Deep Learning Image Transfer by Simulation

Abdussalam Masaud Ammar
Faculty of Electronic Technology
Tripoli, Libya
E. mail Slm2010ly@gmail.com

Amira Youssef Ellafi
Faculty of Electronic Technology
Tripoli, Libya
E mail Amoor85ly@yahoo.com

Kenz A. Bozed
Computer System Design Dept. Faculty of Information Technology, Benghazi University, Benghazi, Libya
E.mail kenz.bozed@uob.edu.ly

Amer R. Zerek
EE Department, Faculty of Engineering, Zawia University
Zawia, – Libya
E-mail anas_94z@yahoo.co.uk
a.zerek@zu.edu.ly

***Abstract* - The image classification is a classical problem of image processing, computer vision and machine learning fields. In this paper we study the image classification using deep learning with Neural style transfer that has been a high risk application for deep learning, make attention from and advertising the effectiveness to both the academic prisons and the general public. However, we have found by removal experiments that optimizing an image in the way neural style transfer does, we can even factor out the deepness (multiple layers of exchange linear and nonlinear transformations) all together and have neural style transfer working to a certain range. We present the VGG-19 Artistic Type Neural Algorithm, which can convert and recombine the image quality and style of natural images. This algorithm allows us to produce new images of high perceptual quality that combine the content of an arbitrary photograph with the appearance of numerous well known portrait. Our results provide new insights into the deep image representations learned by Convolutional Neural Networks(CNN) and demonstrate their potential for high level image synthesis and manipulation.**

Index Terms – Deep Learning CNN, ANN, ReLU.

I. Introduction

This may be considered a problem of structure transfer if the style is transferred from one image to another. The goal of structure conversion is to synthesize a structure from a source image, thus pressuring the structure to merge in order to maintain the semantic substance of the image desired.

There is an enormous expansion of non-parametric algorithms for structure synthesis that can synthesize regular photorealistic structure by recombining the pixels of a given source structure[1,2,3,4].

Most of past structure transfer algorithms depend on these nonparametric ways for structure blend whereas using diverse strategies to preserve the construction of the wanted picture. For occasion, Freeman and Efros presents a correspondence outline that incorporates highlights of the target picture like a picture intensity to constrain the structure blend strategy [4].

Although these algorithms gain exceptional outage, they all experience from the same limitations, they use low-level image characteristics of the target image to advise on the transfer of the structure. A style transfer algorithm must be able to derive the semantic content of the image from the target image (e.g. the objects and the common natural views) and after that inform a structure transfer strategy to made the semantic content of the target picture within the pattern of the source picture.

The key requirement for this is to discover image representations that display variations between the content of the semantic image and the style where it is presented differently. For controlled subsets of normal images, such as faces under different light characters and conditions, such factorized simulations have already been conducted in several textual styles [5] or written by hand digits and house numbers only for controlled subsets of normal images, such as faces under different light characters and conditions [6].

Generally, it is very difficult problem to separate content from style in normal images. However, the recent creation of Deep CNN [7,8], has produced successful computer vision systems that learn how to distinguish high-level semantic data from normal images.

It is clear that CNN, prepared with productive labelled information on basic tasks such as target recognition, learns to derive high-level image material from standard attribute representations that generalize visual data preparation, structure detection counting and innovative style classification practices across datasets and also other activities [9].

In this paper, we explain how it is possible to use the bland highlight representations taught from high-performance CNN to separate the handle and control the content and design of regular images. By introducing A Neural Algorithm of Artistic Style, a modern algorithm to perform the transfer of image style.

Ideally, it is a structure transfer algorithm that limits the technique of structure synthesis by feature caricatures from

978-1-6654-1856-0/21/$31.00 ©2021 IEEE

state-of-the-art CNN. Since the structure model is also based on deep image representations, the style transfer technique can minimize problems with optimization within a single neural network. New images are created by conducting a pre-image search to match sample feature representations. In order to strengthen the relationship between deep image representations and structure synthesis, this popular approach has previously been used. Actually, the style transfer algorithm VGG-19 gathers a parametric structure model based on CNN [7] with a method to invert their picture representations [10].

II The DEEP LEARNING

Deep learning it is a part of machine learning, which could be a sort of artificial intelligent (AI) technology. Particularly, it's a sort of machine learning that look for to teach computers to memorize by repeats examples. And it needs many of information to memorize from.

Deep learning learns from data by itself. Deep learning not like other types of machine learning based on how it works. Basically, It's the side of machine learning that uses messy data to draw its own conclusions.

Deep learning will take thousands of pictures of diverse things. It must spot designs to memorize how to classify those pictures. The meaning behind deep learning is for the computer to memorize independently, without people telling it what to seek for.

Today, deep learning is also used in common applications such as machine vision and image recognition. One of the crucial applications of a typical combination of conventional approach and AI innovation is aesthetic development or computer synthesized representation. Already, how to exchange one portray fashion to other fashion is depleting and time expending.

To illuminate this issue, image style transfer algorithms have been developed and analysts have taken into account the study of this area. Image of deep learning recently, the style transfer ways were ordinarily called surface exchange or structure transfer. These calculations give strategies and motivations for afterward calculations with profound learning. Two kinds of structure transfer techniques, non-photorealistic and photorealistic review, are noted in this paper.

Non-photorealistic recall typically operates on canvases or inventive pictures that are synthesized. When searching for desired pixels, a quick structure transfer algorithm is proposed to use the Different Reference Outline (DRO) technique rather than the comprehensive search [11], so structure synthesis speed can be unmistakably improved. Depend on this tool, Lee et al. [12] recommend a made strides calculation. Their strategy works speedier and more proficient on fashion exchange. Photorealistic reviewing calculations frequently works with genuine photos or synthesized practical pictures.

The picture knitting algorithm which Freeman and Efros [13] displayed for structure transfer and structure synthesis is one of the foremost substantial Photorealistic reviewing strategies. The main thought of their method is to sew structure pieces via finding the leading way which reduces the misfortune. At show time, as GPUs' noteworthy advancement, thinks about on profound learning and neural systems got to be predominant once more.

Creators gathering structure transfer techniques and models of deep learning, then gives the Neural Style Transfer (NST), which presently may be an incredibly vital inquire about region. The first full and efficient neural strategy was identified by two separate component types and images, and it was proposed that the two components be merged to replicate a new image with a pre-prepared CNN. Deep learning techniques that rely on online pre-trained models, such as VGG networks, rely on online models [14].

In view of this approach is rather inconsistent with Photorealistic exchange, other approach suggest an algorithm made enhanced on reality. The first algorithm by modified the equations.

Online picture recreation strategies are generally complex and overwhelming, so a few analysts displayed deep learning strategies depend on offline models in arrange to get rid of the need of adaptability. Offline and pre-trained models prepared by information-driven techniques are used by these algorithms. In reality, exchange speed of these calculations is or maybe quicker than that depend on offline models. Offline models can be assorted into three groups, (along with the capacity of fashion kinds of these models), single-style models, arbitrary models and multi-style models.

Strategies using single-style models are the calculations that could be shared by each prepared demonstration since it was one style of fashion without retraining. The most considerate offline models used are models that are usually limited in scale and have faster speed.

Some researchers exploited Generative Antagonistic Systems (GANs) to show a modern form of loss, providing an image processing technique without matched data. Multi-style models allow techniques to exchange images without needing the demonstration to be retrained in different styles. There are several Style Banks, the main structure of their multi-style capacity support policy. These models are the most adaptable techniques, taking a very short amount of time to prepare for new style. Furthermore, one of the key tactics that met the arbitrary-style and real-time criteria.

III IMAGE CLASSIFICATION WITH DEEP LEARNING.

If some person looks at a picture, he will ordinarily recognize what it represents easily. For example, ears are triangle shape, black circle nose, obviously a cat. Windows, 4

Wheels, 4 doors, it's a car. We learn this talent instinctively, and it has become second nature to us.

This assignment is not as obvious to computers as it is to us. They don't have the same 'perspective' on the universe as we do. Computers may find it difficult to classify images at that stage. This is where deep learning comes in. Deep learning is a way of training a computer to identify and classify images by analysing several different images of the same object from various angles, different looks, different colours, different sizes and so on as in Fig. 1.

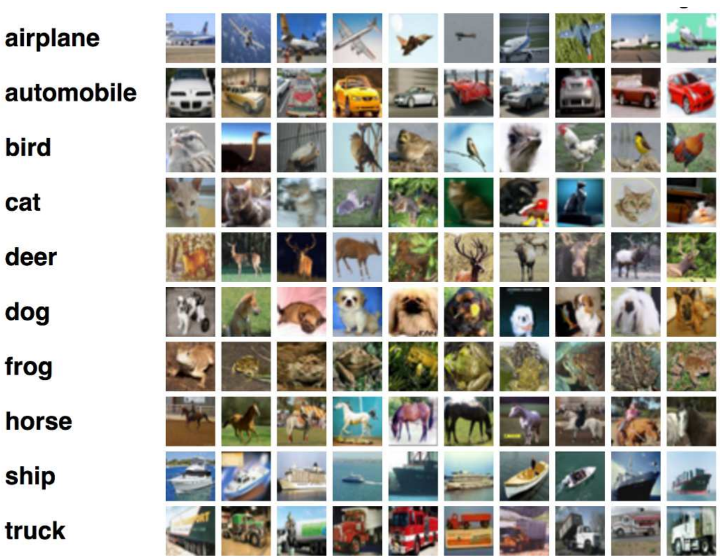

Fig. 1. How to using a computer to identify and recognize pictures

A. *Classification of Image*

Image classification is the process of a machine analyzing a photograph and determining which class it belongs to. For example, suppose you enter a picture of a cow. Image classification is the process where the computer analyzes this picture and determines whether or not it is a cow.

As seen in Fig. 2, early image classification is based on raw of data pixel. This implied that computers would break down pictures into independent pixels. The problem is that two images of the same object will seem to be very different. They may have unique foundations, points, and postures, for example. This made it difficult for computers to accurately recognize and identify certain images.

The image classification has many challenges including: -

1) Viewpoint variety. A single occasion of a thing can be arranged in numerous ways with regard to orientation of the camera.
2) Scale variety. Visual classes regularly show variety in their estimate (measure within the genuine world, not as it were in terms of them extend within the picture).
3) Distortion. Numerous objects of intrigued are not solid bodies and can be distorted in extraordinary ways
4) Occlusion. The objects of intrigued can be blocked. In some cases, as it were a little parcel of an object (as small as few pixels) may well be obvious.

Illumination conditions. The impacts of light are exceptional on the pixel level.

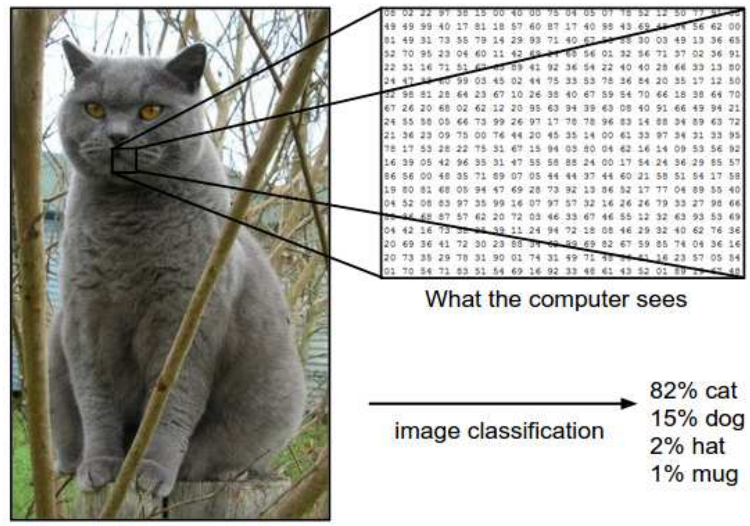

Fig 2. Computer "vision" via data.

B. *Adding to deep learning.*

Deep learning is a form of artificial intelligence, which is a subset of Machine learning that allows computers to memorize data. Deep learning makes use of neural networks, which are computer models.

The input filters through covered up layers of hubs in neural networks. Any of these nodes prepares the input and communicates their result to the next level of nodes. This will repeat until it reaches a final layer or output layer, at which point the computer will respond.

Based on how the hidden layers behave, there are several types of neural networks. Convolutional Neural Networks (CNNs) are the most commonly used deep learning algorithms for image recognition. In CNNs, the hubs inside the hidden layers don't always communicate their contribution with the hubs in the next layer (known as convolutional layers).

The computers were able to identify and extract highlights from images thanks to deep learning. This means they can discover what to look for in pictures by analyzing a large number of them. As a result, software developers do not have to manually enter these filters.

C. *Artificial Neural Network*

Deep learning works by acting just like the human brain. And to do that, it employments something called Artificial Neural Networks (ANN).

ANN could be a recreation of a natural human brain, but for your computer. In other words, they're a way that a machine can prepare information, that's propelled by in spite of the fact that not similar to human and creature brains.

An ANN comprises of a collection of associated 'nodes' known as artificial neurons. These freely take after neurons in a natural brain. Each node can handle the input, and employing a connection a bit like neural connections in a natural brain, communicate their result with the other nodes.

These nodes are regularly organized into layers, known as hidden layers. Each layer of nodes communicates with the layer following to it. An ordinary ANN will likely have two or three hidden layers of nodes as in Fig. 3. But a few huge ANN exist, which have as numerous as 150 layers.

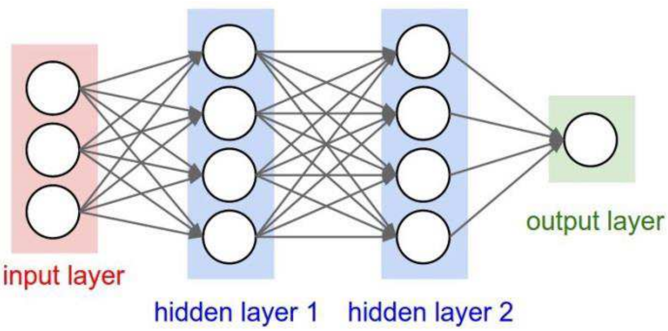

Fig 3. Basic Deep Neural Network

The multilayered neural network is extremely complex in terms of its architecture in a real-world application [15]. At present, the multilayered neural network is described as deep learning. Each node in a deep neural network recognizes its own inputs and sends them to the next layer on the previous layer's behalf.

IV CONVOLUTIONAL NEURAL NETWORK.

The most popular algorithm used to apply the deep learning process is the Convolution Neural Network (CNN). The CNN consists of a layer's classification and the detection feature. In other words, a CNN consists of several layers, such as convolution layers, layers of max-pooling or average-pooling, and fully-connected layers.

CNN were inspired by animal vision. The way the nodes in a CNN communicate with each other resembles the way some animals see the world. So, rather than taking everything in as a whole, small areas of an image are taken. And these small areas overlap to cover the whole image.

The CNN is divided into two categories named LeNet and lexNet. The LeNet, a pioneering 7-level convolutional network by LeCun et al. in 1998 [16] that classifies digits, was applied by several banks to recognize hand-written numbers on checks (British English: cheques) digitized in 32x32 pixel images. The ability to process higher resolution images requires larger and more layers of convolutional neural networks, so this technique is constrained by the availability of computing resources.

The LeNet is expressed as the Shallow CNN that is planned for the hand-written digits to be classified. Two convolutional layers, two subsampling layers, two hidden layers and one output layer are used in the LeNet [17].

AlexNet is a set of deep convolutional neural networks used to assign an input image into one of a thousand groups, i.e. image recognition. AlexNet is used to solve many issues, such as the grouping of indoor senses, which in artificial neural intelligence is strongly seen. It's a good way to understand the image's specifications for pattern recognition in the computer field with more differential vision.

AlexNet is made up of five convolutional layers, three subsampling layers, and three fully connected layers. The main distinction between LeNet and AlexNet is the function extractor form. In AlexNet, we use non-linearity in the function extractor module, while in LeNet, we use log sinusoid. AlexNet utilizes a dropout that isn't found in any of the network's other data sets.

In fact, deep learning CNN models to prepare and test, each input picture will pass it through a many of convolution layers with channels (Kernals), Pooling, completely associated layers and apply Softmax work to classify an object with probabilistic values between 0 and 1. As in the Fig. 4 , may be a full CNN stream to prepare an input image and classify objects based on values.

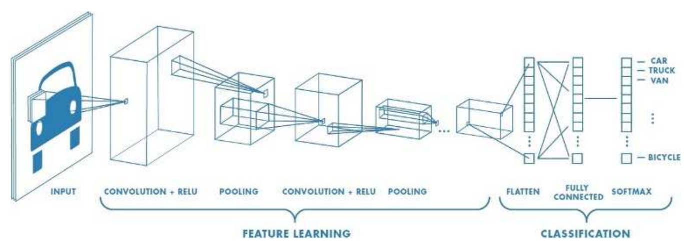

Fig. 4 : Neural network with several layers of convolution.

A. Convolution Layer.

A convolution layer is the main layer for extracting highlights from an input image. Convolution maintains the relationship between pixels by learning image highlights from tiny bits of input data. It can be a scientific operation containing two specific inputs, such as a matrix of images and a kernel or filter.

B. Strides.

The number of pixels shifts over the stride-defined input matrices. The whole issue comes down to shifting the filters to one pixel at a time when the stride at that point is 1. If the stride is 2 at that point, it is appropriate to shift the filters to two pixels at a time, and so on.

C. *Padding.*

It is covering the target picture with a set of zeros or ones, in some cases, filter does not suitable the input picture. So there are two choices:

1) Pad or cover the picture with zeros (zero-padding).
2) Get rid of the image part where the filter didn't suit. Typically referred to as substantial padding that preserves a large portion of the image as it was.

D. *Non Linearity (ReLU).*

Rectified Linear Unit (ReLU) is used for a non-linear active . We get the output as in (1).

$$f(x) = max(0, x). \quad (1)$$

The reason for ReLU is to present in our CNN Non-linearity. Since non-negative linear values would be the real life information that our CNN would like to memorize, Fig. 5 illustrates the process of ReLU.

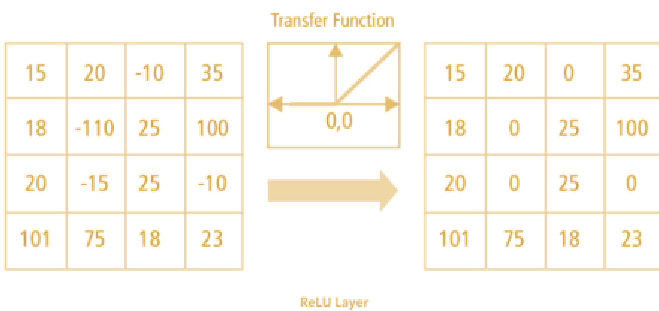

Fig 5. ReLU operation

E. *The Layer of Pooling*

The number of parameters can be reduced by pooling segment layers, since the images are also large. Spatial pooling, which minimizes the dimensionality of each outline, may also be called subsampling or down sampling, but includes essential details. Spatial pooling comes in many types:

1) Sum pooling.
2) Max pooling process.
3) Average pooling process.

Max pooling, as seen in Fig. 6, takes the largest element from the corrected included map. Average pooling can also involve taking the largest component. In the function map, the summation of all components could be called sum pooling.

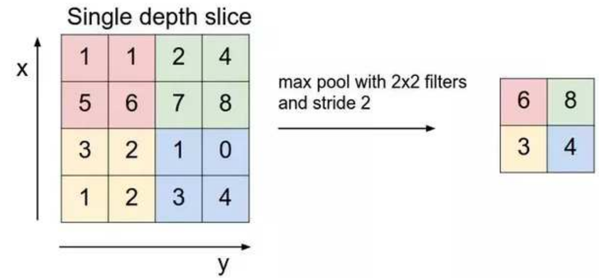

Fig 6. Max Pooling

F. *The Layer of Fully Connected.*

This layer is known as FC layer, we straighten our matrices into vectors and nourish them like a neural network into a fully associated layer. In Fig. 7, the vector matrix (X1, X2, X3, X4, ...) of the highlight map will be modified. We merged these features with the fully associated layers to create a model.

In conclusion, we have an activation job to identify the yields as cow, puppy, bus, house, etc., such as softmax or sigmoid.

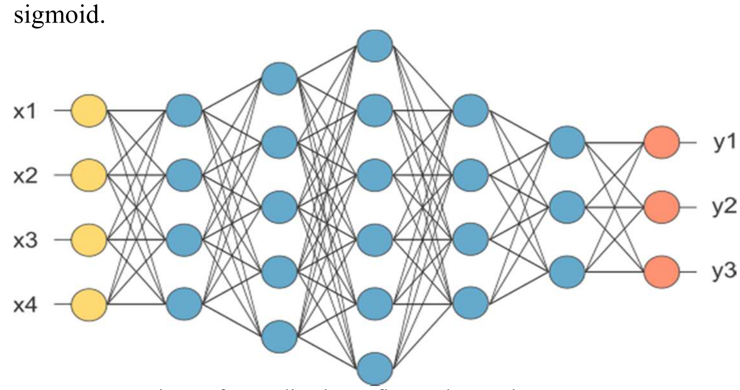

Fig 7. After pooling layer, flattened as FC layer

V NEURAL STYLE TRANSFER USING VGG-19 MODEL

Neural style transfer is an optimization technique that used to take two images, a content image (here we use a photo of Traditional style suburban home) and a style reference image such as magnificent artwork by a historical famous painter (here we use Van Gogh painting "Starry Night" as the style image) as shown in Fig. 8. and mix them together so the output picture looks like the original image, but drowned in the style of the reference image as in Fig. 9.

Fig 8. content image with style image.

Fig 9. the result of transfer

VGG-19 CNN has six basic architectures, each consisting essentially of separate connected convolution layer and completely linked layers. The convolutional component has a 3*3 scale and an input size of 224*224*3. For the most part, the number of layers is clustered between 16 and 19 [18]. The model structure of VGG-19 is shown in Fig. 10 [19].

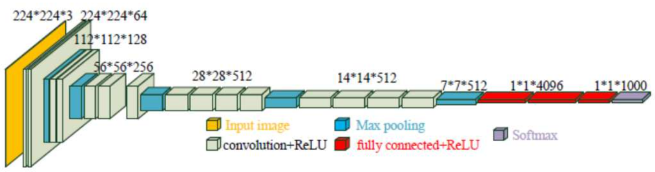

Fig 10. model of VGG-19 network model

As a pre-processing model, VGG-19 CNN is used. Compared to conventional CNN, it has improved in network depth. It uses a replacement system that is far stronger than a single convolution with many convolutional layer and non-linear layers of operation.

Using Max pooling for down sampling and modifying the linear unit (ReLU) as the actuation function, the layer structure can better extricate image highlights, that is, the largest value is chosen as the zone's pooled value within the image zone. The down sampling layer is effectively used to enhance the network's ability to distort the signal, while maintaining the sample's key characteristics and reducing the number of parameters [20].

$$\chi_{pj}^{(n)} = f\left(\tau_j^n\ dwon\ \chi^{(n-1)} + b_j^{(n)}\right) \qquad (2)$$

The expression for the down sampling layer as in (2 Among them, the maximum pooling sampling function is $down\ \chi^{(n-1)}$), τ_j^n is the coefficient corresponding to the j-th feature map of the n -th layer, and $f\left(\tau_j^n\ dwon\ \chi^{(n-1)} + b_j^{(n)}\right)$ is the ReLU activation function.

VI CONCLUTION

Image Style Transfer has significant development within the last few decades. This field has steadily attracted observers after a few scientists introduced the primary neural style transfer algorithm. We present a few picture type transfer algorithms that rely on non-deep learning as well as deep learning techniques in this search. Non-deep learning algorithms include image recreation and style highlighting state, despite the need to think about high-level semantic details. Deep learning approaches have evolved exponentially in recent years as a result of the inspiration of these algorithms. Increasingly inquire about concentrates on progressing the transfer quality and speed. In any event, most participants at the show judge the test results based on their own individual opinions, necessitating a more widespread assessment of picture transfer quality in the future.

REFERENCES

[1] A. Efros and T. K. Leung. Texture synthesis by nonparametricsampling. In Computer Vision, 1999. The Pro-ceedings of the Seventh IEEE International Conference on,volume 2, pages 1033–1038. IEEE, 1999.

[2] L. Wei and M. Levoy. Fast texture synthesis using treestructuredvector quantization. In Proceedings of the 27thannual conference on Computer graphics and interactivetechniques, (2000).

[3] A. A. Efros and W. T. Freeman. Image quilting for texture synthesis and transfer. In Proceedings of the 28th an- nual conference on Computer graphics and interactive tech- niques, (2001).

[4] V. Kwatra, A. Sch¨odl, I. Essa, G. Turk, and A. Bobick. Graphcut textures: image and video synthesis using graph cuts. In ACM Transactions on Graphics (ToG), (2003).

[5] J. B. Tenenbaum and W. T. Freeman. Separating style and content with bilinear models. Neural computation, (2000).

[6] D. P. Kingma, S. Mohamed, D. Jimenez Rezende, and M. Welling. Semi-supervised Learning with Deep Generative Models. In Z. Ghahramani, M.Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, Advances in Neu- ral Information Processing Systems 27, (2014).

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, (2012).

[8] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. Oct. (2013).

[9] S. Karayev, M. Trentacoste, H. Han, A. Agarwala, T. Darrell, A. Hertzmann, and H. Winnemoeller. Recognizing image style. arXiv preprint, (2013).

[10] A. Mahendran and A. Vedaldi. Understanding Deep Image Representations by Inverting Them. arXiv:1412.0035 [cs], Nov. 2014.

[11] Ashikhmin, N. Fast texture transfer. IEEE Computer Graphics and Applications, (2003).

[12] Lee, H., Seo, S., Ryoo, S., & Yoon, K.. Directional texture transfer.In Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering, June (2010).

[13] Efros, A. A., & Freeman, W. T.. Image quilting for texture synthesis and transfer. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques, August (2001).

[14] Simonyan, K., & Zisserman, A. Very deep convolutional networks for large-scale image recognition. (2014).

[15] KISHORE, P.V.V., KISHORE, S.R.C. and PRASAD, M.V.D. Conglomeration of hand shapes and texture information for recognizing gestures of indian sign language using feed forward neural networks. International Journal of Engineering and Tech-nology, (2013).

[16] LeCun, Yann; Léon Bottou; Yoshua Bengio; Patrick Haffner (1998). "Gradient-based learning applied to document recognition" (PDF). Proceedings of the IEEE. October (2016)

[17] https://www.completegate.com/2017022864/blog/deep-machine-learning-images-lenet-alexnet-cnn/all-pages

[18] LUO H,YANGY,TONG B,et al ,Traffic sign recognition using a multi-task convolutional neural network.. (2017).

[19] Liu Tong,Wang Zheng. HiCNN [J], a very deep convolutional neural network to better enhance theresolution of Hi-C data. (2019).

[20] Wen Rui ,Zhang Lin,Yuan Fei Niu, and Zeng Xia Ling, Review of Full ConvolutionalNeural Networks (2019).