Comparaison de la méthode des eigenfaces et des CNN dans le cadre de la reconnaissance faciale

Simon Eveillé
Université de Technologie de Belfort-Montbéliard, Belfort, France

Mots clés : comparaison, ACP, analyse en composantes principales, CNN, réseau de neurones convolutifs, reconnaissance faciale, reconnaissance des visages, deep-learning

Résumé

Cette étude présente une analyse comparative des performances de l'Analyse en Composantes Principales (ACP) et des Réseaux de Neurones Convolutifs (CNN) dans le cadre de la reconnaissance faciale. L'ACP, utilisée avec un classifieur k-NN, se distingue par sa rapidité, avec des temps d'entraînement et de prédiction nettement plus courts, la rendant idéale pour des applications en temps réel. En revanche, les CNN, basés sur l'architecture LeNet-5, offrent une capacité d'apprentissage plus robuste dans des scénarios complexes. Bien que l'ACP ait montré de meilleurs résultats dans des conditions contrôlées, l'étude suggère que les CNN pourraient exceller dans des environnements moins prévisibles. Enfin, l'étude ouvre la voie à une approche hybride combinant l'ACP et les CNN pour optimiser la précision et la rapidité tout en minimisant les ressources de calcul nécessaires.

I. Introduction

Un système de reconnaissance facial est un programme capable de différencier et classifier des visages humains issus de photos ou de vidéos. Afin de pouvoir sélectionner l'algorithme de reconnaissance le plus adapté dans un contexte donné, il est important de connaître les différents avantages et inconvénients des modèles disponibles. En comparant les algorithmes sur divers critères, tels que leur précision, leur temps d'entraînement ou leur temps de prédiction, il sera possible de déterminer quelle méthode adopter dans tel ou tel scénario particulier.

Dans cet article, nous proposons donc une analyse comparative de l'analyse en composante principale (ACP) et des réseaux de neurones convolutifs (CNN). L'ACP, parfois appelée méthode eigenfaces lorsqu'appliquée à la reconnaissance faciale, est une approche statistique et déterministe de la problématique de l'analyse biométrique des visages. Son objectif est de réduire au maximum la taille des données à étudier, dans notre cas : les valeurs d'intensité des pixels d'une image, tout en conservant le maximum d'information dans ces vecteurs réduits. Les réseaux de neurones convolutifs (CNN, pour convolutional neural network en anglais) sont eux issus des récentes avancées dans le domaine de l'IA et plus précisément du deep-learning. Ce sont des modèles de réseaux de neurones spécialisés dans le traitement d'image. Ils sont en effet capables de hiérarchisés les différentes caractéristiques d'images, ce qui les rends extrêmement bons dans les tâches de classification telle que la reconnaissance faciale.

II. Revue de la littérature

La publication « Research on Face Recognition Based on CNN » de Jie Wang et

Zihao Li [1] propose une étude théorique du fonctionnement standard des réseaux de neurones. En particulier elle présente le modèle d'architecture LeNet-5 proposé par le chercheur en intelligence artificielle Yann LeCun en 1998 [2]. C'est sur cette architecture que nous baserons notre modèle de réseau de neurones convolutif dans cette étude. En effet, de nombreuses études, dont les deux précédemment citées démontrent l'efficacité de cette architecture pour les problèmes de classification d'images.

Les auteurs de la publication « A Review Paper Comparing the CNN, LBPH, and PCA Face Recognition Algorithms », Divya Maithreyi Tenneti, Divya A Kittur et Jyothika K Raju [3] ont produit une étude comparative de différents algorithmes, dont les réseaux de neurones convolutifs et l'analyse en composantes principales. Les auteurs évoquent notamment les différences de performances lors de l'usages de datasets (ensembles de données) aux caractéristiques différentes. Par exemple, il est évoqué que des algorithmes comme LBPH (Local Binary Pattern Histogram) obtiennent de meilleurs résultats que les CNN ou que l'ACP lorsque les luminosités des images étudiées ont tendance à être hétérogènes. Cependant, cette étude, très penchée sur la précision des algorithmes testés, rentre moins dans les détails en ce qui concerne les durées requises pour utiliser de tels modèles.

Enfin, l'étude « Face Recognition Using PCA and SVM », menée par Omar Faruqe et Al Mehedi Hasan en 2009 [4] montre bien que la PCA n'est pas suffisante à elle seule pour émettre une décision de classification : un classificateur est ensuite nécessaire pour déterminer si telle image appartient plutôt à telle ou telle classe. Dans le cas de cette étude, c'est l'efficacité des SVM (Machines à Vecteur de Support) en tant que classifieur qui est étudiée. Dans notre étude, nous proposons d'utiliser un mécanisme de classification plus simple, à savoir les classifieurs k-NN basés sur la méthode des k plus proches voisins.

III. Méthodologie

Cette étude vise à réaliser une analyse comparative des méthodes basées sur l'Analyse en Composantes Principales (ACP) et des méthodes basées sur les réseaux de neurones convolutifs (CNN) pour la reconnaissance faciale. Bien que ces deux approches aient démontré leur efficacité dans diverses applications de traitement d'images, il est important de noter qu'elles reposent sur des concepts et des mécanismes de fonctionnement fondamentalement différents. Par conséquent, les résultats qu'elles produisent ne sont pas directement comparables sans ajustements spécifiques. L'ACP, en tant que méthode statistique, est principalement axée sur la réduction de la dimensionnalité des données. Son objectif est de compresser les données tout en conservant l'essentiel des informations pertinentes, c'est-à-dire celles qui permettent de distinguer les différents individus dans un ensemble d'images. En simplifiant les données, l'ACP permet de réduire le bruit et d'améliorer l'efficacité des algorithmes de classification qui sont appliqués par la suite.

D'un autre côté, les réseaux de neurones convolutifs, et en particulier les architectures comme LeNet-5, adoptent une approche entièrement différente. Contrairement à l'ACP, qui extrait manuellement des composantes principales, les CNN sont capables d'apprendre automatiquement les caractéristiques discriminantes des données grâce à leurs multiples couches de convolutions, de pooling et de couches entièrement connectées. Ces réseaux sont conçus pour recevoir une image brute en entrée et, à travers une série d'opérations complexes, ils produisent directement en sortie une classification, c'est-à-dire l'identification de la classe à laquelle l'image appartient. Cela fait des CNN une méthode très puissante pour la reconnaissance

d'objets et de visages, car ils peuvent détecter et apprendre des motifs complexes dans les données sans intervention humaine pour la sélection des caractéristiques.

Cependant, dans le cadre de cette étude, pour permettre une comparaison équitable entre l'ACP et les CNN, il est nécessaire d'introduire un algorithme de classification supplémentaire à l'ACP, puisque celle-ci, en tant que technique de réduction de dimensionnalité, ne réalise pas de classification par elle-même. En d'autres termes, bien que l'ACP puisse projeter les données dans un espace de dimension réduite, elle ne fournit pas une sortie sous forme de catégorie qui pourrait être directement comparée aux résultats des CNN. Pour combler cette lacune, nous avons choisi d'associer l'ACP à la méthode des k plus proches voisins (k-NN), un classifieur populaire qui fonctionne en mesurant la proximité entre les données dans l'espace projeté par l'ACP.

La méthode des k plus proches voisins (k-NN en anglais pour k-Nearest Neighbours) est une technique de classification basée sur la notion de distance entre 2 enregistrements pour émettre une classification. Une fois que toutes les images du dataset d'entrainement ont été vectorisées et projetées dans l'espace défini par l'ACP, une nouvelle image peut être, de la même manière, vectorisée et projetée pour ensuite être classifiée selon k-NN. La classe choisie correspond alors à la classe majoritaire dans l'ensemble des k voisins les plus proches trouvés dans le dataset d'entraînement. La notion de voisin est définie par une fonction permettant d'évaluer la distance entre 2 enregistrements. Dans notre cas nous utiliserons la notion de distance euclidienne « standard » parfaitement applicable aux espaces de dimension N ($N \in \mathbb{N}^*$). Il faut néanmoins garder en tête que l'une des faiblesses de cette méthode de classification repose sur le fait que l'on recherche la classe majoritaire parmi les données présente dans le dataset d'entraînement. Une sur-représentation d'une classe par rapport à une autre peut alors entraîner des conséquences négatives sur la précision de la classification.

Ainsi notre étude comparative sera construite de cette manière :

1. Normalisation de l'ensemble d'images de départ : notre dataset est constitué de 152 images d'un groupe composé de 14 personnes (environ une dizaine d'images par personne), nous allons donc le diviser, en prenant deux tiers pour la base d'entraînement, et un tiers pour la base de test. Ces images seront ensuite toutes converties aux format PNG, et redimensionnées en 64*64.

2. Entraînement des 2 modèles à savoir d'une part, le réseau de neurones convolutifs basé sur l'architecture LeNet-5, et d'autres par celui basé sur la combinaison d'ACP et des k plus proches voisins.

Plusieurs paramètres doivent être déterminés au cours de cette phase, comme le nombre de passes d'entraînements faites sur le réseau de neurones (souvent référencées comme le nombre d'époques ou epochs en anglais), ainsi le paramètre k du classifieur basé sur les k plus proches voisins. Dans les 2 cas, l'objectif est de réduire au maximum ces 2 paramètres tout en conservant une précision satisfaisante. En effet, le nombre d'époques d'entraînement pour le réseau de neurones est linéairement corrélé au temps d'entraînement du réseau. Le paramètre k de la méthode k-NN a lui une influence, moins importante, mais à conserver en considération sur les temps de classification donnés par k-NN. Une fois ces paramètres déterminés, on pourra précisément comparer les temps d'entraînement des deux modèles, à l'issue de cette deuxième étape.

3. Test des deux modèles, en émettant des prédictions pour chacune des images de l'ensemble de test. On pourra alors comparer les taux de précision globaux et par classes

dans des matrices de confusion, ainsi que les temps moyens pour émettre une prédiction pour chacun des deux modèles.

En ce qui concerne l'implémentation technique de l'expérience, le programme sera écrit en Python, avec la partie CNN réalisée avec l'implémentation open-source de la bibliothèque PyTorch, et la partie ACP/k-NN réalisée en utilisant les implémentations mises à disposition par la bibliothèque open-source Scikit-learn.

Le code sera mis à disposition à travers un notebook jupyter disponible sur GitHub : https://github.com/eveillesimon/face-recognition-acp-vs-cnn

IV. Présentation des résultats

La détermination des paramètres utilisés au cours de l'expérience a été un processus itératif et méthodique, réalisé selon les étapes définies précédemment. Les deux principaux paramètres à ajuster étaient, d'une part, le nombre d'époques pour l'entraînement du modèle basé sur les réseaux de neurones convolutifs (CNN), et d'autre part, le nombre k de voisins à prendre en compte lors de l'utilisation du classifieur basé sur la méthode des k plus proches voisins (k-NN) pour le modèle ACP.

Nombre d'époque pour le CNN

Le choix du nombre d'époques, qui correspond au nombre de passages complets sur l'ensemble des données d'entraînement, a été effectué sur la base des résultats présentés dans les annexes 1 et 2. En particulier, on observe clairement sur l'annexe numéro 2 que le taux de perte se stabilise autour de la dixième époque d'entraînement. Cela signifie que, dès ce stade, le modèle a atteint un niveau de convergence acceptable, où les ajustements des poids et des biais cessent d'améliorer significativement la performance du réseau. Plus d'entraînement risquerait de diminuer la qualité des prédiction (voir phénomène d'overfitting pour plus d'informations à ce sujet [5]). Pour ces raisons, la valeur retenue pour le nombre d'époques d'entraînement est 10.

Paramètre k du k-NN

Concernant la méthode k-NN, la valeur optimale de k, qui détermine le nombre de voisins les plus proches pris en compte lors de la classification, a été déterminée après une série de tests de précision. Ces tests ont été réalisés en faisant varier la valeur de k entre 1 et 50, afin d'évaluer l'impact de ce paramètre sur les performances du modèle ACP associé à k-NN. Les résultats de ces expériences, présentés dans l'annexe 3, montrent que la précision du modèle tend à diminuer de manière significative à mesure que la valeur de k augmente. Cela peut s'expliquer par le fait qu'avec un k trop élevé, le modèle prend en compte un trop grand nombre de voisins lors de la classification, y compris des points qui peuvent appartenir à des classes différentes, ce qui augmente le risque d'erreur. Cette chute de précision au fur et à mesure que k augmente indique que le modèle ACP se comporte mieux avec une valeur de k relativement faible. En fin de compte, le choix s'est porté sur k = 1, car c'est avec cette valeur que la précision du modèle est restée la plus élevée.

Temps d'entraînement des modèles

En ce qui concerne l'entraînement des deux modèles, le temps de calcul nécessaire pour déterminer les paramètres du modèle d'ACP a été bien plus court que le temps d'entraînement du réseau de neurones. Cela s'explique par la nature intrinsèque des deux méthodes : l'ACP, étant une technique de réduction de dimensionnalité basée sur des calculs linéaires, est moins coûteuse en termes de calculs. Une fois la réduction de

dimensionnalité effectuée, l'algorithme k-NN est relativement rapide, puisqu'il ne nécessite pas d'entraînement à proprement parler, mais repose simplement sur des calculs de distance entre points. En revanche, le CNN, avec ses multiples couches convolutives, ses paramètres à optimiser et ses itérations répétées à travers les données d'entraînement, nécessite un temps de calcul beaucoup plus important.

	Temps d'entraînement (en s)
CNN	1.585
ACP et k-NN	0.231

En ce qui concerne les temps de prédiction, les résultats suivent la même tendance que ceux observés pour les temps d'entraînement. Plus précisément, les temps de prédiction du modèle basé sur l'ACP sont significativement plus courts que ceux du modèle CNN, avec une différence d'un ordre de grandeur.

	Temps de prédiction pour l'ensemble de test (en ms)
CNN	56.51
ACP et k-NN	5.00

On a ensuite pu calculer le temps moyen de prédiction par image :

	Temps de prédiction moyen pour une image (en ms)
CNN	1.13
ACP et k-NN	0.10

Précision des modèles

À propos des taux de précision des deux modèles, les résultats montrent des taux de précision élevés pour les deux méthodes.

	Taux de précision observé (en %)
CNN	92
ACP et k-NN	94

À savoir que le taux de précision du modèle basé sur le réseau de neurones convolutif varie lors des différentes répétitions de l'expérience, en raison de la nature aléatoire des paramètres initiaux de réseau. Ce taux a varié entre 88% et 96% au cours des différentes expériences.

Enfin, les matrices de confusion présentées dans les annexes 4 et 5 révèlent que les classes confondues par les deux modèles, ACP associé à k-NN d'une part, et le CNN d'autre part, ne semblent pas avoir de lien direct. Autrement dit, les erreurs de classification commises par les deux approches ne concernent pas les mêmes catégories d'images, ce qui pourrait indiquer que chaque modèle a des points faibles distincts en termes de reconnaissance. Cependant, il est important de noter que nos données ne permettent pas de tirer des conclusions définitives à ce stade. Nous disposons de trop peu d'informations pour déterminer si ces différences sont réellement significatives ou si elles sont simplement dues à des facteurs aléatoires, tels que le bruit dans les données ou des imprécisions liées à notre processus de normalisation des images.

V. Conclusion

Nos données mettent en lumière les différences fondamentales entre les deux approches étudiées pour la reconnaissance faciale, à savoir l'analyse en composantes principales (ACP) et les réseaux de neurones convolutifs (CNN). Les résultats révèlent que le modèle basé sur l'ACP offre des temps d'entraînement et de prédiction nettement plus rapides comparés à ceux des CNN, ce qui semble indiquer une meilleure adéquation de l'ACP pour des applications où la réactivité est cruciale, telles que les

systèmes de reconnaissance en temps réel. Ce gain de rapidité est un avantage majeur dans les contextes où les ressources informatiques sont limitées ou lorsque des décisions rapides sont nécessaires, comme dans les dispositifs de sécurité ou les systèmes embarqués.

Cependant, bien que l'ACP se démarque par sa rapidité, il convient de nuancer cette conclusion. Notre expérience a été réalisée sur un ensemble d'images capturées dans des conditions idéales : un bon éclairage et un fond uniforme (blanc). Ces conditions contrôlées facilitent sans doute la tâche des algorithmes, ce qui pourrait influencer les résultats en faveur de l'ACP. En réalité, dans des environnements moins idéaux, où les images sont prises sous divers angles, avec des éclairages variés, ou des arrière-plans complexes, les performances pourraient diverger. Dans ces cas, il est possible que les réseaux de neurones convolutifs, réputés pour leur robustesse et leur capacité à généraliser à des situations complexes, prennent l'avantage. Il serait donc pertinent d'étendre cette étude à un second ensemble d'images capturées dans des conditions plus variées pour évaluer la précision des deux modèles dans un contexte plus réaliste. Cela pourrait permettre de trancher plus clairement sur les avantages et les inconvénients respectifs de chaque méthode, en particulier en termes de robustesse face à la variabilité des données.

Une autre question importante qui émerge de cette étude concerne la possibilité de combiner ces deux approches. Il serait intéressant d'explorer une architecture hybride où les données d'entrée seraient d'abord réduites dimensionnellement par un modèle basé sur l'ACP avant d'être transmises à un réseau de neurones convolutifs pour la classification finale. Cette approche pourrait potentiellement offrir le meilleur des deux mondes : la réduction de la dimensionnalité permettrait de simplifier le travail du CNN en limitant le nombre de neurones et de connexions nécessaires, ce qui se traduirait par un modèle plus léger et plus rapide à entraîner. En théorie, un réseau neuronal moins complexe, fonctionnant sur des données déjà optimisées, pourrait réduire considérablement le temps d'entraînement tout en maintenant une précision élevée.

Une telle approche hybride réduirait également la consommation de ressources computationnelles, un aspect particulièrement pertinent dans des environnements où la capacité de traitement est restreinte, comme les appareils mobiles ou les systèmes embarqués.

VI. Références

[1] Jie Wang, Zihao Li, "Research on Face Recognition Based on CNN", 2018

[2] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, "Gradient-Based Learning Applied to Document Recognition", 1998

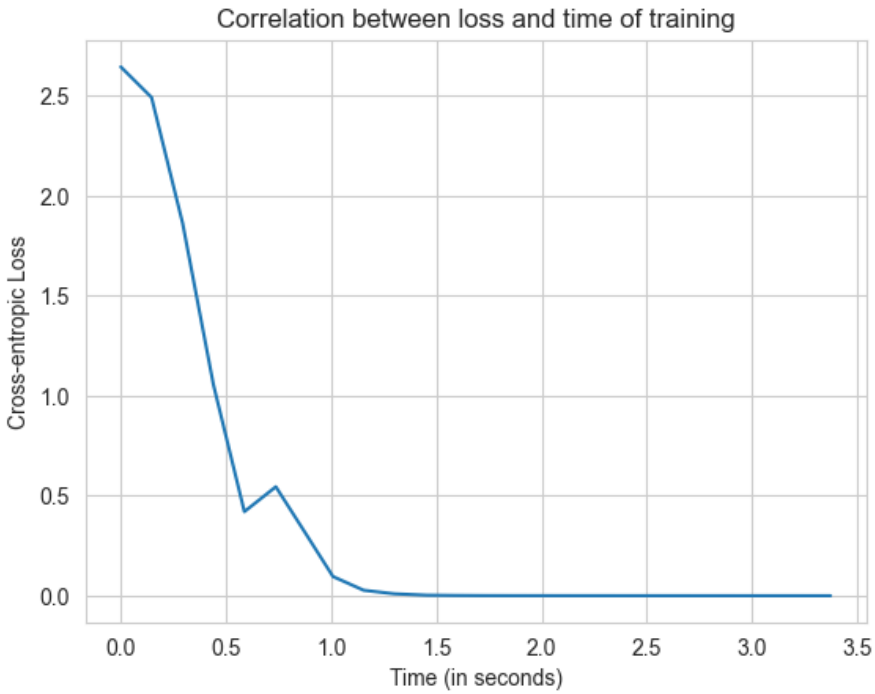
[3] Divya Maithreyi Tenneti, Divya A Kittur, Jyothika K Raju, "A Review Paper Comparing the CNN, LBPH, and PCA Face Recognition Algorithms", 2023

[4] Omar Faruqe, Al Mehedi Hasan, "Face Recognition Using PCA and SVM", 2009

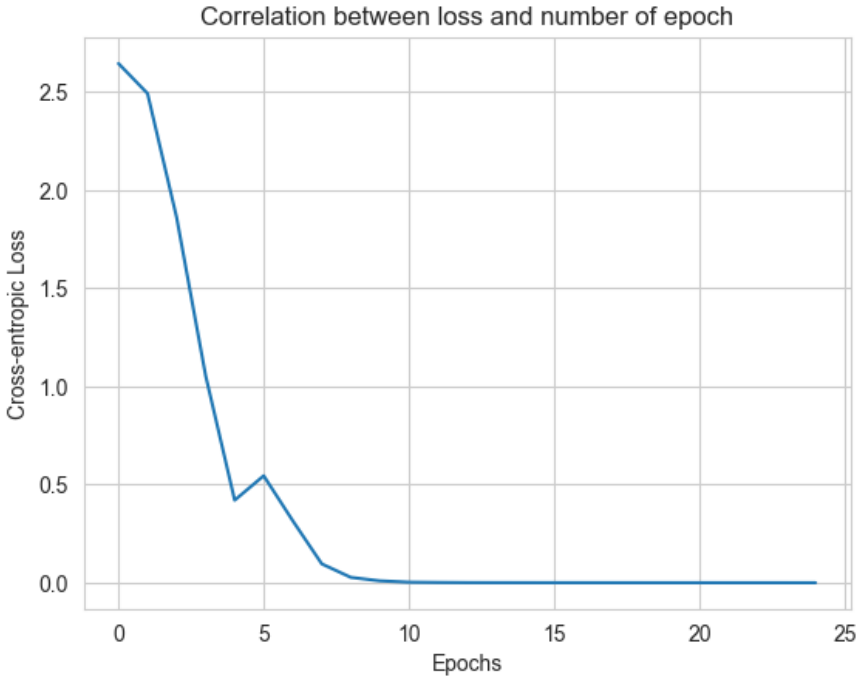
[5] Tom Dietterich, "Overfitting and Undercomputing in Machine Learning", 1995

VIII. Annexes

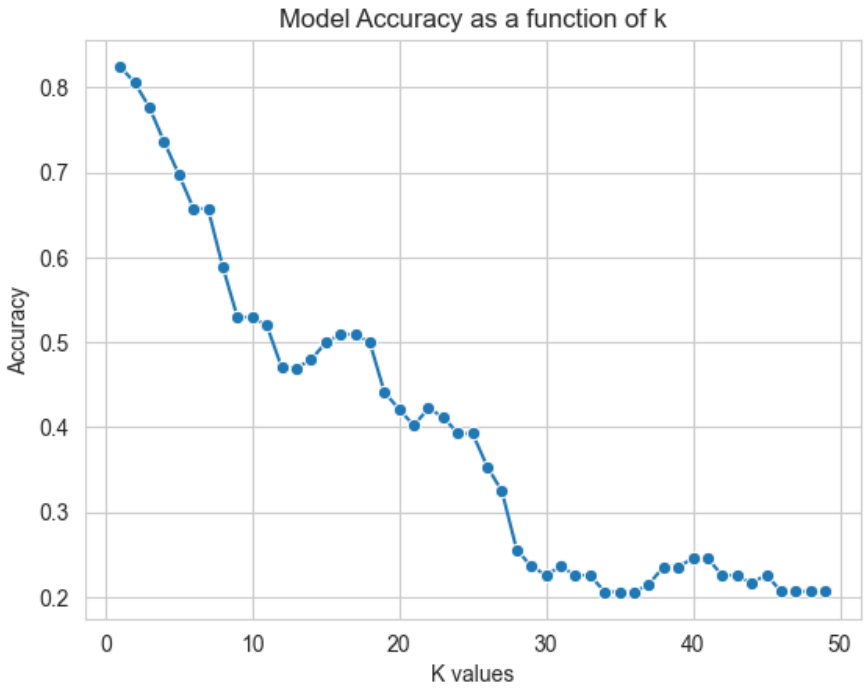
Annexe n°1 : Erreur (calcul selon la méthode cross-entropic loss) du modèle en fonction du temps



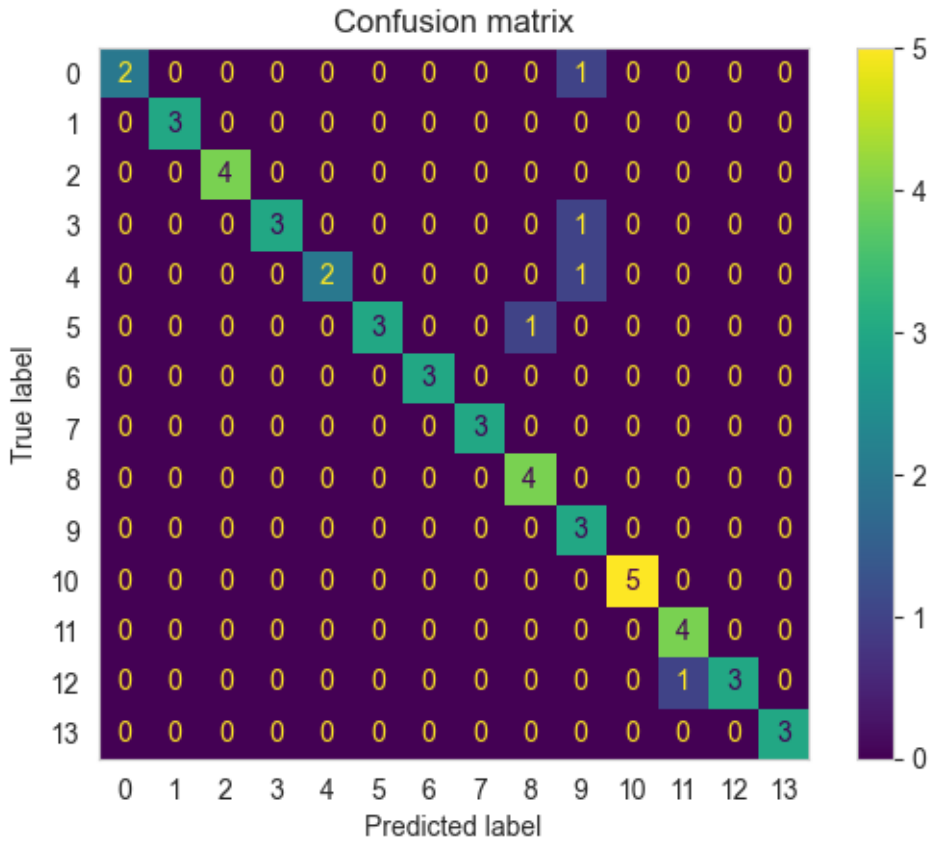
Annexe n°2 : Erreur (calcul selon la méthode cross-entropic loss) du modèle en fonction du nombre d'époques d'entraînement



Annexe n°3 : Taux de précision du modèle ACP + k-NN en fonction de différentes valeurs de k



Annexe n°4 : Matrice de confusion pour le modèle à base de CNN



Annexe n°5 : Matrice de confusion pour le modèle à base d'ACP