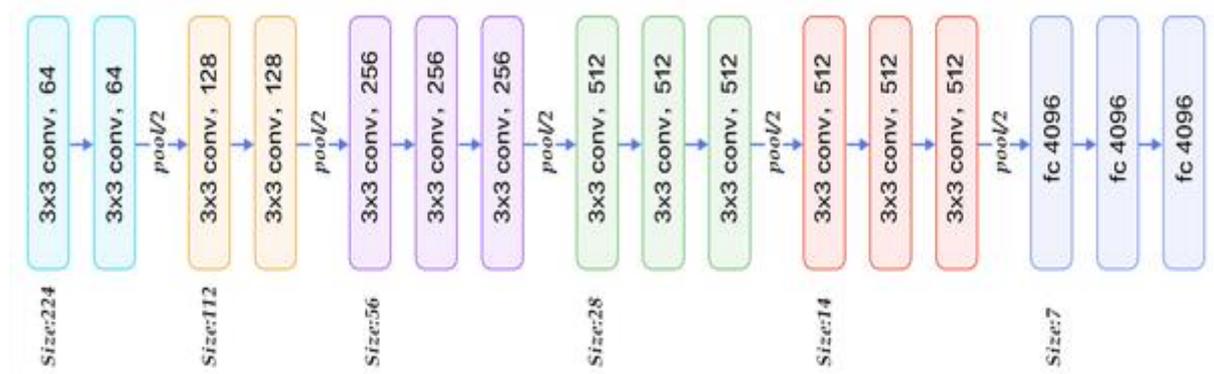


Figure 3- 5. Architecture du modèle VGG16[45]

Pour déterminer si l'utilisation de l'apprentissage par transfert rend le réseau pré-entraîné plus efficaces pour nos données, nous effectuons deux différents types d'expériences :

- **Méthode 1** : Entraîner une partie du modèle pré-entraîné, consiste à Fixer les poids de toutes les couches du modèle pré-entraîné sur ImageNet, et entraîner juste les couches entièrement connectées sur notre dataset (dernière couche).
- **Méthode 2** : Entraîner toutes les couches du modèle pré-entraîné.

3. Environnement de travail

Dans notre étude, nous avons utilisé deux bibliothèques fameuses pour entraîner notre modèle de classification :

- **TensorFlow**

TensorFlow est un framework d'apprentissage automatique open source développé par Google. Il est largement utilisé pour créer et déployer divers modèles d'apprentissage automatique, y compris des modèles d'apprentissage en profondeur. TensorFlow fournit un ensemble complet d'outils, de bibliothèques et de ressources qui permettent aux développeurs et aux chercheurs de créer et de former efficacement des modèles d'apprentissage automatique. Il offre une architecture flexible qui permet un déploiement facile sur différentes plateformes, telles que les ordinateurs de bureau, les serveurs, les appareils mobiles et même les systèmes distribués. TensorFlow prend en charge un large éventail de tâches, notamment la reconnaissance d'images et de la parole, le traitement du langage naturel et l'analyse de séries chronologiques. Il est connu pour son

évolutivité, permettant aux utilisateurs de former des modèles sur de grands ensembles de données à l'aide de ressources informatiques distribuées. TensorFlow fournit également des API de haut niveau pour créer rapidement des modèles, ainsi que des API de niveau inférieur pour plus de personnalisation et de contrôle. Dans l'ensemble, TensorFlow est devenu l'un des frameworks les plus populaires et les plus largement adoptés dans le domaine de l'apprentissage automatique.

- **Keras**

Keras est une bibliothèque de réseaux de neurones open source de haut niveau écrite en Python. Il est conçu pour fournir une interface conviviale et intuitive pour la création et la formation de modèles d'apprentissage en profondeur. Il prend en charge divers moteurs backend, tels que TensorFlow et Theano, et offre une large gamme de couches prédéfinies, de fonctions d'activation, de fonctions de perte et d'algorithmes d'optimisation. Il fournit également des utilitaires pour le prétraitement des données, l'évaluation des modèles et la visualisation. Keras a acquis une grande popularité dans la communauté de l'apprentissage en profondeur et est largement utilisé dans les universités et l'industrie. Sa flexibilité, ses performances et son support communautaire étendu en ont fait un choix privilégié pour développer et déployer des modèles d'apprentissage en profondeur pour un large éventail d'applications, notamment la vision par ordinateur, le traitement du langage naturel et l'apprentissage par renforcement.

4. Expérimentation et résultats

Pour entraîner notre modèle nous avons joué les paramètres suivants :

- **Batch size** : nombre d'échantillons ou de points de données qui sont traités ensemble en un seul passage avant et arrière pendant la formation. Au lieu de mettre à jour les paramètres du modèle après chaque échantillon individuel, il est plus efficace de les mettre à jour après le traitement d'un lot d'échantillons.
- **Epoch** : détermine le nombre de fois que le modèle itérera sur l'ensemble du jeu de données.

Pour évaluer notre modèle nous avons basé sur les métriques suivantes : accuracy et matrice de confusion

- **Accuracy** : quantifie la justesse globale des prédictions du modèle, $Acc = (\text{Nombre d'instances correctement classées}) / (\text{Nombre total d'instances})$

- **Matrice de confusion :** La matrice de confusion est une matrice $K \times K$ où chaque élément (i, j) représente le nombre d'instances appartenant à la classe i mais classées dans la classe j . Elle fournit une ventilation détaillée des prédictions du modèle pour chaque classe, montrant les vrais positifs, les vrais négatifs, les faux positifs et les faux négatifs. Supposons que nous ayons un problème de classification avec K classes.
 - True Positive (TP) : nombre d'instances correctement classées dans la classe i .
 - Vrai négatif (TN) : le nombre d'instances correctement classées comme n'appartenant pas à la classe i (correctement classées dans d'autres classes que i).
 - Faux positif (FP) : le nombre d'instances incorrectement classées dans la classe i (prédites comme appartenant à la classe i mais appartenant en fait à d'autres classes).
 - Faux négatif (FN) : le nombre d'instances incorrectement classées comme n'appartenant pas à la classe i (prédites comme n'appartenant pas à la classe i , mais appartenant en fait à la classe i).

Alors nous avons lancé l'apprentissage du modèle par plusieurs valeurs du batch-size et différent nombre d'époch. Voici les résultats obtenus après la modification des valeurs de ces options pour chaque méthode utilisée :

Méthode 1: CNN from scratch

Tableau 1. Résultat d'évaluation du modèle CNN pour différentes valeurs de batch size.

batch_size	16	32	64
accuracy	98.56%	99.00%	98.36%

Le meilleur résultat est obtenu dans le cas du $batch-size= 32$ et $epochs = 30$. Où nous avons obtenus un taux d'accuracy égale à 99%. Les figures 3-5 et 3-6 montrent les graphes de l'apprentissage de modèle par cette méthode et leur précision ainsi que la matrice de confusion obtenu.

Méthode 2 : Transfer Learning

Nous avons fixé les poids de toutes les couches du modèle pré-entraîné, et entrainer juste la partie de classifieur qui comporte les couches entièrement connectées. Nous avons lancé l'entraînement du modèle avec différentes valeurs de batch-size. Le tableau suivant montre les résultats obtenus pour les différentes valeurs de batch-size.

Tableau 2. Résultat d'évaluation du méthode 1 de transfer learning pour différentes valeurs de batch size.

batch_size	16	32	64
accuracy	54.54%	53.76%	55.47%

On peut remarquer qu'on a obtenu un taux de précision de 55.47 % avec batch-size= 64 pour 30 epochs. Voici le graphe de précision et perte ainsi que la matrice de confusion de ce modèle :

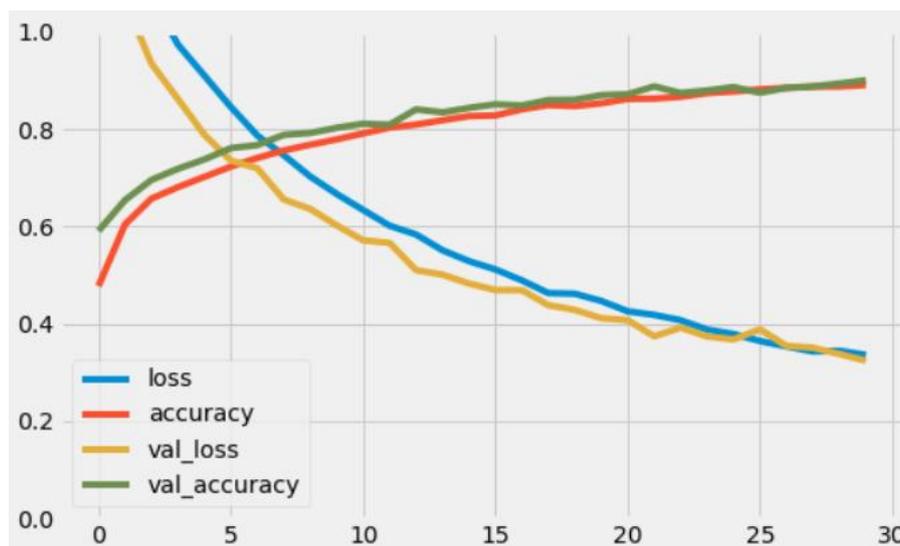


Figure 3- 8. Graphe de précision et de perte du méthode 2.

Les résultats obtenus à l'aide de l'apprentissage par transfert avec un modèle pré-entraîné fixe, où seules les couches entièrement connectées ont été entraînées sur l'ensemble de données spécifique, ont donné une précision modérée. Avec des tailles de lot de 16, 32 et 64, les précisions obtenues étaient respectivement de 54,54 %, 53,76 % et 55,47 %. Ces résultats suggèrent que l'application directe des caractéristiques apprises par le modèle pré-entraîné sur l'ensemble de données ImageNet peut ne pas être optimale pour notre problème de classification. Les précisions relativement faibles obtenues indiquent que les caractéristiques du modèle pré-entraîné peuvent ne pas être directement applicables aux nuances et aux exigences spécifiques du problème de classification à 43 classes.

Tableau 3. Résultat d'évaluation du méthode 2 de transfer learning pour différentes valeurs de batch size.

batch_size	16	32	64
accuracy	96.87%	98.25%	95.30%

Le meilleur résultat est obtenu dans le cas du *batch-size*= 32. Où nous avons obtenus un taux de précision égale à 98.25%.

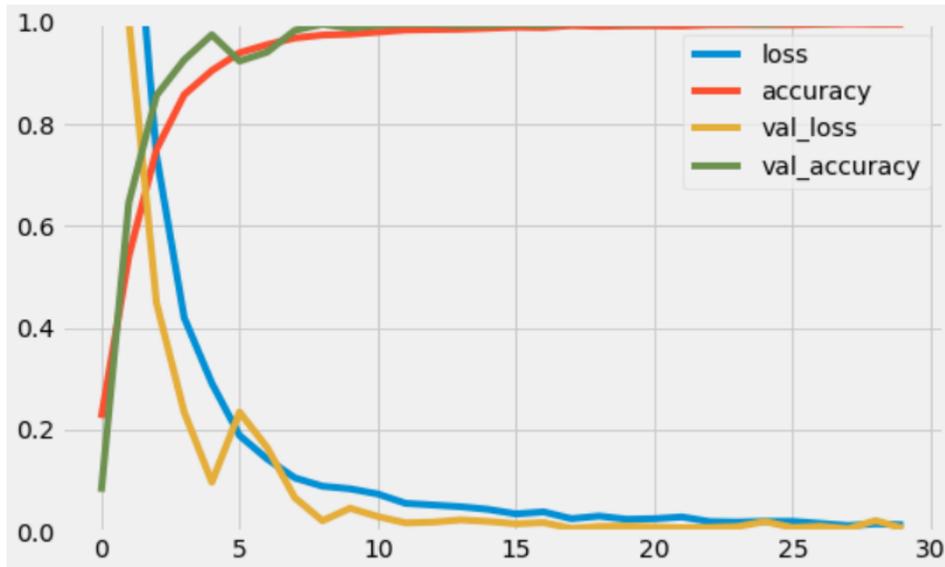


Figure 3- 10. Graphe de précision et de perte du méthode 2.

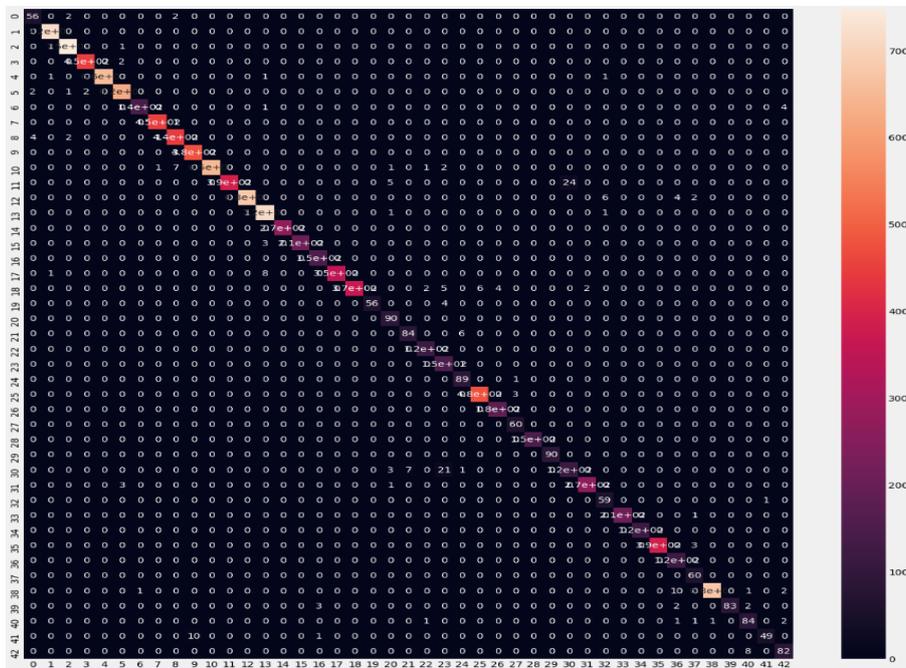


Figure 3- 11. Matrice de confusion du méthode 3.

Nous pouvons résumer les meilleurs résultats de chaque méthode dans le tableau suivant

Tableau 4. Meilleur résultat de chaque méthode

		Batch size	Accuracy
CNN from scratch		32	99.00%
Transfer learning	VGG16 (Méthode 1)	64	53.76%
	VGG16 (Méthode 2)	32	98.25%

D'après ce tableau on peut voir que le meilleur taux de précision est obtenu par notre modèle CNN proposé de la méthode 1 (from scratch) qui a donné un taux de précision de 99% avec 30 epochs et batch_size de 32.

Le tableau 5 montre une comparaison de notre modèle avec différentes méthodes d'apprentissage profond et machine learning utilisées pour la reconnaissance et la classification des panneaux de signalisation qui utilisent la même dataset en fonction de accuracy. Accuracy est un indicateur général du fonctionnement d'un modèle dans toutes les classes. Il est calculé comme le rapport du nombre de prédictions précises au nombre total de prédictions.

Tableau 5. Comparaison de résultats.

	Dataset	Accuracy
Sugiharto et al [34]	GTSDDB	82.01%
Li et al [32]	GTSDDB	96.82%
Liu et al [31]	GTSDDB	90.20%
Serna et al [38]	GTSDDB	99%
Loukmane et al [39]	GTSDDB	99.30%
Vincent et al [40]	GTSDDB	98.44%
Sun et al [42]	GTSDDB	98.20%
Notre modèle	GTSDDB	99%

Il est important de noter que malgré la précision légèrement inférieure par rapport à certains des autres modèles [39][38], notre modèle a atteint une précision compétitive de 99 % sur

l'ensemble de données GTSDDB. Ce qui rend cette réalisation remarquable est le fait que notre modèle a moins de paramètres par rapport aux autres modèles mentionnés.

Avoir moins de paramètres suggère que notre modèle est plus efficace en termes de ressources de calcul et d'utilisation de la mémoire. Cela peut être avantageux dans les scénarios où des contraintes de calcul sont présentes, telles que le déploiement sur des appareils à ressources limitées ou des applications en temps réel. Malgré le nombre réduit de paramètres, notre modèle atteint toujours un haut niveau de précision, ce qui en fait un choix attrayant dans les applications pratiques où l'efficacité est une priorité.

Dans l'ensemble, l'analyse comparative indique que notre modèle fonctionne exceptionnellement bien sur l'ensemble de données GTSDDB tout en présentant une configuration de paramètres plus efficace par rapport aux autres modèles évalués. Cela suggère que "notre modèle" trouve un équilibre favorable entre précision et efficacité de calcul, ce qui en fait une solution prometteuse pour les tâches de détection des panneaux de signalisation.

Conclusion

Dans ce chapitre, nous avons décrit la démarche à suivre afin d'obtenir le classifieur le plus précis possible pour classifier les images des panneaux routiers. Nous avons utilisé deux différentes méthodes, la première consiste à utiliser un réseau CNN from scratch, et dans la deuxième et la troisième méthode, nous avons utilisé l'apprentissage par transfert sur le modèle pré-entraîné VGG16. Après plusieurs tests, nous avons obtenus un taux de précision de 99% par le modèle CNN proposé. Ce travail reste une première contribution qui nécessite encore des améliorations.

CONCLUSION GENERALE

En tant qu'élément important de la conduite intelligente, la détection et la reconnaissance visuelle des panneaux routiers attirent de plus en plus l'attention. À l'heure actuelle, dans des conditions de faible luminosité, par temps de brouillard, par temps de pluie et dans d'autres situations où les appareils photo numériques sont bloqués ou flous, les vidéos ou images collectées deviendront floues, les erreurs de reconnaissance sont très susceptibles d'être générées. Par conséquent, dans ces environnements difficiles, la précision de reconnaissance du TSR est très faible. Par exemple, la reconnaissance précise des panneaux routiers dans un environnement brumeux est affectée par la visibilité des images claires, sa précision de reconnaissance doit encore être améliorée.

C'est dans ce cadre que s'inscrit notre travail, qui a pour objectif de classifier des images des panneaux routiers afin de détecter le type de panneau parmi 43 types de panneaux routiers. Pour réaliser notre travail nous avons utilisé deux différentes méthodes, la première consiste à utiliser un réseau CNN from scratch, et dans la deuxième méthode, nous avons utilisé l'apprentissage par transfert sur le modèle pré-entraîné VGG16. Après plusieurs tests, nous avons obtenu un taux de précision de 99% par notre modèle CNN de la première méthode. Ces avancées dans la reconnaissance informatisée des panneaux de signalisation sont extrêmement importantes pour la sécurité routière, car elles contribuent à la prévention des accidents de la circulation et à la protection des vies et des biens.

Notre travail futur comporte trois volets. Tout d'abord, nous collecterons des ensembles de données dans diverses conditions d'éclairage telles que les jours de brouillard et de pluie. Dans un second temps, nous comparons les méthodes de reconnaissance et de détection d'objets dans TSR telles que YoLo et Faster RCNN. Enfin, nous tirerons parti de nouvelles méthodes d'augmentation de données pour augmenter l'ensemble de données, ce qui rendra notre modèle plus robuste et plus solide.

BIBLIOGRAPHIE

- [1] M. Lu, K. Wevers, and R. Van Der Heijden, “Technical Feasibility of Advanced Driver Assistance Systems (ADAS) for Road Traffic Safety,” *Transp. Plan. Technol.*, vol. 28, no. 3, pp. 167–187, Jun. 2005, doi: 10.1080/03081060500120282.
- [2] V. K. Kukkala, J. Tunnell, S. Pasricha, and T. Bradley, “Advanced Driver-Assistance Systems: A Path Toward Autonomous Vehicles,” *IEEE Consum. Electron. Mag.*, vol. 7, no. 5, pp. 18–25, Sep. 2018, doi: 10.1109/MCE.2018.2828440.
- [3] F. G. A. Cartel and N. L. Ngwa, “CONCEPTION D’UN SYSTÈME D’AIDE A LA CONDUITE POUR VEHICULE DE TOURISME (ANTICOLLISION)”.
- [4] R. Kreuzig, M. Ochs, and R. Mester, “DistanceNet: Estimating Traveled Distance from Monocular Images using a Recurrent Convolutional Neural Network.” arXiv, Apr. 17, 2019. doi: 10.48550/arXiv.1904.08105.
- [5] M. A. Haseeb, D. Ristic-Durrant, A. Graeser, M. Banic, and S. Dušan, “Multi-DisNet: Machine Learning-Based Object Distance Estimation from Multiple Cameras,” 2019, pp. 457–469. doi: 10.1007/978-3-030-34995-0_41.
- [6] “Thermal Infrared vs. Active Infrared,” *IRINFO.org*. <https://irinfo.org/03-01-2007-grossman/> (accessed May 19, 2023).
- [7] D. Lamine, “Un système de détection des objets de la circulation routière et d’estimation de leur distance.”
- [8] J. L. M. Narvaez, “Adaptation of a Deep Learning Algorithm for Traffic Sign Detection”.
- [9] K. Djihene, “Image et Vie Artificielle.”
- [10] M. Boumediene, J.-P. Lauffenburger, J. Daniel, and C. Cudel, “Détection, association et suivi de pistes pour la reconnaissance de panneaux routiers”.
- [11] Y. SAADNA, “Identification des panneaux routiers pour des systèmes d’aide à la conduite.”
- [12] J. Xing, “Traffic Sign Recognition From Digital Images Using Deep Learning”.
- [13] J. Yan, L. Zhao, W. Diao, H. Wang, and X. Sun, “AF-EMS Detector: Improve the Multi-Scale Detection Performance of the Anchor-Free Detector,” *Remote Sens.*, vol. 13, no. 2, Art. no. 2, Jan. 2021, doi: 10.3390/rs13020160.
- [14] E. Sperling, “Deep Learning Spreads,” *Semiconductor Engineering*, Jan. 31, 2018. <https://semiengineering.com/deep-learning-spreads/> (accessed May 19, 2023).
- [15] Z. He, “Deep Learning in Image Classification: A Survey Report,” in *2020 2nd International Conference on Information Technology and Computer Application (ITCA)*, Dec. 2020, pp. 174–177. doi: 10.1109/ITCA52113.2020.00043.
- [16] A. Ouaknine, “Review of Deep Learning Algorithms for Object Detection,” *Zyl Story*, Feb. 05, 2018. <https://medium.com/zylapp/review-of-deep-learning-algorithms-for-object-detection-c1f3d437b852> (accessed May 20, 2023).
- [17] S. Girgin, “Day-40 Deep Learning-5 (CNN-1),” *PursuitData*, Sep. 19, 2019. <https://medium.com/pursuitnotes/day-40-deep-learning-5-cnn-1-5f664998ea86> (accessed May 19, 2023).
- [18] M. Yani, S. Irawan, and C. Setianingsih, “Application of Transfer Learning Using Convolutional Neural Network Method for Early Detection of Terry’s Nail,” *J. Phys. Conf. Ser.*, vol. 1201, p. 012052, May 2019, doi: 10.1088/1742-6596/1201/1/012052.
- [19] Prabhu, “Understanding of Convolutional Neural Network (CNN) — Deep Learning,” *Medium*, Nov. 21, 2019. <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148> (accessed May 19, 2023).
- [20] J. Plested and T. Gedeon, “Deep transfer learning for image classification: a survey.” arXiv, May 19, 2022. Accessed: May 19, 2023. [Online]. Available: <http://arxiv.org/abs/2205.09904>
- [21] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL: IEEE, Jun. 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.
- [22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
- [23] A. Pujara, “Concept of AlexNet:- Convolutional Neural Network,” *Analytics Vidhya*, Jun. 09, 2021. <https://medium.com/analytics-vidhya/concept-of-alexnet-convolutional-neural-network-6e73b4f9ee30> (accessed May 19, 2023).
- [24] C. Szegedy *et al.*, “Going Deeper with Convolutions.” arXiv, Sep. 16, 2014. doi: 10.48550/arXiv.1409.4842.
- [25] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition.” arXiv, Apr. 10, 2015. doi: 10.48550/arXiv.1409.1556.

- [26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition." arXiv, Dec. 10, 2015. doi: 10.48550/arXiv.1512.03385.
- [27] A. Rastogi, "ResNet50," *Medium*, Mar. 14, 2022. <https://blog.devgenius.io/resnet50-6b42934db431> (accessed May 19, 2023).
- [28] J. L. M. Narvaez, "Adaptation of a Deep Learning Algorithm for Traffic Sign Detection".
- [29] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." arXiv, Jan. 06, 2016. Accessed: May 20, 2023. [Online]. Available: <http://arxiv.org/abs/1506.01497>
- [30] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.
- [31] Y. Liu, W. Zhong, W. Wang, Q. Cao, and K. Luo, "A Method for Recognizing Prohibition Traffic Sign Based on HOG-SVM," in *2021 International Conference on Computer Engineering and Application (ICCEA)*, Jun. 2021, pp. 486–490. doi: 10.1109/ICCEA53728.2021.00101.
- [32] W. Li, Q. Chen, T. Dong, L. Wei, and Q. Zhang, "Traffic Signs Classification Based on Local Characteristics and ELM," in *2017 10th International Symposium on Computational Intelligence and Design (ISCID)*, Dec. 2017, pp. 127–130. doi: 10.1109/ISCID.2017.164.
- [33] N. Namyang and S. Phimoltares, "Thai Traffic Sign Classification and Recognition System Based on Histogram of Gradients, Color Layout Descriptor, and Normalized Correlation Coefficient," in *2020 - 5th International Conference on Information Technology (InCIT)*, Oct. 2020, pp. 270–275. doi: 10.1109/InCIT50588.2020.9310778.
- [34] A. Sugiharto and A. Harjoko, "Traffic sign detection based on HOG and PHOG using binary SVM and k-NN," in *2016 3rd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, Oct. 2016, pp. 317–321. doi: 10.1109/ICITACEE.2016.7892463.
- [35] C. Rahmad, I. F. Rahmah, R. A. Asmara, and S. Adhisuwigno, "Indonesian traffic sign detection and recognition using color and texture feature extraction and SVM classifier," in *2018 International Conference on Information and Communications Technology (ICOIACT)*, Mar. 2018, pp. 50–55. doi: 10.1109/ICOIACT.2018.8350804.
- [36] D. Coțovanu, C. Zet, C. Foșalău, and M. Skoczylas, "Detection of Traffic Signs Based on Support Vector Machine Classification Using HOG Features," in *2018 International Conference and Exposition on Electrical And Power Engineering (EPE)*, Oct. 2018, pp. 0518–0522. doi: 10.1109/ICEPE.2018.8559784.
- [37] L. Zhang, S. Zhang, and M. Zhang, "Detection and recognition on traffic sign in complex scene," in *2019 6th International Conference on Information Science and Control Engineering (ICISCE)*, Dec. 2019, pp. 254–258. doi: 10.1109/ICISCE48695.2019.00058.
- [38] C. Gámez Serna and Y. Ruichek, "Traffic Signs Detection and Classification for European Urban Environments," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 10, pp. 4388–4399, Oct. 2020, doi: 10.1109/TITS.2019.2941081.
- [39] A. Loukmane, M. Graña, and M. Mestari, "A Model for Classification of Traffic Signs Using Improved Convolutional Neural Network and Image Enhancement," in *2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS)*, Oct. 2020, pp. 1–8. doi: 10.1109/ICDS50568.2020.9268761.
- [40] M. A. Vincent, V. K. R, and S. P. Mathew, "Traffic Sign Classification Using Deep Neural Network," in *2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, Dec. 2020, pp. 13–17. doi: 10.1109/RAICS51191.2020.9332474.
- [41] P. Sanjeevani, B. Verma, and J. Affum, "A Novel Evolving Classifier with a False Alarm Class for Speed Limit Sign Recognition," in *2021 IEEE Congress on Evolutionary Computation (CEC)*, Jun. 2021, pp. 2211–2217. doi: 10.1109/CEC45853.2021.9504710.
- [42] M. T. Islam, "Traffic sign detection and recognition based on convolutional neural networks," in *2019 International Conference on Advances in Computing, Communication and Control (ICAC3)*, Dec. 2019, pp. 1–6. doi: 10.1109/ICAC347590.2019.9036784.
- [43] E. Sarku, J. Steele, T. Ruffin, B. Gokaraju, and A. Karimodini, "Reducing Data Costs- Transfer Learning Based Traffic Sign Classification Approach," in *SoutheastCon 2021*, Mar. 2021, pp. 1–5. doi: 10.1109/SoutheastCon45413.2021.9401900.
- [44] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition," *Neural Netw.*, vol. 32, pp. 323–332, Aug. 2012, doi: 10.1016/j.neunet.2012.02.016.
- [45] "VGG16 architecture," *OpenGenus IQ: Computing Expertise & Legacy*, Jan. 21, 2019. <https://iq.opengenus.org/vgg16/> (accessed May 20, 2023).

- [46] F. Shao, X. Wang, F. Meng, T. Rui, D. Wang, and J. Tang, “Real-Time Traffic Sign Detection and Recognition Method Based on Simplified Gabor Wavelets and CNNs,” *Sensors*, vol. 18, no. 10, Art. no. 10, Oct. 2018, doi: 10.3390/s18103192
- [47] T. Y. Zheng, N. Wang, and J. Tang, “Traffic signs detection based on ICF and multi-class classifier,” *Transducer Microsyst. Technol.*, Tech. Rep., 2017
- [48] J. Gim, M. Hwang, B. C. Ko, and J.-Y. Nam, “Real-Time Speed-Limit Sign Detection and Recognition Using Spatial Pyramid Feature and Boosted Random Forest,” in *Image Analysis and Recognition*, M. Kamel and A. Campilho, Eds., in *Lecture Notes in Computer Science*, vol. 9164. Cham: Springer International Publishing, 2015, pp. 437–445. doi: 10.1007/978-3-319-20801-5_48.