

Plan prévisionnel

Dataset retenu

Stanford Dogs Dataset est un jeu de données composé de 20 580 images réparties sur 120 races de chiens différentes. Les images couvrent un large éventail de poses, d'angles et de contextes d'environnement. Chaque image est annotée avec l'étiquette de la race, permettant à la fois l'entraînement et l'évaluation rigoureuse de modèles de deep learning. Le dataset est réduit à 10 races représentatives pour accélérer l'itération.

Modèle envisagé

Je propose d'utiliser **EfficientNetV2**, publié par Google Research en 2021. EfficientNetV2 optimise simultanément la profondeur, la largeur et la résolution d'entrée, tout en intégrant des kernels de convolution pour accélérer la convergence et réduire la consommation de calcul. Le modèle atteint 83,2 % de précision Top-1 sur ImageNet tout en restant compact avec seulement 7,1 M de paramètres, ce qui en fait un des meilleurs modèles CNN en termes de rapport performance/coût de calcul. L'introduction des blocs Fused-MBConv accélère la convergence, réduisant significativement le temps d'entraînement. Le modèle visera à classer précisément les races du Stanford Dogs Dataset. EfficientNetV2 peut être déployé pour des applications mobiles ou embarquées grâce à son compromis optimal entre précision et rapidité, dans des domaines vétérinaires et de reconnaissance d'images sur smartphone.

Références bibliographiques

1. Mingxing Tan, Quoc V. Le, « **EfficientNetV2: Smaller Models and Faster Training** », arXiv:2104.00298, 2021.
2. Mooney Paul , « **EfficientNetV2 with TensorFlow** », Kaggle Notebook, 2023, <https://www.kaggle.com/code/paultimothymooney/efficientnetv2-with-tensorflow>.
3. D. Majchrowski, K. Kudrynski, « **NVIDIA Deep Learning Examples for Tensor Cores** », GitHub, 2024, <https://github.com/NVIDIA/DeepLearningExamples>
4. Google Research, Brain Team, « **AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE** », arxiv:2010.11929, 2021

Explication de votre démarche de test du nouvel algorithme (votre preuve de concept) :

Pour évaluer la supériorité d'EfficientNetV2 sur le Stanford Dogs Dataset, je vais commencer par établir une baseline. Pour cela j'utiliserai un ResNet-50 pré-entraîné sur ImageNet, dont les couches seront fine-tunées sur 80 % des images du jeu de données, 10 % étant réservées à la validation et les derniers 10% pour le test. Cette première étape nous permettra de mesurer la précision Top-1, Top-3 ainsi que le F1-score macro, mais également le temps d'entraînement et d'inférence afin d'évaluer la performance atteinte.

La phase suivante portera sur l'implémentation d'EfficientNetV2, en s'appuyant sur le modèle "v2-S" sera chargé avec ses poids pré-entraînés sur ImageNet-1k, puis adapté à notre tâche en remplaçant la couche finale par une couche Dense à 10 neurones (softmax) précédé par une couche de Dropout. Les hyperparamètres qui seront optimisés en utilisant une GridSearch sont le Learning Rate, la Batch Size, le Dropout Rate ainsi que le nombre d'époque.

J'ai entraîné le modèle sur des images redimensionnées en 224×224 et augmentées par la bibliothèque Albumentations qui permet de générer divers augmentations sur nos images. Les augmentations appliquées sont : un Horizontal/Vertical Flip, une égalisation, .

Enfin, une évaluation comparative viendra confirmer les gains apportés par EfficientNetV2 : le jeu de données sera une nouvelle fois découpé selon même la répartition avec une graine fixe pour assurer la reproductibilité. Je comparerai alors les deux modèles, en analysant leurs courbes d'apprentissage pour observer la vitesse de convergence, et en mesurant le temps d'inférence de chacun des modèles.