

Rapport Projet 7

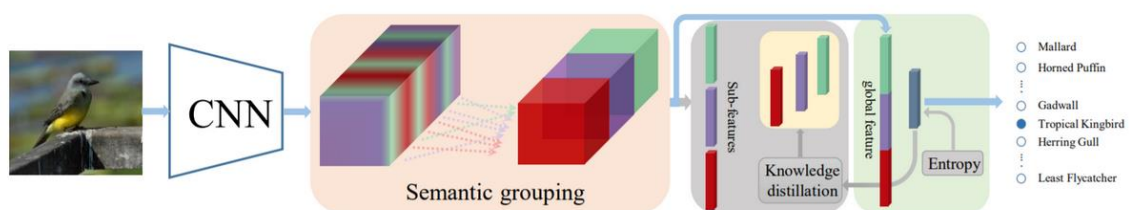
Duleme Méyi, OC

Le dataset est le Stanford Dogs, qui est lui-même un extrait d'ImageNet. Ce dataset contient 120 races pour un total de 20 580 images.

La baseline réalisée était l'utilisation d'un transfer learning de VGG-16, qui est un Very Deep Convolutional Neural Network. Les quatre dernières couches (bloc 5) ont été entraînées sur 10 races auxquelles se cantonnent mon analyse. Un score de 0.90 d'accuracy pour le training et 0.68 pour la validation ont été obtenus.

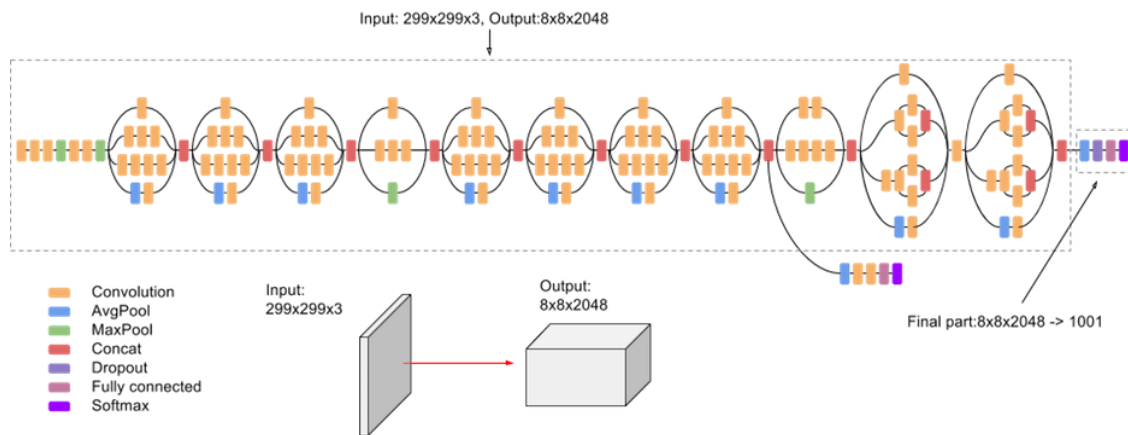
La problématique de ce projet est de prendre en entrée une photographie de chien et de fournir en sortie la race. C'est un problème dit de fine-grained recognition. Par définition, c'est donc la tâche de distinguer des objets visuellement très similaires. Les différences apparentes entre ce type de catégories sont très subtiles. Elles peuvent être brouillées par des facteurs comme le point de vue ou la localisation de l'objet sur l'image. Cela augmente la difficulté de cette tâche.

Plusieurs approches ont été proposées récemment [2]. Elles incluent souvent la localisation des parties pertinentes des objets suivie d'un apprentissage des features ainsi localisées. Pour effectuer cela, plusieurs propositions ont été implémentées. Notamment la semantic segmentation [3] qui consiste à classifier chaque pixel d'une image en une instance. Chaque instance (ou catégorie) correspondant à un objet ou une partie de l'image (roue, ciel, oeil,...). Le modèle qui m'intéresse n'est pas celui qui a obtenu les résultats les plus impressionnants. Cependant, je le trouve intéressant de par son approche. En effet, le but dans cet article [1] a été de simplifier un modèle existant pour obtenir des résultats similaires aux modèles classiquement utilisés. Je suis admirative de ce genre d'avancées. En effet, contrairement aux modèles précédents qui se basent sur des modules complexes de localisations de parties, il s'agit ici d'améliorer la segmentation sémantique des sous-features d'une feature plus large. Le coût computationnel est donc diminué, tout en conservant l'efficacité du modèle (88.8% accuracy pour le Stanford Dog Dataset).



L'algorithme envisagé est le SEF (Semantically enhanced feature for fine-grained image classification, IEEE, 2020 [1]). Le code de cet article est disponible sur GitHub [8]. Il y est indiqué qu'il est possible de télécharger les modèles préentraînés. Malheureusement en allant sur le lien et en suivant les indications, il s'agit de Baidu Netdisk. Ce logiciel n'est accessible qu'avec un code de vérification à envoyer à un numéro de téléphone Chinois ou Américain. Il m'a donc été impossible d'implémenter les modèles pré-entraînés. Je suis donc partie du code de départ, et ai modifié quelques modules comme torch._six, qui exigeaient des ajustements. Malheureusement l'organisation des fichiers du dataset n'étant pas précisés, mon code a rencontré de nombreux bugs. Cette tentative de Benchmarking a donc été un échec.

Une alternative à ce modèle, est le modèle Inception Network V3 de Google [4].



En effet, même si les parties de l'image ne sont pas explicitement sélectionnées puis isolées, elles le sont implicitement [6,7]. En effet, le principe du réseau inception est d'utiliser en parallèle différents filtres puis de les empiler. C'est d'ailleurs pour cela que ce modèle a été proposé dans une problématique de Fine-Grained Catégorisation [5], dans le cas d'un dataset d'espèces de chanvre (DeepWeed, 2019).



Le modèle est disponible sur le github de tensorflow, dans la partie recherche. Je l'ai donc implémenté et ai obtenu 99% d'accuracy sur le training set et 87% d'accuracy sur le validation set. Ces résultats sont effectivement meilleurs que ma baseline.

Références :

1. Luo, Wei, et al. "Learning semantically enhanced feature for fine-grained image classification." IEEE Signal Processing Letters 27 (2020): 1545-1549
2. <https://www.coursera.org/lecture/visual-perception-self-driving-cars/lesson-2-convnets-for-semantic-segmentation-ii7Th>
3. https://medium.com/@arthur_ouaknine/review-of-deep-learning-algorithms-for-image-semantic-segmentation-509a600f7b57

4. Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
5. Olsen, Alex, et al. "DeepWeeds: A multiclass weed species image dataset for deep learning." *Scientific reports* 9.1 (2019): 1-12.
6. <https://github.com/tensorflow/tpu/tree/master/models/experimental/inception>
7. <https://cloud.google.com/tpu/docs/inception-v3-advanced>
8. <https://github.com/cswluo/SEF/blob/master/README.md>