

République algérienne démocratique et populaire.
Ministère de L'enseignement Supérieure de la recherche scientifique.
Université 8 Mai 45 –Guelma-
Faculté des Mathématiques, d'informatique et des Sciences de la Matière
Département d'Informatique



Mémoire de Fin d'études Master

Filière : Informatique

Option : Science et technologie de l'information et de la communication.

Thème :

L'augmentation des données pour les systèmes de reconnaissances des expressions faciales

Encadré par :
Mme.Bordjiba Yamina

Présenté par :
Mehammedia saber abd errahmen

Juin 2022

Remerciements

Tout d'abord, je remercie Dieu Tout-Puissant de m'avoir donné la force et la patience pour accomplir ce travail.

Je tiens à remercier profondément Madame Yamina Bordjiba, mon encadreur, pour sa disponibilité permanente, son soutien tout au long de mes recherches mais aussi pour ses précieux conseils.

Veillez recevoir ma grande estime pour le grand honneur que vous m'avez fait d'accepter la supervision de ce travail.

Je tiens également à exprimer ma gratitude aux membres du jury de soutenance pour avoir accepté de superviser ce travail.

Je voudrais remercier mon ami amine badji, mon cher collègue dahlouk anis pour leur compagnie et leur soutien .

Je tiens à remercier vivement M. Adel Boughida pour son aide.

Enfin, Je tiens à remercier ma famille et plus particulièrement ma mère et mon père pour leur patience et leur soutien tout au long de ma carrière.

Dédicaces

Je dédie ce modeste travail de fin d'étude à mes chers parents qui ont toujours été là pour moi , pour me soutenir et pour m'encourager tout au long de mes études.

À mes chers soeurs rania et houda et sa charmante fille iline.

À mes proches oussama, jalil, badri, amine et seif qui ont été toujours à mes côtés

Résumé

Depuis l'émergence des approches d'apprentissage profond, la reconnaissance des expressions faciales a connu une forte progression, et ces systèmes sont devenus de plus en plus efficaces. Cependant, un apprentissage efficace nécessite de grandes bases de données, lesquelles ne sont pas toujours disponibles.

Pour y remédier, il convient d'augmenter les bases de données à l'aide de techniques appropriées d'augmentation des données (AD), qui sont généralement basées sur des transformations géométriques ou des augmentations par suréchantillonnage. Ces techniques permettent d'une part d'agrandir la taille de la base d'apprentissage, et d'autre part de la rééquilibrer.

Dans ce projet, nous proposons d'explorer et d'évaluer l'utilisation de trois techniques d'augmentation de données, ainsi que l'effet de leur hybridation, sur un système de reconnaissance des expressions faciales. Deux modèles de l'apprentissage profond ont été utilisés pour cette évaluation, le VGG16 et le Resnet50. Les techniques d'augmentation testées sont les transformations géométriques, SMOTE et le Cycle-Gan.

Les résultats obtenus montrent que le fonctionnement des systèmes de FER requiert un grand volume de données pour l'apprentissage et que les GAN offrent une alternative intéressante pour pallier ce déficit de données.

Abstarct

Since the emergence of deep learning approaches, facial expression recognition has grown significantly, and these systems have become increasingly more efficient. However, effective learning requires large databases which are not always available.

To address this, databases need to be augmented using appropriate data augmentation techniques (DA), which are typically based on geometric transformations or oversampling augmentations. These techniques allow to increase the size of the training database and to rebalance it.

In this project, we propose to explore and evaluate the use of three data augmentation techniques, as well as the effect of their hybridization, on a facial expression recognition system. Two deep learning models have been used for this evaluation, the VGG16 and the Resnet50. The augmentation techniques tested are geometric transformations, SMOTE and Cycle-Gan.

The results obtained show that the operation of the RMF systems requires a large volume of data for learning and that GANs offer an interesting alternative to overcome this data deficit.

ملخص

منذ ظهور مناهج التعلم العميق، أصبح التعرف على تعابير الوجه نامياً بشكل ملحوظ، وأصبحت هذه الأنظمة أكثر كفاءة بشكل متزايد. لكن، يتطلب التعلم الفعال قواعد بيانات كبيرة لا تتوفر دائماً. لمعالجة هذا الأمر، تحتاج قواعد البيانات إلى التعزيز باستخدام تقنيات مناسبة لزيادة البيانات، والتي تعتمد عادةً على التحويلات الهندسية أو زيادة العينات، تسمح هذه التقنيات بزيادة حجم قاعدة بيانات التدريب وإعادة توازنها. في هذا المشروع، نقترح استكشاف وتقييم استخدام ثلاثة تقنيات لزيادة البيانات وكذا تأثير تهجينها على أنظمة التعرف على تعابير الوجه. تم استخدام نموذجين للتعلم العميق لهذا التقييم، VGG16 و ResNet50. تقنيات التعزيز التي تم اختبارها هي التحويلات الهندسية، SMOTE و Cycle-Gan. تظهر النتائج التي تم الحصول عليها أن تشغيل أنظمة التعرف على تعابير الوجه يتطلب حجمًا كبيرًا من البيانات للتعلم وأن الشبكات التوليدية الخصومية (GAN) تقدم بديلاً مثيراً للاهتمام للتغلب على عجز هذه البيانات.

كلمات مفتاحية: VGG16 ResNet50 SMOTE Cycle-Gan FER

Table des matières

Liste des tableaux	v
Table des figures	vi
1 L'apprentissage profond	2
1.1 Introduction	2
1.2 L'apprentissage automatique	2
1.2.1 Définition	2
1.2.2 Les méthodes d'apprentissage automatique	2
1.3 Apprentissage en profondeur :	3
1.4 Réseau de neurones	3
1.4.1 Le perceptron multicouche	4
1.4.2 Différentes architectures de réseau de neurones	5
1.4.3 Les domaines d'application des CNN :	9
1.4.4 Conditionnel GAN :	10
1.4.5 Cycle-Gan :	11
1.4.6 L'entraînement des GANs :	11
1.4.7 Les domaines d'application d'un GAN :	11
1.5 Auto-encodeur	12
1.5.1 Domaines d'utilisation d'un auto encodeur :	13
1.6 Les mesures de performances	13
1.6.1 Matrice de confusion :	14
1.6.2 Exactitude 'Accuracy' :	14
1.6.3 La courbe ROC	14
1.6.4 Le Rappel :	14
1.7 conclusion	15
2 L'augmentation de données en reconnaissance des expressions faciales	16
2.1 Introduction	16
2.2 Les expressions faciales	16
2.2.1 La reconnaissance des expressions faciales	16
2.2.2 Les domaines d'application de la reconnaissance des expressions faciales	17
2.2.3 Les difficultés de la reconnaissances d'expressions faciales	17
2.3 L'augmentation des données	18
2.4 Les techniques d'augmentation des données	18
2.4.1 Les Transformations géométriques :	18
2.4.2 Les techniques basés sur l'apprentissage profond (le deep learning)	20
2.5 Bases de données d'expressions faciales :	20

2.6	Quelques travaux connexe sur l'augmentation des données pour la reconnaissance des expressions faciales	22
2.6.1	Article 1 "Evaluation of Data Augmentation Techniques for Facial Expression Recognition Systems" [Porcu <i>et al.</i> , 2020]	22
2.6.2	Article 2 "Facial Expression Recognition with Convolutional Neural Networks via a Data Augmentation Strategy" [Yan <i>et al.</i> , 2021]	23
2.6.3	Article 3 "Facial expression recognition with trade-offs between data augmentation and deep learning features" [Umer <i>et al.</i> , 2022]	23
2.6.4	Article 4 "Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy" [Li <i>et al.</i> , 2020]	24
2.6.5	Article 5 "A Facial Expression Recognition Method Using Deep Convolutional Neural Networks Based on Edge Computing" [Chen <i>et al.</i> , 2020]	24
2.6.6	Article 6 "Real Time Face Expression Recognition along with Balanced FER2013 Dataset using CycleGAN" [Mazen <i>et al.</i> ,	25
2.6.7	Article 7 "Data Augmentation Merging Dataset for Facial Emotion Recognition" [Kim et Han, 2020]	25
2.6.8	Article 8 "Small facial image dataset augmentation using conditional GANs based on incomplete edge feature input" [Hung et Gan, 2021]	26
2.7	conclusion	26
3	Conception	27
3.1	Introduction	27
3.2	Objectif	27
3.3	Architecture du système	27
3.3.1	Détection de visage	29
3.3.2	Choix du modèle	29
3.3.3	Choix de la technique d'augmentation des données	30
3.4	Conclusion	32
4	Implémentation	33
4.1	Introduction	33
4.2	Environnement	33
4.2.1	Kaggle	33
4.2.2	Google Colaboratory	33
4.2.3	Apprentissage sur GPU :	34
4.3	Langage de programmation et bibliothèque utilisée	35
4.3.1	Python :	35
4.3.2	Bibliothèque utilisée :	35
4.3.3	Tensorflow	35
4.3.4	Keras	35
4.3.5	Numpy	35
4.3.6	Matplotlib	36
4.3.7	Tkinter	36
4.4	Base d'apprentissage	36

4.5	Apprentissage et test	37
4.5.1	Les expérimentations réalisées	37
4.5.2	Résultats et discussion	38
4.5.3	Les modèles sans augmentation des données	38
4.5.4	Les modèles avec une des techniques d'augmentation des données .	39
4.5.5	Hybridation des techniques d'augmentation des données	41
4.6	Conclusion	42
	Bibliographie	48
	Webographie	49

Liste des tableaux

1.1	matrice de confusion	14
2.1	FER2013	21
2.2	Tableau récapitulatif sur les bases de données d'expression faciales	22
4.1	Paramètres d'apprentissage des différentes expérimentations	37
4.2	Tableau récapitulatif des précisions obtenues pour les expérimentations effectués	38

Table des figures

1.1	le processus d'apprentissage automatique	2
1.2	Schéma d'un neurone biologique	4
1.3	Réseau de neurones	4
1.4	Quelques topologies standards [Hammouche, 2009]	5
1.5	reseaux de neurones convolutif	6
1.6	opération de pooling	7
1.7	les différents types de pooling	7
1.8	L'architecture de la couche entièrement connectée	8
1.9	La fonction du couche Relu	8
1.10	technique de dropout [Srivastava <i>et al.</i> , 2014]	9
1.11	Architecture de base d'un réseau adversaire génératif	10
1.12	Exemple de transformation d'un objet, du cheval au zèbre et du zèbre au cheval.	11
1.13	entraînement d'un GAN	12
1.14	Architecture d'un auto-encodeur	13
1.15	La courbe ROC	15
2.1	expressions faciales	17
2.2	Augmentation de luminosité	19
2.3	Reffet horizontal	19
2.4	translation d'une image	20
2.5	rotation d'une image	20
2.6	JAFFE dataset	21
3.1	L'architecture de notre système	28
3.2	Architecture de VGG16 [Porcu <i>et al.</i> , 2020]	29
3.3	La technique SMOTE	31
3.4	Cycle gan	32
4.1	Interface de Colab	34
4.2	Activation de GPU	34
4.3	distribution des classes dans la base de données FER2013	36
4.4	Expérimentation 1	39
4.5	Matrice de confusion	39
4.6	Ajouter une seule technique d'augmentation des données	40
4.7	Les Matrices de confusion pour les expérimentations avec une seule technique d'augmentation des données	41
4.8	Hybridation des techniques d'augmentation des données	42
4.9	Hybridation des techniques d'augmentation des données	43
4.10	Hybridation des techniques d'augmentation des données	43

Introduction générale

Les réseaux de neurones à convolution ont permis d'obtenir des résultats remarquables dans de nombreuses tâches de vision par ordinateur. Cependant, ces réseaux sont fortement dépendants de grandes quantités de données pour éviter le surajustement (l'overfitting). Malheureusement, de nombreux domaines d'application n'ont pas accès à de grandes données, comme la reconnaissance des expressions faciales.

Ces dernières années, les applications de la reconnaissance des expressions faciales (FER) ne cessent de se répandre, essentiellement dans les domaines de l'interaction homme-machine, des jeux et du secteur de la santé. Selon Ekman et Friesen, les six émotions de base (colère, peur, dégoût, joie, surprise et tristesse) sont les principales expressions émotionnelles communes aux êtres humains [Ekman et Friesen, 1971].

Bien que de nombreuses études aient été menées sur la reconnaissance des expressions faciales, elle reste l'une des tâches les plus difficiles dans les systèmes de classification d'images en raison du chevauchement important entre les classes des expressions de base, des différences dans la manifestation culturelle de l'émotion et de la nécessité d'une grande quantité de données d'entraînement pour éviter le surajustement. L'augmentation des données est une méthode très puissante généralement utilisée pour résoudre ces problèmes. Les techniques d'augmentation des données sont basées sur des transformations géométriques ou des augmentations par suréchantillonnage.

Dans ce mémoire, nous proposons d'évaluer et de comparer l'effet de l'utilisation de diverses techniques d'augmentation des données sur la précision de la reconnaissance des expressions faciales dans les systèmes de FER basés sur l'apprentissage profond. Deux modèles de réseaux de neurones convolutifs VGG16 et ResNet50 ont été utilisés pour cette évaluation. Les techniques d'augmentation considérées sont les transformations géométriques, SMOTE et le réseau d'adversaire génératif CycleGan.

Notre mémoire est structuré en quatre chapitres qui sont présentés brièvement comme suit :

Chapitre 1 : L'apprentissage profond

Dans ce chapitre, nous présentons les concepts de base de l'apprentissage profond et fournissons une description détaillée des réseaux neuronaux convolutifs et des réseaux adversaires génératifs.

Chapitre 2 : L'augmentation de données en reconnaissance des expressions faciales

Ce chapitre se focalise sur les systèmes de reconnaissance des expressions faciales et les difficultés rencontrées, ainsi que sur les bases de données de visages existantes. Un bref état de l'art de quelques recherches récentes dans le domaine est également présenté.

Chapitre 3 : Conception

Dans ce chapitre nous avons détaillé l'architecture globale de notre système de reconnaissance des expressions faciales et les principales phases de sa conception.

Chapitre 4 : Implémentation

Nous avons abordé l'aspect implémentation de notre application qui consiste à affiner les concepts précédemment développés ainsi que l'environnement de développement, les détails des tests et les principaux résultats obtenus, et leur interprétation.

Nous achevons ce mémoire par une conclusion générale et des suggestions pour les travaux futurs. .

Chapitre 1

L'apprentissage profond

1.1 Introduction

D'après le père de l'intelligence artificielle, John McCarthy. L'intelligence artificielle consiste à concevoir des machines capables de penser pour nous aider à réduire l'erreur humaine et augmenter les chances d'atteindre un degré de précision supérieur, depuis son apparition le siècle précédent, l'IA a un impact considérable sur de nombreux domaines tels que le domaine médical, les jeux, les systèmes experts, la reconnaissance de la parole, la reconnaissance de l'écriture manuscrite et les robots intelligents. Ces systèmes sont adaptables à l'environnement qui les entoure et les humains définissent de mieux en mieux leurs besoins et transforment rapidement ce besoin en réalité [W1].

1.2 L'apprentissage automatique

1.2.1 Définition

L'apprentissage automatique ou Machine Learning (ML) est une branche de l'intelligence artificielle qui consiste à programmer des algorithmes permettant d'apprendre automatiquement à partir des données et d'expériences passées ou par interaction avec l'environnement. Ce qui rend l'apprentissage machine vraiment utile est le fait que l'algorithme peut "apprendre" et adapter ses résultats en fonction de nouvelles données sans aucune programmation à priori [Mifdal, 2019].

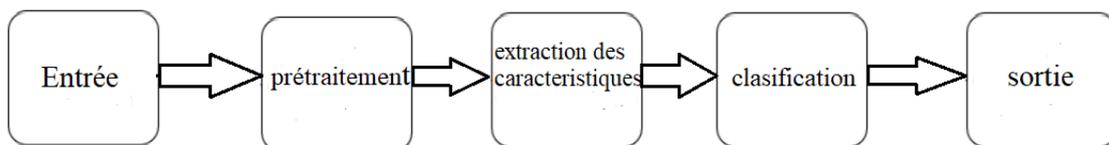


FIGURE 1.1 – le processus d'apprentissage automatique

1.2.2 Les méthodes d'apprentissage automatique

Selon les problèmes à résoudre et les données disponibles, il existe plusieurs façons d'apprendre automatiquement à partir des données [Mifdal, 2019].

1. Apprentissage supervisé :

L'algorithme est formé à l'aide d'une base de données apprise contenant des exemples de cas réels traités et validés. L'objectif est de trouver des corrélations entre les données d'entrée (variables explicatives) et les données de sortie (variables à prédire), puis de déduire les connaissances extraites sur les entrées avec des sorties inconnues.

2. Apprentissage non supervisé :

Pour ce type d'apprentissage, la base de données d'apprentissage ne contient pas la variable cible, seul un ensemble de données est collecté en entrée, l'algorithme doit découvrir la structure par lui-même à partir des données, nous utilisons cette technique pour diviser les données en groupes de éléments homogène. La distance est généralement la mesure de similarité entre les groupes la plus couramment utilisée.

3. Apprentissage semi-supervisé

il se compose d'un mélange d'apprentissage supervisé et non supervisé, utilisant à la fois des données étiquetées et non étiquetées provenant du même ensemble de données. L'avantage de cette approche est que l'étiquetage des données peut être coûteux et prend souvent beaucoup de temps. En outre, elle peut entraîner un biais humain dans les données étiquetées. Dans ce cas, ce type d'apprentissage avec seulement quelques étiquettes est très pratique. L'inclusion de grandes quantités de données non étiquetées dans la formation tend à améliorer les performances du modèle tout en réduisant le temps et le coût nécessaires à sa construction.

4. Apprentissage par renforcement :

L'apprentissage se déroule sans supervision, et en interaction avec l'environnement en observant les résultats des actions entreprises. Chaque action de la chaîne est associée à une récompense. L'objectif est de déterminer la stratégie comportementale optimale pour maximiser la récompense totale. Pour cela, un simple retour est nécessaire pour apprendre le fonctionnement de la machine. C'est ce qu'on appelle le renforcement du signal.

1.3 Apprentissage en profondeur :

L'apprentissage profond est une manière particulière d'apprentissage automatique qui permet d'intégrer des réseaux neuronaux en couches successives pour obtenir des données de façon itérative. L'apprentissage profond est très pratique lorsque vous essayez de découvrir des modèles à partir de données non structurées.

1.4 Réseau de neurones

Les réseaux de neurones, fabriqués de structures cellulaires artificielles, constituent une approche permettant d'aborder sous des angles nouveaux les problèmes de perception, de mémoire, d'apprentissage et de raisonnement. Ils s'avèrent aussi des alternatives très prometteuses pour contourner certaines des limitations des ordinateurs classiques. Grâce à leur traitement parallèle de l'information et à leurs mécanismes inspirés des cellules nerveuses (neurones) (comme le montre la figure 1.2), ils infèrent des propriétés émergentes permettant de solutionner des problèmes auparavant appelé complexe [\[Parizeau, 2004\]](#). C'est en 1943 que Mc Culloch (neurophysiologiste) et Pitts (logicien) ont proposé les

premières notions de neurone formel. Ce concept fut ensuite mis en réseau avec une couche d'entrée et une sortie par Rosenblatt en 1959 pour simuler le fonctionnement rétinien et tacher de reconnaître des formes. C'est l'origine du perceptron [Wira, 2009]

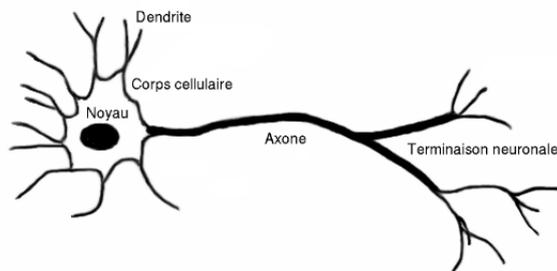


FIGURE 1.2 – Schéma d'un neurone biologique

Un réseau de neurones est un système composé de neurones, généralement répartis en plusieurs couches connectées entre elles, s'utilise pour résoudre divers problèmes, le réseau calcule à partir de l'entrée un score (ou probabilité) pour chaque classe. La classe attribuée à l'objet en entrée correspond à celle de score le plus élevé.

Plusieurs architectures de réseaux de neurones ont été proposées, le plus important fut le perceptron multicouche.

1.4.1 Le perceptron multicouche

Le perceptron multicouche est un des réseaux de neurones les plus utilisés pour des problèmes d'approximation, de classification et de prédiction. Il est habituellement constitué de deux ou trois couches de neurones totalement connectés (Comme le montre la figure 1.3) ce type de réseau est dans la famille générale des réseaux à «propagation vers l'avant», c'est-à-dire que en mode normal d'utilisation, l'information se propage dans un sens unique, des entrées vers les sorties sans aucune rétroaction. Son apprentissage est de type supervisé [Parizeau, 2004]

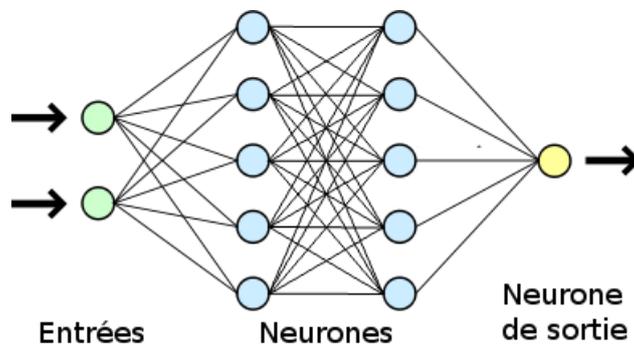


FIGURE 1.3 – Réseau de neurones

- Chaque couche reçoit en entrée des données et les renvoie transformées. Pour cela, elle calcule une combinaison linéaire puis applique éventuellement une fonction non-linéaire, appelée fonction d'activation. Les coefficients de la combinaison linéaire définissent les paramètres (ou poids) de la couche

- La dernière couche calcule les probabilités finales en utilisant pour fonction d'activation la fonction logistique (classification binaire) ou la fonction softmax (classification multi-classes) .

1.4.2 Différentes architectures de réseau de neurones

Il existe plusieurs topologies de réseaux de neurones, on appelle topologie des réseaux la façon dont les neurones sont connectés entre eux à travers les différentes couches. En général, on peut distinguer deux grandes classes de réseaux de neurones artificiels selon leurs topologies : les réseaux à couches qui sont des réseaux de neurones dans lesquels l'information se propage couche par couche sans retour en arrière possible et les réseaux récurrents qui sont des réseaux de neurones dans lesquels il y a une liaison vers l'arrière [Hammouche, 2009]

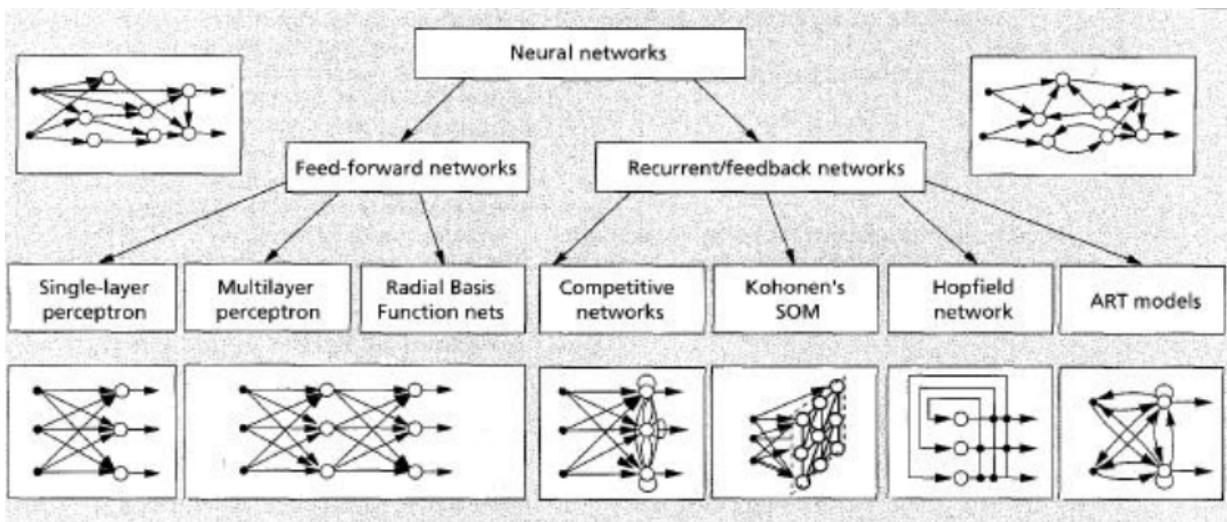


FIGURE 1.4 – Quelques topologies standards [Hammouche, 2009]

Parmi les architectures de réseaux de neurones , nous détaillons quelques unes : Les réseaux de neurones convolutifs , les réseaux antagonistes génératifs , les auto encodeurs

- A) **Les réseaux de neurones convolutifs** Les réseaux de neurones convolutifs (en anglais : convolutional neural network ,CNN) ont une méthodologie similaire à celle des méthodes traditionnelles d'apprentissage supervisé : ils reçoivent des images en entrée, détectent les caractéristiques (dites features en anglais) de chacune d'entre elles, puis entraînent un classifieur, l'architecture spécifique du réseau permet d'extraire des caractéristiques de différentes complexités, des plus simples au plus sophistiquées. Cette extraction et la hiérarchisation automatiques des caractéristiques, qui s'adaptent au problème donné, constituent une des forces des réseaux de neurones convolutifs [W2].

Dans les CNN typiques , il y a un nombre de couches de convolution , suivies de couches de mise en commun et en couche final il y a généralement des couches entièrement connecté (MLP) , comme le montre la figure 1.5

1) La couche de convolution :

Une couche de convolution est un composant fondamental de l'architecture

CNN qui effectue l'extraction de caractéristiques, laquelle consiste généralement en une combinaison d'opérations linéaires et non linéaires, c'est-à-dire une opération de convolution et une fonction d'activation [Yamashita *et al.*, 2018], sous réserve d'appliquer le bon «filtre». En fait, une convolution prend simplement en entrée une image et un filtre (qui est une autre image), effectue un calcul, puis renvoie une nouvelle image (généralement plus petite).

Les filtres d'un réseau de neurones convolutifs ont un noyau (kernel), un pas (stride), un remplissage (padding) définis à l'avance mais leurs valeurs sont générées aléatoirement à l'initialisation. Puis, lorsque le réseau effectue son apprentissage, les valeurs des filtres sont actualisées pour améliorer les résultats du CNN les valeurs des filtres font partie des paramètres (poids, biais...) que le réseau modifie pendant l'apprentissage. Trois hyper paramètres permettent

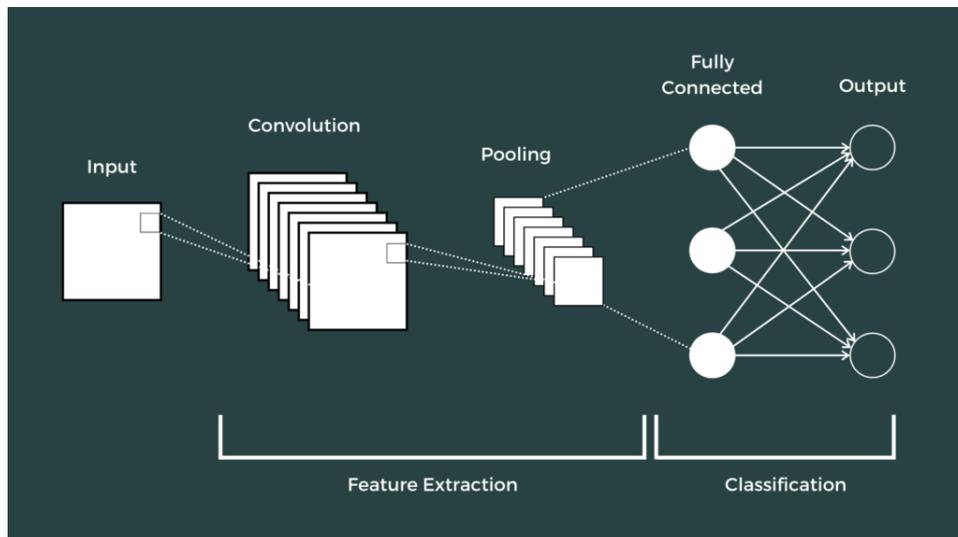


FIGURE 1.5 – reseaux de neurones convolutif

de dimensionner le volume de la couche de convolution (aussi appelé volume de sortie) [Manel,] :

- i. **Profondeur** : Profondeur de la couche est le nombre de noyaux de convolution (ou nombre de neurones associés à un même champ récepteur).
 - ii. **Le pas (Stride)** : Stride est le nombre de pixels par lequel est glissé un filtre sur l'image. Lorsque le stride est 1 alors les filtres sont déplacés d'un pixel. Avoir un plus grand pas produira des cartes de caractéristiques plus petites .
 - iii. **Remplissage nul (Zero-padding)** :il est commode de mettre des zéros à la frontière du volume d'entrée. La taille de ce "zero-padding" est le troisième hyperparamètre. Cette marge permet de contrôler la dimension spatiale du volume de sortie. En particulier, il est parfois souhaitable de conserver la même surface que celle du volume d'entrée .
- 2) **La couche de mise en commun** :
- Une couche de mise en commun (en anglais pooling) fournit une opération typique de sous-échantillonnage qui réduit la dimensionnalité dans le plan des cartes de caractéristiques afin d'introduire une invariance de translation à de

petits décalages et distorsions, et de diminuer le nombre de paramètres apprenables ultérieurs. Il est à noter qu'il n'y a aucun paramètre apprenable dans aucune des couches de mise en commun, alors que la taille du filtre, le pas (stride) et le remplissage (padding) sont des hyperparamètres dans les opérations de mise en commun, similaires aux opérations de convolution. [Yamashita *et al.*, 2018]

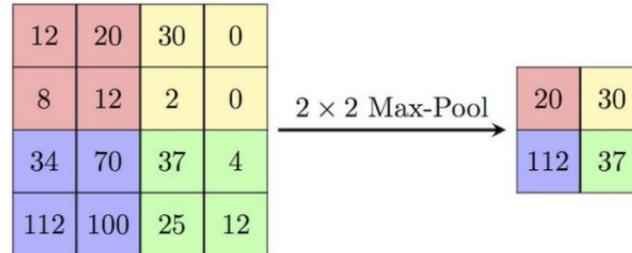


FIGURE 1.6 – opération de pooling

Il existe différents types de pooling :

- i. Le « max pooling », qui revient à prendre la valeur maximale de la sélection. C'est le type le plus utilisé car il est rapide à calculer (immédiat), et permet de simplifier efficacement l'image
- ii. Le « mean pooling » (ou average pooling), soit la moyenne des pixels de la sélection : on calcule la somme de toutes les valeurs et on divise par le nombre de valeurs. On obtient ainsi une valeur intermédiaire pour représenter ce lot de pixels
- iii. Le « sum pooling », c'est la moyenne sans avoir divisé par le nombre de valeurs (on ne calcule que leur somme)
- iv. le « Stochastic pooling » est une méthode probabiliste, qui se situe quelque part entre les deux (max pooling et mean pooling). En donnant la probabilité des points de pixels selon la valeur numérique, puis en sous-échantillonnant selon la probabilité, au sens moyen, elle est similaire à la mise en commun moyenne, et au sens local, elle obéit au max, il obéit aux directives du max pooling. [Jiang *et al.*, 2020]

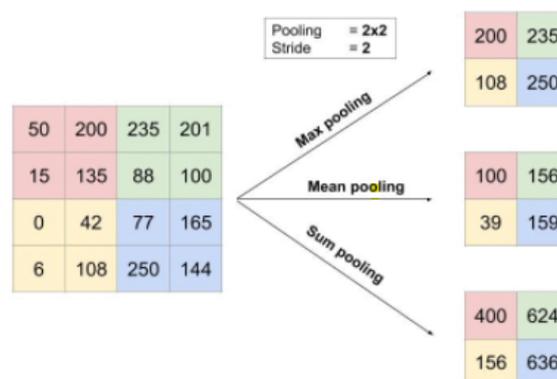


FIGURE 1.7 – les différents types de pooling

3) La couche entièrement connectée :

La couche entièrement connectée (en anglais fully-connected) constitue toujours la dernière couche d'un réseau de neurones convolutif. Ce type de couche reçoit un vecteur en entrée et produit un nouveau vecteur en sortie. Pour cela, elle applique une combinaison linéaire, puis éventuellement une fonction d'activation aux valeurs reçues en entrée [W3].

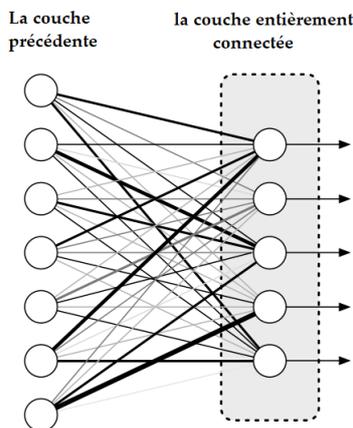


FIGURE 1.8 – L'architecture de la couche entièrement connectée

À ces trois couches de bases, s'ajoutent deux autres couches (la rectification linéaire et le dropout) pour un meilleur résultat.

4) La couche ReLU (rectification linéaire) :

La fonction d'activation ReLU est une fonction dite « rectifier » très utilisée en Deep Learning dans les réseaux de neurones convolutionnels, elle est appliquée très souvent en sortie d'une couche de convolutions.

Le but de ReLU est d'augmenter la non-linéarité du CNN, puisque la sémantique d'une image (par exemple, une personne et un chien Husky assis l'un à côté de l'autre) est évidemment un modèle hautement non linéaire de valeurs de pixels en entrée.

Pour que la cartographie de l'entrée du CNN vers sa sortie soit également hautement non-linéaire. La fonction ReLU, bien que simple, est une fonction non linéaire, comme l'illustre la Figure 1.9 [Wu, 2017].

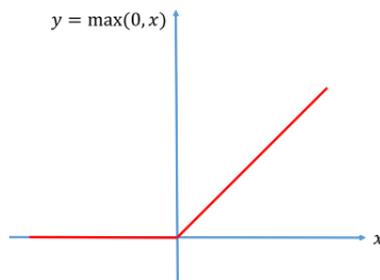


FIGURE 1.9 – La fonction du couche Relu

ReLU, de par sa définition, est une fonction qui vient briser (une partie de)

la linéarité en supprimant une partie des valeurs (toutes celles négatives) , elle permet également d'accélérer les calculs.

5) La couche drop out :

L'abandon (connu généralement par le terme anglais dropout) est une technique qui empêche le surajustement et permet de combiner efficacement un nombre exponentiel d'architectures de réseaux neuronaux différentes de manière efficace. Le terme "dropout" fait référence à l'abandon d'unités (cachées et visibles) dans un réseau neuronal, l'abandon d'une unité, signifie sa suppression temporaire du réseau, avec toutes ses unités entrantes. le réseau, ainsi que toutes ses connexions entrantes et sortantes, comme le montre la figure 1.10, Le choix des unités à éliminer est aléatoire. Dans le cas le plus simple [Srivastava et al., 2014](#)

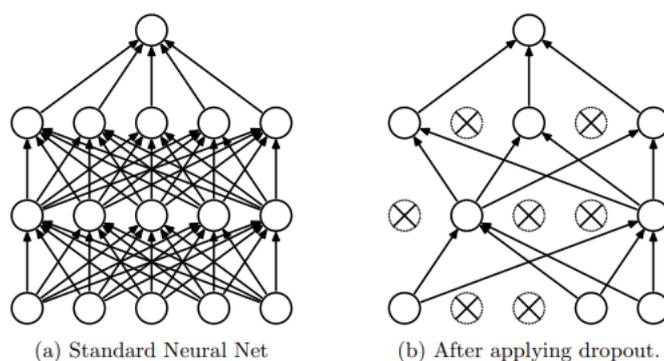


FIGURE 1.10 – technique de dropout [Srivastava et al., 2014](#)

1.4.3 Les domaines d'application des CNN :

Le CNN a fait de brillantes réalisations. Il est devenu l'un des réseaux neuronaux les plus représentatifs dans le domaine de l'apprentissage profond.

La vision par ordinateur basée sur les réseaux de neurones convolutionnels a permis aux gens d'accomplir ce qui était considéré comme impossible au cours des siècles passés, comme : la reconnaissance des visages, les véhicules autonomes, les supermarchés en libre-service et les traitements médicaux [Li et al., 2021](#) .

B) Réseau antagoniste génératif

Un réseau antagoniste génératif (en anglais Generative adversarial networks (GAN)) est une technique de Machine Learning et est une classe puissante de modèles génératifs introduite en 2014 par Goodfellow et al [Goodfellow et al., 2014](#), elle repose sur la mise en compétition de deux réseaux au sein d'un framework , ces deux réseaux sont appelés «**générateur**» et «**discriminateur**» [Wang et al., 2017](#) .

- Le principe de base des GANs est inspiré d'un jeu à somme nulle à deux joueurs, dans lequel le total des gains est égal au total des pertes.
- Les gains des deux joueurs sont nuls, et le gain ou la perte d'utilité de chaque joueur est exactement équilibré par le gain ou la perte d'utilité de l'autre joueur
- Les deux modèles, génératif et discriminatif, sont réalisés sous forme de perceptrons multicouches ,ils apprennent simultanément , Le générateur essaie de capturer la

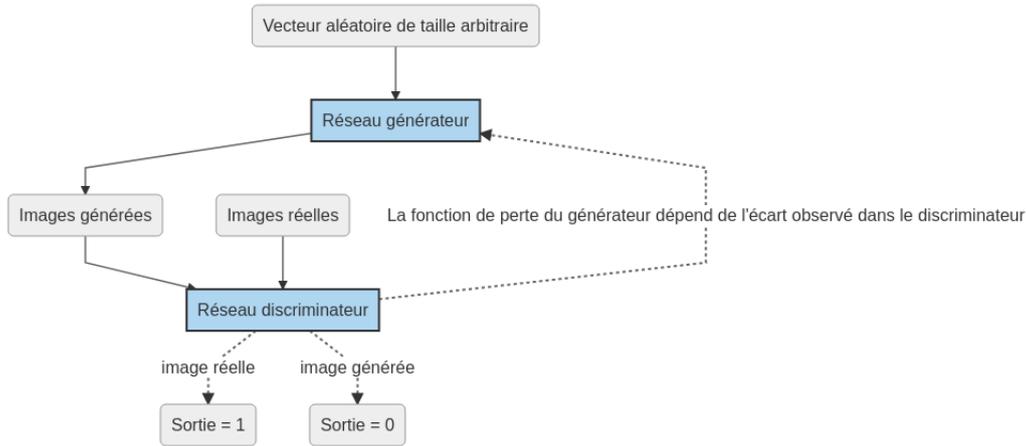


FIGURE 1.11 – Architecture de base d'un réseau adversaire génératif

distribution potentielle d'échantillons réels, et génère de nouveaux échantillons de données, et Le discriminateur est souvent un classificateur binaire, qui discrimine les échantillons réels des échantillons générés aussi précisément que possible.

- Formellement, le GAN est un modèle probabiliste structuré modélisant des variables latentes z et des variables observées x [Ledig *et al.*, 2017]. Le générateur G échantillonne z à partir d'une distribution antérieure prédéfinie $p(z)$ et donne un "faux" échantillon de $G(z)$, associé à un échantillon réel x tiré de la distribution de données réelles p_{data} . à partir de la distribution de données réelles p_{data} , l'objectif d'apprentissage de GAN peut être formulé comme suit [Wang *et al.*, 2020] :

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{data}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))] \quad (1)$$

1.4.4 Conditionnel GAN :

Les réseaux adversariens génératifs peuvent être étendus à un modèle conditionnel si le générateur et le discriminateur sont conditionnés par une information supplémentaire y , y peut être n'importe quel type d'information auxiliaire, comme des étiquettes de classe ou des données provenant d'autres modalités. Nous pouvons effectuer le conditionnement en introduisant y dans le discriminateur et le générateur comme couche d'entrée supplémentaire [Mirza et Osindero, 2014]

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{data}(\mathbf{x})} [\log D(\mathbf{x}|\mathbf{y})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z}|\mathbf{y})))] \quad (2)$$

1.4.5 Cycle-Gan :

Le CycleGAN est une technique qui consiste à entraîner automatiquement des modèles de traduction d'image à image sans exemples appariés. Les modèles sont formés de manière non supervisée à l'aide d'une collection d'images du domaine source et du domaine cible qui n'ont pas besoin d'être liées de quelque manière que ce soit, Cette technique simple est puissante et permet d'obtenir des résultats visuellement impressionnants dans toute une série de domaines d'application, notamment la traduction de photographies de chevaux en zèbres, et inversement [Zhu *et al.*, 2017].

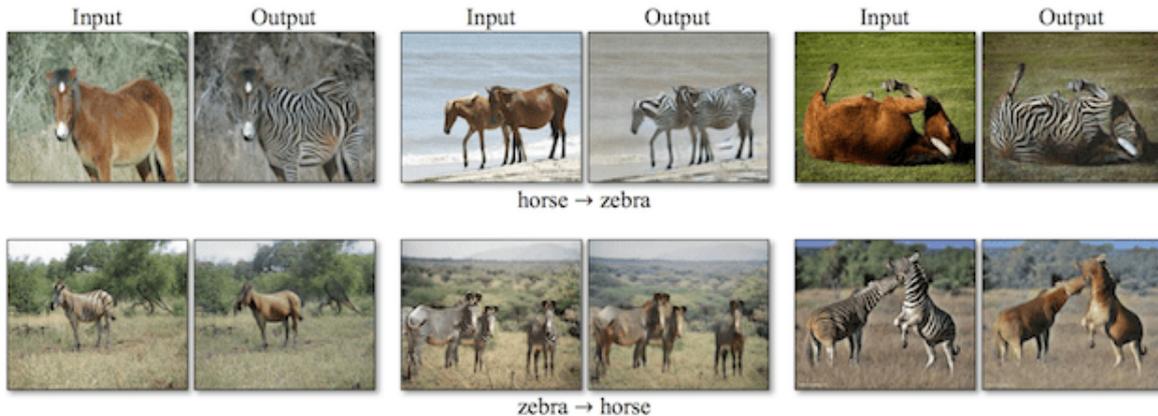


FIGURE 1.12 – Exemple de transformation d'un objet, du cheval au zèbre et du zèbre au cheval.

Un CycleGAN est composé de 2 GAN, ce qui fait un total de 2 générateurs et 2 discriminateurs, il utilise une extension supplémentaire de l'architecture appelée cohérence des cycles. Il s'agit de l'idée qu'une image produite par le premier générateur peut être utilisée comme entrée du second générateur et que la sortie du second générateur doit correspondre à l'image originale. L'inverse est également vrai une sortie du deuxième générateur peut être utilisée comme entrée du premier générateur et le résultat doit correspondre à l'entrée du deuxième générateur [Zhu *et al.*, 2017].

1.4.6 L'entraînemnt des GANs :

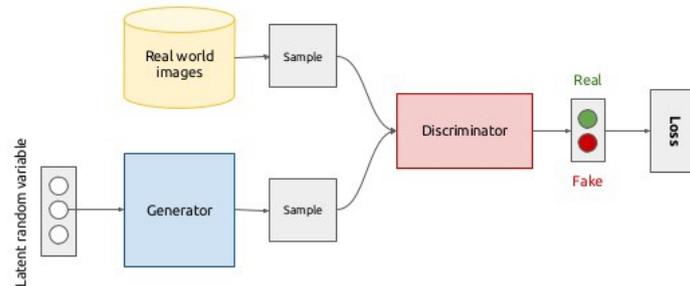
L'entraînement des GAN implique à la fois de trouver les paramètres d'un discriminateur D qui maximisent la précision de sa classification et de rechercher les paramètres d'un générateur G qui confondent au maximum le discriminateur [Abraich *et al.*, 2020], le processus d'entraînement est résumé à la Figure 1.13

1.4.7 Les domaines d'application d'un GAN :

Les réseaux adversaires génératifs sont présents dans des domaines importants ils peuvent également être d'une grande utilité :

- L'étude de techniques puissantes de génération d'images 3D est un point de vue fondamental dans le domaine de la vision par ordinateur et de l'infographie, Pour identifier les objets 3D et améliorer la vision par ordinateur, certaines études ont utilisé des réseaux GAN. Par exemple, Yu *et al.* ont proposé un réseau qui traite les données floues sans étiquetage, et l'idée d'un encodeur de points 3D GAN Point

Generative adversarial networks (conceptual)



5

FIGURE 1.13 – entraînement d'un GAN

encoder a été utilisée en peinture et utilise une couche de maxpooling pour résoudre les points pour le processus d'apprentissage. [Yu *et al.*, 2020]

- Les concepts d'apprentissage automatique et d'apprentissage profond sont fortement impliqués dans le domaine médical notamment pour identifier les maladies chroniques [Battineni *et al.*, 2020], Il est rapporté que le GAN peut produire avec succès des images RM dans des pixels scellés [Baek *et al.*, 2020]
- Différentes enquêtes ont mis en évidence que les détails précis du trafic routier par les détecteurs sont plus précis que les prédictions des différents modèles. Xu et al ont développé un cadre d'estimation du trafic routier à l'aide de l'apprentissage profond appelé GE-GAN. comme étude de cas [Xu *et al.*, 2020]. Les images du trafic peuvent détecter et classer les occurrences du trafic, ce qui aide à contrôler les voitures autonomes sur les routes [Fathi-Kazerooni et Rojas-Cessa, 2020]

1.5 Auto-encodeur

L'auto encodeur AE est un type de réseau neuronal artificiel non supervisé qui est apparu à la fin des années 80 , L'AE a été considéré comme un outil puissant pour l'extraction automatique de caractéristiques non linéaires qui permettent de construire une nouvelle représentation d'un jeu de données . Comme l'illustre la figure 1.14, un AE de base est constitué de trois couches : couche d'entrée, couche cachée et couche de sortie. Le nombre de neurones dans chaque couche est n , m , n . La couche d'entrée et un espace latent construisent un codeur. , L'espace latent et la couche de sortie constituent un décodeur. [Zhang *et al.*, 2020]

- les couches de neurones du décodeur, reçoivent ces représentations et les traitent afin d'essayer de reconstruire les données de départ.
- Les différences entre les données reconstruites et les données initiales permettent de mesurer l'erreur commise par l'auto-encodeur.
- L'entraînement consiste à modifier les paramètres de l'auto-encodeur afin de réduire l'erreur de reconstruction mesurée sur les différents exemples du jeu de données.

1.5.1 Domaines d'utilisation d'un auto encodeur :

il y a plusieurs utilisations des auto encodeur , leur fonctionnement permet de réaliser de nombreuses tâches très utiles [W4] :

- De détecter les anomalies en ne parvenant pas à restituer la valeur d'entrée si celle-ci dévie trop du schéma appris par le réseau de neurones. Si le réseau est capable de restituer un ensemble de données, il doit en effet pouvoir dire si de nouvelles données sont normales ou anormales en étant capable (ou non) de les restituer.
- Réduire le nombre de dimensions nécessaires pour représenter un problème en utilisant un nombre d'unités inférieur à ce qui est nécessaire pour avoir comme entrée (par exemple, en passant de trois à deux dimensions)
- Décoder l'information en prenant des données d'entrée qui contiennent du bruit, et les transformer en informations claires.
- Comprimer l'information sous une forme plus petite, en utilisant la réduction de dimension. Moins de dimensions signifie également moins d'espace pour les données.

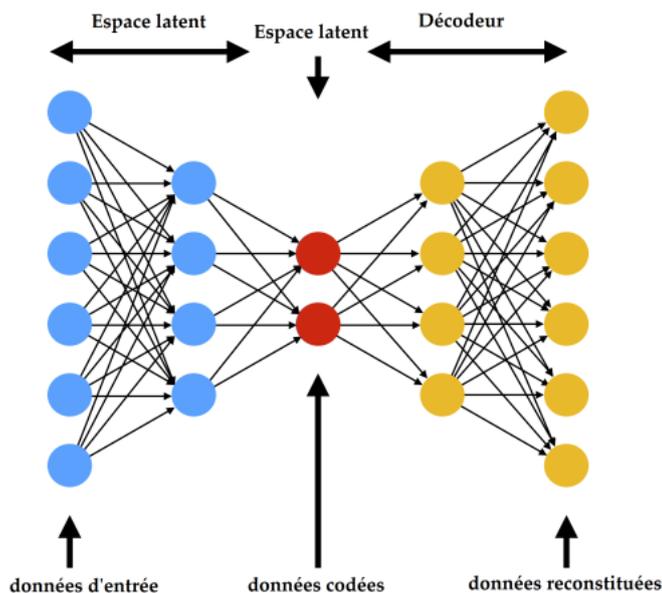


FIGURE 1.14 – Architecture d'un auto-encodeur

1.6 Les mesures de performances

Il est important de pouvoir évaluer la performance d'un modèle afin de mesurer les risques mais aussi de comparer plusieurs algorithmes et/ou modèles. Les indicateurs de performance des modèles ont été développés principalement pour aider à répondre à la question suivante : Dans quelle mesure puis-je faire confiance à un modèle pour prédire des événements futurs ?

Il faut également calculer un certain nombre de paramètres tels que :

- (A) Vrai positif 'True positive TP' : prédiction positive correcte.

- (B) Vrai négatif 'True negative TN' : prédiction negative correcte
- (C) Faux positif 'false positive FP' : prédiction positive incorrecte
- (D) Faux négatif 'false negative FN' : prédiction negative incorrecte

1.6.1 Matrice de confusion :

Une matrice de confusion est un tableau souvent utilisé pour résumer la performance de classification d'un classificateur par rapport à un ensemble de données de test dont les valeurs réelles sont connues [Sammut et Webb, 2011]

	Classe réelle positive	Classe réelle négative
Classe prédite positive	Vrai positif (VP)	Faux positif (FP)
Classe prédite negative	Faux négatif (FN)	Vrai négatif (VN)

TABLE 1.1 – matrice de confusion

1.6.2 Exactitude 'Accuracy' :

La précision de classification peut être obtenue à partir de cette matrice comme suit :

$$\text{Accuracy} = \frac{VP+VN}{VP+VN+FN+FP}$$

1.6.3 La courbe ROC

Le graphe de La courbe ROC (en anglais Receiver Operating Characteristic), c'est un graphe qui indique les performances du modèle de classification. Cette courbe est dessinée avec deux paramètres : taux de vrais positifs (TVP) en fonction du taux de faux positifs (TFP) .

Le taux de vrais positifs (TVP) :

$$\text{TVP} = \frac{VP}{VP+FN}$$

Le taux de faux positifs (TFP) :

$$\text{TFP} = \frac{FP}{FP+VN}$$

1.6.4 Le Rappel :

Le rappel permet de savoir quelle proportion de vrais résultats positifs a été correctement identifiée .

$$\text{Rappel} = \frac{VP}{VP+FN}$$

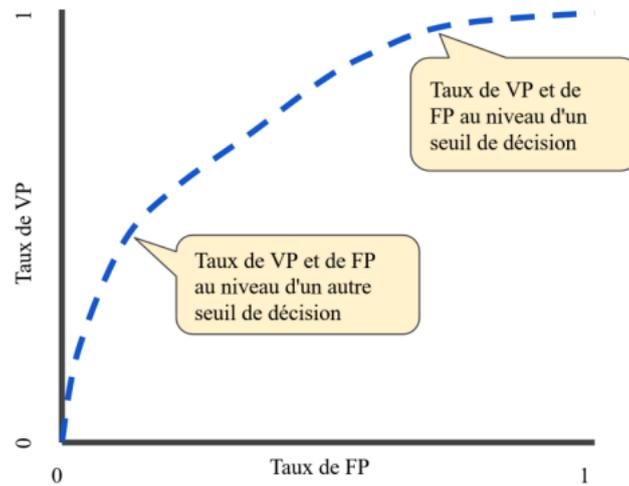


FIGURE 1.15 – La courbe ROC

1.7 conclusion

Le réseau neuronal convolutif est une technique d'apprentissage profond pour les tâches courantes de reconnaissance dans le domaine visuel. Comme toutes les techniques d'apprentissage profond, elles dépendent fortement de la qualité et le volume des données d'entraînement. Avec un ensemble de données bien préparé, les CNN sont capables de surpasser les humains dans les tâches de reconnaissance visuelle. Les GAN constituent un développement intéressant de l'apprentissage non supervisé. Ils ont été capables de générer des images si réalistes que les humains sont incapables de dire qu'elles représentent des images synthétiques .

Chapitre 2

L'augmentation de données en reconnaissance des expressions faciales

2.1 Introduction

La reconnaissance des expressions faciales (en anglais facial expression recognition , FER) est une tâche difficile qui repose sur des approches d'apprentissage automatique, ces dernières nécessitent de grandes bases de données pour un apprentissage efficace. Bien que de nombreuses études aient été menées sur la reconnaissance des expressions faciales, elle reste l'une des tâches les plus difficiles pour les systèmes de classification d'images , et parmi les problèmes rencontrés dans ce domaine : un chevauchement important dans le nombre d'échantillons entre les classes d'émotions de base , les différences dans la manifestation culturelle de l'émotion et le surajustement. une des solutions généralement utilisées pour résoudre ce problème est : l'augmentation des données .

2.2 Les expressions faciales

Le visage est l'un des moyens les plus puissants de communication non verbale. L'expression faciale fournit des indices sur les émotions, les intentions, la nervosité, la douleur, la personnalité, régule le comportement interpersonnel, et communique le statut psychiatrique et biomédicale, entre autres fonctions. Au cours des dernières années, l'analyse automatisée des expressions faciales a bénéficié d'un intérêt croissant dans les communautés de la vision par ordinateur et de l'apprentissage automatique. [Torre et Cohn, 2011](#) Ekman a défini six expressions émotionnelles de base dans son article [Ekman et Friesen, 1971](#), le bonheur, la peur, la surprise, le dégoût, la tristesse et la colère.

2.2.1 La reconnaissance des expressions faciales

Les expressions faciales sont l'une des caractéristiques les plus importantes pour refléter l'état émotionnel de l'être humain car elles transmettent des informations utiles à l'observateur , avec le développement et les applications récentes des systèmes d'interaction homme-machine, la reconnaissance des expressions faciales (FER) est devenue un domaine de recherche populaire, c'est un processus réalisé par les humains ou les ordinateurs, et qui consiste à [PANTIC, 2008](#) :



FIGURE 2.1 – expressions faciales

1. Localiser les visages dans une image, cette étape est également appelée détection des visages
2. Extraction de caractéristiques faciales à partir de la région de visage détectée (tel que la détection de la forme des composants du visage ou la description de la texture de la peau dans une zone du visage .
3. Analyse du mouvement des traits du visage et/ou les changements dans l'apparence des traits du visage et classer ces informations dans des catégories d'interprétation de l'expression faciale, telles que des l'activation des muscles faciaux comme le sourire ou le froncement des sourcils, les catégories d'émotions comme le bonheur ou la colère, les catégories d'attitudes comme le (dés)amour ou l'ambivalence, etc. Cette étape est également appelée interprétation de l'expression faciale.

2.2.2 Les domaines d'application de la reconnaissance des expressions faciales

En réalité, la complexité des expressions faciales humaines rendent leur classification difficile, Cependant, on ne peut ignorer que la technologie de la reconnaissance des expressions faciales est un élément important de l'interaction homme-machine , Elle est utilisée dans

- la recherche sur l'interaction sociale
- le contrôle intelligent
- la médecine
- la communication

et d'autres domaines qui sont prometteurs, comme :

- la recherche psychologique dans le domaine des sciences sociales
- l'analyse du sentiment de l'utilisateur dans le marketing
- les jeux sociaux en ligne

ainsi que d'autres domaines qui incluent l'interaction homme-machine , qui peuvent tous bénéficier de la reconnaissance automatique des expressions faciales [Yan *et al.*, 2021].

2.2.3 Les difficultés de la reconnaissances d'expressions faciales

La reconnaissance des expressions faciales est un problème difficile pour les modèles d'apprentissage automatique et d'apprentissage profond existants, car les images peuvent

varier en termes de luminosité, d'arrière-plan, de pose, etc. Les méthodes d'apprentissage profond nécessitent également l'appui de données volumineuses. Elles ne sont pas performantes lorsque la base de données est petite, mal étiquetée ou déséquilibrée. Et à titre d'exemple la base de données des expressions faciales FER2013 elle contient un déséquilibre concernant la distribution de nombres d'expressions faciales (l'expression "dégoût" est représenté par 547 images tandis que l'expression "joie" qui est représenté par 8989 images).

2.3 L'augmentation des données

L'augmentation des données d'image est appliquée dans de nombreuses applications pour augmenter la taille et la diversité des données. Les réseaux neuronaux profonds ne sont pas faciles à entraîner en utilisant de petits ensembles de données d'entraînement en raison de problèmes de surajustement [Bartlett et al., 2020], l'augmentation des données de formation est l'une des solutions, qui vise à accroître la diversité d'un petit ensemble de données d'apprentissage [Shorten et Khoshgoftaar, 2019], les méthodes traditionnelles telles que la rotation, la réflexion, la translation, la mise à l'échelle, le recadrage, le flou, la mise à l'échelle des gris et la conversion des couleurs, ont été couramment utilisés pour l'augmentation des données d'image afin de réduire le surajustement lors de la formation de réseaux neuronaux profonds pour des applications de classification d'images. Ces techniques traditionnelles peuvent produire des images similaires de haute qualité, elles augmentent rarement la diversité des caractéristiques de l'image originale. Par conséquent, le développement de nouvelles méthodes d'augmentation des données d'image qui peuvent synthétiser des images non seulement diverses mais aussi photoréalistes à partir d'un petit ensemble de données d'image peut éviter les problèmes de sur-apprentissage et améliorer encore les performances du système. [Hung et Gan, 2021]

2.4 Les techniques d'augmentation des données

2.4.1 Les Transformations géométriques :

Il y a plusieurs techniques d'augmentations basées sur les transformations géométriques caractérisées par leur facilité de mise en œuvre. La compréhension de ces transformations fournira une base utile pour des recherches ultérieures sur les techniques d'augmentation des données :

- **Espace couleur :**

Les données d'image numérique sont généralement codées sous la forme d'un tenseur de la dimension (hauteur \times largeur \times canaux couleurs). L'exécution d'augmentations dans l'espace des canaux de couleur est une stratégie très pratique à mettre en œuvre. Des augmentations de couleur très simples consistent à isoler un seul canal de couleur tel que R, G ou B. En outre, les valeurs RGB peuvent être facilement manipulées avec de simples opérations matricielles pour augmenter ou diminuer la luminosité de l'image. Plus encore des augmentations de couleur plus avancées proviennent de la dérivation des histogrammes de couleur décrivant l'image, et la modification des valeurs d'intensité de ces histogrammes permet de modifier l'éclairage [Shorten et Khoshgoftaar, 2019].



FIGURE 2.2 – Augmentation de luminosité

- **Reffet :**

C'est une technique d'augmentation de données qui crée une image miroir à partir de l'image originale dans le sens vertical ou horizontal, Le reflet (en anglais flip) de l'axe horizontal est beaucoup plus courant que le flip de l'axe vertical, Cette technique d'augmentation est l'une des plus faciles à mettre en œuvre et s'est avérée utile sur les ensembles de données [Shorten et Khoshgoftaar, 2019].

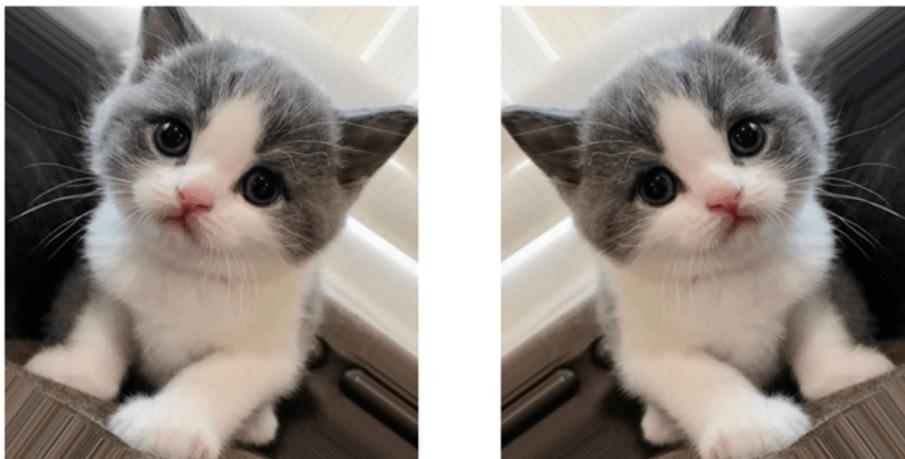


FIGURE 2.3 – Reflet horizontal

- **Recadrage :** C'est une technique d'augmentation des données qui échantillonne de manière aléatoire une section de l'image originale. La taille de l'image recadrée est suffisamment grande pour contenir une partie pertinente de l'image.
- **Injection de bruit :** L'injection de bruit consiste à injecter une matrice de valeurs aléatoires généralement tirées d'une distribution gaussienne. L'injection de bruit est testée par Moreno-Barea et al [Moreno-Barea *et al.*, 2018] sur neuf jeux de données provenant du dépôt UCI . L'ajout de bruit aux images peut aider les CNN à apprendre des caractéristiques plus robustes.
- **Translation :** C'est une transformation qui effectue un déplacement aléatoire de l'image originale dans le sens horizontal ou vertical (ou les deux). [Porcu *et al.*, 2020]



FIGURE 2.4 – translation d'une image

- **Rotation aléatoire** : cette technique applique une rotation aléatoire entre -90° et $+90^\circ$ à l'image [Porcu *et al.*, 2020].



FIGURE 2.5 – rotation d'une image

2.4.2 Les techniques basés sur l'apprentissage profond (le deep learning)

Les GANs : La classification d'images avec de multiples étiquettes de classe en utilisant seulement un petit nombre d'exemples étiquetés est une tâche difficile, multiple surtout lorsque la distribution des étiquettes (classes) est déséquilibrée. Les réseaux adversariaux génératifs (GAN) peuvent être utilisés pour générer des images et créer de nouvelles données avec la même distribution des données d'apprentissage [Zhu *et al.*, 2018], ce qui a permis de réaliser l'augmentation des données .

CycleGAN Il peut faire une transition d'image à image entre deux domaines d'images non appariées, ce qui est utile pour construire le modèle FER , L'idée principale du cycleGAN est la suivante : "Si nous traduisons d'un domaine à un autre et inversement, nous devons arriver là où nous avons commencé" [Zhu *et al.*, 2017].

2.5 Bases de données d'expressions faciales :

La détection, l'interprétation et le traitement des comportements émotionnels se situent au centre de la recherche sur les interfaces naturelles entre l'homme et la machine,

les bases de données sont les outils essentiels pour le développement et l'évaluation des systèmes de reconnaissance des expressions faciales. Plusieurs bases de données d'expressions faciales ont été créées au cours des dix dernières années. Malgré cela, il est toujours nécessaire de disposer de données émotionnelles adéquates pour de nombreux sujets, afin d'améliorer la qualité de la recherche afin d'évaluer statistiquement les algorithmes de reconnaissance, nous présentons quelques bases de données pour la reconnaissance des expressions faciales

1. **FER2013** : la base de données FER2013 (Facial Expression Recognition 2013) qui a été créé par Pierre Luc Carrier et Aaron Courville, fait partie d'un projet plus vaste en cours. Le jeu de données a été créé en utilisant une API de recherche d'images de Google pour rechercher des images de visages qui correspondent à un ensemble de 184 mots-clés liés aux émotions, comme "béat", "enragé", etc. Ces mots-clés ont été combinés avec des mots liés au sexe, à l'âge ou à l'origine ethnique, pour obtenir près de 600 chaînes de caractères qui ont été utilisées comme requêtes de recherche d'images de visages , le jeu de données résultant contient 35887 images distribué comme suit : [Goodfellow et al., 2013](#)

type d'expression	Colère	Dégoût	Peur	Joie	Tristesse	Surprise	Neutre
nombre d'images	4953	547	5121	8989	6077	4002	6198

TABLE 2.1 – FER2013

2. **JAFFE** : Ce jeu de données d'images a été planifié et assemblé par Michael Lyons, Miyuki Kamachi et Jiro Gyoba, à l'université de Kyushu, au Japon, l'ensemble de données JAFFE est composé de 213 images d'expressions faciales différentes provenant de 10 sujets féminins japonais différents. Chaque sujet a fait 7 expressions faciales (6 expressions faciales de base et neutres) et les images ont été annotées avec des notes sémantiques moyennes sur chaque expression faciale par 60 annotateurs. [Lyons et al., 1999](#)

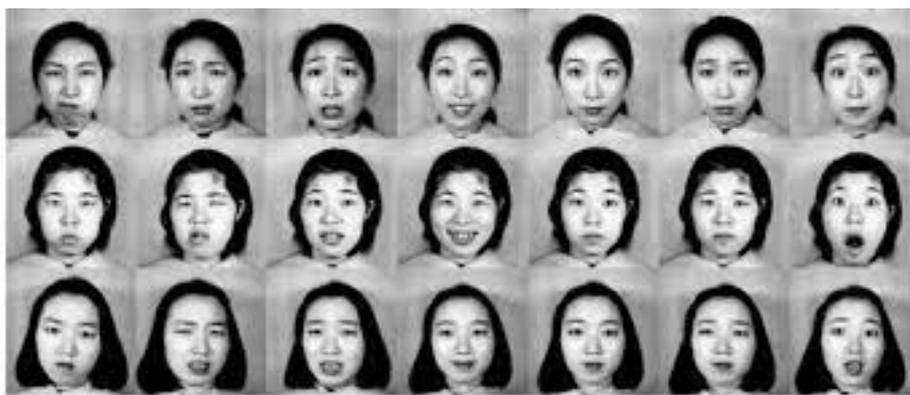


FIGURE 2.6 – JAFFE dataset

3. **CK+** : L'ensemble de données Cohn-Kanade étendu (CK+) contient 593 séquences vidéo provenant d'un total de 123 sujets différents, âgés de 18 à 50 ans, de sexe et d'origine variés. Chaque vidéo montre un changement facial de l'expression neutre à une expression de pointe ciblée, enregistrée à 30 images par seconde (FPS) avec une résolution de 640x490 ou 640x480 pixels. Parmi ces vidéos, 327 sont étiquetées

avec l'une des sept classes d'expression suivantes : colère, mépris, dégoût, peur, joie, tristesse et surprise. La base de données CK+ est largement considérée comme la base de données de classification des expressions faciales contrôlée en laboratoire la plus utilisée, et elle est utilisée dans la majorité des méthodes de classification des expressions faciales. [Lucey et al., 2010]

4. **AffectNet** : C'est une base de données des expressions faciales dans la nature, en collectant et en annotant des images faciales. AffectNet contient plus d'un million d'images faciales collectées sur Internet en interrogeant trois grands moteurs de recherche à l'aide de 1250 mots-clés liés aux émotions dans six langues différentes. Environ la moitié des images récupérées (440K) ont été annotées manuellement pour la présence de sept expressions faciales discrètes (modèle catégoriel) et l'intensité de la valence et de l'excitation (modèle dimensionnel). AffectNet est de loin la plus grande base de données d'expressions faciales, de valence et d'excitation dans la nature, ce qui permet d'effectuer des recherches sur la reconnaissance automatique des expressions faciales [Mollahosseini et al., 2017]

Il existe dans la littérature d'autre base de données utilisés pour la reconnaissance des expressions faciales , nous résumons les plus importants dans le tableau 2.2

nom	échantillons	Condit	sujet	MC
FER2013 [Goodfellow et al., 2013]	35887 images	Web	N/d	P + S
AFEW 7.0 [Lee et al., 2019]	1809 vidéos	film	N/d	P + S
JAFFE [Lyons et al., 1999]	213 images	Lab	10	P
AffectNet [Mollahosseini et al., 2017]	450.000 images (étiquetées)	Web	N/d	P + S
KDEF [Lundqvist et al., 1998]	4900 images	Lab	70	p
MMI [Pantic et al., 2005]	740 images et 2900 vidéos	Lab	25	P
Oulu-Casia [Zhao et al., 2011]	2,880 séquences d'images	Lab	80	p
SFEW 2.0 [Dhall et al., 2015]	1766 images	film	N/d	P + S

TABLE 2.2 – Tableau récapitulatif sur les bases de données d'expression faciales

- condit = condition de collection, MC= Méthode de collection , p=posé , S=spontané , N/D = non disponible

2.6 Quelques travaux connexe sur l'augmentation des données pour la reconnaissance des expressions faciales

Nous élaborons dans ce contexte, quelques travaux sur l'augmentation des données dans le domaine de la reconnaissance des expressions faciales , pour cela nous présentons huit articles récents de la littérature scientifique pour mieux le comprendre.

2.6.1 Article 1 "Evaluation of Data Augmentation Techniques for Facial Expression Recognition Systems" [Porcu et al., 2020]

Dans cette recherche les auteurs ont évalué et comparé l'impact de l'utilisation de techniques d'augmentation de données bien établies sur la précision de la reconnaissances

des émotions d'un système de reconnaissance des expressions faciales

Le système FER proposé basé sur le célèbre réseau neuronal convolutif VGG16 (CNN). les auteurs ont utilisé à la fois les transformations géométriques et le GAN pour augmenter la quantité d'images d'entraînement et ont réalisé des évaluations croisées de bases de données, ils ont fait la formation (l'entraînement) du système avec la base de données 'augmenté' KDEF et le test avec deux bases de données (CK+ , EXPW) la précision du système sans l'augmentation des données été 53% . Les meilleurs résultats ont été obtenus en combinant la réflexion horizontale, la translation et le GAN, ce qui a permis d'augmenter la précision du système proposé de 30%.

Les résultats démontrent que les transformations géométriques d'images, telles que la Réflexion horizontale et la translation, apportent des améliorations de performance limitées ,en revanche l'adoption du GAN enrichit la base de données d'entraînement avec des images nouvelles et utiles qui permettent des améliorations de précision assez significatives.

2.6.2 Article 2 "Facial Expression Recognition with Convolutional Neural Networks via a Data Augmentation Strategy"

[Yan *et al.*, 2021]

Ce travail vise à construire un système FER profond disponible et efficace sur la base d'un réseau neuronal convolutif (CNN)

Les auteurs utilisent le model VGG-16 pour les tâches de reconnaissance des expressions faciales et ils mettent en œuvre un réseau résiduel (ResNET) pour former un réseau plus profond, en utilisant différents types des GAN's (CGAN,DCGAN,WGAN) pour augmenter l'ensemble de données afin de résoudre le problème de sur-apprentissage (Over-Fitting). Les résultats expérimentaux considérables ont été obtenus en utilisant la base de données des expressions faciales FER-2013

Avant d'utiliser les GANs pour l'augmentation des données, ResNET-50 peut atteindre une précision de 64,28% , concernant la génération d'images de visage, WGAN et DCGAN peuvent produire des images de meilleure qualité que CGAN. Parmi eux, la qualité d'image de WGAN est meilleure que celle de DCGAN, mais la précision de l'algorithme du modèle ResNET-50+DCGAN dans les résultats est meilleure que celle de WGAN qui est élevée à 71.75%.

Donc la génération d'images de visage pour augmenter les données est utile pour améliorer la performance de VGG et ResNET sur la tâche FER.

2.6.3 Article 3 "Facial expression recognition with trade-offs between data augmentation and deep learning features"

[Umer *et al.*, 2022]

L'objectif de cet article est de reconnaître les types d'expressions dans la région du visage humain . L'implémentations de ce système a été divisé en quatre étapes : la détection du visage à partir de l'image d'entrée capturé puis la proposition d'une architecture de réseau

neuronal convolutif pour extraire des caractéristiques plus spécifiques et discriminantes et pour améliorer les performances du système proposé , des techniques d'augmentation des données ont été appliquées sur les images faciales .

Un compromis entre l'augmentation des données et les caractéristiques d'apprentissage profond a été effectué pour affiner le modèle CNN formé.

Des résultats expérimentaux approfondis ont été démontrés en utilisant trois bases de données de visages de référence : KDEF, GENKI-4k et CK+. Chaque base de données est partitionnée aléatoirement avec 50% des données dans l'ensemble d'entraînement et 50% dans l'ensemble de test. Après avoir utilisé différentes tailles d'images en entrée, les auteurs ont trouvé que la meilleure taille d'image en entrée en termes de précision pour chaque ensemble de données est la taille 128*128 : 82.79% avec KDEF , 94.33% avec GENKI , 97.69% avec CK+ les performances du système propose ont été les meilleurs par a port au travaux connexe mentionné dans l'article : (97.69% avec CK+ ,94.67% avec GENKI-4k ,83.43% avec KDEF)

2.6.4 Article 4 "Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy" [\[Li et al., 2020\]](#)

Ce travail vise a utiliser un classificateur unique basé sur un CNN , en utilisant un pré-traitement aux images pour faire face au problème de manque de données par l'augmentation de données qui sont des transformations géométriques spécifiques (recadrage et rotation des visages). Pendant la phase d'entraînement, une rotation aléatoire et un retournement horizontal ont été effectués pour augmenter la taille de la base de données. Les données de formation étendues ont été utilisées pour former le CNN. Deux bases de données largement répandues ont été utilisées dans les expériences : la base de données Cohn-Kanade étendue (CK+) et la base de données japonaise japonaise Female Facial Expressions (JAFFE), des précisions moyennes de 97,38 % et 97,18 % ont été obtenues pour des expériences à 7 classes sur les bases de données CK+ et JAFFE, respectivement.

2.6.5 Article 5 "A Facial Expression Recognition Method Using Deep Convolutional Neural Networks Based on Edge Computing" [\[Chen et al., 2020\]](#)

Cet article traite le problème de la reconnaissance des expressions faciales en se concentrant sur l'amélioration des données d'une séquence vidéo. Dans ce processus l'image d'expression neutre est utilisée comme domaine source, et d'autres images d'expression comme domaine cible. En vue de résoudre le problème du déséquilibre des classes dans la base de données d'expressions faciales existante, l'article améliore le Cycle GAN, par la proposition d'une méthode de reconnaissance des expressions faciales basée sur la génération cohérente de cycle de contrainte pour résister au réseau, et introduit une condition de contrainte de classe et une règle de pénalité de gradient. Les résultats expérimentaux montrent que le modèle de génération amélioré peut mieux apprendre les informations

détaillées sur la texture de l'image du visage, et la qualité de l'image générée est élevée. Le réseau discriminant amélioré a un meilleur effet de classification . Les expériences sont réalisées sur le jeu de données JAFFE,CK+ et FER2013 le taux de reconnaissance avec la méthode proposé été :

- 98.46% (un gain de 1.23%) avec la base de données CK+
- 76.96% (un gain de 4.01%) avec la base de données JAFFE
- 78.61 % (un gain de 4.90%) avec la base données FER2013

2.6.6 Article 6 "Real Time Face Expression Recognition along with Balanced FER2013 Dataset using CycleGAN" [Mazen et al.,]

cette recherche présente un modèle de réseau neuronal profond pour la reconnaissance des sept principaux types d'expression faciale. La méthodologie proposée comporte quatre étapes : le prétraitement de l'ensemble de données FER2013 par le biais d'un ré-étiquetage afin d'éviter les résultats inappropriés et de se dégager des expressions non faciales , conception d'un réseau adversarial générateur de cycles (CycleGAN) stable et efficace,qui fournit une translation non supervisée d'expression à expression. Le CycleGAN a été conçu et entraîné avec une nouvelle perte de cohérence de cycle. ensuite les chercheurs ont Générer de nouvelles images pour surmonter le déséquilibre de classe et Enfin,ils ont fait la construction de l'architecture DNN (Deep Neural Network) pour la reconnaissance de l'expression du visage et du signe, en utilisant les images pré-entraînées. le taux de reconnaissance des classes s'est remarquablement amélioré en utilisant le jeu de données FER2013 équilibré,Le modèle conçu a permis d'obtenir une performance optimale et un taux de 91,76 % en seulement 0,44 seconde.

2.6.7 Article 7 "Data Augmentation Merging Dataset for Facial Emotion Recognition" [Kim et Han, 2020]

Dans ce travail , les chercheurs montrent que le fait de ne tester qu'un seul jeu de données, comme FER 2013 ou CK+, a parfois limité l'amélioration de la performance des systèmes FER, même après l'augmentation des données . Ils ont proposé de fusionner des jeux de données comme une autre méthode pour améliorer la performance de système. Autrement que l'utilisation de la technique d'augmentation des données ,en fusionnant différents ensembles de données pour augmenter le nombre d'images faciales d'entraînement

la performance de système a été amélioré de 15,33% , la précision de validation (76.32%)

2.6.8 Article 8 "Small facial image dataset augmentation using conditional GANs based on incomplete edge feature input" [Hung et Gan, 2021]

Cet article propose un GAN conditionnel pour l'augmentation des images faciales en utilisant un très petit ensemble de données d'entraînement et des formes incomplètes ou modifiées. La méthode proposée définit un nouveau domaine ou espace pour le raffinement des images provisoires afin d'éviter le surajustement causé par l'utilisation d'un petit ensemble de données d'entraînement et d'améliorer la tolérance aux distorsions causées par des caractéristiques de bord incomplètes, ceci améliore efficacement la qualité de l'augmentation de l'image faciale avec la diversité, pour la synthèse d'images faciales à partir d'un très petit ensemble de données d'entraînement, Ils ont expérimenté les méthodes suivantes :

- Élargissement la diversité du domaine des sources
- Double translation :
- Réutilisation de l'information conditionnelle
- Utilisation de GAN conditionnel
- Prétraitement et affinage des images
- Extraction des bords

Les résultats expérimentaux ont montré que la méthode proposée peut générer des images de haute qualité avec une bonne diversité lorsque les GANs sont entraînés à l'aide de bords très épars et d'un petit nombre d'échantillons d'entraînement, par rapport à de l'état de l'art de cet article.

2.7 conclusion

La reconnaissance des expressions faciales vise à extraire les émotions des êtres humains à partir de l'image de leur visage ce qui permet d'extraire des informations très utiles , mais parmi les problèmes rencontrés dans la FER le problème de sur-apprentissage . L'augmentation des données permet d'éviter ce problème et de généraliser l'ensemble de données, ce qui permet d'améliorer les performances du système de reconnaissance des expressions faciales.

Chapitre 3

Conception

3.1 Introduction

De nos jours, les modèles basés sur l'apprentissage profond nécessitent l'utilisation de large base de données pour obtenir des résultats performants. La reconnaissance des expressions faciales est une application basée sur l'apprentissage profond, mais malheureusement la plupart des ensembles de données de visage souffrent de limitation dans le nombre d'échantillons, et de chevauchement important entre les différentes classes d'émotions. Ce manque et ce déséquilibre dans les données, provoque généralement un sur-apprentissage. Grâce aux techniques d'augmentation des données, le jeu de données devient de plus en plus riche en termes de variété et de diversité des données, ce qui permet aux systèmes d'apprentissage de faire leur entraînement avec efficacité sur la reconnaissance des expressions faciales .

3.2 Objectif

L'objectif principal de notre travail est de concevoir et réaliser un système de reconnaissance des expressions faciales basé sur l'apprentissage profond, deux modèles ont été sélectionnés, le VGG-16 qui est l'un des modèles les plus utilisés en apprentissage profond, pour sa simplicité et son efficacité, tandis que le ResNet50 est l'un des modèles les plus puissants en classification.

Dans notre projet, nous visons à évaluer et à comparer l'effet de l'utilisation de différentes techniques d'augmentation des données sur la performance des systèmes de reconnaissance des expressions faciales (FER) basés sur l'apprentissage profond.

3.3 Architecture du système

Le système que nous proposons permet de reconnaître l'expression faciale du visage présent dans une image d'entrée. Dans notre travail, nous utilisons deux modèles de base, ResNet50 et VGG-16, avec et sans augmentation des données. Les augmentations utilisées sont : les transformations géométriques, la technique SMOTE et Cycle-Gan. Comme le montre la figure 3.1, l'entrée de notre système est une image (contenant un visage) après une étape de détection des visages et de prétraitement (un redimensionnement) et le choix d'un des deux modèles, une des techniques d'augmentation des données est sélectionnée pour permettre la prédiction des expressions (parmi les sept expressions faciales de base).

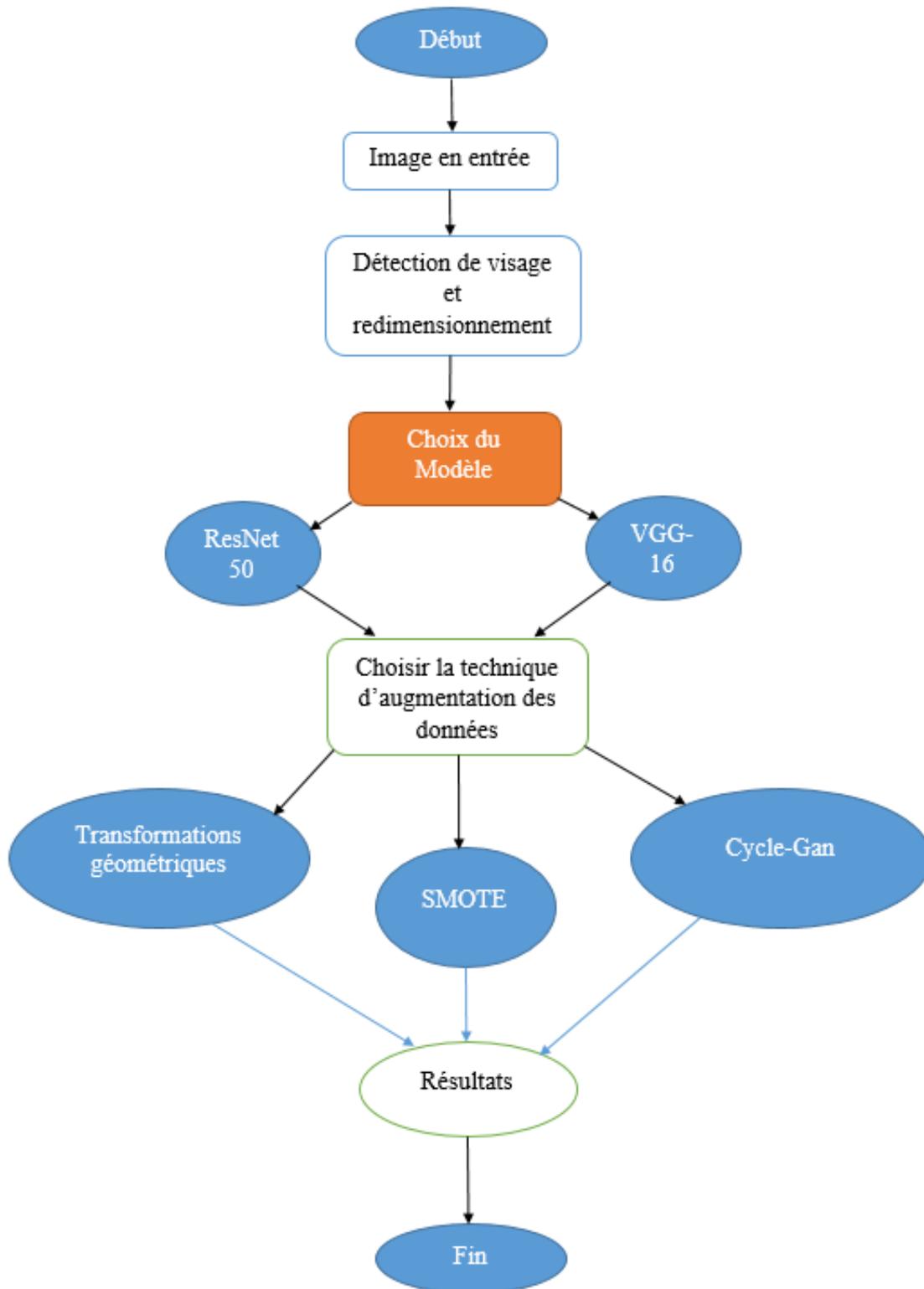


FIGURE 3.1 – L'architecture de notre système

Dans ce qui suit, nous détaillons chacune de ces étapes :

3.3.1 Détection de visage

C'est la première phase de notre système, elle consiste à détecter le visage humain dans l'image d'entrée afin de redimensionner l'image d'entrée en 48×48 , qui est la taille des entrées de nos modèles.

3.3.2 Choix du modèle

Comme nous l'avons déjà mentionné, deux modèles ont été utilisés pour atteindre notre objectif de comparer les trois techniques d'augmentation des données et de choisir la plus appropriée pour les systèmes FER basés sur l'apprentissage profond.

A) Le modèle VGG-16

Il s'agit d'un modèle de réseau neuronal convolutif (CNN) proposé par Karen Simonyan et Andrew Zisserman à l'université d'Oxford [Simonyan et Zisserman, 2014], le VGG16 a apporté des améliorations par rapport au CNN AlexNet en remplaçant les grands filtres de la taille de noyau (11 et 5 dans la première et la deuxième couche convolutive, respectivement) par de multiples filtres de taille noyau 3×3 , l'un après l'autre. L'architecture du VGG16 est simple et profonde, elle se compose de 16 couches : 13 couches convolutionnelles et 3 couches entièrement connectées, toutes les couches cachées sont équipées de la non-linéarité de rectification (ReLU). L'architecture est résumée dans la figure 3.2. Ce réseau avait 138M de paramètres et utilisait environ 500MB d'espace de stockage, ce qui le ralentissait péniblement lors de l'entraînement

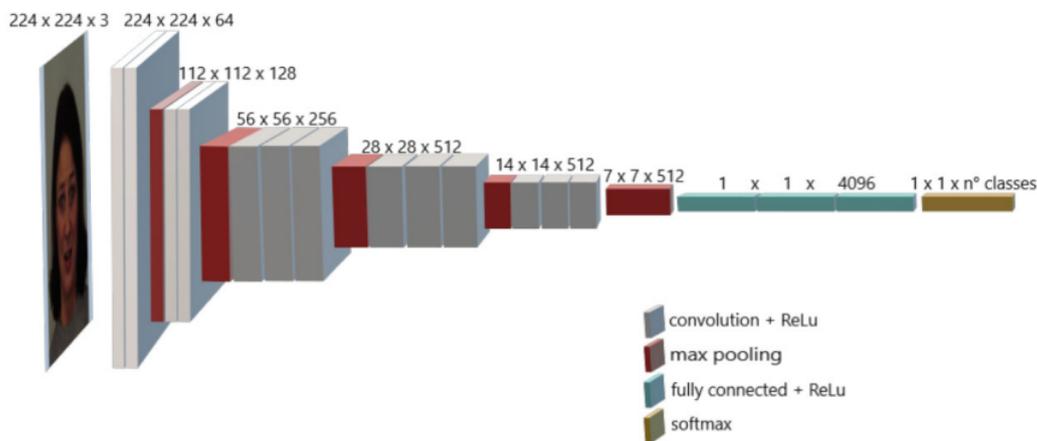


FIGURE 3.2 – Architecture de VGG16 [Porcu *et al.*, 2020]

B) Le modèle ResNet50

Il s'agit d'un réseau innovant qui a été introduit pour la première fois par Kaiming He et Xiangyu Zhang, Shaoqing Ren et Jian Sun Dans leur article de recherche sur la vision par ordinateur de 2015 [He *et al.*, 2016]

- ResNet possède de nombreuses variantes qui fonctionnent sur le même concept mais avec un nombre de couches différent, Resnet50 est utilisé pour désigner la variante qui peut fonctionner avec 50 couches de réseau neuronal.
- Lorsque l'on utilise des réseaux neuronaux convolutifs profonds pour résoudre un problème lié à la vision par ordinateur, les experts en apprentissage automatique empilent des couches supplémentaires. Ces couches supplémentaires

aident à résoudre des problèmes complexes plus efficacement, car les différentes couches peuvent être entraînées pour des tâches différentes afin d'obtenir des résultats très précis.

- Si le nombre de couches empilées peut enrichir les caractéristiques du modèle, un réseau plus profond peut présenter un problème de dégradation. En d'autres termes, à mesure que le nombre de couches du réseau neuronal augmente, les niveaux de précision peuvent être saturés et se dégrader lentement après un certain temps. Par conséquent, les performances du modèle se détériorent à la fois sur les données de formation et de test.
- Cette dégradation n'est pas le résultat d'un surajustement. Elle peut plutôt résulter de l'initialisation du réseau, de la fonction d'optimisation ou, plus important encore, du problème des gradients qui disparaissent ou explosent. [W5]

3.3.3 Choix de la technique d'augmentation des données

Trois techniques d'augmentation ont été utilisées, à savoir : les transformations géométriques, SMOTE et Cycle-Gan :

A) Les transformations géométriques

Nous pouvons appliquer n'importe quelle transformation aléatoire sur chaque image d'entraînement lorsqu'elle est transmise au modèle, cela rendra notre modèle plus robuste. L'augmentation de l'image est une technique qui consiste à appliquer différentes transformations aux images originales, ce qui permet d'obtenir plusieurs copies transformées de la même image. Chaque copie est différente de l'autre sous certains aspects, en fonction des techniques d'augmentation que nous appliquons, comme le décalage, la rotation, le retournement, etc. Les transformations géométriques que nous avons appliquées sur notre jeu de données sont les suivantes :

- La rotation aléatoire
- Le cisaillement
- Le zoom
- Le reflet horizontal
- Le décalage en largeur
- Le décalage en hauteur

B) SMOTE

Le déséquilibre dans les bases de données est un problème qui conduit souvent à un sur-apprentissage dans les systèmes FER. Donc, la nature de l'application exige un taux de détection correcte assez élevé dans la classe minoritaire et permet un faible taux d'erreur dans la classe majoritaire. Pour cela une Technique de sur-échantillonnage des minorités synthétiques (en anglais dit Synthetic Minority Over-sampling TEchnique) SMOTE a été utilisée. Cette technique consiste en plusieurs étapes :

- La classe minoritaire est sur-échantillonnée en prenant chaque échantillon de la classe minoritaire et en introduisant des exemples synthétiques le long de segments reliant tous les k voisins les plus proches de la classe minoritaire.
- En fonction de sur-échantillonnage requis, les voisins parmi les k plus proches sont choisis aléatoirement.

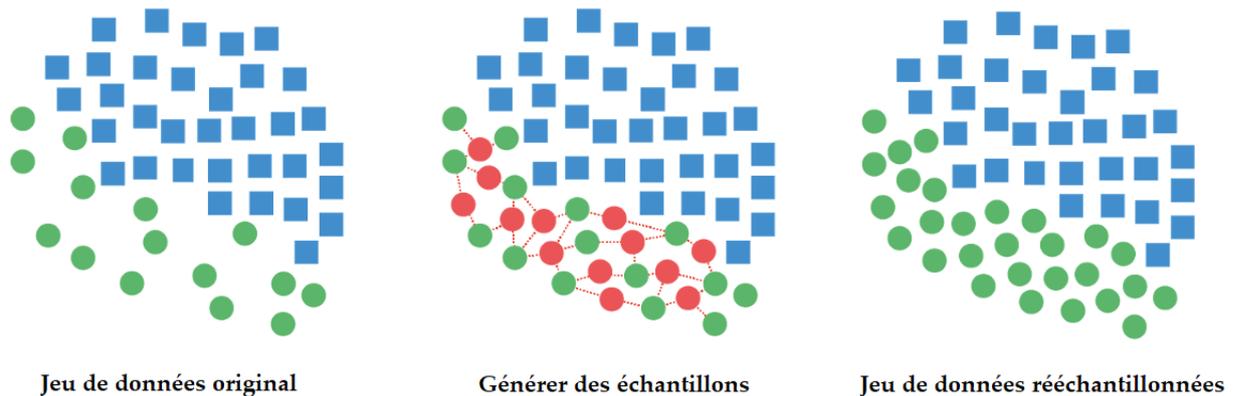


FIGURE 3.3 – La technique SMOTE

Les échantillons synthétiques sont générés de la manière suivante :

- Prendre la différence entre le vecteur de caractéristiques (échantillon) en considération et son plus proche voisin.
- Multiplier cette différence par un nombre aléatoire entre 0 et 1, et l'ajouter au vecteur de caractéristiques considéré.

Cela provoque la sélection d'un point aléatoire le long du segment de ligne entre deux caractéristiques spécifiques. Cette approche force effectivement la région de décision de la classe minoritaire à devenir plus générale. [Chawla *et al.*, 2002]

C) Cycle-Gan

La deuxième technique généralement utilisée pour traiter une base de données déséquilibrées est les réseaux adversaires génératifs, le modèle Gan le plus performant pour l'augmentation des données est le Cycle-Gan qui est entraîné pour générer plus d'images dans les classes les moins nombreuses dans la base de données utilise, afin de rendre notre système de reconnaissance des expressions faciales plus fiable et efficace. La figure 3.4 représente une architecture simplifiée de Cycle-Gan.

Le fonctionnement du cycle gan peut être résumé par les points suivants :

- Le CycleGAN est une technique qui consiste à entraîner automatiquement des modèles de traduction image à image sans exemples appariés.
- La traduction d'image à image consiste à générer une nouvelle version synthétique d'une image donnée avec une modification spécifique, comme la traduction d'un paysage d'été en hiver, l'apprentissage d'un modèle de traduction d'image à image nécessite généralement un grand ensemble de données d'exemples appariés. Ces ensembles de données peuvent être difficiles et coûteux à préparer, et dans certains cas impossibles, comme les photographies de peintures d'artistes décédés depuis longtemps.
- La cohérence du cycle est un concept de la traduction automatique selon lequel une phrase traduite de l'anglais au français doit être traduite du français à l'anglais et être identique à la phrase originale. Le processus inverse doit aussi être vrai.

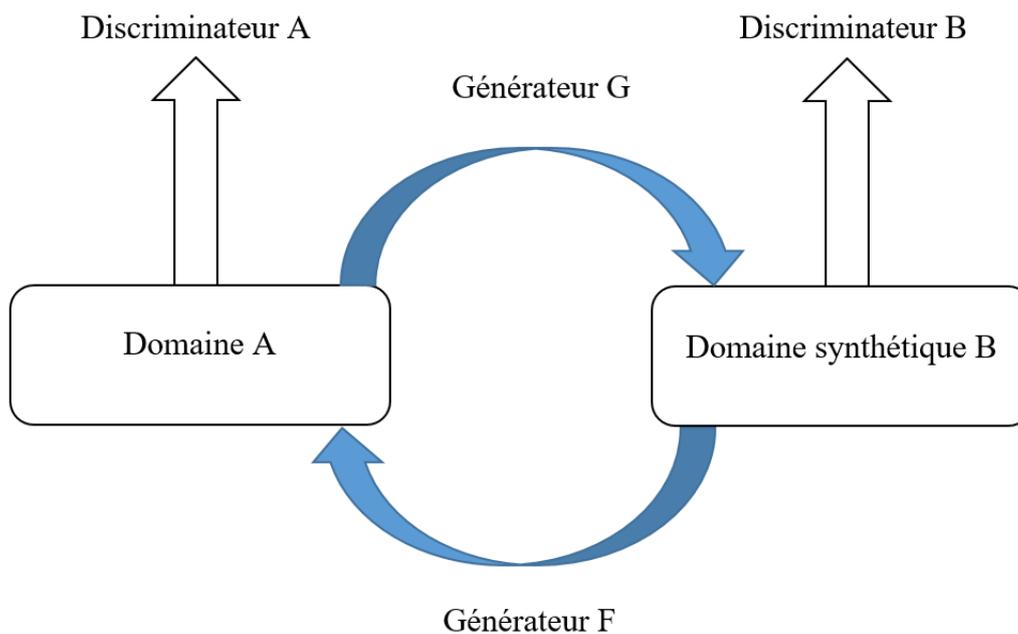


FIGURE 3.4 – Cycle gan

3.4 Conclusion

Dans le but de choisir la méthode d'augmentation des données la plus appropriée aux systèmes FER basés sur l'apprentissage profond, nous avons conçu une application de reconnaissance des expressions faciales basée sur deux modèles performants en classification, qui sont VGG-16 et ResNet 50. Nous avons prévu d'utiliser 3 techniques d'augmentation des données qui sont : les transformations géométriques, SMOTE et Cycle-Gan, plusieurs expérimentations ont été faites, elles seront détaillées dans le chapitre suivant.

Chapitre 4

Implémentation

4.1 Introduction

Après avoir expliqué dans le chapitre précédent la conception détaillée de notre application nous allons présenter dans ce chapitre l'implémentation de notre système, ainsi que l'environnement de développement avec les bibliothèques utilisées et les résultats de nos modèles. La deuxième partie de ce chapitre est consacrée à la présentation des tests des modèles et des résultats obtenus ainsi que leur interprétation.

4.2 Environnement

4.2.1 Kaggle

Kaggle est une plateforme web qui accueille la plus grande communauté de Data Science au monde, avec plus de 536 000 membres actifs dans 194 pays et reçoit près de 150 000 soumissions par mois, et qui lui fournit des outils et des ressources puissants pour aider à atteindre tous les progrès de science des données. Kaggle, offre un environnement Jupyter Notebooks personnalisable et sans configuration. Sont accessibles gratuitement des GPU et une grande quantité de données et de codes publiés par la communauté. À l'intérieur de Kaggle, vous trouverez tout le code et les données dont vous avez besoin pour réaliser vos projets de science des données. Il y a plus de 50 000 jeux de données publics et 400 000 notebooks publics disponibles pour tous.[W6]

En résumé, sur Kaggle un utilisateur peut :

- Trouver et publier des bases de données
- Explorer et construire des modèles sur un environnement web adapté
- Travailler avec d'autres professionnels et passionnés
- Faire des compétitions pour se challenger sur des sujets innovants

4.2.2 Google Colaboratory

Colaboratory, souvent raccourci en "Colab", est un produit de Google Research. Colab permet à n'importe qui d'écrire et d'exécuter le code Python de son choix par le biais du navigateur. C'est un environnement particulièrement adapté au machine learning, à l'analyse de données et à l'éducation [W7]. En termes plus techniques, Colab permet :

- D'améliorer les compétences de codage en langage de programmation Python.

- De développer des applications en Deep Learning en utilisant des bibliothèques Python populaires telles que Keras, TensorFlow, PyTorch et OpenCV.
- D'utiliser un environnement de développement (Jupyter Notebook) qui ne nécessite aucune configuration, Mais la fonctionnalité qui distingue Colab des autres services est l'accès à un processeur graphique GPU, totalement gratuit.

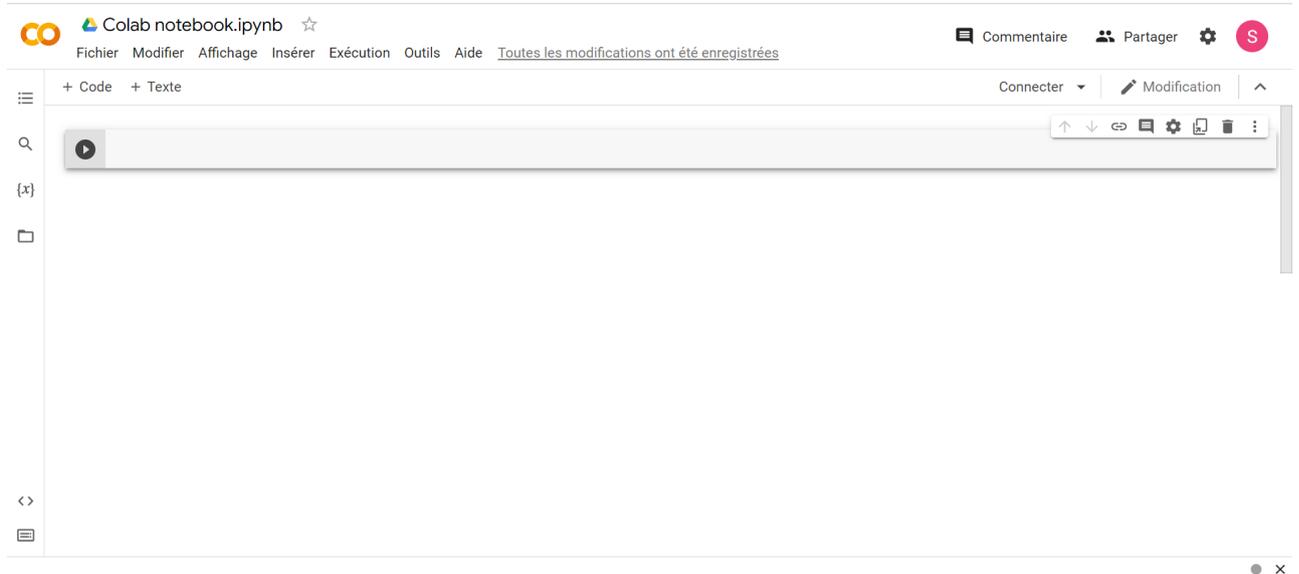


FIGURE 4.1 – Interface de Colab

4.2.3 Apprentissage sur GPU :

Pour configurer le GPU (processeur graphique), il suffit de cliquer sur Exécution > Modifier le type d'exécution et sélectionner GPU comme accélérateur matériel.

Paramètres du notebook

Accélérateur matériel

GPU

Pour tirer le meilleur parti de Colab, évitez d'utiliser un GPU si vous n'en avez pas besoin. [En savoir plus](#)

Exécution en arrière-plan

Vous souhaitez que votre notebook continue de fonctionner même après que vous avez fermé votre navigateur ? [Passer à Colab Pro+](#)

Omettre l'élément de sortie des cellules de code lors de l'enregistrement de ce notebook

Annuler [Enregistrer](#)

FIGURE 4.2 – Activation de GPU

4.3 Langage de programmation et bibliothèque utilisée

4.3.1 Python :

Python est un langage de programmation de haut niveau à usage général largement utilisé. Sa philosophie de conception met en valeur la lisibilité du code. Python prend en charge plusieurs paradigmes de programmation, notamment la programmation orientée objet, impérative et fonctionnelle ou les styles procéduraux. Il dispose d'un système de types dynamiques et une gestion automatique de la mémoire. bibliothèque standard vaste et complète. Python dispose en plus d'une poignée de bibliothèques open source importantes et complètes. à savoir Numpy/Scipy pour les opérations numériques, Cython pour l'optimisation de bas niveau, IPython pour le travail interactif et matplotlib pour le traçage. [Ari et Ustazhanov, 2014]

4.3.2 Bibliothèque utilisée :

4.3.3 Tensorflow

TensorFlow est une bibliothèque open source de Machine Learning , il s'agit d'une boîte à outils permettant de résoudre des problèmes mathématiques extrêmement complexes, Elle permet aux chercheurs de développer des architectures d'apprentissage expérimentales et de les transformer en logiciels, créée par Google ,TensorFlow regroupe un grand nombre de modèles et d'algorithmes de Machine Learning et de Deep Learning. Son API front-end de développement d'applications repose sur le langage de programmation Python . [W8]

4.3.4 Keras

Keras est une bibliothèque open source de prototypage rapide de modèles de deep learning Ecrit en Python, A la portée des débutants en IA, elle s'articule autour d'une API de haut niveau supportant différentes librairies de réseaux de neurones artificiels récurrents,L'objectif de Keras est d'offrir un cadre pour développer au plus vite des réseaux de neurones artificiels. Initiée en 2015, cette technologie repose sur le travail de François Chollet, un développeur de Google , Nous avons utilisé cette bibliothèque pour rendre les projets de deep learning accessibles, Son API se veut relativement simple à prendre en main. Elle est conçue pour réduire au maximum le nombre d'actions à réaliser lors de la création d'un réseau de neurones artificiel . [W9]

4.3.5 Numpy

NumPy, qui signifie Numerical Python, est une bibliothèque composée d'objets de tableaux multidimensionnels et d'un ensemble de fonctions permettant de traiter ces tableaux.[W10] Nous avons utilisé cette bibliothèque pour effectuer les opérations suivantes :

- Opérations mathématiques et logiques sur des tableaux.
- Transformations de Fourier et routines de manipulation de formes.
- Opérations liées à l'algèbre linéaire. NumPy possède des fonctions intégrées pour l'algèbre linéaire et la génération de nombres aléatoires.

4.3.6 Matplotlib

Nous avons utilisé cette bibliothèque pour faire le traçage graphique 2D et 3D matplotlib est un paquet du langage de programmation Python, qui produit des figures de qualité de publication scientifique dans une variété de formats papier et environnements interactifs sur toutes les plateformes. matplotlib est conçu avec la philosophie que l'utilisateur doit pouvoir créer des graphiques simples avec seulement quelques commandes, ou juste une! matplotlib essaie de rendre les choses faciles faciles et les choses difficiles possibles [Ari et Ustazhanov, 2014].

4.3.7 Tkinter

Nous avons utilisé le module Tkinter (Tk interface), qui est un module intégré à la bibliothèque standard de Python, permettant de créer des interfaces graphiques rapidement et facilement comme :

- Des fenêtres.
- Des widgets (boutons, zones de texte cases à cocher, ...).
- Des événements (clavier, souris, ...).

Tkinter est disponible sur Windows et la plupart des systèmes Unix , les interfaces créées avec Tkinter sont donc portables [Grayson, 2000].

4.4 Base d'apprentissage

Dans la phase d'apprentissage nous avons utilisé la base de données FER2013 qui est un ensemble de 35887 images de 7 émotions différentes, Dans ce jeu de données, toutes les images ont la dimension de 48×48 , et sont divisées en jeu de formation (80%) et de test (20%).

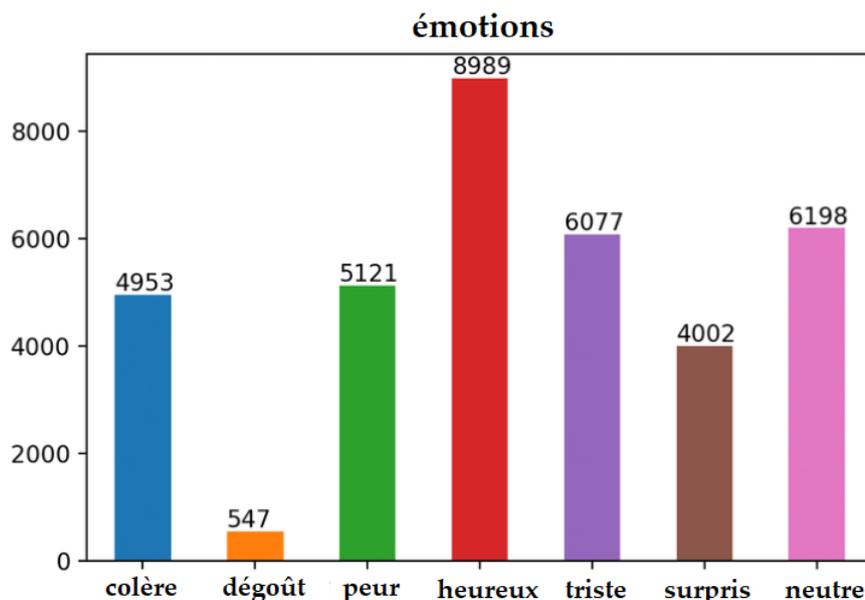


FIGURE 4.3 – distribution des classes dans la base de données FER2013

4.5 Apprentissage et test

Pendant la phase d'apprentissage, nous avons testé plusieurs configurations basées sur la modification de certains paramètres du réseau tels que :

La taille du lot, nombre d'époques, le taux d'apprentissage .

4.5.1 Les expérimentations réalisées

Pour les réseaux de neurones convolutifs utilisés (VGG-16 et ResNet 50), plusieurs expérimentations ont été effectuées :

- Sans augmentation des données .
- Avec augmentation des données :
 - Transformation géométriques
 - SMOTE
 - Cycle-Gan
 - Hybridation des techniques d'augmentation des données

Le tableau suivant résume les expérimentations réalisées sur nos modèles :

Expérimentation	Modèle	Taille du lot	Nombre d'époques	Taux d'apprentissage
1) Modèle sans augmentation	ResNet 50	64	33	0.01
	VGG-16	64	100	0.001
2) Modèle avec transformations géométriques	ResNet 50	128	80	0.01
		128	100	
		64	50	0.01
	VGG-16	64	80	0.001
		64	100	
3) Modèle avec SMOTE	ResNet 50	64	100	0.01
	VGG-16	64	100	0.001
4) Modèle avec transformations géométriques et SMOTE	ResNet 50	64	100	0.01
	VGG-16	64	100	0.001
5) Modèle avec Cycle-Gan	ResNet50	64	33	0.01
6) Modèle avec Cycle-Gan et transformations géométriques	ResNet50	64	100	0.01
7) Modèle avec Cycle-Gan et SMOTE	ResNet50	64	100	0.01
8) Modèle avec Cycle-Gan et SMOTE et transformations géométriques	ResNet50	64	100	0.01

TABLE 4.1 – Paramètres d'apprentissage des différentes expérimentations

4.5.2 Résultats et discussion

Nous pouvons résumer les résultats des expériences que nous avons faites en termes de précision, dans le tableau suivant :

Expérimentation	Modèle	Précision d'entraînement (%)	Précision de test (%)
1) Modèle sans augmentation	ResNet 50	99 %	58 %
	VGG-16	94 %	65 %
2) Modèle avec transformations géométriques	ResNet 50	65 %	61 %
		70 %	64 %
		67 %	62 %
	VGG-16	63 %	60 %
3) Modèle avec SMOTE	ResNet 50	99 %	57 %
	VGG-16	97 %	62 %
4) Modèle avec transformations géométriques et SMOTE	ResNet 50	83 %	68 %
	VGG-16	11%	11%
5) Modèle avec Cycle-Gan	ResNet50	99 %	76 %
6) Modèle avec Cycle-Gan et transformations géométriques	ResNet50	84 %	74 %
7) Modèle avec Cycle-Gan et SMOTE	ResNet50	99 %	75 %
8) Modèle avec Cycle-Gan et SMOTE et transformations géométriques	ResNet50	83 %	72 %

TABLE 4.2 – Tableau récapitulatif des précisions obtenues pour les expérimentations effectués

4.5.3 Les modèles sans augmentation des données

La figure 4.4 représente les résultats des modèles sans augmentation des données. On peut remarquer sur les graphiques de précision des deux modèles et en lisant les résultats dans le tableau, qu'il y a un sur-apprentissage. La figure 4.5 représente les matrices de

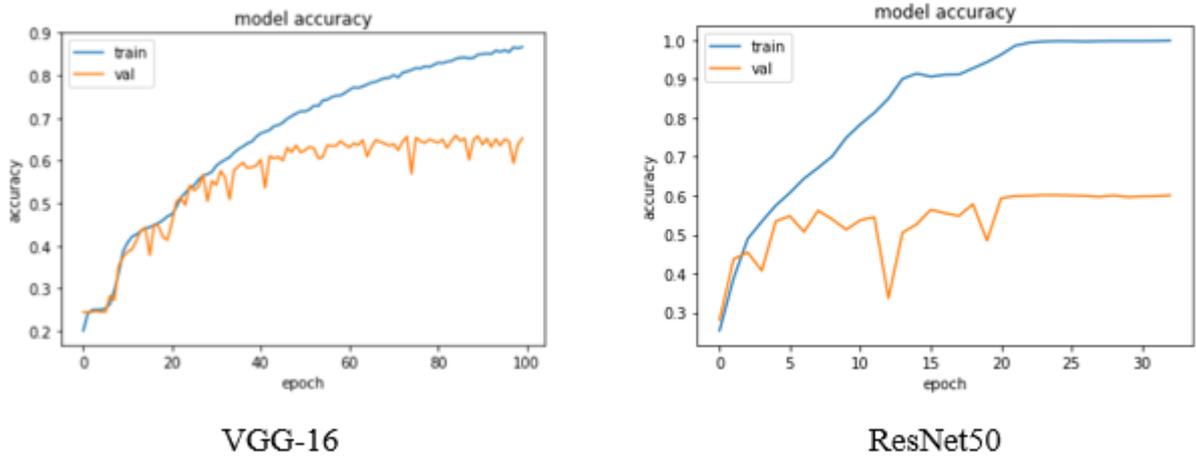


FIGURE 4.4 – Expérimentation 1

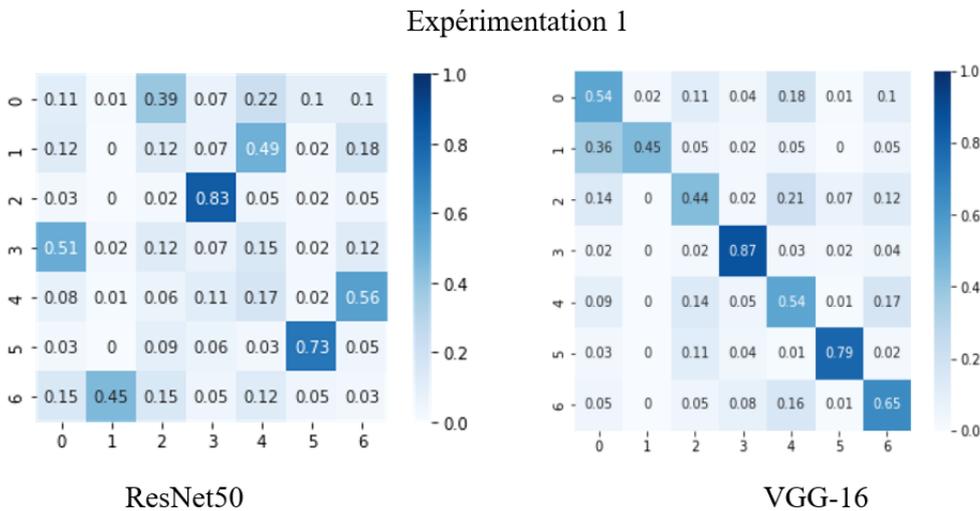


FIGURE 4.5 – Matrice de confusion

confusion des modèles sans aucune augmentation des données, nous pouvons remarquer que les modèles sans augmentation entraînent d'être confus dans la plupart des expressions.

4.5.4 Les modèles avec une des techniques d'augmentation des données

La figure 4.6 représente les graphes de précision des modèles que nous avons, lorsque nous avons ajouté une seule technique d'augmentation des données, et concernant les transformations géométriques, après avoir appliqué différentes techniques de transformation géométrique nous avons trouvé que les meilleures pour nos modèles en terme de précision sont les suivantes :

- Gamme de rotation = 10
- Gamme de cisaillement = 10

- Gamme de décalage en largeur = 0.1
- Gamme de décalage en hauteur = 0.1
- Zoom = 0.1
- Un reflet horizontal

Nous pouvons voir que l'application de SMOTE seul (expérimentation 3) ne résout pas le problème du sur-apprentissage, au contraire qu'avec les transformations géométriques (expérimentation 2) nous pourrions finalement éviter le sur-apprentissage et réduire la différence entre la précision de la formation et la précision du test.

L'expérimentation 5 indique que le Cycle-Gan augmente la précision du système mais ne résout pas vraiment le problème du sur-apprentissage.

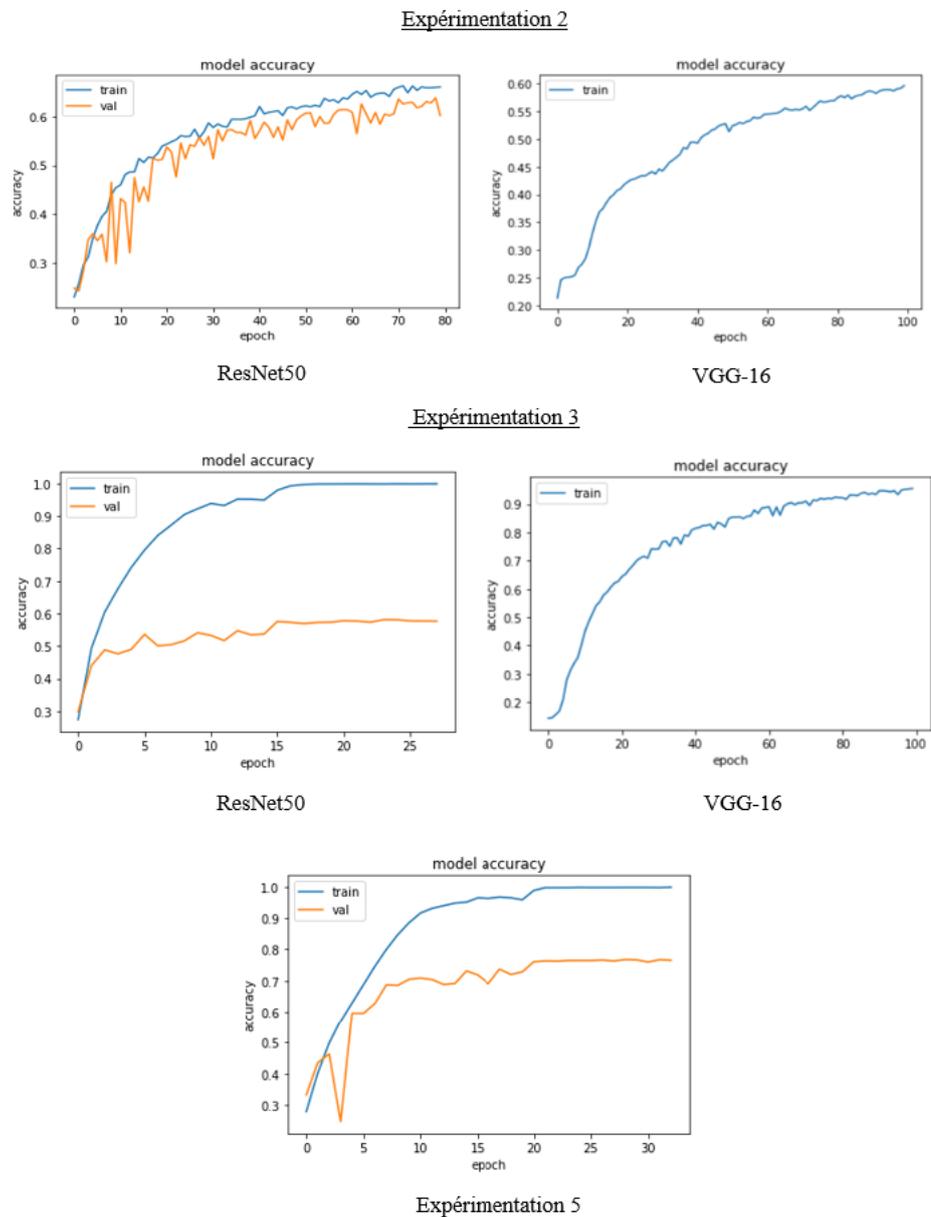


FIGURE 4.6 – Ajouter une seule technique d'augmentation des données

La figure 4.7 représente la matrice de confusion pour certains modèles pour lesquels une seule technique d'augmentation des données est appliquée. On peut voir que la prédiction des modèles des expériences 2 et 3 a été développée par rapport aux expériences 1

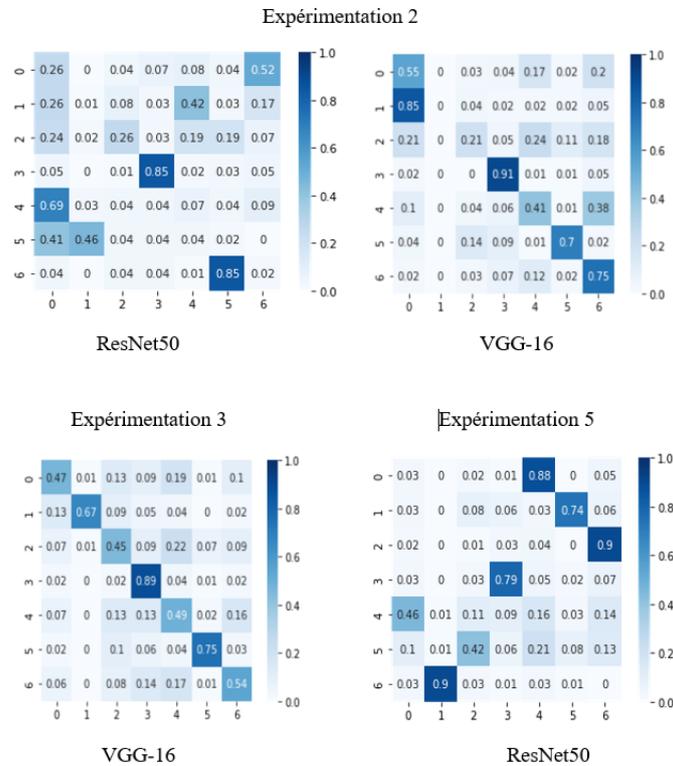


FIGURE 4.7 – Les Matrices de confusion pour les expérimentations avec une seule technique d'augmentation des données

4.5.5 Hybridation des techniques d'augmentation des données

Les figures 4.8 et 4.9 contiennent les graphes des résultats obtenus avec l'application de plus d'une technique d'augmentation des données à la fois. À partir de ces graphes et de la lecture du tableau, nous pouvons remarquer que l'ajout de transformations géométriques chaque fois que nous appliquons l'une des techniques d'augmentation des données est important, surtout avec le Cycle-Gan à partir duquel nous avons obtenu une grande précision et évitons le surapprentissage.

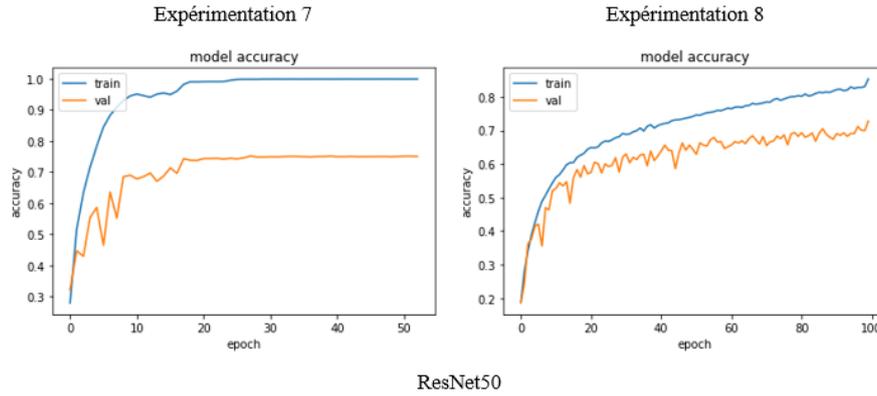


FIGURE 4.8 – Hybridation des techniques d’augmentation des données

La figure 4.10 montre la matrice de confusion de certains modèles pour lesquels nous avons appliqué plus d’une technique d’augmentation des données à la fois, ce qui peut rendre notre modèle plus efficace dans la prédiction et éviter le sur-apprentissage.

4.6 Conclusion

Les expériences réalisées sur les deux modèles ResNet50 et VGG-16 montrent que l’absence d’augmentation des données ou l’application d’une seule technique d’augmentation peut ne pas être suffisante pour éviter le problème du sur-apprentissage et augmenter les performances des systèmes de reconnaissance des expressions faciales. En revanche, lorsque plus d’une technique d’augmentation est utilisée, on constate que le sur-apprentissage disparaît progressivement, en particulier les transformations géométriques qui sont très importantes pour vaincre le sur-apprentissage et augmenter la précision.

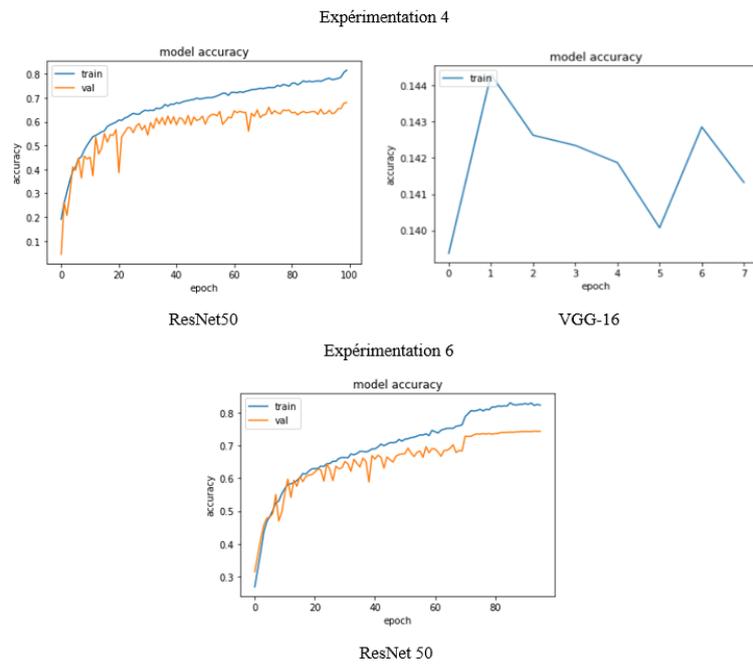


FIGURE 4.9 – Hybridation des techniques d’augmentation des données

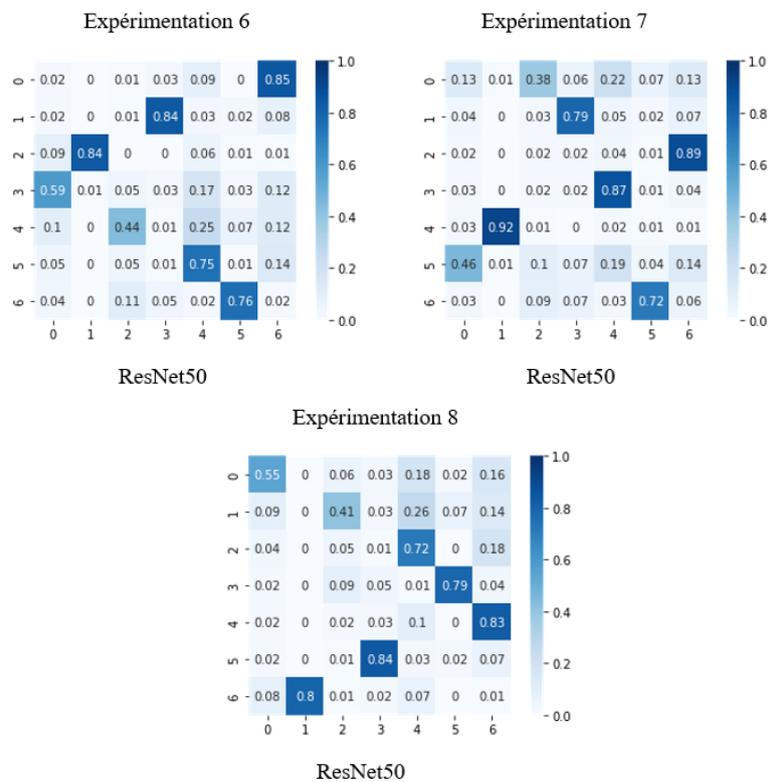


FIGURE 4.10 – Hybridation des techniques d’augmentation des données

Conclusion générale

Les méthodes d'apprentissage profond, qui constituent un sous-ensemble des techniques d'apprentissage automatique, ont permis d'obtenir des résultats remarquables sur des problèmes difficiles de vision par ordinateur, tels que la reconnaissance des expressions faciales. Ces méthodes et applications nécessitent l'utilisation d'un grand volume de données représentatives pour obtenir des résultats de pointe, en raison du nombre considérable de paramètres à optimiser dans les modèles profonds. Cependant, les données sont limitées avec des distributions non équilibrées dans les ensembles de donnée de visage. L'augmentation des données d'image est apparue pour créer des données d'entraînement avec des distributions équilibrées en augmentant les données des classes minoritaires, afin de réduire le sur-apprentissage.

Dans ce mémoire, nous avons exploré l'utilisation de trois techniques pour l'augmentation des données pour la tâche de reconnaissance des expressions faciale, à savoir : les transformations géométriques, la technique SMOTE et la génération de nouvelles images à l'aide de Cycle-Gan.

Les résultats démontrent que les transformations géométriques de l'image apportent que des améliorations à la performance ; en revanche, l'adoption du GAN enrichit la base de données de formation avec des images nouvelles et utiles ce qui offre des améliorations de précision importantes, par rapport au cas où l'augmentation n'est pas utilisée.

En comparant l'utilisation de chacune de ces techniques des données, ainsi que leur hybridation, nous avons constaté que cette dernière est plus efficace que l'utilisation d'une seule technique à la fois, et nous avons remarqué que les meilleures performances du système en termes de précision sont obtenues en utilisant les deux techniques d'augmentation : le cycle gan avec transformations géométriques (74%), ainsi que l'utilisation des trois technique qui a obtenu un précision de (72%) sans avoir de sur-apprentissage.

Ces résultats confirment que les systèmes de FER nécessitent une grande quantité de données pour l'entraînement et que les GANs sont donc une alternative intéressante pour surmonter ce manque de données.

La performance de notre système pourrait être encore améliorée dans des projets futurs. Dans cette optique, nous proposons les points suivants à étudier :

- Validation des résultats sur d'autres bases de données visage.
- Utilisation d'autre Gan récemment proposés tel que StyleGan de Nvidia, qui est considéré actuellement comme le Gan le plus performant en matière de génération de visage photoréaliste.

Bibliographie

- [Abraich *et al.*, 2020] ABRAICH, A., BOUCHOUAT, E. M. et TOUNSI, M. (2020). Improved techniques for gan : Application to mnist data.
- [Ari et Ustazhanov, 2014] ARI, N. et USTAZHANOV, M. (2014). Matplotlib in python. *In 2014 11th International Conference on Electronics, Computer and Computation (ICECCO)*, pages 1–6. IEEE.
- [Baek *et al.*, 2020] BAEK, S., KIM, K. I. et KIM, T.-K. (2020). Weakly-supervised domain adaptation via gan and mesh model for estimating 3d hand poses interacting objects. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6121–6131.
- [Bartlett *et al.*, 2020] BARTLETT, P. L., LONG, P. M., LUGOSI, G. et TSIGLER, A. (2020). Benign overfitting in linear regression. *Proceedings of the National Academy of Sciences*, 117(48):30063–30070.
- [Battineni *et al.*, 2020] BATTINENI, G., SAGARO, G. G., CHINATALAPUDI, N. et AMENTA, F. (2020). Applications of machine learning predictive models in the chronic disease diagnosis. *Journal of personalized medicine*, 10(2):21.
- [Chawla *et al.*, 2002] CHAWLA, N. V., BOWYER, K. W., HALL, L. O. et KEGELMEYER, W. P. (2002). Smote : synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.
- [Chen *et al.*, 2020] CHEN, A., XING, H. et WANG, F. (2020). A facial expression recognition method using deep convolutional neural networks based on edge computing. *IEEE Access*, 8:49741–49751.
- [Dhall *et al.*, 2015] DHALL, A., RAMANA MURTHY, O., GOECKE, R., JOSHI, J. et GEDEON, T. (2015). Video and image based emotion recognition challenges in the wild : Emotiw 2015. *In Proceedings of the 2015 ACM on international conference on multi-modal interaction*, pages 423–426.
- [Ekman et Friesen, 1971] EKMAN, P. et FRIESEN, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124.
- [Fathi-Kazerooni et Rojas-Cessa, 2020] FATHI-KAZEROONI, S. et ROJAS-CESSA, R. (2020). Gan tunnel : Network traffic steganography by using gans to counter internet traffic classifiers. *IEEE Access*, 8:125345–125359.
- [Goodfellow *et al.*, 2014] GOODFELLOW, I., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIR, S., COURVILLE, A. et BENGIO, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.
- [Goodfellow *et al.*, 2013] GOODFELLOW, I. J., ERHAN, D., CARRIER, P. L., COURVILLE, A., MIRZA, M., HAMNER, B., CUKIERSKI, W., TANG, Y., THALER, D., LEE, D.-H. *et al.* (2013). Challenges in representation learning : A report on three machine learning

- contests. *In International conference on neural information processing*, pages 117–124. Springer.
- [Grayson, 2000] GRAYSON, J. E. (2000). *Python and Tkinter programming*. Manning Publications Co. Greenwich.
- [Hammouche, 2009] HAMMOUCHE, Y. (2009). Comparaison de plusieurs méthodes pour la prédiction de la charge électrique nationale. *MEMOIRE Présenté en vue de l’obtention du diplôme de Magistère, université du Badji Mokhtar Annaba, Année*.
- [He et al., 2016] HE, K., ZHANG, X., REN, S. et SUN, J. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- [Hung et Gan, 2021] HUNG, S.-K. et GAN, J. Q. (2021). Small facial image dataset augmentation using conditional gans based on incomplete edge feature input. *PeerJ Computer Science*, 7:e760.
- [Jiang et al., 2020] JIANG, X., LU, M. et WANG, S.-H. (2020). An eight-layer convolutional neural network with stochastic pooling, batch normalization and dropout for fingerspelling recognition of chinese sign language. *Multimedia Tools and Applications*, 79(21):15697–15715.
- [Kim et Han, 2020] KIM, J. H. et HAN, D. S. (2020). Data augmentation & merging dataset for facial emotion recognition. *In Proceedings of the Symposium of the 1st Korea Artificial Intelligence Conference, Jeju, Korea*, pages 12–16.
- [Ledig et al., 2017] LEDIG, C., THEIS, L., HUSZÁR, F., CABALLERO, J., CUNNINGHAM, A., ACOSTA, A., AITKEN, A., TEJANI, A., TOTZ, J., WANG, Z. et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690.
- [Lee et al., 2019] LEE, M. K., CHOI, D. Y., KIM, D. H. et SONG, B. C. (2019). Visual scene-aware hybrid neural network architecture for video-based facial expression recognition. *In 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, pages 1–8. IEEE.
- [Li et al., 2020] LI, K., JIN, Y., AKRAM, M. W., HAN, R. et CHEN, J. (2020). Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy. *The visual computer*, 36(2):391–404.
- [Li et al., 2021] LI, Z., LIU, F., YANG, W., PENG, S. et ZHOU, J. (2021). A survey of convolutional neural networks : analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*.
- [Lucey et al., 2010] LUCEY, P., COHN, J. F., KANADE, T., SARAGIH, J., AMBADAR, Z. et MATTHEWS, I. (2010). The extended cohn-kanade dataset (ck+) : A complete dataset for action unit and emotion-specified expression. *In 2010 IEEE computer society conference on computer vision and pattern recognition-workshops*, pages 94–101. IEEE.
- [Lundqvist et al., 1998] LUNDQVIST, D., FLYKT, A. et ÖHMAN, A. (1998). Karolinska directed emotional faces. *Cognition and Emotion*.
- [Lyons et al., 1999] LYONS, M. J., BUDYNEK, J. et AKAMATSU, S. (1999). Automatic classification of single facial images. *IEEE transactions on pattern analysis and machine intelligence*, 21(12):1357–1362.

- [Manel,] MANEL, B. Reconnaissance de visage 3d par un réseau de neurones à convolution.
- [Mazen *et al.*,] MAZEN, F. M. A., NASHAT, A. A. et SEOUD, R. A. A. A. Real time face expression recognition along with balanced fer2013 dataset using cyclegan.
- [Mifdal, 2019] MIFDAL, R. (2019). *Application des techniques d'apprentissage automatique pour la prédiction de la tendance des titres financiers*. Thèse de doctorat, École de technologie supérieure.
- [Mirza et Osindero, 2014] MIRZA, M. et OSINDERO, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv :1411.1784*.
- [Mollahosseini *et al.*, 2017] MOLLAHOSSEINI, A., HASANI, B. et MAHOOR, M. H. (2017). Affectnet : A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10(1):18–31.
- [Moreno-Barea *et al.*, 2018] MORENO-BAREA, F. J., STRAZZERA, F., JEREZ, J. M., URDA, D. et FRANCO, L. (2018). Forward noise adjustment scheme for data augmentation. In *2018 IEEE symposium series on computational intelligence (SSCI)*, pages 728–734. IEEE.
- [PANTIC, 2008] PANTIC, M. (2008). Facial expression recognition. *Imperial College London, London, UK ; University of Twente, AE Enschede, The Netherlands*, page 7.
- [Pantic *et al.*, 2005] PANTIC, M., VALSTAR, M., RADEMAKER, R. et MAAT, L. (2005). Web-based database for facial expression analysis. In *2005 IEEE international conference on multimedia and Expo*, pages 5–pp. IEEE.
- [Parizeau, 2004] PARIZEAU, M. (2004). Réseaux de neurones. *GIF-21140 et GIF-64326*, 124.
- [Porcu *et al.*, 2020] PORCU, S., FLORIS, A. et ATZORI, L. (2020). Evaluation of data augmentation techniques for facial expression recognition systems. *Electronics*, 9(11):1892.
- [Sammut et Webb, 2011] SAMMUT, C. et WEBB, G. I. (2011). *Encyclopedia of machine learning*. Springer Science & Business Media.
- [Shorten et Khoshgoftaar, 2019] SHORTEN, C. et KHOSHGOFTAAR, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48.
- [Simonyan et Zisserman, 2014] SIMONYAN, K. et ZISSERMAN, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv :1409.1556*.
- [Srivastava *et al.*, 2014] SRIVASTAVA, N., HINTON, G., KRIZHEVSKY, A., SUTSKEVER, I. et SALAKHUTDINOV, R. (2014). Dropout : a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958.
- [Torre et Cohn, 2011] TORRE, F. D. I. et COHN, J. F. (2011). Facial expression analysis. *Visual analysis of humans*, pages 377–409.
- [Umer *et al.*, 2022] UMER, S., ROUT, R. K., PERO, C. et NAPPI, M. (2022). Facial expression recognition with trade-offs between data augmentation and deep learning features. *Journal of Ambient Intelligence and Humanized Computing*, 13(2):721–735.
- [Wang *et al.*, 2017] WANG, K., GOU, C., DUAN, Y., LIN, Y., ZHENG, X. et WANG, F.-Y. (2017). Generative adversarial networks : introduction and outlook. *IEEE/CAA Journal of Automatica Sinica*, 4(4):588–598.

- [Wang *et al.*, 2020] WANG, S., YANG, Y., WU, Z., QIAN, Y. et YU, K. (2020). Data augmentation using deep generative models for embedding based speaker recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:2598–2609.
- [Wira, 2009] WIRA, P. (2009). réseaux de neurones : architecture et applications. Mémoire de D.E.A., université de Haute Alsace.
- [Wu, 2017] WU, J. (2017). Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology. Nanjing University. China*, 5(23):495.
- [Xu *et al.*, 2020] XU, D., WEI, C., PENG, P., XUAN, Q. et GUO, H. (2020). Ge-gan : A novel deep learning framework for road traffic state estimation. *Transportation Research Part C : Emerging Technologies*, 117:102635.
- [Yamashita *et al.*, 2018] YAMASHITA, R., NISHIO, M., DO, R. K. G. et TOGASHI, K. (2018). Convolutional neural networks : an overview and application in radiology. *Insights into imaging*, 9(4):611–629.
- [Yan *et al.*, 2021] YAN, B., XIAO, Z., YUAN, P., CAI, K. et CHEN, Q. (2021). Facial expression recognition with convolutional neural networks via a data augmentation strategy.
- [Yu *et al.*, 2020] YU, Y., HUANG, Z., LI, F., ZHANG, H. et LE, X. (2020). Point encoder gan : A deep learning model for 3d point cloud inpainting. *Neurocomputing*, 384:192–199.
- [Zhang *et al.*, 2020] ZHANG, G., LIU, Y. et JIN, X. (2020). A survey of autoencoder-based recommender systems. *Frontiers of Computer Science*, 14(2):430–450.
- [Zhao *et al.*, 2011] ZHAO, G., HUANG, X., TAINI, M., LI, S. Z. et PIETIKÄINEN, M. (2011). Facial expression recognition from near-infrared videos. *Image and Vision Computing*, 29(9):607–619.
- [Zhu *et al.*, 2017] ZHU, J.-Y., PARK, T., ISOLA, P. et EFROS, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *In Proceedings of the IEEE international conference on computer vision*, pages 2223–2232.
- [Zhu *et al.*, 2018] ZHU, X., LIU, Y., LI, J., WAN, T. et QIN, Z. (2018). Emotion classification with data augmentation using generative adversarial networks. *In Pacific-Asia conference on knowledge discovery and data mining*, pages 349–360. Springer.

Webographie

- [W1] , <https://www.axiocode.com/avantages-inconvenients-intelligence-artificielle>,
Dernier accès au site : 10/02/2022
- [W2] , <https://penseeartificielle.fr/focus-reseau-neurones-convolutifs>, Der-
nier accès au site : 19/02/2022
- [W3] , <https://openclassrooms.com/fr/courses/4470531-classez-et-segmentez-des-donnees-5083336-decouvrez-les-differentes-couches-dun-cnn>, Dernier accès au site :
20/02/2022
- [W4] , <https://ichi.pro/fr/apprentissage-profond-applique-partie-3-autoencodeurs-30821>
Dernier accès au site : 21/02/2022
- [W5] , <http://viso.ai/deep-learning/resnet-residual-neural-network>, Dernier
accès au site : 08/05/2022
- [W6] , <https://datascientest.com/kaggle-tout-ce-quil-a-savoir-sur-cette-plateforme>,
Dernier accès au site : 25/04/2022
- [W7] , <https://research.google.com/colaboratory/faq.html?hl=fr>, Dernier accès
au site : 23/04/2022
- [W8] , [https://www.lebigdata.fr/tensorflow-definition-tout-savoir#TensorFlow_QUEST-CE_QUE_CEST](https://www.lebigdata.fr/tensorflow-definition-tout-savoir#TensorFlow_quest-ce_que_cest), Dernier accès au site : 27/04/2022
- [W9] , <https://www.journaldunet.fr/web-tech/guide-de-l-intelligence-artificielle/1501863-keras-bibliotheque-de-deep-learning/#:~:text=Ecrire>, Dernier ac-
cès au site : 27/04/2022
- [W10] , <https://www.mygreatlearning.com/blog/python-numpy-tutorial>, Dernier
accès au site : 27/04/2022