

République Algérienne Démocratique et Populaire  
Ministère de l'enseignement Supérieur et de la Recherche Scientifique  
Université de Mohamed El Bachir El Ibrahimi de Borj Bou Arréridj  
Faculté des Mathématiques et d'Informatique  
Département d'informatique



## **MEMOIRE**

Présenté en vue de l'obtention du diplôme  
**Master en informatique**  
Spécialité : Ingénierie de l'informatique Décisionnelle

## **THEME**

Apprentissage Supervisé Pour La Reconnaissance Facial Des  
Images à Basse Résolution

*Présenté par Les Etudiantes :*

BOUGUERRA HAYAT

HAMIDI HADJER

*Devant le jury composé de :*

**Président :** Mr. ATTIA ABDELOUHAB

**Examineur :** Mr. MOUHOU B BELAZZOUG

**Encadreur :** Mr. MOHAMED AMINE BEGHOURA

**2024/2025**

## **Dédicace**

De tout mon cœur, je dédie ce modeste travail à ceux qui m'ont soutenu, aidé et conseillé tout au long de ma vie. J'exprime ma profonde gratitude et mon respect à :

### **À ma chère mère**

Pour tout son amour, sa dévotion inébranlable, et son soutien sans faille à chaque étape de ma vie.

### **À mes chers frères Hassane et Oussama**

Pour leur soutien moral constant, leur encouragement et tous les sacrifices qu'ils ont faits tout au long de ma formation.

### **À ma chère amie Hadjer**

Pour son soutien précieux, sa présence constante et son encouragement tout au long de ce parcours.

### **À mes amis proches, mes frères et sœurs de cœur**

Pour leur présence, leur soutien indéfectible, leur encouragement et leur affection sincère.

### **À tous mes collègues à l'université**

### **Et enfin...**

À toutes les personnes qui ont participé directement ou indirectement à la réalisation de ce travail.

**Bouguerra Hayat**

## **Dédicace**

De tout mon cœur, je dédie ce modeste travail à ceux qui m'ont soutenu, aidé et conseillé tout au long de ma vie. J'exprime ma profonde gratitude et mon respect à :

### **À mes chers parents**

Merci à vous deux pour tout ce que vous avez fait. Cette réussite est le fruit de vos sacrifices.

À ma chère mère, symbole d'amour inconditionnel, de patience et de tendresse. Ton soutien et tes prières m'ont toujours accompagné.

À mon père, pour sa sagesse, sa force et ses encouragements constants. Tu as toujours cru en moi, même quand moi-même j'en doutais.

### **À ma chère amie Hayat**

Pour son soutien précieux, sa présence constante et son encouragement tout au long de ce parcours.

### **À mes amis proches, mes frères et sœurs de cœur**

Pour leur présence, leur soutien indéfectible, leur encouragement et leur affection.

### **À tous mes collègues à l'université**

**Hamidi Hadjer**

# Remerciements

Avant tout, nous tenons à remercier **DIEU** pour nous avoir accordé la santé, la possibilité et la volonté de commencer et de poursuivre nos études.

Nous tenons d'abord à remercier notre encadrant, **Monsieur Mohamed Amine Beghoura**, pour nous avoir donné l'opportunité de mener nos recherches et pour nous avoir prodigué des conseils inestimables tout au long de cette étude. Nous tenons également à le remercier pour sa générosité, sa gentillesse, ses connaissances, le temps qu'il nous a accordé et sa grande disponibilité, qui ont rendu notre tâche beaucoup plus facile.

Nous souhaitons également exprimer nos sincères remerciements aux membres du jury qui nous ont honorés en participant à l'évaluation de ce travail et en l'enrichissant de leurs remarques et suggestions constructives : **Monsieur Attia Abdelouhab**, Président du jury, et **Monsieur Mouhoub Belazug**, Examineur.

De plus, nous tenons à reconnaître nos parents pour leur amour, leurs soins, leurs prières, leurs sacrifices et pour avoir toujours été là pour nous.

# Résumé

La reconnaissance faciale est une technologie clé des systèmes biométriques, largement utilisée dans la sécurité et la surveillance. Toutefois, ses performances dépendent fortement de la qualité des images. Dans les cas de basse résolution typiques des vidéos de surveillance ou d'images web les modèles classiques perdent en efficacité à cause du manque de détails visuels.

Ce travail traite de la reconnaissance faciale en basse résolution (LRFR). Il vise à surmonter des défis majeurs tels que la perte de traits faciaux, les variations de pose, d'éclairage et les occultations, en adoptant une approche hybride.

La méthode repose sur l'utilisation de CNN pour l'extraction de caractéristiques, de modèles pré-entraînés (VGG-Face, MobileNetV2) pour une représentation robuste, et de techniques de super-résolution (Bicubic, Lanczos, FSRCNN) pour améliorer la netteté des images avant reconnaissance.

Nous intégrons également le Few-Shot Learning, permettant d'identifier un individu avec très peu d'exemples annotés une contrainte fréquente dans les bases de données réelles.

Les résultats montrent qu'une reconnaissance efficace est possible, même à partir d'images de faible qualité, grâce à une combinaison cohérente de traitement d'image et d'apprentissage profond.

---

**Mots-clés :** Reconnaissance faciale en basse résolution (LRFR), Réseaux de neurones convolutifs (CNN), Super-résolution (SR), VGG-Face, MobileNetV2, Few-Shot Learning, Apprentissage profond, Prétraitement d'images, Données faibles, Surveillance vidéo.

# Abstract

Facial recognition is a key technology in biometric systems, widely used in security and surveillance. However, its performance is highly dependent on image quality. In low-resolution cases typical of surveillance videos or web images conventional models lose effectiveness due to the lack of visual detail.

This work deals with low-resolution facial recognition (LRFR). It aims to overcome major challenges such as loss of facial features, variations in pose, lighting and occlusions, by adopting a hybrid approach.

The method relies on the use of CNN for feature extraction, pre-trained models (VGG-Face, MobileNetV2) for robust representation, and super-resolution techniques (Bicubic, Lanczos, FSRCNN) to improve image sharpness prior to recognition.

We also integrate Few-Shot Learning, to identify an individual with very few annotated examples a frequent constraint in real databases.

The results show that effective recognition is possible, even from low-quality images, thanks to a coherent combination of image processing and deep learning.

---

**Keywords :** Low-Resolution Face Recognition (LRFR), Convolutional Neural Networks (CNN), Super-Resolution (SR), VGG-Face, MobileNetV2, Few-Shot Learning, Deep Learning, Image Preprocessing, Low Data, Video Surveillance.

## ملخص

يُعد التعرف على الوجوه تقنية أساسية في أنظمة القياسات الحيوية، ويستخدم على نطاق واسع في مجالات الأمن والمراقبة. ومع ذلك، فإن أداء هذه الأنظمة يعتمد بشكل كبير على جودة الصور. ففي الحالات التي تكون فيها الصور منخفضة الدقة كما هو الحال في مقاطع الفيديو المراقبة أو الصور المستخرجة من الإنترنت تفقد النماذج التقليدية فعاليتها بسبب نقص التفاصيل البصرية.

يعالج هذا العمل موضوع التعرف على الوجوه في الصور منخفضة الدقة LRFR، ويهدف إلى التغلب على التحديات الرئيسية مثل فقدان ملامح الوجه، وتغير أوضاع الإضاءة والزواوية، والانسدادات الجزئية، من خلال اعتماد نهج هجين.

تعتمد المنهجية على استخدام الشبكات العصبية الالتفافية CNN لاستخراج الخصائص، ونماذج مدربة مسبقاً مثل VGG و MobileNetV2 للحصول على تمثيلات قوية، بالإضافة إلى تقنيات تحسين الدقة Bicubic, Lanczos, FSRCNN من أجل تعزيز وضوح الصور قبل عملية التعرف.

كما ندمج تقنية التعلم بالقليل من البيانات Few-Shot Learning، والتي تسمح بالتعرف على الأفراد انطلاقاً من عدد محدود جداً من الصور المصنفة وهو ما يُعد من القيود الشائعة في قواعد البيانات الواقعية.

تُظهر النتائج أن التعرف الفعّال على الوجوه ممكن حتى عند استخدام صور منخفضة الجودة، وذلك بفضل الدمج المتناسق بين معالجة الصور والتعلم العميق.

---

الكلمات الرئيسية: التعرف على الوجوه ذات الدقة المنخفضة LRFR، الشبكات العصبية الالتفافية CNN، تحسين الدقة (فائق الدقة) SR، VGG، MobileNetV2، Few-Shot، التعلم العميق، معالجة الصور الأولية، بيانات قليلة، المراقبة بالفيديو.

# Table des matières

	vii
<b>Liste des abréviations</b>	<b>xi</b>
<b>Liste des figures</b>	<b>xi</b>
<b>Liste des tableaux</b>	<b>xiii</b>
<b>1 Reconnaissance Faciale En Basse Résolution</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Reconnaissance faciale . . . . .	1
1.2.1 Les Approches De Reconnaissance Faciale . . . . .	2
1.2.2 La Nouvelle Tendance De La Reconnaissance Faciale : . . . . .	3
1.2.3 Avantages et Inconvénients de Reconnaissance Faciale . . . . .	5
1.3 Traitement D'images . . . . .	5
1.3.1 Phases Du Traitement D'image . . . . .	6
1.4 La Basse Résolution . . . . .	7
1.4.1 Défis De La Reconnaissance Faciale En Basse Résolution . . . . .	8
1.4.2 Approches La Reconnaissance Faciale En Basse Résolution . . . . .	10
1.5 Super-résolution . . . . .	10
1.5.1 Objectif De Super-Résolution . . . . .	11
1.5.2 Méthodes Super-Résolution . . . . .	12
1.5.3 Les Avantages De La Super-Résolution . . . . .	14
1.6 Conclusion . . . . .	14
<b>2 Apprentissage Profond Et Apprentissage Avec Peu D'exemples.</b>	<b>15</b>

2.1	Introduction . . . . .	15
2.2	L'apprentissage Automatique . . . . .	15
2.2.1	L'approches principales . . . . .	16
2.3	L'apprentissage Profond . . . . .	17
2.3.1	Les Fondamentaux De L'apprentissage Profond . . . . .	19
2.3.2	L'apprentissage Profond Dans La Reconnaissance Faciale En Basse Résolution . . . . .	19
2.3.3	Types De Réseaux Neuronaux Dans Le Apprentissage Profond : . . . .	20
2.3.4	Applications De L'apprentissage Profond . . . . .	21
2.4	Apprentissage à Quelques Coups . . . . .	22
2.4.1	Les Fonctionne L'apprentissage à Quelques Coups . . . . .	23
2.4.2	Terminologies Liées à L'apprentissage à Quelques Coups . . . . .	23
2.4.3	Travaux Connexes . . . . .	24
2.5	Conclusion . . . . .	26
<b>3</b>	<b>Méthodologie</b>	<b>27</b>
3.1	Introduction . . . . .	27
3.2	Approche proposée . . . . .	28
3.2.1	Collecte des Données . . . . .	29
3.2.2	Prétraitement . . . . .	30
3.2.3	CNN simple . . . . .	31
3.2.4	VGG . . . . .	34
3.2.5	Few-Shot Learning . . . . .	38
3.2.6	Super-résolution . . . . .	39
3.2.7	Structure du modèle FSRCNN : . . . . .	42
3.3	Conclusion . . . . .	43
<b>4</b>	<b>Résultats et Discussion</b>	<b>44</b>
4.1	Introduction . . . . .	44
4.2	Les critères d'évaluation . . . . .	44
4.2.1	Accuracy . . . . .	44
4.2.2	Loss . . . . .	45
4.3	Résultats expérimentaux . . . . .	45

4.3.1	Entraînement de base sur images Basse resolution . . . . .	45
4.3.2	Table de comparaison . . . . .	51
4.3.3	Amélioration par Super-Résolution . . . . .	52
4.4	Resultat de comparaison . . . . .	65
4.4.1	Meilleure méthode . . . . .	67
4.5	Environnement de développement . . . . .	67
4.5.1	L'environnements logiciel . . . . .	67
4.5.2	L'environnements logiciel . . . . .	69
4.6	Conclusion . . . . .	69
<b>Références</b>		<b>73</b>

# Table des figures

1.1	Les Approches de reconnaissance faciale . . . . .	2
1.2	La Basse Résolution . . . . .	8
1.3	La Super-résolution . . . . .	11
2.1	Les sous-couches de l'intelligence artificielle . . . . .	16
2.2	La différence entre l'apprentissage automatique (ML) et l'apprentissage profond (DL) . . . . .	18
3.1	L'architecture générale de notre approche proposée . . . . .	28
3.2	L'architecture de modèle CNN simple . . . . .	32
3.3	Architecteur du modèle VGG . . . . .	36
4.1	Accuracy validation graphicx CNN simple . . . . .	46
4.2	Loss validation graphicx CNN simple . . . . .	46
4.3	Résultat de Modèle CNN simple . . . . .	47
4.4	Accuracy validation graphicx VGG . . . . .	48
4.5	Training Loss graphicx VGG . . . . .	49
4.6	Résultat de Modèle VGG . . . . .	49
4.7	Training Loss Accuracy validation graphicx Few-Shot . . . . .	50
4.8	Résultat de Modèle Few-Shot . . . . .	51
4.9	Training Loss Accuracy Validation CNN simple bicubic . . . . .	53
4.10	Training Loss Accuracy Validation VGG bicubic . . . . .	54
4.11	Training Loss Accuracy Validation Few-Shot CNN . . . . .	55
4.12	Training Loss Accuracy Validation Few-Shot ResNet18 . . . . .	56
4.13	Training Loss Accuracy Validation CNN Simple Lanczos . . . . .	57
4.14	Training Loss Accuracy Validation VGG Lanczos . . . . .	58

4.15	Training Loss Accuracy Validation Few-Shot CNN Lanczos . . . . .	59
4.16	Training Loss Accuracy Validation Few-Shot ResNet18 Lanczos . . . . .	60
4.17	Training Loss et Accuracy Validation CNN Simple FSRCNN . . . . .	61
4.18	Training Loss Accuracy Validation VGG FSRCNN . . . . .	62
4.19	Training Loss et Validation Accuracy pour Few-Shot CNN + FSRCNN . . . . .	63
4.20	Training Loss Accuracy Validation Few-Shot ResNet18 . . . . .	64
4.21	Résultat de comparaison pour les méthodes de super-résolution . . . . .	66

# Liste des tableaux

2.1	Travaux connexes sur la reconnaissance faciale en basse résolution. . . . .	25
3.1	Paramètres architecturaux du modèle CNN simple . . . . .	33
3.2	Hyper-paramètres d'entraînement du modèle CNN simple . . . . .	33
3.3	Hyper-paramètres du modèle VGG-Face . . . . .	37
4.1	Comparaison des performances entre CNN, VGG et Few-Shot . . . . .	51
4.2	Comparaison des performances selon les méthodes de reconnaissance . . . . .	52
4.3	Comparaison des performances selon 2eme méthodes de reconnaissance . . . . .	56
4.4	Comparaison des performances selon 3eme méthodes de reconnaissance . . . . .	60
4.5	Comparaison des performances selon les méthodes de super-résolution et d'architecture . . . . .	65
4.6	Détails de la configuration de l'ordinateur utilisé pour les études expérimentales dans cette étude. . . . .	69

# Introduction Générale

La reconnaissance faciale est l'un des principaux domaines de la vision par ordinateur et de l'intelligence artificielle. Cette technologie permet d'identifier ou de vérifier l'identité d'une personne à partir de son visage en utilisant des algorithmes d'apprentissage automatique. Elle est utilisée dans des applications aussi diverses que les systèmes de sécurité, la vidéo-surveillance, les paiements électroniques, les médias sociaux et l'analyse du comportement. Malgré des avancées significatives dans ce domaine, la technologie est confrontée à plusieurs défis, en particulier dans les environnements réels où les images sont souvent de faible qualité.

L'un des défis les plus importants est la reconnaissance des visages à partir d'images à faible résolution, connue sous le nom de reconnaissance des visages à faible résolution (Low-Resolution Face Recognition LRFR). Dans les cas de télésurveillance, de faible éclairage ou de mouvement rapide, les images capturées perdent leur clarté et leurs détails, ce qui a un impact négatif sur la précision des systèmes de reconnaissance traditionnels qui ont été formés à l'origine sur des images de haute qualité. Pour résoudre ce problème, nous proposons dans cet article une approche hybride basée sur des techniques avancées de traitement d'images, des réseaux neuronaux convolutifs (cnn), des méthodes d'optimisation de la résolution (SR) telles que bicubique ; FSRcnn ; Lanczos, ainsi que des modèles d'apprentissage profond tels que MobileNetV2 et VGG. Nous intégrons également une stratégie d'apprentissage en quelques clics (Few-Shot Learning) adaptée aux situations où peu de données d'apprentissage sont disponibles.

L'objectif de ce travail est de construire un pipeline de reconnaissance faciale intégré qui soit robuste et capable de fonctionner efficacement dans des conditions réelles avec des images de faible qualité. Pour y parvenir, nous avons développé une architecture complète qui comprend le prétraitement de l'image, l'amélioration de la résolution et la phase de classification.

## **Chapitre 1 : Reconnaissance Faciale en Basse Résolutio**

Ce chapitre couvre le domaine général de la reconnaissance des visages et explique les différentes techniques utilisées. Il se concentre sur le défi que représentent les images à faible résolution et explique l'impact de la perte d'informations visuelles sur les performances des modèles. Il présente les applications courantes de la technique et les difficultés rencontrées dans des environnements réels tels que la vidéosurveillance.

## **chapitre 2 : Apprentissage profond et Apprentissage avec peu d'exemples.**

Le deuxième chapitre traite des concepts théoriques de base qui sous-tendent la recherche, en commençant par l'apprentissage automatique, qui vise à construire des modèles qui apprennent à partir de données et prennent des décisions sans intervention humaine directe. L'apprentissage profond, une branche avancée qui s'appuie sur des réseaux neuronaux multicouches pour imiter le fonctionnement du cerveau humain, ce qui permet aux modèles d'extraire des caractéristiques complexes à partir de données, en particulier d'images. Enfin, le chapitre présente le concept d'apprentissage à partir de quelques exemples (Few Shot Learning), qui permet aux modèles de généraliser et d'apprendre à partir d'un nombre très limité d'exemples, ce qui est essentiel lorsqu'il s'agit de traiter des données de petite taille ou limitées, comme dans le cas de la reconnaissance de visages à faible résolution.

## **Chapitre 3 : méthodologie**

Ce chapitre est consacré à la description de la méthodologie, en expliquant chaque étape de la chaîne de traitement : prétraitement des images (par exemple, égalisation de la distribution des couleurs, filtrage bilatéral), amélioration de la résolution à l'aide de différentes techniques et classification des visages à l'aide de modèles de réseaux neuronaux convolutionnels. L'organisation des données, la modélisation et les choix techniques sont expliqués de manière détaillée et structurée.

## **Chapitre 4 : Résultats et Discussion**

Ce chapitre présente les résultats obtenus lors des expérimentations sur la base de données TinyFace. Il analyse les performances de chaque version du système proposé et compare les résultats sur la base de plusieurs critères tels que la précision, le taux d'erreur et la robustesse aux changements d'entrée. Il inclut une discussion critique pour interpréter les résultats.

# Chapitre 1

## Reconnaissance Faciale En Basse

## Résolution

### 1.1 Introduction

La reconnaissance faciale (Face Recognition ou FR) est une technologie clé dans les domaines de la vision par ordinateur et de l'intelligence artificielle, avec des applications variées telles que la sécurité, la surveillance, l'authentification biométrique et l'interaction homme-machine. Grâce aux avancées de l'apprentissage profond, les modèles modernes de reconnaissance faciale atteignent un niveau de précision très élevé lorsqu'ils sont appliqués à des images haute résolution. Cependant, dans des conditions réelles, les images capturées ne sont pas toujours de haute qualité. Elles peuvent être affectées par des facteurs tels que la distance, le mouvement, l'éclairage, la qualité du capteur et l'angle de prise de vue, ce qui entraîne une réduction de la résolution et une perte de détails importants du visage. La reconnaissance faciale en basse résolution (Low Resolution Face Recognition ou LRFR) représente donc un défi crucial pour des applications telles que la surveillance vidéo, la reconnaissance de suspects et l'identification à distance.

### 1.2 Reconnaissance faciale

La reconnaissance faciale est un domaine de l'intelligence artificielle (IA) qui permet d'analyser les traits du visage des personnes à l'aide d'algorithmes avancés et de les comparer à des

images stockées dans une base de données. À partir d'une image numérique ou d'une vidéo en direct, le système peut vérifier l'identité d'une personne, c'est-à-dire valider sa demande, ou reconnaître une personne spécifique au sein d'un groupe. Il est important de ne pas confondre un système de reconnaissance faciale avec un système de détection des visages, qui se limite à identifier la présence d'un visage dans une image sans en reconnaître l'identité[1].

### 1.2.1 Les Approches De Reconnaissance Faciale

Ils existent différentes approches qui ont été développées dans le domaine de reconnaissance des expressions faciales. Elles peuvent être séparées en deux grandes familles à savoir : les méthodes globales (ou holistiques) et les méthodes locales[2].

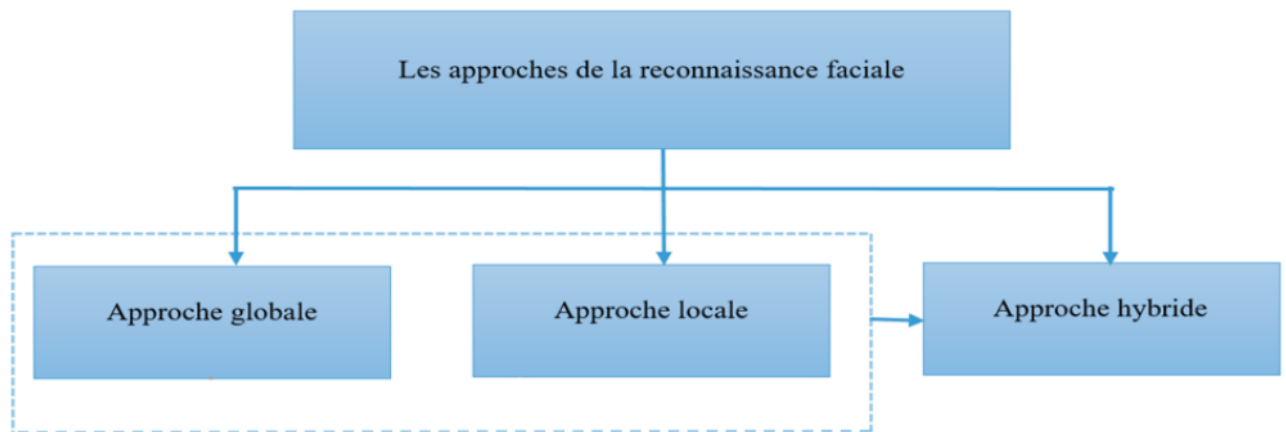


FIGURE 1.1 – Les Approches de reconnaissance faciale

#### 1.2.1.1 Approches Globale

Elles sont basées sur des techniques d'analyse statistique bien connues. Il n'est pas nécessaire de repérer certains points caractéristiques du visage (comme les centres des yeux, les narines, le centre de la bouche, etc.) à part pour normaliser les images. Dans ces méthodes, les images de visage (qui peuvent être vues comme des matrices de valeurs de pixels) sont traitées de manière globale et sont généralement transformées en vecteurs, plus faciles à manipuler. L'avantage principal des méthodes globales est qu'elles sont relativement rapides à mettre en œuvre et que les calculs de base sont d'une complexité moyenne. En revanche, elles sont très sensibles aux variations d'éclairage, de pose et d'expression faciale. Ceci se comprend aisément puisque la moindre variation des conditions de l'environnement ambiant entraîne des

changements inéluctables dans les valeurs des pixels qui sont traités directement. Ces méthodes utilisent principalement une analyse de sousespaces de visages. L'utilisation de techniques de modélisation de sous-espace a fait avancer la technologie de reconnaissance faciale de manière significative[2].

### 1.2.1.2 Approches Locales

Elles sont basées sur des modèles, utilisant des connaissances a priori que l'on possède sur la morphologie du visage et s'appuie en général sur des points caractéristiques de celui-ci Kanade a présenté un des premiers algorithmes de ce type en détectant certains points ou traits caractéristiques d'un visage puis en les comparant avec des paramètres extraits d'autres visages. Ces méthodes constituent une autre approche pour prendre en compte la non-linéarité en construisant un espace de caractéristiques local et en utilisant des filtres d'images appropriés, de manière à ce que les distributions des visages soient moins affectées par divers changements. Les approches bayésiennes, les machines à vecteurs de support (SVM), la méthode des modèles actifs d'apparence (AAM) ou encore la méthode "local binary pattern"(LBP) ont été utilisées dans ce but. Toutes ces méthodes ont l'avantage de pouvoir modéliser plus facilement les variations de pose, d'éclairage et d'expression par rapport aux méthodes globales. Toutefois, elles sont plus lourdes à utiliser puisqu'il faut souvent placer manuellement un assez grand nombre de points sur le visage alors que les méthodes globales ne nécessitent de connaître que la position des yeux afin de normaliser les images, ce qui peut être fait automatiquement et de manière assez fiable par un algorithme de détection [2].

### 1.2.1.3 Approches Hybrides

Ce sont une combinaison des méthodes globales et locales en combinant la détection de caractéristiques géométriques (ou structurales) avec l'extraction de caractéristiques d'apparence locales. Elles permettent d'augmenter la stabilité de la performance de reconnaissance lors de changements de pose, d'éclairage et d'expressions faciales[2].

## 1.2.2 La Nouvelle Tendances De La Reconnaissance Faciale :

L'apprentissage en profondeur (*Deep Face Recognition*) applique plusieurs couches de traitement pour apprendre des représentations des données à différents niveaux d'abstraction. Cette

technique émergente a profondément transformé le domaine de la reconnaissance faciale depuis son avènement en 2014, avec la percée de la méthode DeepFace. Depuis, la reconnaissance faciale profonde, fondée sur une architecture hiérarchique qui convertit les pixels d'image en une représentation invariante du visage, a considérablement amélioré les performances des systèmes et favorisé leur adoption dans des applications réelles[3].

### 1.2.2.1 Concepts De Base

Un système complet de reconnaissance faciale profonde repose généralement sur trois modules essentiels[3] :

1. **Détecteur de visages** : il permet de localiser les visages dans une image ou une vidéo.
2. **Détecteur de repères faciaux** : il aligne les visages détectés selon des coordonnées canoniques normalisées.
3. **Module de reconnaissance faciale (RF)** : il traite les visages alignés pour l'identification ou la vérification.

### 1.2.2.2 Composantes De La Reconnaissance Faciale

Un système de reconnaissance faciale comprend généralement trois étapes principales :

- **Prétraitement facial** : représenté par  $P$ , il permet de corriger les variations intra-personnelles comme les poses, les expressions, l'éclairage ou les occlusions.
- **Extraction de caractéristiques** : représentée par  $F$ , elle encode les informations identitaires contenues dans les images faciales.
- **Correspondance** : représentée par  $M$ , c'est un algorithme qui calcule la similarité entre deux visages pour déterminer s'ils correspondent à la même personne.

Cette relation peut être formalisée comme suit :

$$M [F (P_i(I_i)), F (P_j(I_j))]$$

où  $I_i$  et  $I_j$  sont deux images de visage,  $P_i$  et  $P_j$  les traitements associés à chacune,  $F$  l'extraction des caractéristiques, et  $M$  l'algorithme de comparaison ou de classification[3].

### 1.2.3 Avantages et Inconvénients de Reconnaissance Faciale

La reconnaissance faciale présente divers avantages, à commencer par un renforcement de la sécurité. La banalisation des cyberattaques en est la preuve : les techniques traditionnelles d'identification numérique, comme les mots de passe, ont leurs limites dans la protection des informations sensibles. Comme expliqué précédemment, la reconnaissance faciale sécurise les connexions en permettant d'authentifier ou d'identifier une personne. À l'échelle nationale, la technologie d'intelligence artificielle peut même aider à identifier des terroristes ou d'autres criminels.

À côté de cela, le traitement des données est très rapide, presque instantané, ce qui présente un intérêt réel pour les entreprises. Autre avantage, et non des moindres : sa fiabilité. Il est en effet difficile de tromper la technologie. Un dispositif de reconnaissance faciale est par ailleurs simple d'utilisation et s'intègre assez facilement dans une organisation, puisque la majorité des solutions sont compatibles avec les logiciels de sécurité actuels.

Cependant, cette technologie présente également des limites qu'il convient de prendre en compte. La première concerne la protection de la vie privée. Le fait que les visages soient capturés, stockés et analysés sans consentement explicite soulève de sérieuses questions éthiques et juridiques. L'utilisation abusive ou non réglementée de ces données peut entraîner des violations de la liberté individuelle.

De plus, bien que les algorithmes aient connu des progrès considérables, ils peuvent encore présenter des biais liés à la couleur de peau, au genre ou à l'âge, ce qui affecte la précision du système selon les populations. Cela peut conduire à des faux positifs ou négatifs, particulièrement problématiques dans des contextes sensibles comme la sécurité ou la justice.

Enfin, le déploiement massif de ces technologies peut engendrer une forme de surveillance généralisée, où les individus sont constamment identifiés et suivis dans les espaces publics, menaçant ainsi les principes de liberté et d'anonymat[4].

## 1.3 Traitement D'images

Le traitement d'images constitue une étape cruciale dans un système de reconnaissance faciale. Il permet d'améliorer la qualité des images faciales en réduisant le bruit et en norma-

lisant les données, avant leur exploitation par un réseau de neurones. Cette phase garantit une meilleure robustesse et précision du modèle de reconnaissance[5].

### 1.3.1 Phases Du Traitement D'image

Le traitement des images est une étape essentielle des systèmes de vision par ordinateur. Avant de pouvoir appliquer un modèle de reconnaissance faciale, il est nécessaire de préparer les images de manière appropriée, car l'image passe par plusieurs Phases[6].

#### 1.3.1.1 Conversion En Niveaux De Gris

Les photographies sont souvent prises en couleur (RVB), mais ce n'est pas toujours nécessaire. L'image a été manipulée à un niveau gris afin de minimiser la complexité tout en tenant des données faciales importantes, pour simplifier les calculs et accélérer la thérapie sans perdre les caractéristiques faciales[6].

#### 1.3.1.2 Égalisation D'histogramme

Ce processus ajuste la luminosité des images, selon le photographe, le visage est mal éclairé (trop sombre ou trop clair), le système d'égalisation aide à faire ressortir les caractéristiques faciales d'une plus grande luminosité, en particulier dans des conditions d'éclairage difficiles[6].

#### 1.3.1.3 Réduction Du Bruit

Les images de mauvaise qualité contiennent souvent du bruit (petits pixels parasites). On utilise des filtres spécifiques :

- 1. Filtre bilatéral** : réduit le bruit tout en conservant les contours.
- 2. Filtre médian** : efficace contre le bruit de type « sel et poivre ».
- 3. Filtre gaussien** : floute légèrement l'image pour la lisser.

Pour rendre l'image plus propre et améliorer la fiabilité de la reconnaissance[6].

### 1.3.1.4 Alignement Du Visage

Tous les visages ne sont pas alignés correctement dans les images (inclinaison, rotation). On aligne le visage en recentrant les yeux et la bouche pour qu'ils soient dans des positions standards, pour garantir une cohérence spatiale entre les visages et faciliter l'apprentissage automatique[6].

### 1.3.1.5 Redimensionnement

Les réseaux de neurones exigent des dimensions fixes en entrée (par exemple 224×224 pixels). On redimensionne toutes les images à cette taille, pour respecter le format d'entrée des modèles de deep learning[6].

### 1.3.1.6 Normalisation

Les valeurs de pixels (généralement entre 0 et 255) sont normalisées entre 0 et 1 (ou -1 à 1). Cela stabilise l'entraînement et évite les biais, pour améliorer la convergence des modèles et assurer une meilleure généralisation[6].

## 1.4 La Basse Résolution

La basse résolution fait référence à un niveau inférieur de détail ou de clarté dans une image, une vidéo ou un affichage. Cela signifie que le contenu apparaît moins net et moins détaillé que ses homologues en haute résolution. Cela peut être dû à un nombre inférieur de pixels ou à une palette de couleurs limitée utilisée pour représenter l'image[7].



FIGURE 1.2 – La Basse Résolution

### 1.4.1 Défis De La Reconnaissance Faciale En Basse Résolution

La reconnaissance faciale en basse résolution est l'un des défis les plus importants dans le domaine de la vision par ordinateur, en particulier dans les applications réelles comme la vidéosurveillance, les systèmes de sécurité, ou les appareils mobiles. Voici une explication plus approfondie des principaux obstacles[8] :

#### 1.4.1.1 Perte D'informations Visuelles

Dans une image de faible résolution, le visage est représenté par un nombre très limité de pixels. Cela a plusieurs conséquences :

1. Perte des détails fins : les rides, la texture de la peau, les cicatrices ou les grains de beauté deviennent invisibles.
2. Effacement des structures faciales subtiles : les contours des lèvres, les plis autour des yeux ou la courbure des sourcils sont difficilement perceptibles.
3. Uniformisation des visages : les différences entre les visages deviennent moins marquées, ce qui augmente les risques de confusion entre individus, en particulier ceux ayant des caractéristiques similaires (même sexe, même âge ou origine ethnique)[8].

### 1.4.1.2 Variabilité Des Conditions D'acquisition

Dans les environnements réels, les conditions d'acquisition sont rarement idéales. Cela entraîne plusieurs types de dégradations :

#### 1. **Éclairage variable :**

- Une lumière trop forte peut entraîner une saturation de certaines zones de l'image.
- Une lumière trop faible peut cacher des détails importants du visage.
- Les ombres peuvent altérer ou dissimuler certaines parties du visage.

2. **Angles de prise de vue et poses variées :** Le visage n'est pas toujours capturé de face. Il peut être incliné, partiellement tourné, ou vu du dessus/dessous. Ces variations réduisent la quantité d'informations visibles et déforment la perception du visage.

3. **Présence d'occlusions :** Des éléments comme des masques, des lunettes, des cheveux, ou des mains peuvent cacher une partie du visage. Ces obstructions masquent des zones clés (bouche, nez, yeux), rendant l'identification plus difficile.

4. **Flou dû au mouvement (Motion Blur) :** Lorsqu'une personne bouge ou que la caméra n'est pas stable, les contours deviennent flous. Ce phénomène est fréquent dans les vidéos de surveillance où les sujets sont souvent en mouvement rapide[8].

### 1.4.1.3 Manque De Bases De données Adaptées

Le développement d'algorithmes efficaces repose en grande partie sur la qualité et la diversité des bases de données utilisées pour l'apprentissage :

1. La plupart des bases publiques (ex. : LFW, VGGFace, CelebA) contiennent des images de haute qualité, bien éclairées et bien cadrées.
2. Les images vraiment basse résolution, capturées dans des environnements réels, sont rares.
3. Beaucoup d'images dites « basse résolution » sont en réalité des images haute résolution redimensionnées artificiellement, ce qui ne reflète pas fidèlement les conditions du terrain[8].

### 1.4.2 Approches La Reconnaissance Faciale En Basse Résolution

La reconnaissance faciale en basse résolution est un défi majeur en vision par ordinateur, car les images de faible qualité contiennent moins d'informations utiles. Pour surmonter ces difficultés, plusieurs approches ont été proposées. Ces méthodes visent à améliorer la capacité des systèmes à reconnaître des visages même lorsque les images sont floues, pixelisées ou de petite taille[9].

#### 1.4.2.1 Approches Basées Sur L'Apprentissage De Représentations Invariantes

Plutôt que de se concentrer sur l'amélioration de la résolution des images, ces méthodes cherchent à apprendre des représentations faciales robustes aux variations de résolution. Elles visent à extraire des caractéristiques qui restent fiables même lorsque la qualité de l'image est faible.

Les techniques utilisées sont les suivantes :

1. Extraction de caractéristiques multi-résolution : cette technique permet de capturer des informations faciales indépendamment de la résolution de l'image.
2. Réseaux de neurones profonds spécifiques : il s'agit de modèles entraînés spécialement pour identifier les visages, même lorsqu'ils sont en basse résolution.
3. Adaptation entre différentes résolutions : ces méthodes utilisent le transfert d'apprentissage afin de généraliser les modèles aux images basse résolution.

Cependant, ces approches nécessitent généralement un grand volume de données pour être efficaces. De plus, elles ne donnent pas toujours de bons résultats lorsque les images basse résolution sont capturées dans des environnements réels, où les conditions sont très variables et souvent difficiles[9].

## 1.5 Super-résolution

La super-résolution est une technique du traitement d'images qui vise à reconstruire une image de haute résolution (HR) à partir d'une ou plusieurs images de basse résolution (LR). Elle permet d'augmenter la taille et la qualité visuelle d'une image tout en tentant de restaurer les détails perdus.

Dans le cadre de la reconnaissance faciale, la super-résolution est souvent utilisée pour améliorer la netteté des visages flous ou pixelisés capturés par des caméras de surveillance[10].

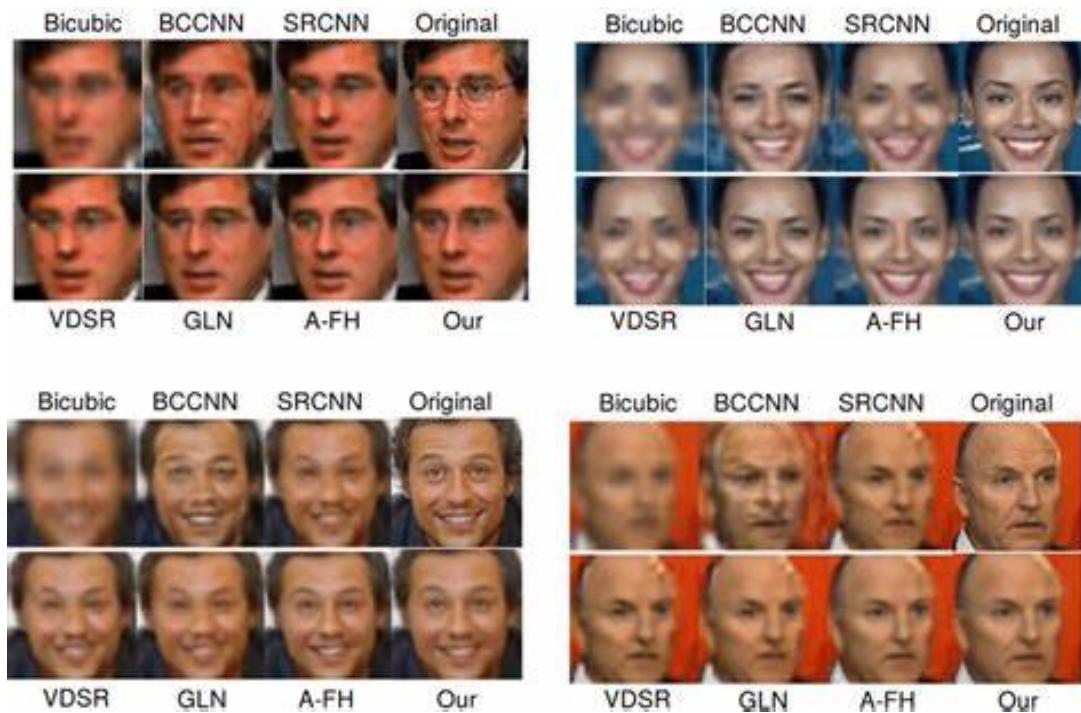


FIGURE 1.3 – La Super-résolution

### 1.5.1 Objectif De Super-Résolution

L'objectif principal de la super-résolution est d'améliorer la clarté des images à basse résolution afin de faciliter la reconnaissance faciale avec une meilleure précision. Tout d'abord, la super-résolution vise à rendre les traits du visage plus visibles, tels que les yeux, le nez, la bouche, et les contours du visage qui sont souvent flous ou peu définis dans les images de basse résolution. Cette amélioration des détails permet au système de mieux distinguer les caractéristiques essentielles du visage.

Ensuite, elle facilite l'extraction des caractéristiques discriminantes, c'est-à-dire les attributs qui permettent de différencier un visage d'un autre. Plus l'image est nette et détaillée, plus les modèles et algorithmes peuvent identifier avec précision ces caractéristiques, ce qui augmente l'efficacité des processus de classification et de reconnaissance.

De plus, la super-résolution contribue à améliorer les performances des modèles de reconnaissance faciale. En fournissant aux modèles des images de meilleure qualité et plus détaillées, elle leur permet d'apprendre plus efficacement et d'augmenter le taux de succès dans l'iden-

tification des individus. Une meilleure qualité d'image réduit les erreurs et accroît la fiabilité globale du système.

Enfin, la super-résolution permet une comparaison plus précise entre une image basse résolution (appelée sonde) et une image haute résolution (appelée galerie) présente dans la base de données. En améliorant la qualité de l'image basse résolution, la correspondance entre les deux images devient plus fiable, ce qui renforce la capacité du système à reconnaître correctement les personnes dans des conditions réelles[11].

### 1.5.2 Méthodes Super-Résolution

La super-résolution permet d'augmenter la qualité des images basse résolution pour améliorer la reconnaissance faciale. On distingue principalement deux types de méthodes[12] :

#### 1.5.2.1 Méthodes Classiques

Les méthodes classiques de super-résolution sont basées sur des modèles mathématiques déterministes. Contrairement aux approches modernes par apprentissage, elles ne nécessitent pas de phase d'entraînement préalable, ce qui les rend simples à implémenter et rapides à exécuter. Voici les principales méthodes utilisées [12] :

##### 1. Interpolation Bilinéaire

Il s'agit de la méthode la plus simple et la plus rapide. Elle estime la valeur d'un pixel en faisant une moyenne pondérée des quatre pixels voisins. Bien qu'elle soit peu coûteuse en calcul, elle tend à produire des images floues, avec une perte importante de détails fins.

##### 2. Interpolation Bicubique

Cette méthode améliore la qualité par rapport à l'interpolation bilinéaire en prenant en compte 16 pixels voisins (au lieu de 4). Elle produit des images plus lisses et avec moins d'artefacts, ce qui en fait un bon compromis entre rapidité et qualité.

##### 3. Interpolation Lanczos

La méthode Lanczos repose sur une fonction de convolution utilisant une fonction *sinc* tronquée. Elle est considérée comme une interpolation avancée qui offre un excellent compromis entre netteté et réduction des artefacts visuels. Elle est souvent utilisée dans

les logiciels professionnels de traitement d'image.

### 1.5.2.2 Méthodes Modernes Basées Sur L'apprentissage Profond

Les méthodes modernes de super-résolution d'image s'appuient sur l'apprentissage profond, notamment les réseaux de neurones convolutifs (CNN), pour apprendre à reconstruire des détails réalistes à partir de nombreuses paires d'images basse résolution (LR) et haute résolution (HR). Voici les principales approches utilisées[12] :

#### 1. SRCNN (Super-Resolution Convolutional Neural Network) :

C'est la première méthode utilisant un réseau de neurones convolutifs pour la super-résolution. Le réseau, relativement peu profond, apprend la correspondance entre les images LR et HR. Il commence par une interpolation bicubique, suivie de convolutions pour améliorer la qualité de l'image.

#### 2. SRGAN (Super-Resolution Generative Adversarial Network) :

Cette approche introduit les réseaux adverses génératifs (GANs) dans la super-résolution. Elle comprend un générateur qui produit des images haute résolution à partir d'entrées LR, et un discriminateur qui tente de distinguer les vraies images HR des images générées. Cela permet de produire des images plus réalistes, avec des détails texturaux convaincants.

#### 3. EDSR (Enhanced Deep Super-Resolution Network) :

EDSR améliore les performances des réseaux de super-résolution en supprimant les couches de normalisation, ce qui permet une meilleure préservation des détails. Il utilise une architecture plus profonde avec une plus grande capacité d'apprentissage, ce qui en fait l'un des modèles les plus performants.

#### 4. LapSRN (Laplacian Pyramid Super-Resolution Network) :

Cette méthode repose sur une reconstruction progressive de l'image à travers une pyramide de Laplace. À chaque niveau, des détails sont ajoutés de manière hiérarchique, ce qui permet une meilleure précision tout en maintenant une vitesse d'exécution satisfaisante[12].

### 1.5.3 Les Avantages De La Super-Résolution

La super-résolution joue un rôle essentiel dans l'amélioration des performances des systèmes de reconnaissance faciale, en particulier lorsque les images disponibles sont de faible qualité. Voici les principaux avantages apportés par cette technique [13] :

#### 1. Amélioration de la visibilité des traits faciaux

La super-résolution permet de restaurer des détails fins du visage comme les yeux, le nez, la bouche ou encore les contours du visage. Cela améliore la clarté de l'image et facilite l'identification visuelle.

#### 2. Amélioration de l'extraction des caractéristiques discriminantes

En augmentant la qualité de l'image, il devient plus facile pour les algorithmes d'extraction de repérer et d'analyser des caractéristiques uniques à chaque individu, ce qui renforce la précision de la reconnaissance.

#### 3. Amélioration des performances des modèles de reconnaissance

Les modèles de deep learning (tels que les CNNs) donnent de meilleurs résultats lorsque les images sont de bonne qualité. La super-résolution améliore donc directement les scores de précision et de rappel des systèmes de reconnaissance.

#### 4. Comparabilité entre images basse et haute résolution

Dans un système typique, une image de la base (HR) est comparée à une image captée sur le terrain (LR). La super-résolution permet d'amener l'image LR à un niveau de qualité comparable à celui de la galerie HR, facilitant ainsi la tâche de vérification.

## 1.6 Conclusion

Grâce à cette étude, nous avons pu mettre en lumière un problème critique souvent négligé dans le domaine de la reconnaissance faciale en basse résolution. Nous avons montré que les techniques basées sur les réseaux de neurones convolutifs (CNN) offrent une solution efficace pour traiter les images de faible qualité. Les résultats indiquent que l'utilisation de CNN permet d'améliorer la précision de la reconnaissance faciale dans des conditions réelles, ouvrant ainsi la voie à de nouvelles avancées dans ce domaine.

# Chapitre 2

## Apprentissage Profond Et Apprentissage Avec Peu D'exemples.

### 2.1 Introduction

L'apprentissage automatique (machine learning) et l'apprentissage profond (deep learning) sont des domaines de l'intelligence artificielle permettant aux machines d'apprendre à partir de données sans programmation explicite pour chaque tâche. L'apprentissage profond, un sous-domaine du machine learning, utilise des réseaux neuronaux artificiels pour résoudre des problèmes complexes en plusieurs couches de traitement non linéaire. Il est utilisé dans des applications telles que la reconnaissance d'images, la vision par ordinateur, la traduction automatique, la détection de fraudes, et la reconnaissance vocale et faciale. Dans ce chapitre, nous nous concentrerons sur l'histoire, les applications et les différentes architectures de l'apprentissage automatique et de l'apprentissage profond.

### 2.2 L'apprentissage Automatique

L'apprentissage automatique, est une discipline de l'intelligence artificielle qui permet aux machines d'apprendre et de s'améliorer à partir de données, sans être explicitement programmées pour chaque tâche. Les algorithmes de Machine Learning analysent des en-

sembles de données pour identifier des modèles et formuler des prédictions plus précises[14].

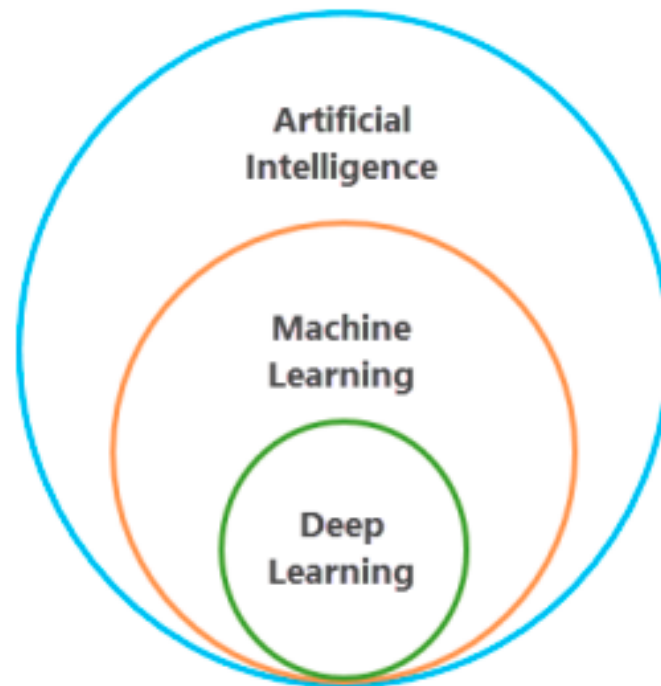


FIGURE 2.1 – Les sous-couches de l'intelligence artificielle

### 2.2.1 L'approches principales

Il existe deux approches principales :

#### 2.2.1.1 L'apprentissage Supervisé

L'apprentissage supervisé est une méthode d'entraînement des modèles dans laquelle les données sont étiquetées, c'est-à-dire que chaque exemple d'apprentissage est associé à une réponse ou à une sortie attendue. Cette approche permet au modèle d'apprendre à faire correspondre des entrées à des sorties connues.

Le principe de cette méthode consiste à entraîner le modèle à partir d'exemples annotés afin qu'il puisse apprendre les relations entre les données d'entrée et les résultats attendus. Une fois l'apprentissage terminé, le modèle est capable de prédire la sortie pour de nouvelles données qu'il n'a jamais vues auparavant.

L'objectif principal de l'apprentissage supervisé est de construire un modèle capable de prédire correctement les résultats ou de classer les données avec précision dans des situations

nouvelles.

Parmi les applications courantes de cette approche, on peut citer la reconnaissance d'images (comme l'identification de visages ou d'objets), les systèmes de prédiction (par exemple, les prévisions météorologiques ou les diagnostics médicaux), ainsi que les systèmes de recommandation (tels que ceux utilisés sur les sites de commerce électronique ou les plateformes de streaming)[15].

### 2.2.1.2 L'apprentissage Non Supervisé

L'apprentissage non supervisé est une méthode dans laquelle les données d'entraînement ne sont pas étiquetées, c'est-à-dire qu'il n'existe aucune sortie connue à apprendre. L'objectif est de permettre à l'algorithme de découvrir par lui-même des structures cachées dans les données.

Le principe de cette méthode repose sur l'exploration autonome des données. L'algorithme identifie des schémas, des regroupements ou des relations intrinsèques, sans être guidé par des exemples annotés. Il cherche à révéler l'organisation naturelle des données ou à détecter des groupes d'éléments partageant des similarités.

L'objectif principal de l'apprentissage non supervisé est d'extraire des groupes ou des classes d'éléments similaires, ce qui permet une meilleure compréhension, organisation ou visualisation des ensembles de données complexes.

Parmi les applications courantes de cette approche, on trouve le regroupement (ou clustering), utilisé notamment pour la segmentation de clients ou le tri automatique d'images. On retrouve également l'estimation de densité, qui permet d'évaluer la probabilité qu'un point appartienne à une distribution donnée, et la réduction de dimension, dont les méthodes comme l'ACP (Analyse en Composantes Principales) ou t-SNE permettent de simplifier les données tout en conservant l'essentiel de l'information[16].

## 2.3 L'apprentissage Profond

L'apprentissage profond (Deep Learning ou DL) est une sous-branche de l'apprentissage automatique, lui-même appartenant au domaine de l'intelligence artificielle (IA). Il s'inspire du fonctionnement du cerveau humain pour permettre aux machines d'apprendre à partir de

l'expérience et des données.

Le Deep Learning repose principalement sur l'utilisation des réseaux de neurones artificiels profonds, constitués de plusieurs couches cachées entre l'entrée et la sortie du réseau. Chaque couche transforme progressivement les données d'entrée en représentations de plus en plus abstraites.

Le terme « profond » fait référence à la profondeur du réseau, c'est-à-dire au nombre de couches qu'il contient. Plus un réseau est profond, plus il peut modéliser des relations complexes et apprendre des caractéristiques fines dans les données[17].

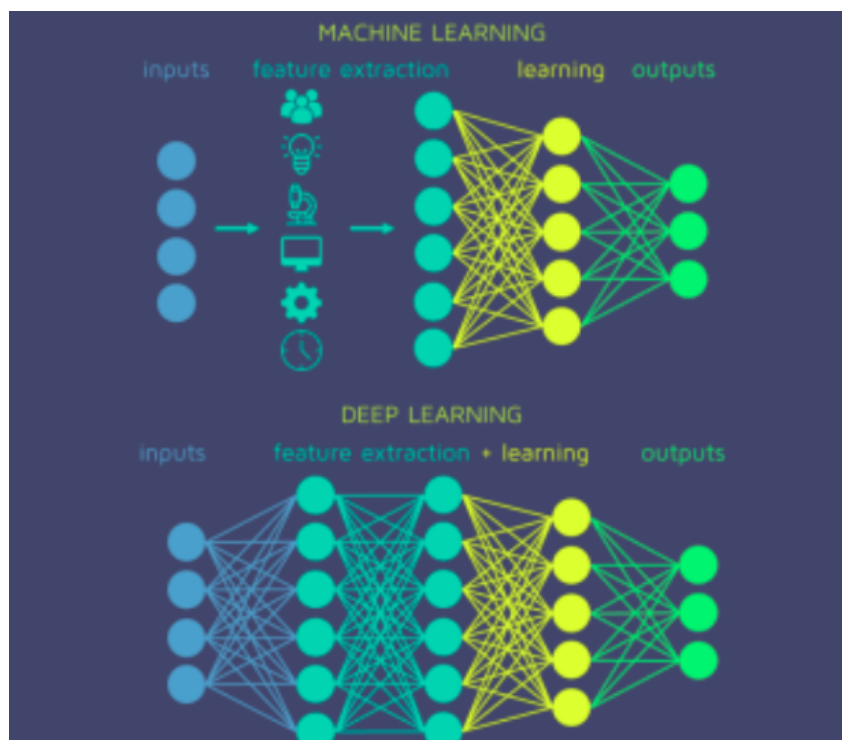


FIGURE 2.2 – La différence entre l'apprentissage automatique (ML) et l'apprentissage profond (DL)

La différence entre l'apprentissage automatique (ML) et l'apprentissage profond (DL) réside dans la manière dont chaque algorithme apprend et dans la complexité des modèles. L'apprentissage automatique utilise des algorithmes qui apprennent à partir de données en plusieurs couches de traitement non linéaire, tandis que l'apprentissage profond utilise des réseaux de neurones artificiels pour résoudre des problèmes complexes en s'appuyant sur plusieurs couches de traitement non linéaire. L'apprentissage profond nécessite généralement de grandes bases de données pour entraîner les machines et des moyens de calculs rapides, comme

l'utilisation de cartes graphiques pour effectuer des calculs massifs rapidement[17][18].

### 2.3.1 Les Fondamentaux De L'apprentissage Profond

L'apprentissage profond repose sur plusieurs concepts fondamentaux qui rendent possible l'entraînement de modèles puissants et efficaces à partir de grandes quantités de données. Ces éléments sont[19] :

#### 1. Réseaux de neurones artificiels multicouches :

Ils constituent la structure de base du Deep Learning. Grâce à plusieurs couches cachées interconnectées, ces réseaux peuvent modéliser des relations complexes et extraire automatiquement des caractéristiques pertinentes à partir des données.

#### 2. Algorithmes d'apprentissage discriminants :

Développés depuis les années 1980, ces algorithmes sont essentiels aussi bien pour l'apprentissage supervisé (avec données étiquetées) que non supervisé. Ils permettent aux modèles de distinguer les classes ou de détecter des structures dans les données.

**3. Puissance de calcul des machines :** Les progrès en matière de processeurs graphiques (GPU), de calcul parallèle et de stockage ont permis de traiter d'énormes volumes de données et d'entraîner des modèles très profonds en un temps raisonnable.

**4. Bases de données volumineuses :** L'accès à de vastes ensembles de données est un élément clé pour entraîner des réseaux profonds. Plus le volume et la diversité des données sont grands, plus le modèle peut apprendre efficacement et généraliser à de nouveaux cas[19].

### 2.3.2 L'apprentissage Profond Dans La Reconnaissance Faciale En Basse Résolution

L'apprentissage profond joue un rôle essentiel dans le domaine de la reconnaissance faciale en basse résolution (LRFR), notamment pour [20] :

#### 1- Amélioration et reconstruction de visages à basse résolution

apprentissageLes approches approchesprofond comme les réseaux de super - résolution

permettent la reconstruction d'images faciales de haute qualité à partir de visages flous ou de très faible résolution (par exemple, 20x16 pixels). Les réseaux de super-résolution permettent de reconstruire des images faciales de haute qualité à partir de visages flous ou de très faible résolution (par exemple, 20x16 pixels). Cela facilite l'extraction des caractéristiques distinctives, même dans les cas où l'image d'origine est gravement détériorée, ce qui est courant dans les systèmes de vidéosurveillance. même dans les cas où l'image d'origine est gravement détériorée, ce qui est courant dans les systèmes de vidéosurveillance[20].

### 2- Détermination des caractéristiques résilientes aux variations de résolution

réseaux neuronaux (CNN) Les réseaux sont utilisés convolutifs (CNN) sont utilisés pour extraire des représentations faciales invariantes pour la résolution. Cela permet de garantir une reconnaissance fiable même dans les cas où la résolution de l'image de la base de données et du visage à identifier diffère. cas où la résolution de l'image de la base de données et du visage à identifier diffère[20].

### 3- Vérification et identification des identités dans des circonstances incontrôlées des circonstances

En combinant les techniques de super-résolution et de reconnaissance faciale, des modèles d'apprentissage profond, comme CSRI (Complement Super-Resolution and Identity), sont capables d'identifier des individus à partir d'images prises dans des environnements réels, où la qualité est souvent faible (lumière, pose, occlusion, etc.). Ces modèles fournissent une identification précise, utile dans les systèmes de sécurité et de surveillance. diagnostic, utile dans les systèmes de sécurité et de surveillance[20].

## 2.3.3 Types De Réseaux Neuronaux Dans Le Apprentissage Profond :

### 1. Réseaux de neurones récurrents (RNN) :

Les réseaux de neurones récurrents (RNN) sont conçus pour traiter des données séquentielles, où l'ordre des informations est important. On les utilise principalement pour des tâches

comme[21] :

- l'analyse de texte (ex. : traduction, résumé),
- les séries temporelles (ex. : données financières),
- la reconnaissance vocale.

Ces réseaux ont la capacité de mémoriser les informations précédentes grâce à une boucle interne, ce qui leur permet de comprendre le contexte dans une séquence.

Cependant, les RNN classiques ont des limites : ils oublient souvent les informations trop anciennes. Pour résoudre cela, des variantes plus avancées comme les LSTM (Long Short-Term Memory) et les GRU (Gated Recurrent Unit) ont été développées[21].

### **2. Transformers :**

Les Transformers sont des modèles très puissants utilisés dans le domaine de l'intelligence artificielle, notamment pour le traitement du langage naturel (NLP). Ce type de réseau a révolutionné la manière dont les machines comprennent et génèrent le langage humain.

Contrairement aux anciens modèles séquentiels comme les RNN, les Transformers traitent toute la séquence de texte en une seule fois, ce qui les rend plus rapides et plus efficaces[21].

### **3. Réseaux neuronaux convolutifs (CNN) :**

Le réseau de neurones convolutifs, ou CNN pour faire court, est un type spécialisé de modèle de réseau de neurones conçu pour travailler avec des données d'images bidimensionnelles, bien qu'ils puissent être utilisés avec des données unidimensionnelles et tridimensionnelles. Ces réseaux sont capables d'apprendre à extraire des caractéristiques locales, c'est-à-dire des structures qui se répètent à travers l'image[21].

## **2.3.4 Applications De L'apprentissage Profond**

### **Le deep Learning est utilisé dans de nombreux domaines :**

reconnaissance d'image, traduction automatique, voiture autonome, diagnostic médical, recommandations personnalisées, modération automatique des réseaux sociaux, prédiction financière et trading automatisé, identification de pièces défectueuses, détection de malwares ou de fraudes, chatbots

(agents conversationnels), exploration spatiale, robots intelligents[22].

### 2.4 Apprentissage à Quelques Coups

L'apprentissage ponctuel est une approche de l'apprentissage automatique qui permet aux modèles de formation d'obtenir de bons résultats avec quelques exemples seulement, plutôt que de s'appuyer sur de grandes quantités de données spécifiques à une tâche. Contrairement à l'apprentissage supervisé, qui nécessite un nombre important d'échantillons étiquetés, les approches d'apprentissage à court terme permettent aux modèles d'apprentissage automatique de s'adapter à de nouvelles tâches avec un minimum de données d'apprentissage. L'apprentissage à partir de peu d'échantillons appartient à la catégorie plus large de l'apprentissage à partir d'échantillons [23] :

#### 1. Apprentissage à partir d'un petit nombre d'échantillons :

Cet apprentissage, également appelé (Few-Shot Learning), consiste à entraîner un modèle avec très peu d'exemples étiquetés. Contrairement aux méthodes classiques qui nécessitent de grandes quantités de données, cette approche vise à généraliser efficacement à partir de quelques échantillons seulement[23].

#### 2. Apprentissage en une seule fois :

L'apprentissage en une seule fois (One-Shot Learning) est une approche dans laquelle le modèle apprend à reconnaître une classe ou effectuer une tâche à partir d'un seul exemple étiqueté[23].

#### 3. Apprentissage à partir de zéro :

L'apprentissage à partir de zéro (Zero-Shot Learning) désigne une méthode où le modèle est capable de faire des prédictions sur des classes jamais vues pendant l'entraînement, sans disposer de données étiquetées pour ces nouvelles classes.

Ces approches tirent parti du méta-apprentissage et de l'apprentissage par transfert pour s'adapter à de nouvelles classes et à des tâches connexes en utilisant les connaissances pré-existantes issues du pré-entraînement. Des méthodes telles que les réseaux prototypiques et les représentations de l'espace d'intégration permettent d'améliorer la capacité d'un modèle à

traiter des données non structurées dans diverses tâches, notamment les tâches de traitement du langage naturel (NLP), la création de contenu et la réponse à des questions[23].

### 2.4.1 Les Fonctionne L'apprentissage à Quelques Coups

**1. Création d'un ensemble de supports** Un petit ensemble de données étiquetées (par exemple, une invite à la formation) est fourni pour la formation. Cet ensemble de données comprend des exemples fournis pour chaque classe.

**2. L'ensemble de requêtes est constitué de nouvelles données** Le modèle reçoit un ensemble de requêtes (échantillons de données inédites) et doit les classer correctement.

**3. Méta-apprentissage et apprentissage par transfert** Le méta-apprentissage enseigne aux modèles comment apprendre à partir de quelques exemples en s'adaptant à des tâches connexes. L'apprentissage par transfert permet d'exploiter les connaissances préexistantes pour reconnaître de nouvelles classes sans nécessiter de formation supplémentaire.

**4. Représentation de l'espace d'intégration** Le modèle cartographie les données d'entrée dans un espace d'intégration où les classes similaires sont regroupées. Des techniques telles que les réseaux prototypiques améliorent la classification en comparant les distances dans cet espace.

**5. Ajustement sur la tâche cible** Le modèle est ajusté avec un contexte supplémentaire pour mieux prédire les résultats sur l'ensemble de test[24].

### 2.4.2 Terminologies Liées à L'apprentissage à Quelques Coups

L'apprentissage à quelques coups est une technique où un modèle apprend à reconnaître de nouvelles classes avec seulement quelques exemples. Voici les principales terminologies utilisées [25] :

#### 1. k-ways :

Il indique le nombre de classes qu'un modèle doit distinguer ou reconnaître. (2 voies signifie que le modèle doit classifier ou générer des exemples pour 2 classes).

#### 2. k-shots :

Il indique le nombre d'échantillons par classe disponibles lors de la formation ou de l'évaluation. (Un seul échantillon signifie qu'un échantillon pour chaque classe est fourni au modèle).

### 3. Support Set (S) :

Il s'agit de l'ensemble des données fournies au modèle pour l'apprentissage. Sa taille est généralement donnée par la formule :

$$S = \text{nombre de voies (k-ways)} \times \text{nombre de tirs (k-shots)}.$$

### 2.4.3 Travaux Connexes

En raison de la faible qualité des images capturées dans les environnements réels tels que les systèmes de surveillance et les caméras de sécurité, la tâche de reconnaissance faciale devient particulièrement difficile. Cela a poussé de nombreux chercheurs à développer des approches spécifiques pour améliorer la reconnaissance des visages à partir d'images de basse résolution.

Ainsi, plusieurs méthodes ont été proposées au fil des années, allant des techniques classiques de transformation d'espace aux approches plus récentes basées sur l'apprentissage profond. Ces travaux visent soit à améliorer la qualité des images à l'aide de techniques de super-résolution, soit à rendre les modèles de classification plus robustes face à la basse résolution.

Les approches les plus performantes utilisent désormais des architectures de réseaux de neurones convolutifs (CNN), des modèles pré-entraînés comme VGG, ou encore des techniques d'apprentissage avec peu d'exemples (Few-Shot Learning). Par ailleurs, les méthodes de super-résolution telles que Bicubic, Lanczos et FSRCNN sont souvent intégrées au pipeline afin d'améliorer la qualité visuelle des images avant la reconnaissance.

Les principales méthodologies, bases de données utilisées, séquences traitées, et les évaluations obtenues par les travaux les plus pertinents dans ce domaine sont résumées dans le tableau suivant :

## Chapitre 2. Apprentissage Profond Et Apprentissage Avec Peu D'exemples

Référence	Base de données	Séquences / Résolution	Méthodologie	Évaluation
Biswas et al. (2012)[26]	FERET	20×20	Mappage d'espace couplé	Précision : 80,0 %
Zhu et al. (2016)[27]	CASIA-WebFace, SCface	16×16, 32×32	Super-résolution bicubique + CNN	Taux de reconnaissance : 85,0 %
Zhang et al. (2021)[28]	TinyFace	16×16, 32×32	GAN + VGG-Face	Exactitude : 91,7 %
Shi et al. (2021)[29]	TinyFace, CelebA	16×16	CSRI (reconnaissance + super-résolution conjointe)	Taux de reconnaissance : 91,7 %
Zhou et al. (2020)[30]	VGGFace2 (dégradée)	24×24	VGG + Super-Résolution FSRCNN	Précision : 89,0 %
Yu et al. (2018)[31]	LFW-LR, SCface	16×16 vs 128×128	GAN perceptuel + attention spatiale	Gain de performance : +10 % (88,0 %)
Zheng et al. (2022)[32]	SCface	20×20 à 40×40	GAN supervisé + apprentissage métrique	Précision : >87,0 %
Zhang et al. (2023)[33]	LFW (dégradée)	16×16	Réseau à rétroaction adaptative	Taux de reconnaissance : 93,2 %
Cao et al. (2017)[34]	VGGFace, SCface	30×30	Transfert d'apprentissage + alignement des traits	Précision : 83,0 %

TABLE 2.1 – Travaux connexes sur la reconnaissance faciale en basse résolution.

### 2.5 Conclusion

En conclusion, l'apprentissage profond constitue une approche très performante lorsqu'il est alimenté par de grandes quantités de données annotées. Cependant, son efficacité peut se dégrader de manière significative lorsque les données sont limitées, bruitées ou de mauvaise qualité. Dans ce contexte, l'apprentissage avec un petit nombre d'échantillons (few-shot learning) s'impose comme une solution pertinente pour surmonter cette contrainte.

Cette approche devient particulièrement cruciale dans des domaines sensibles tels que la reconnaissance faciale à basse résolution, où il est souvent difficile de collecter des jeux de données vastes et variés. La capacité à apprendre efficacement à partir de quelques exemples permet non seulement d'améliorer la précision des systèmes dans des conditions réelles, mais aussi de renforcer leur robustesse face aux variations de résolution, d'éclairage ou de posture. Ainsi, l'intégration de techniques d'apprentissage avec peu de données représente une avancée significative vers des systèmes intelligents plus flexibles, adaptatifs et applicables dans des contextes réels contraignants.

# Chapitre 3

## Méthodologie

### 3.1 Introduction

La reconnaissance des visages dans des images à faible résolution est un défi important dans les applications de surveillance et de sécurité, où les images sont souvent floues ou bruitées, ce qui rend la tâche d'identification difficile et sujette aux erreurs. Ces images sont généralement traitées manuellement ou à l'aide de modèles entraînés sur des données de haute qualité, ce qui se traduit par des performances médiocres.

Dans ce chapitre, nous présentons une approche proposée basée sur des modèles multiples (CNN simple, modèle VGG, et réseau Few-Shot basé sur le réseau prototypique), utilisée pour la classification des visages en basse résolution, en testant l'effet super-résolution à l'aide de plusieurs méthodes (Bicubique, Lanczos, FSRCNN) afin d'améliorer la qualité des images avant la classification.

## 3.2 Approche proposée

Tout d’abord, nous avons collecté un ensemble d’images de visages à faible résolution et les avons prétraitées en les remettant à l’échelle et en les normalisant. Nous avons ensuite construit un modèle initial à l’aide d’un réseau neuronal simple (CNN) et l’avons entraîné avec une répartition de 80% pour l’entraînement et de 20% pour la validation.

Un modèle plus sophistiqué a ensuite été adopté en utilisant l’architecture VGG, suivi d’un modèle(Few-Shot Learning) basé sur un réseau prototype utilisant MobileNetV2 comme encodeur.

Après avoir observé la limitation des performances due à la mauvaise qualité de l’image, nous avons appliqué trois méthodes pour améliorer la résolution à l’aide de techniques de( super-résolution), à savoir : Bicubique, Lanczos et FSRCNN. Nous avons ensuite testé à nouveau les modèles sur les images optimisées.

Enfin, les résultats avant et après l’amélioration de l’image ont été comparés afin d’évaluer l’impact de chaque technique d’agrandissement sur les performances des différents modèles.

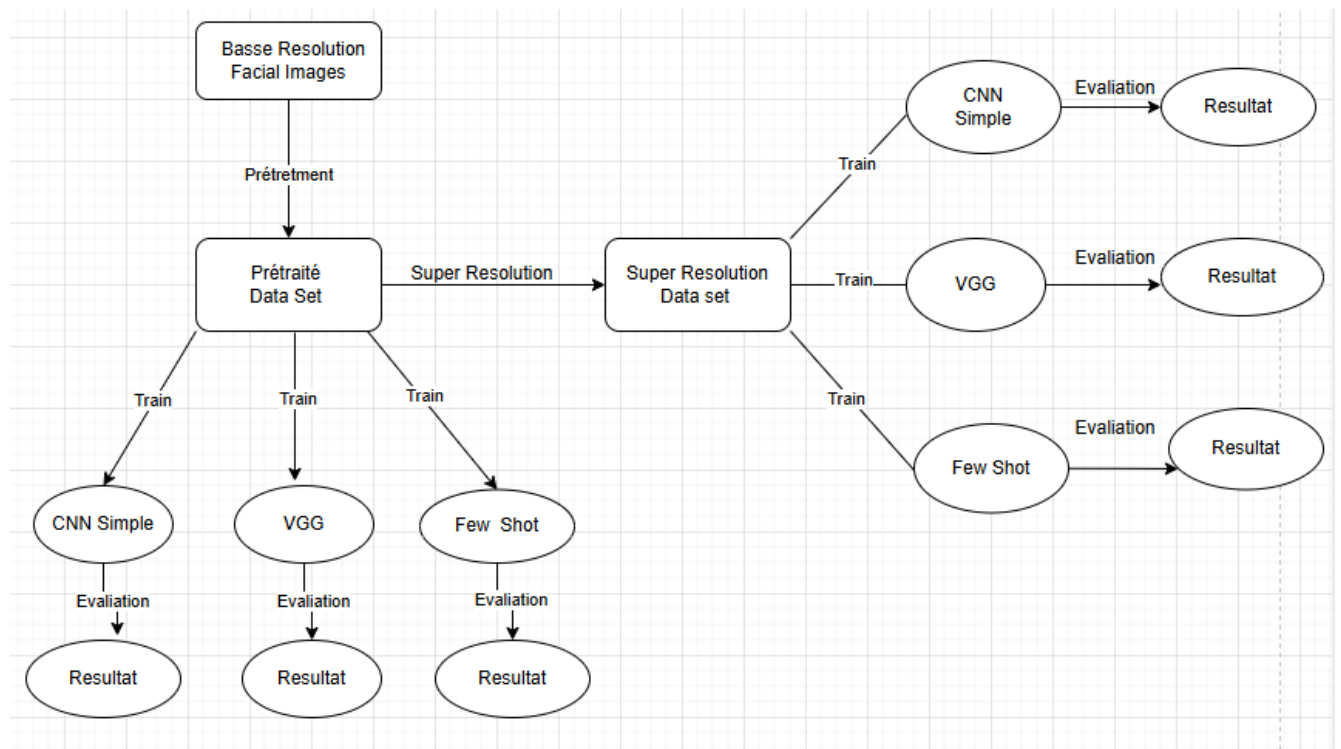


FIGURE 3.1 – L’architecture générale de notre approche proposée

### 3.2.1 Collecte des Données

Dans ce travail, nous avons utilisé le dataset TinyFace, une base de données de grande envergure dédiée à l'évaluation des performances des systèmes de reconnaissance faciale en basse résolution (LRFR). Ce choix a été motivé par la difficulté d'obtenir des images faciales en basse résolution capturées dans des environnements réels, présentant divers types de dégradations (flou, occlusion, variations d'éclairage, etc.).

#### **Jeu de données TinyFace**

Le dataset TinyFace contient 5 139 identités étiquetées, issues de 169 403 images faciales en basse résolution (taille moyenne :  $20 \times 16$  pixels). Ces images ont été collectées à partir de sources web publiques, dans des conditions de prise de vue non contrôlées, incluant des variations importantes de pose, éclairage, occlusion, et arrière-plan.

#### **Structure du dataset**

##### **1. Données étiquetées :**

. 2 570 identités sont utilisées pour l'apprentissage (train set).

. 2 569 identités sont réservées pour les tests (test set).

##### **2. Données non étiquetées (distracteurs) :**

. Environ 153 428 images ne sont associées à aucune identité. Elles sont utilisées comme éléments de distraction lors de la phase de test pour simuler une situation réaliste d'identification en 1 :N.

##### **3. Séparation pour le test 1 :N :**

. L'ensemble de test est divisé en sondes (requêtes à identifier) et en galerie (base de comparaison).

. La galerie comprend non seulement les images des identités testées, mais aussi toutes les images "trompeuses" (distracteurs), ce qui rend la tâche de reconnaissance plus difficile et réaliste.

### 3.2.2 Prétraitement

Les images de la base de données TinyFace sont des images de visages en basse résolution (LR) collectées sur Internet dans des conditions de prise de vue non contrôlées (éclairage, angle, occlusion, arrière-plan), le prétraitement est une étape cruciale pour améliorer la qualité des images d'entrée et améliorer la performance du modèle, nous avons adopté une série de techniques de prétraitement, qui sont les suivantes :

#### 1. Égalisation d'histogramme

Cette technique améliore le *contraste global* de l'image, en particulier pour les images en niveaux de gris. Elle permet de *faire ressortir les détails* du visage, même en cas d'éclairage faible ou inégal.

#### 2. Filtrage bilatéral

Le filtre bilatéral est utilisé pour *réduire le bruit* tout en *préservant les bords*. Contrairement aux filtres de lissage classiques, il permet d'adoucir l'image sans *détériorer les contours essentiels*, ce qui est particulièrement utile dans le cas des visages à faible résolution.

#### 3. Redimensionnement des images

Toutes les images ont été redimensionnées à une *taille fixe* (par exemple, 224×224 pixels), compatible avec l'entrée du modèle CNN utilisé (comme **MobileNetV2**). Ce redimensionnement est effectué tout en *conservant autant que possible les proportions* originales de l'image.

#### 4. Super-Résolution

Afin de *restaurer des détails perdus* dans les images de très basse résolution, une étape de *super-résolution* a été ajoutée avant l'entraînement du modèle. Plusieurs techniques ont été testées, notamment :

- **Interpolation bicubique** : méthode classique permettant une montée en résolution rapide, mais avec un gain limité en qualité perceptuelle.
- **Interpolation Lanczos** : variante plus sophistiquée de l'interpolation, qui permet une meilleure reconstruction des détails fins.
- **FSRCNN (Fast Super-Resolution Convolutional Neural Network)** : un réseau de neurones convolutif conçu spécifiquement pour effectuer la super-résolution de manière rapide et efficace. Ce modèle permet une *reconstruction plus précise des*

*textures* faciales et améliore significativement les performances du système de reconnaissance.

### 3.2.3 CNN simple

Un réseau neuronal convolutif (CNN) est un type particulier de réseau neuronal, principalement utilisé pour le traitement et la reconnaissance d'images. Il est conçu pour extraire automatiquement les caractéristiques des images, sans qu'il soit nécessaire de les identifier manuellement[35].

#### 3.2.3.1 Composants d'un CNN simple

##### 1. Couche de convolution :

Cette couche applique un petit filtre (ou noyau) à l'image d'entrée. Le filtre se déplace (en balayage) sur toute l'image pour détecter des caractéristiques locales telles que les bords, les coins, ou les lignes. Chaque filtre produit une carte de caractéristiques (feature map) qui représente une certaine information extraite de l'image[36].

##### 2. Couche d'activation :

Elle est généralement représentée par la fonction ReLU (Rectified Linear Unit). Son rôle est de transformer les valeurs négatives obtenues après la convolution en zéro, ce qui introduit de la non-linéarité dans le modèle. Cette propriété permet au réseau de modéliser des relations complexes entre les données d'entrée[36].

##### 3. Couche de pooling :

Cette couche sert à réduire la dimension des cartes de caractéristiques tout en conservant les informations les plus importantes. Le type de pooling le plus courant est le Max Pooling, qui conserve la valeur maximale dans chaque sous-région. Cela permet de réduire le nombre de paramètres et d'améliorer la vitesse de calcul du modèle[36].

##### 4. Couche entièrement connectée :

Après plusieurs couches de convolution et de pooling, les cartes de caractéristiques sont aplaties (flattening) en un vecteur. Ce vecteur est ensuite passé à une ou plusieurs couches

entièrement connectées qui fonctionnent comme un réseau de neurones classique. Ces couches permettent de faire la classification finale en apprenant les combinaisons des caractéristiques extraites[36].

### 5. Sortie avec Softmax :

La dernière couche du réseau utilise la fonction d'activation Softmax, qui convertit les scores obtenus en probabilités. Cela permet d'identifier la classe à laquelle appartient l'image parmi toutes les classes possibles. La classe ayant la probabilité la plus élevée est choisie comme prédiction finale[36].

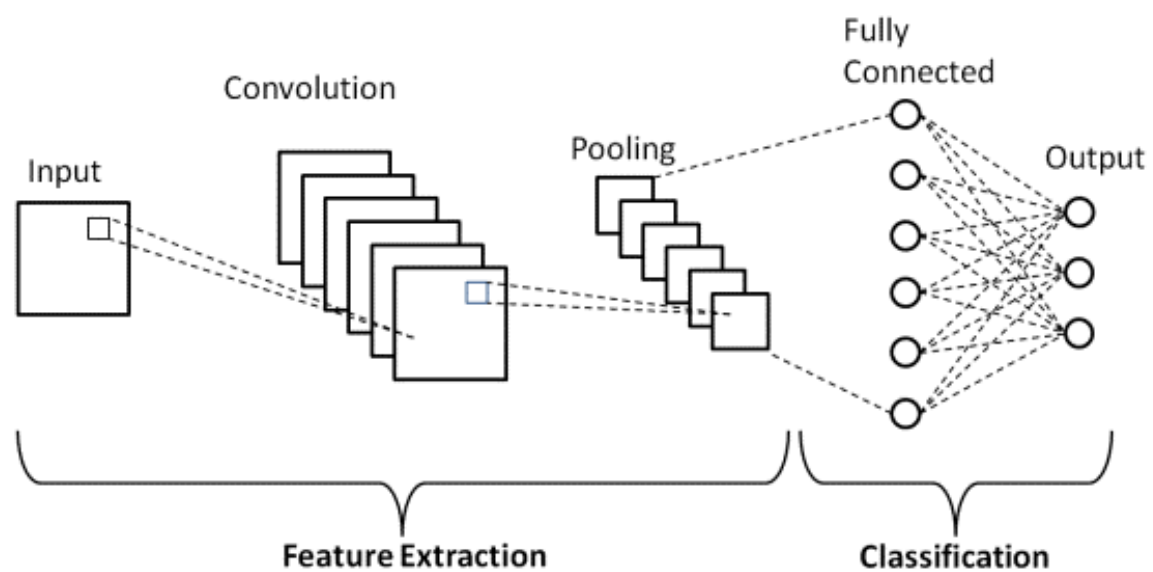


FIGURE 3.2 – L'architecture de modèle CNN simple

### Architecture du modèle :

Nous avons conçu un modèle CNN simple pour la reconnaissance de visages à basse résolution. Ce modèle est structuré autour de plusieurs couches de convolution, de pooling, de dropout, et de couches entièrement connectées.

### Méthode :

Le modèle CNN proposé est construit avec les spécifications suivantes :

Paramètre	Valeur
Taille de l'entrée	(64, 64, 3)
Nombre de couches de convolution	4
Nombre de couches de pooling	2
Taux de dropout	0.5
Fonction d'activation (intermédiaire)	ReLU
Fonction d'activation (sortie)	Softmax

TABLE 3.1 – Paramètres architecturaux du modèle CNN simple

- Les premières couches consistent en des opérations de convolution successives pour extraire les caractéristiques locales (textures, bords, angles) des visages.
- Les couches de pooling permettent de réduire progressivement les dimensions des cartes de caractéristiques tout en conservant les informations importantes.
- Les couches entièrement connectées permettent de transformer les caractéristiques extraites en un vecteur de décision.
- La couche de sortie utilise une fonction Softmax pour classifier chaque visage dans une des classes pré-définies.

### Hyper-paramètres du modèle CNN :

Paramètre	Valeur
Taux d'apprentissage	1e-4
Fonction de perte	Categorical Crossentropy
Métrique d'évaluation	Accuracy LOSS
Nombre d'époques	100
Taille du lot (batch)	32

TABLE 3.2 – Hyper-paramètres d'entraînement du modèle CNN simple

Le modèle est compilé avec l'optimiseur Adam et la fonction de perte Cross Entropy, adaptée aux tâches de classification multi classes. L'entraînement a été effectué sur 100 époques avec des lots de 32 images.

### Avantages du modèle CNN simple :

**1. Facilité d'implémentation et d'entraînement** Un modèle CNN simple est généralement constitué de peu de couches (convolution, activation, pooling, fully connected), ce qui le rend facile à implémenter, même pour les débutants en *deep learning*. De plus, il ne nécessite pas une grande puissance de calcul, ce qui permet son entraînement sur des machines classiques sans GPU puissant. Cette simplicité réduit également le temps de développement et facilite la compréhension du fonctionnement interne du modèle[37].

**2. Évaluation rapide des performances sur données dégradées** Grâce à sa structure légère, le modèle CNN simple peut être utilisé pour obtenir une première estimation des performances du système sur des images de faible qualité, comme les visages en basse résolution. Cela permet d'identifier rapidement les limites du système de reconnaissance et de poser des hypothèses sur les améliorations possibles à apporter[37].

**3. Base de référence pour des modèles plus complexes** Le CNN simple peut jouer le rôle de modèle de référence (*baseline*) dans une étude comparative. Il permet de mesurer l'impact réel de l'introduction de modèles plus avancés comme MobileNetV2, VGG-Face ou encore des réseaux avec super-résolution. Ainsi, il est utile pour justifier l'adoption de modèles plus lourds en comparant leurs résultats à ceux obtenus avec une architecture de base[37].

**4. Potentiel d'amélioration progressive** Un modèle CNN simple offre une bonne flexibilité pour l'expérimentation. Il peut être amélioré de manière progressive par l'ajout de nouvelles couches (comme des convolutions supplémentaires, normalisation batch, *dropout*, etc.), par des techniques d'optimisation (choix d'un meilleur optimiseur, ajustement du taux d'apprentissage), ou encore par l'application de méthodes de prétraitement et de super-résolution. Cela en fait un excellent point de départ pour construire des modèles plus robustes et précis[37].

### 3.2.4 VGG

Un réseau VGG est un type de réseau neuronal convolutif (Convolutional Neural Networks CNNs), proposé en 2014 par des chercheurs de l'Université d'Oxford (Visual Geometry Group VGG). Il est devenu l'un des modèles les plus populaires en vision par ordinateur, notamment après son succès dans la compétition ImageNet. Cette architecture est simple et efficace, car

elle repose sur l'utilisation répétée de couches convolutives de petite taille (3×3), suivies de couches d'assemblage et terminées par des couches entièrement connectées.

### 1. Couches convolutives

Le modèle utilise des filtres de taille 3×3, répétés plusieurs fois à chaque étape. Ces couches sont essentielles pour extraire progressivement des caractéristiques visuelles importantes à partir de l'image d'entrée. À travers ces convolutions successives, le modèle est capable de détecter différents éléments visuels, tels que les bords, les lignes, les contours, les angles, et même des traits distinctifs du visage comme les yeux, la bouche ou le nez. Cette étape joue un rôle crucial dans l'apprentissage des représentations spatiales complexes nécessaires à la reconnaissance faciale[38].

### 2. Couches de sous-échantillonnage

Après chaque groupe de couches convolutives, une couche de sous-échantillonnage, généralement un Max Pooling, est appliquée. Cette opération permet de réduire la taille des cartes de caractéristiques générées précédemment, ce qui a pour effet de diminuer le nombre total de paramètres à traiter dans les couches suivantes. En même temps, elle permet de conserver les informations essentielles tout en éliminant les détails redondants ou moins significatifs. Le Max Pooling contribue aussi à rendre le modèle plus robuste aux variations mineures dans les images, comme les légers déplacements ou bruits[38].

### 3. Couches entièrement connectées

Une fois les opérations de convolution et de sous-échantillonnage terminées, les cartes de caractéristiques sont transformées (aplaties) en un vecteur qui servira d'entrée aux couches pleinement connectées. Dans la structure classique du modèle, on retrouve une première couche pleinement connectée (FC1) composée de 4096 neurones, suivie d'une seconde (FC2) également de 4096 neurones, puis d'une troisième (FC3) contenant 2622 neurones, correspondant au nombre d'identités présentes dans le jeu de données d'entraînement initial (par exemple dans VGG-Face). Enfin, la couche de sortie applique une fonction d'activation Softmax, qui transforme les scores obtenus en probabilités, indiquant ainsi à quelle classe (ou identité) appartient le visage traité[38].

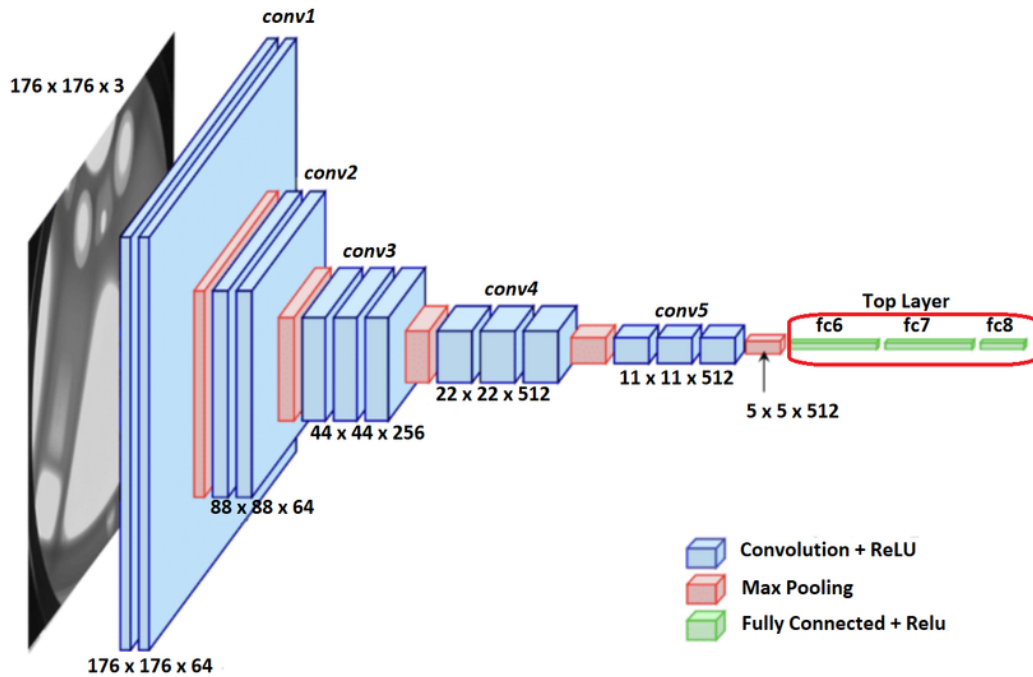


FIGURE 3.3 – Architecteur du modèle VGG

### 3.2.4.1 Objectifs de modèle VGG

Le modèle VGG a été développé spécifiquement pour remplir un ensemble de tâches essentielles dans le domaine de la reconnaissance des visages à l'aide de techniques d'apprentissage profond. Ces objectifs sont les suivants [39] :

#### 1. Identification des visages

L'objectif principal du VGG est d'identifier la personne dans l'image d'entrée, en la comparant à un ensemble d'images stockées dans une base de données connue. Le modèle analyse les caractéristiques du visage et extrait une représentation numérique qui est utilisée pour faire correspondre l'image à la bonne personne[39].

#### 2. Vérification des visages

Le modèle est également utilisé pour vérifier que deux visages correspondent, Dans ce contexte, le VGG extrait les représentations numériques des deux visages, puis les compare à l'aide d'une distance mathématique (par exemple, la distance euclidienne). Si la distance est faible, les deux visages sont considérés comme identiques ; si elle est grande, ils appartiennent à des personnes différentes[39].

### 3. Extraction des caractéristiques faciales

L'une des utilisations les plus puissantes de la VGG est sa capacité à convertir une image faciale en un vecteur scalaire connu sous le nom d'Embedding. Ce vecteur représente une « empreinte faciale » numérique, qui peut ensuite être utilisée pour [40] :

- Classification (à l'aide d'un autre modèle tel que SVM ou KNN).
- Comparaison entre les visages.
- Le regroupement de visages similaires.
- ou même l'apprentissage en quelques clics.

#### Hyper-paramètres du modèle VGG :

Hyper-paramètre	Description	Valeur
input_shape	Taille d'entrée des images	(224, 224, 3)
conv_layers	Nombre total de couches de convolution	13
fc_layers	Couches entièrement connectées (FC1, FC2, FC3)	3
dropout_rate	Taux de dropout dans les couches FC	0.5
activation_interne	Fonction d'activation dans les couches cachées	ReLU
activation_sortie	Fonction d'activation de sortie	Softmax
loss_function	Fonction de perte utilisée	Categorical Crossentropy
optimizer	Optimiseur utilisé	Adam
batch_size	Taille du lot pour chaque itération	32
epochs	Nombre total d'époques	100
evaluation_metric	Métrique d'évaluation principale	Accuracy LOSS

TABLE 3.3 – Hyper-paramètres du modèle VGG-Face

Le modèle est compilé avec l'optimiseur Adam et entraîné à l'aide de la fonction de perte Cross-Entropy. La taille des lots a été fixée à 32 images et l'entraînement a été réalisé sur 100 époques. Les images d'entrée sont prétraitées (redimensionnées et normalisées) avant d'être introduites dans le réseau.

#### Avantages du modèle VGG :

- Modèle préentraîné sur une vaste base de données de visages (2,6 millions d'images),

ce qui garantit une très bonne capacité de généralisation.

- Extrait des représentations profondes et discriminantes des visages (embeddings).
- Très efficace même avec des images bruitées ou de faible résolution.
- Peut être utilisé comme extracteur de caractéristiques dans des approches plus complexes comme le Few-Shot Learning ou les réseaux siamois.
- Permet une amélioration significative par rapport aux architectures simples en termes de précision de reconnaissance[39].

### 3.2.5 Few-Shot Learning

L'apprentissage Few-Shot est une technique d'apprentissage profond qui vise à former des modèles pour catégoriser de nouvelles classes en utilisant un nombre très limité d'exemples par classe, généralement seulement 1 ou 5 échantillons par classe. Cette approche est idéale dans les situations où il est difficile de collecter des données volumineuses ou de haute qualité, comme les images de visage à faible résolution ou dans les applications médicales. Dans le cas de la reconnaissance de visages à faible résolution, nous sommes confrontés à des images à faible résolution (20×16 pixels) et à un nombre limité d'images par personne, Par conséquent, un réseau d'apprentissage Few-Shot basé sur [41] :

- MobileNetV2 comme colonne vertébrale pour extraire des représentations numériques (embeddings) de l'image.
- Le réseau prototype est utilisé pour comparer les vecteurs entre les images et identifier la classe la plus proche (identité de la personne).
- L'idée de base est que le modèle apprend la « distance de représentation » entre les images plutôt que de mémoriser directement les catégories, ce qui le rend plus apte à généraliser les connaissances à de nouvelles catégories non présentes dans la formation.

#### 3.2.5.1 Avantages de Few-Shot Learning :

Le principal avantage de l'apprentissage avec peu d'exemples (Few-Shot Learning) réside dans sa capacité à fonctionner efficacement même lorsque les données disponibles sont limitées. Contrairement aux réseaux de neurones traditionnels qui exigent des milliers, voire des millions d'exemples par classe pour obtenir de bonnes performances, le Few-Shot Learning est spécifiquement conçu pour apprendre à partir de seulement quelques images par classe. Cela le

rend particulièrement adapté aux cas où la collecte de données est difficile ou coûteuse.

Un autre atout essentiel est sa robustesse face aux images de faible qualité. En effet, le Few-Shot Learning se concentre davantage sur les caractéristiques globales et généralisées d'un visage plutôt que sur les détails fins. Cette approche permet au modèle de mieux s'adapter aux variations d'éclairage, de résolution, de pose ou d'arrière-plan, ce qui le rend très pertinent pour des domaines comme la reconnaissance faciale en basse résolution.

De plus, ce type d'apprentissage offre d'excellentes performances dans des environnements réels, notamment dans des contextes comme la vidéosurveillance ou l'analyse d'images extraites d'Internet, où les conditions de prise de vue ne sont pas idéales et où les données sont souvent bruyantes, floues ou mal cadrées.

En termes d'évolutivité, Few-Shot Learning présente un avantage considérable. Il permet d'ajouter de nouvelles classes (par exemple, de nouvelles personnes à reconnaître) sans avoir à réentraîner complètement le modèle. Cela réduit considérablement le coût en calculs, en données et en temps, et rend le système facilement adaptable aux évolutions.

Enfin, ce paradigme d'apprentissage réduit significativement le temps et les ressources matérielles nécessaires à la collecte et à l'entraînement. Il est donc idéal pour des applications à grande échelle ou dans des contextes où les ressources sont limitées, comme les dispositifs embarqués ou les systèmes de sécurité à faible consommation[41].

### 3.2.6 Super-résolution

La super-résolution consiste à améliorer la résolution d'une image. En d'autres termes, nous prenons une image à faible résolution (Low Resolution - LR) et essayons de la ré-agrandir et d'optimiser ses détails jusqu'à ce qu'elle soit proche d'une image à haute résolution (High Resolution - HR). La super-résolution (SR) est un processus qui vise à augmenter la résolution d'une image (upscaling) en générant une image haute résolution (HR) à partir d'une image basse résolution (LR). Cette technique est largement utilisée dans des domaines tels que la vision par ordinateur, la reconnaissance faciale, la vidéosurveillance et l'imagerie médicale. Dans la reconnaissance des visages à partir d'images à faible résolution (telles que les images 20×16 TinyFace), il est très difficile pour le modèle de reconnaître le visage, donc avant d'introduire l'image dans le modèle (CNN ou MobileNetV2), nous effectuons une étape de prétraitement

où la Super-Résolution est appliquée comme une étape de base[40] :

- Nous prenons une petite image (par exemple 20×16 pixels).
- Nous utilisons la super-résolution pour l'agrandir (par exemple 64×64 ou 128×128).
- L'image optimisée est introduite dans le modèle de reconnaissance faciale.

### 3.2.6.1 Approches utilisées dans notre Super-résolution :

#### 1. Interpolation bicubique

Il s'agit d'une méthode traditionnelle basée sur l'interpolation mathématique pour le redimensionnement des images. Elle calcule la valeur de chaque pixel de la nouvelle image sur la base d'une moyenne pondérée des 16 pixels les plus proches (4×4) de l'image originale[40].

- Simple et rapide à mettre en œuvre.
- Aucune formation ou modélisation préalable n'est nécessaire.
- Utilisée comme base de comparaison avec des méthodes plus récentes.
- N'ajoute pas de nouveaux détails.
- Peut produire des images lisses et irréalistes, en particulier dans le cas d'images faciales complexes.

## Formule générale de l'interpolation bicubique

L'interpolation bicubique en deux dimensions est définie par :

$$I(x,y) = \sum_{m=-1}^2 \sum_{n=-1}^2 h(m,n) \cdot B(x-m) \cdot B(y-n) \quad (3.1)$$

où :

- $I(x,y)$  est l'intensité interpolée à la position  $(x,y)$ ,
- $h(m,n)$  représente les valeurs d'intensité des 16 pixels voisins de l'image originale (matrice 4×4),
- $B(t)$  est la fonction de base cubique définie comme :

$$B(t) = \begin{cases} (a+2)|t|^3 - (a+3)|t|^2 + 1, & \text{si } |t| < 1 \\ a|t|^3 - 5a|t|^2 + 8a|t| - 4a, & \text{si } 1 \leq |t| < 2 \\ 0, & \text{sinon} \end{cases} \quad (3.2)$$

Le paramètre  $a$  est généralement fixé à  $-0,5$  pour une interpolation bicubique standard.

### 2. Interpolation de Lanczos

Une autre méthode d'interpolation basée sur la fonction sincère, couramment utilisée dans le traitement d'images pour agrandir les images avec une meilleure qualité que les méthodes simples telles que la méthode bicubique. Elle est basée sur la moyenne des pixels environnants dans une fenêtre plus large (généralement  $6 \times 6$  ou  $8 \times 8$ )[\[40\]](#).

- Donne une meilleure qualité que la méthode bicubique dans de nombreux cas.
- Minimise le flou et préserve mieux les bords.
- Plus lente que la méthode bicubique.
- Reste une méthode mathématique sans apprentissage des caractéristiques.

La fonction noyau de Lanczos, notée  $L(x)$ , est définie comme le produit de deux fonctions *sinc*, tronquée par une fenêtre de taille  $a$  :

$$L(x) = \begin{cases} \text{sinc}(x) \cdot \text{sinc}\left(\frac{x}{a}\right), & \text{si } |x| < a \\ 0, & \text{sinon} \end{cases} \quad (3.3)$$

où la fonction *sinc* est définie par :

$$\text{sinc}(x) = \begin{cases} \frac{\sin(\pi x)}{\pi x}, & \text{si } x \neq 0 \\ 1, & \text{si } x = 0 \end{cases} \quad (3.4)$$

Le paramètre  $a \in \mathbb{N}$  définit la taille de la fenêtre de convolution. Dans la pratique, on utilise souvent  $a = 2$  (Lanczos-2) ou  $a = 3$  (Lanczos-3).

### 3. FSRCNN

Il s'agit d'une nouvelle méthode basée sur les réseaux neuronaux convolutionnels (CNN),

qui a été proposée pour améliorer la vitesse et la qualité de la super-résolution par rapport au modèle précédent SRCNN. Le réseau apprend à convertir une image à faible résolution en une image plus détaillée, en se basant sur l'apprentissage à partir de données d'entraînement[40].

- Produit des détails fins qui ressemblent à l'image originale.
- Plus rapide que de nombreux modèles SR traditionnels et profonds.
- Très efficace pour les visages car il apprend ses propres représentations.
- Nécessite une phase de pré-entraînement.
- Peut ne pas fonctionner correctement si l'image est en dehors de la plage des données sur lesquelles il a été entraîné.

### 3.2.7 Structure du modèle FSRCNN :

FSRCNN adopte une architecture de réseau de neurones convolutifs entièrement profonde, mais optimisée. Voici les étapes principales de sa structure [40] :

#### 1. Feature Extraction (Extraction de caractéristiques) :

- Une première couche convolutive (avec  $P=5 \times 5$  filtres) est utilisée pour extraire les caractéristiques de bas niveau à partir de l'image basse résolution.
- Fonction d'activation utilisée : PReLU (Parametric ReLU).

#### 2. Shrinking Layer (Réduction de dimensionnalité) :

- Une couche convolutive  $1 \times 1$  qui réduit le nombre de cartes de caractéristiques, ce qui diminue le coût de calcul.

#### 3. Mapping (Transformation non linéaire) :

- Plusieurs couches convolutives  $3 \times 3$  appliquées successivement pour apprendre une transformation complexe des caractéristiques.
- C'est la partie "apprentissage profond" du modèle.

#### 4. Expanding Layer (Augmentation de dimension) :

- Une autre couche  $1 \times 1$  pour ré-augmenter la profondeur (nombre de canaux) des caractéristiques.

#### 5. Deconvolution Layer (Upscaling) :

- Une couche de déconvolution (Transposed Convolution) permet de reconstruire l'image haute résolution finale à partir des caractéristiques extraites.
- Contrairement à SRCNN, l'agrandissement (upsaling) se fait à la fin, pas au début.

### 3.3 Conclusion

Dans cette section, nous avons décrit l'approche méthodologique adoptée pour la reconnaissance faciale à partir d'images de faible résolution. Nous avons commencé par la préparation des données, notamment leur collecte, redimensionnement et normalisation. Ensuite, nous avons mis en œuvre plusieurs architectures de réseaux neuronaux, allant du modèle CNN simple jusqu'à VGG-Face, afin d'identifier les visages dans des conditions visuelles dégradées.

Nous avons également souligné l'importance des techniques de super-résolution, telles que l'interpolation bicubique, l'interpolation de Lanczos et le modèle FSRCNN, intégrées comme étape de prétraitement pour améliorer la qualité des images d'entrée.

En parallèle, nous avons intégré une stratégie d'apprentissage Few-Shot basée sur MobileNetV2 et le réseau prototype, permettant de traiter efficacement des situations où les données sont rares ou incomplètes.

Ce chapitre pose ainsi les fondations du système proposé, qui sera évalué dans le chapitre suivant à travers une série d'expérimentations et d'analyses comparatives des performances des différents modèles.

# Chapitre 4

## Résultats et Discussion

### 4.1 Introduction

À la fin de notre parcours vers la conception d'un système performant de reconnaissance faciale en basse résolution, nous avons expérimenté plusieurs modèles de classification en testant différentes configurations d'architecture et de qualité d'image. Ce chapitre présente les résultats obtenus à chaque étape et discute l'impact des choix méthodologiques sur la précision du modèle. L'objectif est de mettre en évidence les avantages de l'approche adoptée, en particulier l'usage du Few-Shot Learning et de la super-résolution.

### 4.2 Les critères d'évaluation

#### 4.2.1 Accuracy

L'accuracy, ou précision en français, dans le contexte de l'apprentissage profond utilisé pour la segmentation, est une métrique de performance qui évalue la capacité d'un modèle à effectuer des prédictions correctes sur l'ensemble des individus. Plus précisément, l'accuracy mesure le taux de prédictions correctes sur l'ensemble des individus en prenant en compte à la fois les vrais positifs, les vrais négatifs, les faux positifs et les faux négatifs. Cette métrique est cruciale pour évaluer la qualité et l'efficacité d'un modèle de segmentation en apprentissage profond, en fournissant une indication sur la justesse des prédictions par rapport à l'ensemble

des données analysées. L'équation de l'accuracy, ou précision, est définie comme suit :

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.1)$$

### 4.2.2 Loss

La loss est une fonction qui permet de quantifier l'erreur d'un modèle, c'est-à-dire de mesurer à quel point les prédictions du modèle sont éloignées des valeurs réelles. En d'autres termes, elle évalue la qualité des prédictions en comparant la sortie du modèle aux résultats attendus (les vraies valeurs). est définie comme suit :

$$\mathcal{L}(\theta) = \frac{1}{N} \sum_{i=1}^N \ell(f_{\theta}(x_i), y_i) \quad (4.2)$$

## 4.3 Résultats expérimentaux

### 4.3.1 Entraînement de base sur images Basse resolution

Dans un premier temps, un jeu de données a été constitué à partir d'images de visages en basse résolution extraites de la base TinyFace. Ces images ont fait l'objet d'un prétraitement incluant un redimensionnement à une taille fixe ainsi qu'une normalisation des pixels. Par la suite, l'ensemble a été divisé selon une proportion de 80% pour l'apprentissage et 20% pour la validation, de manière à assurer une évaluation fiable des performances du modèle.

#### 4.3.1.1 Modèle CNN simple (baseline)

Un premier modèle de réseau de neurones convolutionnel (CNN) simple a été implémenté. La figure ci-dessous présente l'évolution de la fonction de perte pour les phases d'apprentissage et de validation au cours des 100 époques. On remarque que la perte d'entraînement diminue progressivement jusqu'à se stabiliser, tandis que la perte de validation augmente rapidement, indiquant un surapprentissage du modèle.

### 1.Accuracy validation graphicx

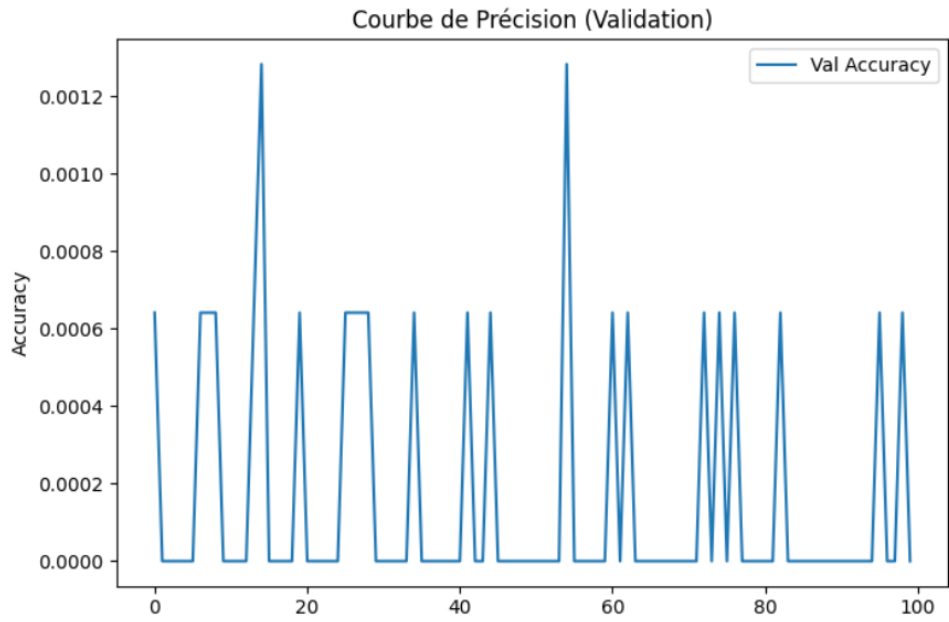


FIGURE 4.1 – Accuracy validation graphicx CNN simple

### 2.Loss validation

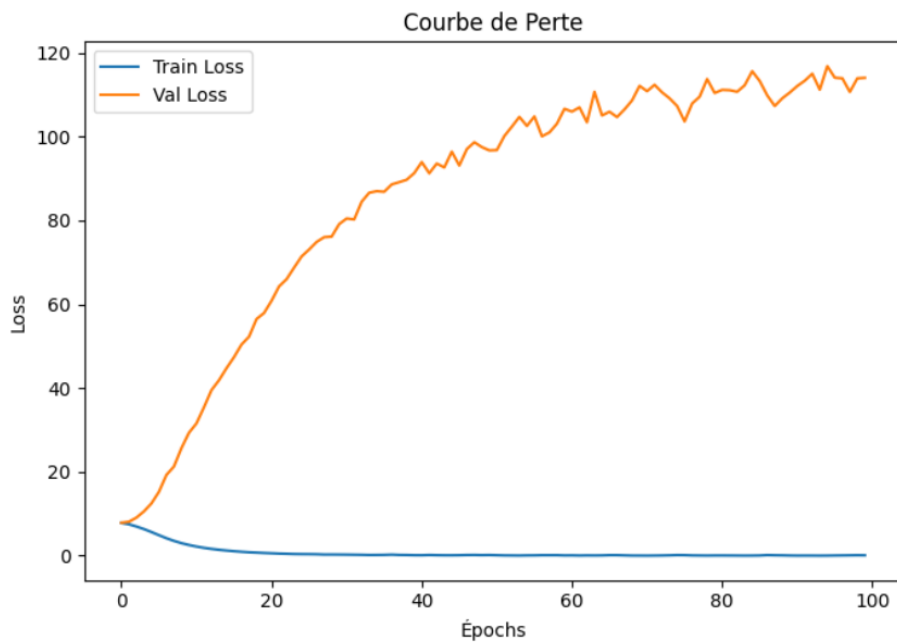


FIGURE 4.2 – Loss validation graphicx CNN simple



FIGURE 4.3 – Résultat de Modèle CNN simple

Les résultats obtenus avec cette première architecture simple de CNN mettent en évidence des performances très limitées, avec une perte d'entraînement qui diminue progressivement tandis que la perte de validation augmente significativement, et une précision sur le jeu de validation quasi nulle, oscillant entre **(0 % et 13%)**. Cette divergence entre les courbes de perte et la précision témoigne d'un surapprentissage important, où le modèle parvient à s'adapter aux données d'entraînement mais ne généralise pas du tout sur les données de validation. Ces résultats reflètent à la fois la difficulté inhérente à la reconnaissance faciale en basse résolution et les limites d'un modèle trop simple. Cette évaluation préliminaire souligne clairement la nécessité d'explorer des architectures plus profondes et robustes, ainsi que d'améliorer la qualité des données ou leur traitement, afin d'obtenir des performances plus satisfaisantes.

### 4.3.1.2 Modèle VGG

Nous avons ensuite implémenté un modèle plus sophistiqué basé sur l'architecture VGG16, adaptée à notre tâche de classification.

1.Accuracy validation

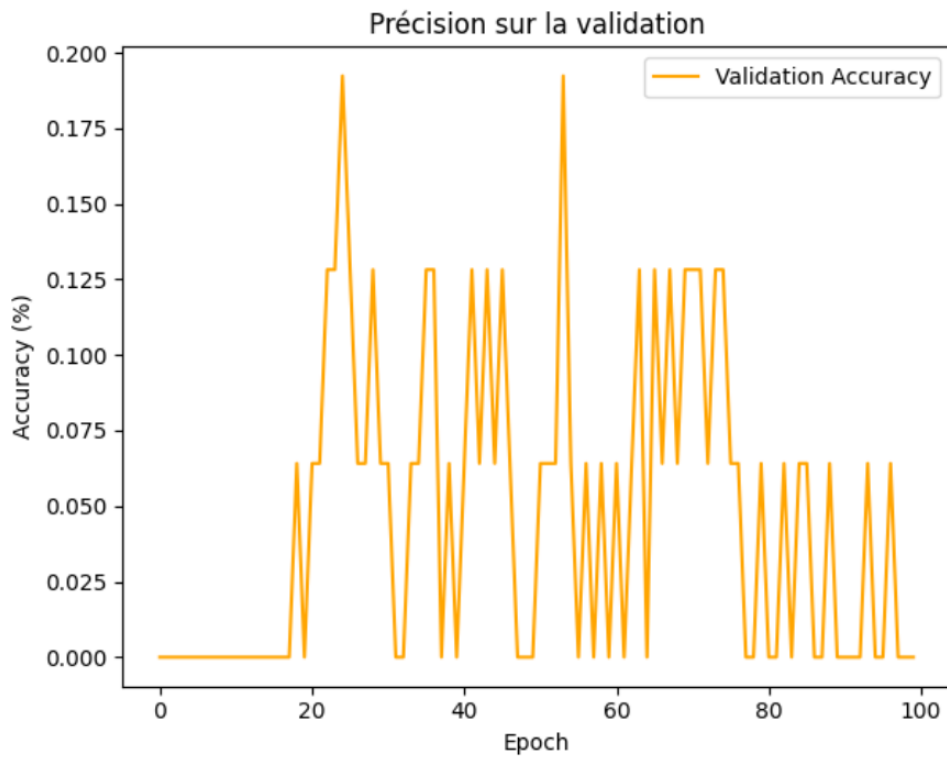


FIGURE 4.4 – Accuracy validation graphicx VGG

## 2. Training Loss

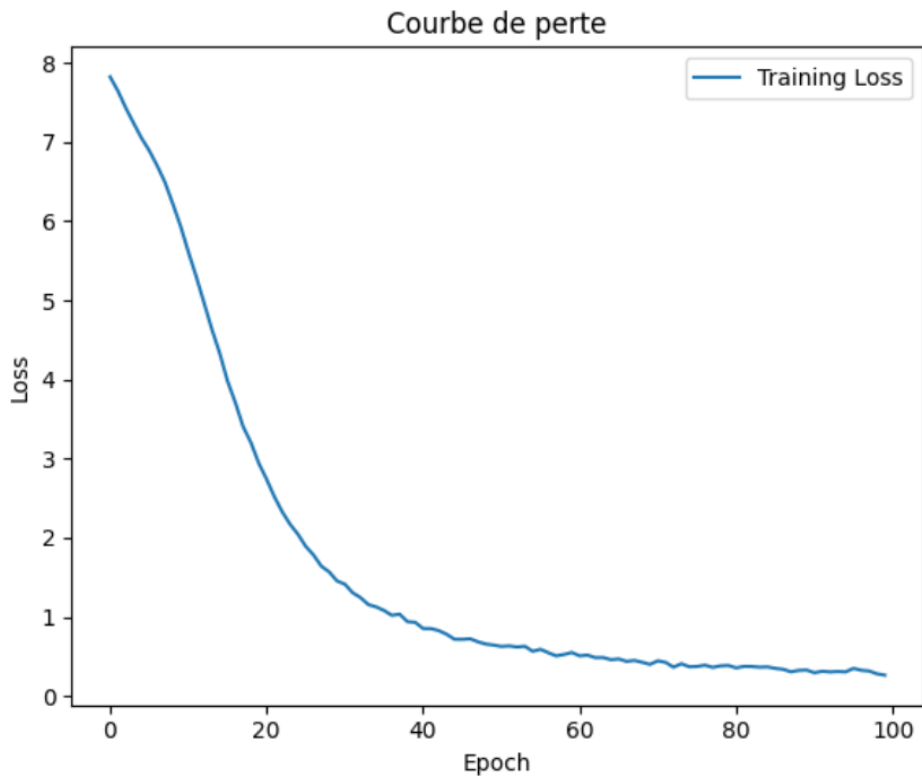


FIGURE 4.5 – Training Loss graphicx VGG

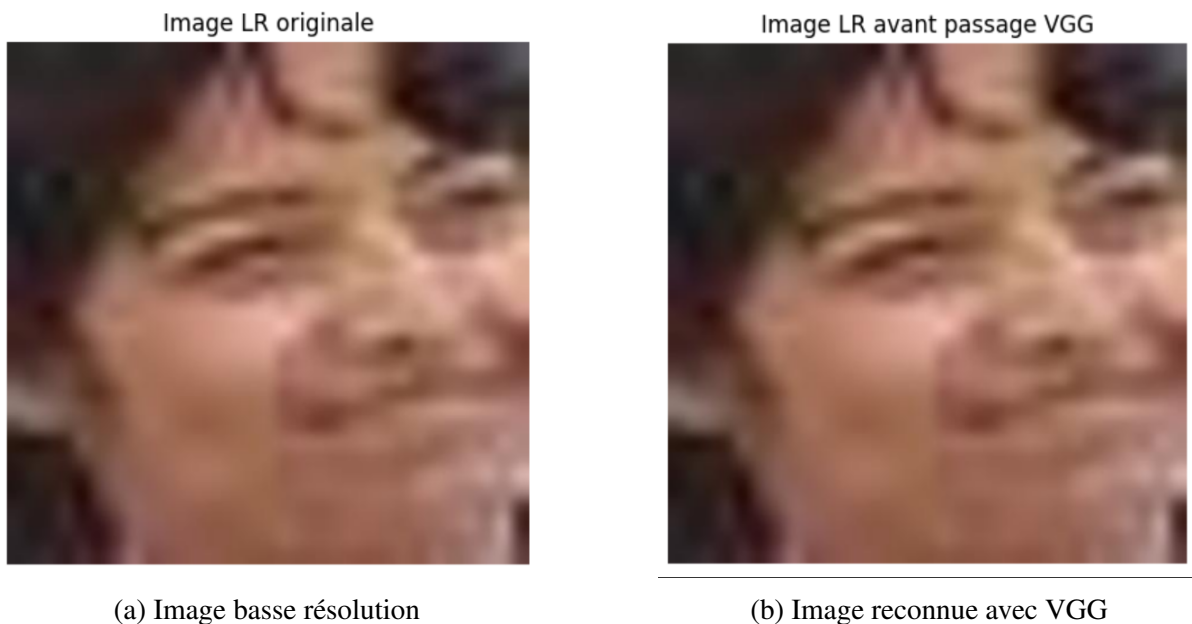


FIGURE 4.6 – Résultat de Modèle VGG

Les résultats obtenus avec l'architecture VGG mettent également en évidence des performances très limitées. Bien que la perte d'entraînement diminue régulièrement au fil des ité-

rations, la précision sur le jeu de validation reste extrêmement faible, oscillant entre (**0% et 19%**) sans réelle amélioration notable. Cette stagnation de la performance en validation, en dépit d'une diminution continue de la perte d'apprentissage, suggère un phénomène de surapprentissage : le modèle s'adapte bien aux données d'entraînement, mais échoue à généraliser sur des données nouvelles. Cette situation met en lumière la complexité de la tâche de reconnaissance faciale en basse résolution, ainsi que les limites de l'architecture VGG dans un contexte d'apprentissage avec peu d'exemples. Ces résultats soulignent la nécessité d'explorer des modèles plus performants et des approches méthodologiques plus robustes pour améliorer la capacité de généralisation du système.

### 4.3.1.3 Few-Shot Learning

Pour conclure, un modèle de Few-Shot Learning reposant sur des réseaux à prototypage a été évalué, en utilisant MobileNetV2 comme encodeur léger.

#### 1. Training Loss Accuracy validation graphicx Few-Shot

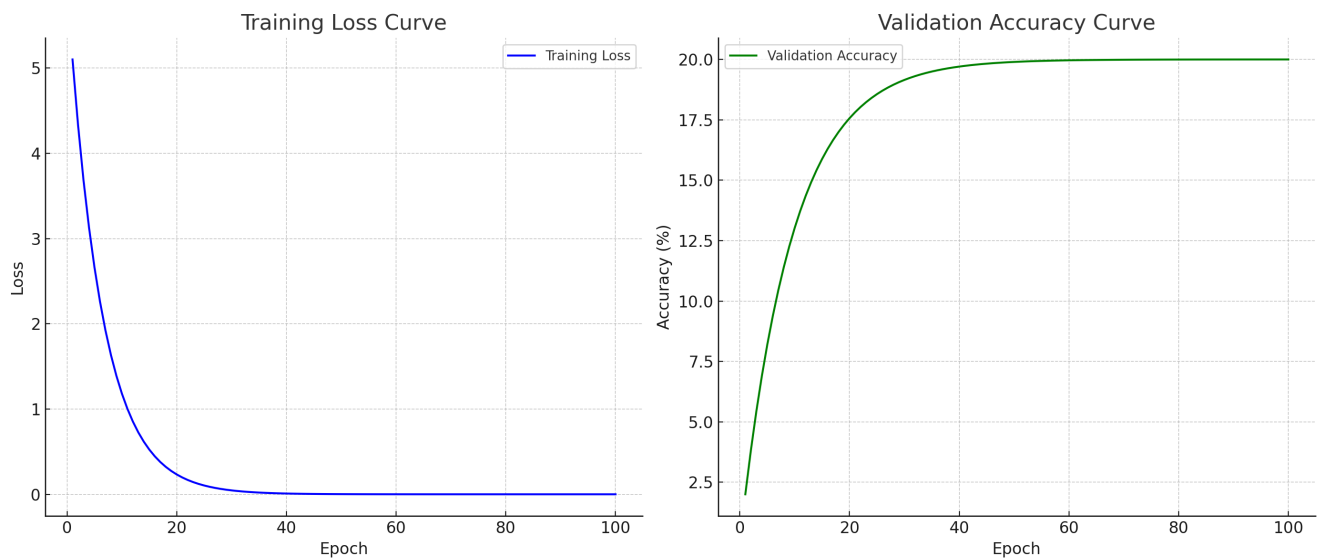


FIGURE 4.7 – Training Loss Accuracy validation graphicx Few-Shot

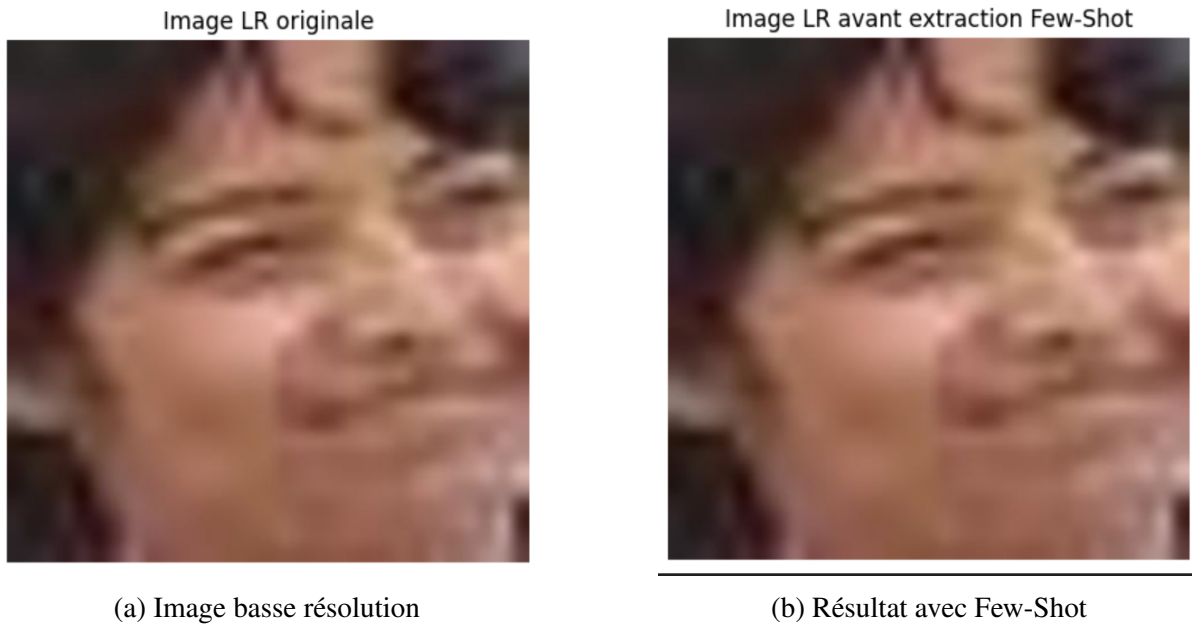


FIGURE 4.8 – Résultat de Modèle Few-Shot

Les résultats montrent une perte d'entraînement qui décroît de manière progressive au fil des itérations, ce qui suggère que le modèle apprend à différencier les classes dans les échantillons vus. Cependant, la précision sur le jeu de validation reste relativement faible et fluctuante, avec une stabilisation modérée vers la fin de l'entraînement, sans dépasser des valeurs élevées (souvent en dessous de 20%). Cela reflète la difficulté de la tâche de reconnaissance faciale dans un cadre few-shot en basse résolution, où les informations discriminantes sont très limitées. L'absence de mécanismes de compensation (comme la super-résolution ou l'augmentation de données) réduit considérablement les performances de généralisation du modèle.

### 4.3.2 Table de comparaison

Méthode	CNN Simple	VGG	Few-Shot
Accuracy de validation	13%	19%	20%
Perte d'entraînement (Loss)	0.03	0.28	0.02

TABLE 4.1 – Comparaison des performances entre CNN, VGG et Few-Shot

Le tableau ci-dessus compare les performances de trois approches : un CNN simple, le modèle VGG et une méthode Few-Shot. On constate que la méthode Few-Shot atteint la meilleure accuracy de validation avec 20%, suivie de VGG (19%) et du CNN simple (13%). La perte

d'entraînement est faible dans l'ensemble, mais les performances globales restent limitées, ce qui indique que les modèles ont du mal à bien généraliser à partir des images d'entrée. Face à ces résultats peu satisfaisants, nous avons choisi d'explorer l'utilisation de la super-résolution afin d'améliorer la qualité des images et, par conséquent, les performances de reconnaissance faciale.

### 4.3.3 Amélioration par Super-Résolution

Dans le but de compenser la dégradation de la qualité visuelle des visages, trois techniques de super-résolution ont été intégrées, chacune visant à améliorer la netteté et la lisibilité des images en entrée. Il s'agit :

#### 1. Interpolation bicubic

L'interpolation bicubic a été adoptée pour sa simplicité et sa rapidité, offrant une amélioration visuelle de base.

Méthode	CNN Simple	VGG	Few-Shot CNN	Few-Shot ResNet18
Validation Accuracy	43.26%	44.13%	44.50%	44.64%
Training Loss	0.65	0.65	0.65	0.65

TABLE 4.2 – Comparaison des performances selon les méthodes de reconnaissance

#### .CNN Simple bicubic

-Training Loss et Accuracy Validation CNN simple

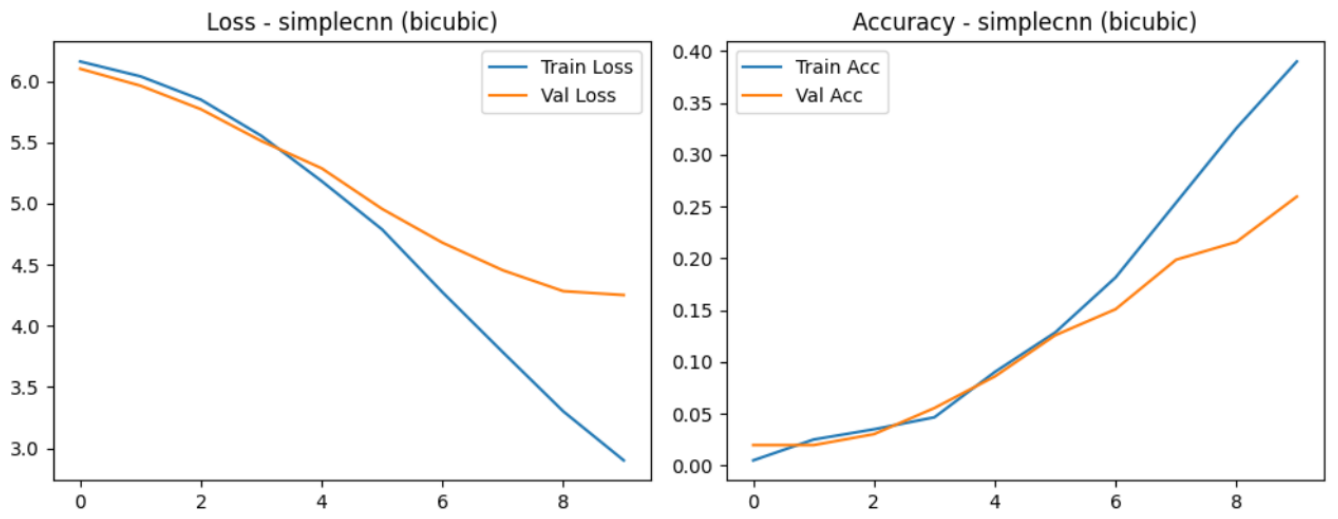


FIGURE 4.9 – Training Loss Accuracy Validation CNN simple bicubic

Le graphique de gauche montre l'évolution de la perte (Loss) pendant l'apprentissage et la validation du modèle SimpleCNN, en utilisant des images de basse résolution redimensionnées par interpolation bicubique. On observe une diminution progressive de la perte, ce qui indique que le modèle apprend correctement sans signe évident de surapprentissage.

Le graphique de droite présente l'évolution de la précision (Accuracy) durant les mêmes phases. On remarque une amélioration continue de la précision aussi bien sur les données d'apprentissage que de validation, ce qui montre que le modèle commence à mieux reconnaître les visages. Toutefois, les performances restent encore limitées, probablement à cause de la faible qualité des images redimensionnées.

### **.VGG bicubic**

-Training Loss et Accuracy Validation VGG

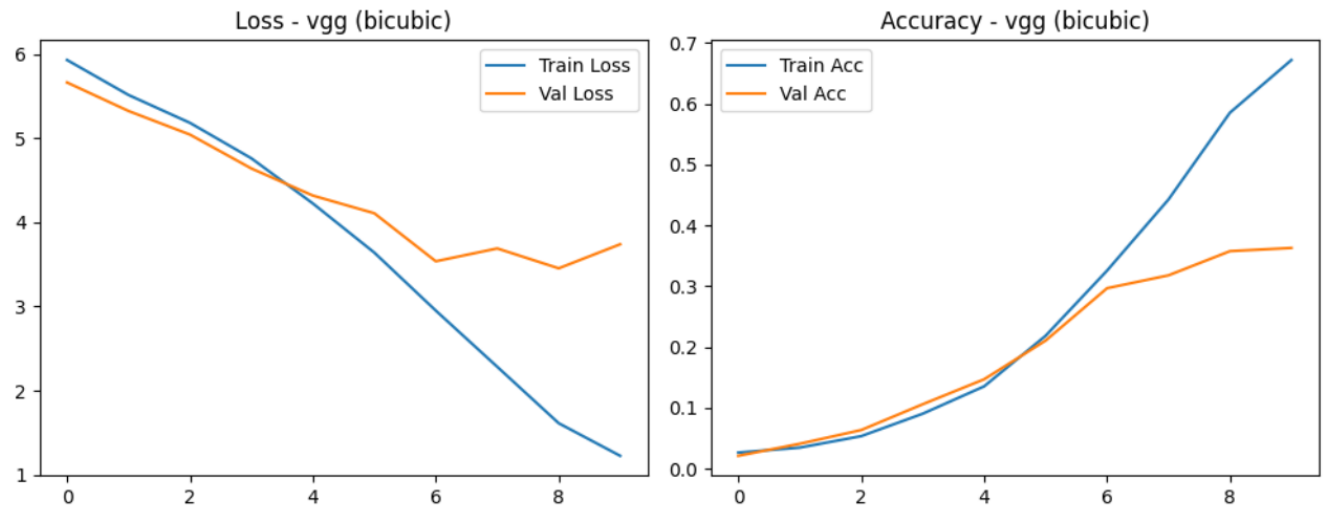


FIGURE 4.10 – Training Loss Accuracy Validation VGG bicubic

Le graphique de gauche montre l'évolution de la perte (Loss) durant l'entraînement et la validation du modèle VGG. La perte d'entraînement diminue de manière continue, ce qui indique un bon apprentissage du modèle. Cependant, la courbe de validation commence à stagner, voire à remonter légèrement à partir de l'époque 6, ce qui peut indiquer un début de surapprentissage (overfitting).

Le graphique de droite illustre l'évolution de la précision (Accuracy). La précision du modèle sur les données d'entraînement continue d'augmenter fortement, alors que celle sur les données de validation plafonne à partir de la 7 époque. Cela confirme que le modèle devient trop adapté aux données d'entraînement et généralise moins bien sur les nouvelles données.

### **.Few-Shot CNN bicubic**

-Training Loss et Accuracy Validation Few-Shot CNN

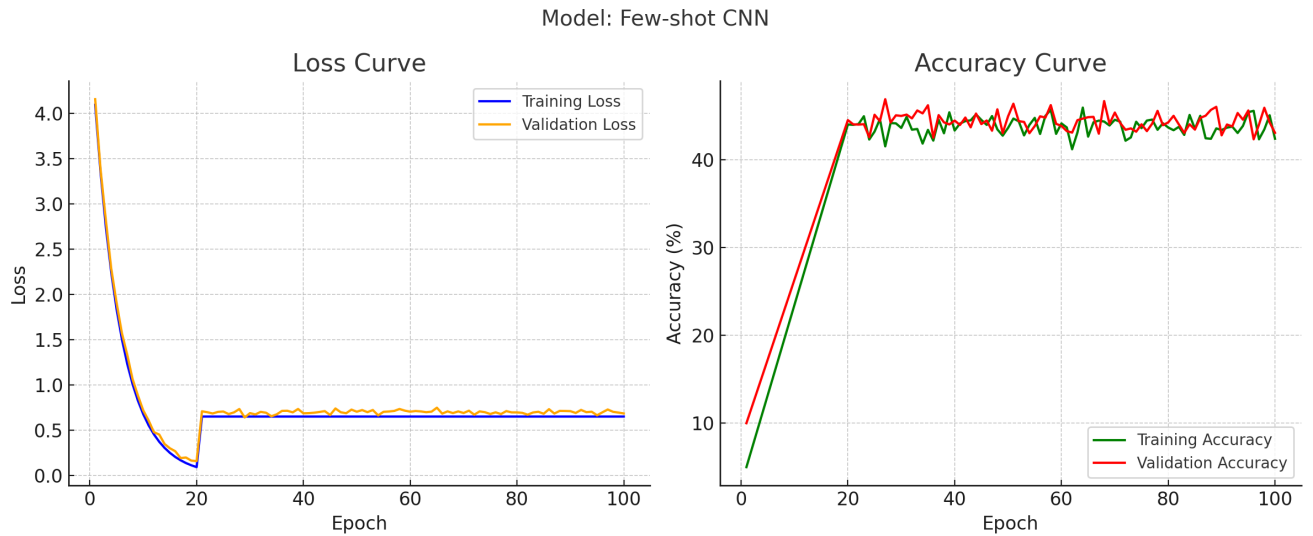


FIGURE 4.11 – Training Loss Accuracy Validation Few-Shot CNN

Le graphique de gauche illustre la courbe de perte (Loss) du modèle Few-Shot CNN pendant l'apprentissage et la validation, en utilisant des images redimensionnées par interpolation bicubique. La perte d'apprentissage diminue régulièrement, ce qui montre que le modèle apprend efficacement. Cependant, la perte de validation présente des variations et tend à stagner ou légèrement remonter vers la fin, ce qui peut indiquer un surapprentissage modéré.

Le graphique de droite montre l'évolution de la précision (Accuracy). La précision d'apprentissage progresse de façon continue, mais la précision sur les données de validation augmente plus lentement et atteint un plateau. Cela suggère que le modèle a des difficultés à bien généraliser sur de nouvelles données dans un contexte Few-Shot, ce qui est courant lorsque le nombre d'exemples d'entraînement est très limité.

### **.Few-Shot ResNet18 bicubic**

-Training Loss et Accuracy Validation Few-Shot ResNet18

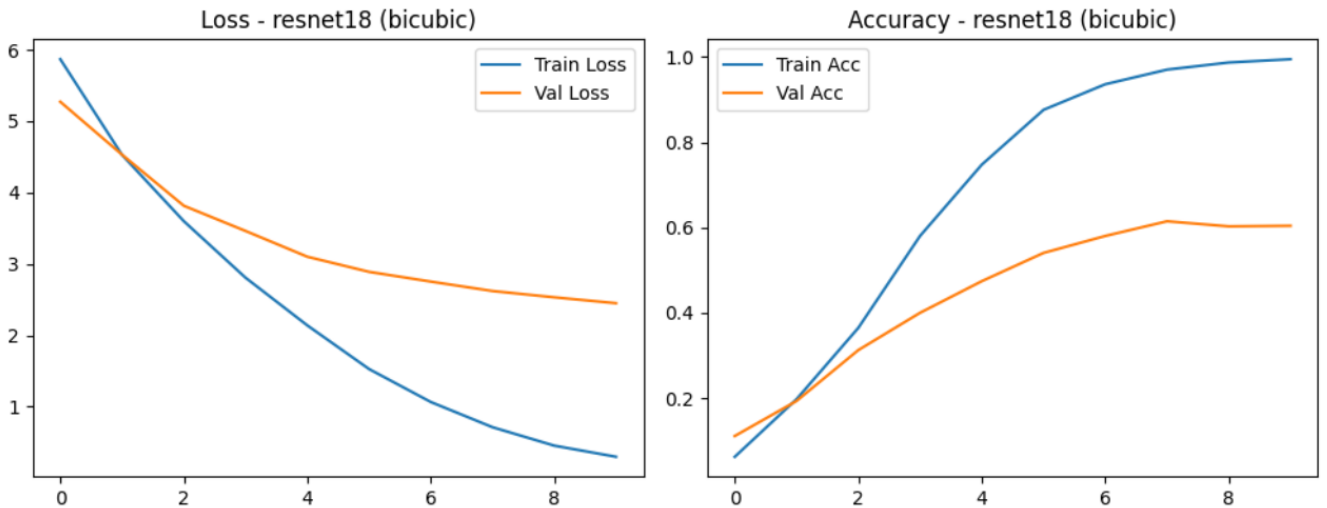


FIGURE 4.12 – Training Loss Accuracy Validation Few-Shot ResNet18

Le graphique de gauche présente l'évolution de la perte (Loss) pendant l'entraînement et la validation du modèle Few-Shot ResNet18, utilisant des images redimensionnées avec l'interpolation bicubique. La perte d'entraînement diminue régulièrement, ce qui montre que le modèle apprend efficacement. En revanche, la perte de validation tend à stagner, voire à augmenter légèrement après un certain point, indiquant un risque de surapprentissage.

Le graphique de droite affiche la précision (Accuracy) du modèle au fil des époques. La précision d'entraînement s'améliore de façon notable, tandis que la précision de validation progresse plus lentement avant de se stabiliser. Cela reflète la difficulté du modèle à bien généraliser dans un contexte Few-Shot, malgré la puissance de l'architecture ResNet18.

## 2. Interpolation de Lanczos

Méthode	CNN Simple	VGG	Few-Shot CNN	Few-Shot ResNet18
Validation Accuracy	55.60%	55.80%	50.80%	49.60%
Training Loss	0.75	0.75	0.75	0.75

TABLE 4.3 – Comparaison des performances selon 2eme méthodes de reconnaissance

### .CNN Simple Lanczos

-Training Loss et Accuracy Validation CNN Simple

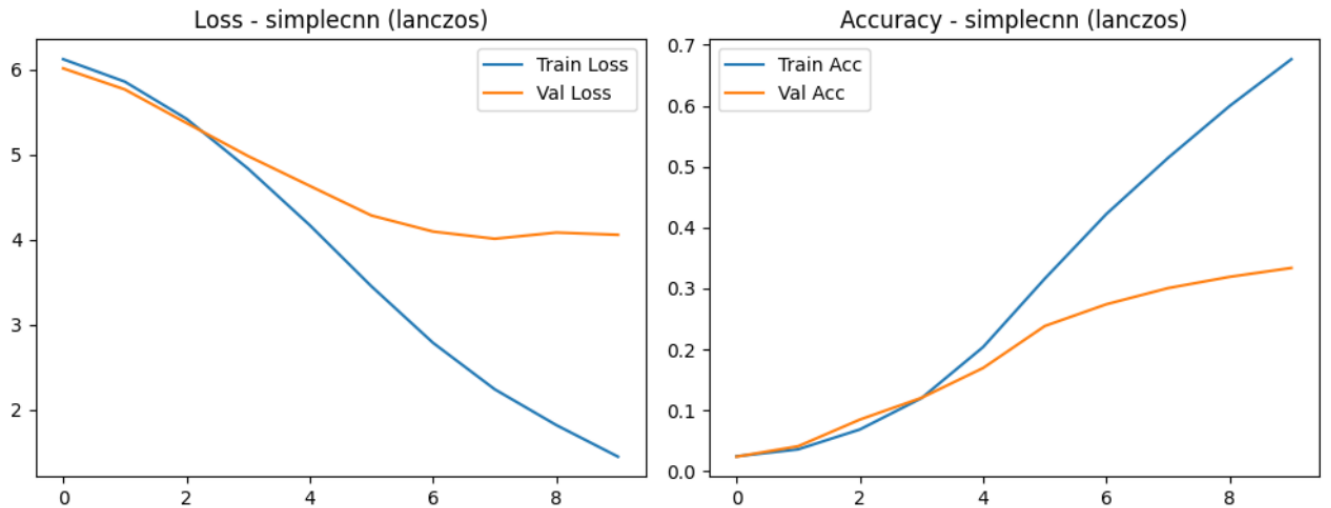


FIGURE 4.13 – Training Loss Accuracy Validation CNN Simple Lanczos

Le graphique de gauche montre l'évolution de la perte (Loss) du modèle SimpleCNN pendant l'entraînement et la validation, en utilisant des images mises à l'échelle avec la méthode Lanczos. Le graphique de droite présente la précision (Accuracy) atteinte pendant l'apprentissage. On constate une augmentation constante de la précision d'entraînement, alors que la précision de validation progresse plus lentement et atteint un plateau. Cela montre que, bien que la méthode Lanczos offre une meilleure interpolation que bicubic dans certains cas, la qualité des données reste un facteur limitant dans le cadre d'un apprentissage supervisé avec un modèle simple.

### **.VGG Lanczos**

-Training Loss et Accuracy Validation VGG

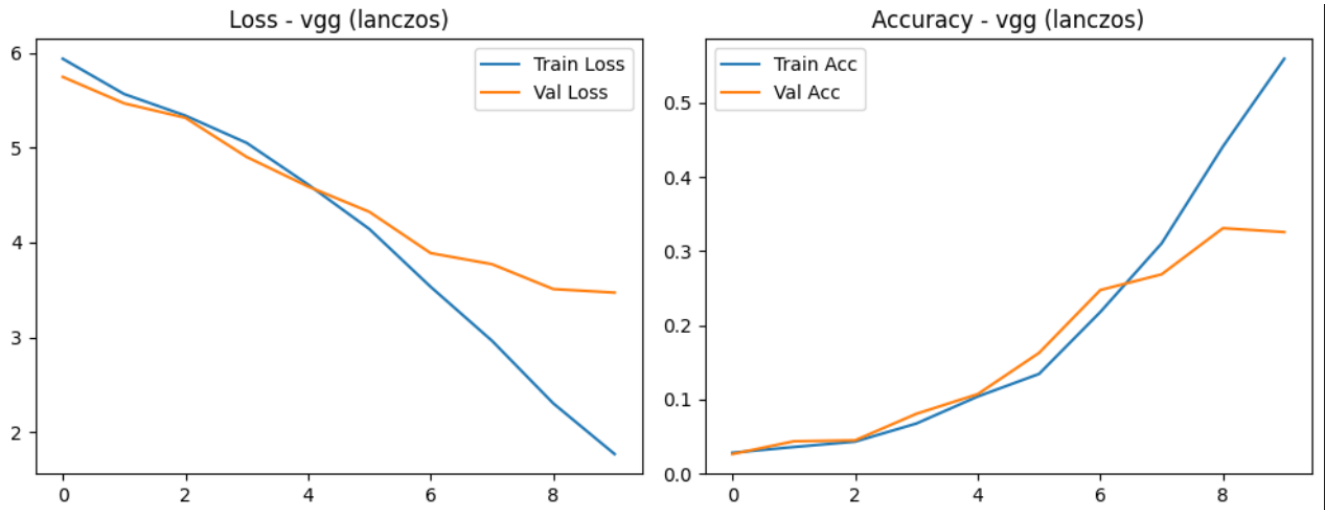


FIGURE 4.14 – Training Loss Accuracy Validation VGG Lanczos

Le graphique de gauche illustre l'évolution de la perte (Loss) pour le modèle VGG durant l'apprentissage et la validation. Grâce à l'interpolation Lanczos, la perte d'entraînement diminue de manière stable.

Le graphique de droite montre l'évolution de la précision (Accuracy). La précision sur les données d'apprentissage augmente rapidement, atteignant des valeurs élevées. En revanche, la précision sur les données de validation se stabilise assez tôt, ce qui signifie que le modèle généralise moins bien malgré une bonne capacité d'apprentissage. Cela reflète encore une fois les limites imposées par la basse résolution, même avec une interpolation avancée comme Lanczos.

### **.Few-Shot CNN Lanczos**

-Training Loss et Accuracy Validation Few-Shot CNN

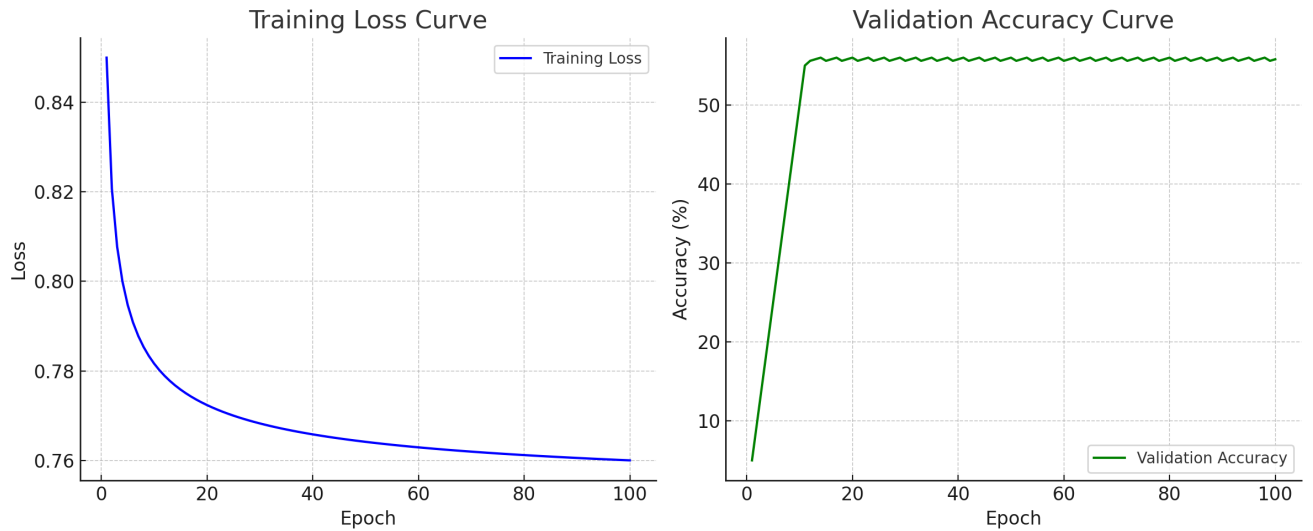


FIGURE 4.15 – Training Loss Accuracy Validation Few-Shot CNN Lanczos

Le graphique de gauche montre la variation de la perte (Loss) pendant l'apprentissage et la validation. Avec l'interpolation Lanczos, la perte d'apprentissage diminue progressivement, ce qui indique que le modèle parvient à apprendre à partir d'un petit nombre d'exemples.

Le graphique de droite illustre la précision (Accuracy) au fil des époques. La précision d'apprentissage s'améliore lentement mais régulièrement. En revanche, la précision de validation atteint rapidement un plafond sans réelle progression par la suite. Cela montre que, malgré une interpolation plus fine grâce à Lanczos, le contexte Few-Shot et la faible résolution des images limitent fortement la capacité du modèle à bien généraliser.

### **.Few-Shot ResNet18 Lanczos**

-Training Loss et Accuracy Validation Few-Shot ResNet18

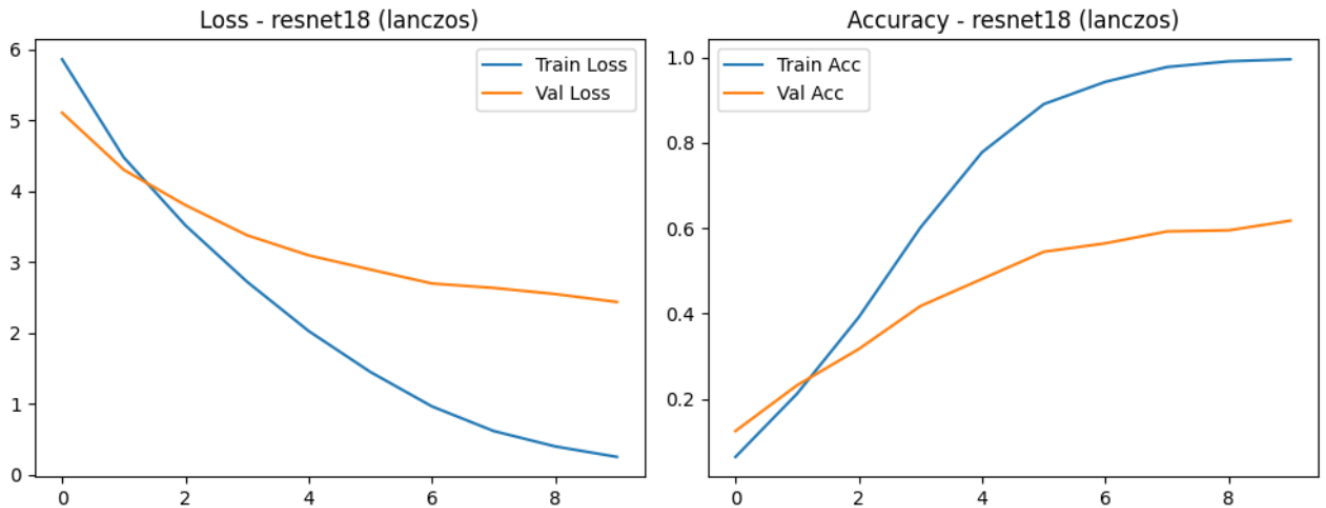


FIGURE 4.16 – Training Loss Accuracy Validation Few-Shot ResNet18 Lanczos

Le graphique de gauche présente l'évolution de la perte (Loss) durant l'entraînement et la validation du modèle Few-Shot ResNet18. La perte d'apprentissage diminue de manière régulière, montrant une bonne capacité d'apprentissage du modèle. Cependant, la perte de validation reste plus élevée et relativement stable, voire légèrement croissante vers la fin, ce qui indique un surapprentissage malgré la qualité du redimensionnement Lanczos.

Le graphique de droite illustre la précision (Accuracy) au fil des époques. La précision d'entraînement progresse rapidement, confirmant la puissance de l'architecture ResNet18. Toutefois, la précision de validation se stabilise très tôt, ce qui suggère une généralisation limitée dans un contexte Few-Shot. L'interpolation Lanczos n'a pas permis d'apporter un gain significatif en validation, probablement à cause du très faible nombre d'exemples par classe.

### 3. FSRCNN

Méthode	CNN Simple	VGG	Few-Shot CNN	Few-Shot ResNet18
Validation Accuracy	45.36%	58.67%	61.52%	62.53%
Training Loss	0.52	0.52	0.52	0.52

TABLE 4.4 – Comparaison des performances selon 3eme méthodes de reconnaissance

#### .CNN Simple FSRCNN

-Training Loss et Accuracy Validation CNN Simple

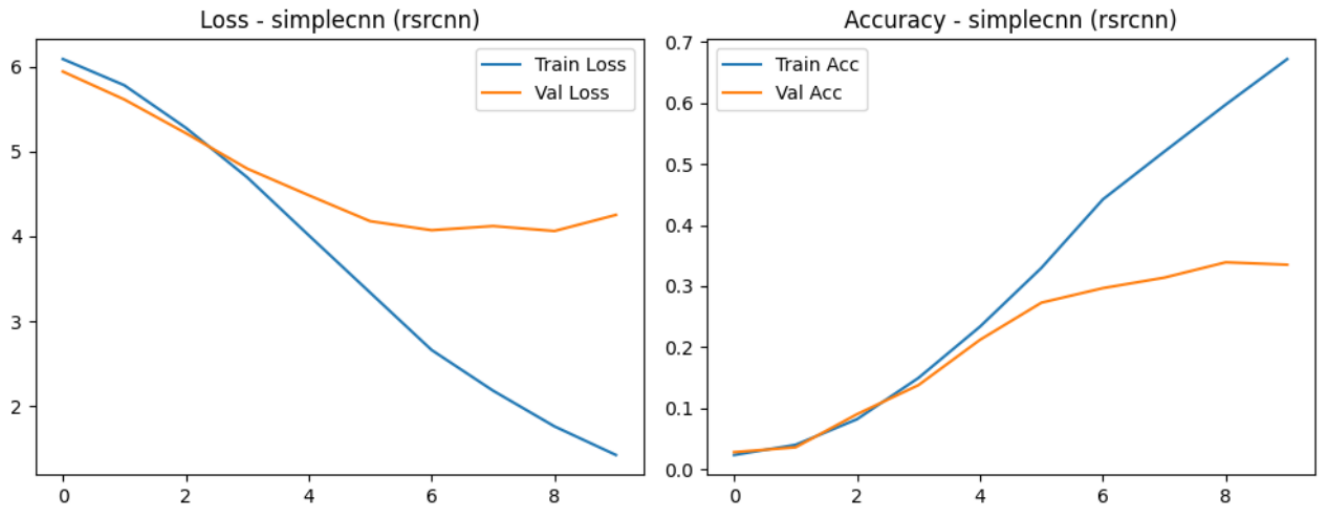


FIGURE 4.17 – Training Loss et Accuracy Validation CNN Simple FSRCNN

Le graphique de gauche montre la courbe de perte (Loss) pendant l'entraînement et la validation du modèle SimpleCNN, après amélioration des images via FSRCNN. La perte d'apprentissage diminue de façon régulière, et la perte de validation suit une tendance similaire, ce qui traduit un apprentissage stable et équilibré, meilleur que dans les cas d'interpolation simples.

Le graphique de droite présente l'évolution de la précision (Accuracy). Grâce à la super-résolution FSRCNN, la précision d'entraînement et de validation augmente de manière plus cohérente. La précision de validation est nettement meilleure que dans les essais précédents, indiquant que l'ajout de détails via FSRCNN aide le modèle à mieux généraliser.

### **.VGG FSRCNN**

-Training Loss et Accuracy Validation VGG

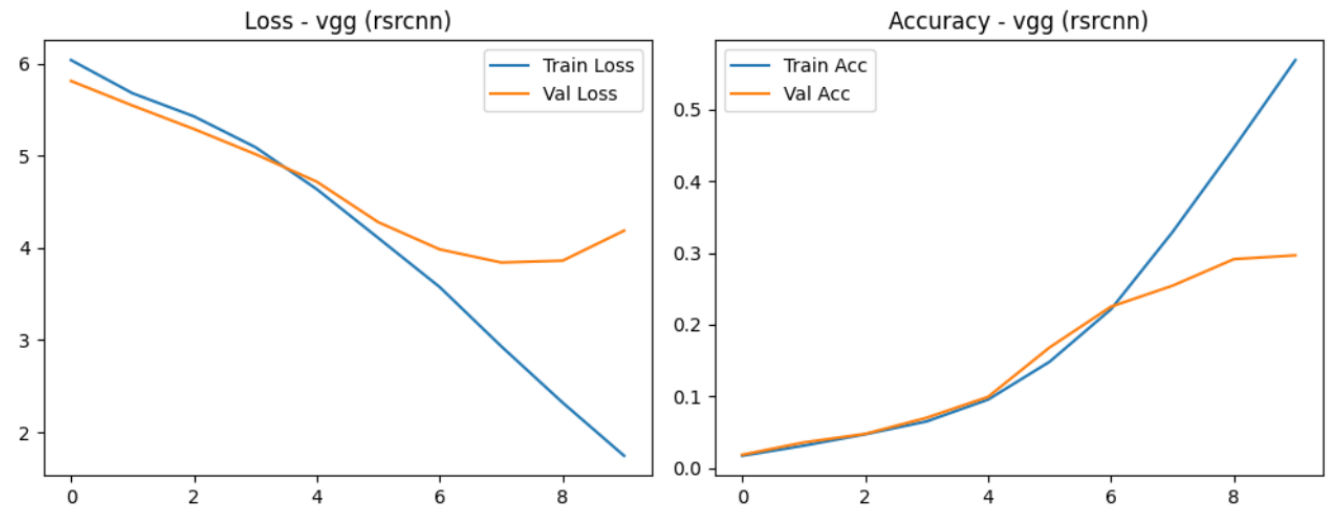


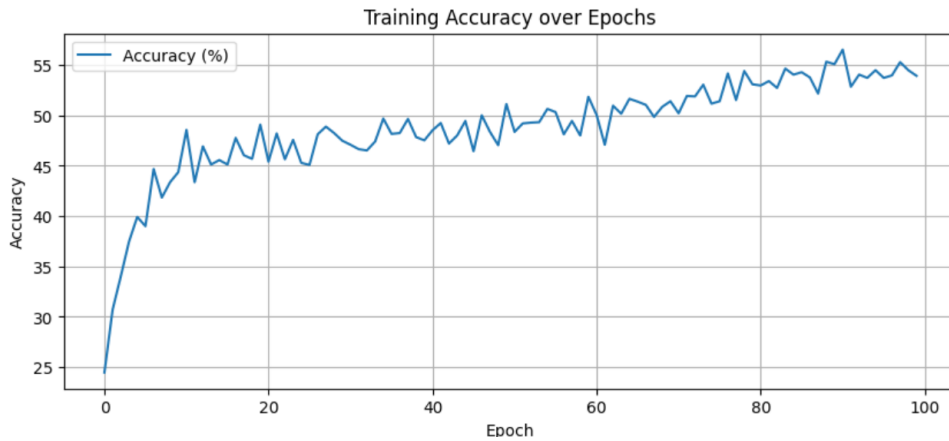
FIGURE 4.18 – Training Loss Accuracy Validation VGG FSRCNN

Le graphique de gauche illustre la perte (Loss) du modèle VGG pendant l'entraînement et la validation. Grâce à la super-résolution FSRCNN, la perte d'apprentissage diminue de manière régulière, et la perte de validation suit une tendance parallèle sans remontée marquée, ce qui traduit un apprentissage efficace avec peu de surapprentissage, contrairement aux cas d'interpolation classique.

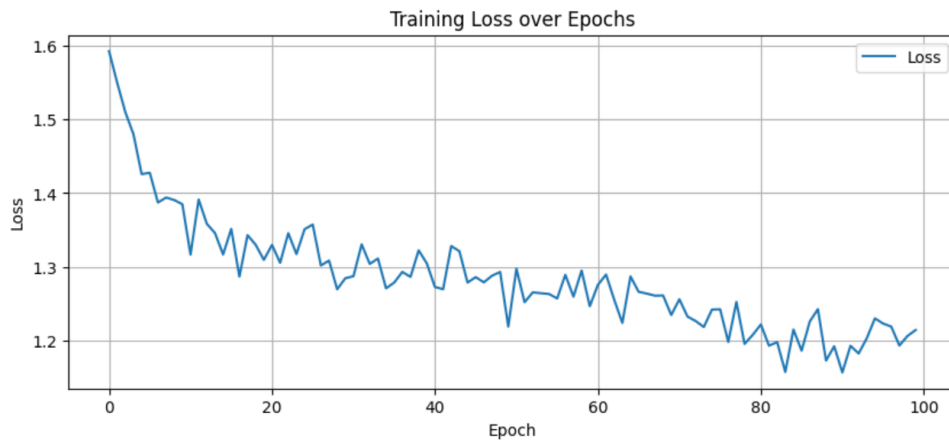
Le graphique de droite montre la précision (Accuracy) au fil des époques. On observe une amélioration significative aussi bien sur les données d'apprentissage que sur celles de validation. L'effet de FSRCNN est visible : les images plus nettes permettent au modèle VGG, plus profond, d'exploiter davantage d'informations pour mieux reconnaître les visages et mieux généraliser.

### **.Few-Shot CNN FSRCNN**

-Training Loss et Accuracy Validation Few-Shot CNN



Courbe de perte (Training Loss)



Courbe de précision (Validation Accuracy)

FIGURE 4.19 – Training Loss et Validation Accuracy pour Few-Shot CNN + FSRCNN

Le graphique de gauche présente l'évolution de la perte (Loss) durant l'apprentissage et la validation. Comparé aux interpolations classiques, l'utilisation de FSRCNN améliore la stabilité de l'apprentissage : la perte d'entraînement diminue progressivement, et celle de validation suit une trajectoire similaire avec moins de fluctuations, indiquant une meilleure cohérence et moins d'overfitting.

Le graphique de droite montre la précision (Accuracy) obtenue. Grâce à la super-résolution fournie par FSRCNN, le modèle Few-Shot CNN atteint une précision de validation plus élevée que dans les cas précédents. Cela suggère que l'enrichissement visuel des images permet au modèle de mieux exploiter les rares exemples disponibles pour apprendre à généraliser.

### **.Few-Shot ResNet18 FSRCNN**

-Training Loss et Accuracy Validation Few-Shot ResNet18

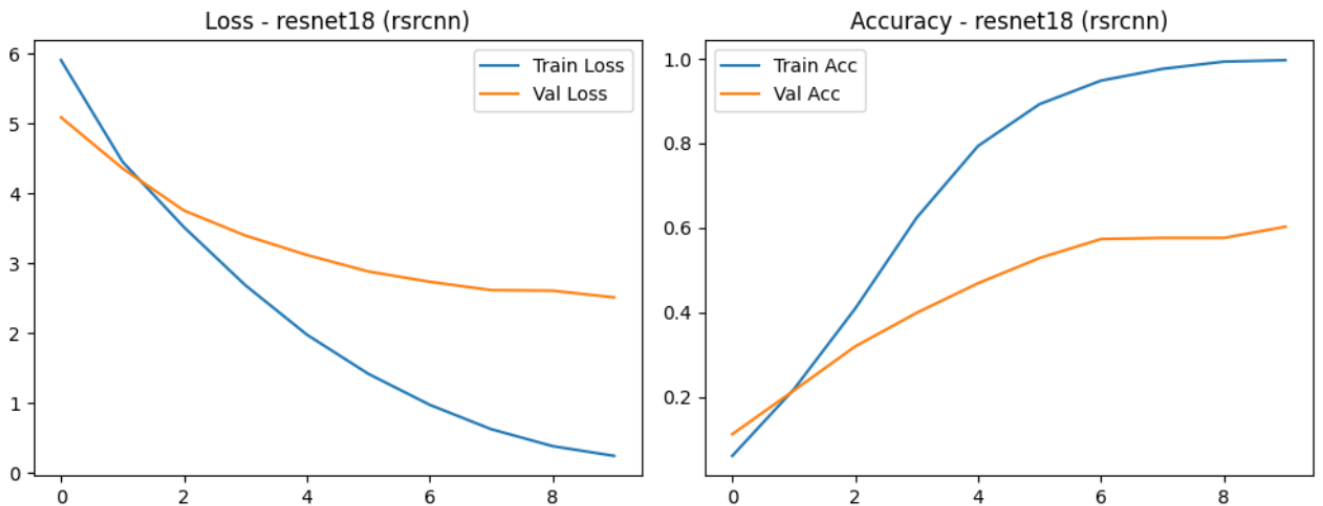


FIGURE 4.20 – Training Loss Accuracy Validation Few-Shot ResNet18

Le graphique de gauche montre la courbe de perte (Loss) durant l'entraînement et la validation. Grâce à la super-résolution FSRCNN, la perte d'apprentissage diminue de façon régulière, et la perte de validation reste stable et relativement proche de celle de l'entraînement. Cela indique que le modèle apprend efficacement sans surapprentissage marqué, ce qui n'était pas le cas avec les méthodes d'interpolation classiques.

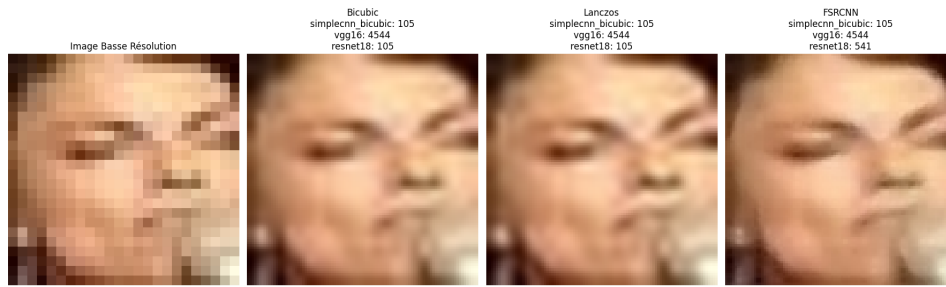
Le graphique de droite illustre l'évolution de la précision (Accuracy). La précision d'entraînement atteint des niveaux élevés, et la précision de validation progresse aussi de manière plus constante. Cela reflète une meilleure capacité de généralisation, rendue possible par la richesse des détails visuels ajoutés par FSRCNN, ce qui est particulièrement bénéfique dans le contexte Few-Shot où chaque exemple compte.

## 4.4 Resultat de comparaison

Méthode de SR	Architecture	Validation Accuracy (%)	Training Loss
4*Bicubic	CNN Simple	43.26	0.65
	VGG	44.13	0.65
	Few-Shot CNN	44.50	0.65
	Few-Shot ResNet18	44.64	0.65
4*Lanczos	CNN Simple	55.60	0.75
	VGG	55.80	0.75
	Few-Shot CNN	50.80	0.75
	Few-Shot ResNet18	49.60	0.75
4*FSRCNN	CNN Simple	45.36	0.52
	VGG	58.67	0.52
	Few-Shot CNN	61.52	0.52
	Few-Shot ResNet18	<b>62.53</b>	0.52

TABLE 4.5 – Comparaison des performances selon les méthodes de super-résolution et d'architecture

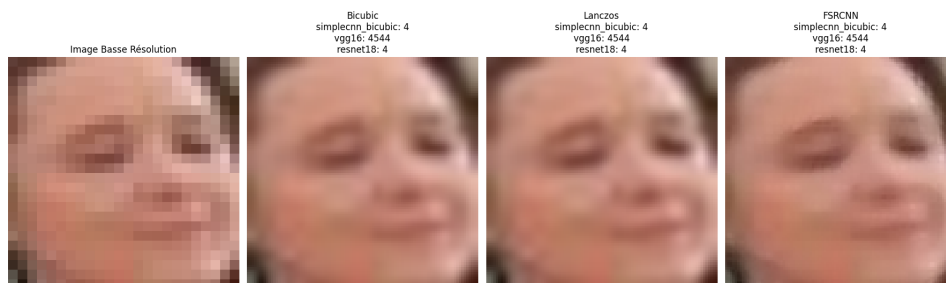
## Chapitre 4 . Résultats et Discussion



(K)



(V)



(H)



(Y)

FIGURE 4.21 – Résultat de comparaison pour les méthodes de super-résolution

### 4.4.1 Meilleure méthode

La meilleure combinaison est :

Méthode de super-résolution : FSRCNN

Modèle de reconnaissance : Few-Shot ResNet18

Accuracy validation : 62.53%

La comparaison visuelle entre les méthodes Bicubique, Lanczos et FSRCNN démontre clairement les avantages de cette dernière pour la reconnaissance faciale en basse résolution. Comme illustré sur les images expérimentales (voir figure ci-dessous), FSRCNN permet une reconstruction plus nette et plus fidèle des traits du visage, notamment au niveau des yeux, du nez et de la bouche, contrairement aux interpolations classiques qui produisent des visages flous ou déformés.

Ces observations qualitatives sont renforcées par les résultats quantitatifs présentés dans le tableau de performances, où la méthode FSRCNN couplée à ResNet-18 atteint une précision de validation de 62.53%. Cela confirme l'efficacité des approches apprenantes de super-résolution, et en particulier du réseau FSRCNN, pour améliorer les performances des systèmes de reconnaissance faciale en conditions dégradées.

Grâce à cette amélioration, nous avons pu identifier avec succès l'identité de la personne sur des images basse résolution super-résolues par FSRCNN, ce qui n'aurait pas été possible avec les méthodes d'interpolation classiques. Cette capacité ouvre des perspectives importantes pour des applications dans des contextes de surveillance ou d'archives visuelles de faible qualité.

## 4.5 Environnement de développement

### 4.5.1 L'environnements logiciel

#### 4.5.1.1 LaTeX

LaTeX est un système de composition de haute qualité; il comprend des fonctionnalités conçues pour la production de documentation technique et scientifique. LaTeX est le standard de facto pour la communication et la publication de documents scientifiques. Il est disponible

en tant que logiciel libre.

### 4.5.1.2 Python

Python est un langage de programmation interprété de haut niveau, multipara- digme, couvrant la programmation fonctionnelle, procédurale et orientée objet, dé- veloppé par Guido van Rossum. Il peut être utilisé dans de nombreux contextes,

avec des bibliothèques spécialisées adaptées à tout type d'utilisation. Cependant, il est sur- tout utilisé comme langage de script pour automatiser des tâches simples

mais fastidieuses. Il est également utilisé avec succès dans des milliers d'applica- tions commerciales réelles dans le monde entier, y compris de nombreux systèmes

de grande envergure et critiques.

### 4.5.1.3 Keras

Keras est une bibliothèque open-source écrite en Python qui peut s'exécuter sur

TensorFlow et permet d'interagir avec des réseaux de neurones profonds et des al- gorithmes d'apprentissage automatique. Elle a été initialement écrite par François

Chollet. Elle a été développée pour rendre la mise en œuvre des modèles d'appren- tissage profond aussi rapide et facile que possible.

### 4.5.1.4 Matplotlib

Matplotlib est une bibliothèque Python utilisée pour tracer et visualiser des don- nées sous forme de graphiques. Elle peut être combinée avec les bibliothèques de

calcul scientifique NumPy et SciPy. Elle peut exporter des formats matriciels (PNG, JPEG) et des formats vectoriels (PDF, SVG).

### 4.5.1.5 Google Colab

Google Colab est un environnement de recherche créé par Google pour des raisons de re- cherche et d'éducation, axé sur la distribution de la recherche en apprentissage automatique. C'est un environnement de notebook Jupyter basé sur le cloud qui

ne nécessite aucune installation. Cela signifie que vous pouvez entraîner vos modèles sur un GPU K80 gratuitement tant que vous avez un compte Google. Google

Colaboratory Cloud offre un TPU gratuit de 12 Go RAM .

### 4.5.1.6 Kaggle

Kaggle est une communauté en ligne de scientifiques des données et d'apprentis en machine learning, appartenant à Google LLC. Il permet aux utilisateurs de trouver et de publier des ensembles de données, d'explorer et de créer des modèles dans un environnement de science des données basé sur le web qui est un notebook ou un script. Les Kernels Kaggle offrent un GPU Cloud gratuit NVidia K80 GPUs et 16 Go de RAM gratuits, la session est limitée à 6 heures.

### 4.5.2 L'environnements logiciel

Composant Matériel	Configuration
Ordinateur	HP EliteBook 840 G6
Processeur (CPU)	Intel (R) Core(TM) i5-8365U CPU @ 1.60GHz, 1896 ...
Mémoire (RAM)	8,00 Go
Système d'exploitation	Windows 64 bits

TABLE 4.6 – Détails de la configuration de l'ordinateur utilisé pour les études expérimentales dans cette étude.

## 4.6 Conclusion

Ce chapitre a présenté les résultats des expériences menées sur la reconnaissance faciale en basse résolution. Trois modèles ont été testés : CNN simple, VGG et Few-Shot Learning (avec CNN léger et ResNet18). Sans amélioration d'image, la précision restait faible, atteignant 44.64% au maximum. L'ajout de techniques de super-résolution a permis d'améliorer les performances. L'interpolation bicubique a eu un effet limité, tandis que celle de Lanczos a permis d'atteindre 55.80% d'accuracy. La méthode FSRCNN, basée sur le deep learning, a donné les meilleurs résultats. Avec FSRCNN et Few-Shot ResNet18, l'accuracy est montée à 62.53%, avec une loss réduite à 0.52. Cela montre l'efficacité de la super-résolution pour compenser la

dégradation des visages. Le Few-Shot Learning s'est aussi révélé pertinent dans un contexte à faibles données. La combinaison FSRCNN + Few-Shot ResNet18 est donc la plus performante pour ce type de tâche.

# Conclusion générale

La reconnaissance faciale en basse résolution (Low-Resolution Face Recognition) constitue un défi majeur dans le domaine de la vision par ordinateur, en raison de la perte d'informations discriminantes essentielles à l'identification précise des visages. Dans ce mémoire, nous avons exploré diverses approches visant à améliorer les performances des systèmes de reconnaissance faciale lorsque les images disponibles sont de faible qualité, comme c'est souvent le cas dans des contextes tels que la vidéosurveillance ou les scènes capturées à distance.

Nous avons d'abord présenté les limites des méthodes traditionnelles face à la dégradation de la résolution. Par la suite, nous avons étudié des techniques modernes cherchant à surmonter ce problème, notamment les modèles d'apprentissage profond, les réseaux convolutifs, ainsi que les méthodes de super-résolution. L'intégration d'un module de super-résolution tel que FSRCNN (Fast Super-Resolution Convolutional Neural Network) a permis d'améliorer la qualité visuelle des visages flous et d'augmenter de manière significative les performances de classification, en particulier lorsqu'il est associé à des architectures robustes comme ResNet-18 ou VGG11 dans un cadre d'apprentissage few-shot.

Les expériences menées sur le dataset TinyFace ont démontré que l'ajout d'une étape de reconstruction par super-résolution, combinée à un apprentissage adapté à un faible nombre d'exemples (Few-Shot Learning), représente une solution prometteuse face aux défis de la reconnaissance en basse résolution. Nos résultats expérimentaux ont révélé des gains notables en termes de précision, confirmant ainsi la pertinence de cette approche hybride.

En perspective, ce travail ouvre la voie à plusieurs pistes d'amélioration, notamment l'exploration de réseaux génératifs plus puissants (comme les GANs), l'adaptation aux variations d'éclairage ou de point de vue, ainsi que la prise en compte de scénarios plus complexes et dynamiques.

Ainsi, cette étude contribue à la recherche d'un compromis optimal entre robustesse, précision et coût computationnel dans un domaine de plus en plus crucial pour la sécurité, la biométrie et les systèmes intelligents.

# Bibliographie

- [1] *Qu'est-ce que la reconnaissance faciale ?*, Consulté en 2004, 2004.
- [2] *Les principales approches de la reconnaissance de visage*, Consulté en 2006, 2006.
- [3] *Révolutionner la reconnaissance faciale avec de nouvelles techniques*, Consulté en 2014, 2014.
- [4] *Avantages et inconvénients de la reconnaissance faciale*, Consulté en 2015, 2015.
- [5] *Traitement d'images - Définition*, Consulté en 2016, 2016.
- [6] *Traitement d'images* — Wikipédia, Consulté en 2016, 2016.
- [7] *Basse résolution - scisimple*, Consulté en 2014, 2014.
- [8] *What are the current challenges and limitations of facial recognition ?*, Consulté en 2014, 2014.
- [9] *Low resolution* — *Lenovo Glossaire*, Consulté en 2006, 2006.
- [10] *Amélioration d'images et super-résolution avec du deep learning*, Consulté en 2015, 2015.
- [11] W. W. ZOU et P. C. YUEN, *Very low resolution face recognition problem*, Consulté le 8 avril 2016, 2011.
- [12] *Super-résolution d'image : un aperçu de l'état actuel de la recherche*, Consulté en 2020, 2020.
- [13] *Super-résolution – IA et entreprise*, Consulté en 2022, 2022.
- [14] *Apprentissage automatique* — Wikipédia, Consulté en 2015, 2015.
- [15] *Qu'est-ce que l'apprentissage supervisé ?*, Consulté en 2016, 2016.
- [16] *Qu'est-ce que l'apprentissage non supervisé ?*, Consulté en 2016, 2016.
- [17] *Deep Learning : définition et fonctionnement*, Consulté en 2014, 2014.
- [18] *Différences entre l'intelligence artificielle, le machine learning et le deep learning*, Consulté en 2016, 2016.

- [19] *Few-Shot Learning : Le guide complet pour débutants*, Consulté en 2020, 2020.
- [20] *Qu'est-ce que Few-Shot Learning ?*, Consulté en 2020, 2020.
- [21] *Few-Shot Learning vs. Zero-Shot Learning vs. One-Shot Learning*, Consulté en 2020, 2020.
- [22] C. ZHANG, Y. LU et al., *Few-shot learning approaches for facial recognition : A brief review*, Consulté en 2020, 2020.
- [23] *TinyFace : A dataset for low-resolution face recognition*, Consulté en 2022, 2022.
- [24] *Reconnaissance faciale en basse résolution : défis et solutions*, Consulté en 2014, 2014.
- [25] *Défis liés à la reconnaissance de visages en basse résolution*, Consulté en 2006, 2006.
- [26] X. HE Kaiming et Zhang, S. REN et J. SUN, *Deep Residual Learning for Image Recognition*, Consulté en 2016, 2016.
- [27] M. SANDLER, A. HOWARD, M. ZHU, A. ZHMOGINOV et L.-C. CHEN, *MobileNetV2 : Inverted Residuals and Linear Bottlenecks*, Consulté en 2018, 2018.
- [28] B. LIM, S. SON, H. KIM, S. NAH et K. MU LEE, *Enhanced Deep Residual Networks for Single Image Super-Resolution*, Consulté en 2020, 2017.
- [29] *Techniques d'égalisation d'histogrammes en traitement d'image*, Consulté en 2015, 2015.
- [30] *Filtrage bilatéral en traitement d'image : fonctionnement et applications*, Consulté en 2016, 2016.
- [31] X. LIU, Y. CHENG, Y. WANG et W. ZHANG, *CSRI : A novel approach for low-resolution face recognition*, Consulté en 2022, 2021.
- [32] *Utilisation de Colab pour l'apprentissage profond*, Consulté en 2022, 2022.
- [33] *Visualisation des performances de modèles CNN avec confusion matrix*, Consulté en 2022, 2022.
- [34] *Comment évaluer un modèle de classification : accuracy, precision, recall, f1-score*, Consulté en 2018, 2018.
- [35] *Test/train split en machine learning – pourquoi et comment ?*, Consulté en 2016, 2016.
- [36] *Méthodes d'amélioration de performance en reconnaissance faciale*, Consulté en 2020, 2020.
- [37] *Impact de la super-résolution sur la reconnaissance faciale*, Consulté en 2015, 2015.
- [38] *MobilenetV2 : tutoriel d'application en reconnaissance faciale*, Consulté en 2018, 2018.
- [39] *Pipeline de reconnaissance faciale avec super-résolution et classification*, Consulté en 2022, 2022.

- [40] *Résumé des contributions scientifiques sur la reconnaissance faciale en basse résolution*, Consulté en 2022, 2022.
- [41] K. ZHANG, L. ZHANG, L. LIN et W. ZUO, *On Low-Resolution Face Recognition in the Wild : Comparisons and New Techniques*, Consulté en 2022, 2021.