الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

Ministère de L'enseignement Supérieure de la recherche scientifique

Université 8 Mai 45 –Guelma-

Faculté des Mathématiques, d'informatique et des Sciences de la Matière

Département d'Informatique

Mémoire de Fin d'études Master

Filière : Informatique

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

Thème :

Système de Discrimination Visages / Faux Visages par Réseaux de Neurones Convolutifs (CNN)

Encadré par : Dr.Bencheriet Chemesse Ennahar

Présenté par :

Yahiaoui Abdennour

Septembre 2021

Remerciement

Tout d'abord, Je tiens à remercier chaleureusement Madame Chemmse Ennahar Benchriet et Madame Yamina Bourjiba, mes encadreurs de mémoire pour leurs disponibilités permanentes, leurs soutiens tout au long de mes recherches mais aussi leurs conseils très précieux.

Veuillez bien mesdames recevoir ma grande estime pour le grand honneur que vous m'avez fait d'accepter l'encadrement de ce travail.

Je tiens également à adresser toute ma gratitude aux membres de jury de soutenance d'avoir accepté de superviser ce travail.

Enfin je remercie toute ma famille et tout particulièrement ma mère pour sa patience et son soutien tout au long de mon parcours.

Dédicace

Je dédie ce modeste mémoire de fin de master à mes chers parents qui ont été à mes côtés pour me soutenir et m'encourager le long de mon cursus.

A mes frères et plus particulièrement mon très cher frère cadet Achraf et ma chère sœur Ibtissam à qui je souhaite un avenir plein de réussite.

A tous mes proches qui m'ont apporté leurs sollicitudes pour accomplir ce travail surtout mes grands-parents, mes oncles, ma tante, mes cousins et mes cousines.

Sans oublier tous mes amis et l'ensemble des étudiants de ma promotion.

Résumé

Un réseau d'adversaire génératif (GAN) est un modèle génératif de premier plan qui est largement utilisé dans diverses applications. Il est possible d'obtenir des images de faux visages d'une grande qualité visuelle grâce à ce modèle. Ces visages fictifs ont le potentiel d'avoir un impact significatif sur la manière dont les gens déterminent la légitimité.

L'identification des fausses images devient une question très importante dans le domaine de l'analyse d'images qui feront l'objet d'un apprentissage profond.

L'objectif principal de notre application est la détection des faux visages (Fake Faces) en utilisant les réseaux adversaire génératifs (GAN).

La capacité de l'algorithme des GAN s'améliorera au fil du temps ce qui rend la détection des faux visages difficile. Tout détecteur aura une courte durée de vie sur le fait que la détection des faux visages peut résoudre le problème à court terme, mais à long terme, la solution pratique sera les techniques d'authentification face à la robustesse des générateurs des faux visages par rapport aux discriminateurs.

Nous proposons donc dans notre application une architecture basée sur l'entrainement de trois discriminateurs du DCGAN à l'élément clé, c'est la méthode d'entrainement utilisée pour chaque cas.

L'entrainement du système a été effectué sur 140000 échantillons confondus entre les vrais (70000) et les faux visages (70000) où la précision de l'apprentissage obtenue était de 98% et celle du test de 94%

Ces résultats considéré comme très prometteurs vu la robustesse des générateurs qui ont fait un grand saut par rapport aux discriminateurs et détecteurs des fausses images.

Mots-clés: faux visage, vrais visages, CNN, GAN, DCGAN, Deepfake.

Abstract

A generative adversary network (GAN) is a leading generative model that is widely used in various applications. It is possible to obtain images of fake faces with high visual quality using this model. These fake faces have the potential to have a significant impact on how people determine legitimacy.

The identification of fake images is becoming a very important issue in the field of image analysis that will be subject to deep learning.

The main objective of our application is the detection of Fake Faces using Generative Adversarial Networks (GAN).

The capability of the GAN algorithm will improve over time making the detection of fake faces difficult. Any detector will have a short lifetime on the fact that false face detection can solve the problem in the short term, but in the long term, the practical solution will be authentication techniques in the face of the robustness of false face generators compared to discriminators.

We therefore propose in our application an architecture based on the training of three discriminators from the DCGAN to the key element, which is the training method used for each case.

The training of the system was carried out on 140,000 samples between the real (70,000) and false (70,000) faces where the accuracy of the training obtained was 98% and that of the test of 94%

These results are considered very promising given the robustness of the generators that have made a great leap compared to the discriminators and detectors of false images.

Keywords: fake face, real face, CNN, GAN, DCGAN, Deepfake.

ملخص

الشبكة التوليدية الخصومية (GAN) هي نموذج توليدي رائد يستخدم على نطاق واسع في مختلف التطبيقات. من الممكن الحصول على صور لوجوه مزيفة بجودة بصرية عالية باستخدام هذا النموذج. يمكن لهذه الوجوه الوهمية أن يكون لها تأثير سلبي بارز على الناس.
أصبح التعرف على الصور المزيفة قضية مهمة للغاية في مجال تحليل الصور والتي ستكون موضوع التعلم العميق.
الهدف الرئيسي لتطبيقنا هو اكتشاف الوجوه المزيفة باستخدام الشبكات التوليدية الخصومية (GAN).
ستتحسن قدرة خوارزمية GAN بمرور الوقت مما يجعل اكتشاف الوجوه المزيفة أمرًا صعبًا. سيكون لأي كاشف عمر قصير حيث يمكن لاكتشاف الوجه المزيف أن يحل المشكلة على المدى القصير، ولكن على المدى الطويل سيكون الحل العملي هو تقنيات المصادقة في مواجهة متانة مولدات الوجه المزيفة ضد أدوات التمييز.
لذلك نقترح في تطبيقنا بنية تعتمد على تدريب ثلاثة مميّزات لـ DCGAN على العنصر الأساسي، وهذه هي طريقة التدريب المستخدمة لكل حالة.
تم تنفيذ تدريب النظام على 140.000 عينة مشوشة بين الوجوه الحقيقية (70.000) والوجوه الزائفة (70.000) حيث كانت دقة التعلم التي تم الحصول عليها 98٪ وتلك الخاصة بالاختبار 94٪ تعتبر واعدة جداً نظراً لقوة المولدات التي صنعت قفزة كبيرة مقارنة بالمميزات وكاشفات الصور الزائفة.
أن تعتبر واعدة جدا بالنظر إلى متانة المولدات التي حققت قفزة كبيرة مقارنة بأجهزة كشف الصور الخاطئة.

كلمات مفتاحية : faux visage, vrais visages, CNN, GAN, DCGAN, Deepfake

Table des matières

Liste des figures

Liste des tableaux

Introduction générale

Les GAN sont des modèles génératifs qui sont largement utilisés dans différentes applications. Récemment, des études ont montré qu'il est possible d'avoir des images de faux visages d'une grande qualité visuelle grâce à ce modèle. Ces visages fictifs ont le potentiel d'avoir un impact significatif sur la manière dont les gens déterminent la légitimité.

Cette technologie de génération et de modification peut affecter le public et la sauvegarde des droits de l'homme. En plus, les faux visages peuvent être utilisés de façon malveillante comme source de manipulation de d'harcèlement.

L'identification des images de faux visages devient une question très importante dans le domaine de l'analyse d'images et un défi techniquement exigeant. Il semble donc important de concevoir un système pouvant discriminer les faux visages des vrais.

Plusieurs recherches ont été effectuées pour l'amélioration du processus de détection de faux visages et cela nécessite des collaborations dans l'ensemble du secteur technologique pour mettre sur pied la détection de ces faux visages et d'inciter les chercheurs du monde entier à créer des technologies innovantes qui permettent de détecter les deepfakes.

Nous avons donc focalisé notre objectif sur ce dernier point qui consiste à la mise au point d'une architecture à base de réseau GAN pour la détection des faux visages.

Pour cela et face à la robustesse des générateurs disponibles aujourd'hui, capable de générer des visages que même les êtres humains ne peuvent les distinguer des vrais. Nous avons donc mis au point et entraîné un réseau discriminateur du DCGAN de trois manières différentes afin d'obtenir le meilleur taux de reconnaissances des faux visages.

Notre mémoire est organisé en quatre chapitres comme suit :

Chapitre 1 : Réseaux de neurones convolutionnels

Ce chapitre fait l'objet d'un travail détaillé sur les réseaux de neurones convolutionnels et les réseaux adversaires génératifs.

Chapitre 2 : Etat de l'art

Dans ce chapitre, nous nous sommes concentrés sur les notions de base et types des faux visages suivies d'un état de l'art de quelques études récentes sur la détection des faux visages.

Chapitre 3 : Conception

Dans ce chapitre, nous présentons la conception et l'architecture détaillée du système composé.

Chapitre 4 : Implémentation

Nous avons évoqué l'aspect implémentation de notre application et les principaux résultats obtenus.

Nous terminons cette étude par une conclusion générale et des points de vus pour des travaux futures qui seront élaborés pour d'autres étudiants.

Chapitre 1
Réseaux de Neurones Convolutionnels

1. Introdution

Les réseaux neuronaux et l'apprentissage en profondeur donnent de nos jours de bonnes solutions à plusieurs problèmes de reconnaissance et de classification d'image, raison pour laquelle ils sont devenus sujets importants en informatique et dans l'industrie économique. Inspirés de la structure du cerveau de l'être humain, les réseaux de neurones artificiels font partie des moyens qui permettent de rendre les ordinateurs plus humains et aider ainsi les machines à raisonner comme des êtres humains. Le cerveau humain apprend d'une manière essentielle par l'expérience. Il s'agit donc d'une preuve naturelle que certains problèmes peuvent dépasser la portée des ordinateurs.

2. Système neurophysiologique

Un réseau de neurone artificiel est un ensemble interconnecté d'éléments de traitement simple dont la fonctionnalité s'inspire du neurone biologique. Le cerveau humain est constitué d'environ 100 milliards de cellules nerveuse ou neurones (figure 1.1). Les neurones communiquent par des signaux électriques et les connections entre elles sont assurées par des jonctions électrochimiques appelées synapse, située sur des branches de la cellule appelée dendrites. Chaque neurone reçoit généralement plusieurs milliers de connection avec d'autres neurones et reçoit ainsi en permanence une multitude de signaux entrants qui finissent par atteindre le corps cellulaire. Si le signal resultant dépasse un certain seuil, le neurone se déclenche ou génère une impulsion de tension en réponse. Celle-ci est ensuite transmise à d'autres neurones par une fibre ramifieé appeleé axone [1].

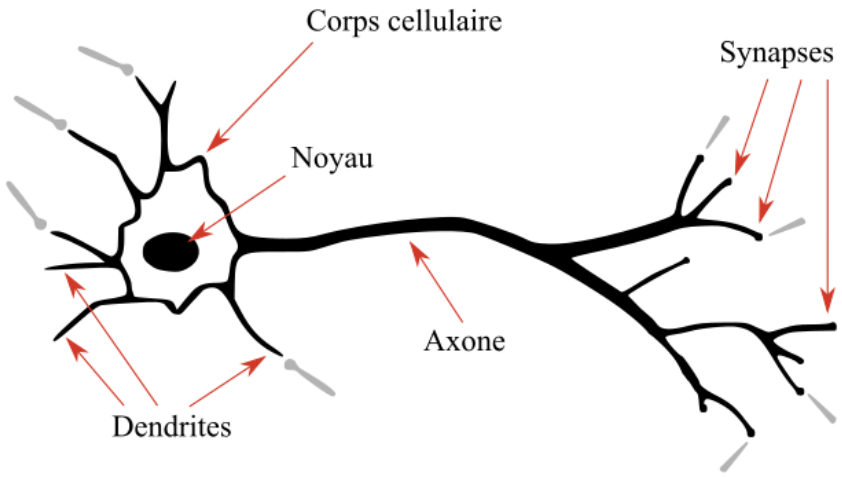

Figure 1.1 Neurone biologique

3. Réseaux de neurones artificiels

Un réseau de neurone artificiel est un modèle de calcul inspiré des neurones naturels. La complexité des neurones réels est fortement abstraite lors de la modélisation des neurones artificiels. Ceux-ci consistent essentiellement en des entrées (comme les synapses) qui sont multipliées par des poids (force des signaux respectifs), puis calculées par une fonction mathématique qui détermine l'activation du neurone. Une autre fonction calcule la sortie du neurone artificiel. Les ANN combinent des neurones artificiels pour traiter l'information [2]

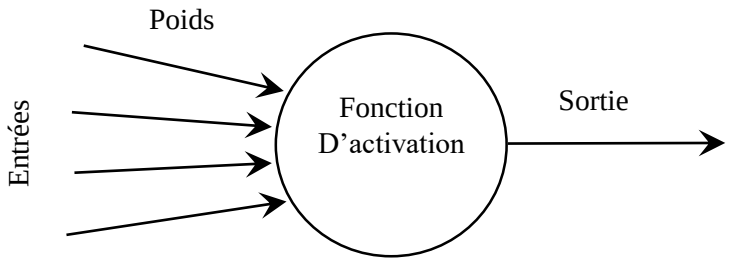

Figure 1.2 Neurone artificiel

4. Structure d'interconnexion

La connexion à travers les neurones qui constituent le réseau décrit la topologie du modèle. Elle peut être quelconque, mais le plus souvent il est possible de distinguer une certaine régularité [3].

4.1.Réseau à connexion locale (feedforward)

Un réseau neuronal à connexion locale est un réseau neuronal artificiel dans lequel les connexions entre les nœuds ne forment pas un cycle. Le réseau neuronal à connexion locale est le modèle le plus simple de réseau neuronal, car l'information n'est traitée que dans un seul sens. Bien que les données passent par plusieurs nœuds, elles se déplacent toujours dans une seule direction et jamais dans le sens contraire [w1].

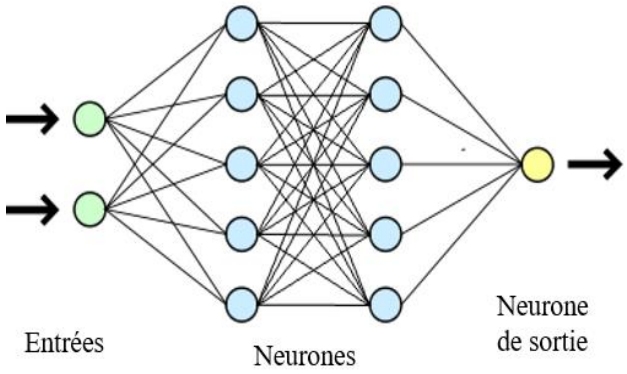

Figure 1.3 Réseau à connexions locales

4.2.Réseau à connexion récurrente

Le réseau neuronal récurrent (RNN) est une généralisation du réseau neuronal à connexion locale qui possède une mémoire interne. Le RNN est récurrente par nature parce qu'il réalise le même travail pour chaque entrée de donné, alors que la sortie de l'entrée actuelle résulte du calcul précédent. Après avoir produit la sortie, celle-ci est copiée et renvoyée dans le réseau récurrent. Contrairement aux réseaux neuronaux à connexion locale, les RNN peuvent utiliser leur mémoire pour examiner des séquences d'entrées.

Dans les autres réseaux neuronaux, toutes les entrées sont indépendantes les unes des autres. Mais dans un RNN, toutes les entrées sont liées les unes aux autres [w2].

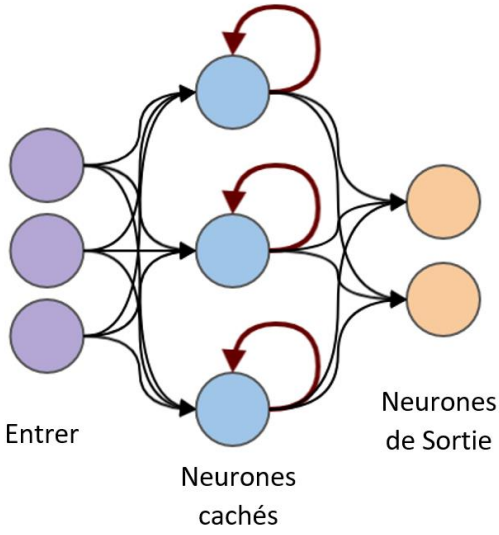

Figure 1.4 Réseau à connexions récurrentes

4.3.Longue mémoire à court terme

Les réseaux de longue mémoire à court terme LSTM (Long Short Term Memory) sont une version modifiée des réseaux neuronaux récurrents, qui facilite la mémorisation des données passées. Les réseaux LSTM sont bien adaptés à la classification, au traitement et à la prédiction de séries temporelles avec des décalages temporels de durée inconnue. Il entraîne le modèle en utilisant la rétropropagation [w2].

Dans un LSTM, nous avons trois portes: porte entrée, porte d'oubli et porte sortie. Ces portes définissent s'il faut ou non laisser entrer une nouvelle entrée (porte d'entrée), supprimer l'information parce qu'elle n'est pas importante (porte oubli) ou la laisser influencer la sortie au pas de temps courant (porte de sortie).

Les portes d'un LSTM sont analogiques, sous la forme de sigmoïdes, c'est-à-dire qu'elles vont de 0 à 1. Cette analogie leur permet de faire une rétropropagation avec elles. Le problème posé par la disparition des gradients est résolu grâce à LSTM [w6].

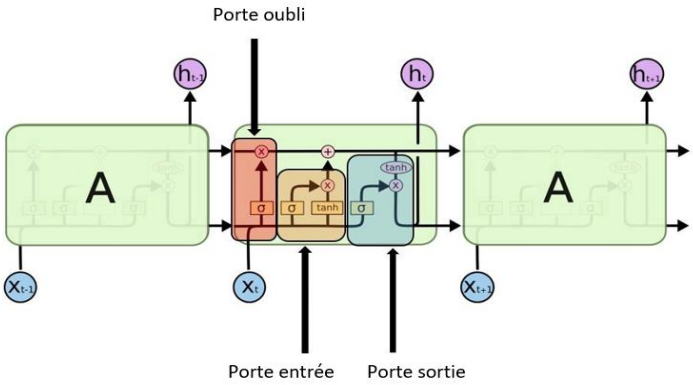

Figure 1.5 Réseau de longue mémoire à court terme

5. Types Apprentissage

5.1.Apprentissage supervisé

L'apprentissage supervisé est une technique d'apprentissage d'utilisation des données étiquetées. Dans le cas de l'apprentissage supervisé, l'environnement dispose d'un ensemble d'entrées et de sorties correspondantes. Il modifiera donc de façon itérative les paramètres du réseau pour avoir une bonne approximation des sorties préférées. Il y a plusieurs approches d'apprentissage pour l'apprentissage profond principalement les réseaux de neurones profonds (DNN), les réseaux de neurones convolutifs (CNN), les réseaux de neurones récurrents (RNN) [4].

5.2.Apprentissage semi-supervisé

L'apprentissage semi-supervisé est un apprentissage basé sur des ensembles de données partiellement étiquetés. Dans certains cas L'apprentissage par renforcement profond et les réseaux génératifs adversarial sont utilisés comme techniques d'apprentissage semi-supervisé [4].

5.3.Apprentissage non supervisé

L'apprentissage non supervisé est un système d'apprentissage qui peut être exécuté sans étiquettes de données. Dans ce cas, l'agent apprend des représentations internes ou d'importantes fonctionnalités pour découvrir des relations ou des structures non connues dans les données d'entrée.

Généralement, le regroupement, la réduction des dimensions et les techniques génératives sont souvent envisagés comme des méthodes d'apprentissage non supervisées. De nombreux membres de la famille du deep learning peuvent être utilisés de manière efficace pour le clustering et la réduction de la dimensionnalité non linéaire y compris les auto-encodeurs (AE), les machines Boltzmann restreintes (R.B.M) et les GAN nouvellement développés.

Par ailleurs, les RNN sont aussi utilisés pour l'apprentissage non supervisé dans de nombreux domaines d'applications[4].

6. Modèles de l'apprentissage profond

L'apprentissage profond est une manière d'apprentissage automatique qui permet aux ordinateurs d'apprendre de l'expérience et de comprendre le monde selon une hiérarchie conceptuelle. Du fait que les ordinateurs collectent des connaissances à partir de l'expérience. Il n'est pas utile qu'un opérateur informatique humain ait besoin de spécifier d'une manière formelle toute les connaissances requises par l'ordinateur. La hiérarchie des concepts permet aux ordinateurs d'apprendre à partir de concepts plus simples en construisant des concepts complexes. Ce graphique hiérarchique serait de plusieurs couches [5].

6.1.Réseau neuronal à convolution

Le réseau neuronal à convolution profonde (CNN) est un type particulier de réseau neuronal qui a bien fonctionné dans certaines compétitions liées à la vision par ordinateur et au traitement d'images. Parmi les domaines d'application les plus célèbres du CNN, on peut citer la classification, la segmentation et la détection d'objets [6].

Cnn a obtenu des résultats révolutionnaires dans différents domaines au cours des dix dernières années. Ces résultats ont incité les chercheurs et les développeurs à utiliser de grands modèles

pour résoudre des taches complexes, ce qui n'était pas possible avec les réseaux ANN classique [7].

6.1.1. Structure globale

Cnn est un modèle d'apprentissage profond pour le traitement des données qui présente des motifs de grille (tels que des images), Ce modèle vise à apprendre automatiquement et à s'adapter à la hiérarchie spatiale des images dans un cadre adaptif. CNN (fig 1.6) est une construction mathématique composée essentiellement de trois types de couches (la convolution, pooling, et fully connected layers). les deux premières sont des couches qui effectuent l'extraction des caractéristiques, tandis que la troisième (fully connected) est une couche entièrement connectée pour classifier les images [8].

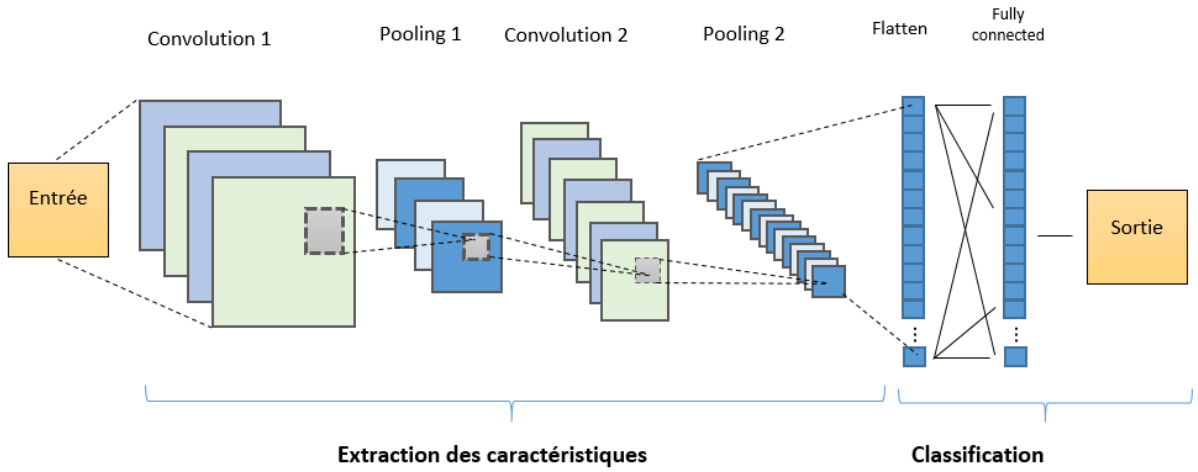

Figure 1.6 Architecture d'un réseau de neurones à convolution

6.1.2. La couche de convolution

La couche de convolution est l'élément essentiel des réseaux de neurones convolutifs et forme presque toujours leur première couche. Son rôle essentiel est de trouver la présence d'un ensemble de caractéristiques dans les images reçues en entrée. C'est pour cela qu'on doit réaliser un filtrage par convolution.

La règle (fig 1.7) est de faire « glisser » une fenêtre représentant la caractéristique sur l'image et de calculer ainsi le produit de convolution entre la caractéristique et chaque partie de l'image balayée.

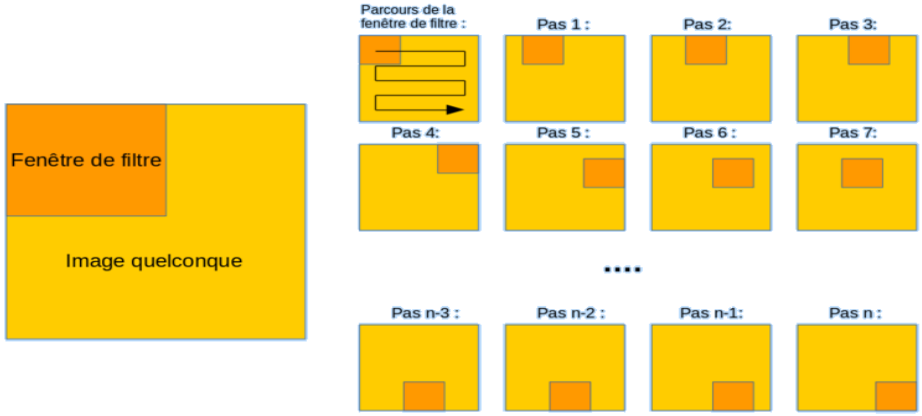

Figure 1.7 Exemple sur l'opération de convolution

Trois grands paramètres permettant alors de dimensionner le volume de la couche de convolution [w3].

-La profondeur de la couche : C'est le nombre de noyaux de convolution voire le nombre de neurones assemblés à un même champ récepteur.

-Le pas (stride) contrôle le chevauchement des champs récepteurs. Nous définissons la distance à laquelle le filtre se déplace d'une position à la suivante par "stride" (fig 1.8).

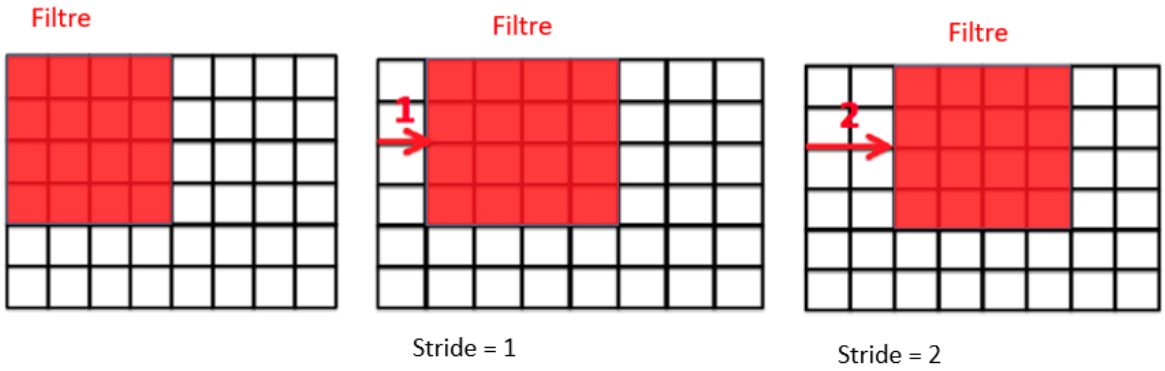

Figure 1.8 Exemple sur le pas

- La marge à zéro (Zéro padding) : Consiste à rajouter des zéro à la frontière du volume d'entrée. Cette marge accorde le contrôle de la dimension spatiale du volume de sortie.

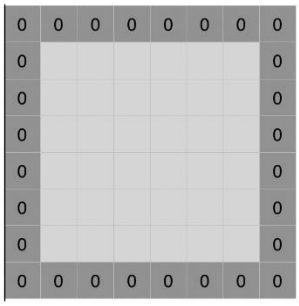

Figure 1.9 Image avec la marge à zéro

6.1.3. Couche de pooling

La couche pooling placé généralement entre deux couches de convolution, cette couche mise en commun vise à réduire progressivement la dimensionnalité de la représentation, et donc à réduire encore le nombre de paramètres et la complexité de calcul du modèle, mais conserve les caractéristiques importantes. Cette opération remplace un carré de pixels (généralement 2×2 ou 3×3) par une seule valeur [8].

Type de pooling :

-pooling moyen : il prend la moyenne de tous les pixels de la sélection (figure 1.10)

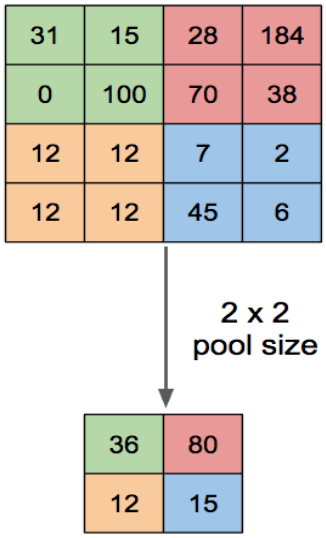

Figure 1.10 Pooling moyen

-pooling max : il prend le pixel qui a la valeur maximale entre tous les pixels de la sélection (figure 1.11)

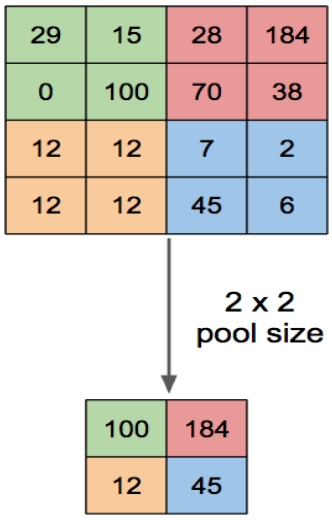

Figure 1.11 Pooling maximal

6.1.4. Couche entièrement connecteé

La sortie de la couche finale de la convolution ou de mise en commun est généralement aplatie. Autrement dit convertie en une carte de caractéristiques. C'est-à-dire transformée en un tableau de nombres (ou de vecteurs) unidimensionnel (1D) qui est connectée à une ou plusieurs couches entièrement connectées [8].

D'autres couches peuvent être utilisé pour améliorer les résultats parmi les quelles, nous citons :

a. Couche de correction

Afin d'améliorer l'efficacité du traitement en introduisant une couche qui exécutera une fonction mathématique (fonction d'activation sur le signal de sortie).

Nous trouverons, dans ce cas

- ReLU (Rectified Linear Units) qui remplace d'une manière significative toutes les valeurs négatives reçues en entrée par des zéros définie par

$$ReLU(x) = \max(0, x)$$

- La couche de correction joue cependant le rôle de fonction d'activation.

La correction ReLU est préférable, mais il existe d'autres comme la correction par tangente hyperbolique et la correction par la fonction sigmoïde, etc [w4].

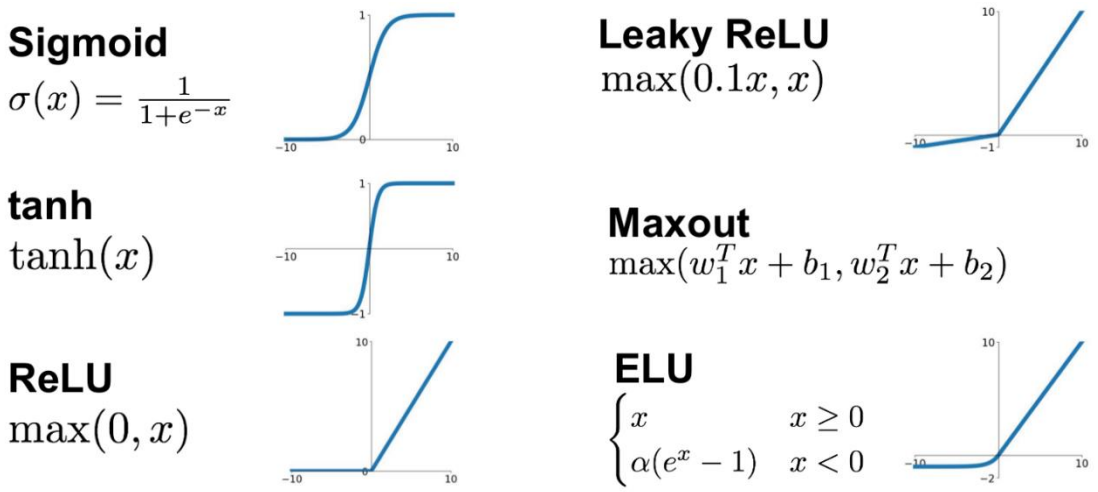

Figure 1.12 Fonctions d'activation [w7]

b. Dropout

Le dropout consiste à ignorer des unités autrement dit des neurones tout au long de la phase de formation d'un certain ensemble de neurones qui est choisi au hasard. Par "ignorer", on veux dire que ces unités ne sont pas prises en compte pendant un passage particulier en avant ou en arrière. A chaque étape de la formation, les nœuds individuels sont soit éliminés du réseau avec une probabilité de 1-p, soit conservés avec une probabilité de p, de telle manière qu'il reste un réseau réduit; les arêtes entrantes et sortantes vers un nœud éliminé sont également supprimées [w5].

6.1.5. Architecture

De nos jours, les CNN sont considérés comme les algorithmes les plus utilisés dans la technologie d'intelligence artificielle (IA) d'inspiration biologique et durant les dix dernières années, plusieurs efforts ont été déployés pour améliorer les performances des CNN et c'est ainsi que différentes améliorations ont été apportées de 1989 jusqu'à nos jours, et selon le type de modification architecturale, CNN (fig 1.12) peut être divisé en sept classes à savoir : exploitation spatiale, profondeur, multi trajets, largeur, exploitation de la carte caractéristique, amplification des canaux et CNN basés sur l'attention [6].

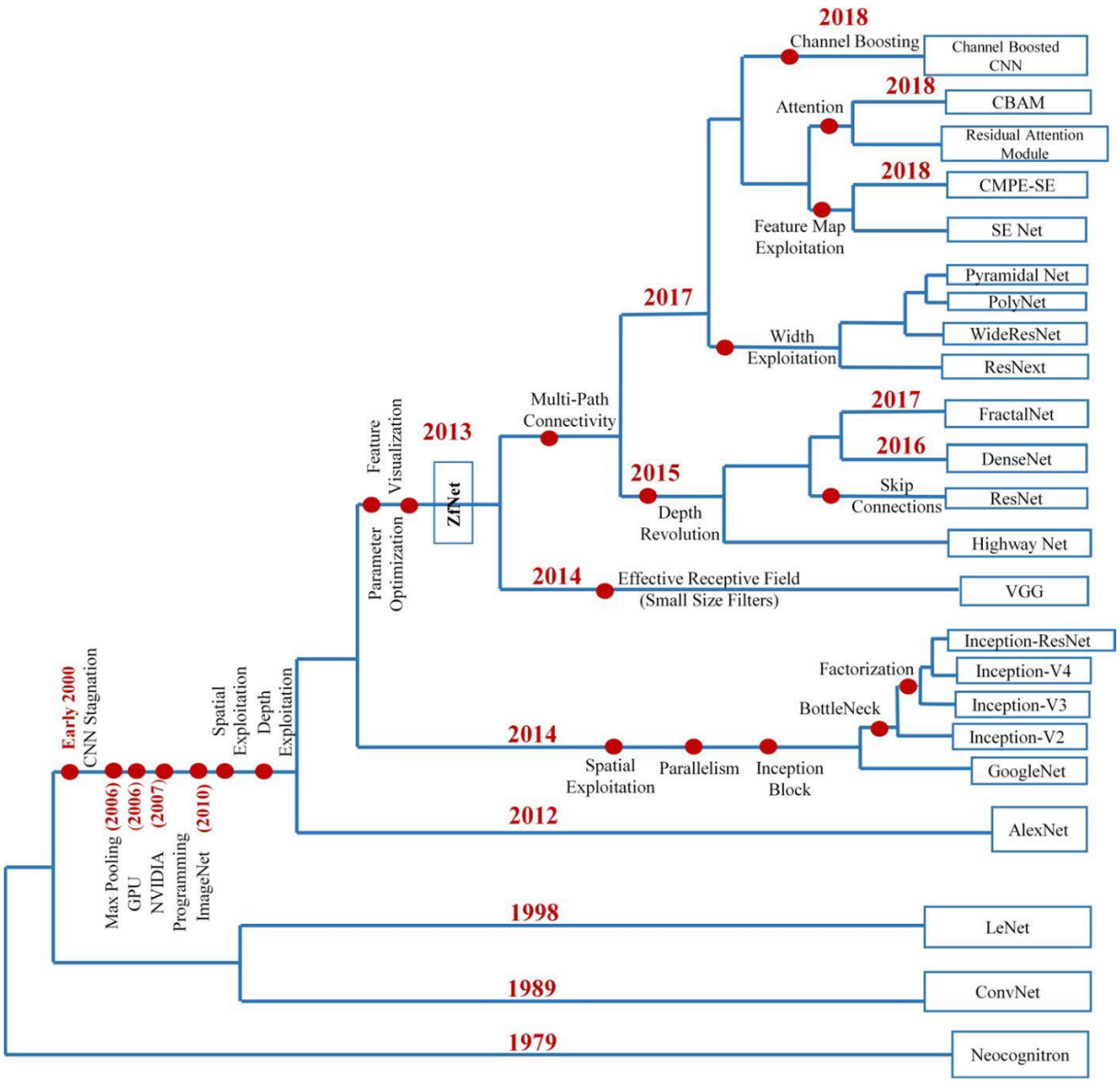

Figure 1.13 Histoire de l'évolution des CNN [6]

6.2.Réseaux adversaires géneratifs

Sur le plan technologique, les Deepfakes et la falsification des vidéos et des photos sont le produit d'un Générative Adversarial Network (GAN) qui consiste en deux neurones artificiels travaillant ensemble pour créer des médias ou supports d'apparence réelle.

Ces deux réseaux appelés « le générateur » et « le discriminateur » sont entraînés sur le même jeu de données d'images, de vidéos ou de sons.

Ensuite le premier essai est de créer de nouveaux échantillons capables de tromper le second réseau qui tente de déterminer si le nouveau média qu'il voit est réel.

Un GAN peut vérifier des milliers de photos d'une personne et générer un nouveau portrait qui se rapproche de ces photos sans être une copie exacte de l'une d'entre elles.

Dans un proche avenir, les GAN seront formés à partir de moins d'informations et seront capables d'échanger des têtes, des corps entiers et même des voix [9].

6.2.1. Le générateur

le générateur prend un bruit aléatoire comme entrée et génère des échantillons comme sortie son but est de générer de tels échantillons qui tromperont le discriminateur en lui faisant croire qu'il voit de vraies images alors qu'en réalité il voit des faux. on peut considérer le générateur comme une contrefaçon [10].

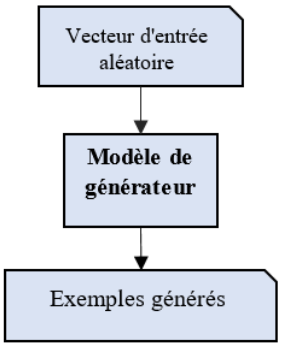

Figure 1.14 Exemple de modèle de générateur de GAN

6.2.2. Le discriminateur

Le discriminateur est un modèle de classification qui prend en exemple du domaine en entrée réel ou générer et prévoit une étiquette de classe binaire réelle ou fausse. Le modèle réel provient de l'ensemble de données d'apprentissage tandis que les exemples générés sont produit par le modèle généré.

Après le processus d'apprentissage, le discriminateur est écarté puisque nous nous interessons au générateur.

Le générateur est parfois réutilisé parce qu'il a appris à extraire d'une manière efficace des caractéristique a partir d'exemples dans le domaine du problème. Ainsi tout ou une partie des couches d'extraction de caractéristiques peuvent être utilisées dans des applications d'apprentissage par transfert [11].

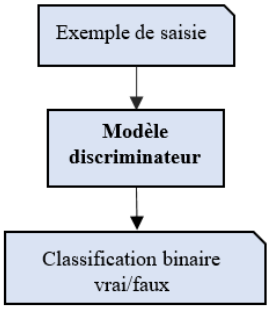

Figure 1.15 Exemple du modèle de discriminateur GAN

6.2.3. Les GAN comme un jeu à deux joueurs

La modélisation générative est un problème d'apprentissage non supervisé quoiqu'une propriété de l'architecture GAN soit conçue comme un problème d'apprentissage supervisé.

Le générateur génère un ensemble d'échantillon ainsi que des exemples réels du domaine sont fournis au discriminateur et répartis comme réels ou faux puisque, le générateur et le discriminateur sont formés ensemble.

Le discriminateur est par la suite mis à jour pour améliorer sa capacité à caractériser les vrais et les faux échantillons au tour suivant et, surtout, le générateur est mis à jour en fonction de l'efficacité avec laquelle les échantillons générés ont trompé le discriminateur. Ainsi, le discriminateur et le générateur sont en compétition l'un contre l'autre, ils sont alors adversaires. Puisqu'ils jouent un jeu à somme nulle qui signifie que lorsque le discriminateur identifie avec succès les vraies et les faux échantillons, il est récompensé ou aucune modification n'est indispensable au paramètres du modèle, tandis que le générateur est pénalisé par de grandes mises à jour des paramètres du modèle.

Tour à tour, quand le générateur trompe le discriminateur, il est récompensé et aucun changement n'est essentiel aux paramètres du modèles mais le discriminateur est pénalisé et ses paramètres de modèle sont mise à jour.

Le discriminateur ne peut pas faire la différence et prévoit « incertain » c'est-à-dire 50% pour le vraie et le faux dans tous les cas il peut alors être éliminé [11].

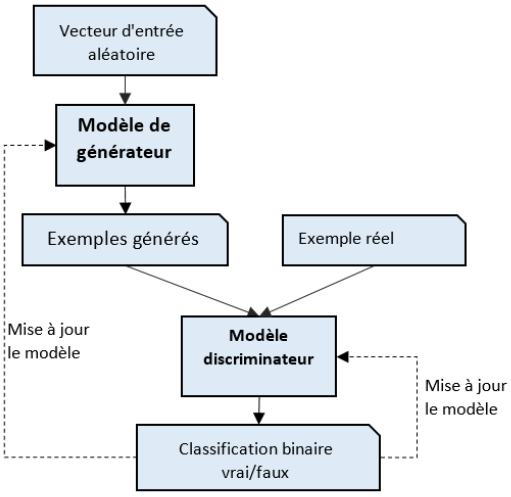

Figure 1.16 Exemple d'architecture du modèle de réseau adversarial génératif

7. Mesures de performances

Pour évaluer les réseaux CNN, il faut calculer un certain nombre de paramètres tels que :

a. Vrai positif 'True positive Tp' : prédiction positive correcte.
b. Vrai négatif 'True negative TN': prédiction negative correcte
c. Faux positif 'false positive FP': prédiction positive incorrecte
d. Faux negatif 'false negative FN': prédiction negative incorrecte

7.1.matrice de confusion

Une matrice de confusion est un tableau souvent utilisé pour résume la performance de classification d'un classificateur par rapport à un ensemble de données de test dont les valeurs réelles sont connues [12].

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

Tableau 1.1 Matrice de confusion

7.2.Exactitude ‘Accuracy’

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

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

7.3.Le Rappel

Le **rappel** permet de répondre à la question suivante :

Quelle proportion de résultats positifs réels a été identifiée correctement ?

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

7.4.La courbe ROC

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

Le taux de vrais positifs (TVP):

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

Le taux de faux positifs (TFP):

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

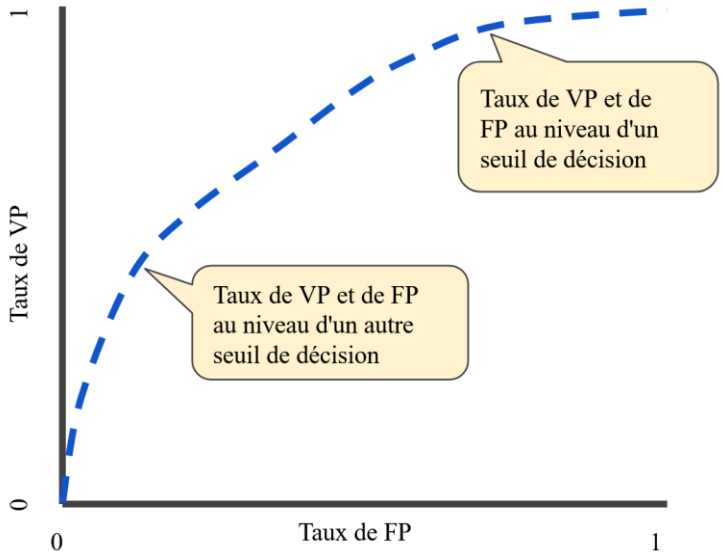

Figure 1.17 La courbe ROC

8. Conclusion

Les CNN est une technique d'apprentissage profond pour les tâches actuelles de reconnaissance visuelle. Comme toutes les techniques d'apprentissage profond, ils sont très dépendants de la taille et de la qualité des données d'entraînement. Avec un jeu de données bien préparé, les CNN sont capables de surpasser les humains dans les tâches de reconnaissance visuelle.

Les GAN constituent un développement intéressant d'apprentissage non supervisé. Ils ont été capables de générer des photos si réalistes que les humains sont incapables de dire qu'elles représentent des objets, des scènes et des personnes qui n'existent pas dans la vie réelle.

Chapitre 2
L'état de l'art

1. Introduction

Les fausses images sont devenues un problème primordial au cours des dernières années, car elles peuvent facilement tromper les êtres humains et entrainer inévitablement de graves risques sociaux comme les fausses preuves et les fausses informations. Ils pourront aussi porter atteinte à des réputations et nuire à des personnes. Générer les faux visages est désormais possible avec un minimum d'efforts et par n'importe qui. C'est pourquoi la détection des deepfackes est de plus en plus importante.

2. Deepfake

La récente avancée technologique a facilité la création des Deepfakes qui sont des vidéos ou des images hyperréalistes consistant à utiliser des échanges de visages ne laissant que peu de traces de manipulation.

Ces Deepfakes sont le résultat d'application d'intelligence artificielle (IA) qui mélangent et combinent des images et des clips vidéo pour créer de fausses vidéos qui apparaissent réelles et naturelles. Cette technologie peut générer par exemple une vidéo humoristique ou politique d'une personne sans l'accord et la permission de la personne dont l'image et la voix sont concernées. Les facteurs qui changent les règles du jeu pour les contrefaçons sont la portée, l'ampleur et la complexité de la technologie impliquée puisque de nos jours n'importe qui avec un ordinateur peut créer des fausses vidéos qui sont presque impossibles à distinguer des vraies vidéos [9].

3. Types des faux visages

La manipulation relative au visage est généralement divisée en deux catégories, à savoir le changement facial et la reconstitution de l'expression du visage. Faceswap et DeepFakes appartiennent à la même catégorie : le visage de l'image source est remplacé par un visage cible. Faceswap est une approche graphique tandis que DeepFakes est basé sur les GANs.

Les méthodes Face2face et NeruralTextures transfèrent les expressions faciales de l'image cible à l'image source. La méthode Face2Face est basée sur le transfert de l'expression de la source en utilisant les coefficients de blendshape à la vidéo cible, tandis que la méthode NeuralTextures utilise des GANs pour apprendre la texture de la personne cible [13].

Figure 2.1 Un exemple de Faceswap

Figure 2.2 Un exemple de reconstitution faciale

4. Historique

Les travaux sur la production de fausse information (images/vidéos) ne sont pas nouveaux et le domaine de la vision par ordinateur existe depuis les années 1990. Ce domaine associe l'intelligence artificielle et le traitement des images et des vidéos numériques afin de créer de nouveaux médias artificiels. En 1997, l'un des premiers projets universitaires qui s'appelait le programme vidéo rewrite [14], il a utilisé l'apprentissage automatique pour automatiser d'une manière complète la réanimation du visage.

Les travaux universitaires et amateurs modernes se sont concentrés sur la rapidité, la simplicité et l'accessibilité du processus. Par exemple le programme face2face [15] qui a été publié en 2016, modifie les séquences vidéo du visage d'une personne pour la représenter en train d'imiter les expressions faciales d'une autre personne en temps réel.

En 2017, le programme synthesizing obama [16] a montré le potentiel de cette technologie mais on peut dire que la plus grande partie des travaux les plus importants sont le fait d'amateurs.

5. Avantages et inconvénients

Les avantages de la technologie Deepfake sont nombreux vu la popularité que présente cette technologie dans le monde entier [W9].

Ces avantages sont :

- Le divertissement qui va nous permettre d'expérimenter des choses qui n'existent plus, par exemple changer la fonction du film ou de la conversation vidéo sans reprendre la prise de vue.
- Susciter une attention extraordinaire parmi le public en ligne, ce qui rend la page web populaire sur le moteur de recherche.
- Prendre conscience des fausses choses et ne pas croire à tout ce que nous voyons autour de nous.

Les inconvénients des Deepfakes sont :

- La fraude et l'escroquerie : les Deepfakes vidéo ont été utilisés dans le cas de chantage
- La crédibilité et l'authenticité des films Deepfakes ont été difficiles à créer et l'utilisation du Face swapping augmente la difficulté de classer les vidéos comme véridiques ou non
- Création de fausses nouvelles et de propagande.

6. Dataset

Nous présentons dans ce qui suit les bases de données les plus utilisées pour la détection des vraies et des faux visages.

6.1.CelebA

Celeb faces attributes Dataset (CelebA) est un jeu de données comprenant plus de 2000000 images de célébrités. Les images de ce jeu de données couvrent de grandes variations de poses et de fouillis d'arrière-plan.

CelebA présente une grande diversité, de grandes quantités et de riches annotations, notamment 10177 identités et 202599 images de visages.

Le jeu de données est utilisé comme jeu d'entraînement et de test pour les tâches de vision par ordinateur tel que la reconnaissance d'attributs de visage, la détection de visage, la localisation des parties du visage et édition de visage [17].

Figure 2.3 Examples de la base CelebA

6.2.CelebA-HQ

Version de meilleure qualité du jeu de données (CelebA) permettant d'expérimenter avec des résolutions de sortie allant jusqu'à 1024x1024 pixels [18]

Figure 2.4 Examples de la base CelebA-HQ

6.3.Flickr-faces-HQ (FFHQ)

Flickr-Faces-HQ est un ensemble d'images de grande qualité de visages de personnes qui a été créé comme références pour les réseaux adversiaux génératifs. L'ensemble de données comprend 70000 images png de haute qualité à une résolution de 1024*1024 et comporte une variation considérable en ce qui concerne l'âge, l'ethnicité et le fond d'image [19].

Figure 2.5 Examples de la base FFHQ

6.4.FaceForensics++

FaceForensics++ est un ensemble de données composé de 1000 séquences vidéo originales qui ont été manipulées à l'aide de quatre méthodes automatisées de manipulation des visages : Deepfakes, Face2Face, FaceSwap et NeuralTextures. Les données proviennent de 977 vidéos youtube et toutes les vidéos contiennent un visage principalement frontal sans occlusions, ce qui permet aux méthodes de manipulation automatisées de générer des faux réalistes [20].

Figure 2.6 Examples de la base FaceForensics++

6.5.IMD2020

Cet ensemble de données contient 35 000 images réelles capturées par 2 322 modèles de caméras différents. Ces modèles de caméras constituent la majorité des caméras existantes sur le marché. L'ensemble de données fournit un ensemble riche et diversifié de bruit de capteur. De plus, la création d'un ensemble d'images manipulées en utilisant une grande variété d'opérations de manipulation, y compris des techniques de traitement d'image ainsi que des méthodes avancées telles que le GAN, ce qui a donné lieu à un ensemble de données de 70 000 images au total [21].

Figure 2.7 Examples de la base IMD2020

6.6.DEFACTO

Le jeu de données a été généré automatiquement à l'aide de la base de données Microsoft common object in context (MSCOCO) afin de produire des contrefaçons sémantiquement significatives. Quatre catégories de falsifications ont été générées. Les contrefaçons par "Splicing", qui consistent à insérer un élément externe dans une image, les contrefaçons par déplacement de copie, qui consistent à dupliquer un élément dans une image, les contrefaçons par suppression d'objet, qui consistent à supprimer des objets dans des images, et enfin les contrefaçons par "morphing", qui consistent à déformer et à mélanger deux images. Plus de 200 000 images ont été générées et chaque image est accompagnée de plusieurs annotations permettant une localisation précise de la falsification et des informations sur le processus de falsification [22].

Figure 2.8 Examples de la base defacto

6.7.Autre base de donnée d'image

Base de données	Année	manipulations	Nombre d'images vrais/ faux	Dimension d'image	format
NC2016 [23]	2016	Splicing,copy-move,removal	560 / 564	500 x 500 – 5616x3744	JPG
FaceSwap [24]	2017	Face swapping	1,758 / 1,927	450 x 338	JPG
MFC2018[23]	2018	divers	14,156 / 3,265	128 x 104 – 7952 x 5304	RAW, PNG, BMP, JPG, TIF
GAN collection [25]	2019	GAN generated	356,000 / 596,000	256 x 256 – 1024 x 1024	PNG

Tableau 2.1 Bases de données images vrais visages/faux visages [26]

6.8.Base de donnée des Deepfake Vidéo

Base de données	année	manipulations	Nb des vidéos vrai/ faux	Dimension du Frame	format
DF-TIMIT [27]	2018	Deepfake	- / 620	64 x 64 – 128 x 128	JPG
DFDC-preview [28]	2019	Deepfake	1.131 / 4.113	180p – 2160p	H.264
Celeb-DF [29]	2020	Deepfake	590 / 5.639	divers	MPEG4
DeeperForensics-1.0 [30]	2020	deepfake	50.000/10.000	1080p	–

Tableau 2.2 Bases de données vidéos deepfakes [26]

7. Travaux connexe sur la détection des faux visages

Nous élaborons dans ce contexte, les derniers travaux sur la détection des faux visages et nous abordons dans ce sujet huit articles récents de la littérature scientifique pour mieux le comprendre.

Article 1 "Deepfake video detection using recurrent neural networks" [31]: ' Détection de vidéos Deepfake à l'aide de réseaux neuronaux récurrents'

Dans cette recherche, les auteurs proposent un système de détection Deepfake vidéo en utilisant un réseau de neurone convolutif reccurent pour extraire les caractéristiques de chaque image qui seront ensuite concaténées et transmisent au LSTM pour analyse. Ils ont produit une estimation de la probabilité que la séquence soit un Deepfake ou une vidéo non manipulée. Les auteurs ont collecté 600 Deepfake vidéo dont 300 vidéos à partir de plusieurs sites internet et 300 vidéos supplémentaires à partir de la base de données HOHA [32]. Ils ont présenté les performances de leur système en termes de précision en employant des sous séquences de longueur 20, 40 et 80 frames.

Pour 20 frames, ils ont obtenu 96.7% de précision, pour 40 frame un pourcentage de 97.1,

Et finalement pour 80 frames 97.1% de précision.

Article 2 "Global Texture Enhancement for Fake Face Detection in the Wild" [33]:
'Amélioration de la texture globale pour la détection des faux visages dans la nature'

Dans ce travail, les chercheurs ont montré que les modèles CNN sont fortement basés sur les textures plutôt que sur les formes et se situent principalement dans les régions de texture comme la peau et les cheveux et que les autres régions n'apportent que peu de contribution. Ils ont effectué alors des expériences sur les régions de la peau car elles contiennent de riches informations de texture et moins d'information structurelles.

Et à partir de cela, ils ont conclu que la texture des faux visages est différente de celle des vrais visages. Ils ont développé alors une nouvelle architecture Gram-Net qui améliore la robustesse et la capacité de généralisation des CNN dans la détection de faux visages. La précision de la détection de faux visages est estimée à 80.72% avec les bases de données styleGan [19] et CelebA-HQ [18] dans la phase du test. Ces résultats montrent une direction prometteuse et encourageante pour la compréhension des fausses images à partir des GAN.

Article 3 "DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern" [34] :
'Détection de Deepfakes à l'aide du modèl de clignement des yeux humains'

Ce travail vise à détecter les deepfakes générés par réseaux adversaires génératifs (GAN) en utilisant l'analyse du clignement des yeux. La méthode proposée est appelée Deepvision avec l'utilisation de l'apprentissage automatique.

L'architecture de DeepVision possède un pré-processeur, qui reçoit des informations en entrée. Grâce à ce processus, des données telles que le sexe, l'âge, l'activité et l'heure sont saisis comme des paramètres importants qui peuvent vérifier les changements dans les clignements d'yeux de l'homme. C'est ainsi que Deepvision effectue des mesures grâce au détecteur (Target détector) pour la détection du visage. Et ensuite l'algorithme Eye Tracker qui suit le clignement des yeux et mesure le nombre de répétition et la durée du clignement des yeux.

Ces données mesurées sont comparées à la base de données des mouvements naturels pour vérifier que le clignement des yeux est naturel ou non.

Deepvision détecte les vidéos deepfakes avec une précision de 87,5%.

Cependant, il n'est pas approprié si le sujet de vidéo souffre d'une maladie mentale, car on constate souvent un clignement anormal des yeux chez ces personnes.

Article 4 "Vulnerability assessment and detection of Deepfake vidéos" [35] 'Évaluation des vulnérabilités et détection des Deepfake vidéos'

Dans cette étude, les chercheurs ont présenté deux ensembles de vidéos Deepfake générés à partir de la base de données VidTIMIT. Ils ont généré des vidéos avec des qualités visuelles différentes (64*64) et (128*128).

Puis, ils ont évalué deux systèmes basés sur les réseaux de neurones VGG [36] et facenet 6 [37] qui n'ont pas pu parvenir à distinguer les Deepfakes vidéos des vraies avec un taux d'erreur égal à 95%.

En utilisant ensuite l'approche audiovisuelle pour détecter les incohérences entre les mouvements des lèvres et la parole en audio, ils ont réalisé que l'approche basée sur la synchronisation des lèvres ne parvient pas à détecter les décalages des lèvres et la parole.

Seules les approches basées sur les images sont capables de détecter efficacement les Deepfakes vidéos.

Ils ont utilisé des techniques basées sur les mesures de qualité d'image IQM avec un classificateur SVM pouvant détecter les Deepfakes avec un taux d'erreur de 8.97%.

Article 5 "Exposing Deepfake videos by detecting face Warping Artifacts" [38]

Cette recherche vise à décrire une nouvelle méthode basée sur l'apprentissage profond qui permet de distinguer d'une manière efficace les vidéos Deepfake des vraies.

Vu la limitation des ressources de calcul et de temps de production, l'algorithme Deepfake ne peut pas repérer les vraies vidéos des vidéos fausses. Les images subissent une déformation affine qui correspond à la configuration du visage de la source. C'est pourquoi, les auteurs de cet article ont proposé un modèle de réseau neuronal convolutif CNN pour détecter la présence d'artfacts à partir des régions du visage et des zones environnantes. L'apprentissage du CNN est basé sur des images collectées sur internet (24442 images de visage JPEG).

Pour élargir la diversité d'apprentissage, les auteurs ont modifié les informations de couleur c'est-à-dire luminosité, contraste et netteté pour tous les exemples de formation et d'entrainement, Ils ont validé leur travail sur le jeu de données vidéos Deepfake UADFV [39]. Ce jeu contient 98 vidéos (32752 frames) dont 49 vraies vidéos et 49 fausses vidéos.

Les tests effectués sur les vidéos Deepfake démontrent l'efficacité de la méthode proposée dans la pratique avec un taux de précision allant jusqu'à 97%.

Article 6 "Fake face detection via adaptive manipulation traces extraction network" [40] 'Détection de faux visage via un réseau d'extraction de traces de manipulation adaptative'

Les auteurs de cet article proposent un module de prétraitement AMTEN qui exploite la couche de convolution pour servir de prédicteur afin d'obtenir une manipulation d'image. Ce qui

signifie qu'il peut être transféré aux modèles basés sur CNN pour détecter d'autres falsifications d'images.

En appliquant certaines opérations de post-traitement (y compris la compression avec perte). Ils ont simulé la réelle analyse des images de visage dans les scènes complexes possible. Pour prouver l'efficacité du projet AMTENnet, une série d'expériences ont été menées pour obtenir un taux de détection plus élevé. En utilisant 116000 images de visages à partir de leur base HFF pour la formation qui comprend des images réelles de différentes résolutions et cinq types de fausses images, cette base est divisée en trois sous-ensembles de données pour la formation (75%), la validation (5%) et le test (20%). AMTENnet atteint donc une précision de 95.17%.

Article 7 "Recurrent Convolutional Structures for Audio Spoof and Video Deepfake Detection" [41] 'Structures convolutives récurrentes pour la détection de fausses pistes audio et vidéo'

Dans cet article, les chercheurs ont proposé un réseau de neurones récurrents convolutifs xceptemporal pour détecter les Deepfakes

Premièrement, ils ont utilisé un détecteur de visage appelé dlib pour déterminer le visage principal sur chaque image de la vidéo. Les visages sont encodés en utilisant l'architecture xceptionNet [42]. Les caractéristiques spatiales du réseau xceptionNet sont passées dans une première couche LSTM bidirectionnelle dont les sorties de la première couche LSTM sont transmises à une deuxième couche LSTM bidirectionnelle pour produire des caractéristique secondaires.

Le vecteur caractéristique provenant de la dernière unité LSTM de cette deuxième couche bidirectionnelle est transmis à une couche entièrement connecté puis à une couche de classification. Un dropout est ajouté à la couche entièrement connecté pour la régularisation.

Le modèle est entrainé avec les fonctions de perte traditionnelle cross-entropy et KL divergence.

Du côté de l'audio, les chercheurs ont introduit une architecture complémentaire en empilant plusieurs modules de convolution afin d'avoir des représentations de caractéristiques audio. Ces incorporations audio sont aussi passées dans une couche récurente bidirectionnelle.

L'entrainement du modèle sur les bases FaceForensics++ [20] et Celeb-DF [29] a donné entre 84.8% et 100% de précision pour le visuel. Pour la détection audio, les auteurs ont démontré la robustesse de leur méthode où ils ont obtenu 0.1424 t-DCF (fonction de coût de détection en tandem). Ce qui indique que ces méthodes se généralisent bien aux attaquent.

Article 8 "Detecting handcrafted facial image manipulations and GAN-generated facial images using Shallow-FakeFaceNet" [43] 'Détection des manipulations d'images faciales artisanales et des images faciales générées par GAN à l'aide du réseau Shallow-FakeFaceNet'

Cette recherche vise à détecter une nouvelle manipulation faciale fabriquée manuellement, en introduisant un nouveau jeu de données qui contient des images de visages photoshoppées, créées avec des outils comme Adobe Photoshop, afin de détecter à la fois les images manipulées par l'homme mais aussi générées par des machines. Ce jeu de données contient 1527 images

développées avec plusieurs niveaux de complexité d'édition et 621 images originales utilisées pour créer les faux visages. Ce travail consiste à développer un modèle avec une architecture de réseau convolutionnel CNN appelé Shallow-FakeFaceNet (SFFN). Ce modèle montre des résultats prometteurs dans la détection de fausses images créées par l'homme, en obtenant un taux de 72,52% au ROC et en utilisant moins de 2500 images fausses pour l'entraînement. Les auteurs de cet article ont évalué également leur approche sur d'autres faux visages générés par le GAN où le taux de reconnaissance obtenu était de 93,99 % pour les images à basse résolution.

8. Conclusion

La détection de deepfakes est une opération très essentielle face à la prolifération des contre façons. Grâce à l'intelligence artificielle il est désormais possible de modifier une photo ou une vidéo et cela au moyen des deepfakes crées à l'aide du deep learning, Cette technologie peut être exploitée à des fins malintionneés. Pour cette raison, les ordinateurs ont besoin d'un système efficace pour détecter la falsification des photos et des vidéos afin de réduire les fausses informations.

Chapitre 3
Conception

1. Introduction

Grace aux techniques de l'intelligence artificielle (IA), les images numériques deviennent de plus en plus réalistes, rendant plus difficile de distinguer les vraies des fausses à l'œil nu.

Ces fausses images et surtout les faux visages inquiètent les experts qui préviennent que ces faux visages peuvent être utilisés par des acteurs malveillants pour diffuser des informations erronées.

Notre projet se focalise autour de ce dernier point où nous utilisons les réseaux de neurones à convolutions pour détecter les fausses images plus particulièrement les faux visages.

2. Objectif

L'objectif principal de notre travail est de concevoir un système intelligent capable de discriminer les faux visages des vrais.

Pour cela, nous proposons un discriminateur de base, qui est le discriminateur du réseau adversarial génératif convolutif profond (DCGAN). Notre principale contribution vise la phase d'apprentissage ou nous proposons trois méthodes différentes d'apprentissage.

Un deuxième point clé de notre travail est les données utilisées pour l'apprentissage que nous détaillerons dans le chapitre 4.

3. Architecture du système

Comme nous l'avons déjà mentienner notre système est basé sur le modèle DCGAN. Ce choix est justifié par les performances connues de ce dernier.

Comme le montre la figure 3.1, l'entrée de notre système est une image (contenant un visage) après une étape de prétraitement (un redimensionnement) l'un des trois modèle est sélectionné pour permettre la prédiction du visage (vrai ou faux)

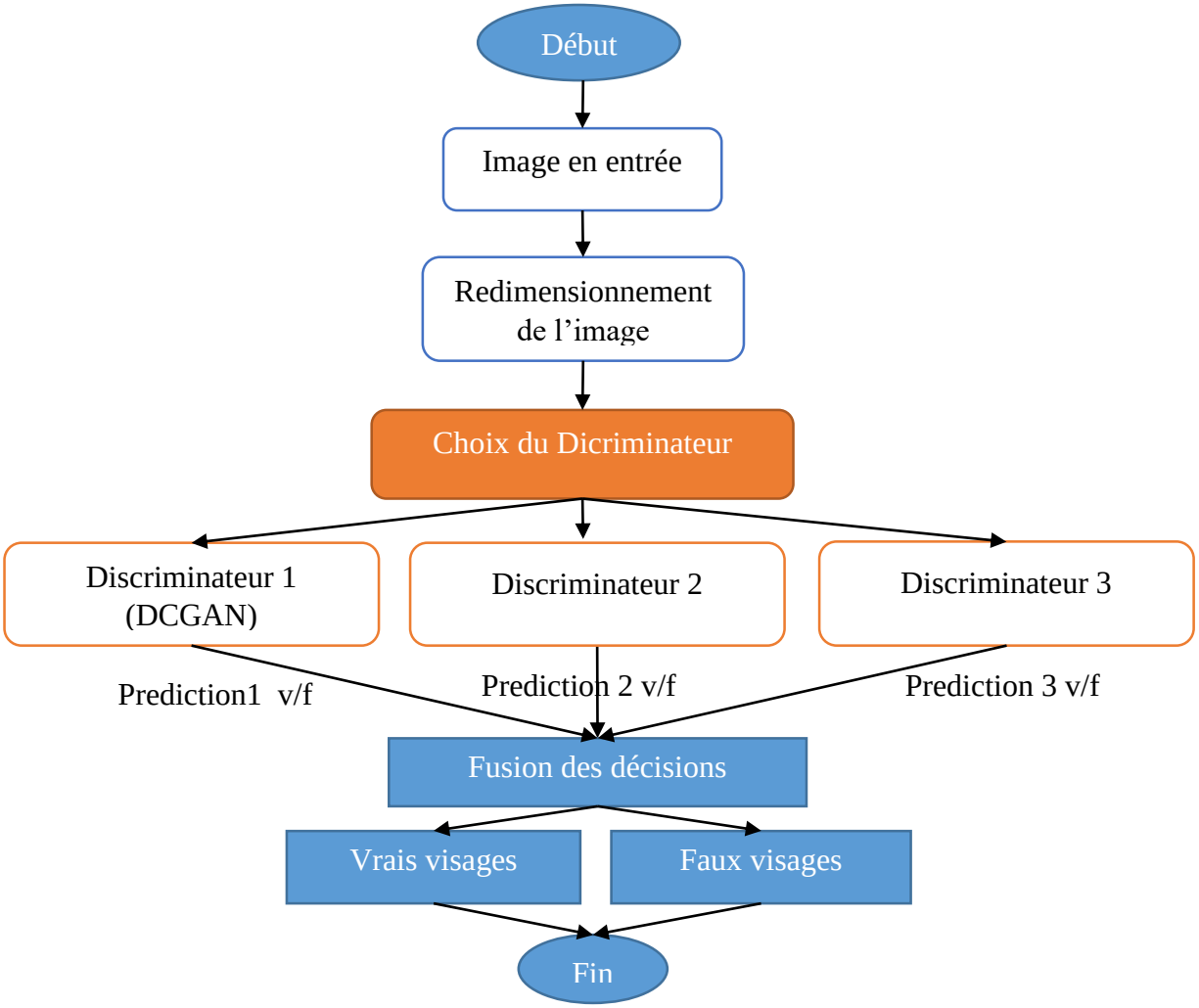

Figure 3.1 Architecture générale du système

3.1.Redimensionnement des images

Le générateur génère des images de taille 128x128 pixel tandis que le discriminateur en évalue l'authenticité. Alors l'entrée du réseau discriminateur est fixée au début de l'apprentissage. Il est donc essentiel de redimensionner les images pour que le discriminateur fasse la prédiction.

3.2.Réseau discriminateur

Nous avons mis au point trois réseaux discriminateurs dans le but de comparer leurs performances et déduire le meilleurs et/ou fusionner les décisions des trois.

3.2.1. Discriminateur 1 (DCGAN)

Dans la première méthode (présentée par la figure 3.2), nous utilisons un apprentissage par adversaire, c'est-à-dire nous formons le DCGAN en entrée (générateur et discriminateur).

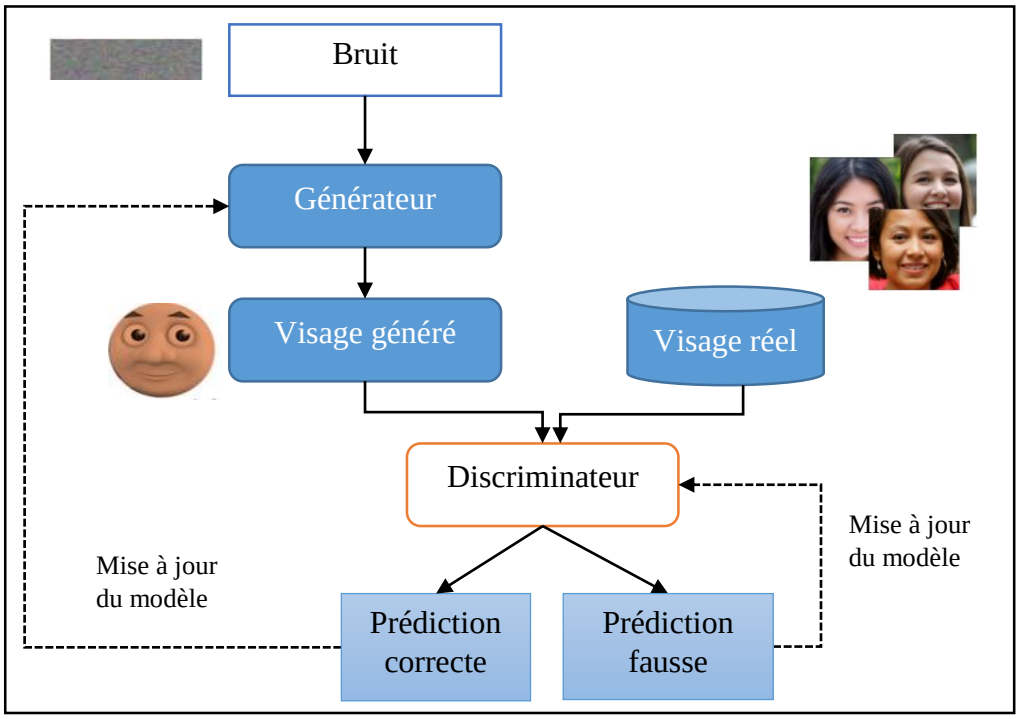

Figure 3.2 Apprentissage du discriminateur 1 (DCGAN)

3.2.2. Discriminateur 2

Dans la deuxième méthode, nous utilisons le discriminateur du DCGAN uniquement et nous faisons un apprentissage classique de CNN avec une base de données des vrais et faux visages.

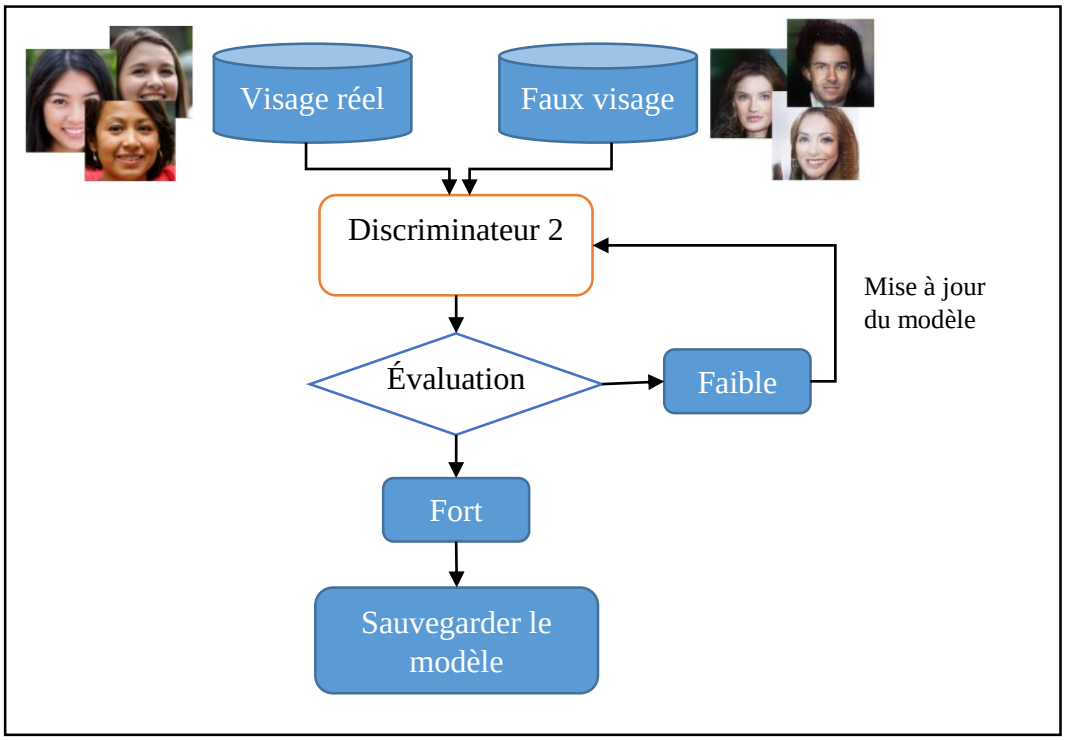

Figure 3.3 Apprentissage du discriminateur 2

3.2.3. Discriminateur 3

Dans le troisième modèle, nous avons utilisé le discriminateur pré entrainé du GAN de la première méthode et nous l'avons renforcé par l'apprentissage avec une base de données de faux visages différent de la première.

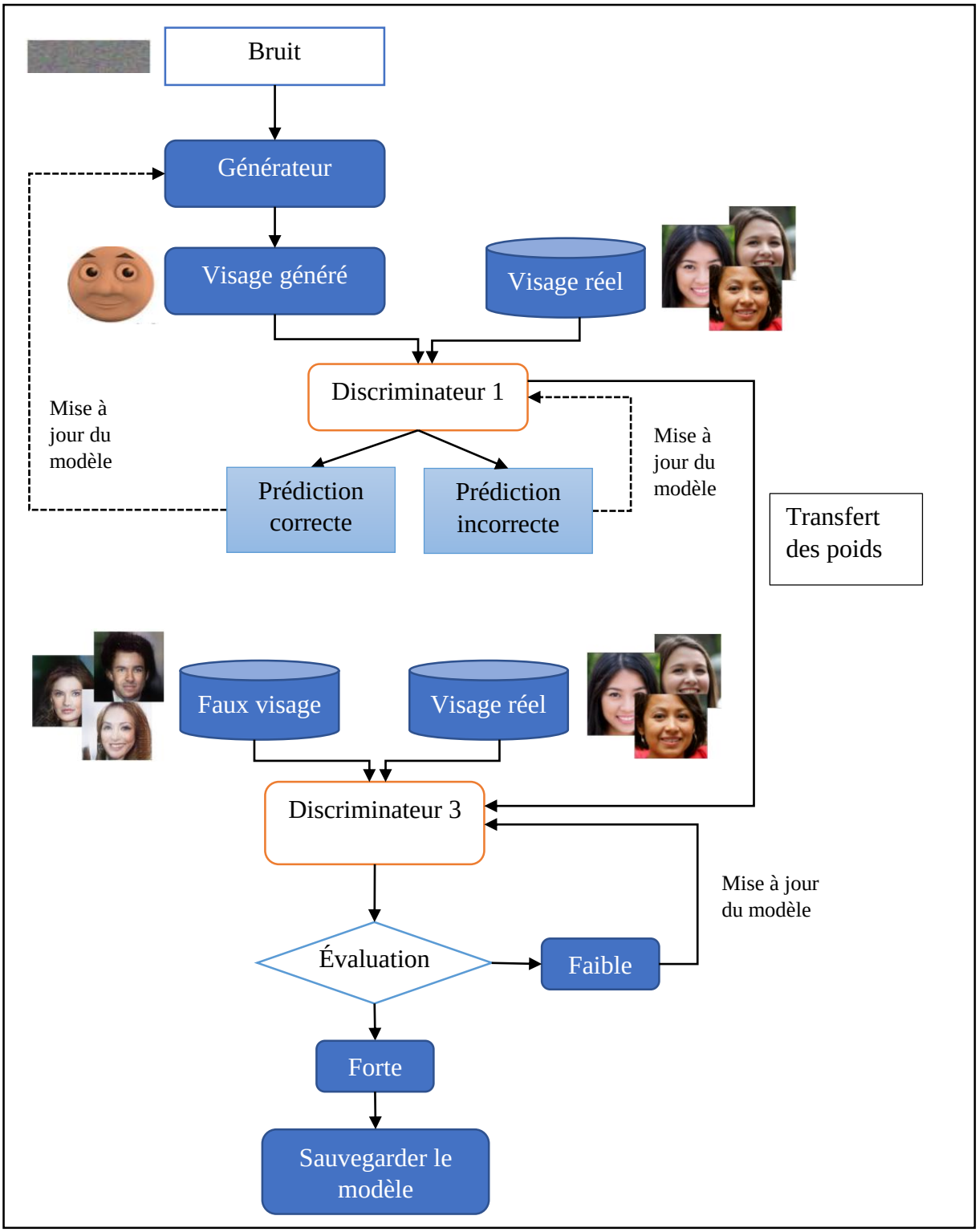

Figure 3.4 Apprentissage du discriminateur 3

3.3.Fusion des décisions

Pour un résultat optimal, nous proposons la fusion des décisions des trois réseaux qui peut être la moyenne des prédictions ou la probabilité maximale.

4. Architecture détaillée du réseau utilisé

Le modèle de base que nous avons utilisé dans ce travail est le DCGAN qui est une extension directe du GAN, sauf qu'il utilise explicitement des couches convolutionnelles et convolutionnelles-transposées dans le discriminateur et le générateur.

Les couches convolutionnelles préservent la structure spatiale d'une image, ce qui signifie que les caractéristiques les plus précises et les plus détaillées seront extraites d'une image. Cela donne au générateur et au discriminateur des capacités de raisonnement spatial plus avancées sur la sortie qu'ils vont générer et sur la façon de discriminer entre les caractéristiques des images réelles et celles des images truquées. L'amélioration de la qualité des caractéristiques extraites est généralement la raison pour laquelle les DCGAN sont utilisés pour traiter des images [44]. Comme tous les GAN, le DCGAN est composé d'un générateur et un discriminateur.

4.1.Architecture de réseau générateur

Le générateur est composé de couches de convolution transposées qui permettent de sur-échantillonner le vecteur de bruit pour le transformer en une image. Dans un réseau CNN classique, les couches convolutionnelles chercheront à extraire des caractéristiques de plus en plus petites qui seront ensuite classées. Tandis que dans un générateur, les couches de convolutions transposées chercheront à inverser les opérations de la couche convolutive.

Dans le sens le plus simple, une transposition fait en sorte qu'au moins deux choses changent de place l'une par rapport à l'autre. Le vecteur de bruit et l'espace de l'image vont changer de place l'un par rapport à l'autre. Cela veut dire que nous changeons l'ordre de leurs dimensions et donc nous renverserons les valeurs des matrices par rapport à la diagonale. Ce processus de sur-échantillonnage agrandit et remplit les détails du résultat final (l'image). C'est la partie du générateur qui "dessine" l'image réelle.

L'entrée du réseau est un vecteur de bruit. Le bruit est une version compressée de ce que l'image deviendra. Les couches de convolution transposées du générateur vont décompresser le bruit pour qu'il devienne une image 128x128 avec tous les détails au bon endroit dans l'image [w12].

Les détails de chaque couche sont représentés par la figure 3.5 et le tableau 3.1. Notons l'activation de LeakyReLU pour chaque couche, à l'exception de la couche de sortie qui utilise tanh.

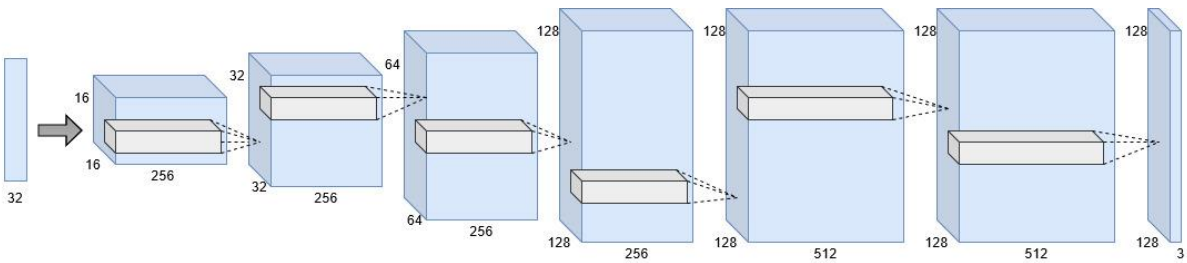

Figure 3.5 Schéma de l'architecture du réseau générateur

Couche (type)	dimension	Taille noyau	Stride	Activation
Entrée	32	-	-	-
dense	-	-	-	LeakyReLU
Conv2D	(16, 16, 256)	5	1	LeakyReLU
Conv2DTranspose	(32, 32, 256)	4	2	LeakyReLU
Conv2DTranspose	(64, 64, 256)	4	2	LeakyReLU
Conv2DTranspose	(128, 128, 256)	4	2	LeakyReLU
Conv2D	(128, 128, 512)	5	1	LeakyReLU
Conv2D	(128, 128, 512)	5	1	LeakyReLU
Conv2D	(128, 128, 3)	7	1	tanh

Tableau 3.1 Tableau détaillé sur le réseau générateur

4.2.Architecture de réseau discriminateur

Le discriminateur est composé de couches strided convolutions et d'activations LeakyReLU. L'entrée est une image 128x128x3 et la sortie est une probabilité entre 0 et 1.

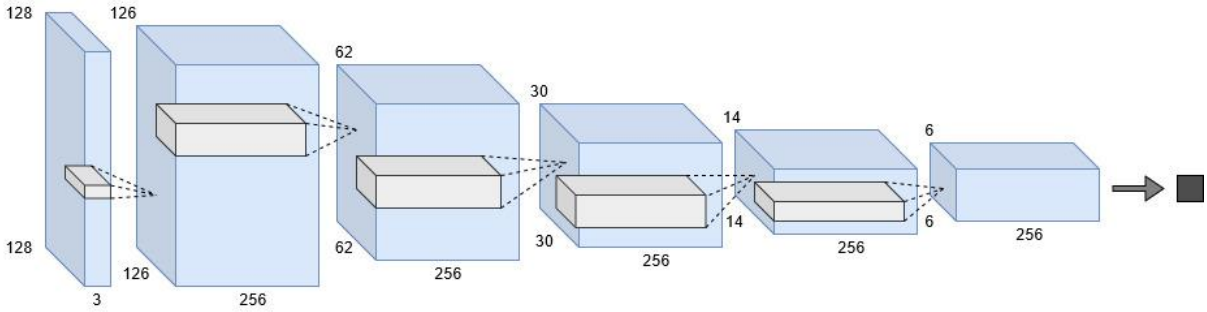

Figure 3.6 Schéma de l'architecture du discriminateur

Couche (type)	dimension	Kernel size	Stride	Activation
Input	(128, 128, 3)	-	-	-
Conv2D	(126, 126, 256)	3	1	LeakyReLU
Conv2D	(62, 62, 256)	4	2	LeakyReLU
Conv2D	(30, 30, 256)	4	2	LeakyReLU
Conv2D	(14, 14, 256)	4	2	LeakyReLU
Conv2D	(6, 6, 256)	4	2	LeakyReLU
Dense	1	-	-	sigmoid

Tableau 3.2 Tableau détaillé sur le réseau discriminateur

4.3.Conditions de stabilité de DCGAN

Pour avoir un DCGAN stable il faut :

- Le pooling est remplacé par un stride convolution. Cela permet au réseau d'apprendre son propre sous-échantillonnage spatial (en changeant la taille de l'entrée).
  Les CNN utilisent des couches de pooling pour réduire les dimensions. Par exemple, une couche de max pooling 2x2 prendrait un tableau 2x2 de pixels et le mettrait en correspondance avec un nombre, qui est le maximum parmi eux. stride convolution peut diminuer la dimension en sautant plusieurs pixels entre les convolutions au lieu de faire

glisser le noyau un par un. De même, elle peut augmenter la dimension en ajoutant des pixels vides entre les pixels réels.

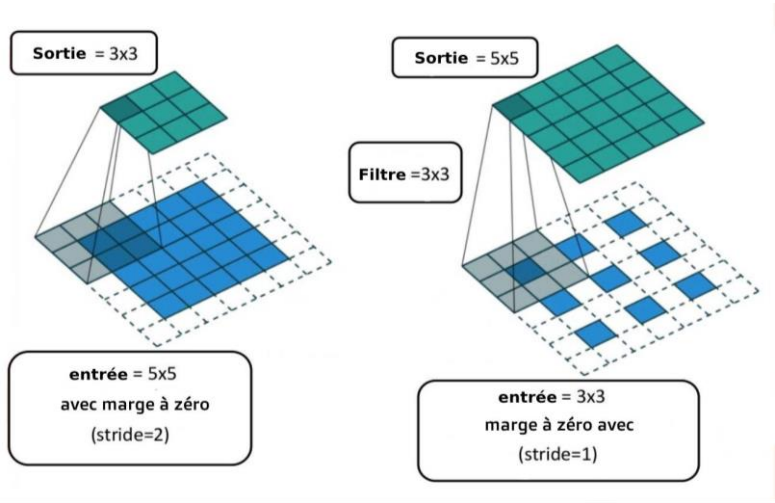

Figure 3.7 Exemple de Stride convolution

- Pas de couche entièrement connectée à la fin du CNN. Le générateur n'est pas un classificateur donc cette partie n'est pas nécessaire.
- Utiliser leakyReLU dans le générateur sauf pour la sortie qui utilise tanh. La symétrie de la fonction tanh permet au modèle d'apprendre plus rapidement, et leakyReLU pour le discriminateur.

La sortie de la fonction d'activation Leaky ReLU sera positive si l'entrée est positive, et elle sera une valeur négative contrôlée si l'entrée est négative. La valeur négative est contrôlée par un paramètre appelé alpha, qui introduit une tolérance du réseau en permettant à certaines valeurs négatives de passer [w13].

$$f(y) = \max(\alpha \times y, y)$$

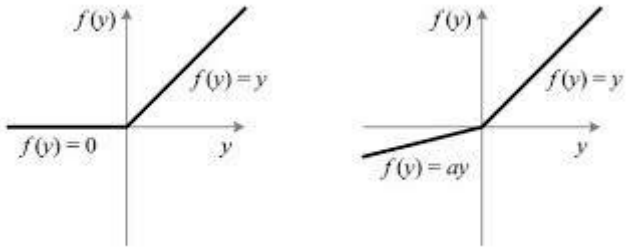

Figure 3.8 Fonction d'activation ReLU et Leaky ReLU

5. Conclusion

Nous avons utilisé le réseau du discriminateur de DCGAN afin de détecter les faux visages mais la fiabilité de ce discriminateur dépend en grande partie du jeu de donnés utilisé et nous avons fait plusieurs expériences pour déterminer un classificateur acceptable qui est nécessaire pour l'amélioration des résultats finaux.

Chapitre 4
Implémentation

1. Introduction

L'architecture proposée de notre système qui a fait l'objet d'une description détaillée dans le chapitre précédent sera validée par une série de tests sur chaque partie du système représenté principalement par les trois discriminateurs mis au point.

2. Environnement

Google colab et kaggle sont des services en nuage très populaires pour l'apprentissage automatique qui offre un accès gratuit aux GPU et TPU. Nous avons utilisé kaggle car il dispose de nombreux ensembles de donneés que nous pouvons importer, il fournit jusqu'à quarante heures de GPU/TPU par semaine et gratuitement ce qui n'est pas le cas pour google colab.

2.1.Kaggle

Kaggle, est une communauté en ligne de praticiens de l'apprentissage automatique. Il permet aux utilisateurs de trouver et de publier des ensembles de données, d'explorer et de construire des modèles dans un environnement de science des données basé sur le Web, de travailler avec d'autres ingénieurs d'apprentissage automatique, et de participer à des concours pour résoudre des défis de sciences des données.

Kaggle a démarré en 2010 en proposant des concours d'apprentissage automatique et offre désormais également une plateforme de données publique, un banc de travail en nuage pour la science des données et une formation à l'intelligence artificielle [w8].

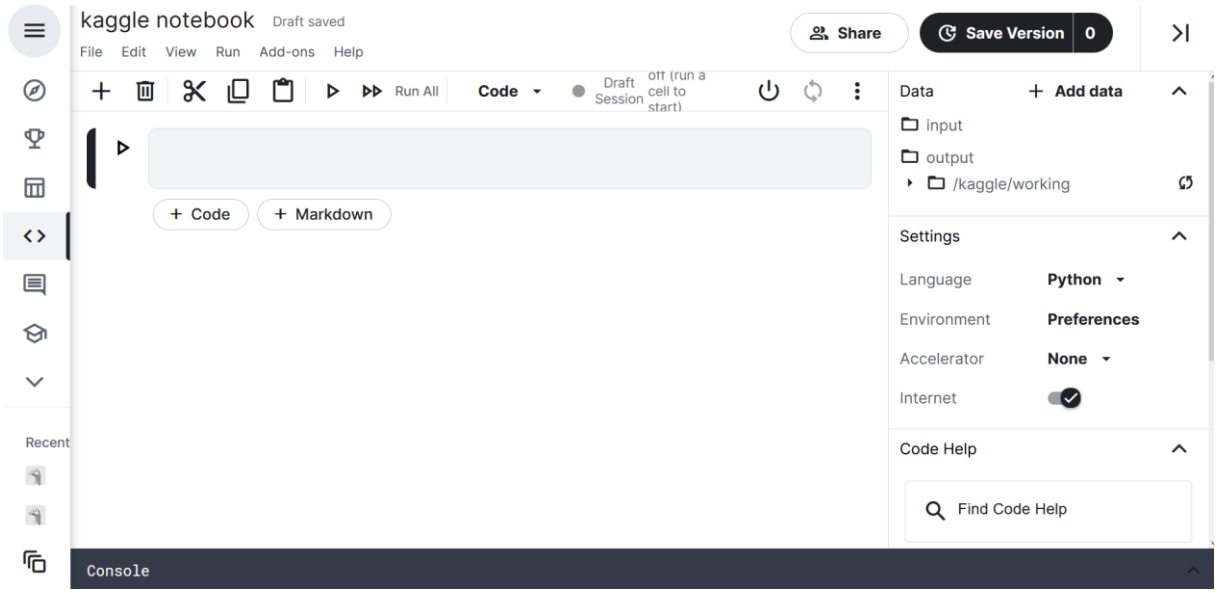

Figure 4.1 Interface de kaggle

2.2.Entrainer sur GPU

Kaggle fournit un accès gratuit aux GPU NVIDIA K80 et TESLA P100. Ces GPU sont utiles pour la formation des modèles d'apprentissage profond, l'intégration d'un GPU à votre noyau permet d'accélérer de 13 fois plus rapide qu'un entrainement classique sur un CPU.

Pour passer en mode GPU, dans la barre des options mettre l'option accélérateur en mode GPU.

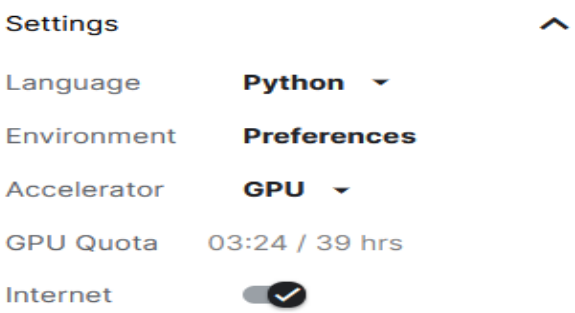

Figure 4.2 Paramètre du notebook

3. Langage de programmation et bibliothèque utilisée

3.1.Python

Python est un langage de programmation puissant et facile à apprendre. Il possède des structures de données de haut niveau et une approche simple mais efficace de la programmation orientée objet. La syntaxe élégante et le typage dynamique de Python, ainsi que sa nature interprétée, en font un langage idéal pour le développement des scripts et d'applications rapides dans de nombreux domaines, et sur la plupart des plateformes.

L'interpréteur Python et la bibliothèque standard étendue sont disponibles gratuitement sous forme source ou binaire pour toutes les principales plates-formes sur le site Web de Python, et peuvent être distribués librement. Le même site contient également des distributions et des pointeurs vers de nombreux modules, programmes et outils gratuits, ainsi que de la documentation supplémentaire [45].

3.2.Bibliothèque utilisée

3.2.1. Tensorflow

Nous avons utilisé Cette bibliothèque open source créé par google afin de définir les composants de base de l'architecture CNN. Cette bibliothèque dispose d'un écosystème complet et flexible d'outils, de bibliothèques et de ressources communautaires qui permettent aux chercheurs de faire progresser l'état de l'art de l'apprentissage automatique et aux développeurs de définir, former et déployer facilement des applications qui exploitent cette technologie [46].

3.2.2. Keras

Nous avons utilisé cette bibliothèque afin de créer les couches pour les réseaux neuronaux tout en s'occupant des détails minutieux des tenseurs, de leurs formes et de leurs détails mathématiques. Cette bibliothèque Python de haut niveau, compacte et facile à apprendre, pour l'apprentissage profond, qui peut s'exécuter au-dessus de TensorFlow. Elle permet aux développeurs de se concentrer sur les principaux concepts de l'apprentissage profond. TensorFlow doit être le back-end pour Keras. on peut utiliser Keras pour des applications d'apprentissage profond sans interagir avec le relativement complexe TensorFlow [47].

3.2.3. Numpy

Nous avons utilisé cette bibliothèque pour la programmation des tableaux pour python .Elle fournit une syntaxe puissante, compacte et expressive pour accéder, manipuler et opérer sur des données dans des vecteurs, des matrices et des tableaux de dimensions supérieures [48].

3.2.4. OS

Nous avons utilisé le module OS qui fournit des fonctions qui permettent d'interagir avec le système d'exploitation. Ce module offre un moyen utilisant les fonctionnalités dépendantes du système d'exploitation. Les modules *os* et *os.path* ont plusieurs fonctions permettant d'interagir avec les fichiers du système [w10].

3.2.5. Tkinter

Nous avons utilisé le module Tkinter qui fournit un moyen rapide et facile pour créer des applications graphiques. Tk et Tkinter sont tous deux disponibles sur la plupart des plateformes Unix, ainsi que sur les systèmes Windows et Macintosh. À partir de la version 8.0, Tk offre un aspect et une convivialité natifs sur toutes les plateformes [w11].

3.2.6. Matplotlib

Matplotlib est un paquet Python capable de produire des graphes de qualité. matplotlib est conçu pour pouvoir créer des graphiques simples et complexes avec quelques commandes.

matplotlib est écrit pour utiliser NumPy et d'autres d'extension. Avec quelques lignes de code, il génère des graphiques, des histogrammes, des diagrammes, des nuages de points, etc [49].

4. Base d'apprentissage

Dans la phase d'apprentissage , nous avons utilisé une base composée de 70 000 visages réels provenant du jeu de données Flickr [19] collecté par Nvidia, ainsi que de 70 000 faux visages générés en utilisant StyleGAN [19] de Nvidia.

Dans ce jeu de données, toutes les images ont la dimension 256px, et sont divisées en jeu de formation, de validation et de test.

4.1.Base de 'DCGAN'

Classe	Apprentissage	Test
Visages réels	50000	10000
Faux visages	-	10000

Tableau 4.1 Jeu de données du DCGAN

4.2.Base de Discriminateur 2 et 3

Classes	Apprentissage	Validation	Test
Visages réels	50000	10000	10000
Faux visages	50000	10000	10000
Total	100000	20000	20000

Tableau 4.2 Jeu de données utilisé pour le discriminateur 2 et 3

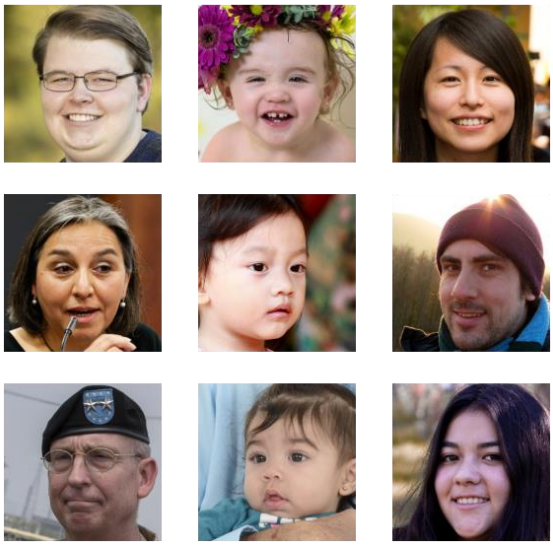

Figure 4.3 Exemples de vrais visages

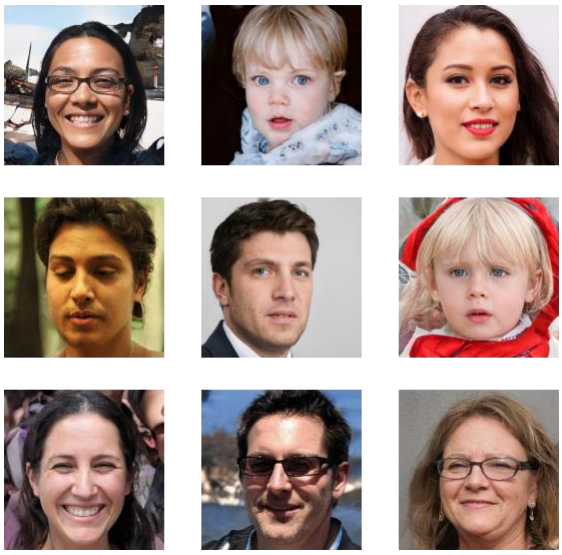

Figure 4.4 Exemples de faux visages

5. Apprentissage et test

Durant la phase d'apprentissage, nous avons essayé plusieurs configurations basées sur le changement de certains paramètres du réseau tel que:
L'optimiseur, nombre d'itération, nombre d'epochs, batch size.

5.1. Apprentissage du 'DCGAN'

Durant l'apprentissage de ce réseau, nous avons retenu la configuration suivante:

- Optimiseur = RMSProp qui génère des fausses images plus réalistes par rapport à Adam.
- Batch size =128.
- Nombre d'itération =10000.

La figure (4.5) représente les résultats d'apprentissage du générateur et du discriminateur.

Le discriminateur voudrait maximiser la probabilité logarithmique indiquant que les données sont réelles ou fausses. Le générateur, quant à lui, essaie de minimiser la probabilité logarithmique que le discriminateur ait raison.

Nous constatons que le réseau discriminateur a atteint un taux de perte de (47%) pendant l'apprentissage et le générateur a atteint un taux de perte de 96%.

Donc le taux de perte du discriminateur n'est pas bon pour discriminer les faux visages, et le générateur peut générer des visages plus au moins réels.

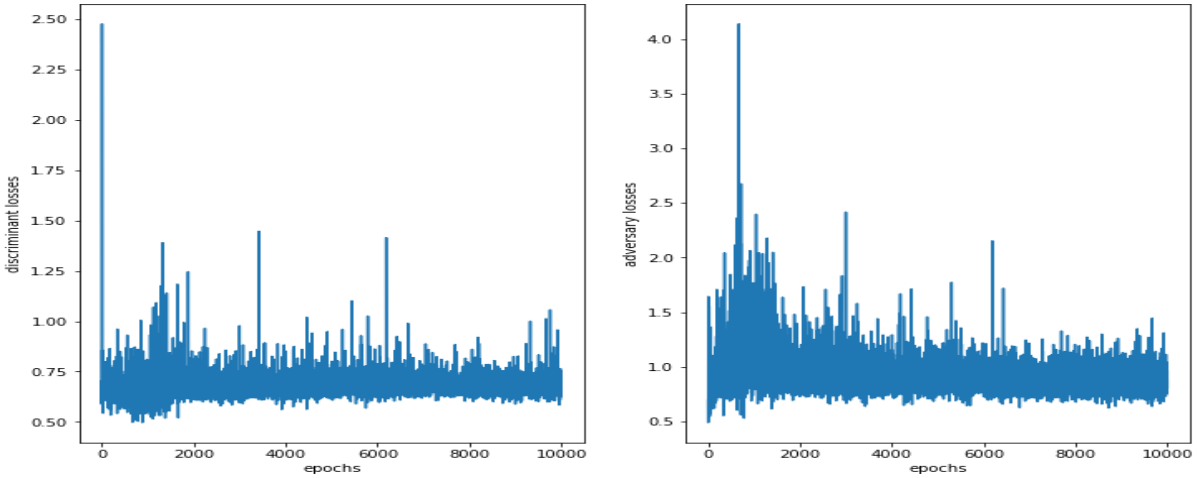

Figure 4.5 Les graphes de 'loss' pour apprentissage de DCGAN

5.2.Test du discriminateur du DCGAN

Nous avons fait un test sur 20000 images n'appartenant pas à la base de données utilisée dans l'entrainement et les résultats étaient les suivants:

Taux	pourcentage
VP	53%
VN	49%
FP	47%
FN	51%

Tableau 4.3 Tableau des taux de reconnaissance du Discriminateur 1 (DCGAN)

```
faux negative 5134 vrai negative 4866
vrai positive 5280 faux positive 4720
```

Figure 4.6 Résultat du test du discriminateur 1 (DCGAN)

	precision	recall	support
fake	0.51	0.49	10000
real	0.51	0.53	10000
accuracy		0.51	20000

Figure 4.7 Mesure de performances du Discriminateur 1 (DCGAN)

5.3.Apprentissage du discriminateur 2

Durant l'apprentissage de ce réseau, nous avons retenu la configuration suivante:

- Optimiseur = RMSProp.
- Batch size =128.
- Nombre d'epochs =45.

La figure (4.8) représente les résultats d'apprentissage du générateur et du discriminateur.

Nous constatons que le réseau a atteint un taux de précision de 97% pendant l'apprentissage et 94% pendant le test avec une perte de 0.06% pendant l'apprentissage et 0.14% pour le test estimé comme acceptable pour détecter les faux visages.

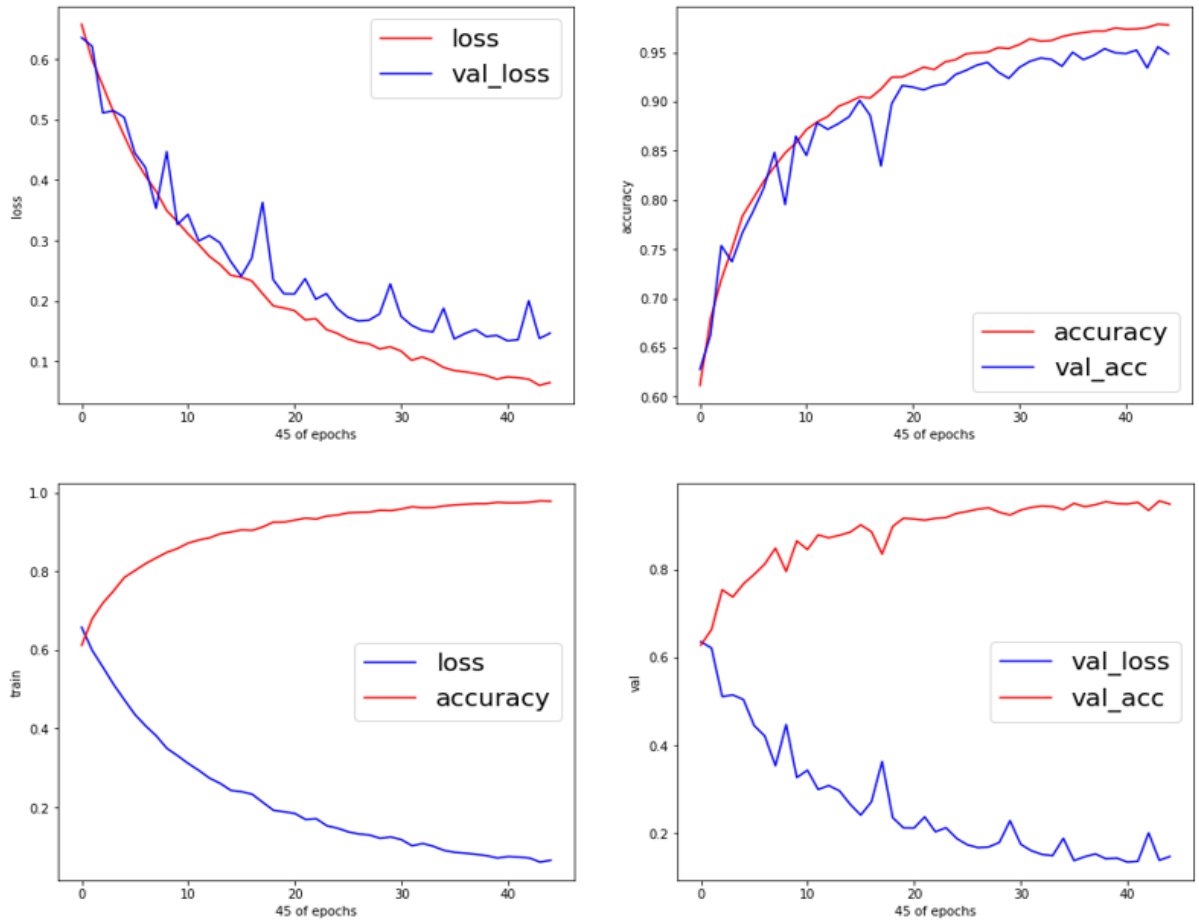

Figure 4.8 Les graphes de 'accracy' et 'loss' pour l'apprentissage du discriminateur 2

5.4.Test du discriminateur 2

Nous avons fait un test sur 20000 images n'appartenant pas à la base de données utilisée dans l'entrainement et les résultats étaient les suivants:

Taux	pourcentage
VP	98%
VN	93%
FP	2%
FN	7%

Tableau 4.4 Tableau des taux de reconnaissance du discriminateur 2

```
faux negative 723 vrai negative 9277
vrai positive 9763 faux positive 237
```

Figure 4.9 Résultat de test du discriminateur 2

	precision	recall		support
fake	0.98	0.93		10000
real	0.93	0.98		10000
accuracy			0.95	20000

Figure 4.10 Mesure de performances du discriminateur 2

5.5.Apprentissage du discriminateur 3

Durant l'apprentissage de ce réseau, nous avons retenu la configuration suivante:

- Optimiseur = RMSProp.
- Batch size =128.
- Nombre d'epochs =30.

La figure (4.11) représente les résultats d'apprentissage du générateur et du discriminateur.

Nous constatons que le réseau a atteint un taux de précision de 98% pendant l'apprentissage et 94% pendant le test avec une perte de 0.05% pendant l'apprentissage et 0.12% pour le test estimé comme acceptable pour détecter les faux visages.

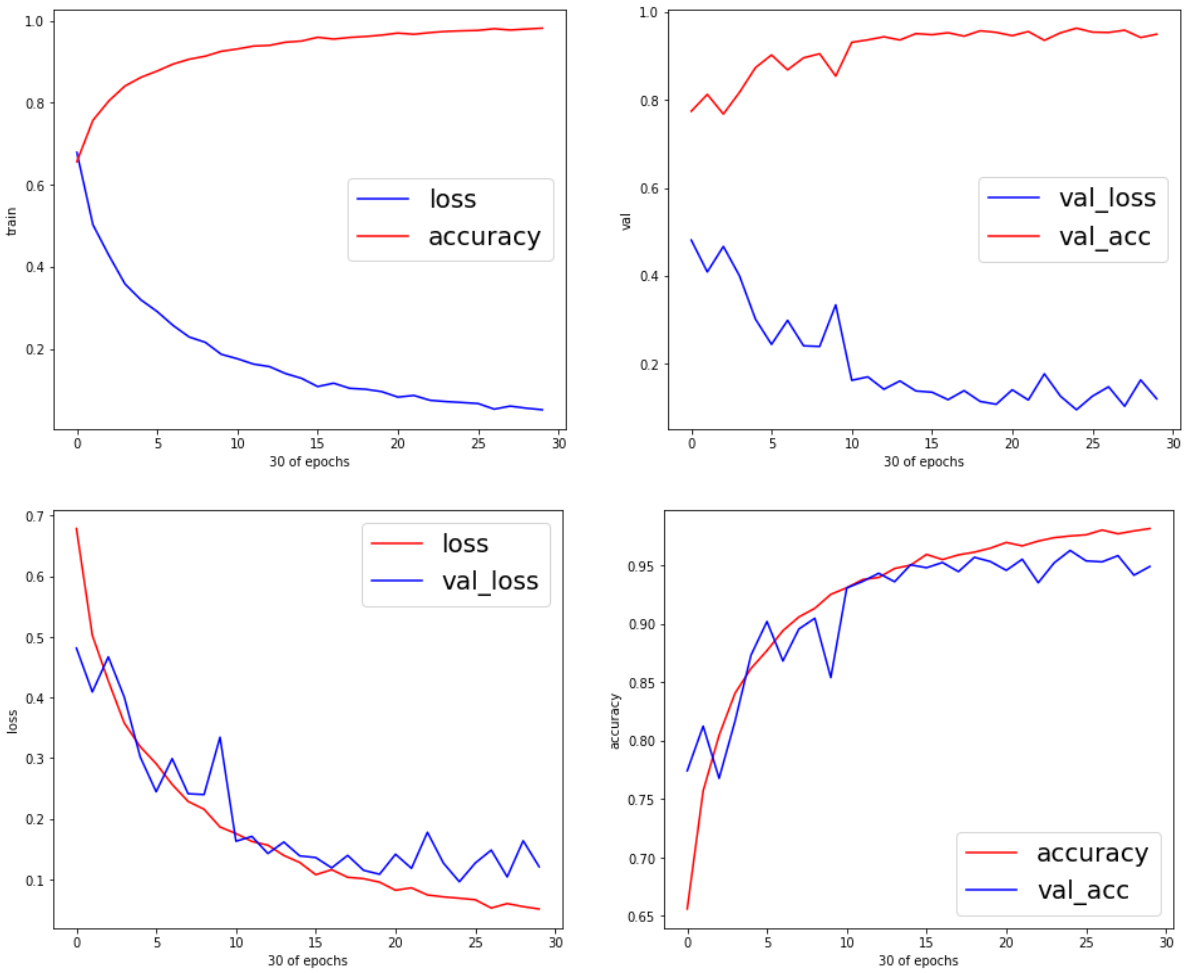

Figure 4.11 Les graphes de 'accracy' et 'loss' pour apprentissage du discriminateur 3

5.6.Test du modèle 3

Taux	pourcentage
VP	92%
VN	98%
FP	8%
FN	2%

Tableau 4.5 Tableau des taux de reconnaissance du discriminateur 3

```
faux negative 177 vrai negative 9823
vrai positive 9205 faux positive 795
```

Figure 4.12 Résultat de test du discriminateur 3

	precision	recall		support
fake	0.93	0.98		10000
real	0.98	0.92		10000
accuracy			0.95	20000

Figure 4.13 Rapport de classification du discriminateur 3

6. Comparaison

Afin de comparer les performances des trois réseaux, nous avons présenté les tests de ces réseaux via leurs courbes ROC dans la figure 4.14 où nous pouvons constater que le discriminateur 3 donne les meilleurs résultats qui sont légèrement supérieurs à ceux du discriminateur 2, alors que le discriminateur 1 ne distingue presque pas les visages vrais des faux visages

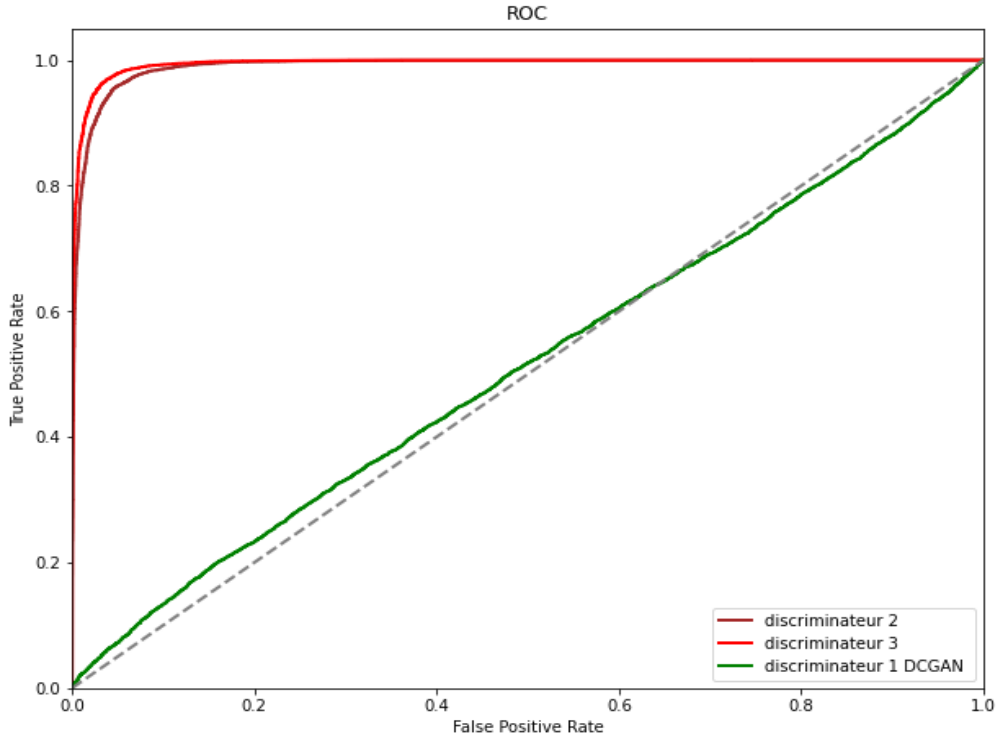

Figure 4.14 la Courbe ROC du trois modèle

7. Test sur quelques images aléatoires

Nous avons élaboré quelques tests sur des images de vrais visages et de faux visages introduites d'une manière aléatoire au système, les résultats sont illustrés par le tableau 4.6 avec indication pour chaque image de la prédiction et de la classe laquelle appartient (vrais ou faux visages).

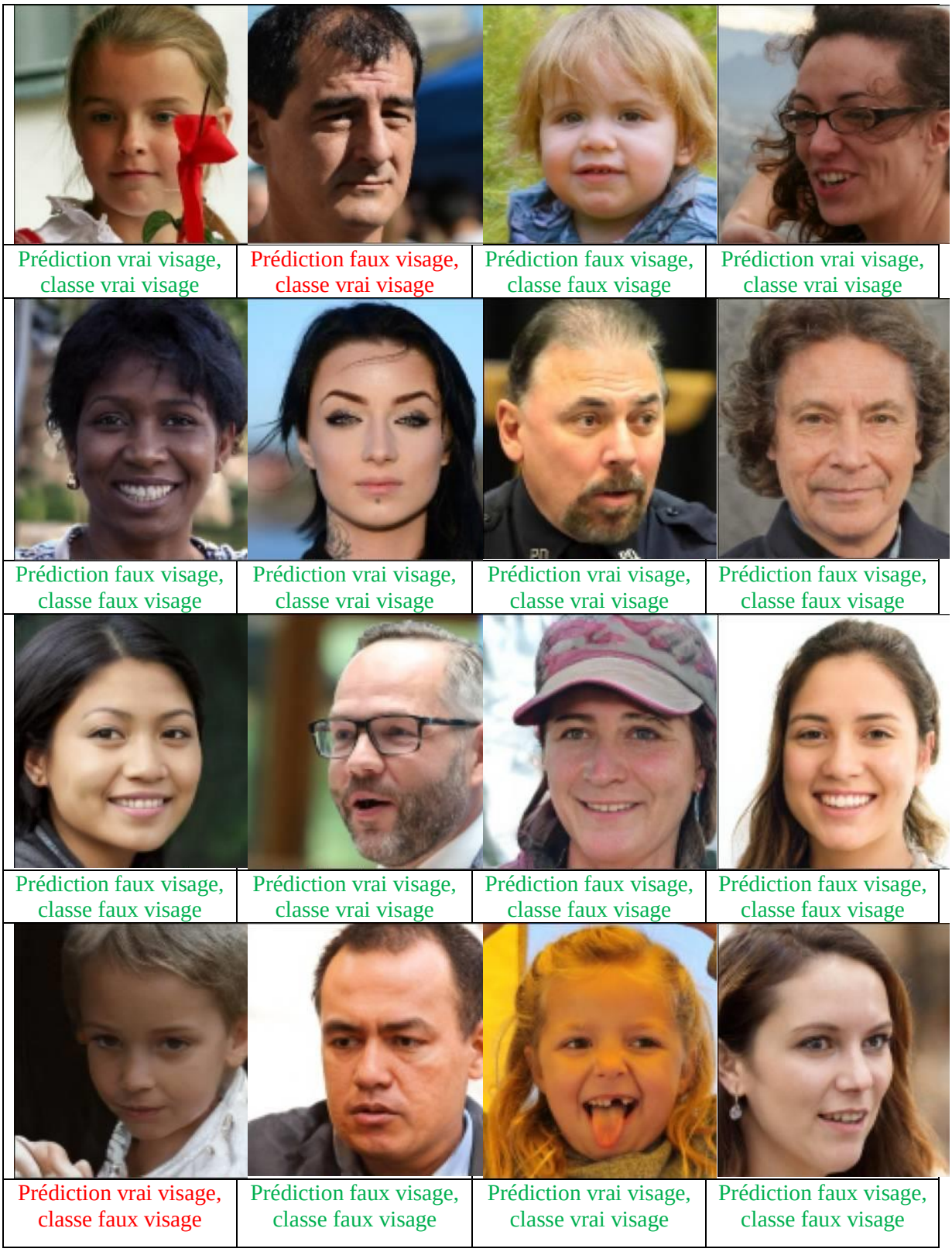

Tableau 4.6 Test sur quelques images aléatoires

8. Conclusion

Les tests effectués sur la détection des faux visages sont encourageant et les performances de notre système peuvent être améliorées en utilisant des images de haute qualité et des

machines qui possèdent des capacités remarquables pour reconnaitre si les vrais visages et les faux visages seront plus faciles à distinguer.

Conclusion générale

Les mécanismes de détection des faux visages sont devenus très indispensables face à la propagation rapide d'images falsifiées grâce à l'intelligence artificielle. Il est désormais possible de modifier une image, et donc de créer des images fausses ayant l'apparence des vraies que même les êtres humains n'arrivent pas à les distinguer.

Ces dernières années, ce domaine a suscité une importance primordiale pour les chercheurs afin d'augmenter les techniques d'apprentissage automatique et d'intelligence artificielle.

L'objectif principal de notre application est la détection des faux visages (Fake Faces) en utilisant les réseaux adversaire génératifs (GAN).

La capacité de l'algorithme des GAN s'améliorera au fil du temps ce qui rend la détection des faux visages difficile. Tout détecteur aura une courte durée de vie sur le fait que la détection des faux visages peut résoudre le problème à court terme, mais à long terme, la solution pratique sera les techniques d'authentification.

Face à la robustesse des générateurs des faux visages par rapport aux discriminateurs.

Nous proposons donc dans notre application une architecture basée sur l'entrainement de trois discriminateurs du DCGAN, où l'élément clé est la méthode d'entrainement utilisée pour chaque cas.

L'entrainement du système a été effectué sur 140000 échantillons confondus entre les vrais (70000) et les faux visages (70000) où la précision de l'apprentissage obtenue était de 98% et celle du test de 94%

Ces résultats considérés comme très prometteurs vu la robustesse des générateurs qui ont fait un grand saut par rapport aux discriminateurs et détecteurs des fausses images.

L'efficacité de ce système de détection pourrait être perfectionnée. Pour cet objectif nous proposons le suivant :

- Utiliser d'autres architectures de discriminateur et de générateur.
- Laisser les réseaux s'entrainer plus longtemps.
- Utiliser un jeu de données d'entrainement plus étendu.
- Booster l'entrainement du discriminateur sur les faux visages.

Bibliographies

[1] Gurney, K. (1997). Introduction to Neural Networks. CRC press.

[2] Gershenson, C. (2003). Artificial Neural Networks for Beginners. arXiv:cs/0308031. http://arxiv.org/abs/cs/0308031

[3] Touzet, C. (1992). Les reseaux de neurones artificiels, introduction au connexionnisme. 130.

[4] Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., … Asari, V. K. (2019). A State-of-the-Art Survey on Deep Learning Theory and Architectures. Electronics, 8(3), 292. https://doi.org/10.3390/electronics8030292

[5] Kim, K. G. (2016). Book Review : Deep Learning. Healthcare Informatics Research, 22(4), 351. https://doi.org/10.4258/hir.2016.22.4.351

[6] Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review, 53(8), 5455-5516. https://doi.org/10.1007/s10462-020-09825-6

[7] Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. 2017 International Conference on Engineering and Technology (ICET), 1-6. https://doi.org/10.1109/ICEngTechnol.2017.8308186

[8] Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks : An overview and application in radiology. Insights into Imaging, 9(4), 611-629. https://doi.org/10.1007/s13244-018-0639-9

[9] Westerlund, M. (2019). The Emergence of Deepfake Technology : A Review. Technology Innovation Management Review, 9(11), 39-52. https://doi.org/10.22215/timreview/1282

[10] Manaswi, N. K. (2020). Generative Adversarial Networks with Industrial Use Cases : Learning How to Build GAN Applications for Retail, Healthcare, Telecom, Media, Education, and HRTech. BPB Publications.

[11] Brownlee, J. (2019). Generative Adversarial Networks with Python : Deep Learning Generative Models for Image Synthesis and Image Translation. Machine Learning Mastery.

[12] Ting, K. M. (2010). Confusion Matrix. In C. Sammut & G. I. Webb (Éds.), Encyclopedia of Machine Learning (p. 209-209). Springer US. https://doi.org/10.1007/978-0-387-30164-8_157

[13] Santha, A. (2020). Deepfakes Generation using LSTM based Generative Adversarial Networks. 64.

[14] Bregler, C., Covell, M., & Slaney, M. (1997). Video Rewrite : Driving visual speech with audio. Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '97, 353-360. https://doi.org/10.1145/258734.258880

[15] Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., & Niessner, M. (2016). Face2Face : Real-Time Face Capture and Reenactment of RGB Videos. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2387-2395. https://doi.org/10.1109/CVPR.2016.262

[16] Suwajanakorn, S., Seitz, S. M., & Kemelmacher-Shlizerman, I. (2017). Synthesizing Obama : Learning lip sync from audio. ACM Transactions on Graphics, 36(4), 1-13. https://doi.org/10.1145/3072959.3073640

[17] Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep Learning Face Attributes in the Wild. ArXiv:1411.7766 [Cs]. http://arxiv.org/abs/1411.7766

[18] Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2018). Progressive Growing of GANs for Improved Quality, Stability, and Variation. ArXiv:1710.10196 [Cs, Stat]. http://arxiv.org/abs/1710.10196

[19] Karras, T., Laine, S., & Aila, T. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks. ArXiv:1812.04948 [Cs, Stat]. http://arxiv.org/abs/1812.04948

[20] Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++ : Learning to Detect Manipulated Facial Images. ArXiv:1901.08971 [Cs]. http://arxiv.org/abs/1901.08971

[21] Novozamsky, A., Mahdian, B., & Saic, S. (2020). IMD2020 : A Large-Scale Annotated Dataset Tailored for Detecting Manipulated Images. 2020 IEEE Winter Applications of Computer Vision Workshops (WACVW), 71-80. https://doi.org/10.1109/WACVW50321.2020.9096940

[22] Mahfoudi, G., Tajini, B., Retraint, F., Morain-Nicolier, F., Dugelay, J. L., & Pic, M. (2019). DEFACTO : Image and Face Manipulation Dataset. 2019 27th European Signal Processing Conference (EUSIPCO), 1-5. https://doi.org/10.23919/EUSIPCO.2019.8903181

[23] Guan, H., Kozak, M., Robertson, E., Lee, Y., Yates, A. N., Delgado, A., … Fiscus, J. (2019). MFC Datasets : Large-Scale Benchmark Datasets for Media Forensic Challenge Evaluation. 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), 63-72. https://doi.org/10.1109/WACVW.2019.00018

[24] Zhou, P., Han, X., Morariu, V. I., & Davis, L. S. (2017). Two-Stream Neural Networks for Tampered Face Detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 1831-1839. https://doi.org/10.1109/CVPRW.2017.229

[25] Marra, F., Gragnaniello, D., Verdoliva, L., & Poggi, G. (2019). Do GANs Leave Artificial Fingerprints? 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 506-511. https://doi.org/10.1109/MIPR.2019.00103

[26] Verdoliva, L. (2020). Media Forensics and DeepFakes : An Overview. IEEE Journal of Selected Topics in Signal Processing, 14(5), 910-932. https://doi.org/10.1109/JSTSP.2020.3002101

[27] Korshunov, P., & Marcel, S. (2018). DeepFakes : A New Threat to Face Recognition? Assessment and Detection. ArXiv:1812.08685 [Cs]. http://arxiv.org/abs/1812.08685

[28] Dolhansky, B., Howes, R., Pflaum, B., Baram, N., & Ferrer, C. C. (2019). The Deepfake Detection Challenge (DFDC) Preview Dataset. ArXiv:1910.08854 [Cs]. http://arxiv.org/abs/1910.08854

[29] Li, Y., Yang, X., Sun, P., Qi, H., & Lyu, S. (2020). Celeb-DF : A Large-scale Challenging Dataset for DeepFake Forensics. ArXiv:1909.12962 [Cs, Eess]. http://arxiv.org/abs/1909.12962

[30] Jiang, L., Li, R., Wu, W., Qian, C., & Loy, C. C. (2020). DeeperForensics-1.0 : A Large-Scale Dataset for Real-World Face Forgery Detection. ArXiv:2001.03024 [Cs]. http://arxiv.org/abs/2001.03024

[31] Guera, D., & Delp, E. J. (2018). Deepfake Video Detection Using Recurrent Neural Networks. 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 1-6. https://doi.org/10.1109/AVSS.2018.8639163

[32] Laptev, I., Marszalek, M., Schmid, C., & Rozenfeld, B. (2008). Learning realistic human actions from movies. 2008 IEEE Conference on Computer Vision and Pattern Recognition, 1-8. https://doi.org/10.1109/CVPR.2008.4587756

[33] Liu, Z., Qi, X., & Torr, P. H. S. (2020). Global Texture Enhancement for Fake Face Detection in the Wild. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 8057-8066. https://doi.org/10.1109/CVPR42600.2020.00808

[34] Jung, T., Kim, S., & Kim, K. (2020). DeepVision : Deepfakes Detection Using Human Eye Blinking Pattern. IEEE Access, 8, 83144-83154. https://doi.org/10.1109/ACCESS.2020.2988660

[35] Korshunov, P., & Marcel, S. (2019). Vulnerability assessment and detection of Deepfake videos. 2019 International Conference on Biometrics (ICB), 1-6. https://doi.org/10.1109/ICB45273.2019.8987375

[36] Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep Face Recognition. Procedings of the British Machine Vision Conference 2015, 41.1-41.12. https://doi.org/10.5244/C.29.41

[37] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet : A unified embedding for face recognition and clustering. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 815-823. https://doi.org/10.1109/CVPR.2015.7298682

[38] Li, Y., & Lyu, S. (2019). Exposing DeepFake Videos By Detecting Face Warping Artifacts. 7.

[39] Yang, X., Li, Y., & Lyu, S. (2019). Exposing Deep Fakes Using Inconsistent Head Poses. ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 8261-8265. https://doi.org/10.1109/ICASSP.2019.8683164

[40] Guo, Z., Yang, G., Chen, J., & Sun, X. (2021). Fake face detection via adaptive manipulation traces extraction network. Computer Vision and Image Understanding, 204, 103170. https://doi.org/10.1016/j.cviu.2021.103170

[41] Chintha, A., Thai, B., Sohrawardi, S. J., Bhatt, K., Hickerson, A., Wright, M., & Ptucha, R. (2020). Recurrent Convolutional Structures for Audio Spoof and Video Deepfake Detection. IEEE Journal of Selected Topics in Signal Processing, 14(5), 1024-1037. https://doi.org/10.1109/JSTSP.2020.2999185

[42] Chollet, F. (2017). Xception : Deep Learning with Depthwise Separable Convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1800-1807. https://doi.org/10.1109/CVPR.2017.195

[43] Lee, S., Tariq, S., Shin, Y., & Woo, S. S. (2021). Detecting handcrafted facial image manipulations and GAN-generated facial images using Shallow-FakeFaceNet. Applied Soft Computing, 105, 107256. https://doi.org/10.1016/j.asoc.2021.107256

[44] Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. ArXiv:1511.06434 [Cs]. http://arxiv.org/abs/1511.06434

[45] Van Rossum, G., & Fred L, D. J. (1995). Python tutorial (Vol. 620). Centrum voor Wiskunde en Informatica Amsterdam.

[46] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., … Zheng, X. (2016). TensorFlow : Large-Scale Machine Learning on Heterogeneous Distributed Systems. ArXiv:1603.04467 [Cs]. http://arxiv.org/abs/1603.04467

[47] Manaswi, N. K. (2018). Understanding and working with Keras. In Deep Learning with Applications Using Python (p. 31-43). Apress.

[48] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., … Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585(7825), 357-362. https://doi.org/10.1038/s41586-020-2649-2

[49] Ari, N., & Ustazhanov, M. (2014). Matplotlib in python. 2014 11th International Conference on Electronics, Computer and Computation (ICECCO), 1-6. https://doi.org/10.1109/ICECCO.2014.6997585

Webgraphie

[w1] Abhipraya, K. D. *Feed Forward Neural Networks*. OpenGenus IQ: Learn Computer Science. https://iq.opengenus.org/feed-forward-neural-networks/. Dernier accès: 30/05/2021

[w2] Mittal, A. *Understanding RNN and LSTM*. Medium. https://aditi-mittal.medium.com/understanding-rnn-and-lstm-f7cdf6dfc14e. Dernier accès: 30/05/2021

[w3] Deshp, M. & e. *Introduction to Convolutional Neural Networks for Vision Tasks*. Python Machine Learning. https://pythonmachinelearning.pro/introduction-to-convolutional-neural-networks-for-vision-tasks/. Dernier accès: 15/04/2021

[w4] Habib, B. L. E. *Les réseaux de neurones convolutifs*. https://datasciencetoday.net/index.php/en-us/deep-learning/173-les-reseaux-de-neurones-convolutifs. Dernier accès: 04/05/2021

[w5] Budhiraja, A. *Learning Less to Learn Better—Dropout in (Deep) Machine learning*. Medium. https://medium.com/@amarbudhiraja/https-medium-com-amarbudhiraja-learning-less-to-learn-better-dropout-in-deep-machine-learning-74334da4bfc5. Dernier accès: 01/016/2021

[w6] HOURRANE, O. *Réseaux neuronaux récurrents et LSTM*. https://datasciencetoday.net/index.php/fr/machine-learning/148-reseaux-neuronaux-recurrents-et-lstm. Dernier accès: 01/06/2021

[w7] Sagar, R. What Are Activation Functions And When To Use Them. *Analytics India Magazine*. https://analyticsindiamag.com/what-are-activation-functions-and-when-to-use-them/. Dernier accès: 12/07/2021

[w8] Witalij, R. *SAP Tech Bytes : Your first Predictive Scenario in SAP Analytics Cloud | SAP Blogs*. https://blogs.sap.com/2021/07/09/first-predictive-scenario-sac-kaggle-titanic/. Dernier accès: 13/07/2021

[w9] Jain, N. *What is a Deepfake—How Deepfakes Work, How to FaceSwap using App*. Electricalfundablog.Com. https://electricalfundablog.com/deepfake-faceswap/. Dernier accès: 05/05/2021

[w10] Krunal, L. Python OS Module : How to Use OS Module In Python 3. *AppDividend*. https://appdividend.com/2019/02/06/python-os-module-tutorial-with-example/. Dernier accès: 25/07/2021

[w11] Lundh, F. An introduction to tkinter. *URL: www. pythonware. com/library/tkinter/introduction/index.* Dernier accès: 26/07/2021

[w12] Elia, E. *Drawing Architecure : Building Deep Convolutional GAN's In Pytorch*. Medium. https://towardsdatascience.com/drawing-architecure-building-deep-convolutional-gans-in-pytorch-5ed60348d43c. Dernier accès: 19/08/2021

[w13] Nayak, M. *Deep Convolutional Generative Adversarial Networks(DCGANs)*. Medium. https://medium.datadriveninvestor.com/deep-convolutional-generative-adversarial-networks-dcgans-3176238b5a3d Dernier accès: 19/08/2021