Projet 7Preuve de concept

Classification d'images de style cartoon avec un Vision Transformer (ViT) Comparaison avec un réseau de neurones convolutionnels (CNN)

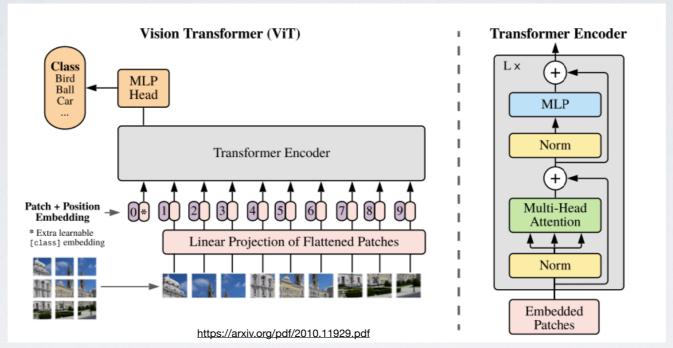
INTRODUCTION

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



· 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

















Création des données

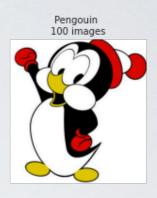
· Récupération sur Google Image

- → 1000 images
- → 10 classes
- → 100 images par classe















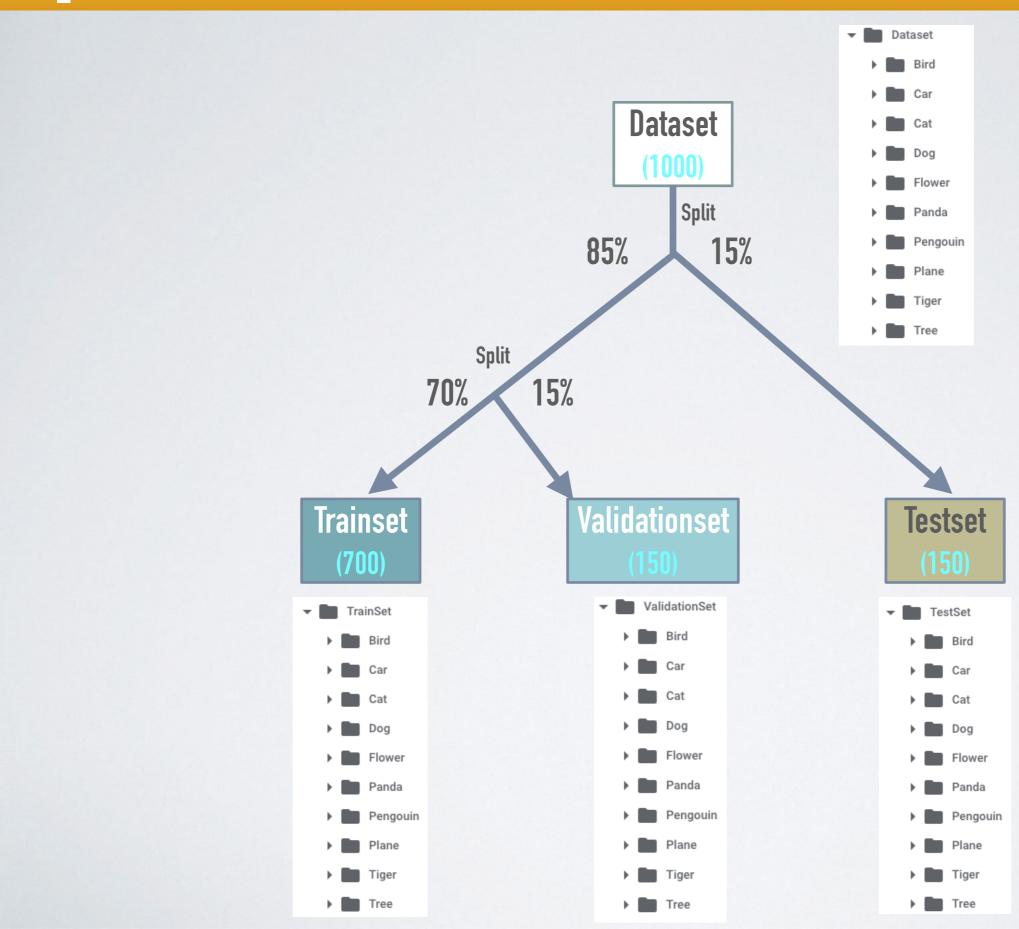






- · 211 images encodées RGBA
 - Passage en RGB

Split des données



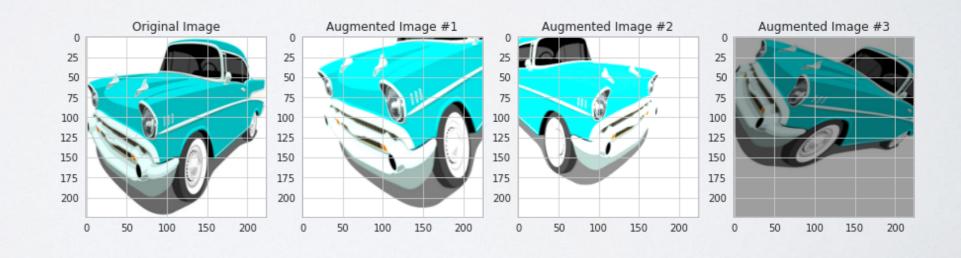
Préprocessing et Data Augmentation

· Préprocessing

- Redimensionnement images
 - → 224 × 224
- Normalisation valeurs pixels
 - → | à |

Data augmentation

- Effet miroir
- Rotation
- Décalage
- Luminosité
- Zoom



Modèles utilisés

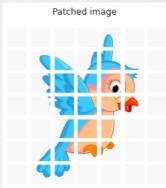
· Référence

- 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

Transfer learning
Optimizer Adam 10-3
50 epochs
Early stopping

ViT

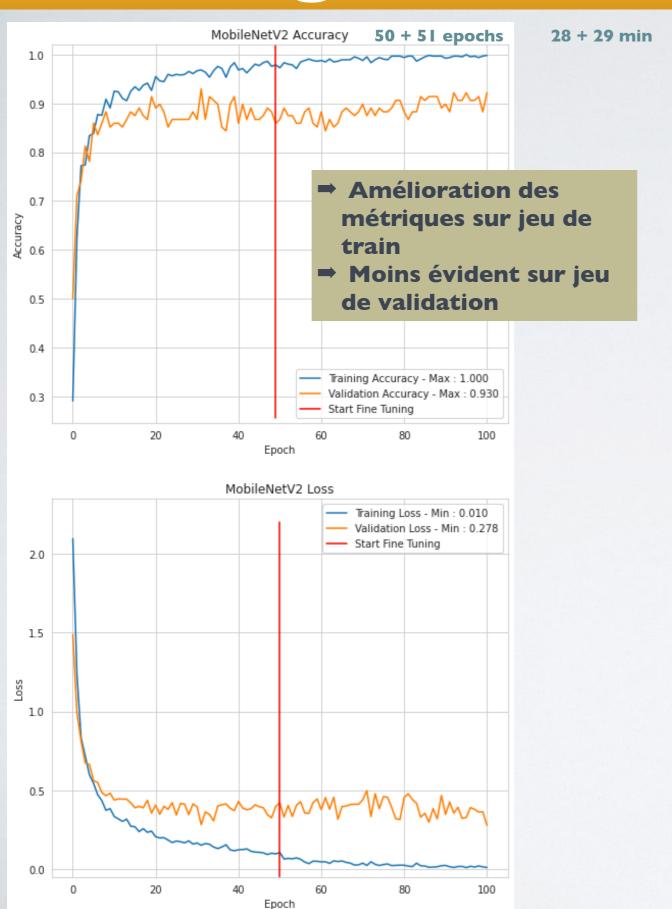
(Fine) tuned MobileNetV2 Fine tuning
Optimizer Adam 10⁻⁵
50 epochs
Early stopping

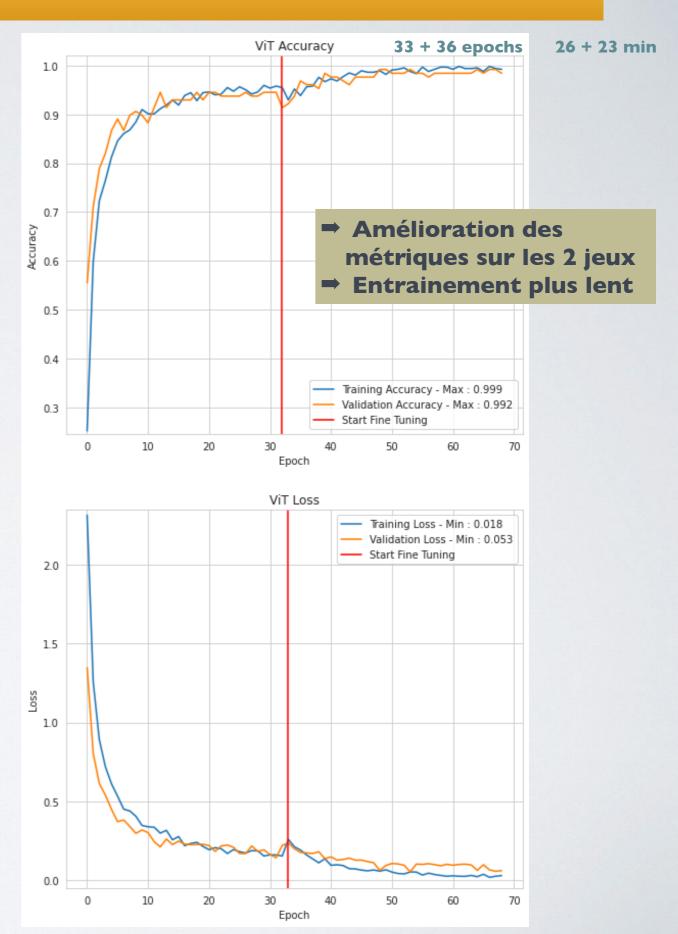
(Fine) tuned ViT

Accuracy Loss

RÉSULTATS

Learning Curves





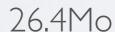
Accuracy et Loss



→ Amélioration avec fine tuning sur jeu de test

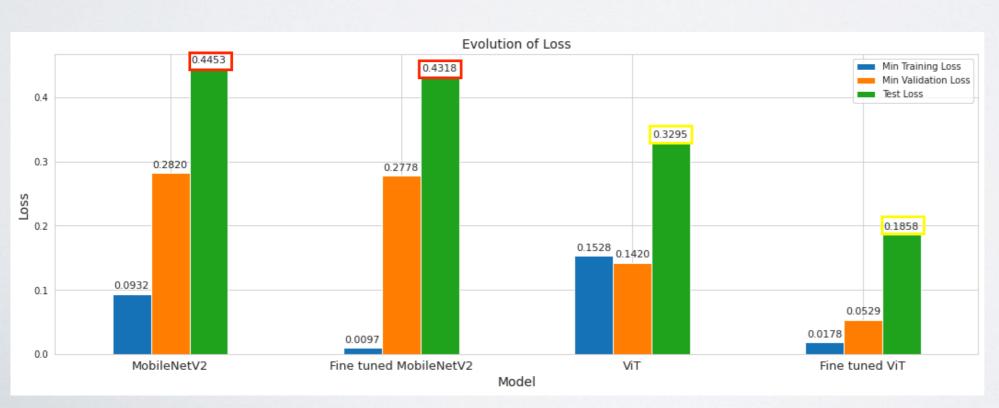
→ Fine tuned ViT





333.9Mo

1001.6Mo!



- → MobileNetV2 nettement plus léger
- → Amélioration
 avec fine tuning
 sur jeu de test
 → Fine tuned ViT

Exemple de prédictions

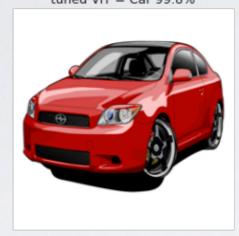
10 images au hasard

Actual VS predicted classes for tested models

Actual = Bird MobileNetV2 = Tiger 49.5% tuned MobileNetV2 = Cat 71.3% ViT = Bird 50.3% tuned ViT = Bird 65.6%



Actual = Car MobileNetV2 = Car 99.9% tuned MobileNetV2 = Car 100.0% ViT = Car 99.6% tuned ViT = Car 99.8%



Actual = Cat MobileNetV2 = Cat 93.7% tuned MobileNetV2 = Cat 99.9% ViT = Cat 99.5% tuned ViT = Cat 100.0%



Actual = Dog MobileNetV2 = Tiger 41.9% tuned MobileNetV2 = Cat 68.8% ViT = Tiger 48.9% tuned ViT = Dog 85.9%



Actual = Flower MobileNetV2 = Flower 86.7% tuned MobileNetV2 = Flower 99.8% ViT = Flower 98.7% tuned ViT = Flower 99.5%



Actual = Panda MobileNetV2 = Panda 98.5% tuned MobileNetV2 = Panda 99.9% ViT = Panda 71.5% tuned ViT = Panda 99.0%



Actual = Pengouin MobileNetV2 = Pengouin 96.7% tuned MobileNetV2 = Pengouin 99.8% ViT = Pengouin 92.4% tuned ViT = Pengouin 99.8%



Actual = Plane MobileNetV2 = Plane 98.4% tuned MobileNetV2 = Plane 100.0% ViT = Plane 74.3% tuned ViT = Plane 99.4%



Actual = Tiger MobileNetV2 = Tiger 100.0% tuned MobileNetV2 = Tiger 100.0% ViT = Tiger 97.9% tuned ViT = Tiger 98.8%

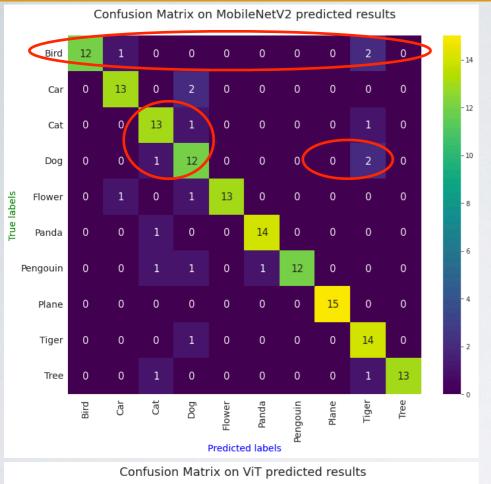


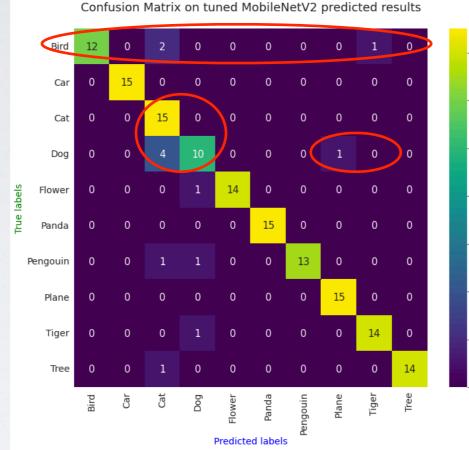
Actual = Tree MobileNetV2 = Tree 95.2% tuned MobileNetV2 = Tree 99.5% ViT = Tree 99.9% tuned ViT = Tree 100.0%



- **→** Quand classe prédite correcte, fine tuning améliore score
- **⇒** Scores « perfect »

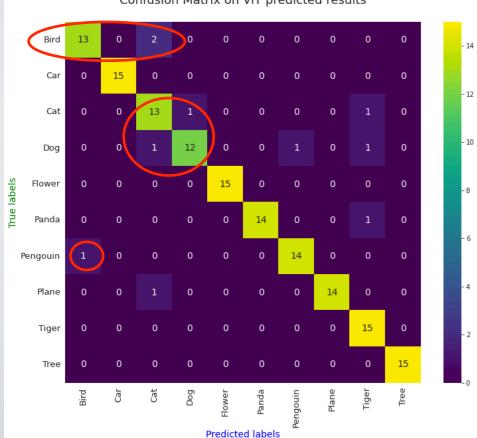
Analyse des prédictions

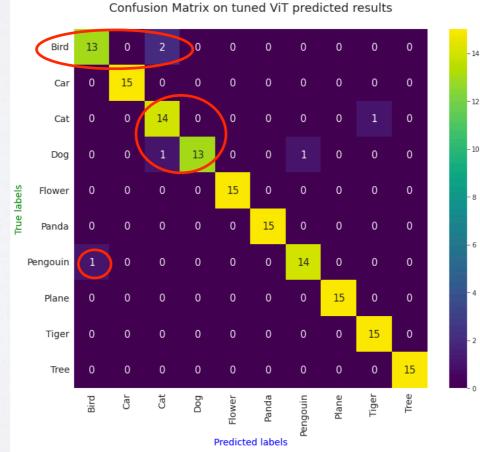






- Cas de mauvaises prédictions qui changent de classe
- → Classes Dog et Bird plus difficiles





- Amélioration de 4 valeurs sur diagonale
- → Pas de mauvaises prédictions qui changent de classe
- → Classes Dog et Bird plus difficiles

Precision, recall et fl-score

Classification Report MobileNetV2 model				Classification Report tuned MobileNetV2 model			1.00
Biro	d 1	8.0	0.89	1	0.8	0.89	-1.00
Ca	r 0.87	0.87	0.87	1	1	1	- 0.95
Ca	t 0.76	0.87	0.81	0.65	1	0.79	0.55
Dog	g 0.67	0.8	0.73	0.77	0.67	0.71	- 0.90
Flowe	r 1	0.87	0.93	1	0.93	0.97	
Panda	a 0.93	0.93	0.93	1	1	1	- 0.85
Pengouii	n 1	0.8	0.89	1	0.87	0.93	
Plane	e 1	1	1	0.94	1	0.97	- 0.80
Tige	er 0.7	0.93	0.8	0.93	0.93	0.93	
Tre	e 1	0.87	0.93	1	0.93	0.97	- 0.75
accurac	y 0.87	0.87	0.87	0.91	0.91	0.91	
macro av	g 0.89	0.87	0.88	0.93	0.91	0.92	- 0.70
weighted av	g 0.89	0.87	0.88	0.93	0.91	0.92	0.55
	precision	recall	f1-score	precision	recall	f1-score	- 0.65
Classification Report ViT model				Classification Report tuned ViT model			-1.00
Bire	d 0.93	0.87	0.9	0.93	0.87	0.9	
Ca	ar 1	1	1	1	1	1	- 0.95
Ca	ot 0.76	0.87	0.81	0.82	0.93	0.87	
Dog	g 0.92	0.8	0.86	1	0.87	0.93	- 0.90
Flowe	r 1	1	1	1	1	1	
Panda	a 1	0.93	0.97	1	1	1	- 0.85
Pengoui	n 0.93	0.93	0.93	0.93	0.93	0.93	
Plane	e 1	0.93	0.97	1	1	1	- 0.80
Tige	er 0.83	1	0.91	0.94	1	0.97	
Tre	e 1	1	1	1	1	1	- 0.75
accurac	y 0.93	0.93	0.93	0.96	0.96	0.96	
macro av	g 0.94	0.93	0.93	0.96	0.96	0.96	- 0.70
weighted av	g 0.94	0.93	0.93	0.96	0.96	0.96	0.55
	precision	recall	f1-score	precision	recall	f1-score	- 0.65

- → 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