



RAPPORT

Amélioration des performances en classification d'images

Présenté par SEKPONA Kokou Sitsopé, Etudiant
En Ingénierie Machine Learning



Table de Matières

- I. Introduction**
- II. Contexte du projet de stage**
- III. Dataset du travail**
- IV. Contraintes du Projet**
- V. Etat de l'Art**
- VI. Mission**
- VII. Implémentation des algorithmes**
- VIII. Comparaison et choix d'algorithme de production**
- IX. Bilan**
- X. Conclusion**



Introduction

La classification d'images est un sujet très ancien dans l'histoire du Deep Learning. Depuis la sortie des réseaux de neurones convolutif en 2012, de nouveaux algorithmes ne cessent d'apparaître, dans le but d'augmenter les précisions et de mieux prédire que les algorithmes existants.

Mais le problème de classification est un problème complexe en soi, fonction de plusieurs paramètres, ce qui fait qu'on ne peut qu'essayer d'améliorer les performances pour être plus proche de la réalité au lieu de chercher à avoir un modèle parfait.

Comment donc améliorer les performances d'un modèle de classification d'images?

Contexte

Nous avons aidé l'association de protection de notre quartier en leur implémentant un algorithme capable de classer les images des animaux.

Mais le modèle n'a pas une bonne précision sur les données, ce qui peut les induire en erreur.

Nous voulons trouver le bon modèle et/ou les bons paramètres pour pouvoir améliorer la précision de la classification.

Choix du Jeu de données

Comme il s'agira d'améliorer les performances de ce problème sur lequel nous avons travaillé au projet précédent, nous allons reprendre la même base de données (Stanford Dog Dataset) qui est répartie en 120 classes et ayant au total 20580 images avec en moyenne 150 images par classe.



Contraintes imposés:

La réalisation de projet impose les contraintes suivantes :

- ✓ Choix de l'algorithme : Doit provenir d'un article de recherche publié au cours des 18 derniers mois et la méthode utilisée doit être publiée dans les 5 dernières années.
- ✓ Obtenir une meilleure précision et un faible cout (facultatif)
- ✓ S'appuyer sur des sources sérieuses.
- ✓ Soumission du plan prévision de travail au plus dans 2 jours après le début du projet

Nous avons donc pris en compte ces contraintes dans l'Etat de l'art et le reste du projet.

Etat de l'art

Ce contexte a poussé la curiosité de plusieurs chercheurs qui ont développés de nouveaux algorithmes et méthodes.

Nous allons nous baser sur les plus récents, qui sont les Visual Transformers.

Sources :

- ✓ **Vision Transformers provably learn spatial structure:**
<https://arxiv.org/abs/2210.09221>
- ✓ **Understanding the Vision Transformer and Counting Its Parameters**
<https://medium.com/analytics-vidhya/understanding-the-vision-transformer-and-counting-its-parameters-988a4ea2b8f3>
- ✓ **Image classification with Vision Transformer**
https://keras.io/examples/vision/image_classification_with_vision_transformer/



- ✓ **Visual Transformers Documentation :**
https://huggingface.co/docs/transformers/model_doc/vit
- ✓ **Fine-Tune Visual Transformers for Image Classification with Transformers :**
<https://huggingface.co/blog/fine-tune-vit>
- ✓ **Vision Transformer (ViT) Fine-Tuning :**
<https://www.kaggle.com/code/raufmomin/vision-transformer-vit-fine-tuning>

Ces articles, blog, sont les plus récents qui ont de bonnes performances en classification d'images.

En fonction des bases de données, les résultats de ces modèles changent.

Le premier article affirme clairement qu'en général, les ViT ont une bonne performance par rapport aux CNN et aux réseaux existants. Mais comme cet article ne contient pas de code, il n'a servi qu'à nous spécifier quel algorithme utiliser.

Le Deuxième article explique l'algorithme et les formules mathématiques sur lesquels il est basé

L'article 3 est une implémentation du ViT en python sur le dataset de cifar100.

Les L'article 4 est une implémentation en python du ViT avec Fine Tuning

Le premier article introduit les Visual Transformers, le 2 ème et le 3 ème sont des documentations pour l'implémentation de ce model donc contiennent du code sur lequel nous sommes appuyés et enfin les 2 derniers parlent du Fine Tuning sur le ViT qui est une manière d'augmenter les performances du ViT en ajoutant à la version pré entraîné d'autres couches.

Notre travail se situe dans le même cadre de ces articles qui est la classification multi classes.

Mission

- ✓ Etude général du projet, état de l'art
- ✓ Soumission du plan de projet prévisionnel au mentor
- ✓ Implémentation des différents modèles selon les revues scientifiques
- ✓ Choix de différents paramètres des modèles et prise de note des résultats



- ✓ Comparaison avec le résultat précédent.
- ✓ Décision finale

Implémentation des algorithmes

Nous réalisons :

- ✓ L'étude de la base de données
- ✓ L'EDA
- ✓ La répartition en Train set et en Test set
- ✓ Implémentation des algorithmes de classification

❖ Implémentation du premier algorithme : Visual Transformers



Les Visual Transformers divisent l'image en différents patches avant d'y passer le model, ce qui n'est pas le cas chez les autres modèles.

Exemple d'image du jeu de données divisé en patches





- Avec 100 epochs, input_image_size=(72, 72, 3)

Test accuracy: 0.89%

- Avec 200 epochs, input_image_size=(72, 72, 3)

Test accuracy: 0.86%

- Avec 100 epochs, input_image_size = 32, 32, 3

Test accuracy: 0.97%

La documentation du ViT suggère la méthode fine-Tuning qui peut augmenter la précision sur certaines bases de données. Comme nous n'avons pas eu d'amélioration par rapport au résultat précédent, nous allons implémenter le ViT Fine-Tuning

❖ Implémentation du ViT Fine Tuning

Nous utilisons le model pré entraîné Model Pré entraîné ViT B32 avec 10 epochs et nous lui ajoutons à la sorties quelques couches de normalisation, de Flatten et d'activation.

L'entraînement a été rapide.

Perte sur le Test_set: 3.128481149673462

Précision sur le Test_set: 50.116 %

Comparaison et choix d'algorithme pour la production

Le model Vgg16 avec Fine Tuning jugé comme meilleur model par rapport au CNN dans le projet précédent, a eu en effet de faibles performances:

Perte sur le Test_set: 8.028903007507324

Précision sur le Test_set: 22.86 %



Mais ce modèle a pu réaliser une bonne prédiction sur une image prise au hasard.

Décision Finale

Nous retenons le « ViT with Fine Tuning » qui sera alors le nouveau modèle de la production pour l'association à cause du faible cout et de sa précision de 50.1%

Bilan

Ce projet nous a permis d'agir positivement sur la performance de la classification pour cette association en augmentant le score de précision de près de 30% et en réduisant également le cout

Ces résultats obtenus prouvent que la précision dans le problème de classification est relative à la répartition et la taille de base de données, aux hyper paramètres choisis, parfois au nombre d'epochs etc...

Conclusion

Ce projet nous a rapproché de notre projet professionnel, celui de travailler en tant qu'ingénieur dans une industrie de conception de voiture autonome ; ces voitures autonomes qui utilise la segmentation et la classification pour la prise de décision.

Nous avons appris à nous adapter aux changements nouveaux en implémentant un algorithme datant de moins de 18 mois, tout en respectant certaines contraintes comme le temps pour établir et soumettre un plan de travail prévisionnel.