
Projet long : Transfer learning using VGG16

— Youness AKOURIM —

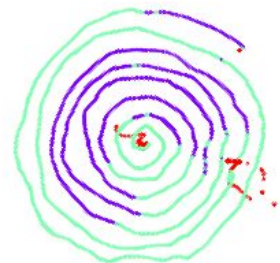
Dataset

75 images de dessins (handwriting Alzheimer task)
composées en deux parties : 45 control et 30 AD.

Types des images:

- Raw : (image brute)
- Penups Raw
- Velocity
- Altitude
- Pressure

Problème : manque de données pour les classifier



Pressure control



Alz control

Data augmentation

- pour chaque image : on génère une série de rotation de pas de 10° afin d'avoir 15 exemples pour alz et de pas de 15° afin d'avoir 10 exemples pour control

- En total :

#Alz = 450 exemples = $30 * 15$

#Control = 450 exemples = $45 * 10$

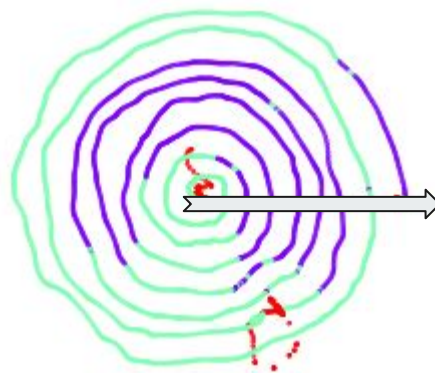


image initiale

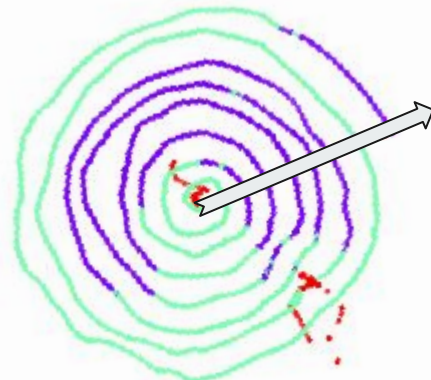
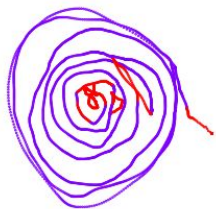


image tournée de
 15°

Visualisation des features MAP

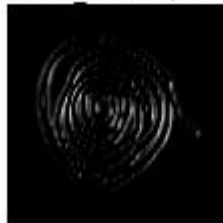


A_0009_0.png

block1 conv2 (41)



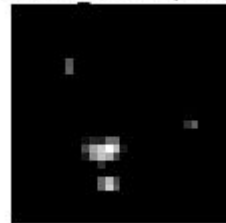
block2 conv2 (111)



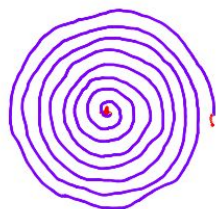
block3 conv3 (51)



block4 conv3 (401)



block5 conv3 (507)



C_0009_0.png

block1 conv2 (41)



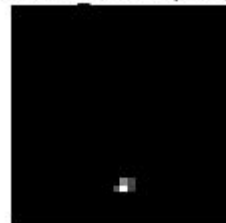
block2 conv2 (111)



block3 conv3 (51)



block4 conv3 (401)



block5 conv3 (507)



Visualisation des features MAP



C_0009_0.png

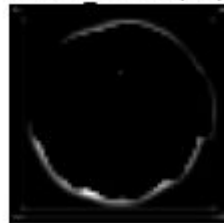
block1 conv2 (41)



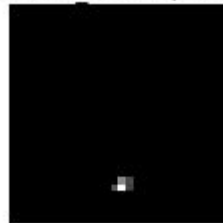
block2 conv2 (111)



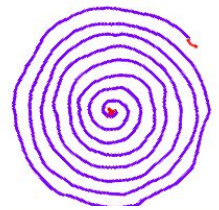
block3 conv3 (51)



block4 conv3 (401)



block5 conv3 (507)



C_0009_3.png

block1 conv2 (41)



block2 conv2 (111)



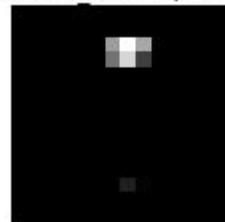
block3 conv3 (51)



block4 conv3 (401)

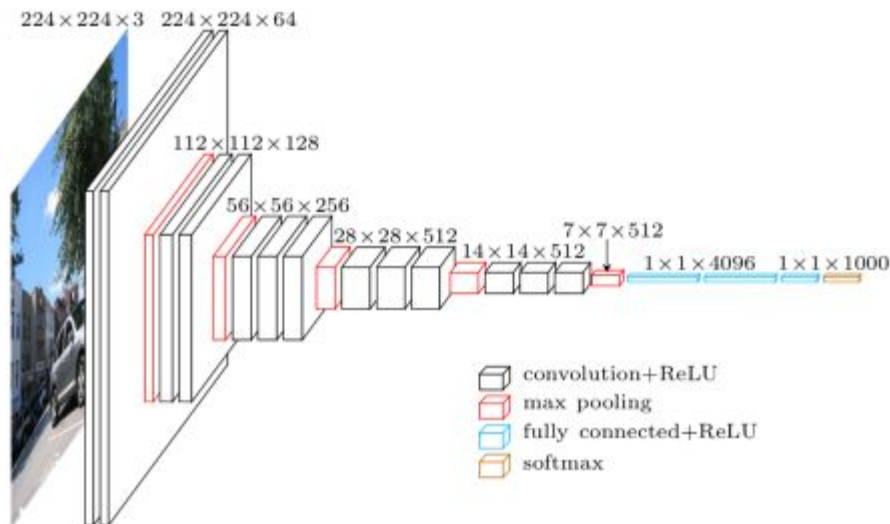


block5 conv3 (507)



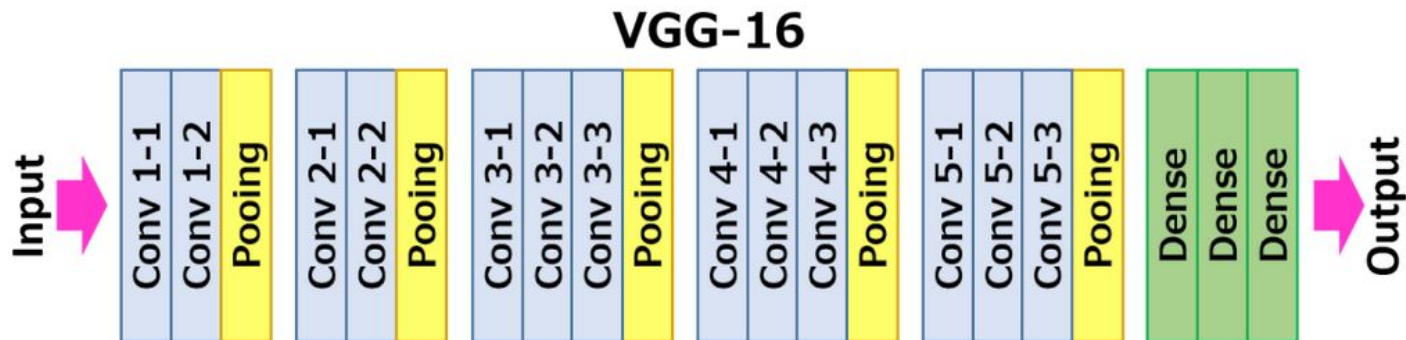
Présentation de VGG16

- A model proposed by K. Simonyan and A. Zisserman from the University of Oxford
- It achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes.



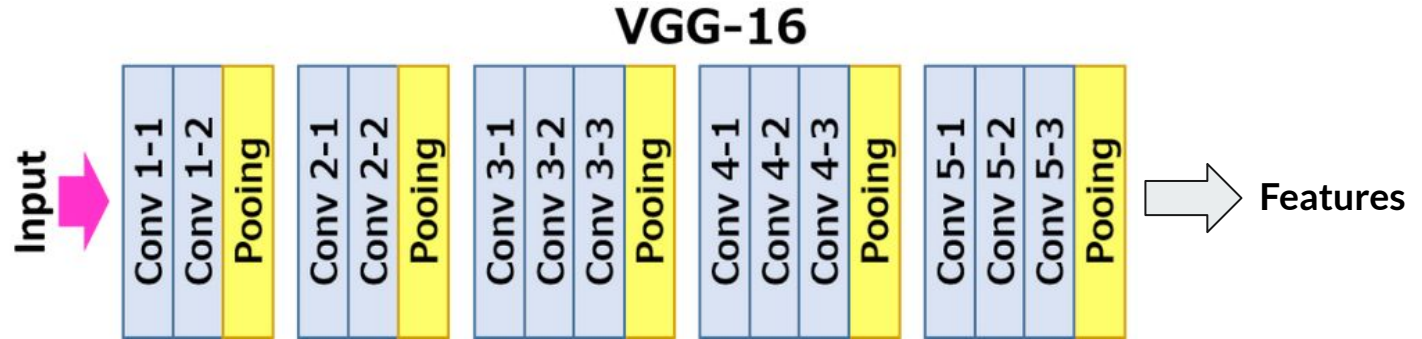
<https://towardsdatascience.com/step-by-step-vgg16-implementation-in-keras-for-beginners-a833c686ae6c>

VGG16 : Zoom In



- 5 blocs de couches de convolutions
- Input Shape : (224,224,3)
- Output shape : (1, 1000)

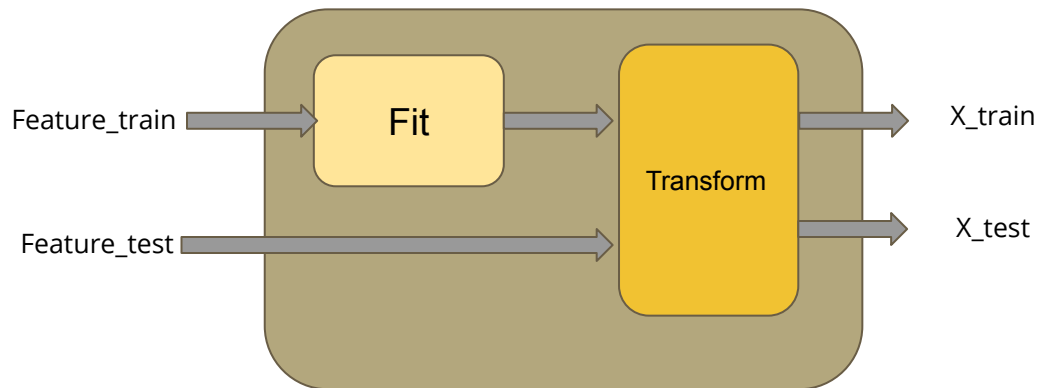
Transfert Learning



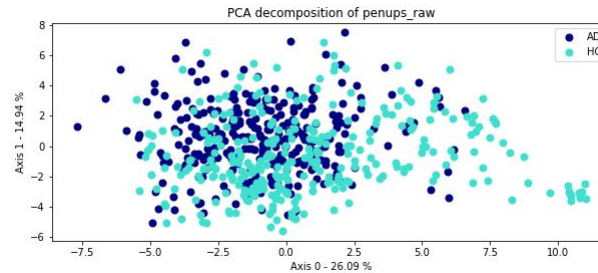
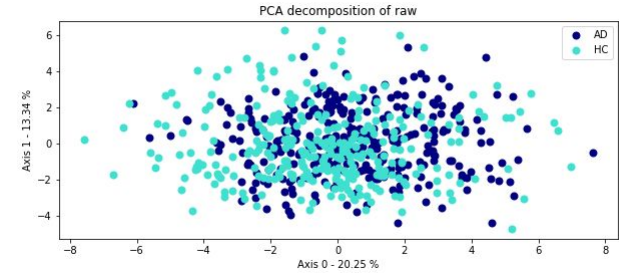
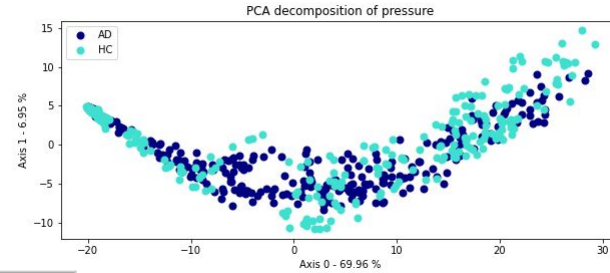
- Extraire les features grâce à un transfert learning
- Taille de l'image de l'entrée : (224, 224, 3)
- Les images sont normalisées (image / 225 pour chaque couche).
- Taille des features : (1, 25088)

Décomposition ACP (1/3)

- On partage les features obtenus après l'étape transfer learning en :
 - feature train qui représente $\frac{2}{3}$ du dataset
 - feature test qui représente $\frac{1}{3}$ du dataset
- Les données sont équilibrées par classe pour feature train et feature test

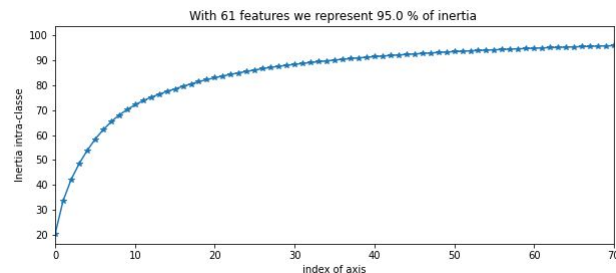
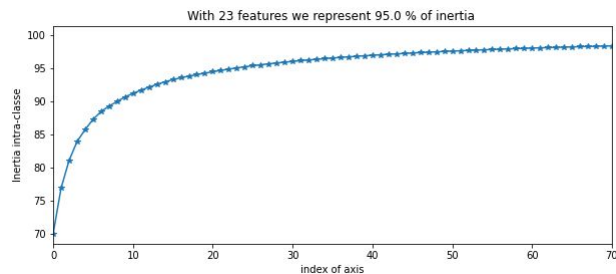
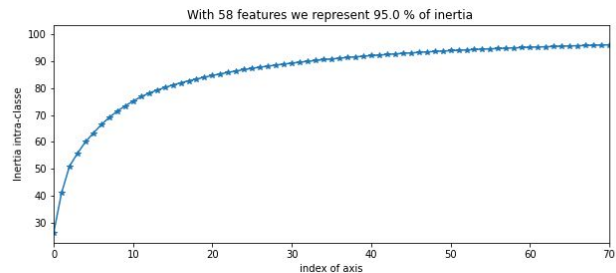


Décomposition ACP (2/3)



	% d'inertie portée par le plan factoriel
Raw	33,59
Penups Raw	41,03
Pressure	76,91

Décomposition ACP (3/3)



	nombre d'axes pour 95% d'info
Raw	61
Penups Raw	58
Pressure	23

- Pressure : une bonne décomposition !

comparaison des classifieurs

- Les classifieurs :
 - SVM
 - MLP
 - KNN
- Entraînement 10-Kfold CV
- Métriques : sur l'ensemble de test

raw :

- SVM est le meilleur estimateur

	SVM	MLP	KNN
Accuracy	0.77 \pm 0.13	0.75 \pm 0.11	0.74 \pm 0.14
sensibility	0.77 \pm 0.15	0.76 \pm 0.12	0.75 \pm 0.22
specificity	0.78 \pm 0.15	0.74 \pm 0.16	0.74 \pm 0.15
AUC	0.85 \pm 0.09	0.84 \pm 0.11	0.85 \pm 0.13
parameters	C: 100.0, gamma: 0.01, kernel: 'rbf'	'alpha': 0.1, 'hidden_layer': 14, 'max_iter': 15000, 'solver': 'lbfgs'	'algorithm': 'ball_tree', 'n_neighbors': 3, 'weights': 'distance'

comparaison des classifieurs

Penups raw :

- SVM est le meilleur estimateur
- KNN est le modèle le plus spécifié

	SVM	MLP	KNN
Accuracy	0.84 \pm 0.087	0.82 \pm 0.07	0.81 \pm 0.13
sensibility	0.86 \pm 0.12	0.83 \pm 0.09	0.74 \pm 0.25
specificity	0.82 \pm 0.15	0.82 \pm 0.11	0.88 \pm 0.10
AUC	0.92 \pm 0.07	0.90 \pm 0.07	0.92 \pm 0.09
parameters	C: 100.0, gamma: 0.01, kernel: 'rbf'	'alpha': 0.1, 'hidden_layer': 12, 'max_iter': 5000, 'solver': 'lbfgs'	'algorithm': 'ball_tree', 'n_neighbors': 3, 'weights': 'distance'

comparaison des classifieurs

pressure:

- KNN est le meilleur estimateur
- SVM est le modèle le plus sensible
- KNN est le modèle le plus spécifié

	SVM	MLP	KNN
Accuracy	0.89 \pm 0.06	0.89 \pm 0.11	0.91 \pm 0.05
sensibility	0.94 \pm 0.13	0.89 \pm 0.07	0.89 \pm 0.07
specificity	0.85 \pm 0.09	0.88 \pm 0.17	0.93 \pm 0.09
AUC	0.97 \pm 0.04	0.94 \pm 0.08	0.97 \pm 0.05
parameters	C: 10.0, gamma: 0.01, kernel: 'rbf'	'alpha': 0.1, 'hidden_layer': 10, 'max_iter': 5000, 'solver': 'lbfgs'	'algorithm': 'ball_tree', 'n_neighbors': 3, 'weights': 'distance'

comparaison des variables

Avec Pressure, la classification est optimale

	Raw	Penups Raw	Pressure
Accuracy	0.77 ± 0.13	0.84 ± 0.09	0.91 ± 0.05
sensibility	0.77 ± 0.15	0.86 ± 0.12	0.89 ± 0.07
specificity	0.78 ± 0.15	0.82 ± 0.15	0.93 ± 0.09
AUC	0.85 ± 0.09	0.92 ± 0.07	0.97 ± 0.05
classifier parameters	SVM C: 100.0, gamma: 0.01, kernel: 'rbf'	SVM C: 10.0, gamma: 0.01, kernel: 'rbf'	KNN 'algorithm': 'ball_tree', 'n_neighbors': 3, 'weights': 'distance'