Artist Style Classification using Transfer Learning

Overview:

Artist identification is traditionally performed by art historians and curators who have expertise with different artists and styles of art. This is a complex and interesting problem for computers because identifying an artist does not just require object or face detection; artists can paint a wide variety of objects and scenes. Additionally, many artists from the same time will have similar styles, and some artists have painted in multiple styles and changed their style over time. Previous work has attempted to identify artists by explicitly defining their differentiating characteristics as features. Instead of hand-crafting features, we use transfer learning where a model trained on one set of data is employed and adapted to our dataset of choice.

Objectives:

1. Train neural networks using transfer learning to obtain better artist identification performance compared to traditional SVM classification [1]
2. Explore and visualize the learned feature representation for identifying artists

Related Work:

Li et al. developed a method of automatically extracting brushstrokes by combining edge detection and clustering-based segmentation [1]. Jou et.al explored the use of k-nearest neighbours and hierarchical clustering for artist identification [3]. However there has not been notable exploration using CNNs for artist identification.

Dataset, Pre-Processing and Data Augmentation:

For this work we use a subset of the dataset compiled by Kaggle that is based on the WikiArt dataset for a balanced pool of artists and styles. Since, art comes in a variety of shapes and sizes we perform pre-processing before passing the image to the CNN. We reprocess the images to be zero mean and unit standard deviation before training. Then we take a 224x224 crop of each input image and randomly flip the input image horizontally. Our hypothesis is that artist style is present everywhere in an image and not limited to specific areas, so crops of paintings should still contain enough information for a CNN to determine style.

Methods:

We develop and train five different CNN architectures using transfer learning for artist identification. In this project, we will consider two types of transfer learning: a feature-extraction based method and a fine-tuning-based method. We will be using networks that have been pre-trained on the ImageNet dataset and adapt them for our datasets.

1. **Fine-tuning of AlexNet, VGG16 and ResNet18:** Fine-tuning is aimed to adapt the existing filters to our data, but not move the parameters so far from the pre-trained parameters. We start with a pre-trained network to test whether or not a feature representation from ImageNet is a valuable starting point for artist identification. Some artists, for example Renaissance painters, used shapes and objects that you would expect to find in ImageNet since they usually painted lifelike scenes. However, other artists such as Cubists did not paint scenes as directly representative of the real world.
2. **Feature Extraction +SVM on AlexNet and VGG: We** will use the base network as a feature extractor. This means that we simply run the images through the pre-trained base network and take outputs from layer(s) as a feature representation of the image. These

features are then used for classification using SVM. All our models are implemented in PyTorch. All experiments were implemented in Amazon Web Services using a machine with 4vCPUs, p2xlarge instance and 60GB storage.

Quantitative Results:

The performance of the model was evaluated based on the top-1 classification accuracy (the fraction of paintings whose artists are identified correctly), we compare our networks against each other as well as against [1], and the results are given below:

MODEL	Train Accuracy	Test Accuracy
Baseline SVM[1]	(Not reported)	0.58
Alexnet-Feature Extraction	0.78	0.60
VGG16 -Feature Extraction	0.79	0.64
Alexnet-Finetune	0.87	0.70
VGG16-Finetune	0.91	0.71
Res-Net	**0.93**	**0.74**

Table 1: Accuracy results for various models test dataset

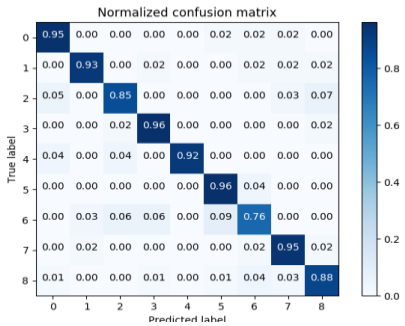

Fig 1: Confusion matrix for top-1 classification accuracy on the

Qualitative Results

Saliency maps allow us to visualize which pixels in an image contribute most to the predicted score for that image. We examined saliency maps of a few paintings and saw that in most but not all, the important pixels were spread all over the image and not focused around objects or people in them. Thus, evidently the network does not focus on any single area of the image to perform classification.

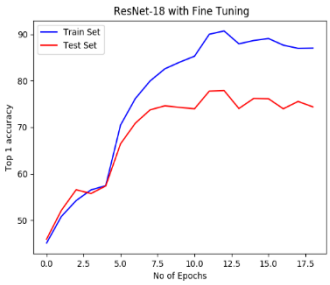

Fig.2: Accuracy vs epoch for training and test set

Conclusions

The transfer learning-based networks performed better than the feature extraction method traditionally used. Extensive tuning of the hyper parameters led to maximised performance on the ResNet that was pre-trained on ImageNet. Also extracting the features before the last fully connected layer and then performing classification yields lower performance than fine-tuning. Also, the high probability along the diagonal of the confusion matrix is indicative of the high accuracy in classification. By viewing the saliency maps of the images we observe the network does not focus on a single area of the image to perform classification.

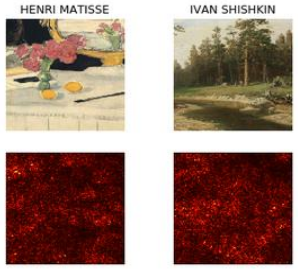

Fig.3: Saliency Map

References

[1] E. H. J. Li, L. Yao and J. Z. Wang. Rhythmic brushstrokes distinguish van gogh from his contemporaries: Findings via automated brushstroke extractions. IEEE Trans. Pattern Anal. Mach. Intell., 2012.

[2] J. Jou and S. Agrawal. Artist identification for renaissance paintings.

[3]B. Saleh and A. M. Elgammal. Large-scale classification of fine-art paintings: Learning the right metric on the right feature. CoRR, abs/1505.00855, 2015.