

## Research Article

# Deep Convolutional Nets Learning Classification for Artistic Style Transfer

**R. Dinesh Kumar,<sup>1</sup> E. Golden Julie,<sup>2</sup> Y. Harold Robinson ,<sup>3</sup> S. Vimal,<sup>4</sup> Gaurav Dhiman ,<sup>5</sup> and Murugesh Veerasamy <sup>6</sup>**

<sup>1</sup>CSE Department, Siddhartha Institute of Technology and Science, Hyderabad, India

<sup>2</sup>Department of Computer Science and Engineering, Anna University Regional Campus, Tirunelveli, India

<sup>3</sup>School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India

<sup>4</sup>Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, India

<sup>5</sup>Department of Computer Science, Government Bikram College of Commerce, Patiala, Punjab, India

<sup>6</sup>Department of Computer Science, College of Computer Science, Bule Hora University, Blue Hora, Ethiopia

Correspondence should be addressed to Murugesh Veerasamy; [murugesh@bhu.edu.et](mailto:murugesh@bhu.edu.et)

Received 25 June 2021; Accepted 16 December 2021; Published 10 January 2022

Academic Editor: Antonio J. Peña

Copyright © 2022 R. Dinesh Kumar et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Humans have mastered the skill of creativity for many decades. The process of replicating this mechanism is introduced recently by using neural networks which replicate the functioning of human brain, where each unit in the neural network represents a neuron, which transmits the messages from one neuron to other, to perform subconscious tasks. Usually, there are methods to render an input image in the style of famous art works. This issue of generating art is normally called nonphotorealistic rendering. Previous approaches rely on directly manipulating the pixel representation of the image. While using deep neural networks which are constructed using image recognition, this paper carries out implementations in feature space representing the higher levels of the content image. Previously, deep neural networks are used for object recognition and style recognition to categorize the artworks consistent with the creation time. This paper uses Visual Geometry Group (VGG16) neural network to replicate this dormant task performed by humans. Here, the images are input where one is the content image which contains the features you want to retain in the output image and the style reference image which contains patterns or images of famous paintings and the input image which needs to be style and blend them together to produce a new image where the input image is transformed to look like the content image but “sketched” to look like the style image.

## 1. Introduction

A decade ago, when machine learning was an emerging application of artificial intelligence providing the ability to automate learning process from foregoing experiences without being explicitly programmed, the only limitation was assumed that a good computer program can never replace a human in creativity [1]. But as the exploration in the field grew, this gave rise to many other subfields like deep learning, which threw the limelight on the solution for replacing humans for their creativity or their process of recognizing objects or people [2]. One of such problems

which characterized human from a machine was art. Generating art has no rules which could be used to replicate a man’s imagination. This paper renders an input image in the range of well-known art works [3]. This is conceptually close to texture transfer for style transfer. Today, there are few existing systems which replicate art from famous painters. One of these methods is style transfer using a neural network [4].

A neural network can be defined as a circuit of neurons which stimulates the behavior of a human brain. Mathematically, a neural network is a chain of functions which maps the inputs with their respective outputs based on their

weights, which defines the amount of influence one function has on the other (where each function represents a neuron) [5]. A neural network engages a larger amount of functions operating similar layers. One layer generates the input and every adjacent layer gathers the output from the previous layers. The final layer generates the output of the system, which is also known as visible layer. As the layers other than the input and output layers are not accessible, they are known as hidden layers [6].

The transformation of the style from one image to another image could be called the texture transfer issue. The aim of doing this is to manufacture an element from an input image and also gather the semantic component to the output image [7]. Several texture transfer methods used the texture synthesis functions of using dissimilar way to conserve the targeted output image. The features of the output image like intensity is computed for effective texture transfer. The frequency related texture data for preserving the target image combined the edge orientation data. A style transfer is utilized to extort the contents from the image for implementing the texture transfer. Hence, a normal procedure is used to analyze the image independent representation to observe the semantic contents and style related structures were demonstrated to control the subset of natural image characterization [8].

The computer vision is used to extract the semantic contents from the images. The labelled information for the concerned tasks like object recognition is to extract the quality image element in general feature extraction from the datasets and other visual elements [9]. The texture transformation algorithm comprises a texture formation technique with feature demonstration based convolutional neural networks (CNN). The style transfer technique minimizes and obtains solutions to the optimization issues in a neural network. The new images are constructed by implementing a feature extraction for the dataset images. The texture synthesis approach is used to increase the deep image representation [10].

The main contribution of our proposed work is as follows:

- (i) Deep learning approach is applied to real-time images after implementing the object detection
- (ii) The proposed methodology executes the style transfer efficiently in which the captured depth image is from the dataset
- (iii) Validation is performed using the proposed training methodology which may produce the style transfer with reduced complexity
- (iv) The visual object recognition is applied to the feature space with the usage of max-pooling layers

## 2. Related Work

The style images and the recombination of the images are separated by creating the artistic images with high quality using neural representations from the neural algorithm; this utilizes the artificial intelligence technique [11] which

presents a way to understand the artistic images algorithmically. Adversarial networks are created by generating the art based learning styles and conflicting from style standards; the system generates imagination based art using art distributions. It focuses on realistic texture rather than pixel accuracy [12]. The essential factors in neural style transfer demonstrate control over position and color data across spatial scale which improves by permitting high quality control over style and also assists in reducing normal faulty situation such as imaginary options.

A fast generator combined with feedforward network for initial style transfer has been proposed to get the styled output as an image with the produced unnoticed element with style image as the input throughout the inference period. The generator is constructed to perform the encoder and decoder by transferring the deep features. The spatial components are utilized to measure the level of styles that have been incorporated. The perceptual loss for every element has been reduced to identify the properties of the image styles from the artistic images. The multidomain related images are created using the mask element to implement the stabilization and evade the collapse in the real-time scenario. The selective stylized images are basically increasing the effectiveness of optimization related style transfer [13]. The neural techniques combined with patch related synthesis approach have been implemented to attain the stylization feature with high resolution. The neural methods have been trained to increase the stylization level in global standard and predict the output for relevant patch related synthesis at different level. The original artistic media is used in better way to improve the dependability of the image in stylish manner. Without incrementing the pixel size the generated image is in high quality. The visual quality is checked with the related style transfer method and the response is better with feature extraction metrics [14].

The deep learning based feature extraction has been implemented using the methods in which the identification of high level image features segregates the style and image contents. The image style adaptation methodology produces the style characteristics of several paintings and pertained styles for learning from the other images also. The incorporation of artificial intelligence concept with the art strengthens this method which has the improved one [15]. Neural learning techniques were implemented to be efficient for style transfer. A new image has been synthesized to keep the high level image with the low-level features for the style image. Moreover, the convolutional methods are used to implement the feature extraction for style images with semantic features identified from the images. The natural style elements are incorporated to improve the low-level features into the high level feature using machine learning approach [16]. Every painting may confine the integral enhancement for natural stylizations. The style transfer technique is combined with the direction based style transfer for improved texture extraction. An innovative field loss with direction is enhanced to the synthesized images. The loss function is used to identify if there is any loss in the time of style transfer and the incorporated method is constructed to reduce the loss itself without any specific methodology. A

simple interaction technique is developed to manage the generated fields and the direction of the texture in generated images. The texture enhancement method is used to implement the style transfer to the synthesized images [17].

The photographic related style transfer shows the improved output for spatial based semantic segmentation for input. The segmentation has been done within the region of the input image and the style referenced images are dissimilar in spatial contents. A spatial transformation technique is applied before implementing style transfer technique. The cross-correlation based feature maps are applied to compute the affine transformation to the generated image to implement the active style transfer for the semantically aligned regions. The pretrained CNN model is implemented to identify the reference images when the shadows are removed from the produced images [18]. The semantic related style transfer technique has been identified for getting the solution for the semantic based issues. The semantic matching is implemented to increase the style transfer quality. Every image is segregated into various regions with semantic values and improved painting [19]. The divided regions are further arranged related to the semantic elucidation. The source region is trained using the learning elements to produce an output in stylish manner. The semantic matches have been identified within the regions and the guaranteed semantic matching is ascertained within the source and the target outputs. The semantic gaps are identified whenever dividing the regions based on the semantic values for the real paintings and also the photographs. The domain adaptation method is developed to decrease semantic gap for regional segregation [20].

The synthesis images have been mostly used in the real-time applications with learning and training model developed to minimize the cost of the resources and also the human values. There are several differences identified for original images and synthesis images in real-time characteristics [21]. The style transfer is one of the solutions to minimize the gap within these kinds of images. The indoor synthesis images are converted into the improved images by minimizing the light influences. Hence, the content data of the real image is achieved by the style information. The convergence speed is converted to identify the real images in complex situations [22]. The visual artworks have been applied in a photo related style transfer method which may produce the realistic images. The outcome of this technique is used to assist the normal users to expand the style transfer images which are affected by the various real-time issues. The autonomous sky segmentation approach is used to segregate the input image with sky background. The background colors are characterized into the sky segmentation. The natural color transformation technique is adopted with the sky background and correction method to guarantee the quality output production [23].

The style transfer approach is also used in the field of medical images. The CNN and combined deep learning approaches are implemented in the computer vision tasks. The CNN model needs a huge amount of data sets that are very difficult in medical based image processing. For this problem, the image generation is used to increase the

computer vision [24]. A new engine is developed to exploit the network to confine the synthetic information. The style transfer method is utilized to enhance the visual realism in the public dataset combined to the semantic features using CNN methodology [25]. The optical images are applied in the deep learning concept under water image information to reduce the bottleneck. The detection of objects in real-time sonar images is utilized for the network based training [26]. The 3D artistic face modeling with controllable manner [27] enables the face geometry space and a face texture space based on 3D face dataset. The experiment is carried out in real time without GPU acceleration to achieve different cartoon characters. The deconstructed integer function [28] is applied to have different attributes as biomorphism, beauty, and symmetry. The random geometric graphs enable the creative artistic composition. The algorithm is introduced with new outline image [29] extracted from the content image. The variation regularization is applied to reduce the noise and to smoothen the boundary region and outline loss function applied on the outline image. The results experience the better design clothing shape which is reserved perfectly. The virtual space technology [30] enables the immersive experience which analyzes the multisensory and multitechnical spatial art style transformation form. The results show the better experience on art style transfer.

The neural style transfer has been constructed with several semantic representations with dual semantic loss which has been maintained with the particular values for the stylized outputs to every technique of the computed content images [31]. The color cast has been established according to the illumination modification and the temperature for increasing the productivity. The color calibration technique is used to transfer the exact color with semantic representation. The global attention functionality has removed the color cast from the input image for style transfer [32]. The neural style transfer technique has been implemented to convert the portrait image into the specific realistic image with some style and the pixel motion parameter with the color displacement from specific frames in semantic representation [33]. The optimization problems have the solutions through the metaheuristic techniques as the torus walk bat algorithm and modified bat algorithm are used to enhance the local search capacity ahead of utilizing the standard process [34, 35].

### 3. Proposed Method

The semantically useful style in a global level without user interaction and the low-level elements is removed and regaining the color related information is done to keep the fidelity without affecting the originality. The proposed system is constructed to diminish the computational complexity by producing the full resolution based images. The exploitation of previously used style transfer technique will enhance the professionalism. The image analogies have been minimized by using the neural related style transfer technique. The VGG16 is used to pretrain the convolutional neural networks in efficient manner. The dissimilar colors are mixed and produced the single region and are utilized to

perform the style transfer in semantically dissimilar regions. The max-pooling operation enhances the flow of the images in gradient level. An important element is implemented for producing art generation which is related with the creativity of art in real-time scenario.

VGG16, a pretrained neural network, is utilized to implement the content and style representations from its layers. Figure 1 represents the block diagram for art generation using neural algorithm using VGG-16 with 16 convolutional layers and 5 pooling layers. Whenever the input image and style image are passed through the network, input image is initialized to the content image. Style representation is extracted at the initial layers of the network as they extract the pixelated features. In the extraction, the layers extract the content in the image. Once the style and content representations are extracted, the output is generated by reducing the losses between them. The generated output image consists of the content representations of the input image and style representation. Convolution neural networks are used in image processing. They are made up of layers where every neuron delivers an input, computes a dot product, and pursues with nonlinearity optionally. Here, unlike any particular neural network, the neurons in the layers of CNN are prearranged in 3 dimensions: height, breadth, and depth, where depth refers to activation volume not the depth of the network.

It consists of layers of minute computational units which produce visual data randomly in a feedforward approach. A CNN is mainly built of convolutional based layers, these layers are the compilation of image filters that extract a particular feature from the input, where the feature map will be the targeted output. When CNN are trained for performing the image processing, they construct a representation of image which takes the object data with the processing functionality of the input image modified into the pixel based image representations for producing the quality images. The higher level contents and objects are arranged to execute the pixel values from the original image; the lower layers are reconstructed for the pixel values of the input image.

Convolution layer is the central unit of constructing the CNN that is mathematically operated to combine the group of data. It is implemented on the input data with convolution filter to create the feature map. Feature space is framed to confine the texture data and it is utilized to get the style of the image. The feature space is developed for delivering the filter responses in every layer. It contains the correlations within the dissimilar filter representations. By incorporating the feature correlations of various layers, the multiscale representation of the input image recognizes the texture components. The operation is computed by sliding the filter across the input image. At the particular location, there is a component related matrix formation and the resultant that creates the feature map. The common area of producing the convolution operation is performed by the standard filter which is shown in Figure 2.

Figure 3 demonstrates the generating feature map using convolutional input and filter. The filter across the aggregated input is used to produce the convolution results with

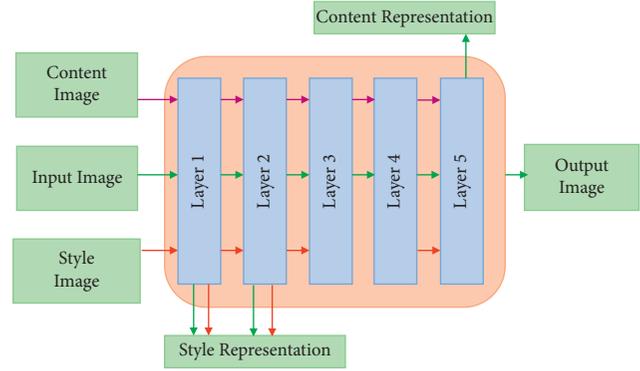


FIGURE 1: Block diagram of proposed system which takes input image.

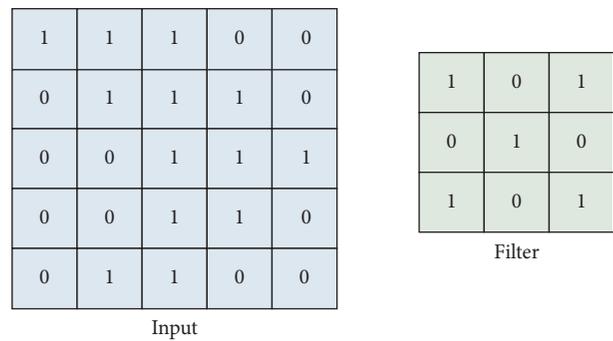


FIGURE 2: Images and filter.

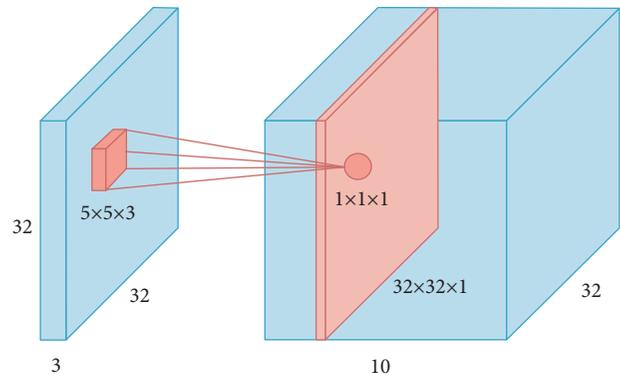


FIGURE 3: Generating feature map by convoluting input and filter.

feature map. The convolution operations are computed in 3 dimensions and the image is demonstrated as the 3-dimensional matrixes like depth, height, and width. A convolution filter holds the particular width and height as  $5 \times 5$  and  $3 \times 3$ ; it may cover the requirements of 3 dimensions. The feature maps are constructed to produce the final output with different kinds of filters and generated the output from the convolution layer. The convolution filter requires being determined in the input. If equal dimension requires maintenance, the padding concept is utilized to enclose the input with zeros.

In Figure 4, there is an input image with  $32 \times 32 \times 3$  dimensions and a filter layer with  $5 \times 5 \times 3$  dimensions. The filter has improved depth that matches the depth of the input

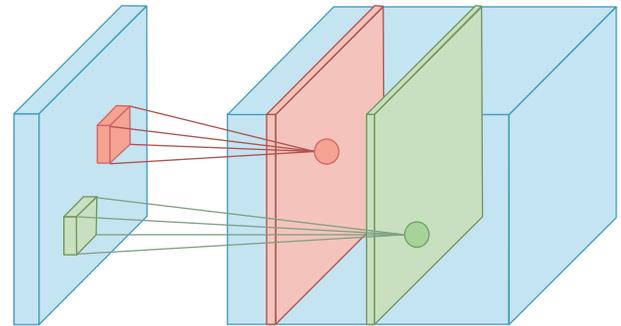
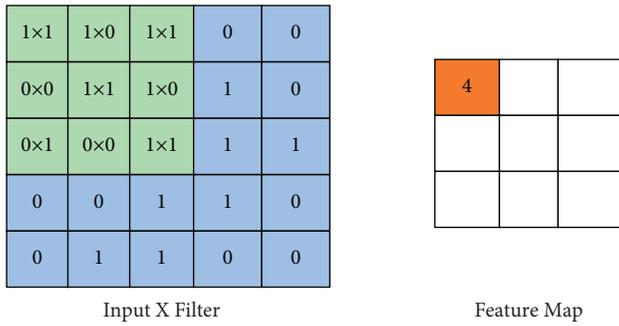


FIGURE 5: Generating a stack of feature maps using multiple filters.

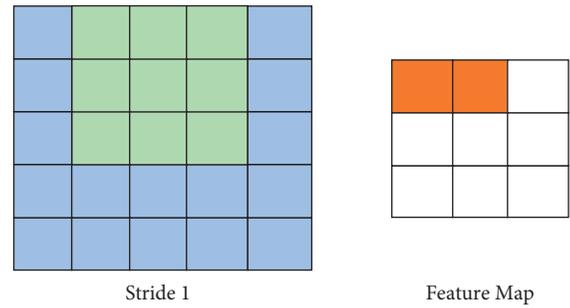
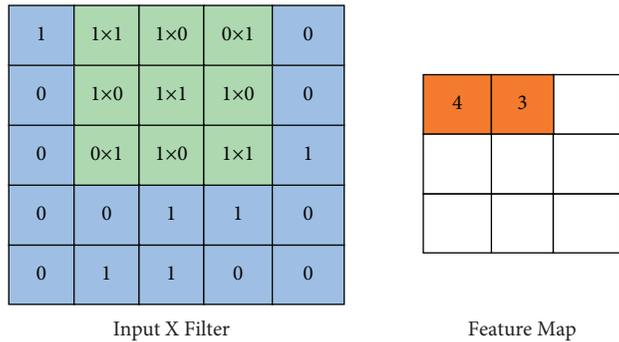


FIGURE 6: Stride value when it equals 1.

FIGURE 4: Generating a feature map with a single filter by convoluting the input layer with the filter.

and the value for both is 3. Whenever the filter is located in a specific position, it wraps a volume of input and produces the output using the convolution operations. If 10 dissimilar filters are used, 10 different feature maps are used when computing them together combined with the dimension of depth for computing the size of  $32 \times 32 \times 10$  convolution layers. The convolution operation for every filter is computed individually and results in a disjoint set. Nonlinearity could be achieved by producing the sum of weighted inputs along with the activation function; the convolution operation is passed along with the nonsaturating activation function called ReLu. The constructed final feature maps are the network with ReLu operation.

The sum of the matrix multiplication is performed in 3 dimensions as shown in Figure 5, but the result is scalar with the feature map of size  $32 \times 32 \times 10$ . The multiple filters are used to decrease the computational cost; it needs to utilize the specific filters within the particular time for learning purpose. The filters are mapped to the input image and learn every part of the image with the tiny filter sliding from one end to another end to produce the output image. The convolution process has the individual value in spite of using the style filter and the convolutions are joined to frame the output through the total amount of filters.

Stride demonstrates the convolution filter required to be moved at every step. Initially, the value of stride is 1 which is demonstrated in Figure 6. The stride is recommended to use the initial value 1. The padding could be frequently using layers in the convolution operation but not in pooling operation. The bigger strides are overlapped within the output fields that generate the output feature map of very minimum value.

From Figure 7, it can be observed that the size of the feature map is shorter than the input stride value increased to 2. After the convolution and pooling layers, few connected layers are included to conclude the CNN architecture. It is known that the output of the convolution layers and pooling layers is in 3 dimensions but fully connected layers expect a 1 dimension vector. Hence, the final output of the pooling layer is flattened to a vector which befalls the input to the fully connected layer, where destruction is organizing the 3-dimension volumes of numbers into a 1-dimension vector.

Figure 8 represents the padding in grey; the input can be embedded with zeros. The padding is the concept of merging the pixels into the image whenever the kernel processing padding value is 0. The testing process has the similar dimensions and the length to highest length will affect the accuracy. The dimensionality of the feature map equals the input value; then the padding is normally used to hold the feature map size that would process at every layer in the convolutional neural networks.

Pooling layers behind sample every feature map individually, diminishing the height and width maintenance with the depth together. Pooling is performed after the convolution operation to diminish the dimensionality, which enable us to reduce the amount of parameters which abbreviates the training time and overfitting. Pooling layer slides a window across its input and takes the highest value in the window where the window size and stride are in particular similar to a convolution. The most general type of pooling is max-pooling which receives the maximum assessment in the window as shown in Figure 9.

From Figure 10, the window and stride configurations segregate the feature map pooling size to minimum that the

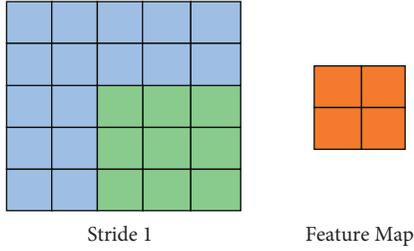


FIGURE 7: Stride value when it equals 2.

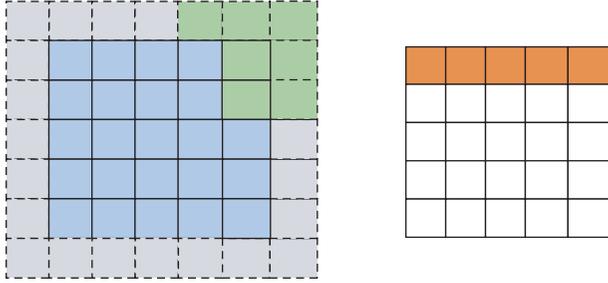


FIGURE 8: Padding.

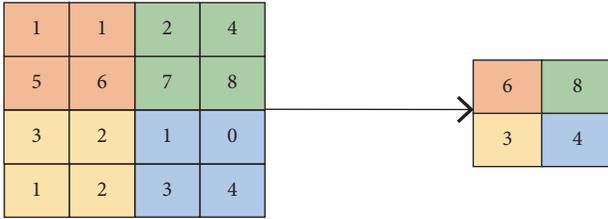
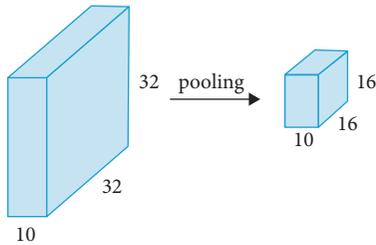
FIGURE 9: Max-pooling with  $2 \times 2$  window and stride 2.

FIGURE 10: Reducing the feature map size by pooling.

feature map is maintaining the significant data. If the input to the pooling layer dimension is  $32 \times 32 \times 10$ , it could be reduced to  $16 \times 16 \times 10$ . There are 4 parameters of  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$  filter sizes that can be used depending on application. It should be noted that filters have depth which is equivalent to the depth of its input. The filter count is a power of 2 within 32 and 1024. Utilizing additional filters might result in a dynamic model, but these risk overfitting according to the improved parameter value. Frequently, the network starts with minimum amount of filters at the arbitrary layers and progressively increases the value once the network increments. The width and the height are changed but the depth does not change because of the individuality of every layer for pooling.

The VGG network was trained to implement the object recognition with the feature space constructed using convolutional and max-pooling layers. The network is positioned to improve the scaling and weight of the activated filters within the images. With the arrangement of VGG network for producing the better output, the activation functions are generated for increasing the efficient feature maps. This model contains the fully connected layers for performing the pooling operating to increase the efficiency.

The CNN learns the styles through the parameters for training the restructuring loss  $Loss_R$  for the targeted output image ( $O$ ) from an input image ( $I$ ) computed using the following equation:

$$Loss_R = \|O - I\|^2. \quad (1)$$

The *perceptual* loss  $Loss_P$  within the style branch is computed using the following equation:

$$Loss_P = x Loss_R(O_i, I_i) + y Loss_{st}(O_i, St_i) + z Loss_N(O_i) + p Loss_{st}(O_i, St_i), \quad (2)$$

where  $St_i$  is the styled image,  $Loss_{st}$  is restructuring loss for style,  $Loss_N$  is the restructuring loss for the normal, and  $Loss_{in}$  is the restructuring loss for intensity. The restructuring loss for style ( $Loss_{st}$ ) is computed using the following equation:

$$Loss_{st}(O_i, St_i) = \sum_l \left\| GM(FM^l(O_i)) - GM(FM^l(St_i)) \right\|^2. \quad (3)$$

The restructuring loss for intensity is computed using the following equation:

$$Loss_{in}(O_i, St_i) = \frac{1}{\alpha} \sum_{\alpha=1}^k \|O_i^\alpha - St_i^\alpha\|^2. \quad (4)$$

The complexity is analyzed in the network that uses the nonlinear values for fixing the layer position in the network. The input image  $\vec{a}$  is prearranged in every layer of the CNN using filter reaction to the image. A layer is mapped with the individual filters for producing the feature maps of every matrix and identifies the position of the particular layer.

The style representation is obtained from an input image; the feature space is identified to identify the texture data. The feature space has been constructed to reply the filter in the layers. It contains the correlations within the dissimilar filter replies that are expected for spatial related feature maps. These correlations framed the generalized matrix ( $Ge_{ij}^l$ ) using feature map  $Fe_{ik}^l$  in the following equation:

$$Ge_{ij}^l = \sum_k Fe_{ik}^l Fe_{jk}^l. \quad (5)$$

#### 4. Results and Discussion

The experiments have been done using the MATLAB software and the WikiArt data set has been utilized to implement the performance evaluation. The style classification is performed using the style ambiguity. The proposed

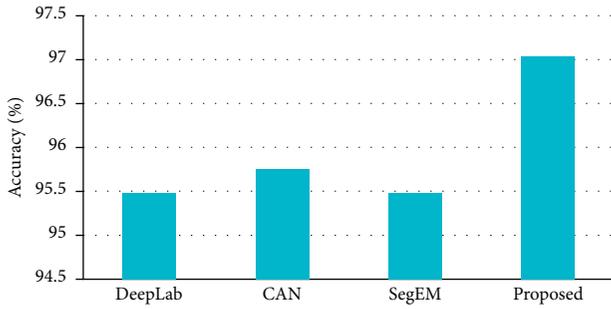


FIGURE 11: Accuracy %.

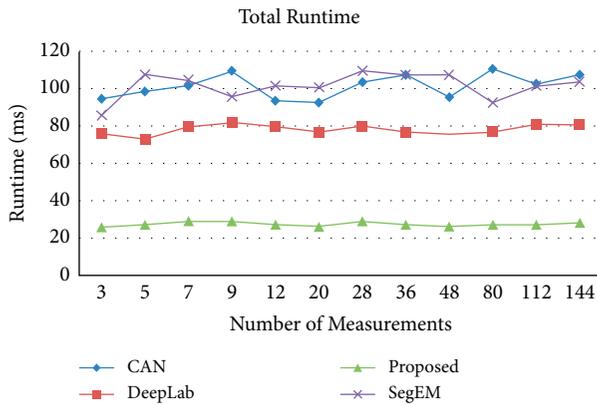


FIGURE 12: Total runtime.

methodology is compared with the related methods of DeepLab [1], CAN [12], and SegEM [23], and the performance metrics used for these experiments are processing time, restructuring loss, and accuracy.

Figure 11 demonstrates the accuracy % for successfully performing style transfer; the proposed methodology has efficient functionality of improving the accuracy compared to the related methods.

Figure 12 demonstrates the runtime for the measurement of the style transfer and the results illustrated that the proposed methodology has diminished amount of runtime compared with the other methods. The restructuring loss is one of the primary parameters to increase the functionality of the proposed methodology.

Figure 13 illustrates that the proposed methodology has reduced amount of restructuring loss percent compared with the related methodologies. The curve is the plot for discovering the performance over time period which has been produced by the machine learning approach.

Our proposed method has produced significantly good results from the neural network based technique. The large-scale based artefacts are generated using this proposed approach. The produced result has high quality in resolution and it looks the same as the input image after successfully implementing the style transfer. Eliminating the style ambiguity loss is used to improve the accuracy of implementing the style transfer. Our proposed system has proved producing the new artefacts without affecting the resolution of the images.

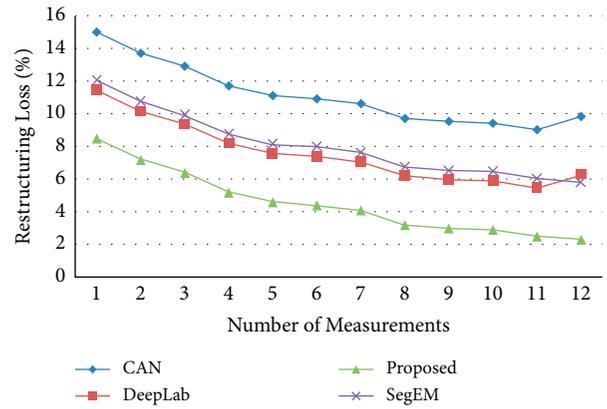


FIGURE 13: Restructuring loss %.

## 5. Conclusion

The images generated through style transfer are dependent on iterations. The images generated with minimum iterations have more style features while the images generated with maximum iterations have more content features. This traces how humans produce and recognize artistic imagery which gives an algorithmic understanding. In conclusion, that pretrained neural network can be used for not only image recognition but also painting. The feature space demonstrates the larger levels of the input image through the VGG16 neural network for replicating the task while the input image has executed the style transfer to capture the depth image. The validation process is used for producing the style transfer with reduced amount of complexity and the max-pooling layers are involved into the feature space for visual object recognition.

## Data Availability

Data will be made available upon request.

## Disclosure

This research does not involve any human or animal participation.

## Conflicts of Interest

The authors declare that they do not have any conflicts of interest.

## Authors' Contributions

R. Dinesh Kumar contributed to writing original draft, writing review and editing, conceptualization, data curation, and validation. E. Golden Julie contributed to conceptualization, formal analysis, and supervision. Y. Harold Robinson contributed to conceptualization, formal analysis, writing original draft, writing review and editing, and supervision. S. Vimal contributed to conceptualization, writing original draft, formal analysis, and supervision. Gaurav Dhiman contributed to conceptualization, writing original

draft, formal analysis, and supervision. Murugesh Veerasamy contributed to conceptualization, formal analysis, writing original draft, writing review and editing, and supervision. All authors have checked and agreed the submission.

## References

- [1] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2017, Article ID 2699184, 848 pages, 2017.
- [2] T. Li, R. Qian, C. Dong et al., “Beautygan: instance-level facial makeup transfer with deep generative adversarial network,” in *Proceedings of the 2018 ACM Multimedia Conference on Multimedia Conference*, pp. 645–653, ACM, Seoul, Korea, October 2018.
- [3] W. Wu, Y. Zhang, C. Li, C. Qian, and C. C. Loy, “Reenactgan: learning to reenact faces via boundary transfer,” in *Proceedings of the European Conference on Computer Vision*, pp. 622–638, Springer, Munich, Germany, September 2018.
- [4] H. Chang, J. Lu, F. Yu, and A. Finkelstein, “Pairedcyclegan: asymmetric style transfer for applying and removing makeup,” in *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, June 2018.
- [5] Z. Ma, N. Wang, X. Gao, and J. Li, “From reality to perception: genre-based neural image style transfer,” in *Proceedings of the IJCAI*, pp. 3491–3497, Stockholm, Sweden, July 2018.
- [6] M. Lu, F. Xu, H. Zhao, A. Yao, Y. Chen, and L. Zhang, “Exemplar-based portrait style transfer,” *IEEE Access*, vol. 6, pp. 58532–58542, 2018.
- [7] N. Wang, X. Gao, and J. Li, “Random sampling for fast face sketch synthesis,” *Pattern Recognition*, vol. 76, pp. 215–227, 2018.
- [8] N. Wang, M. Zhu, J. Li, B. Song, and Z. Li, “Data-driven vs. model-driven: fast face sketch synthesis,” *Neurocomputing*, vol. 257, pp. 214–221, 2017.
- [9] F. Luan, S. Paris, E. Shechtman, and K. Bala, “Deep photo style transfer,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, July 2017.
- [10] L. A. Gatys, A. S. Ecker, and M. Bethge, “Texture synthesis using convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 28, 2015.
- [11] L. A. Gatys, A. S. Ecker, and M. Bethge, “A neural algorithm of artistic style,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–16, Boston, MA, USA, June 2015.
- [12] E. Ahmed, B. Liu, M. Elhoseiny, and M. Mazzone, “CAN: creative adversarial networks generating ‘art’ by learning about styles and deviating from style norms,” in *Proceedings of the ACC Eighth International Conference on Computational Creativity (ICCC)*, pp. 1–22, Atlanta, Georgia, USA, 2017.
- [13] Z. Xu, M. Wilber, C. Fang, A. Hertzmann, and H. Jin, “Adversarial training for fast arbitrary style transfer,” *Computers & Graphics*, vol. 87, 2020.
- [14] O. Texler, D. Futschik, J. Fišer et al., “Arbitrary style transfer using neurally-guided patch-based synthesis,” *Computers & Graphics*, vol. 87, 2020.
- [15] L. Liu, Z. Xi, R. Ji, and W. Ma, “Advanced deep learning techniques for image style transfer: a survey,” *Signal Processing: Image Communication*, vol. 78, pp. 465–470, 2019.
- [16] H. Wu, Z. Sun, Y. Zhang, and Q. Li, “Direction-aware neural style transfer with texture enhancement,” *Neurocomputing*, vol. 370, pp. 39–55, 2019.
- [17] M. Guo and J. Jiang, “A robust deep style transfer for headshot portraits,” *Neurocomputing*, vol. 361, pp. 164–172, 2019.
- [18] J. H. Park, S. Park, and H. Shim, “Semantic-aware neural style transfer,” *Image and Vision Computing*, vol. 87, pp. 13–23, 2019.
- [19] T. Zhao, Y. Yan, I. S. Shehu, X. Fu, and H. Wang, “Purifying naturalistic images through a real-time style transfer semantics network,” *Engineering Applications of Artificial Intelligence*, vol. 81, pp. 428–436, 2019.
- [20] H. Li, L. Wan, S. Wang, and Z. Cui, “Photo2Silhouette: transferring silhouette photo style for images,” *Optics & Laser Technology*, vol. 110, pp. 30–35, 2019.
- [21] P. Andreini, S. Bonechi, M. Bianchini, A. Mecocci, and F. Scarselli, “Image generation by GAN and style transfer for agar plate image segmentation,” *Computer Methods and Programs in Biomedicine*, vol. 184, Article ID 105268, 2020.
- [22] S. Lee, B. Park, and A. Kim, “Deep learning based object detection via style-transferred underwater sonar images,” *IFAC-PapersOnLine*, vol. 52, no. 21, pp. 152–155, 2019.
- [23] M. Berning, K. M. Boergens, and M. Helmstaedter, “SegEM: SegEM: efficient image analysis for high-resolution connectomics,” *Neuron*, vol. 87, no. 6, pp. 1193–1206, 2015.
- [24] J. J. Virtusio, A. Talavera, D. S. Tan, K.-L. Hua, and A. Azcarraga, “Interactive style transfer: towards styling user-specified object,” in *Proceedings of the 2018 IEEE Visual Communications and Image Processing (VCIP)*, December 2018.
- [25] A.-D. Nguyen, S. Choi, W. Kim, and S. Lee, “A simple way of multimodal and arbitrary style transfer,” in *Proceedings of the ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, 2019.
- [26] S. Chelaramani, A. Jha, and A. M. Namboodiri, “Cross-modal style transfer,” in *Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP)*, Athens, Greece, October 2018.
- [27] T. Suontphunt, “A practical approach for identity-embodied 3D artistic face modeling,” *International Journal of Computer Games Technology*, vol. 2014, Article ID 781950, 10 pages, 2014.
- [28] E. Estrada and P. Pereira-Ramos, “Spatial ‘artistic’ networks: from deconstructing integer-functions to visual arts,” *Complexity*, vol. 2018, Article ID 9893867, 8 pages, 2018.
- [29] H. Wang, H. Xiong, and Y. Cai, “Image localized style transfer to design clothes based on CNN and interactive segmentation,” *Computational Intelligence and Neuroscience*, vol. 2020, Article ID 8894309, 12 pages, 2020.
- [30] L. Zeng and X. Dong, “Artistic style conversion based on 5G virtual reality and virtual reality visual space,” *Mobile Information Systems*, vol. 2021, Article ID 9312425, 8 pages, 2021.
- [31] W. Ye, X. Zhu, Z. Xu, Y. Liu, and C.-C. Chang, “A comprehensive framework of multiple semantics preservation in neural style transfer,” *Journal of Visual Communication and Image Representation*, vol. 82, Article ID 103378, 2021.
- [32] H. Huang, A. Yang, Yu Tang et al., “Deep color calibration for UAV imagery in crop monitoring using semantic style transfer with local to global attention,” *International Journal*

*of Applied Earth Observation and Geoinformation*, vol. 104, 2021.

- [33] J. Cui, Y. Q. Liu, H. J. Lu et al., "PortraitNET: photo-realistic portrait cartoon style transfer with self-supervised semantic supervision," *Neurocomputing*, vol. 465, 2021.
- [34] H. B. Waqas, J. Ahmed, and T. R. Hafiz, "A modified bat algorithm with torus walk for solving global optimisation problems," *International Journal of Bio-Inspired Computation*, vol. 15, no. 1, pp. 1–13, 2020.
- [35] W. Haider Bangyal, A. Hameed, J. Ahmad et al., D. Rawat, R. Etengu, D. B. Rawat, and R. Etengu, "New modified controlled bat algorithm for numerical optimization problem," *Computers, Materials & Continua*, vol. 70, no. 2, pp. 2241–2259, 2022.