Deep Learning – MAI

Autonomous lab – Transfer Learning

MAMe: Museum Art Medium dataset

Alberto Becerra
Riccardo Corsiglia

Table of Contents

Baseline Results

Model #	Train	Validation	Test	Overfitting
1	0.97	0.307	0.27	High
2	0.66	0.45	0.43	High
3-4	0.72	0.58	0.58	Medium
5-6	0.86	0.62	0.53	Medium
8	0.77	0.60	0.64	Medium
10	0.73	0.64	0.63	Medium
11	0.54	0.52	0.51	Low
12	0.59	0.63	0.62	Low
13	0.75	0.67	0.67	Medium

Fine-Tuning Baselines

	FS	VS	Architecture
R65k	81.35%	81.39%	VGG11
	81.20%	81.21%	VGG16
	83.33%	82.66%	ResNet18
	84.29%	84.07%	ResNet50
	73.14%	76.06%	DenseNet121
	83.73%	82.38%	EfficientNet-B0
	85.11%	83.48%	EfficientNet-B3

Feature Extraction

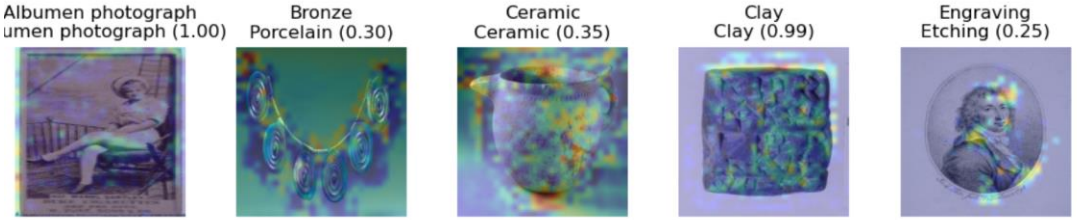
Task Similarity

Features: seem to activate in a smarter and focused way than Scratch model

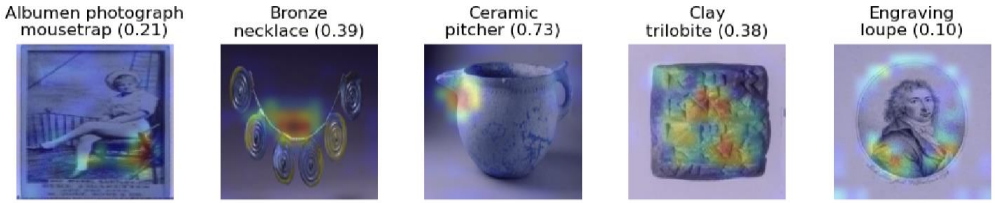
Task: Similar to ours. It focuses on the content

Goal: Use these quality features and change the way labels are assigned

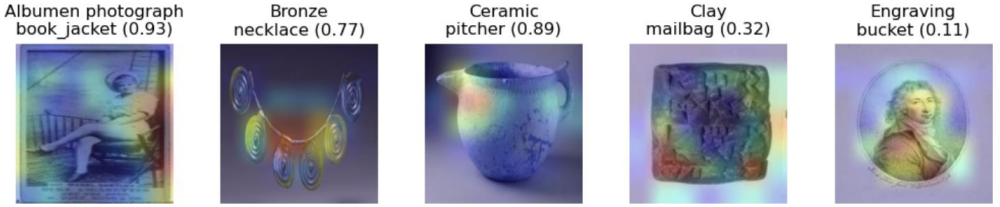
Predictions from ScratchModel



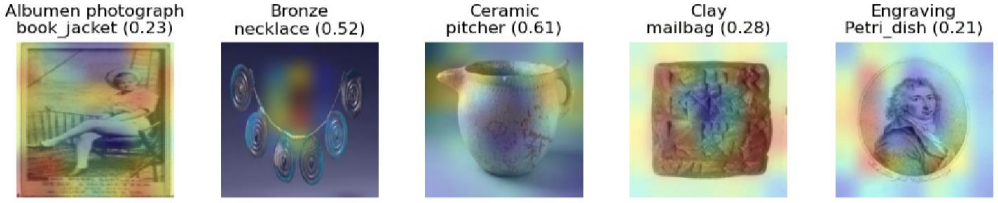
Predictions from VGG16



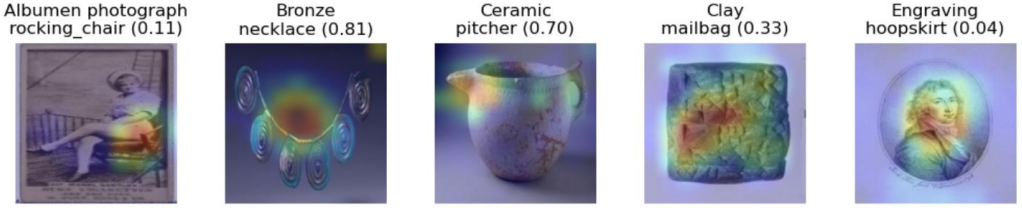
Predictions from ResNet50



Predictions from EfficientNetB0

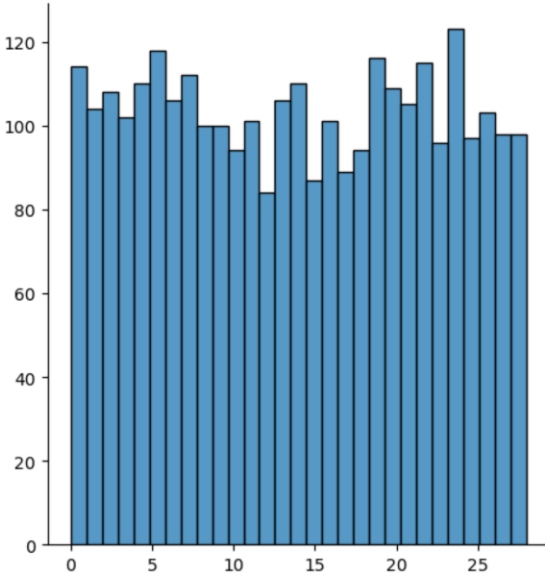


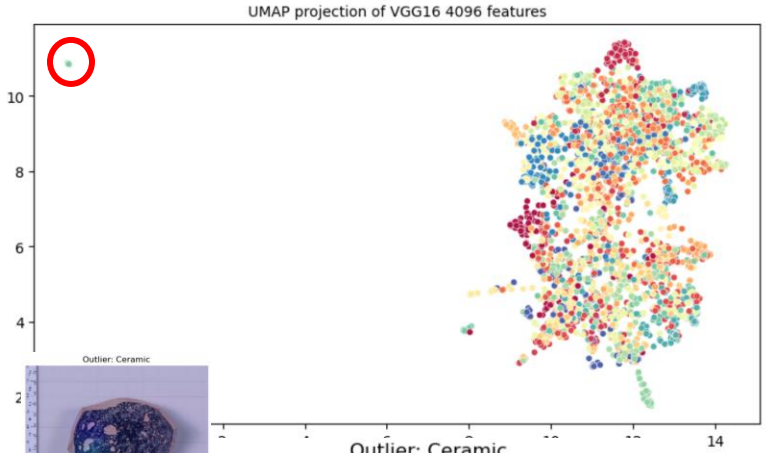
Predictions from EfficientNetB3

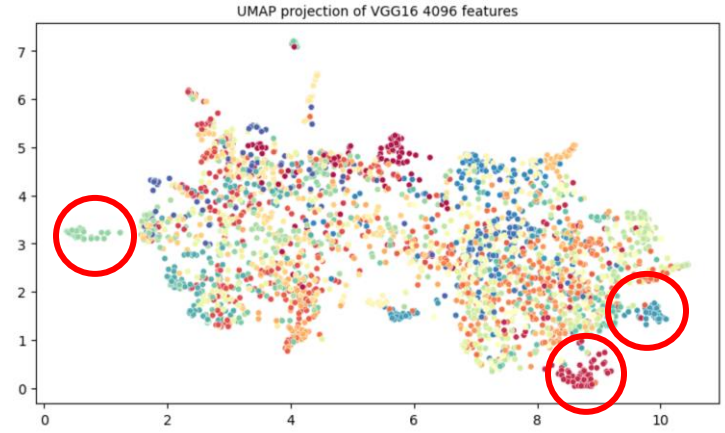


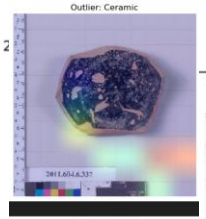
VGG16: Visualization and Insights

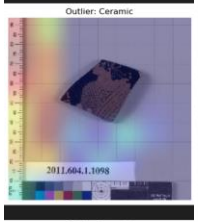
Sample of approximately
100 points per class

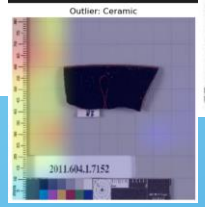


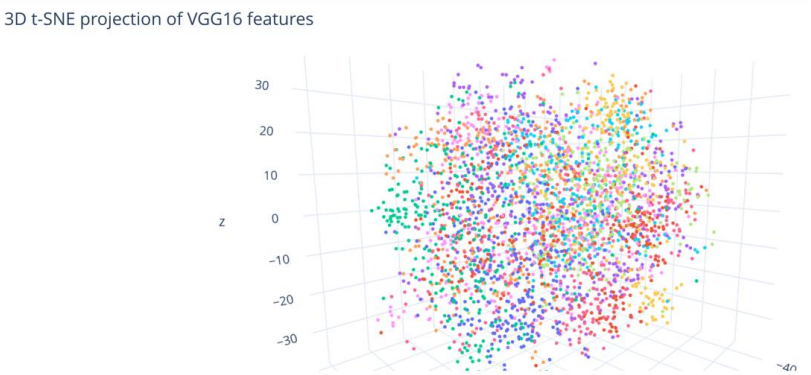




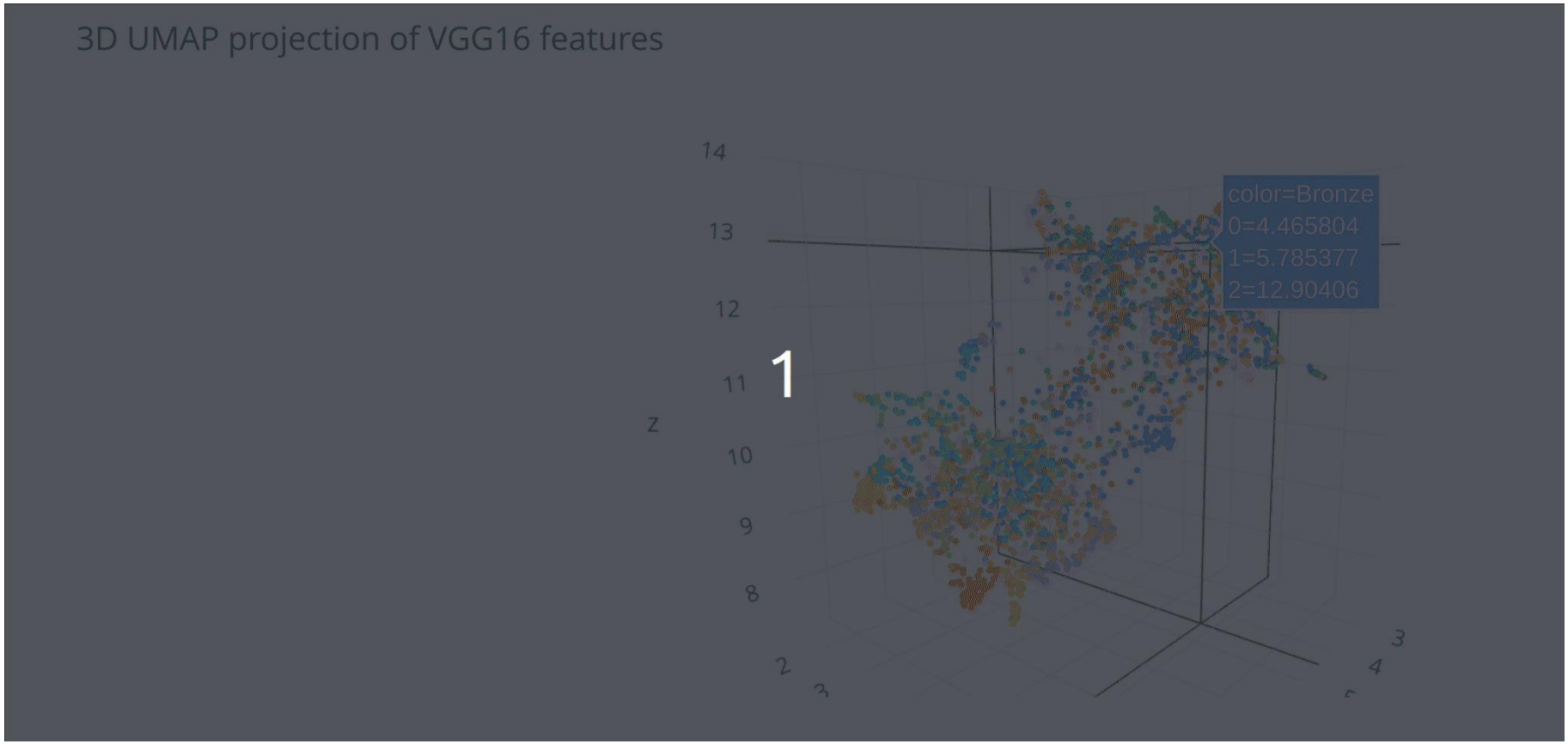






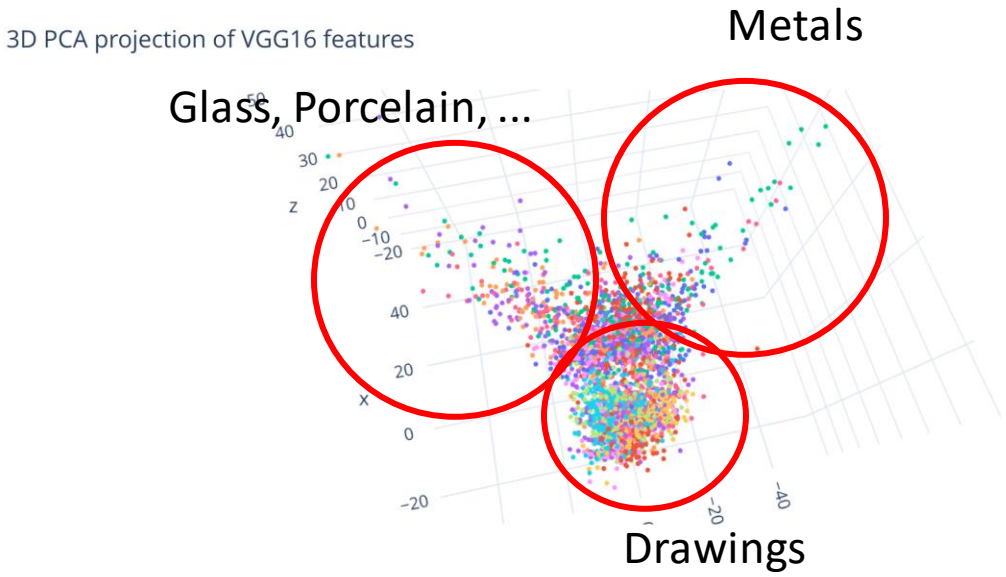


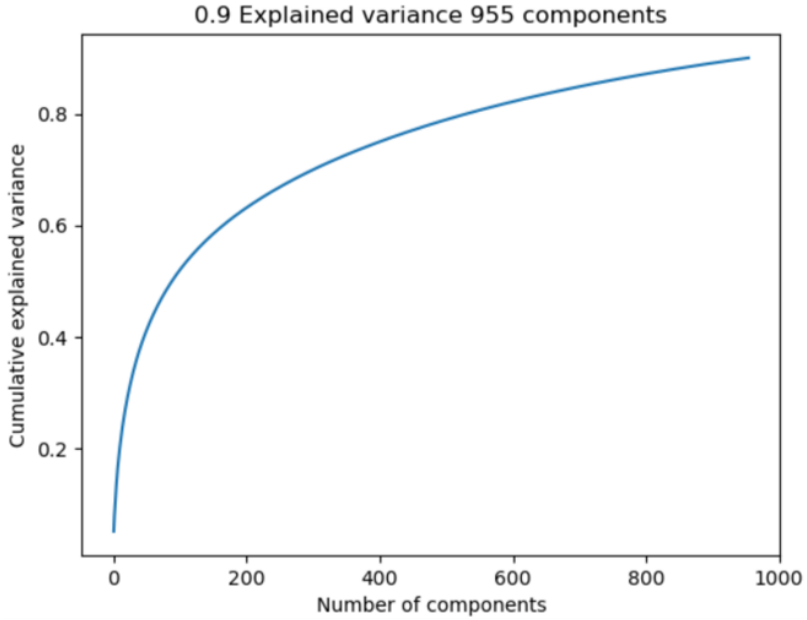
VGG16: 3D Projection



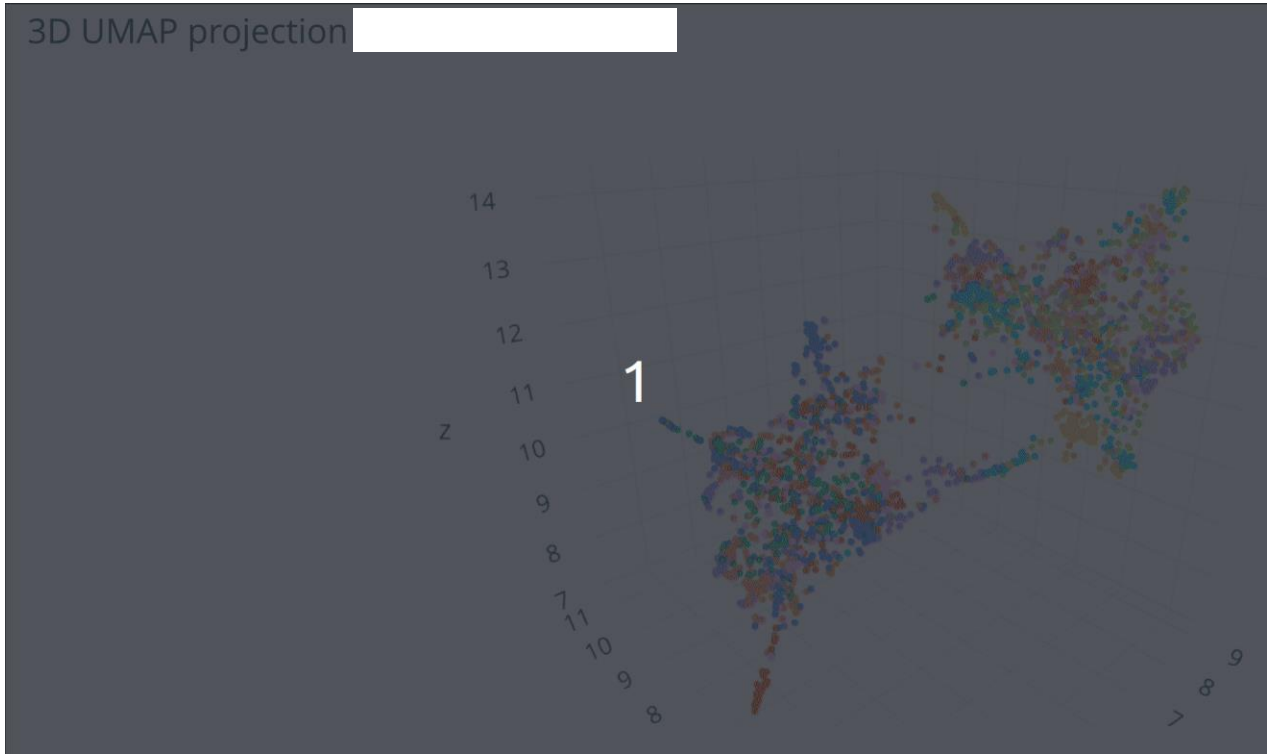
VGG16: PCA and Linear Separability

There are 3 groups linearly separable





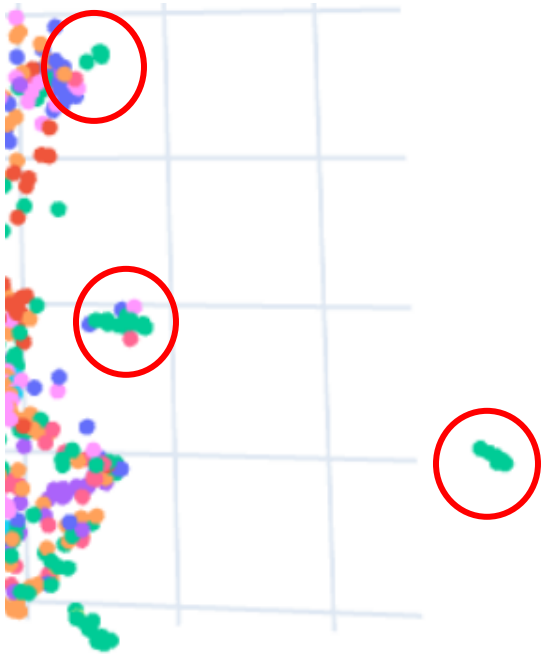
ResNet50: 3D Features Structure



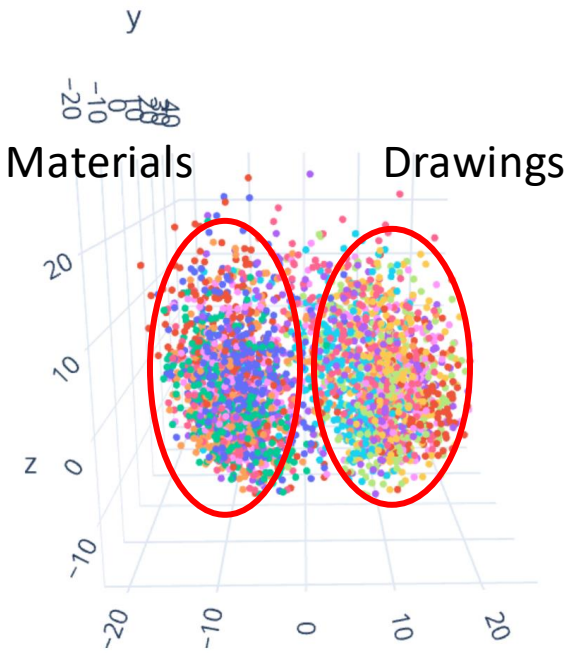
- Connected by Hand-Colored engraving (between Albumen Photograph and Woven Fabric
- Albumen Photograph perfectly clustered
- Woven Fabric divided between both big groups: materials and drawings

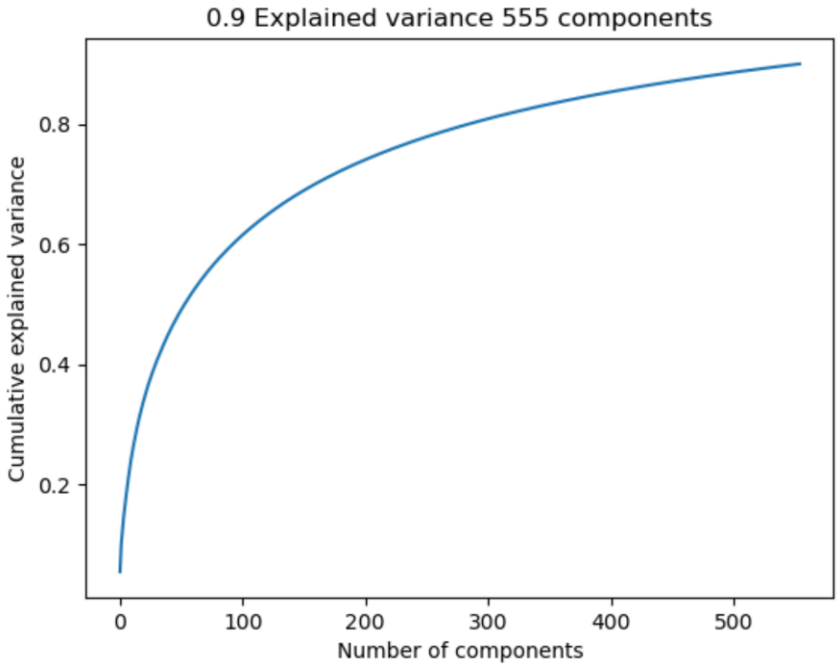
ResNet50: Outliers and particularities

Should it be a specific category?



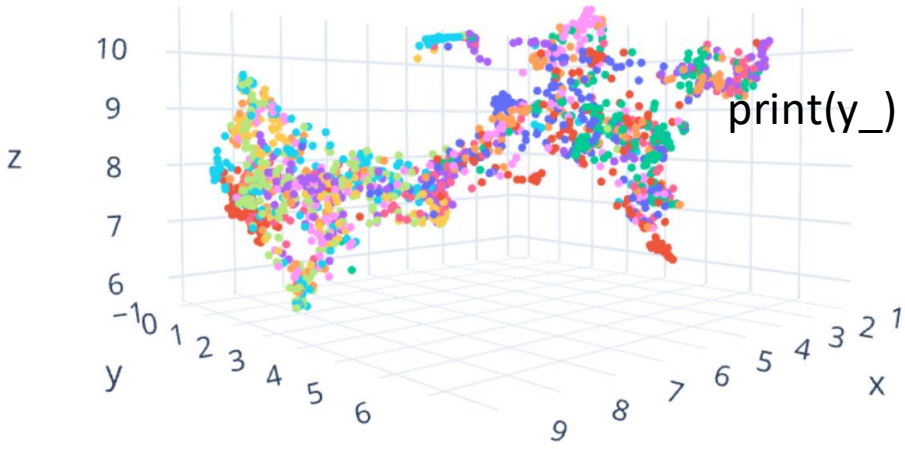
Features Visualization: ResNet50



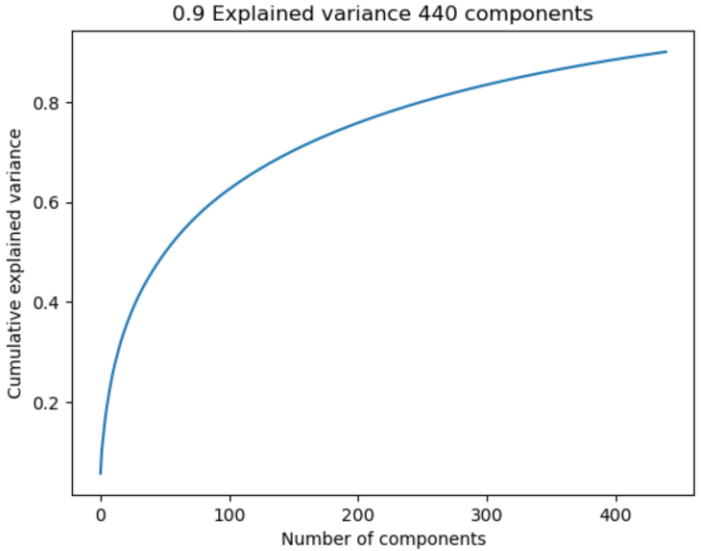


Features Visualization: EfficientNetB0

UMAP for EfficientNetB0

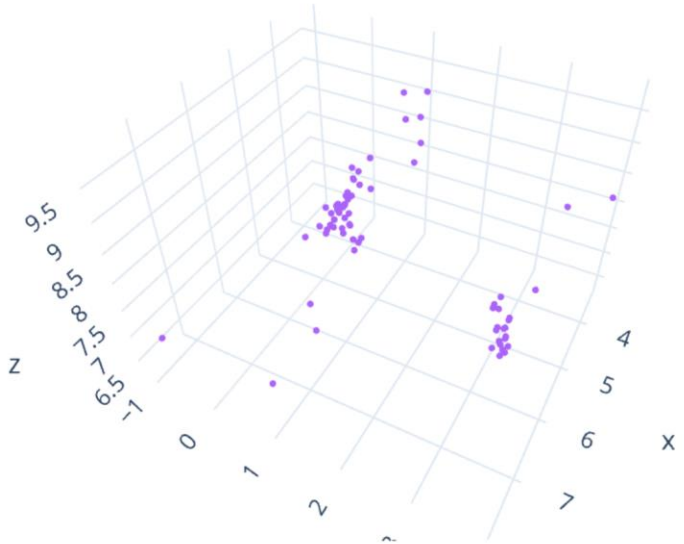


print(y_)

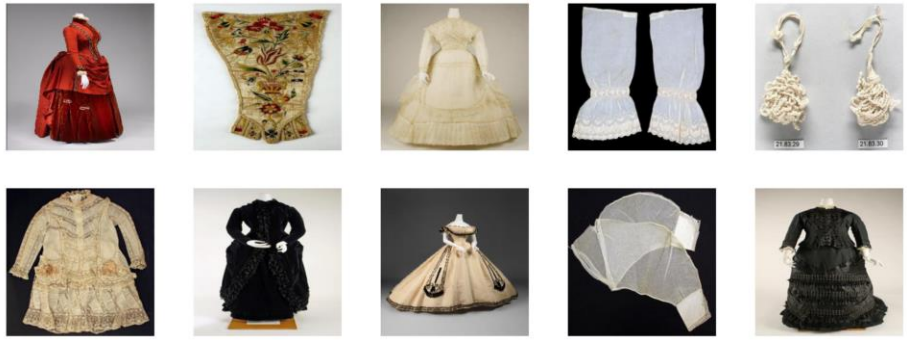


Features Visualization: EfficientNetB0

Woven Fabric



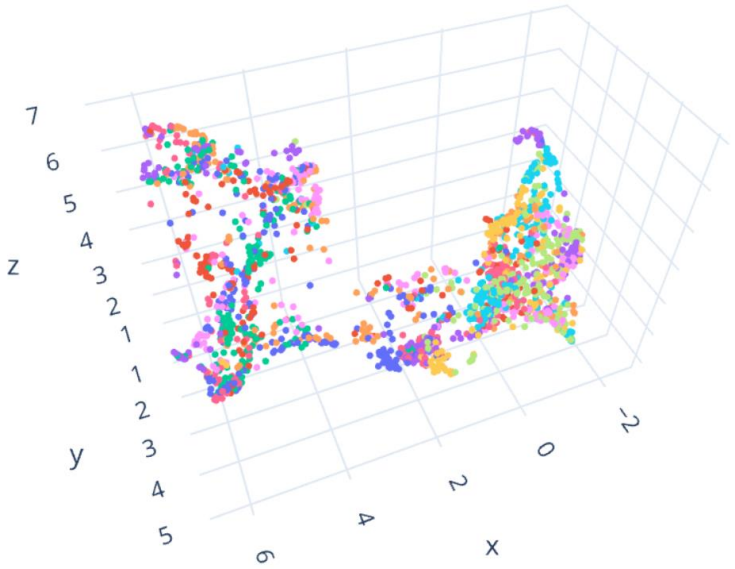
Near Hand-colored engraving and Porcelain

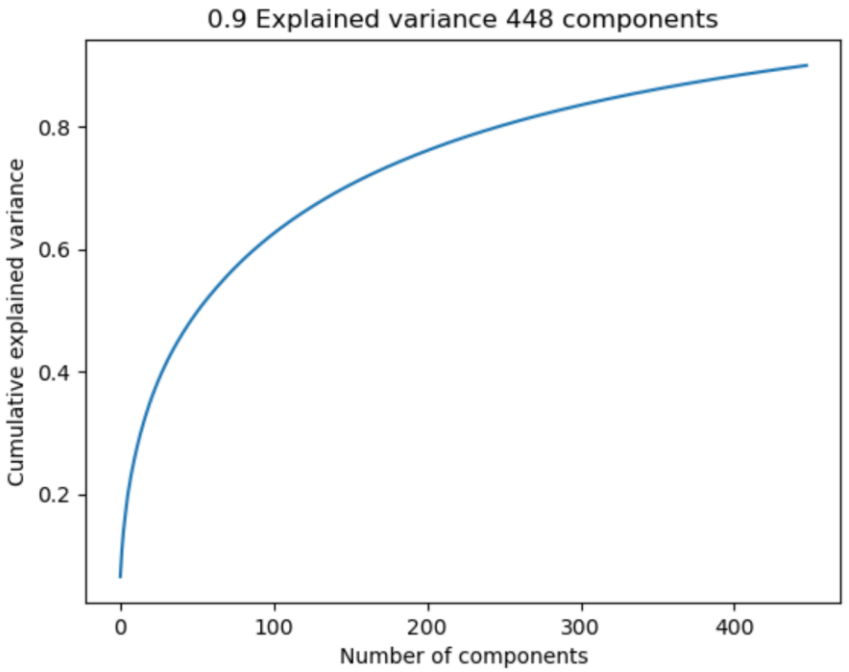


Near silk&metal thread

Features Visualization: EfficientNetB3

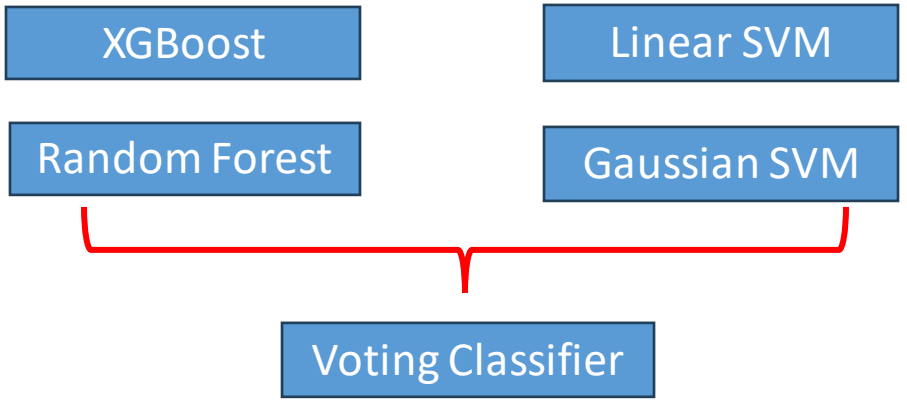
UMAP for EfficientNetB3





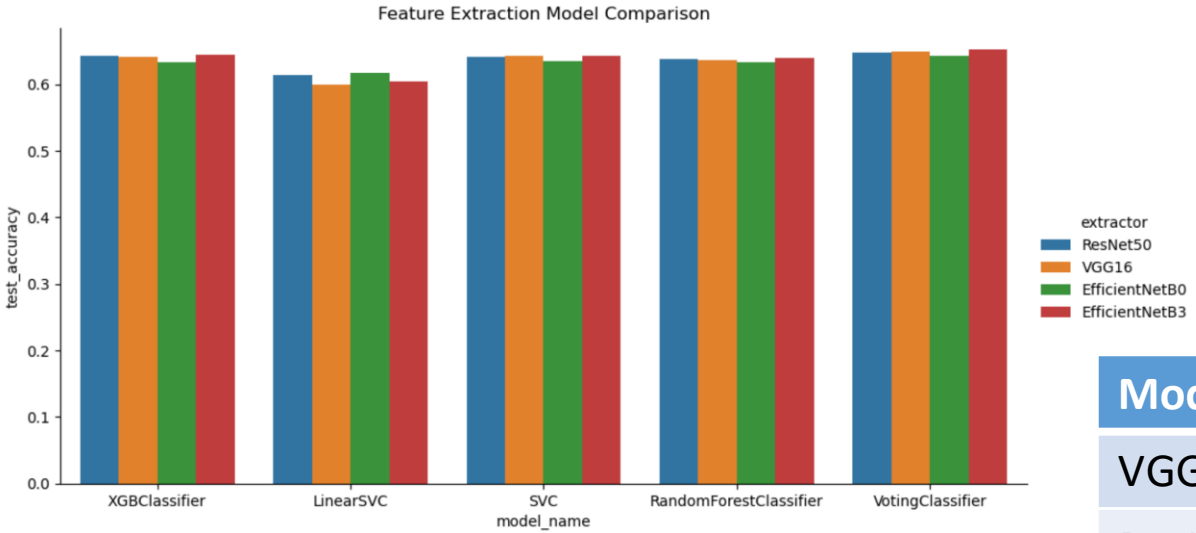
Classification Setup

Width
Height
Product size
Aspect ratio
umap_0
umap_1
umap_2
umap_3
umap_4
umap_5
umap_6
umap_7
Museum_Los Angeles County Museum of Art
Museum_Metropolitan Museum of Art
Museum_The Cleveland Museum of Art



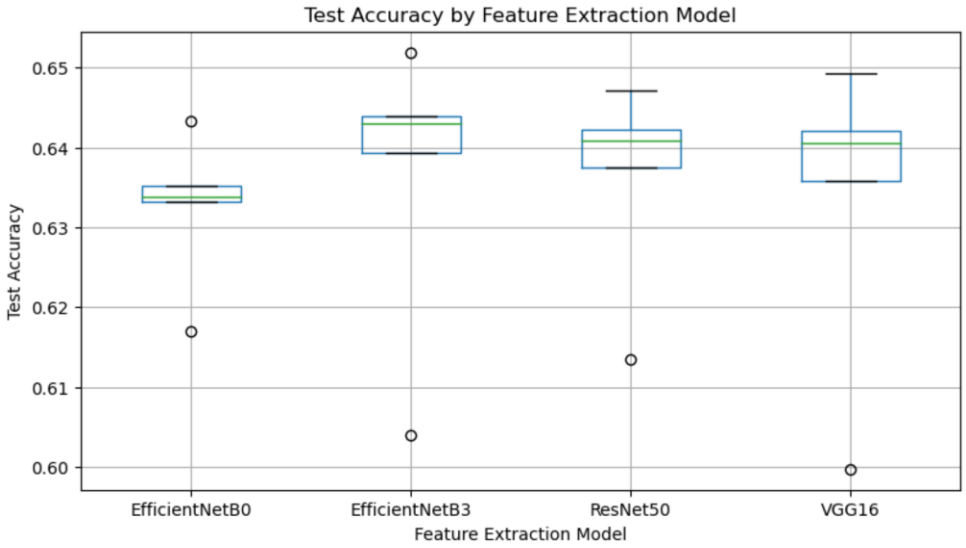
Model	N params	N features
VGG16	102.764.544	4096
ResNet50	23.587.712	2048
EfficientNetB0	4.049.571	1280
EfficientNetB3	10.783.535	1536

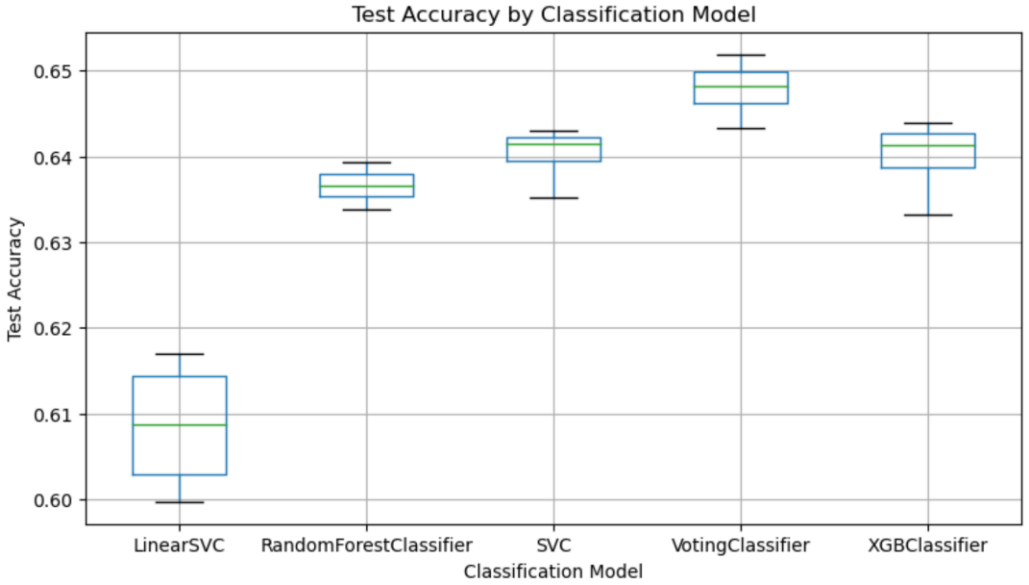
Classification CNN output features



Model	Best Model	Test Score
VGG16	Voting Class	0.649
ResNet50	Voting Class	0.644
EfficientNetB0	Voting Class	0.643
EfficientNetB3	Voting Class	0.644

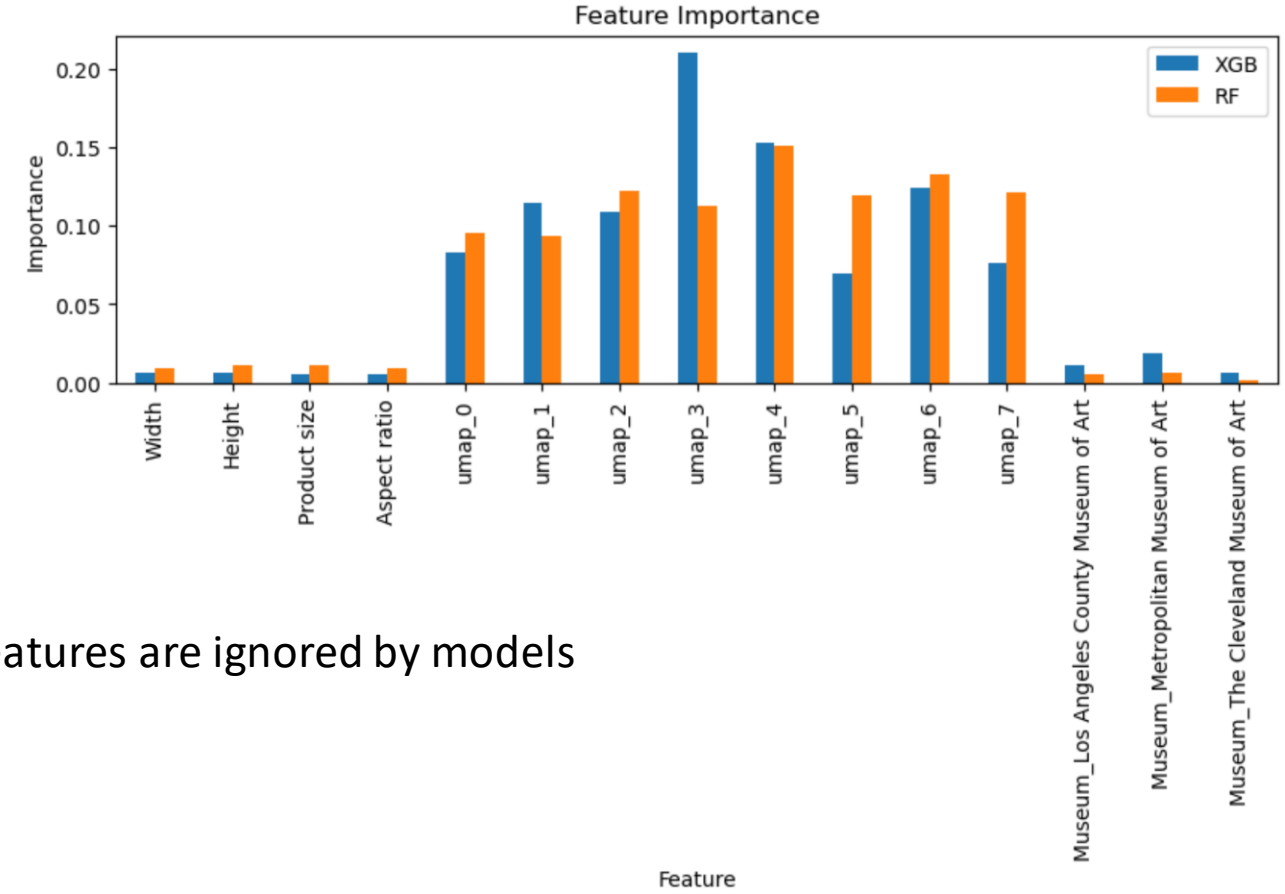
Results





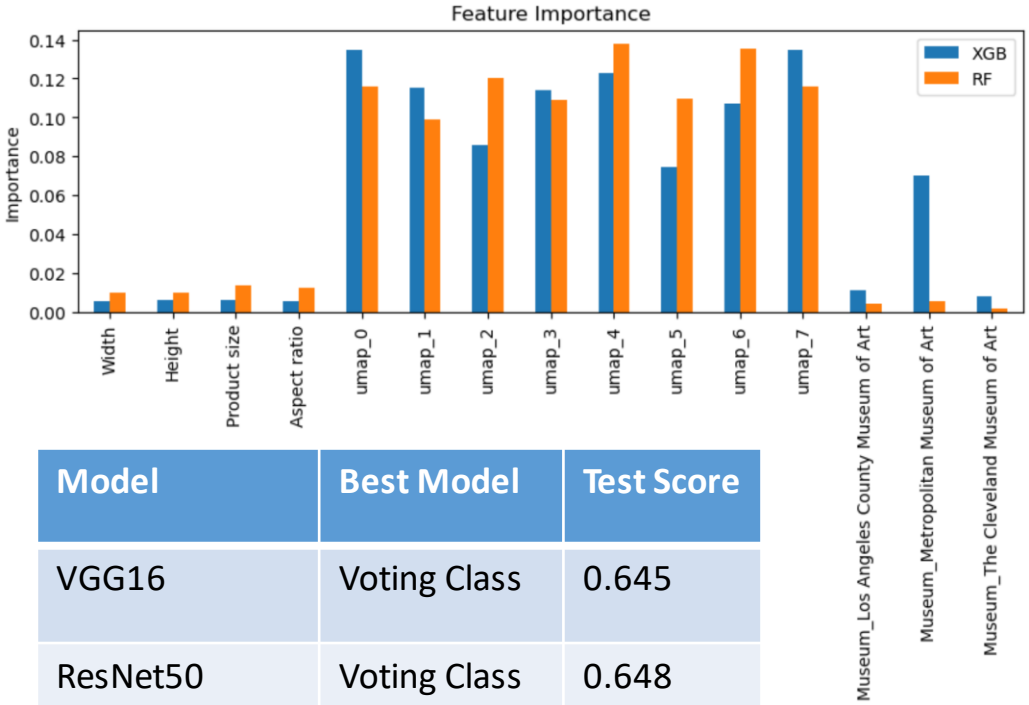
- EfficientNetB0 performs worse than the rest
- LinearSVC seems to underfit
- Difficult for models to learn and separate classes using these features
- Around 64% of overlapping in tasks

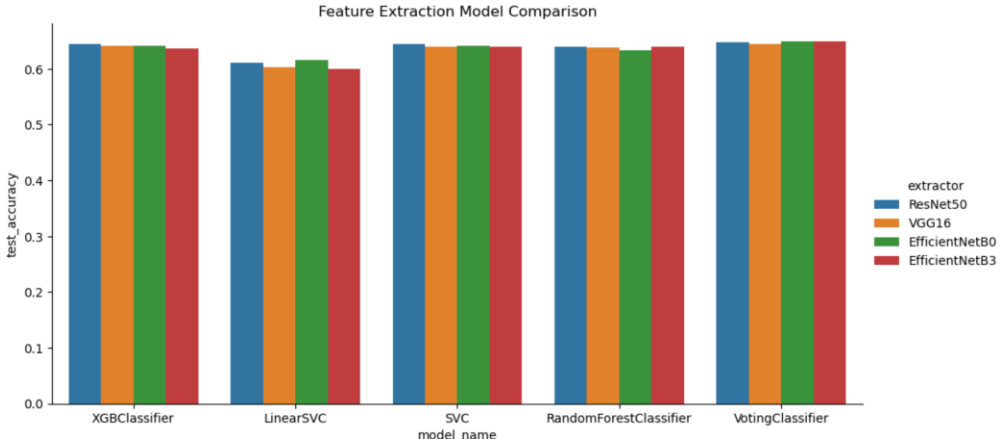
Results



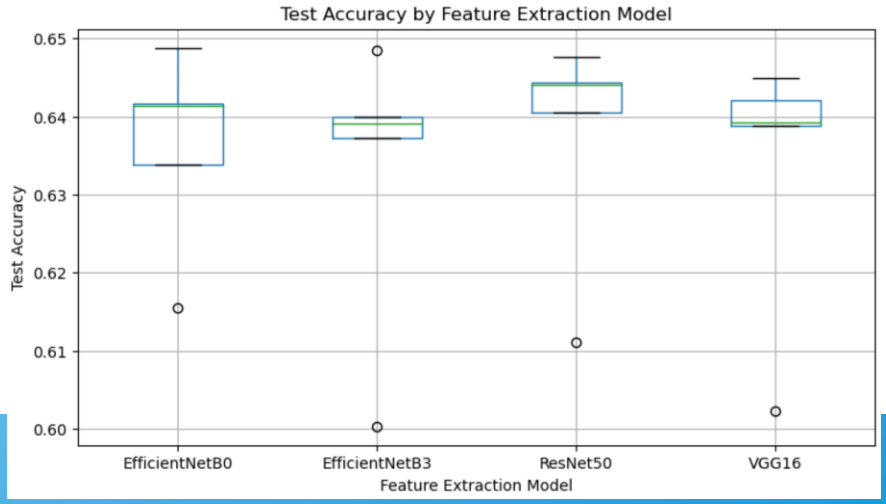
- External features are ignored by models

Use pretrained dense classifier

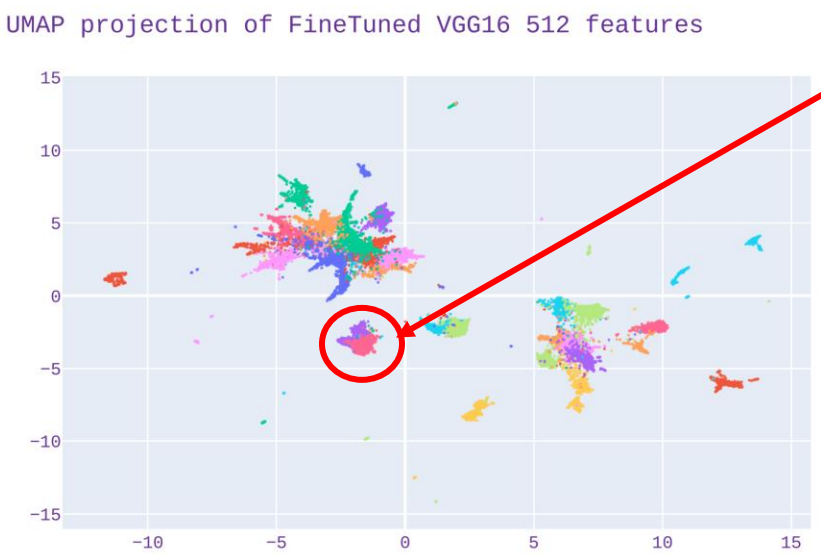




Model	Best Model	Test Score
VGG16	Voting Class	0.645
ResNet50	Voting Class	0.648
EfficientNetB0	Voting Class	0.649
EfficientNetB3	Voting Class	0.649

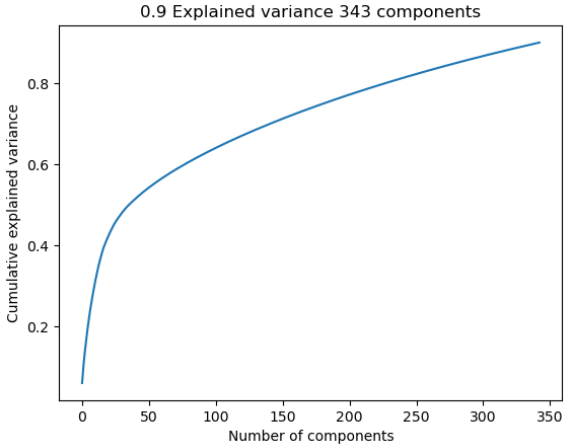


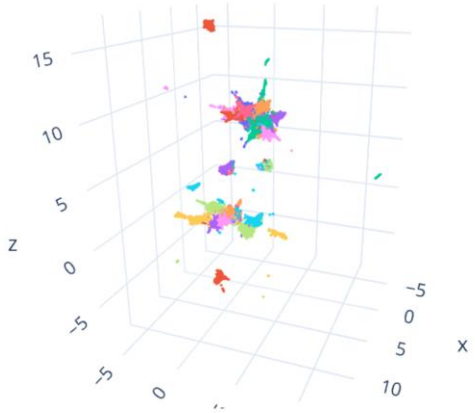
FineTuned VGG16 Features



SIlk and metal thread
Woven Fabric

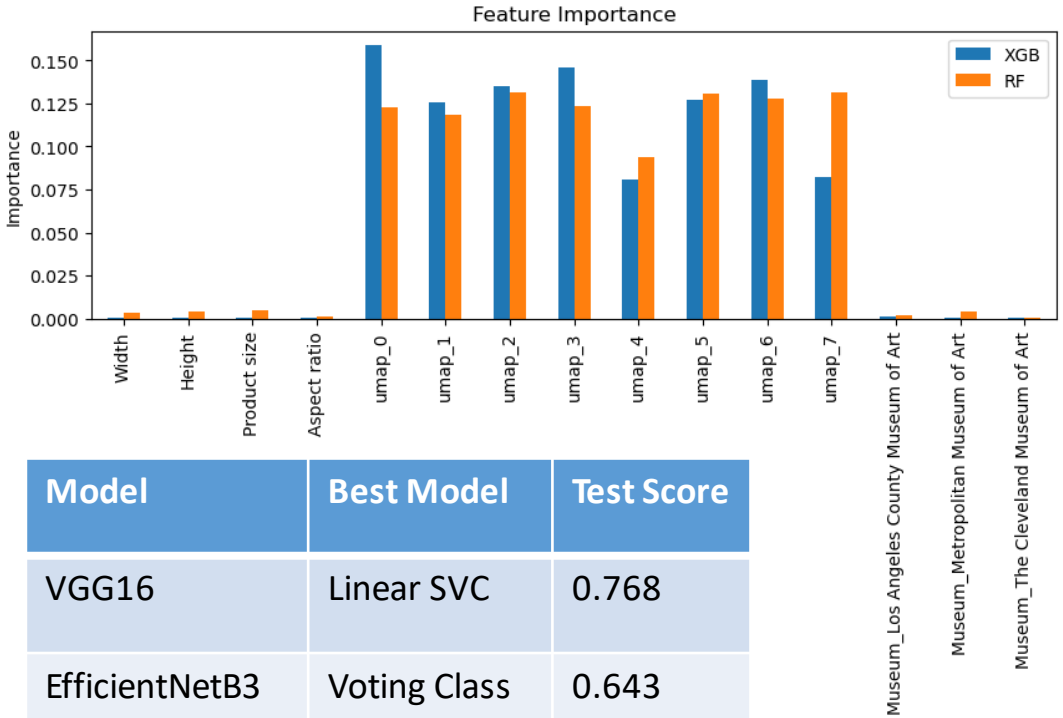
3D UMAP projection of FineTuned VGG16 512 features



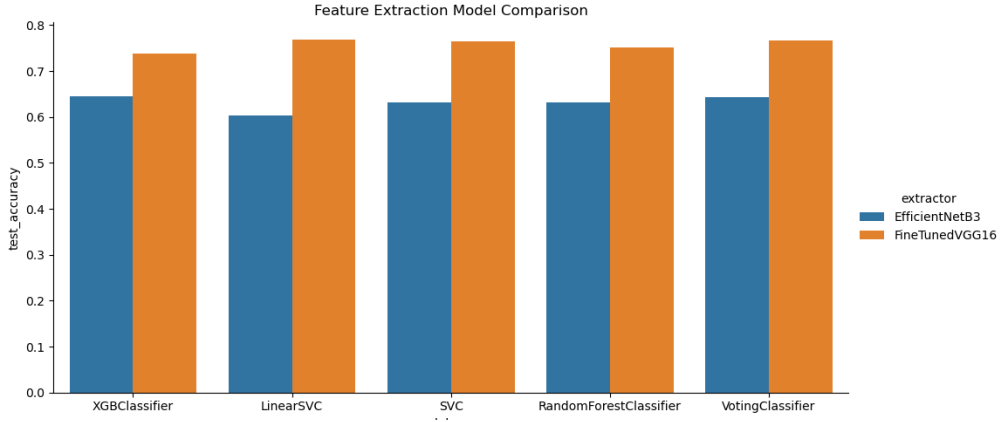


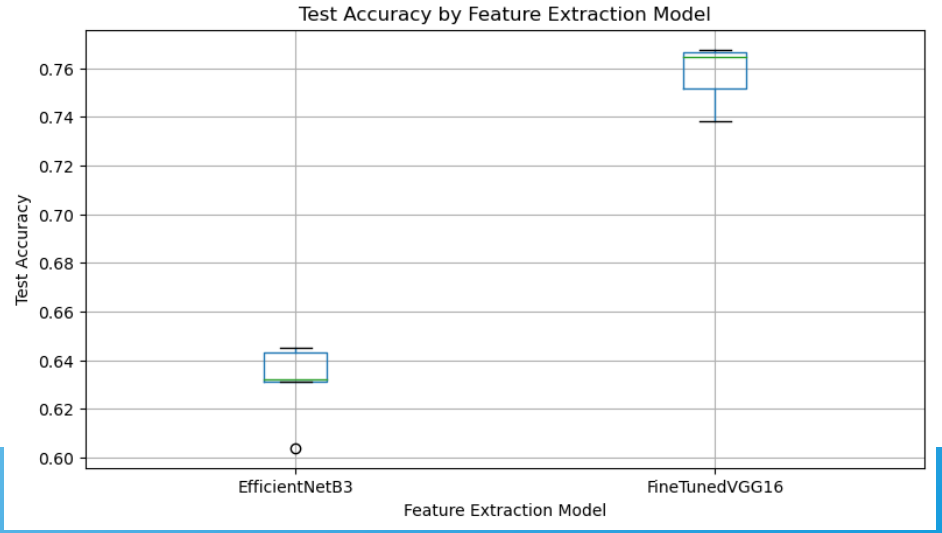
- Much clearer class separability
- Problems with Silk and Metal Thread aren't solved

FineTuned VGG16 Features

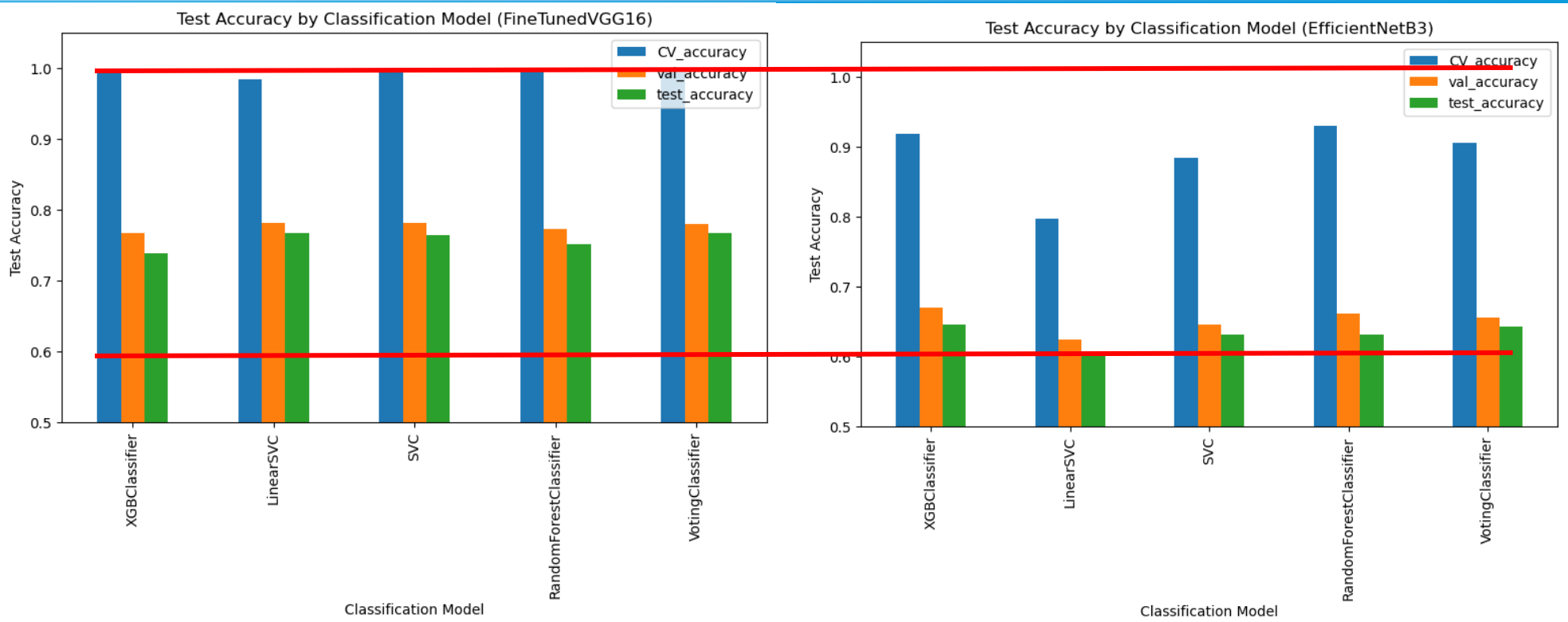


Model	Best Model	Test Score
VGG16	Linear SVC	0.768
EfficientNetB3	Voting Class	0.643





FineTuned VGG16 Features



- LinearSVC from zero to hero!

Conclusions on Feature Extraction

- There is overlap between ImageNet classification and MAME
- Introduction of a Task Similarity Measure with 64% of similarity in this case
- Feature extraction allows a better understanding of images distribution and outliers
- Fine tuning on feature extraction allows the use of pretrained basic features in addition to task specific ones
- Final result of 77% of accuracy
- It's not about the classifier but about the features

Fine Tuning

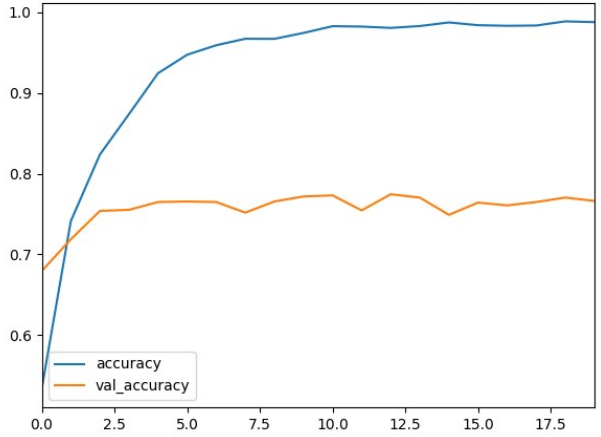
VGG16: First try #1

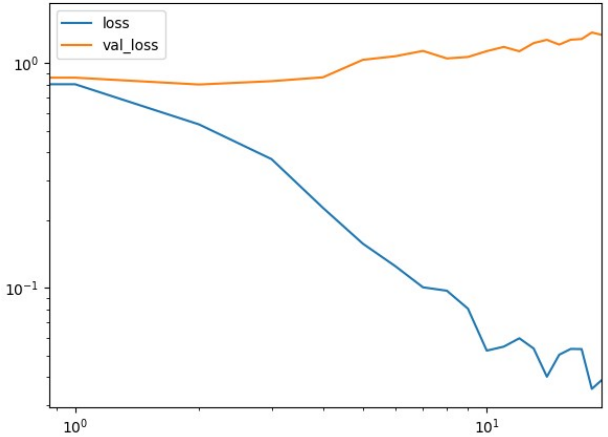
Our first try consisted in freezing the frist three layers and fine tune the other two.
We did this just because we followed the classic rule of thumb of keeping domain related part of the network and training the more task specific layers.
We didnt keep the top of the layer, instead we initially put two 512 fc layers.

Learning rate = 1e-4
Batch size = 128
Epochs = 20

Val Acc = 74%

Improvements:
-smaller learning rate
-regularize





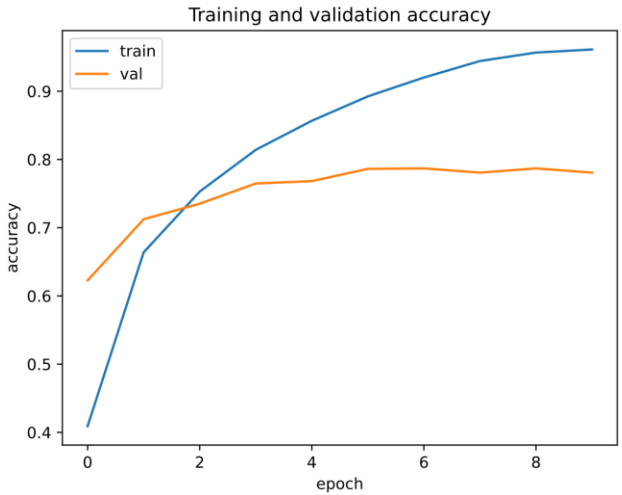
VGG16: Smaller learning rate #2

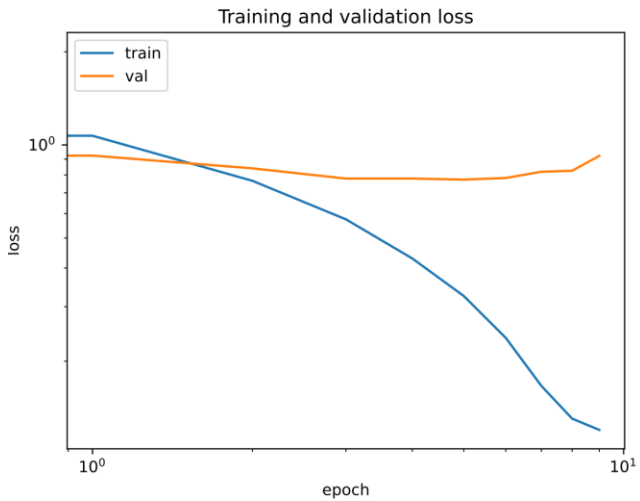
We got an improvement reducing by then the learning rate and shrinking the last fully connected layer to 256 neurons, which is a way to reduce the complexity of the model, thus regularizing.
The validation loss is still not decreasing, thus there is still room for decreasin the learning rate

Learning rate = 5e-5
Batch size = 192
Epochs = 10

Val Acc = 76%

Improvements:
-smaller learning rate
-introduce regularizer





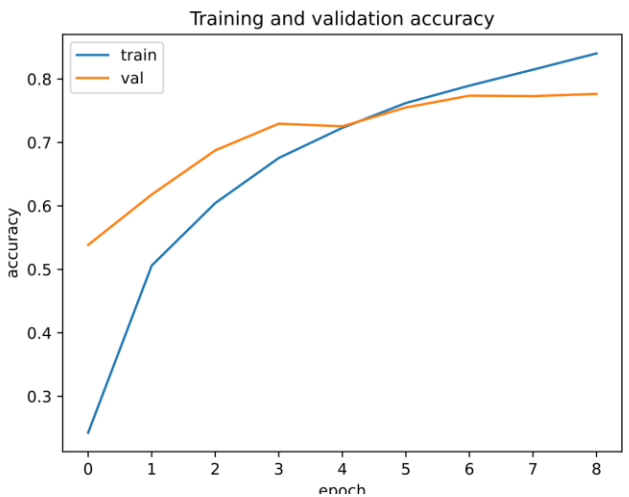
VGG16: L2 and Dropout #3

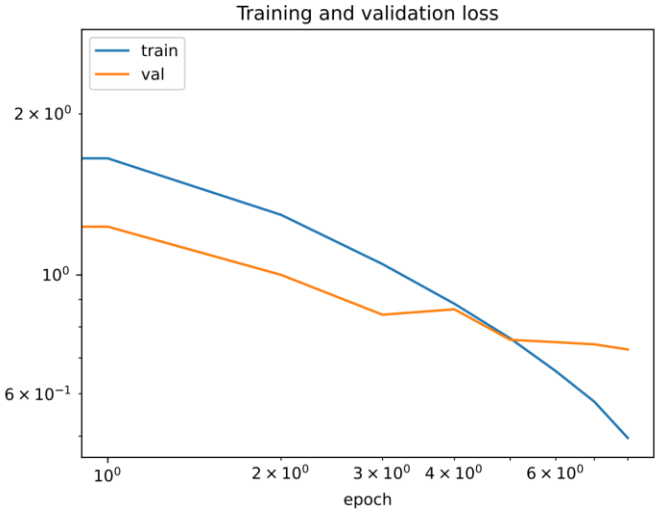
Introducing L2 normalization helped the optimization process to find the right way for minimizing the loss. Even if this was a small experiment (9 ep) it's possible to see an improvement in the loss plot even if the accuracy is more of less still the same. Here we also thought to add more capacity but more neurons in the fc layers performed significantly worse

Learning rate = 5e-5
Batch size = 250
Epochs = 9
L2 = 0.1
Dropout = 0.5

Val Acc = 75%

Improvements:
-Even smaller learning rate



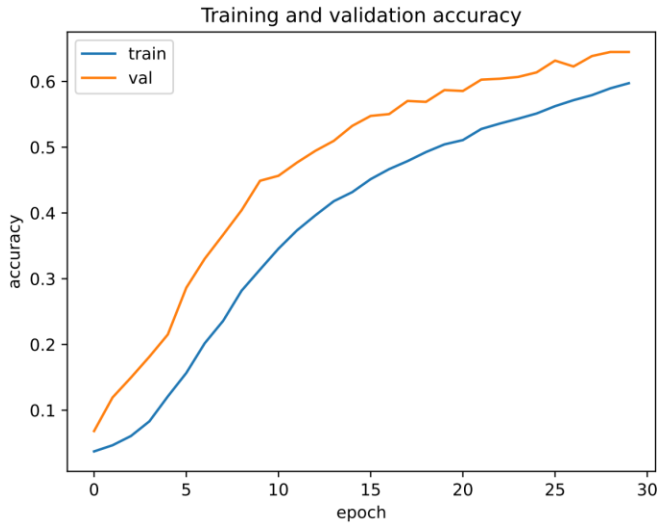


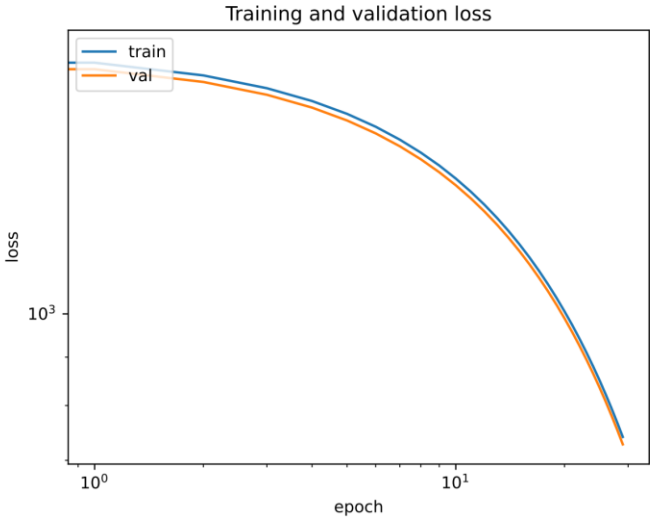
VGG16: Small LR but need more epochs #4

Reducing more the learning rate shown a nice improvement in the learning process even though the training time increase substantially due to more epochs been done.
In this particulare example we also added more regularization to test some values and we can see that we have sign of a too hard training. Indeed the validation accuracy is always better than training.

Learning rate = 2e-6
Batch size = 300
Epochs = 30
L2 = 0.5

Improvements:
-Longer training
-Less L2/Dropout or more capacity





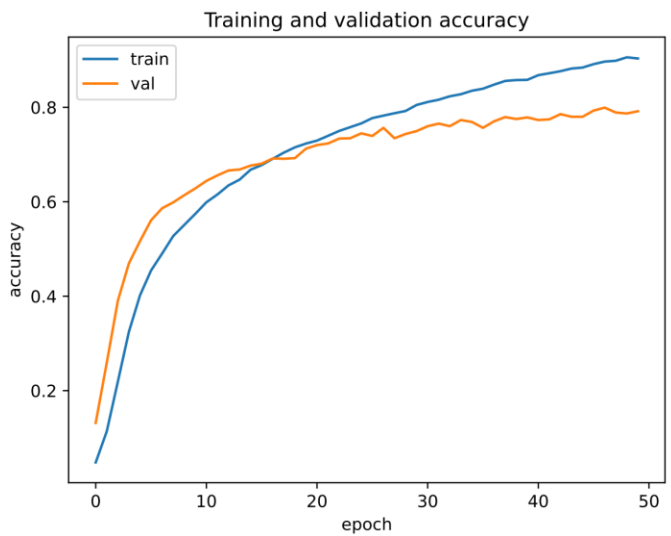
VGG16: A good result #5

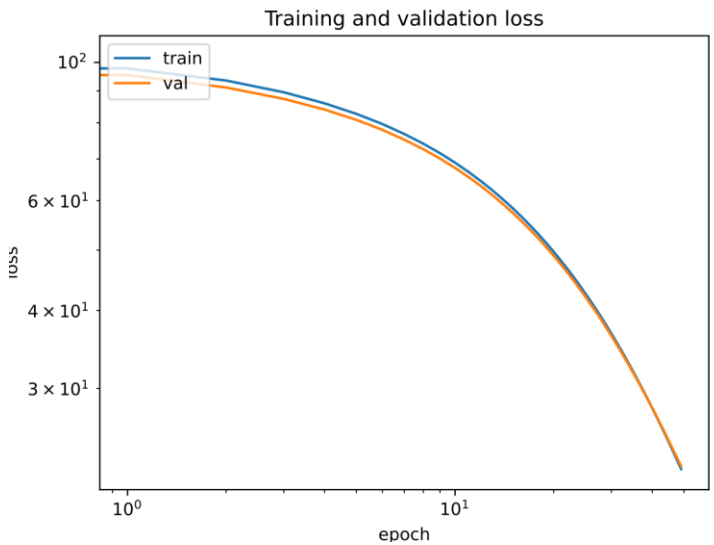
The result we got with this configuration is quite nice.
Its main problems are the still “too hard training” at the beginning, but it becomes a “could be regularized more” in the end.
A really nice aspect is that the loss is still decreasing in log scale and that we can probably improve more

Learning rate = 3e-6
Batch size = 300
Epochs = 50
L2 = 0.1

Val Acc = 77%

Improvements:
-Longer training
-L2: more or less?





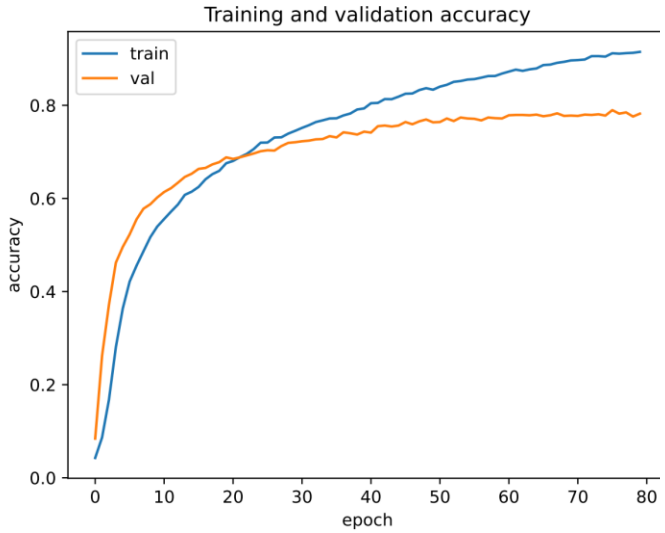
VGG16: Final Result #6

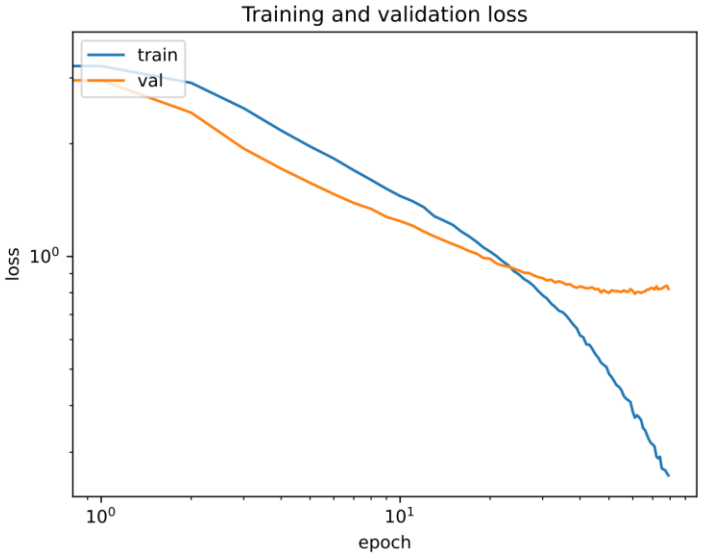
We extended the training until the loss started even to increase. We noticed that after the loss started to increase, the accuracy was still slowly improving. We addressed this phenomenon to the nature of the task and the loss. Even in a prediction vector with a low maximum value, it counts ad correct for the accuracy but has a high loss for the cross entropy because the model is less sure. Do we want a higher accuracy and lower confidence or vice versa?

Learning rate = 2.5e-6
Batch size = 300
Epochs = 80
L2 = 0.15

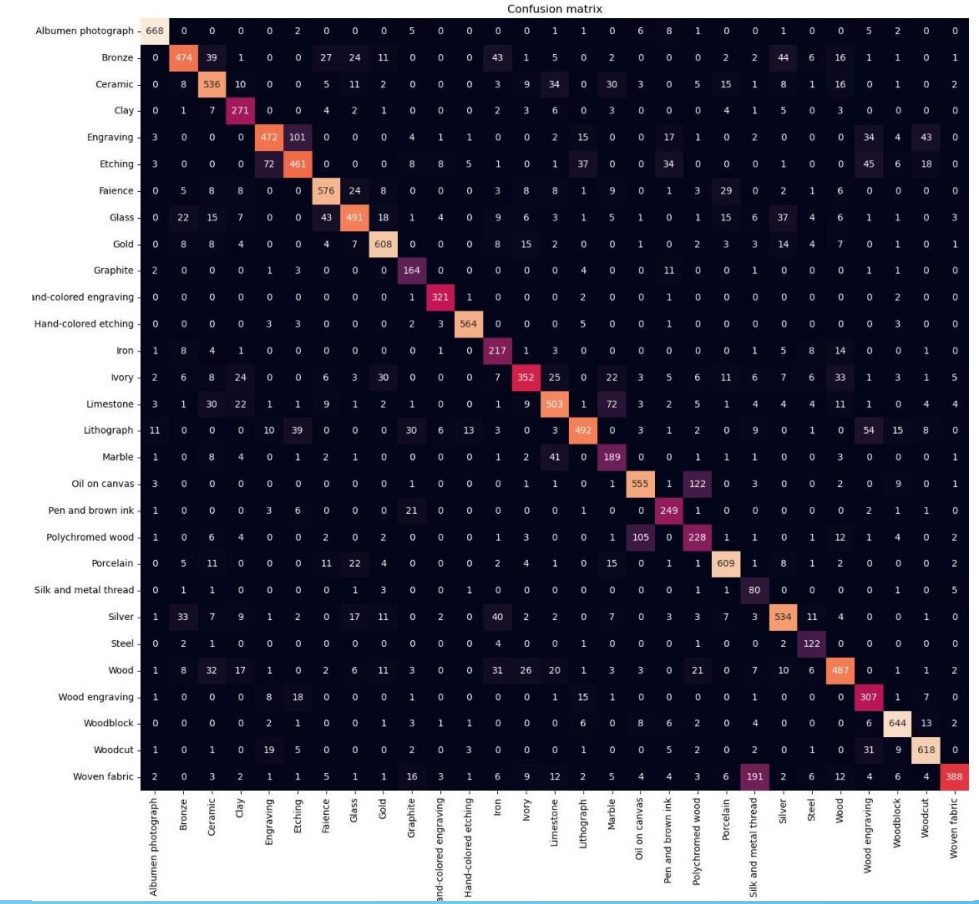
Val acc= 78%

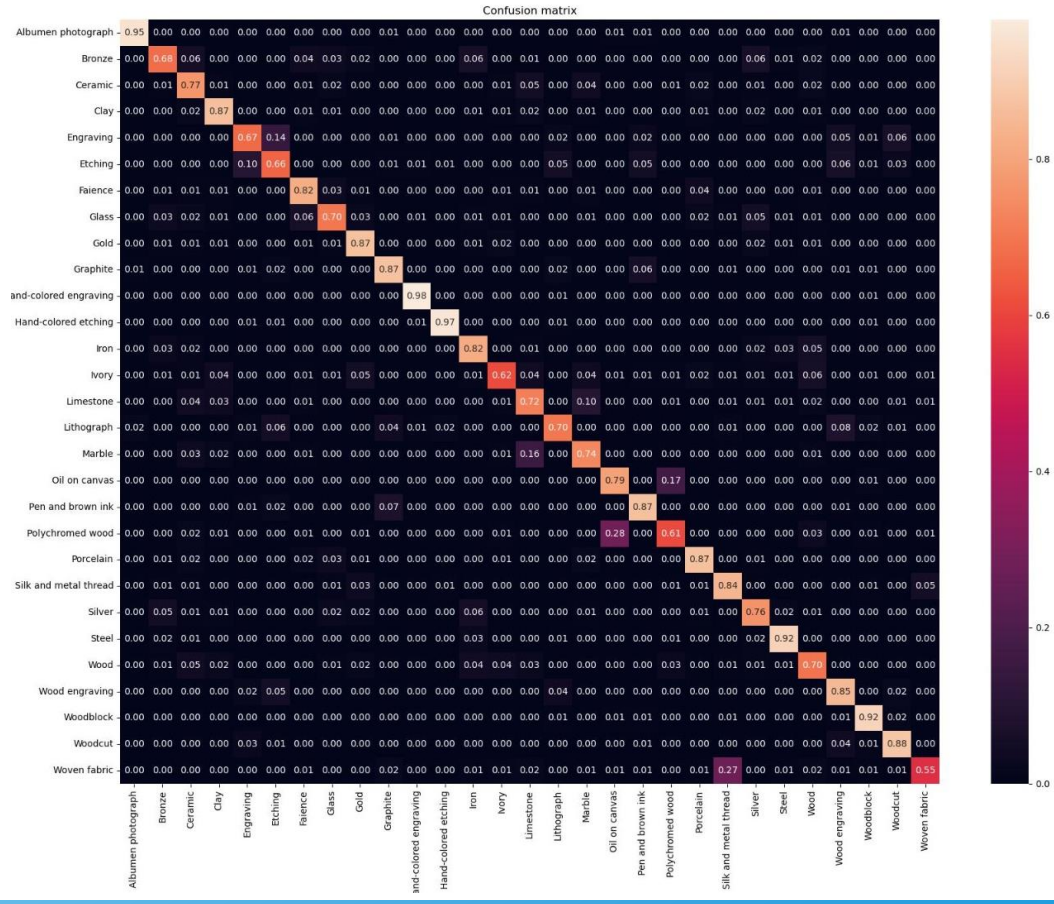
Improvements:
-Longer training
-Less L2 or more capacity





Test Confusion Matrix





Classification Report

	P	R	F	Support
Albumen photograph	0.95	0.95	0.95	700
Bronze	0.81	0.68	0.74	700
Ceramic	0.74	0.77	0.75	700
Clay	0.71	0.87	0.78	313
Engraving	0.79	0.67	0.73	700
Etching	0.72	0.66	0.69	700
Faience	0.83	0.82	0.83	700
Glass	0.80	0.70	0.75	700
Gold	0.85	0.87	0.86	700
Graphite	0.62	0.87	0.73	188
Hand-colored engraving	0.92	0.98	0.95	328
Hand-colored etching	0.96	0.97	0.96	584
Iron	0.57	0.82	0.67	265
Ivory	0.78	0.62	0.69	572
Limestone	0.75	0.72	0.73	700
Lithograph	0.84	0.70	0.77	700
Marble	0.52	0.74	0.61	257
Oil on canvas	0.80	0.79	0.80	700
Pen and brown ink	0.71	0.87	0.78	286
Polychromed wood	0.55	0.61	0.58	375
Porcelain	0.86	0.87	0.87	700
Silk and metal thread	0.24	0.84	0.38	95
Silver	0.78	0.76	0.77	700
Steel	0.67	0.92	0.77	133
Wood	0.77	0.70	0.73	700
Wood engraving	0.62	0.85	0.72	361
Woodblock	0.90	0.92	0.91	700
Woodcut	0.86	0.88	0.87	700
Woven fabric	0.93	0.55	0.69	700
accuracy			0.79	15657
macro avg	0.77	0.79	0.77	15657
weighted avg	0.80	0.79	0.79	15657

Further Improvements

- Use information from feature analysis to deal with outliers and difficult classes
- Fine Tuning more model in literature and extract features from them
- Ensemble classifiers based on different fine tuned networks
- Use models pretrained with more similar images such as Wikiart.

References

[1] Parés, F., Arias-Duart, A., Garcia-Gasulla, D. et al. The MAMe dataset: on the relevance of high resolution and variable shape image properties. Appl Intell 52, 11703–11724 (2022). https://doi.org/10.1007/s10489-021-02951-w

[2] Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. J Big Data 6, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0

[3] Chollet, F. Grad-CAM class activation visualization (2021) https://keras.io/examples/vision/grad_cam/

[4] Shima, Y. Image Augmentation for Object Image Classification Based On Combination of Pre-Trained CNN and SVM. Journal of Physics: Conference Series. https://dx.doi.org/10.1088/1742-6596/1004/1/012001

References

[5] Yang, H., & Min, K. (2019). Classification of basic artistic media based on a deep convolutional approach. The Visual Computer. doi:10.1007/s00371-019-01641-6

[6] Sun, T., Wang, Y., Yang, J., & Hu, X. (2017). Convolution Neural Networks With Two Pathways for Image Style Recognition. IEEE Transactions on Image Processing, 26(9), 4102–4113. doi:10.1109/tip.2017.2710631

[7] Taheri, S., Ezoji, M., & Sakhaei, S. M. (2020). Convolutional neural network based features for motor imagery EEG signals classification in brain–computer interface system. SN Applied Sciences, 2(4). doi:10.1007/s42452-020-2378-z