Deep Learning for Art Characterization

Niraj Sonje, Dr. Rupesh C. Jaiswal, Dr. Girish P. Potdar, Prof. Manish R. Khodaskar

1Student, 2Professor, 3Professor, 4Professor

1,2Department of E & Telecommunication Engineering,
3. Department of Computer Engineering,
4. Department of IT Engineering,
1,2,3,4SCTR’s Pune Institute of Computer Technology, Pune, India.

ABSTRACT

Art in the form of paintings have been fascinating humans since ages. With the digitization of art museums, it becomes imperative to devise tools that can automatically classify artists based on paintings to help visitors, buyers and museum curators gain more information effortlessly. Image classification models, in general, aim at extracting distinct information from images and assign them a particular category. In this paper we use transfer learning to fine tune the existing CNN models such as ResNet, that has been pre-trained on ImageNet, and apply multiple different data augmentation techniques to tackle the challenge of classifying 50 artists. We intend to observe and analyse the hyper-parameters and augmentation technique giving the best results. We also make some attempts of inventing new architecture of neural network dedicate to painting understanding and new approach for data augmentation.

1. INTRODUCTION

Convolutional Neural Networks, in recent years, have proved to achieve state of the art results on certain target-oriented tasks such as image classification, object recognition, segmentation. CNNs have found applications in aerial imaging, autonomous mobility, robotics, facial recognition etc. The power of CNNs to learn filters similar to what a human eye learns is startling and has led researchers to develop award winning architectures. One such example is the ResNet[8] developed in 2015, and has been highly praised for its architecture that achieved state of the art results at classifying images for 1000 classes at the famous ImageNet challenge. This paper uses the idea of applying CNNs on an art classification problem. Paintings and its different forms have been a part of art history and there hasn’t been much research carried out for this particular application. We aim to classify paintings of about 50 renowned artists from the Best Artworks of All Time [7] dataset. We introduce data augmentation techniques such as step-crop and diagonal crop and use them to improve results. We use accuracy and confusion matrix as the evaluation metric and observe the performance of the different ResNet[8] models on our dataset. Besides that, according to the features of artwork, we introduced a new idea about tackling this challenge via multiple-channel images and new model based on Xception Net.
2. BACKGROUND

CNNs have been previously applied to the problem of art classification [cite the papers]. Balakrishan et al., in [1] use transfer learning to fine tune VGG and ResNet CNN models to classify 20 artists and achieve accuracy of 87%. They observe that the ResNet performs better than VGG. Other work such as [2] predicts the style of the painting of 20 artists, achieving highest accuracy of 63.7% on the pretrained and fine-tuned for ResNet model. In addition to this, the authors mention that they study the correlation between the modes of variation and Wolfflin’s suggested pairs to find that only a small number of factors can explain most of the variance in the data. Cetinic et al., in [4] obtain features of paintings for the purpose of genre classification, analyse different features and experiment with various classifiers. In conclusion, they mention that CNN extracted features outperform all other image feature types for genre classification task. Also, they further mention that the best classification accuracy they achieved was by using SVM with RBF kernel. On analysing the above contribution’s, we use data augmentation as a part of data pre-processing and use transfer learning fine tune the CNN models. Also, role of ML and ESPs [12-73] are becoming important in recent applications, recognition and control.

For hyperchannel, [10] utilized multiple-channel images to obtain motion information and object detection, where multiple-channel images consist of raw, motion, MWR image. [11] proposed a methodology is converting raw data into spatial-temporal multiple-channel matrices and using a designed CNN to speed traffic speeds.

3. METHODS

3.1. Data Augmentation:

Data augmentation has efficiently been used as a data pre-processing technique that can increase the existing number of images for training. We separated the data into train and validations sets.

Horizontal Flip: As a part of our first data augmentation technique, as seen in Figure 1., we horizontally flipped the image and used a zoom of 0.5 on our data, further using transfer learning to train the models.

Diagonal Crop: In this data augmentation method, the original image is cropped along the diagonal into “n” small images and these cropped images are zoomed in with a zoom range of 3x. The resulting zoomed in patches are stitched onto the original image such that there are “n” patches along the last column and “n-1” patches along the last row. Then that image is resized to 224x224 in order to match the input size of our models. Figure 2. shows the original, diagonal crop with zoom and the new generated stitched image.
Step Crop: In this method, the original image is cropped along the rows into smaller images with a step in between and these cropped patches of images are zoomed in with a zoom range of 3x and stitched onto the original image similarly as described in the diagonal crop method. Figure 3. shows the original, step-crop with zoom and the newly generated stitched image.

3.2. Transfer Learning:
We add 2 dense layers and a batch normalization layer after each of the dense layer. We freeze the rest of the layers and train only on the top layers. To transfer learn we used the ‘ImageNet’ pre-trained weights. The hyper-parameters we used for training were as follows: a) epochs = 30, b) optimizer = Adam, c) learning rate = 1e-4.

3.3. Multiple channel:
To let neural network focus on brushed strokes and lines, we manipulated images by cropping the paintings at the upper right, centre and bottom right. Then zoom in them three times to obtain more detailed information. All of them along with the original image was resize to 224 which is the most adopted input size of popular neural networks and was concatenated them together to save in tiff file.

3.4. Network Architecture:
The first method is 21 channels images was fed into network together. Since the first three channel include all content of the original image. More information is expected to get from here. To tackle this, we specify the number of output channel of each input channel, rather than specifying the number of output channel of all channel in a lump. Therefore, it needs to concatenate them together after every convolution until the end. The rest parameters and pooling way keep
the same as Xception net. Since some unrelated images are combined to together, machine may get confused when it carry out 1x1 pointwise convolution which calculate all pixel in depth wise to generate new information. This’s where the second method comes in. The second method was designed to convoluted through images and cropped images on diagonal by utilizing parallel network. Since the centre of a painting usually contains richer information than other regions of a painting, the small and large cropped image at centre region was picked. Along with Original image were fed into Xception net separably, the extracted feature map was sent to several combination of the fully connection layer and the dropout layer, finally it performs sigmod activation to get final prediction.

4. RESULTS

4.1. Dataset:

We have used the 'Best artworks from all time' dataset [7]. It has paintings in the form of images for 50 different artists. While training the model for individual data augmented technique we had 6735 training images and 1711 validation images and while collectively using the (Original + Step crop + Diagonal crop) images, we had 20325 training and 5177 validation images.
### Table 1. Results Summary

<table>
<thead>
<tr>
<th>Experiments</th>
<th>Model Used</th>
<th>Validation Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal Flip</td>
<td>ResNet 101</td>
<td>69.75</td>
</tr>
<tr>
<td>Step Crop</td>
<td>ResNet 101</td>
<td>73.00</td>
</tr>
<tr>
<td>Diagonal Crop</td>
<td>ResNet 101</td>
<td>74.55</td>
</tr>
<tr>
<td>Original + Step Crop + Diagonal Crop</td>
<td>ResNet 101</td>
<td>76.70</td>
</tr>
<tr>
<td>Horizontal Flip</td>
<td>EfficientNet B5</td>
<td>71.83</td>
</tr>
<tr>
<td>Step Crop</td>
<td>EfficientNet B5</td>
<td>72.07</td>
</tr>
<tr>
<td>Diagonal Crop</td>
<td>EfficientNet B5</td>
<td>71.00</td>
</tr>
<tr>
<td>Original + Step Crop + Diagonal Crop</td>
<td>EfficientNet B5</td>
<td>74.76</td>
</tr>
</tbody>
</table>

4.2. Experimental Analysis

We perform experimental analysis using the data augmentation techniques on ResNet-101 and EfficientNet-B5 [9] models available in Keras [6]. Table 1. lists the results summary for both these models. We observe best performance for ResNet model using the original, step-crop and diagonal crop images with a validation accuracy of 76.7 % improving it by 7% from the basic horizontal flip technique.

For, new proposed neural network, after several trails and experiments, both method failed to outperform the performance of existing classification methods. They all suffered from overfitting issue. Different optimizers, he initializer and l2 regularizer was used though, it still presented a hard time of converging.

**CONCLUSION**

We showed how data augmentation can be helpful in achieving better validation accuracy while addressing image classification problems such as painting classification. Using our proposed data augmentation techniques, the model learns intricate details from the paintings such as similarity in the paintings of a particular artist, brush strokes. We observe that when multiple augmentation techniques are used together along with the original data, we achieve significant improvement in validation accuracy.

**ACKNOWLEDGEMENT**

I would like to thank all the members who have helped me in completing this research work and paper. I express my heartiest gratitude to Dr. Rupesh Jaiswal Sir, Dr. G.P. Potdar Sir and Prof. M.R. Khodaskar Sir for providing me all guidance to complete our research work. Finally last but not least we would like to thank all our friends and family members and other all members who have directly or indirectly contributed in successful completion of this research work.


