Neural Style Transfer

Srishti Sharma[1] and Bhumiti Gohel[2]

Abstract— **The process of taking a content and a style image as input and outputting an image that has the content same as content image along with style of the style image is known as Neural Style Transfer. This is made possible with the help of a popular deep learning algorithm known as convolutional neural network(CNN). This paper will start by applying major techniques for neural style transfer using VGG16 and VGG19 pretrained models.**

Keywords– Neural Style Transfer, VGG16, VGG19, Data Augmentation, Convolutional Neural Networks

I. Introduction

In layman's language, Neural Style Transfer can be said as an art of creating images with any style applied to any content. Content can be recognized as the layout or sketch and Style is the painting or colors. It can also be said as Image Transformation using Deep Learning. Almost every artist must have dreamed of painting his ideas in the style of famous painter, Picasso. With the help of deep learning that is possible now and termed as Neural Style Transfer.[1]Several humans have mastered the skill of creating unique visual experiences In the domain of painting by composing a complex interplay between the content and style of the image. There doesn't exist an artificial system that can perform this thing. Hence, with recent advances in deep learning, we introduce an artificial system that creates artistic images that are high perceptual in quality. This system used neural representations to separate and recombine content and style of arbitrary images. This also provides a neural algorithm for creation of artistic images. Along with optimization in performance, we offer a path forward which helps humans in perceiving images of different context and art.[2] The base idea on which Neural Style Transfer is proposed is "it is possible to separate the style representation and content representations in a CNN, learned during a computer vision task (e.g. image recognition task)."[2] The state of the art models in this domain includes AlexNet, VGG and RESNET and many others. One thing common about them is that the training of all these models is done on a ImageNet dataset containing 14 million images with 1000 classes which makes them very efficient. Along with this, we increase the quality of these models by segregating the content and style part and providing optimized results with minimum loss. The principle of neural style transfer is to define two distance functions, one that describes how different the content of two images are, Lcontent, and one that describes the difference between the two images in terms of their style, Lstyle.Then, given three images, a desired style image, a desired content image, and the input image (initialized with the content image), we try to transform the input image to minimize the content distance with the content image and its style distance with the style image.we define a pre-trained convolutional model and loss functions which blends two images visually, therefore we would be requiring the following inputs : (1) A Content Image – image on which we will transfer style. (2) A Style Image – the style we want to transfer. (3) An Input Image(generated) – The final content plus the required style image

II. Literature Review

Convolutional Neural Networks have a wide applicability in the field of image analysis and classification purposes of images. Remedies to the poor generalization abilities of the deep neural networks are dropout, batch normalization, transfer learning, data augmentation, early stopping and weight decay [3]. VGGNet are CNN architectures

[1]Srishti Sharma, AU2049002, PhD in Engineering, School of Engineering and Applied Sciences

[2]Bhumiti Gohel,AU1841051, Department of Information and Communication Technology

that apply various cropping and flipping operation over each of the training images in order to boost the size of the training set and to improve the model performance [4]. There architectures of VGGNet gained popularity after demonstrating their effectiveness through excellent results in the ImageNet challenge [5]. In [3], the prime focus was using style transfer for data augmentation for synthesizing a skin lesion dataset. To synthesize the images, neural style transfer was used implementing VGG16 [6] with input image size 224*224. In [6], so as to add more variation to a dataset, neural style transfer was used as a data augmentation technique as it helps applying artistic style to an image without changing its high level semantics. They applied neural style transfer on Caltech101 [7] and Caltech256 [8] dataset and showed an increase in classification accuracy from 83.34% to 85.26% using VGG16 and an increase from 84.50% to 85.81% using VGG19.

Prisma photo editing application has around 250 modern filters and around 110 million users. Since there are a huge amount of images that are being synthesized, CNN architectures such as VGG16, VGG19 are introduced in order to solve the heavy computation problems [9]. In [10], an efficient way of classifying flowers using convolutional neural networks and neural style transfer is proposed. Classification task using models such as VGG16, VGG19, Inception-V3 and ResNet50 avoid overfitting issue and increase the generalization ability thereby increasing the accuracy. This experiment was conducted over the Oxford flowers dataset. The effects of neural style transfer learning were studied in the context of automobile detection under adverse condition application. This task is a tough one due to the large amount of noise in extreme weather conditions. The car images dataset was applied neural style transfer using VGG16, VGG19 and InceptionNet and the effectiveness of each along with the advantages and disadvantages of use of style transfer learning for data augmentation were discussed [11].

III. Architecture

For implementation, we have made use of VGG16 and VGG19 architecture which is a pre-trained image classification network.These CNN architectures are improvements over AlexNet wit kernel-sized filters (11 filters in layer one and 5 filters in layer two) getting replaced with multiple 3×3 kernel-sized filters one after another [14].

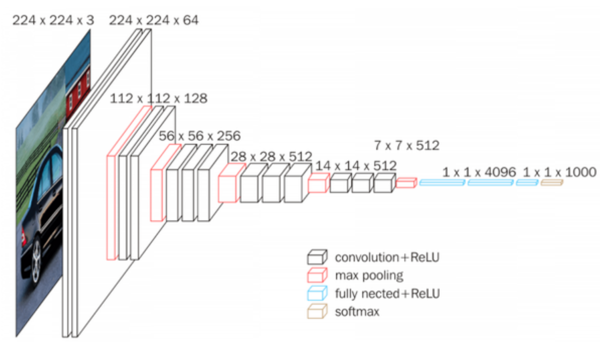

Fig. 1. VGG16 [12]

In this deep neural network, the initial few layers starting from the network's input layer generate activations that represent the features such as edges, texture and other low level features while the final layers generate activations representing features such as eyes, nose, car wheel and other high level features. The input is 224*224 RGB image to Conv1 layer [14]. This is passed through multiple receptive layers with size 3*3 so as to capture the smallest features. 1*1 convolution filters are used for performing dimensionality reduction. Stride 1 and padding of 1 pixel is used. There is spatial pooling of 5 max pool layers performed on a 2*2 window with stride 2.

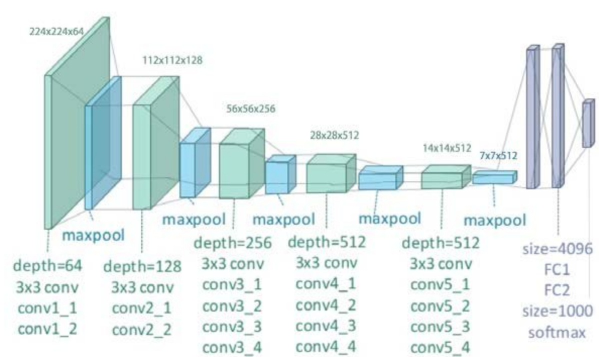

Fig. 2. VGG19 [13]

There are three fully connected layers wherein the first one has 4096 channels and second one has 1000 channels respectively. Third layer is the Softmax layer. Each hidden layer is equipped with a rectification (ReLU) non linearity [14]. It is the intermediate layers that help defining the content and style of an image.While there are 16 deep neural layers in VGG16, there are 19 dense layers in VGG19.

IV. Implementation

In general, neural style transfer was implemented following a certain steps. Initially there were two images taken, one being the content image and the other one being the style image. Content image was the one to be styled artistically using the style image. These images were first loaded and their dimension was restricted to 512 pixels. Next, the VGG16 and VGG19 models were respectively loaded and their intermediate layers were selected so as to represent the style and content. While first few layers represented low level features and last few layers represented high level features, intermediate layers were used. These were selected as for image classification at high level, understanding image and drawing an internal representation that would take raw pixels of image and convert them to high level complex features was needed. The VGG16 or VGG19 model is next built which returns a list of intermediate layers.

Now, the content image is represented by values of these intermediate feature maps. The style can be represented by means and correlation across these various feature maps. Gram matrix, which is the outer product of feature vector with itself at every location and an average of this outer product over all of these locations is calculated as follows:

$$G^l_{cd} = \frac{\sum_{ij} F^l_{ijc}(x) F^l_{ijd}(x)}{IJ}$$

The style and content are then extracted and the style transfer algorithm is run. Mean square error of image's output relative to the target is compared and a sum of losses is taken. Adam optimizer is used for optimization.

$$R_{V_\beta}(\mathrm{y}) = \lambda_{V_\beta} \sum_{i,j,k} \left(\left(y_{i,j+1,k} - y_{i,j,k} \right)^2 + \left(y_{i+1,j,k} - y_{i,j,k} \right)^2 \right)^{\beta/2}$$

Fig. 3. Total Variation Regulizer [14]

Variation loss is next computed and the optimization is again performed. Initializing the optimization variable with weights of variation loss, the style transfer algorithm is re run and resulting target image is generated.

V. Results

Image 1: Yellow Labrador 700 × 577 pixels

Fig. 4. Content Image 1

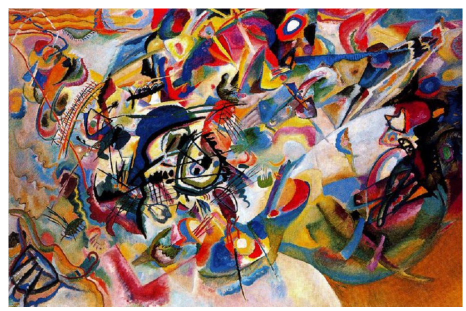

Fig. 5. Style Image 1

Fig. 6. Augmented Image VGG16

Image 2: Paris 910 × 607 pixels

Fig. 7. Augmented Image VGG19

Fig. 8. Content Image 2

Fig. 9. Style Image 2

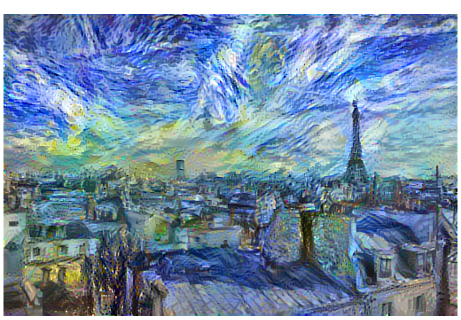

Fig. 10. Augmented Image VGG16

Fig. 11. Augmented Image VGG19

VI. Conclusion

Making use of neural style transfer techniques for data augmentation helps improving the accuracy of image classification. VGG16 and VGG19 are pre-trained image classification deep neural networks that reduce the complexity and computational overhead incurred in training a specific deep neural network for classification problem. The requirement of a very less size dataset for training and faster convergence makes VGG16 and VGG19 favourable for data augmentation for accurate image classification.

VII. References

1 . Luan, Fujun, et al. "Deep photo style transfer." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. 2. Nikulin, Yaroslav, and Roman Novak. "Exploring the neural algorithm of artistic style." arXiv preprint arXiv:1602.07188 (2016). 3. Mikołajczyk, Agnieszka, and Michał Grochowski. "Style transfer-based image synthesis as an efficient regularization technique in deep learning." 2019 24th International Conference on Methods and Models in Automation and Robotics (MMAR). IEEE, 2019.

4. Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

5. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 248–255. Ieee.

6. Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "A neural algorithm of artistic style." arXiv preprint arXiv:1508.06576 (2015).

7. Fei-Fei, L., Fergus, R., and Perona, P. (2006). One-shot learning of object categories. IEEE transactions on pattern analysis and machine intelligence, 28(4):594– 611
8. Griffin, G., Holub, A., and Perona, P. (2007). Caltech-256 object category dataset.
9. Vishwakarma, Dinesh Kumar. ”A State-of-the-Arts and Prospective in Neural Style Transfer.” 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN). IEEE, 2019.
10. Wu, Yong, et al. ”Convolution neural network based transfer learning for classification of flowers.” 2018 IEEE 3rd international conference on signal and image processing (ICSIP). IEEE, 2018.
11. Xu, Yijie, and Arushi Goel. ”Cross-Domain Image Classification through Neural-Style Transfer Data Augmentation.” arXiv preprint arXiv:1910.05611 (2019).
12. Rohit Thakur, ”Step by step VGG16 implementation in Keras for beginners”, Towards Data Science, August 2019.
13. Zheng, Yufeng, Clifford Yang, and Alex Merkulov. ”Breast cancer screening using convolutional neural network and follow-up digital mammography.” Computational Imaging III. Vol. 10669. International Society for Optics and Photonics, 2018.
14. Shaha, Manali, and Meenakshi Pawar. ”Transfer learning for image classification.” 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE, 2018.