

Image Classification of Artworks



Classifying Artworks with the WikiArt Dataset from Huggingface

Federica Andreazza

Academic Year
2023/2024

Project Overview

- ❖ Objective: Classify artworks by **artists** using the WikiArt dataset from Huggingface.
- ❖ Approach:
 - Use pre-trained **ResNet** and **AlexNet** architectures for artist classification.
 - Initially attempted to classify the entire dataset.
 - Sampled a subset due to dataset size.
 - Further refined dataset to focus on artworks of four major artists for a more balanced representation.
 - Tune hyperparameters (learning rate, dropout).
- ❖ Outcomes:
 - Evaluate and compare ResNet and AlexNet performance in artist classification.

WikiArt Dataset

❖ Structure:

- "artist" : 129 artist classes, including a "Unknown Artist" class
- "genre" : 11 genre classes, including a "Unknown Genre" class
- "style" : 27 style classes

❖ Dataset Details:

```
IterableDataset({  
    features: ['image', 'artist', 'genre', 'style'],  
    n_shards: 72  
})
```

❖ Example Entry:

```
<PIL.JpegImagePlugin.JpegImageFile image mode=RGB size=750x597 at 0x7D9564E07C40> 22 4 21
```

Dataset Handling

- ❖ Original dataset size and splits:

Total items:	81444
Train data size:	48866 (60%)
Validation data size:	16288 (20%)
Test data size:	16290 (20%)

- ❖ Sampled subset:

Train data size:	5000
Validation data size:	1667
Test data size:	1667

- ❖ Refined dataset to focus on artworks of **four major artists** for balanced representation:

Train data size:	3865
Validation data size:	1288
Test data size:	1289

Four major artists



Vincent Van Gogh
9.8%

Nicholas Roerich
9.5%



Pierre Auguste Renoir
7.4%

Claude Monet
7.1%



Image Pre-Processing Steps

❖ Training set (Data Augmentation):

- RandomResizedCrop (size=256, scale=(0.8, 1.0))
- RandomRotation (degrees=15) -> rotates the image by a random angle in the range of -15 to 15 degrees.
- RandomHorizontalFlip (randomly flips with a default probability of 50%)
- CenterCrop (size=224)
- Each channel in the tensor is normalized as $T = (T - \text{mean}) / (\text{standard deviation})$

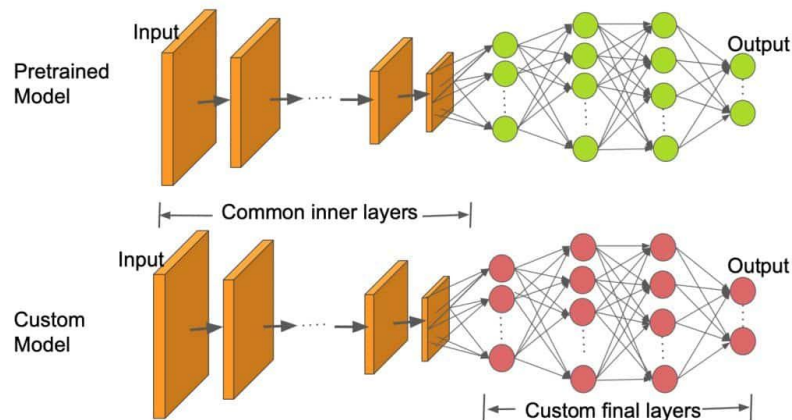
❖ Validation and Test Sets:

- Resize (size=256)
- CenterCrop (size=224)
- Each channel in the tensor is normalized as $T = (T - \text{mean}) / (\text{standard deviation})$

Model Architecture

❖ Transfer Learning in PyTorch:

- Transfer learning is a machine learning technique where a model trained on one task is re-purposed on a second related task.
- We use a pre-trained model as our base and change the last few layers to classify images according to our desirable classes.



Model Architecture

❖ ResNet50 vs ResNet18:

- Different number of layers (ResNet50 has 50 layers, whereas ResNet18 has only 18 layers).
- ResNet18 is suitable for tasks with limited computational resources.
- ResNet50 achieves higher accuracy and it is expected to perform better with more complex data.

❖ AlexNet:

- One of the first CNN architectures.
- 8 layers total (5 convolutional layers followed by 3 fully connected layers).

❖ All these models have been pre-trained on the ImageNet dataset.

❖ Both use the Negative Loss Likelihood (NLLLoss) loss function as it is useful for classifying multiple classes.

Overall Structure of ResNet18

- ❖ Input Size: [32, 3, 224, 224]
- ❖ Total Parameters: 11,308,868
 - Trainable: 132,356
 - Non-trainable: 11,176,512
- ❖ Memory Size: 2070.83 MB
- ❖ ResNet50:
 - Total params: 24,033,604
 - Trainable params: 525,572
 - Non-trainable params: 23,508,032

Layer (type)	Output Shape	Param #
Conv2d-1	[32, 64, 112, 112]	9,408
BatchNorm2d-2	[32, 64, 112, 112]	128
ReLU-3	[32, 64, 112, 112]	0
MaxPool2d-4	[32, 64, 56, 56]	0
...		
BasicBlock-50	[32, 256, 14, 14]	0
Conv2d-51	[32, 512, 7, 7]	1,179,648
BatchNorm2d-52	[32, 512, 7, 7]	1,024
ReLU-53	[32, 512, 7, 7]	0
Conv2d-54	[32, 512, 7, 7]	2,359,296
BatchNorm2d-55	[32, 512, 7, 7]	1,024
Conv2d-56	[32, 512, 7, 7]	131,072
BatchNorm2d-57	[32, 512, 7, 7]	1,024
ReLU-58	[32, 512, 7, 7]	0
BasicBlock-59	[32, 512, 7, 7]	0
Conv2d-60	[32, 512, 7, 7]	2,359,296
BatchNorm2d-61	[32, 512, 7, 7]	1,024
ReLU-62	[32, 512, 7, 7]	0
Conv2d-63	[32, 512, 7, 7]	2,359,296
BatchNorm2d-64	[32, 512, 7, 7]	1,024
ReLU-65	[32, 512, 7, 7]	0
BasicBlock-66	[32, 512, 7, 7]	0
AdaptiveAvgPool2d-67	[32, 512, 1, 1]	0
Linear-68	[32, 256]	131,328
ReLU-69	[32, 256]	0
Dropout-70	[32, 256]	0
Linear-71	[32, 4]	1,028
LogSoftmax-72	[32, 4]	0

Results on the subset of the four major artist

dropout = 0.3
learning rate = 1e-5

Epoch: 1/7

Training: Loss - 1.4151, Accuracy - 30.4269%,
Validation : Loss - 1.2434, Accuracy - 44.2547%, Time: 3149.5057s

Epoch: 2/7

Training: Loss - 1.3172, Accuracy - 45.4334%,
Validation : Loss - 1.1727, Accuracy - 46.8168%, Time: 3209.0047s

Epoch: 3/7

Training: Loss - 1.2678, Accuracy - 50.0647%,
Validation : Loss - 1.1128, Accuracy - 51.1646%, Time: 3209.6855s

Epoch: 4/7

Training: Loss - 1.2290, Accuracy - 52.7814%,
Validation : Loss - 1.0682, Accuracy - 52.8727%, Time: 3217.4729s

Epoch: 5/7

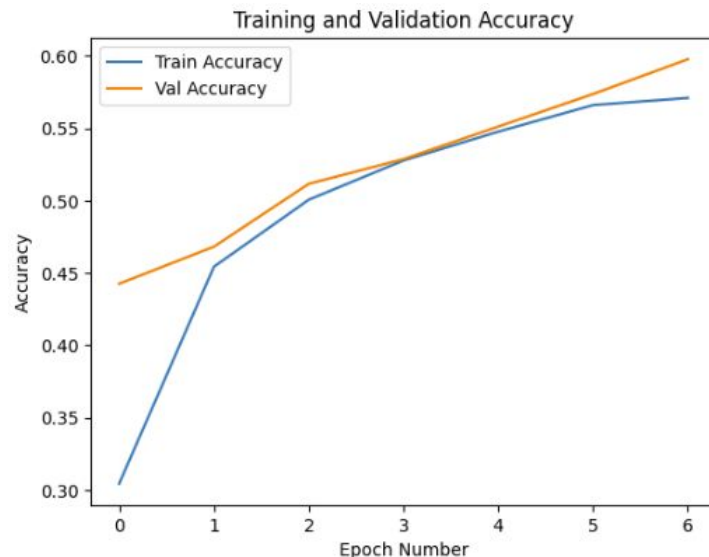
Training: Loss - 1.1904, Accuracy - 54.7736%,
Validation : Loss - 1.0310, Accuracy - 55.1242%, Time: 3176.0553s

Epoch: 6/7

Training: Loss - 1.1634, Accuracy - 56.6106%,
Validation : Loss - 0.9913, Accuracy - 57.3758%, Time: 3199.5789s

Epoch: 7/7

Training: Loss - 1.1432, Accuracy - 57.1022%,
Validation : Loss - 0.9601, Accuracy - 59.7826%, Time: 3299.8582s



Test Accuracy: 0.6067

Results on the subset of the four major artist

dropout = 0.3
learning rate = 1e-5

Epoch: 1/7

Training: Loss - 1.4062, Accuracy - 26.6494%,
Validation : Loss - 1.2992, Accuracy - 38.3540%, Time: 3014.6335s

Epoch: 2/7

Training: Loss - 1.3438, Accuracy - 36.0931%,
Validation : Loss - 1.2379, Accuracy - 45.1087%, Time: 3061.7574s

Epoch: 3/7

Training: Loss - 1.2958, Accuracy - 42.7167%,
Validation : Loss - 1.1877, Accuracy - 48.7578%, Time: 2987.0105s

Epoch: 4/7

Training: Loss - 1.2561, Accuracy - 47.6843%,
Validation : Loss - 1.1457, Accuracy - 51.0093%, Time: 2988.0616s

Epoch: 5/7

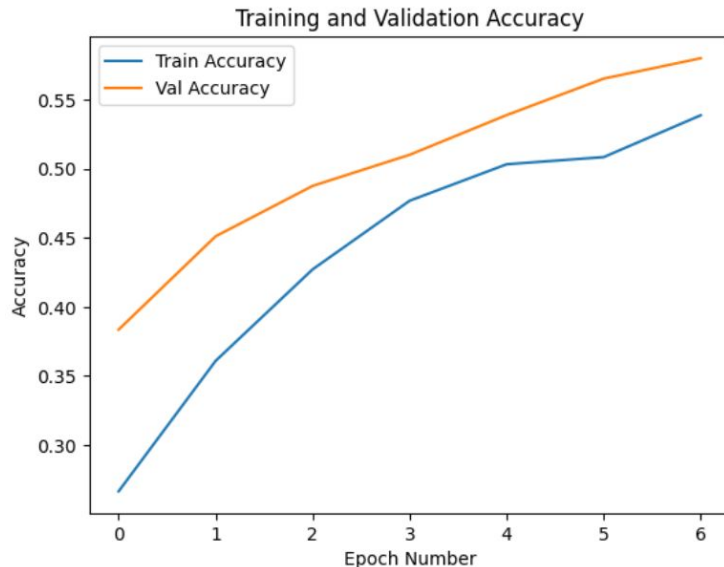
Training: Loss - 1.2304, Accuracy - 50.3234%,
Validation : Loss - 1.1082, Accuracy - 53.8820%, Time: 3055.0701s

Epoch: 6/7

Training: Loss - 1.2059, Accuracy - 50.8409%,
Validation : Loss - 1.0770, Accuracy - 56.5217%, Time: 3008.2806s

Epoch: 7/7

Training: Loss - 1.1859, Accuracy - 53.8680%,
Validation : Loss - 1.0476, Accuracy - 57.9969%, Time: 2997.3567s



Test Accuracy: 0.6175

Results on a subset of the original dataset

dropout = 0.4

learning rate = $1e-4$

Epoch: 1/5

Training: Loss - 3.3040, Accuracy - 14.1600%,

Validation : Loss - 2.6132, Accuracy - 28.6743%, Time: 3971.7131s

Epoch: 2/5

Training: Loss - 2.5964, Accuracy - 26.5800%,

Validation : Loss - 2.2169, Accuracy - 37.6725%, Time: 3201.6121s

Epoch: 3/5

Training: Loss - 2.2803, Accuracy - 34.8800%,

Validation : Loss - 1.9849, Accuracy - 46.1308%, Time: 3204.7930s

Epoch: 4/5

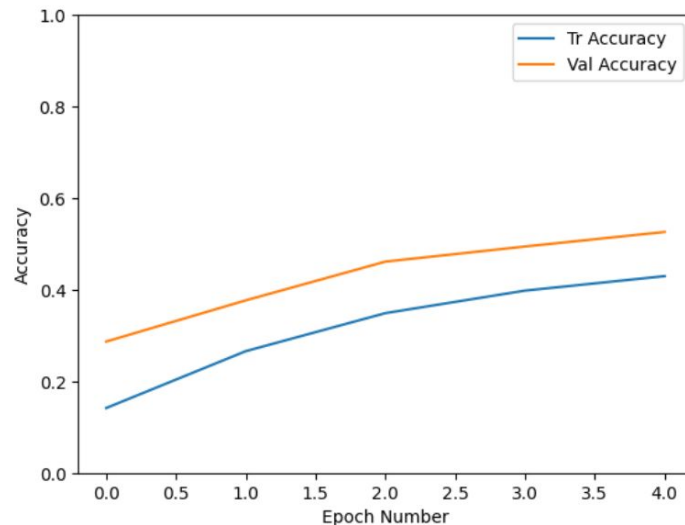
Training: Loss - 2.1021, Accuracy - 39.8000%,

Validation : Loss - 1.8382, Accuracy - 49.4301%, Time: 3251.7533s

Epoch: 5/5

Training: Loss - 1.9701, Accuracy - 42.9600%,

Validation : Loss - 1.7330, Accuracy - 52.6095%, Time: 3290.1512s



Test Accuracy: 0.5111

Results on the subset of the four major artist

dropout = 0.4
learning rate = 1e-4

Epoch : 1/7

Training: Loss - 1.4506, Accuracy - 47.6067%,
Validation : Loss - 1.1920, Accuracy - 33.3851%, Time: 3756.7104s

Epoch: 2/7

Training: Loss - 1.2793, Accuracy - 50.3234%,
Validation : Loss - 1.1520, Accuracy - 35.9472%, Time: 3154.8903s

Epoch: 3/7

Training: Loss - 1.2127, Accuracy - 53.4282%,
Validation : Loss - 1.0935, Accuracy - 39.7516%, Time: 3190.8342s

Epoch: 4/7

Training: Loss - 1.1549, Accuracy - 52.8072%,
Validation : Loss - 1.0598, Accuracy - 51.5528%, Time: 3242.8902s

Epoch: 5/7

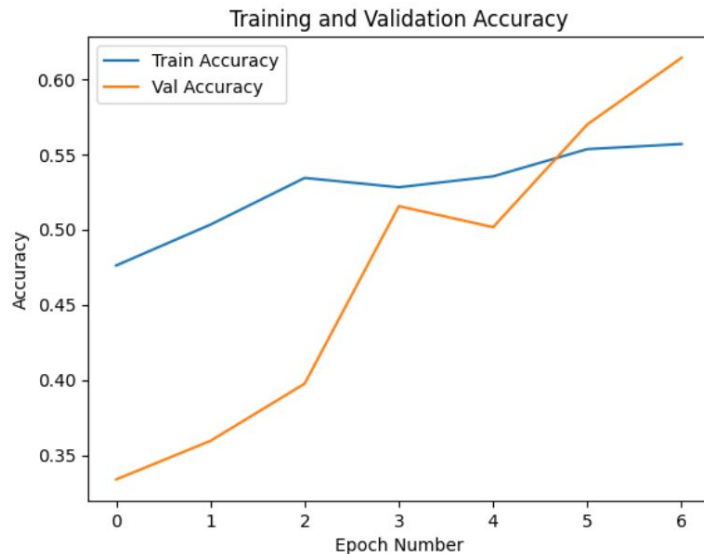
Training: Loss - 1.1264, Accuracy - 53.5317%,
Validation : Loss - 1.0424, Accuracy - 50.1553%, Time: 3192.3024s

Epoch: 6/7

Training: Loss - 1.1143, Accuracy - 55.3428%,
Validation : Loss - 1.0059, Accuracy - 56.9876%, Time: 3140.0257s

Epoch: 7/7

Training: Loss - 1.0866, Accuracy - 55.6792%,
Validation : Loss - 0.9817, Accuracy - 61.4130%, Time: 3086.1602s



Test Accuracy: 0.6137

Results on the subset of the four major artist

dropout = 0.4
learning rate = 1e-3

Epoch: 1/7

Training: Loss - 1.1918, Accuracy - 62.8461%,
Validation : Loss - 1.4361, Accuracy - 43.4006%, Time: 3228.9567s

Epoch: 2/7

Training: Loss - 1.3048, Accuracy - 55.1876%,
Validation : Loss - 1.3138, Accuracy - 33.5404%, Time: 3195.9910s

Epoch: 3/7

Training: Loss - 1.1698, Accuracy - 52.0052%,
Validation : Loss - 1.3318, Accuracy - 34.0062%, Time: 3073.2239s

Epoch: 4/7

Training: Loss - 1.1654, Accuracy - 43.4153%,
Validation : Loss - 1.2780, Accuracy - 50.0000%, Time: 3044.3162s

Epoch: 5/7

Training: Loss - 1.0796, Accuracy - 48.3829%,
Validation : Loss - 1.1280, Accuracy - 45.2640%, Time: 3018.2400s

Epoch: 6/7

Training: Loss - 1.0437, Accuracy - 43.0789%,
Validation : Loss - 1.1711, Accuracy - 45.3416%, Time: 3637.3680s

Epoch: 7/7

Training: Loss - 1.1354, Accuracy - 48.6934%,
Validation : Loss - 1.1371, Accuracy - 49.3789%, Time: 3583.2733s



Test Accuracy: 0.4818

Results on the subset of the four major artist

dropout = 0.5
learning rate = 1e-4

Epoch: 1/7

Training: Loss - 1.4563, Accuracy - 45.0970%,
Validation : Loss - 1.2425, Accuracy - 30.2795%, Time: 4776.4301s

Epoch: 2/7

Training: Loss - 1.2848, Accuracy - 49.3661%,
Validation : Loss - 1.1631, Accuracy - 36.4130%, Time: 4618.0030s

Epoch: 3/7

Training: Loss - 1.2488, Accuracy - 50.4787%,
Validation : Loss - 1.1224, Accuracy - 43.0124%, Time: 4739.1384s

Epoch: 4/7

Training: Loss - 1.1875, Accuracy - 53.8422%,
Validation : Loss - 1.0985, Accuracy - 46.8944%, Time: 4518.4628s

Epoch: 5/7

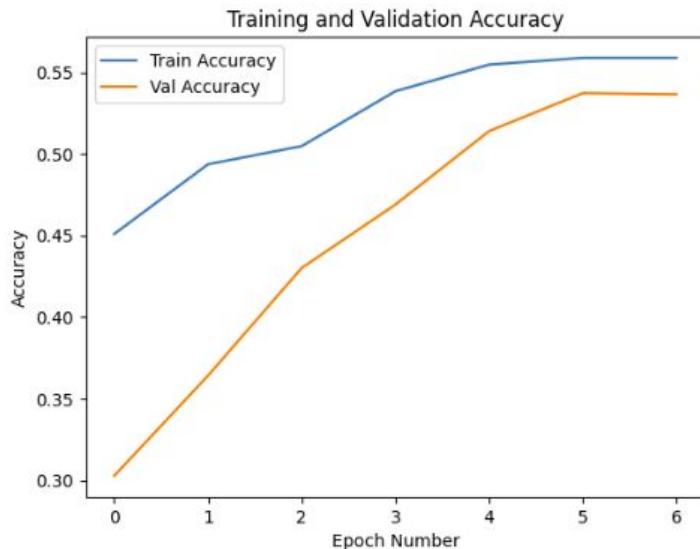
Training: Loss - 1.1650, Accuracy - 55.4722%,
Validation : Loss - 1.0595, Accuracy - 51.3975%, Time: 4548.2054s

Epoch: 6/7

Training: Loss - 1.1193, Accuracy - 55.8862%,
Validation : Loss - 1.0207, Accuracy - 53.7267%, Time: 4645.0653s

Epoch: 7/7

Training: Loss - 1.1165, Accuracy - 55.8862%,
Validation : Loss - 1.0182, Accuracy - 53.6491%, Time: 4600.5640s



Test Accuracy: 0.568

Results on the subset of the four major artist

dropout = 0.5
learning rate = 1e-3

Epoch: 1/7

Training: Loss - 1.2317, Accuracy - 62.5614%,
Validation : Loss - 1.3966, Accuracy - 31.3665%, Time: 4849.2197s

Epoch: 2/7

Training: Loss - 1.2383, Accuracy - 54.2044%,
Validation : Loss - 1.3367, Accuracy - 30.2795%, Time: 4085.3094s

Epoch: 3/7

Training: Loss - 1.1986, Accuracy - 33.8939%,
Validation : Loss - 1.2943, Accuracy - 29.5031%, Time: 4105.3934s

Epoch: 4/7

Training: Loss - 1.2286, Accuracy - 38.7322%,
Validation : Loss - 1.2708, Accuracy - 48.8354%, Time: 4082.0614s

Epoch: 5/7

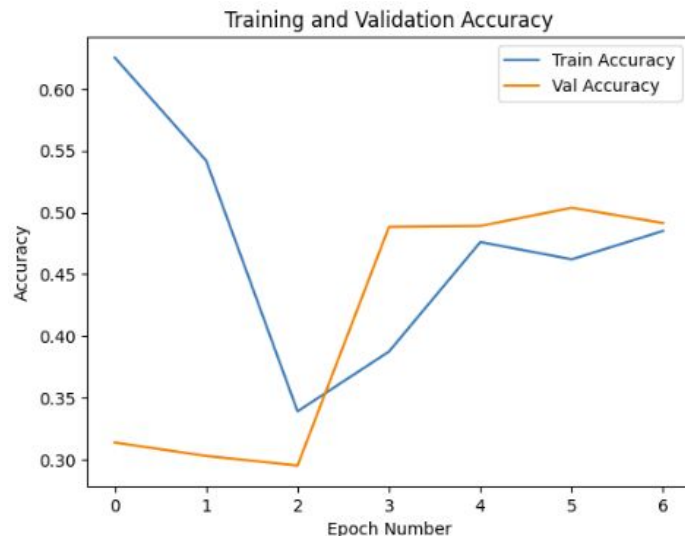
Training: Loss - 1.2173, Accuracy - 47.6067%,
Validation : Loss - 1.2313, Accuracy - 48.9130%, Time: 4094.9527s

Epoch: 6/7

Training: Loss - 1.1562, Accuracy - 46.2096%,
Validation : Loss - 1.2166, Accuracy - 50.3882%, Time: 4121.5588s

Epoch: 7/7

Training: Loss - 1.1394, Accuracy - 48.5123%,
Validation : Loss - 1.2301, Accuracy - 49.1460%, Time: 4118.6807s



Test Accuracy: 0.4732

Overall Structure of AlexNet

- ❖ Input Size: [32, 3, 224, 224]
- ❖ Total Parameters: 58,053,700
 - Trainable: 1,049,860
 - Non-trainable: 57,003,840
- ❖ Memory Size: 507.90 MB

Layer (type)	Output Shape	Param #
Conv2d-1	[32, 64, 55, 55]	23,296
ReLU-2	[32, 64, 55, 55]	0
MaxPool2d-3	[32, 64, 27, 27]	0
Conv2d-4	[32, 192, 27, 27]	307,392
ReLU-5	[32, 192, 27, 27]	0
MaxPool2d-6	[32, 192, 13, 13]	0
Conv2d-7	[32, 384, 13, 13]	663,936
ReLU-8	[32, 384, 13, 13]	0
Conv2d-9	[32, 256, 13, 13]	884,992
ReLU-10	[32, 256, 13, 13]	0
Conv2d-11	[32, 256, 13, 13]	590,080
ReLU-12	[32, 256, 13, 13]	0
MaxPool2d-13	[32, 256, 6, 6]	0
AdaptiveAvgPool2d-14	[32, 256, 6, 6]	0
Dropout-15	[32, 9216]	0
Linear-16	[32, 4096]	37,752,832
ReLU-17	[32, 4096]	0
Dropout-18	[32, 4096]	0
Linear-19	[32, 4096]	16,781,312
ReLU-20	[32, 4096]	0
Linear-21	[32, 256]	1,048,832
ReLU-22	[32, 256]	0
Dropout-23	[32, 256]	0
Linear-24	[32, 4]	1,028
LogSoftmax-25	[32, 4]	0

Results on the subset of the four major artist

dropout = 0.3
learning rate = 1e-5

Epoch: 1/7

Training: Loss - 1.0875, Accuracy - 59.6119%,
Validation : Loss - 0.7867, Accuracy - 77.6398%, Time: 3221.2956s

Epoch: 2/7

Training: Loss - 0.7388, Accuracy - 75.6533%,
Validation : Loss - 0.5945, Accuracy - 81.2112%, Time: 3075.2370s

Epoch: 3/7

Training: Loss - 0.6128, Accuracy - 78.6287%,
Validation : Loss - 0.5093, Accuracy - 83.6180%, Time: 3008.8806s

Epoch: 4/7

Training: Loss - 0.5459, Accuracy - 80.5692%,
Validation : Loss - 0.4621, Accuracy - 84.2391%, Time: 3077.2337s

Epoch: 5/7

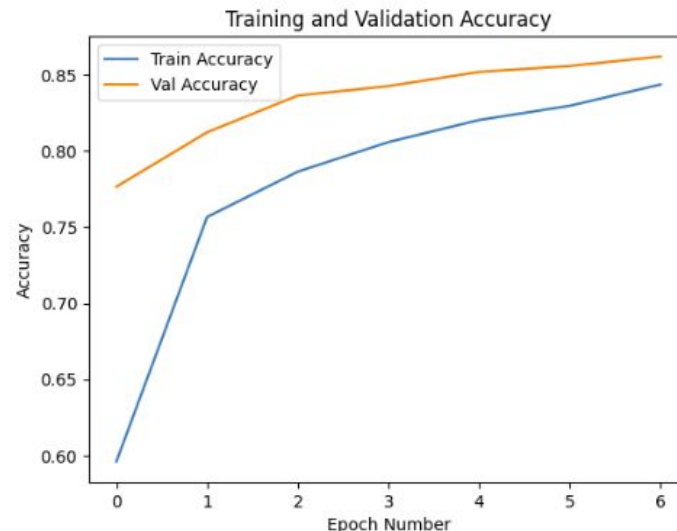
Training: Loss - 0.5041, Accuracy - 82.0181%,
Validation : Loss - 0.4321, Accuracy - 85.1708%, Time: 3058.9680s

Epoch: 6/7

Training: Loss - 0.4721, Accuracy - 82.9495%,
Validation : Loss - 0.4100, Accuracy - 85.5590%, Time: 3038.1487s

Epoch: 7/7

Training: Loss - 0.4378, Accuracy - 84.3467%,
Validation : Loss - 0.3943, Accuracy - 86.1801%, Time: 3041.4791s



Test Accuracy: 0.8604

Results on the subset of the four major artist

dropout = 0.4
learning rate = 1e-4

Epoch: 1/7

Training: Loss - 0.6472, Accuracy - 76.7400%,
Validation : Loss - 1.0935, Accuracy - 57.4534%, Time: 3767.8167s

Epoch: 2/7

Training: Loss - 0.4966, Accuracy - 81.5783%,
Validation : Loss - 0.7836, Accuracy - 67.4689%, Time: 3798.0113s

Epoch: 3/7

Training: Loss - 0.4108, Accuracy - 85.0453%,
Validation : Loss - 0.6683, Accuracy - 71.9720%, Time: 3804.2353s

Epoch: 4/7

Training: Loss - 0.3767, Accuracy - 86.6235%,
Validation : Loss - 0.5631, Accuracy - 77.4068%, Time: 3813.0985s

Epoch: 5/7

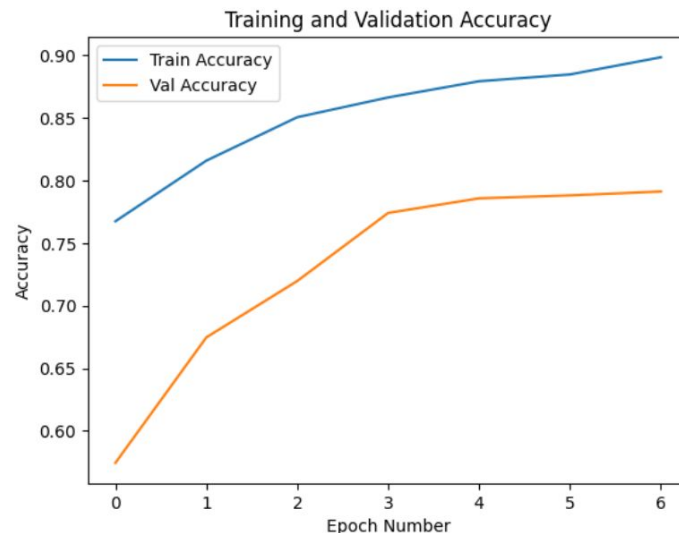
Training: Loss - 0.3237, Accuracy - 87.9172%,
Validation : Loss - 0.5372, Accuracy - 78.5714%, Time: 3782.7887s

Epoch: 6/7

Training: Loss - 0.3078, Accuracy - 88.4605%,
Validation : Loss - 0.5461, Accuracy - 78.8043%, Time: 3779.8250s

Epoch: 7/7

Training: Loss - 0.2777, Accuracy - 89.8318%,
Validation : Loss - 0.5314, Accuracy - 79.1149%, Time: 3789.7982s



Test Accuracy: 0.7913

Results on the subset of the four major artist

dropout = 0.4
learning rate = 1e-3

Epoch : 1/7

Training: Loss - 0.6654, Accuracy - 79.3014%,
Validation : Loss - 4.1417, Accuracy - 34.2391%, Time: 4469.5030s

Epoch: 2/7

Training: Loss - 0.5596, Accuracy - 82.8461%,
Validation : Loss - 2.4305, Accuracy - 50.0000%, Time: 4508.7952s

Epoch: 3/7

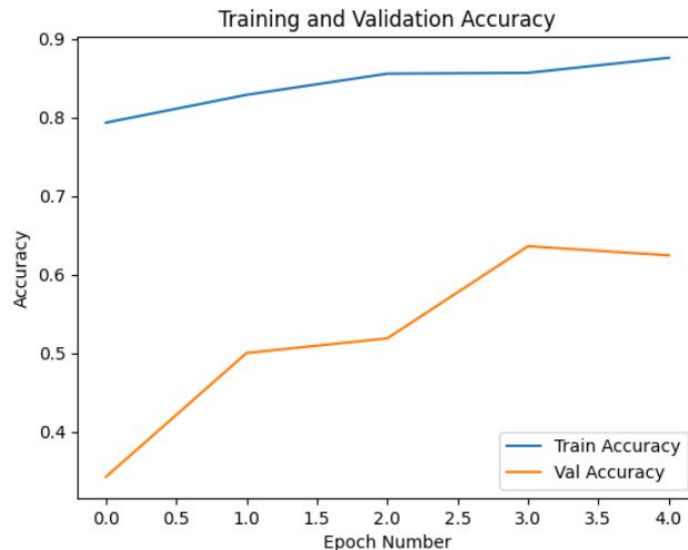
Training: Loss - 0.4551, Accuracy - 85.5369%,
Validation : Loss - 2.0161, Accuracy - 51.8634%, Time: 4623.2712s

Epoch: 4/7

Training: Loss - 0.4237, Accuracy - 85.6404%,
Validation : Loss - 1.1818, Accuracy - 63.5870%, Time: 4600.9963s

Epoch: 5/7

Training: Loss - 0.3705, Accuracy - 87.5550%,
Validation : Loss - 1.2605, Accuracy - 62.4224%, Time: 4528.5422s



Results on the subset of the four major artist

dropout = 0.5
learning rate = 1e-4

Epoch: 1/7

Training: Loss - 0.6961, Accuracy - 74.9288%,
Validation : Loss - 1.0289, Accuracy - 59.7050%, Time: 5832.2118s

Epoch: 2/7

Training: Loss - 0.4941, Accuracy - 81.9664%,
Validation : Loss - 0.6926, Accuracy - 71.6615%, Time: 4417.2765s

Epoch: 3/7

Training: Loss - 0.4264, Accuracy - 84.4761%,
Validation : Loss - 0.6105, Accuracy - 75.0000%, Time: 4302.6227s

Epoch: 4/7

Training: Loss - 0.3871, Accuracy - 86.4424%,
Validation : Loss - 0.5729, Accuracy - 76.8634%, Time: 4345.6633s

Epoch: 5/7

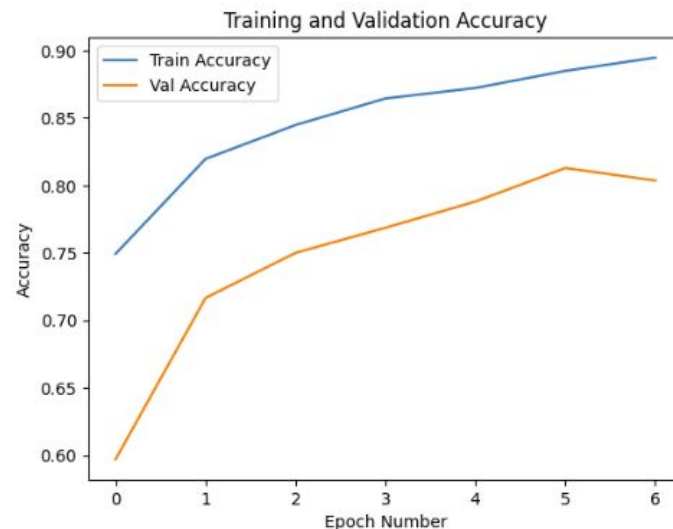
Training: Loss - 0.3464, Accuracy - 87.2186%,
Validation : Loss - 0.5099, Accuracy - 78.8043%, Time: 4442.9466s

Epoch: 6/7

Training: Loss - 0.3205, Accuracy - 88.4864%,
Validation : Loss - 0.4771, Accuracy - 81.2888%, Time: 4413.4108s

Epoch: 7/7

Training: Loss - 0.2931, Accuracy - 89.4696%,
Validation : Loss - 0.4952, Accuracy - 80.3571%, Time: 4534.5440s



Test Accuracy: 0.8076

Results on the subset of the four major artist

dropout = 0.5
learning rate = 1e-3

Epoch: 1/7

Training: Loss - 0.6605, Accuracy - 79.0944%,
Validation : Loss - 4.1712, Accuracy - 36.5683%, Time: 4878.0181s

Epoch: 2/7

Training: Loss - 0.5705, Accuracy - 82.3803%,
Validation : Loss - 2.4512, Accuracy - 49.7671%, Time: 4632.6517s

Epoch: 3/7

Training: Loss - 0.4693, Accuracy - 84.4502%,
Validation : Loss - 1.5655, Accuracy - 56.7547%, Time: 4474.2122s

Epoch: 4/7

Training: Loss - 0.4193, Accuracy - 85.2523%,
Validation : Loss - 1.2648, Accuracy - 60.9472%, Time: 4450.4920s

Epoch: 5/7

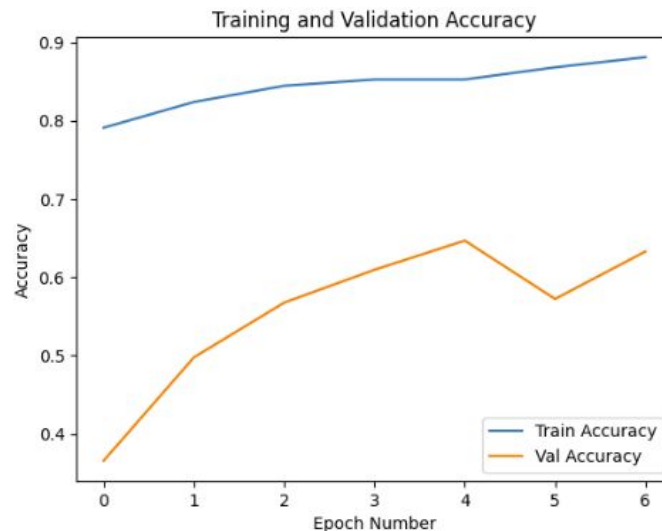
Training: Loss - 0.3942, Accuracy - 85.2523%,
Validation : Loss - 1.2837, Accuracy - 64.6739%, Time: 4419.6420s

Epoch: 6/7

Training: Loss - 0.3799, Accuracy - 86.8047%,
Validation : Loss - 1.9536, Accuracy - 57.2205%, Time: 4515.2084s

Epoch: 7/7

Training: Loss - 0.3666, Accuracy - 88.0983%,
Validation : Loss - 1.4052, Accuracy - 63.2764%, Time: 4609.5027s



Test Accuracy: 0.6330

Conclusions

❖ Best accuracy using ResNet18:

- Dropout = 0.3
- Learning rate = $1e-5$
- Test accuracy = 0.6175
- Average time per epoch ≈ 3001.7386 seconds

❖ vs ResNet50:

- Dropout = 0.3
- Learning rate = $1e-5$
- Test accuracy = 0.6067
- Average time per epoch ≈ 3208.7387 seconds

❖ Best accuracy using AlexNet:

- Dropout = 0.3
- Learning rate = $1e-5$
- Test accuracy = 0.8604
- Average time per epoch ≈ 3074.4632 seconds

Note: all these results
have been computed
using GPU T4x2