Projet 7

Preuve de concept

Classification d'images de style cartoon avec un Vision Transformer (ViT)

Comparaison avec un réseau de neurones convolutionnels (CNN)

Marie-France LAROCHE-BARTHET

20/03/2022

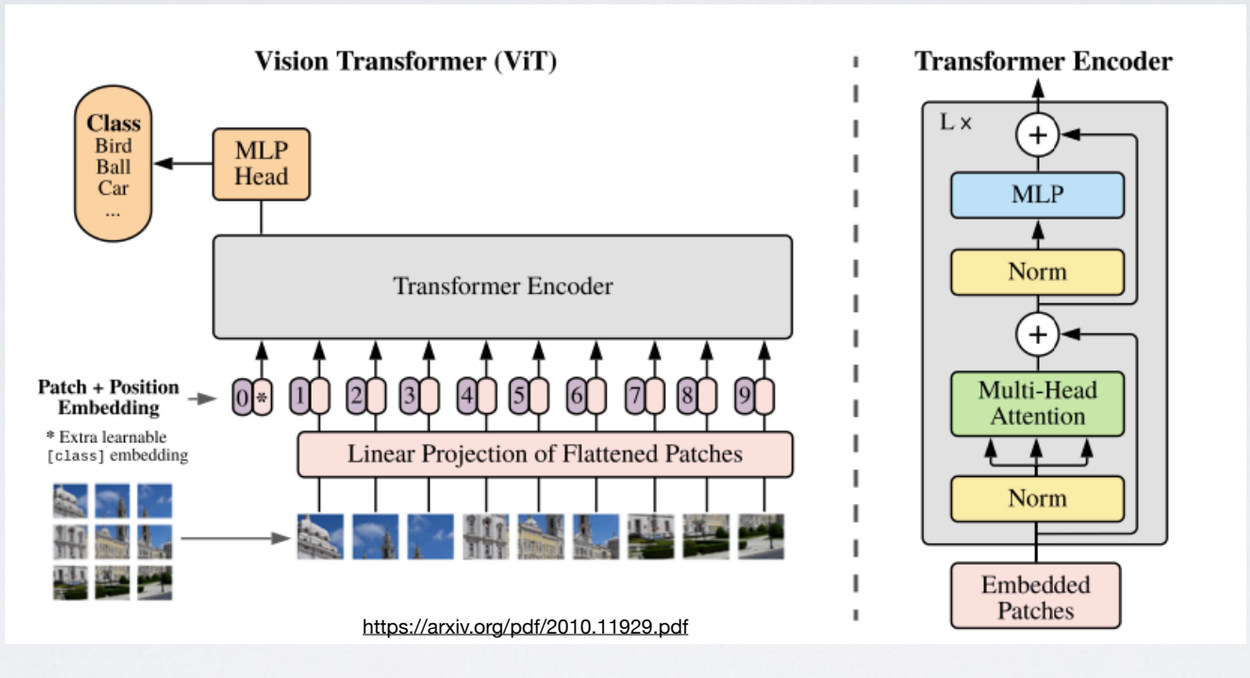
INTRODUCTION

Contexte

- **Classification d'images**
  - Prédominance des CNN depuis 2012
  - Regain d'intérêt Deep Learning
- **Natural Language Processing**
  - Arrivée des Transformers en 2017
    - ➡ Modèles de référence désormais
- **Transformers pour classification images ?**
  - Travaux Dosovitskiy *et al.* en Octobre 2020
    - ➡ Vision Transformers (ViT)

ViT

- **Architecture**



https://arxiv.org/pdf/2010.11929.pdf

- **Innovation**
  - Image découpée en patchs
    - ➡ Patch = « atome » de l'image et non plus le pixel
- **Meilleures performances que meilleurs CNN**
- **Modèles pré-entrainés disponibles**
  - ImageNet-21k, fine tuning sur ImageNet 1k
  - Compatibles Keras (vit-keras)

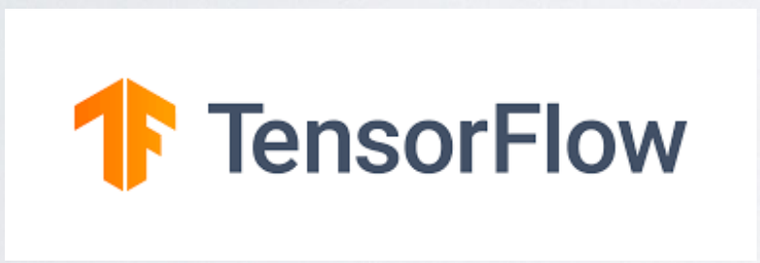
Objectif

- **Classification d'images**
  - Comparaison ViT / CNN
    - ➡ Transfer learning
    - ➡ Fine tuning
  - Images de style cartoon

MÉTHODOLOGIE

Outils utilisés

colab



Création des données

- **Récupération sur Google Image**

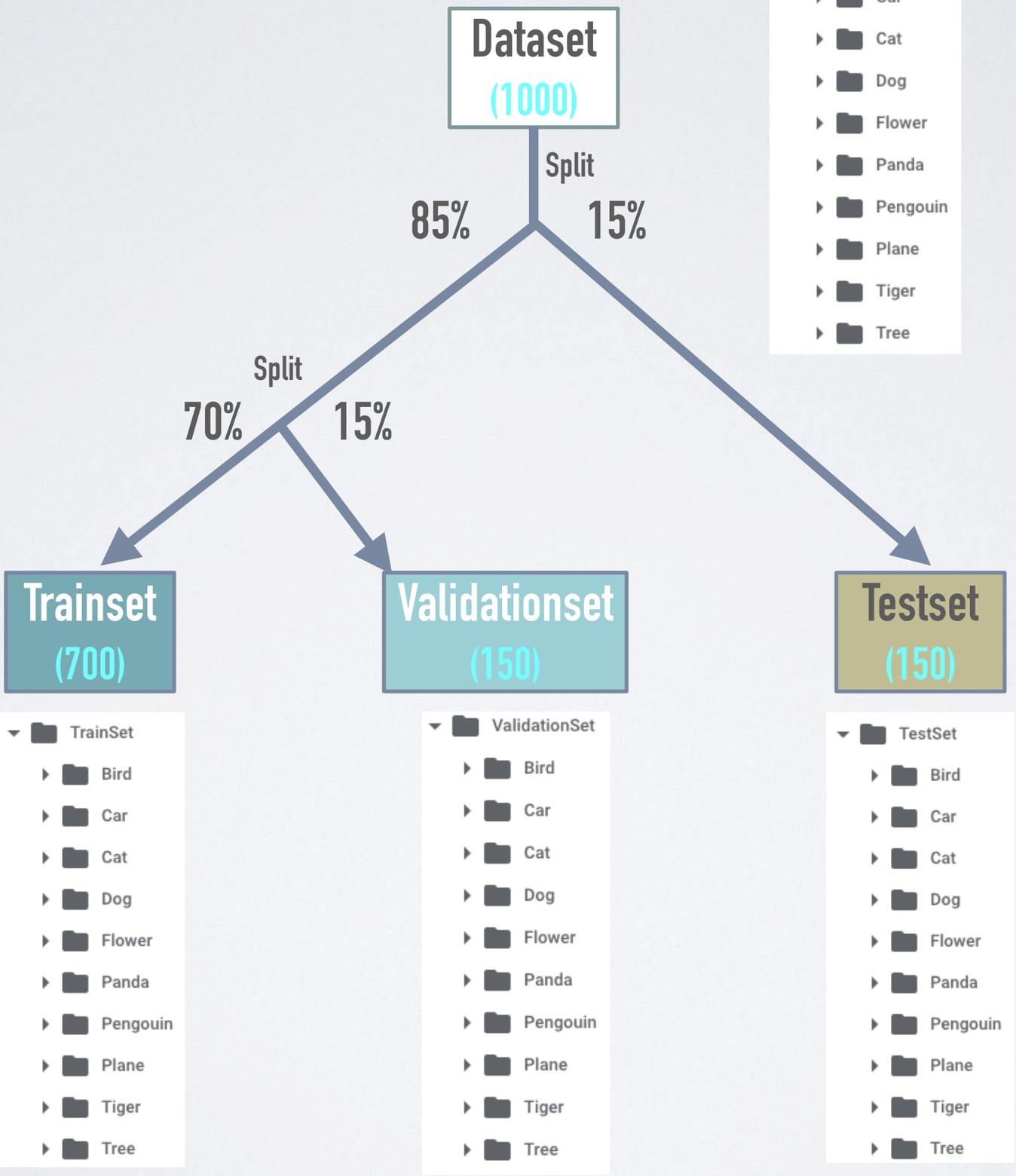
➡ 1000 images

➡ 10 classes

➡ 100 images par classe

- **211 images encodées RGBA**
  - Passage en RGB

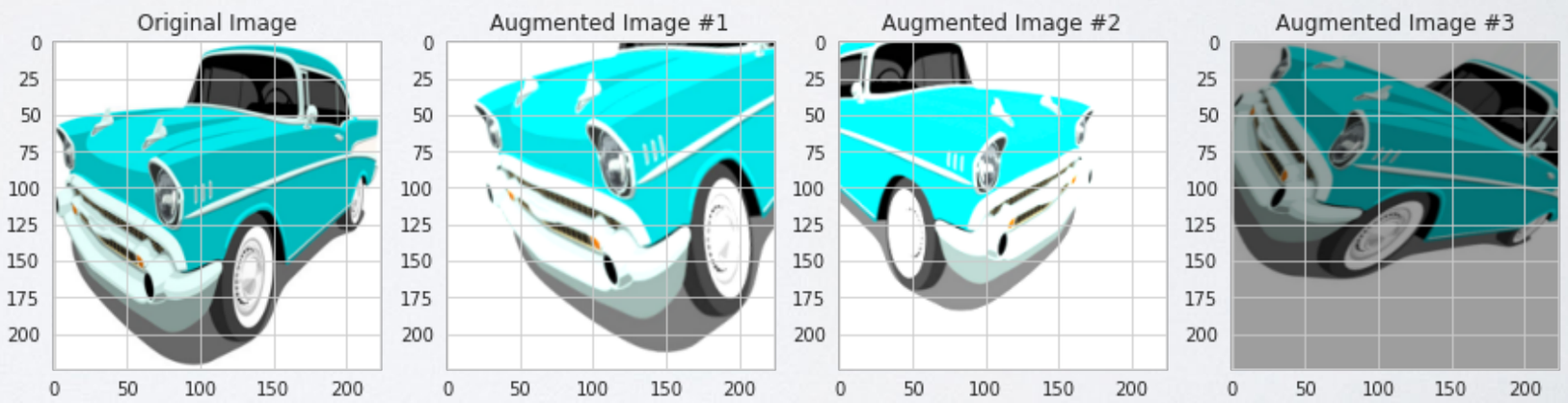
Split des données



Préprocessing et Data Augmentation

- **Préprocessing**
  - Redimensionnement images
    ➡ 224 x 224
  - Normalisation valeurs pixels
    ➡ -1 à 1

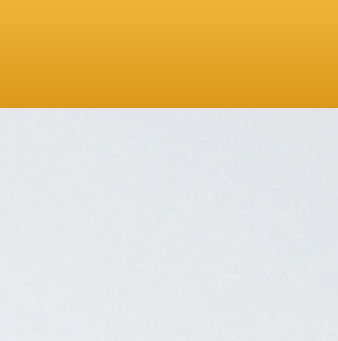
- **Data augmentation**
  - Effet miroir
  - Rotation
  - Décalage
  - Luminosité
  - Zoom

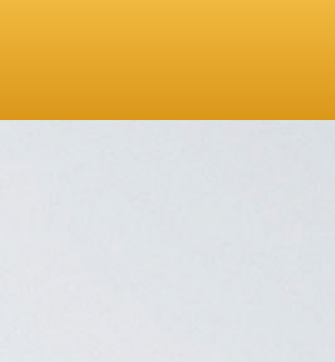


Modèles utilisés

- **Référence**
  - MobileNetV2
  - Dernières couches
  - ➡ GlobalAveragePooling2D
  - ➡ Classifieur 10 classes

- **Modèle testé**
  - Vit-b32
    - ➡ 12 couches de transformers
    - ➡ Taille patchs : 32





  - Dernière couche
    - ➡ Classifieur 10 classes

Entrainement des modèles

MobileNetV2

1
Transfer learning
Optimizer Adam 10^{-3}
50 epochs
Early stopping

ViT

(Fine) tuned MobileNetV2

2
Fine tuning
Optimizer Adam 10^{-5}
50 epochs
Early stopping

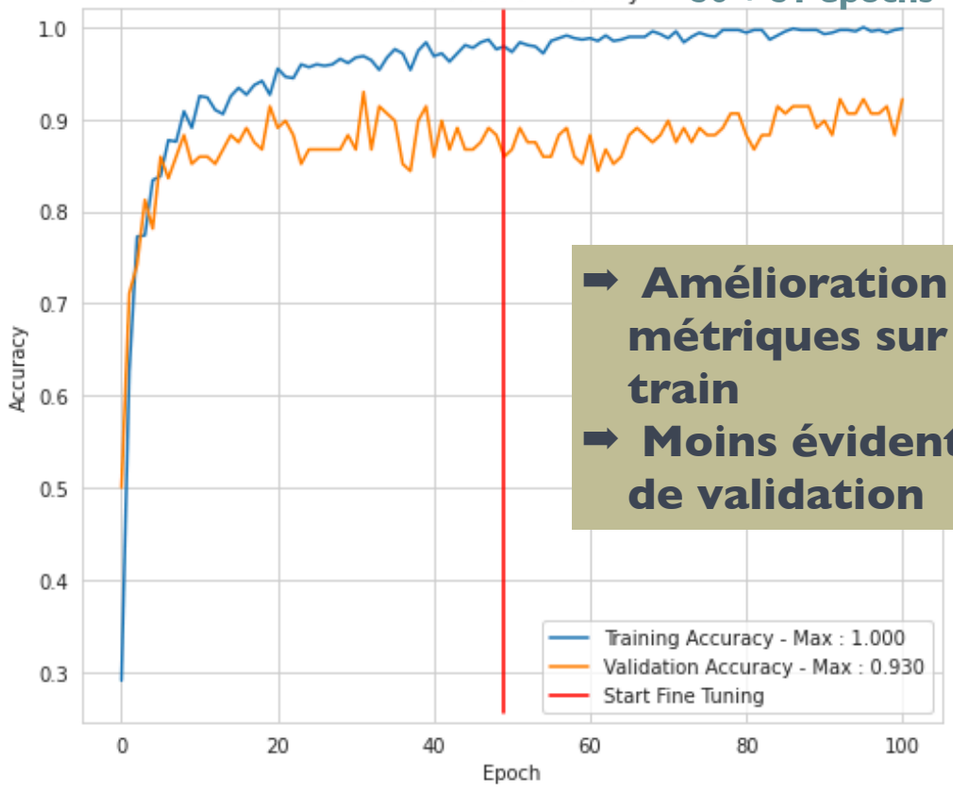
(Fine) tuned ViT

Accuracy
Loss

RÉSULTATS

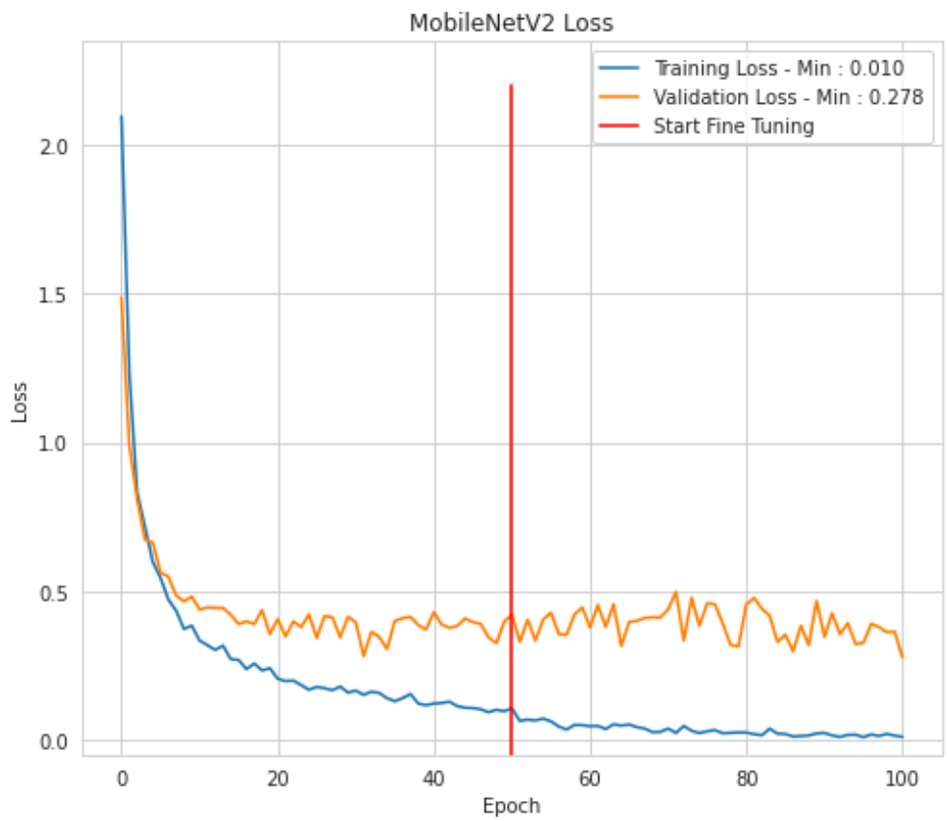
Learning Curves

MobileNetV2 Accuracy — 50 + 51 epochs — 28 + 29 min

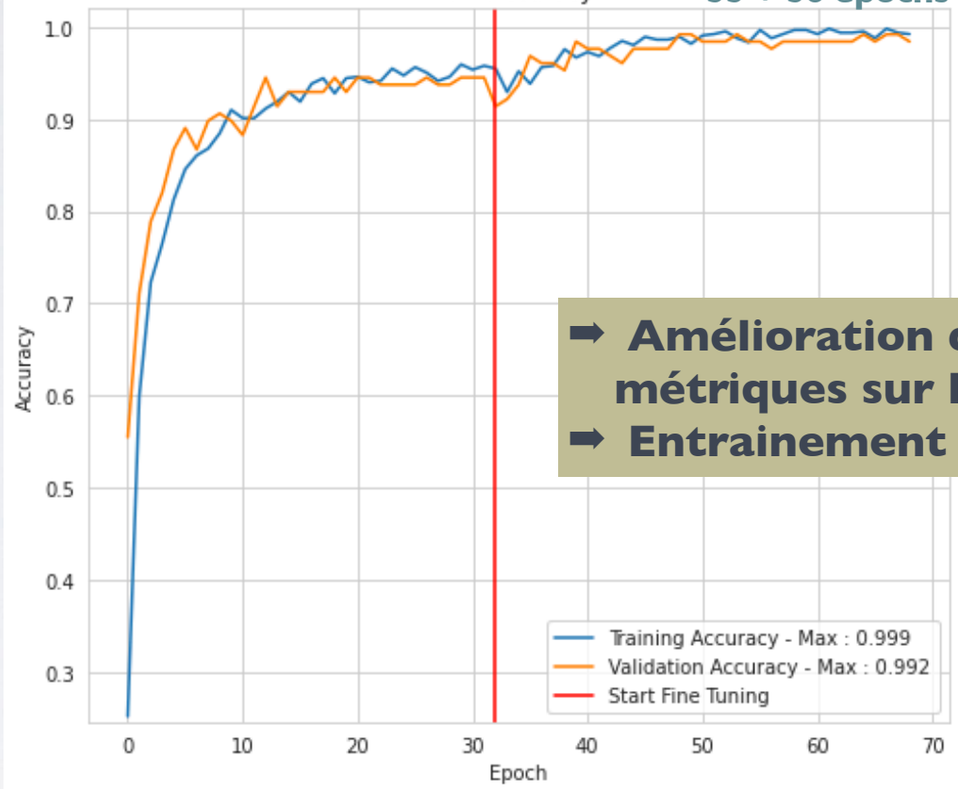


➡ **Amélioration des métriques sur jeu de train**

➡ **Moins évident sur jeu de validation**

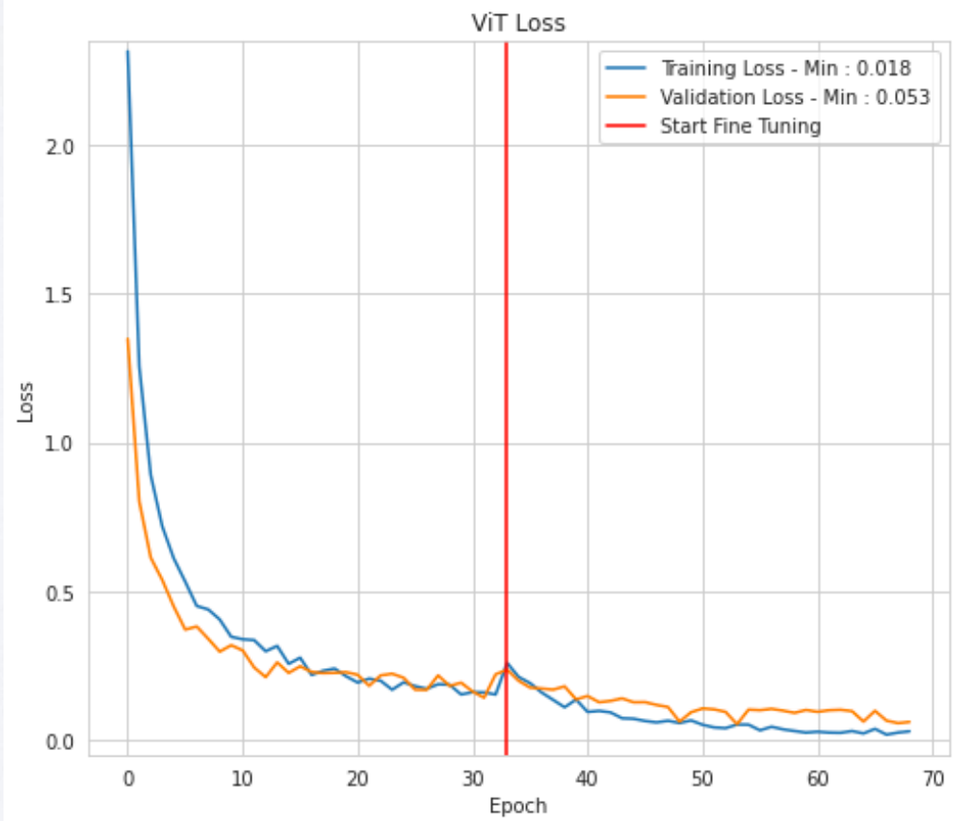


ViT Accuracy — 33 + 36 epochs — 26 + 23 min

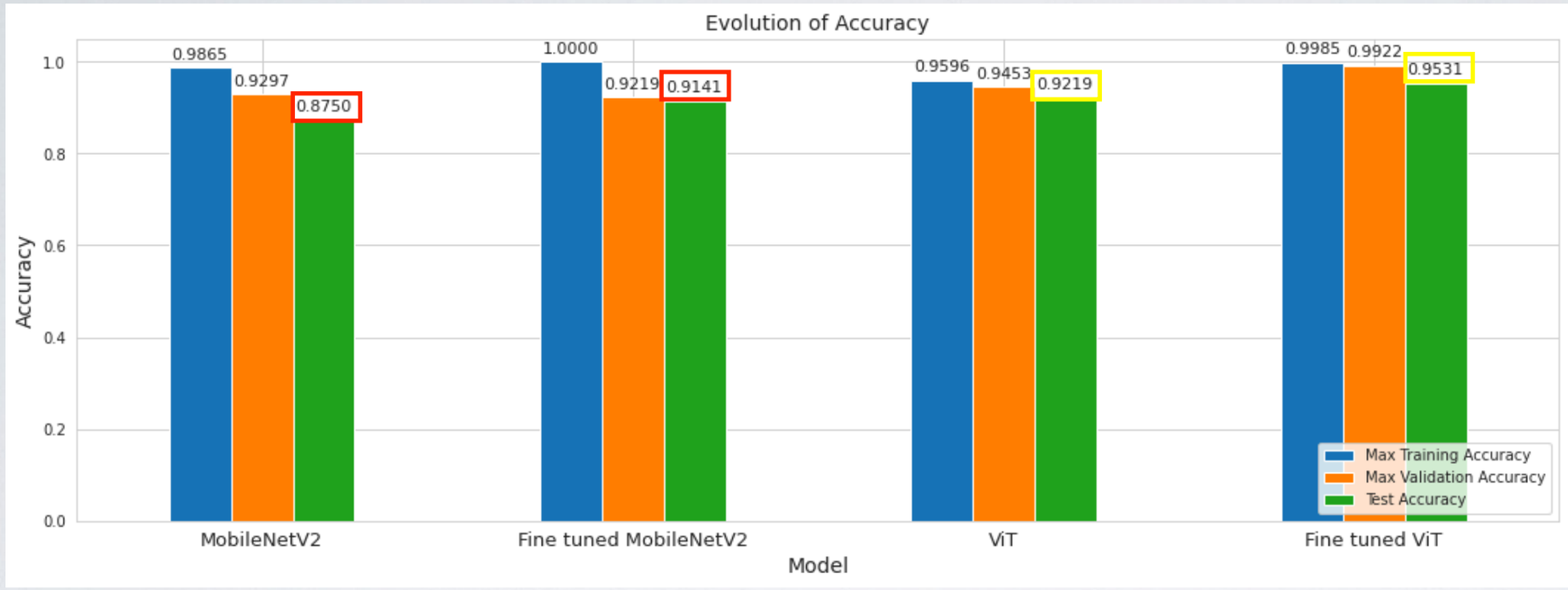


➡ **Amélioration des métriques sur les 2 jeux**

➡ **Entrainement plus lent**

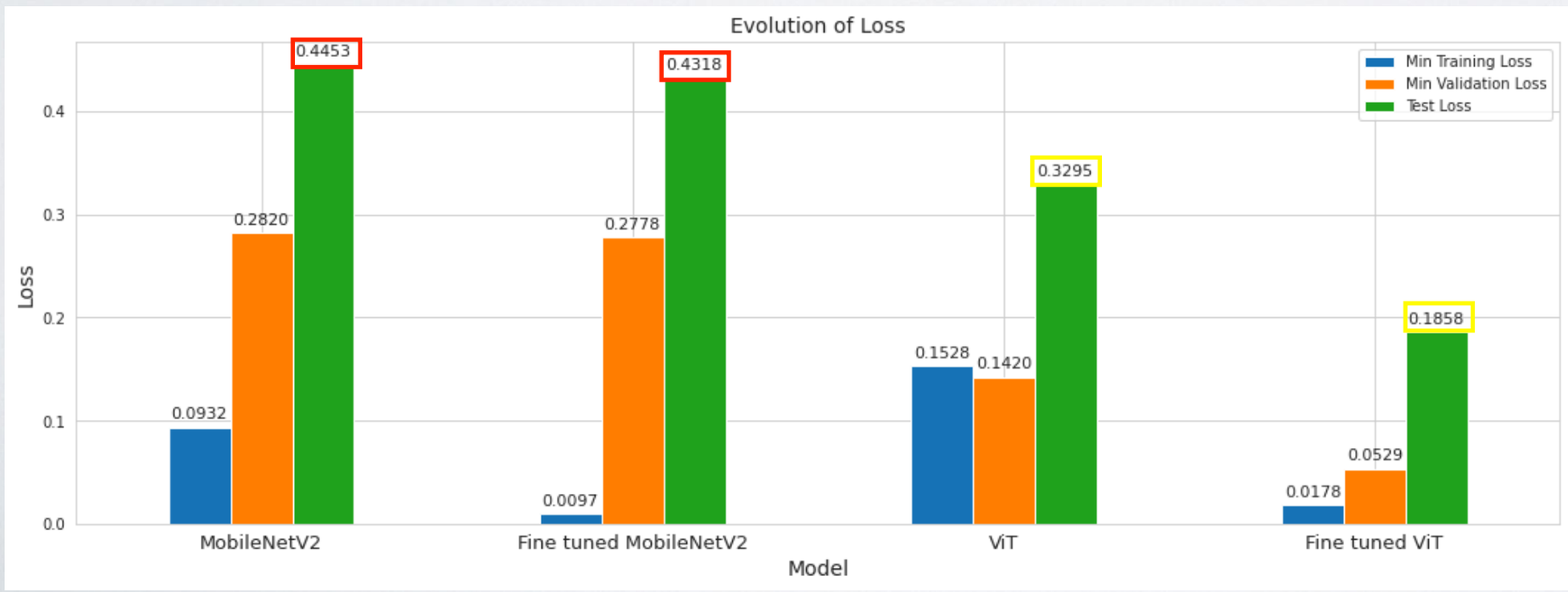


Accuracy et Loss



➡ **Amélioration avec fine tuning sur jeu de test**
➡ **Fine tuned ViT**

Taille	9.1Mo	26.4Mo	333.9Mo	1001.6Mo !

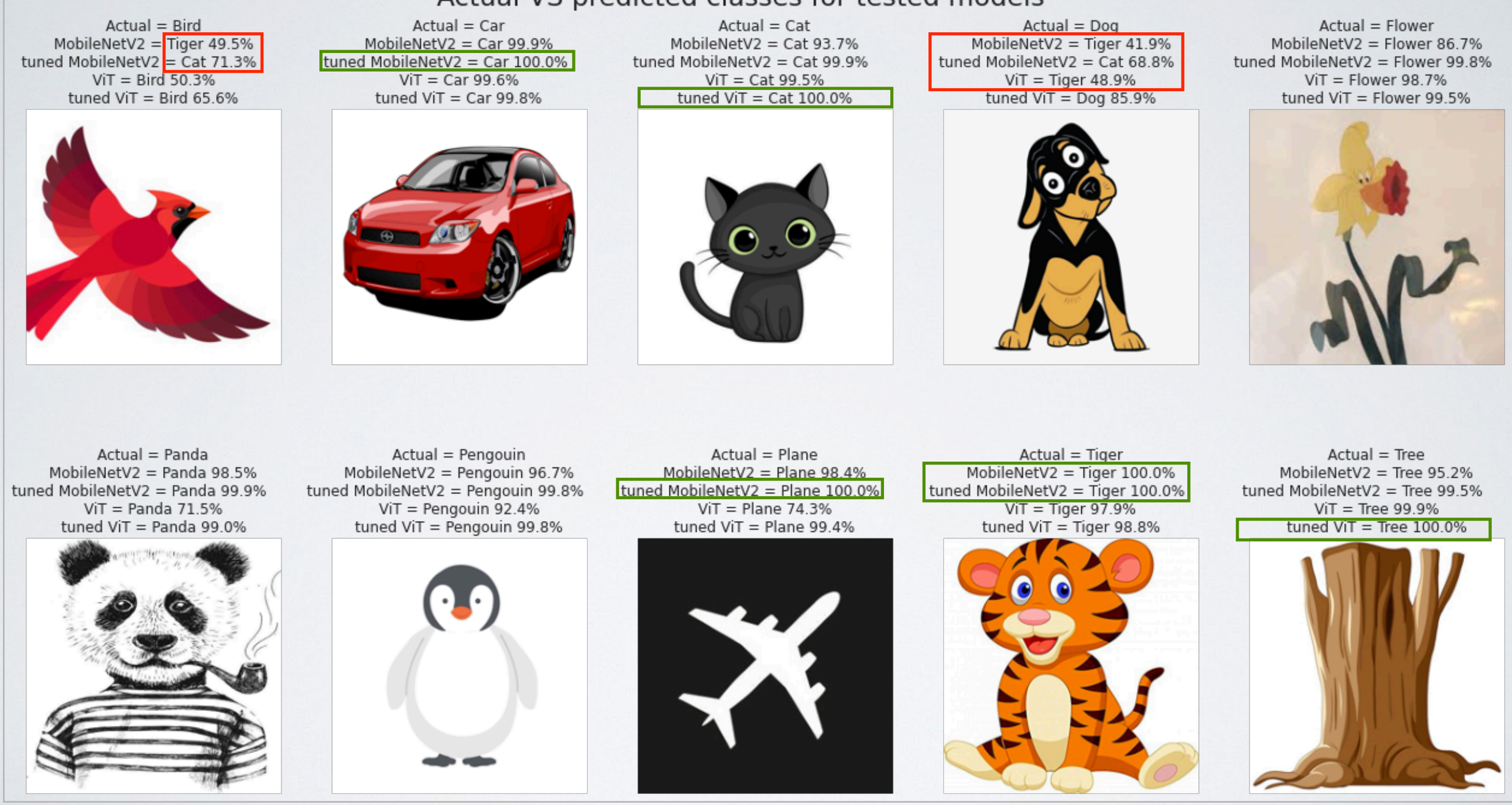


➡ **MobileNetV2 nettement plus léger**

➡ **Amélioration avec fine tuning sur jeu de test**
➡ **Fine tuned ViT**

Exemple de prédictions

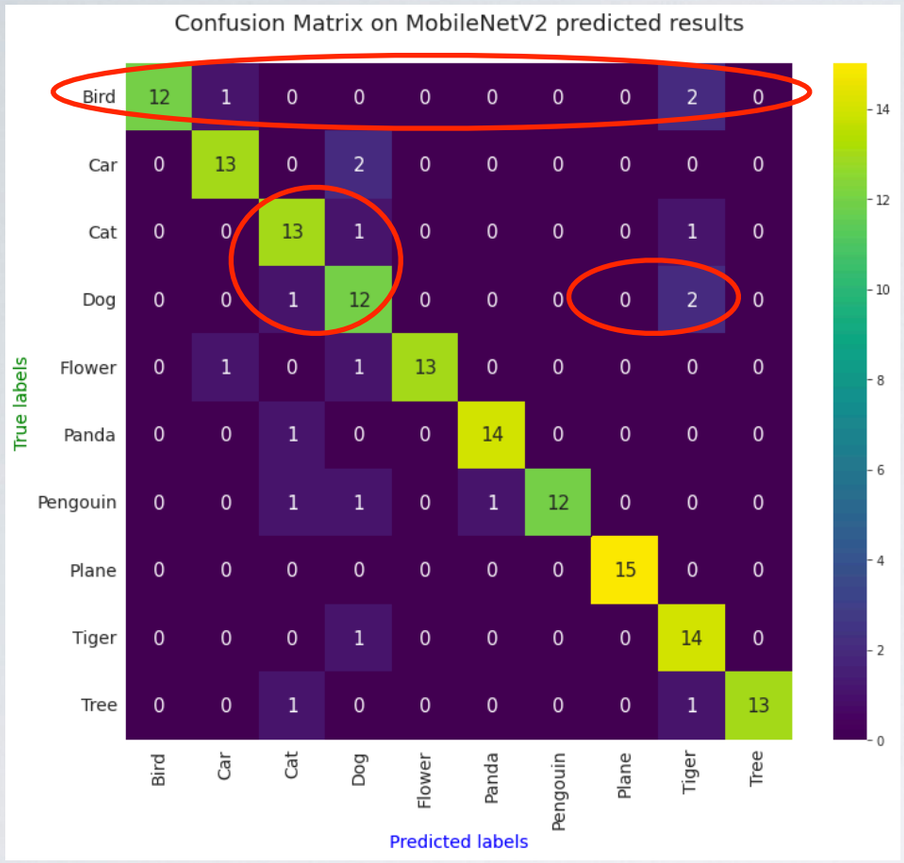
10 images au hasard

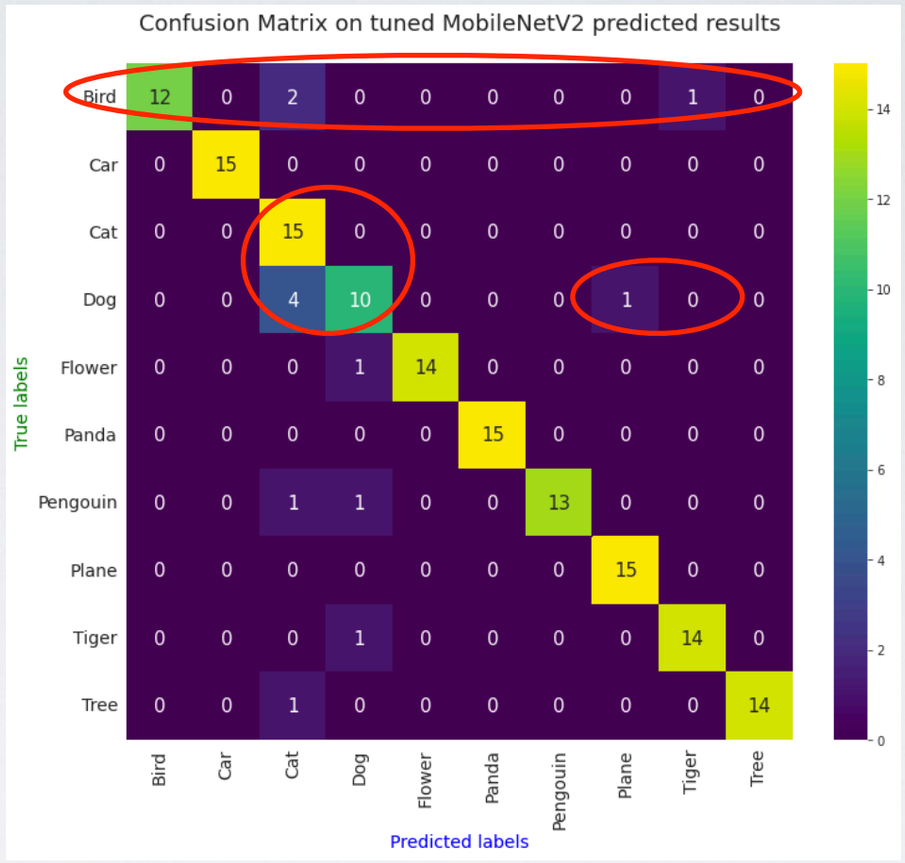


➡ **Quand classe prédite correcte, fine tuning améliore score**

➡ **Scores « perfect »**

Analyse des prédictions

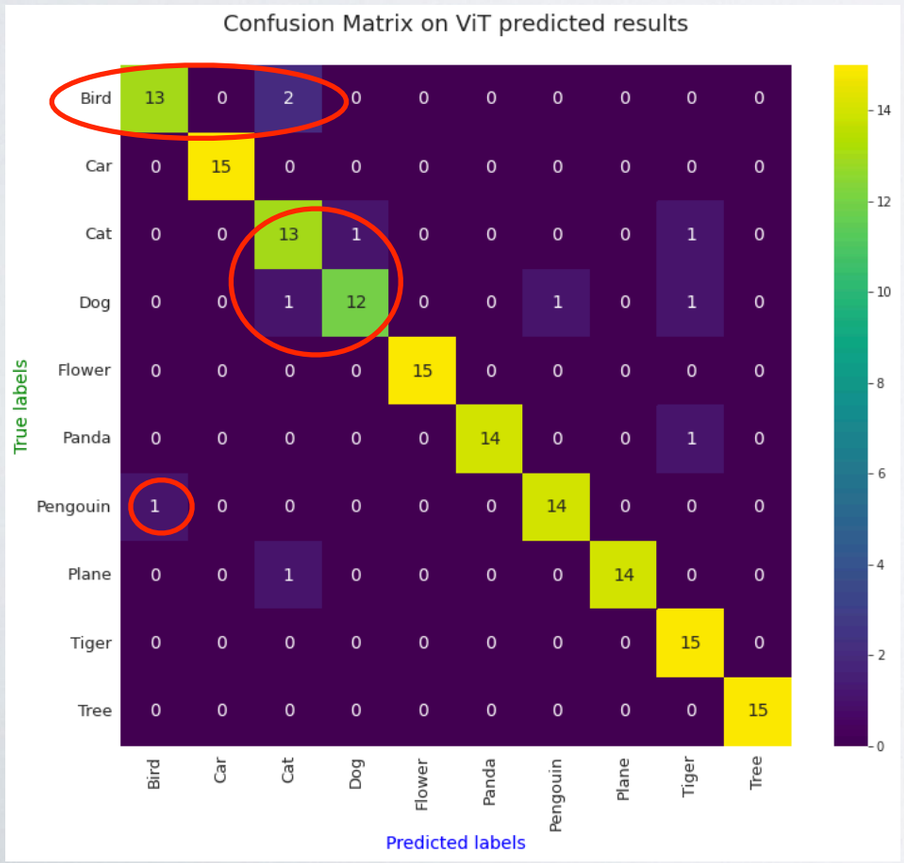


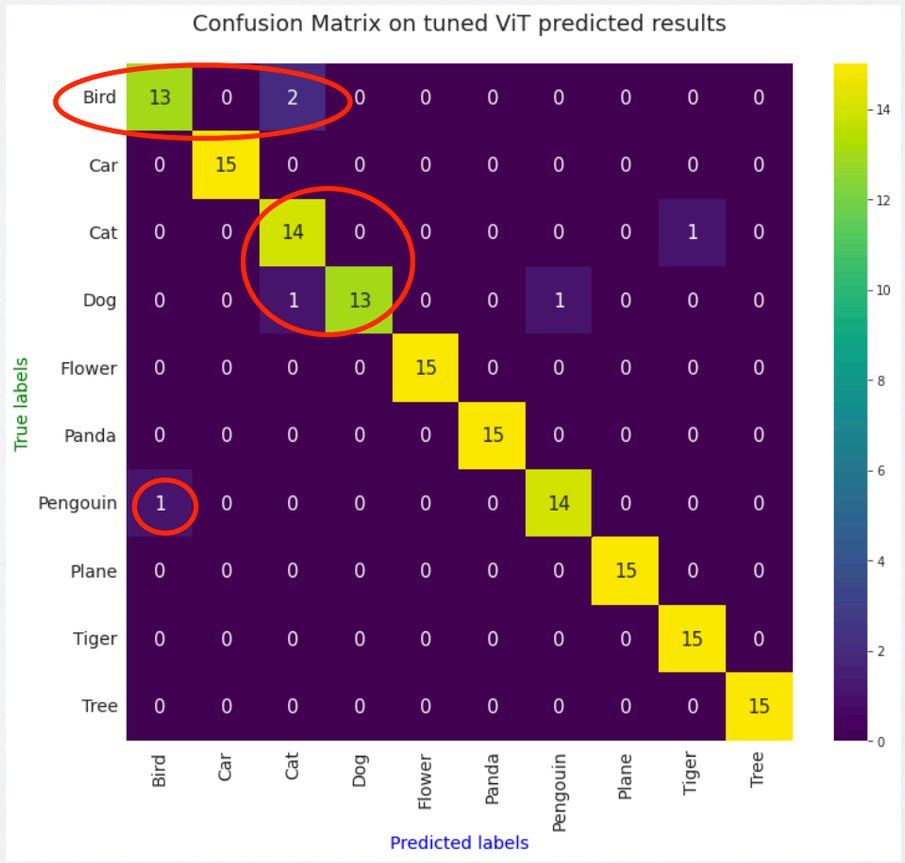


➡ **Amélioration de 6 valeurs sur diagonale**

➡ **Cas de mauvaises prédictions qui changent de classe**

➡ **Classes Dog et Bird plus difficiles**



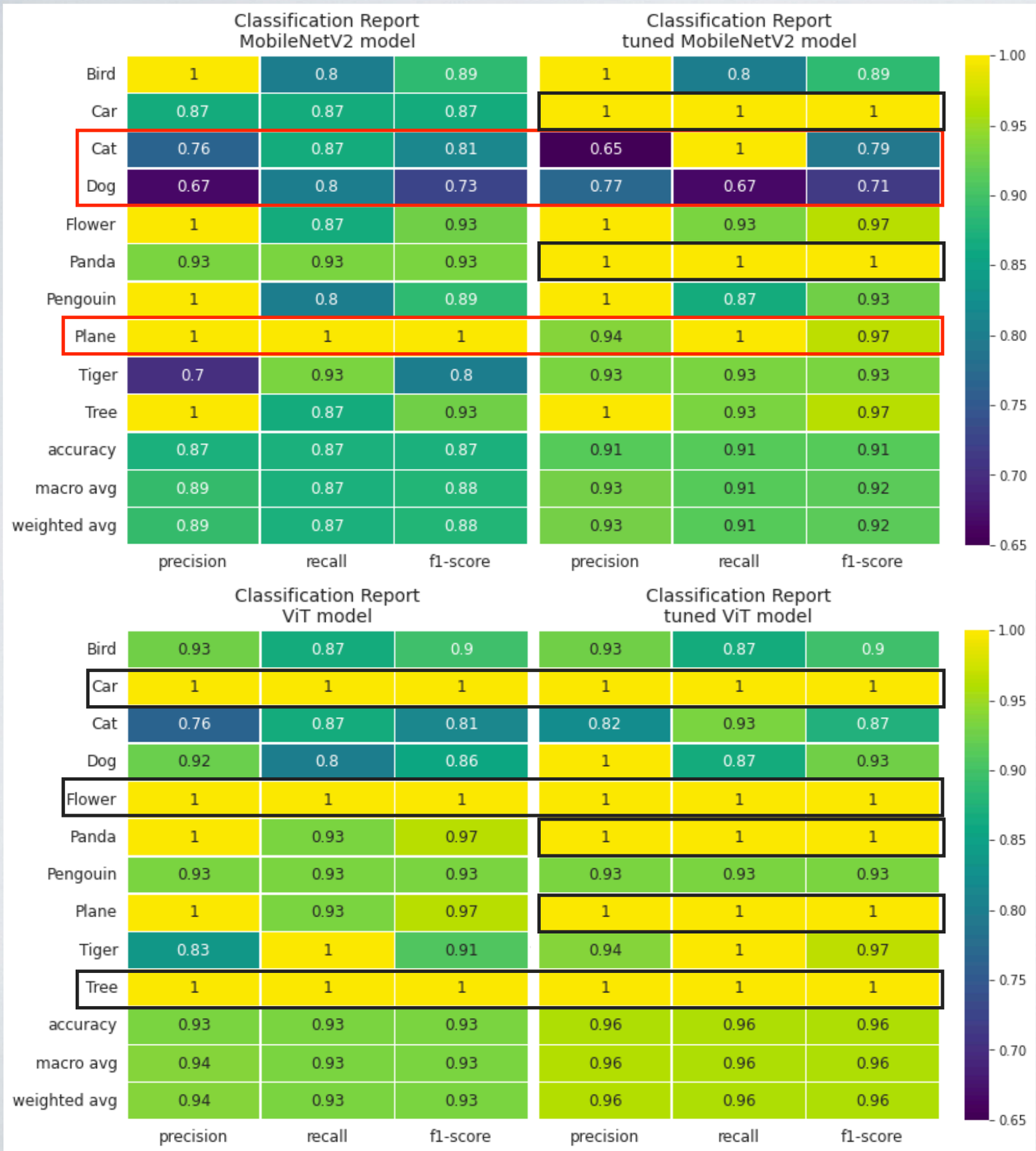


➡ **Amélioration de 4 valeurs sur diagonale**

➡ **Pas de mauvaises prédictions qui changent de classe**

➡ **Classes Dog et Bird plus difficiles**

Precision, recall et f1-score



➡ **Classe Cat recall amélioré au détriment precision**

➡ **Classe Dog precision améliorée au détriment recall**

➡ **Classe Plane perte de scores « perfect »**

➡ **Scores « perfect » passant de 10% à 20% des classes**

Intérêt du fine tuning
Fine tuned ViT
meilleur modèle

➡ **Scores équivalents ou améliorés**

➡ **Scores « perfect » passant de 30% à 50% des classes**

➡ **Meilleur apprentissage**

CONCLUSION

Conclusion

- **Classification d'images**
  - Comparaison ViT / CNN
    - ➡ Transfer learning
    - ➡ Fine tuning
  - Images de style cartoon
- **Meilleur modèle**
  - ViT après fine tuning
    - ➡ Métriques > 95%
    - ➡ Lourd néanmoins
- **Pistes d'amélioration**
  - Dataset plus conséquent
  - Autre modèle CNN
    - ➡ Meilleurs performances
    - ➡ Poids similaires

Objectif atteint

Perspectives

- **Réseaux hybrides CNN/ViT**
  - ResNet/ViT
  - MobileNet/ViT Metha *et al.* Oct 2021
    - ➡ Poids comparables à MobileNetV2
    - ➡ Meilleures performances
    - ➡ Plus lent néanmoins
- **Développements et améliorations permanents**

MERCI

QUESTIONS