Audio Style Transfer Learning

A Survey on Neural Style transfer using Deep Neural Networks

Shreya Prakash
shreya.prakash10d@gmail.com

Bangaru Mohnish
mohnish.b1@gmail.com

Mrs. M. Harini
harini.m@ktr.srmuniv.edu.in

Computer Science and Engineering,
SRM Institute of Science and Technology
Tamil Nadu, India

***Abstract*—Neural Style Transfer refers to a set of algorithms that manipulate and modify images . It is a concept which has been applied to image domain mostly with the example of creating a Van Gogh painting from any given input image. Aim of this review analysis is to understand this concept on audio domain as well. We have consolidated works related to how the style of an audio or image can be transferred to another content audio or image. The artificial neural network models mentioned in these works pave the way to generate a new audio or image with the general features of "style" by also remaining loyal to the "content". Our survey deals with neural models on images, audio processing techniques and existing models on audio. We can play with the features of music and create something new.**

***Keywords*— Neural Style Transfer, Artificial Neural Network, style audio, content audio**

I. Introduction

Music and art play important roles in our lives. Neural style transfer enables any person to become an artist. People can transfer the style of their chosen artist's audio/image to another content audio/image and generate a new output. Other than music, Neural Style Transfer can be applied on audio to enable people with speech problems become an orator. We need to understand this concept on images before we experiment it on audio. Audio processing techniques have to be applied to convert audio to spectrogram and vice versa. A spectrogram is a pictorial representation of a one-dimensional audio signal. Artificial Neural Networks have to be employed to perform the blending of both content and style audios.

II. *LITERATURE SURVEY*

A. Universal Style Transfer via Feature Transforms (2017)

Authored by: *Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, Ming-Hsuan Yang*

The study demonstrates a model to transfer styles to content images and produce high quality output. Feed-forward based methods are able to blend content and style images but they are unable to work on unseen styles or low visual quality. The model presented in this paper deals with such problems and it does not need to train on any pre-defined styles.

Approach

The main problem is extracting the true representation of the style and then blend it in the content image. Work by Gatys et al. prove that the Gram matrix or covariance matrix generated by a trained deep neural network has the desired ability of extracting visual styles.

First, feature extraction is done by using the VGG-19 network (19 layered CNN) , then these features are inverted to the original image by a symmetric decoder (image reconstruction) next WCT (whitening and coloring transforms) is applied. The blended features are further fed forward into the decoder layers in order to generate the stylized image. This algorithm does not need to train on style images unlike the old algorithms. Extraction of feature gram matrices and application to the content features through WCT is the way to deal with new styles.

Inference

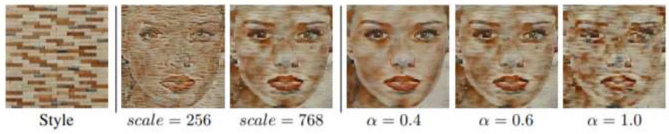
The efficiency of WCT is shown by comparing it with a popular feature adjustment technique, HM (histogram matching). It is evident in their output images that the HM technique does the job but could not capture prominent visual patterns like patterns that are torn into several bits and local structures are not shown properly. On the other hand, WCT technique is able to understand patterns that represent the style image better. This is because the HM technique is unable to see the correlations among features, which is what the gram matrix is made for.

Table 1: Differences between our approach and other methods.

	Chen et al. [3]	Huang et al. [15]	TNet [27]	DeepArt [9]	Ours
Arbitrary	√	√	×	√	√
Efficient	√	√	√	×	√
Learning-free	×	×	×	√	√

Table 2: Quantitative comparisons between different stylization methods in terms of the covariance matrix difference (L_s), user preference and run-time, tested on images of size 256×256 and a 12GB TITAN X.

	Chen et al. [3]	Huang et al. [15]	TNet [27]	Gatys et al. [9]	Ours
$\log(L_s)$	7.4	7.0	6.8	6.7	**6.3**
Preference/%	15.7	24.9	12.7	16.4	**30.3**
Time/sec	2.1	0.20	0.18	21.2	**0.83**



As art is highly subjective, they conducted a user study to evaluate all the different methods. They used five content and thirty style images to obtain one hundred and fifty different result output images on each pair of content/style for every technique. Then randomly chose fifteen style images for every subject to analyze. They showed stylized images by the compared techniques together in random order. People are asked to vote their favorite for each style, and their algorithm was the most popular choice.

As art is highly subjective, they conducted a user study to evaluate all the different methods. They used five content and thirty style images to obtain one hundred and fifty different result output images on each pair of content/style for every technique. Then randomly chose fifteen style images for every subject to analyze. They showed stylized images by the compared techniques together in random order. People are asked to vote their favorite for each style. They collected feedback and it showed that their algorithm received the most votes for better stylized results

B. Image Style Transfer Using Convolutional Neural Networks (2016)

***Authored by:** Leon A. Gatys, Alexander S. Ecker, Matthias Bethge*

This study demonstrates an algorithm that is able to separate and recombine the style and content images. It generates new images of high visual quality that blends the content of an input image with the style of various popular paintings. Their results show that the Convolutional Neural Networks are able to perform high level image synthesis and manipulation. A style transfer algorithm has to obtain the effective representation from the target image (e.g a landscape). It has to further perform a texture synthesis to render the target image in the style of the artwork.

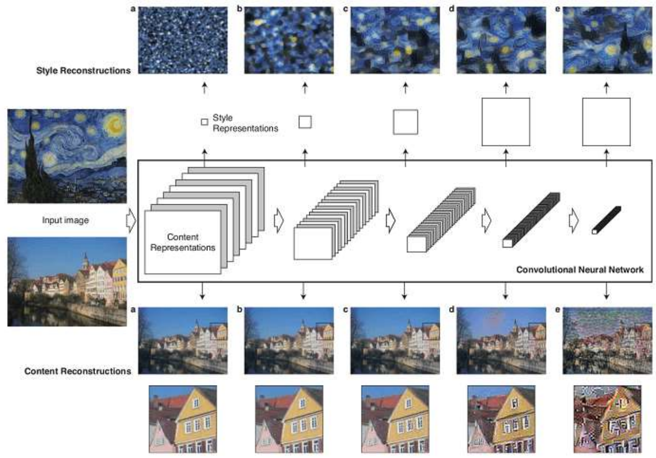
Approach

The transfer process is carried out in two modules one for content and the other for style.

The source images are passed through a convolutional neural network with custom filters to extract local and global features of the content image and similarly for the style image, we design a feature space to capture the texture information it consists of relations between the different filter responses.

They match the content representation of a photograph depicting the riverfront of the Neckar river in T¨ubingen,
The total loss is computed with the weighted sum of the content loss and the style loss, this loss is further minimized using an optimizer.

The flow diagram of the proposed system:



Inference

The key finding of this paper is that the representations of content and style in the Convolutional Neural Network are well separable to independently to produce new, perceptually meaningful Images.

C. Neural Style Transfer: A Review (2018)

***Authored by**: Yongcheng Jing, YezhouYang, ZunleiFeng, JingwenYe, Mingli Song*

This study presents an overview of Neural Style Transfer Models. It evaluates and compares various algorithms used for NST. Finally it discusses about the several applications and scope for future development. The authors help us choose the optimal method for style transfer. It shows the numerical results of every method helping in accurate comparison. Gatys et al. first demonstrated the use of Convolutional Neural Network to regenerate popular artwork styles on natural images.

Approach/Information Gained

The study focuses on image-based artistic rendering (IB-AR)

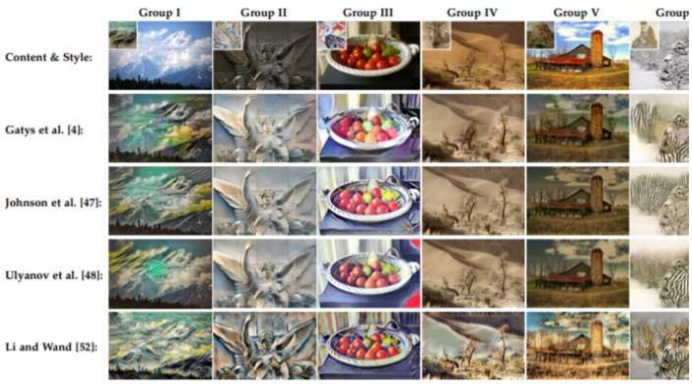
Stroke-Based Rendering(SBR)- It denotes a procedure of putting strokes on a canvas digitally to create a painting with a particular style. It is successful in showing a particular style but it can only follow one style. It is perfect for oil paintings, water colours and sketches.

Region-Based Techniques- It simplifies shape rendering effects by placing canonical shapes instead of regions. It can only follow one style like SBR.
Example-Based Rendering- It maps between a content images pair and target stylized images pair in a supervised way. They can only capture low level features. Hence they are not ideal.

Image Processing and Filtering- It simplifies and abstracts images. They are effective but cannot adopt different styles.

All these limitations of IB-AR techniques paved way for the rise of Neural Style Transfer. It is the procedure of extracting style from an image then using it for content image reconstruction.

Here are 5 different NST algorithms tried on six different content and style images. This is the qualitative comparison.



Average speed comparison for sizes in pixels

Methods	Time(s)			Styles/Model
	256×256	512×512	1024×1024	
Gatys et al. [10]	14.32	51.19	200.3	∞
Johnson et al. [47]	0.014	0.045	0.166	1
Ulyanov et al. [48]	0.022	0.047	0.145	1
Li and Wand [52]	0.015	0.055	0.229	1
Zhang and Dana [56]	0.019 (**0.039**)	0.059 (**0.133**)	0.230 (**0.533**)	$k(k \in Z^+)$
Li et al. [55]	0.017	0.064	0.254	$k(k \in Z^+)$
Chen and Schmidt [57]	0.123 (**0.130**)	1.495 (**1.520**)	–	∞
Huang and Belongie [51]	0.026 (**0.037**)	0.095 (**0.137**)	0.382 (**0.552**)	∞
Li et al. [59]	0.620	1.139	2.947	∞

Comparison for efficiency, arbitrary style and learning-free

Methods	E	AS	LF
Gatys et al. [4]	×	√	√
Ulyanov et al. [47]	√	×	×
Johnson et al. [50]	√	×	×
Li and Wand [52]	√	×	×
Dumoulin et al. [53]	√	×	×
Chen et al. [54]	√	×	×
Li et al. [55]	√	×	×
Zhang and Dana [56]	√	×	×
Chen and Schmidt [57]	√	√	×
Ghiasi et al. [58]	√	√	×
Huang and Belongie [51]	√	√	×
Li et al. [59]	√	√	√

Prisma is an app for converting a natural image into a painting in a style of famous artwork. It uses Neural Style Transfer and it has achieved a lot of popularity. Ostagram is a similar app providing faster speed.
NST can be useful for painters, fashion designers and artists who want to experiment with their ideas.
Creation of animations will become simpler and easier if NST is used for it.

Inference

NST is rising but it still has a long way to go. There is a need to refine the new techniques, looking for properly fitting different kinds of styles. It performs well on paintings but for some styles, it produces irregular output due to image construction from CNN. It works on natural content images perfectly but fails when it takes an abstract image as content.

D. Deep Photo Style Transfer (2017)

Authored by: *Fujun Luan, Sylvain Paris, Eli Shechtman, Kavita Bala.*

The authors of this research paper points the transfer learning concept in a new direction by applying transformations from the input to the output to be locally affine in colorspace of the image, and to express this constraint as a custom fully differentiable energy term.

Approach

Introduces a deep-learning approach that faithfully transfers style from a reference image for a wide variety of image content.
The transfer had to address the major challenges:

1. Structure Preservation
2. Semantic accuracy

The transfer of style to the image must be done in a manner in which the image doesn't lose its basic edges and the structural details of the content this is achieved by restricting the results to be photorealistic hence differentiating the Photographs from the paintings.

The Semantic accuracy plays an important role when the data being styled is a photograph such as the different objects in the picture are not mismatched during the transfer

The style transfer uses a photorealism regularization parameter in the objective function for optimization to prevent distortions. This approach is compared against the CNNMRF

The total loss term computed in this process is governed by the following equation (CNNMRF):

$$\mathcal{L}_{\text{total}} = \sum_{\ell-1}^{L} \alpha_\ell \mathcal{L}_c^\ell + \Gamma \sum_{\ell-1}^{L} \beta_\ell \mathcal{L}_s^\ell \qquad (1a)$$

Equation with the photorealistic regularization parameter (λ) (New Approach):

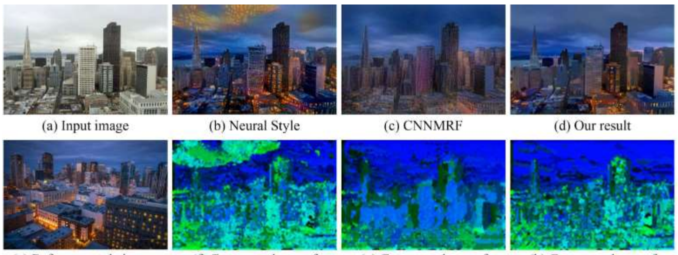
$$\mathcal{L}_{\text{total}} = \sum_{l=1}^{L} \alpha_\ell \mathcal{L}_c^\ell + \Gamma \sum_{\ell=1}^{L} \beta_\ell \mathcal{L}_{s+}^\ell + \lambda \mathcal{L}_m \qquad (4)$$

The results achieved by using and changing the photorealistic regularization parameter:

(a) Input and Style (b) $\lambda = 1$ (c) $\lambda = 10^4$, our result (d) $\lambda = 10^8$

Figure 3: Transferring the dramatic appearance of the reference style image onto an ordinary flat shot in (a) is challenging. We produce results using our method with different λ parameters. Too small a value of λ cannot prevent distortions, and thus results in a non-photorealistic look in (b). Conversely, too large a value of λ suppresses the style to be transferred yielding a half-transferred look in (d). We found the best parameter $\lambda = 10^4$ to be the sweet spot to produce our result (c) and all the other results in this paper.

Comparative analysis of different methods:



(a) Input image (b) Neural Style (c) CNNMRF (d) Our result

(e) Reference style image (f) Correspondence of (b) and (e) (g) Correspondence of (c) and (e) (h) Correspondence of (d) and (e)

Inference

The Neural Style algorithm doesn't succeed in isolating the different patches of the image for the style transfer and causes distortion in the resultant image the original image CNNMRF model fails to identify the different portions of the image and causes improper or blurred resultant image, therefore adding the photorealistic regularization factor along with the Structure Preservation and the Semantic accuracy filters dramatically increases the effectiveness of the transfer.

E. Audio Style Transfer (2018)

Authored by: *Eric Grinstein, Ngoc Q. K. Duong, Alexey Ozerov, Patrick P´erez*

The authors of this research paper propose a flexible framework for the task, which uses a sound texture model to extract texture statistics followed by optimization and the stylized signal generation.

Approach

The authors of this research paper have approached the style transfer of audio signals using a new approach by synthesizing a custom audio texture model inspired by the human hearing mechanism.

The transfer is carried out in two stages:

1. Style extraction
2. Style transfer

The content and the style waveforms are transformed into 2D spectrogram using the Short-term Fourier Transform and the texture statistics are extracted by the sound texture model (Neural Network) after the extraction of the texture statistics the optimization algorithm is used to generate the final signal and finally the post-processing is done to convert the signal into its waveform.

This approach stands out as it does not accommodate content loss and the final signal is optimized to minimize the style loss.

Style Loss Function:

$$\mathcal{L}(\mathbf{x}; \mathbf{x}_{\text{sty}}) = \sum_{\ell \in L} \left\| \mathsf{G}_\ell(\mathbf{x}) - \mathsf{G}_\ell(\mathbf{x}_{\text{sty}}) \right\|_F^2, \qquad (1)$$

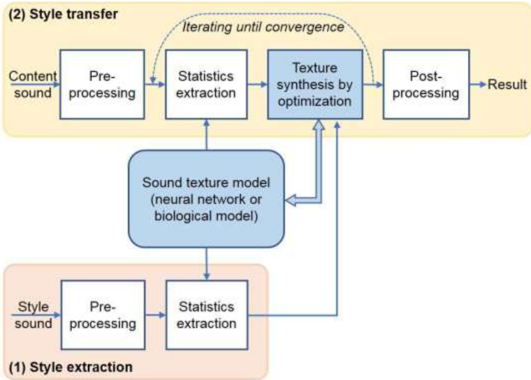


Fig. 2. Proposed audio style transfer framework: Given an audio texture extraction model (artificial neural net or auditory model), the *content* sound is iteratively modified such that its audio texture matches well the one of the *style* sound. If required by texture model, raw signals are mapped to and from a suitable representation space by pre/post-processing.

Inference

Experiments conducted by the authors point that the sound texture model based on the human auditory system and the shallow random net models produced significant results. The author also concludes that initializing the iterative optimization by the content sound, plays an important role in obtaining these results.

F. Neural Style Transfer for Audio Spectrograms (2018)

Authored by: *Prateek Verma, Julius O. Smith*

The Authors of this Research Paper aim to present a machine learning technique for style transfer on audio signals analogous to the approach of Gatys et al on images by using a convolutional neural network, and investigates the generation of spectrograms from noise

Approach

The authors approached the problem of style transfer on audio signals by conducting two fundamental experiments:

1.Imposing the style of a tuning fork on a harp
2.transferring the style of a violin to a voice

The authors explored the problem thoroughly and noted the various hyper-parameters that need to be tuned to achieve the results for the two experiments and aimed to have one parameter setting that can perform well in both the scenarios. Such that hand tuned parameters for each scenario can be avoided.

The objective equation used:

$$X_{recon} = \operatorname{argmin}_X \mathcal{L}_{\text{total}} = \operatorname{argmin}_X \alpha L_c(x, x_c) + \beta L_s(x, x_s) + \gamma L_e(x_e, e_s) + \delta L_t(x_t, t_s).$$

Where X denotes the reconstructed signal, L_c is the content loss and L_s is the style loss the goal is to minimize the sum of the weighted loss terms of both the signals.

Results:

Experiment 1:

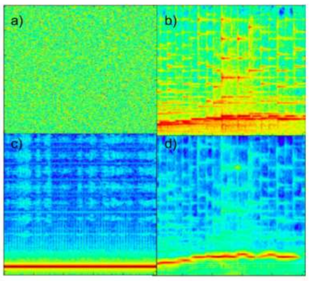


Figure 1: a) shows the Gaussian noise from which we start the input to optimize, b) Harp sound (content) c) Tuning Fork (style) and d) Neural Style transferred output with having content of harp and style of tuning fork `https://youtu.be/UlwBsEigcdE`

Experiment 2:

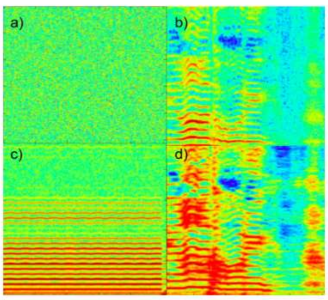


Figure 2: a) shows the Gaussian noise from which we start the input to optimize, b) Singing sound (content) c) Violin note (style) and d) Neural Style transferred output with having content of singing and style of violin. `https://youtu.be/RpGBkfs24uc`

Inference

The authors of the research paper have proposed a novel way to handle audio signals by considering them to be style transfer problems, by using the back propagation technique to optimize the output sound to conform to the filter outputs of a pre-trained neural network architecture. The approach proposed is very flexible and can be applied to many varieties of scenarios as discussed above.

G. Signal Estimation from Modified Short-Time Fourier Transform (1984)
1

Authored by: *Daniel W. Griffin, Jae S. Lim*

The authors of this research paper aim to present an algorithm to estimate a analogue signal from its modified Short-term Fourier transform.

Approach

The proposed algorithm is the iterative short-term Fourier transform it works on the concept of decreasing the ea squared error between the Short-term Fourier transform of the signal and the Modified Short-term Fourier transform of the signal. In the time scale modification of speech.

The major computation of the proposed algorithm is the Discrete Fourier Transform.

The Authors present comparative study between the existing algorithm and the proposed algorithm

Existing algorithm: LSEE-MSTFT
Proposed algorithm: OA-MSTFT
The LSEE-MSTFT algorithm can be applied in two methods
Overlap add method

1. Iteratively

The goal of these algorithms is to minimize the mean squared distance between the STFT and the MSTFT of the analogue signal.

Distance Equation:

$$D[x(n), Y_w(mS, \omega)] = \sum_{m=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} [x_w(mS, l) - y_w(mS, l)]^2 .$$

Results:
The results have been generated for the following audio signal:

Fig. 2. STFTM of "line up at the screen door."

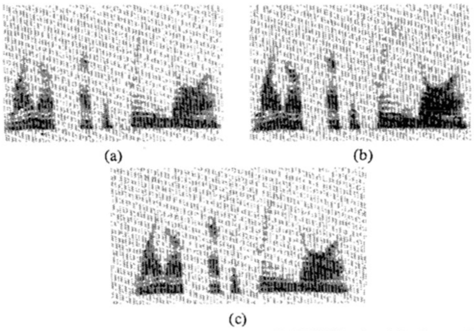


Fig. 3. (a) 128:64 time-scale compressed STFTM of original speech. (b) STFTM of LSEE-MSTFTM estimate. (c) STFTM of OA-MSTFTM estimate.

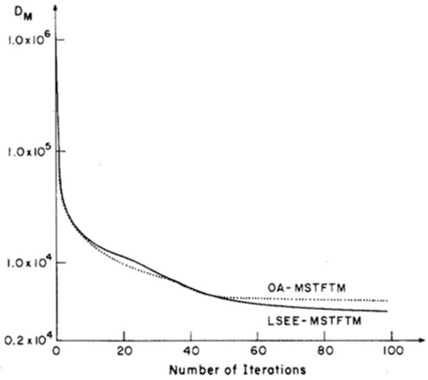


Fig. 4. D_M versus iteration number of LSEE-MSTFTM and OA-MSTFTM.

Inference

The paper presents 3 new algorithms and presents the comparative study between them as shown above
It can be inferred that even though LSEE-MSTFTM and OA-MSTFTM had similar performance LSEE-MSTFTM has always produced minimal DM as compared to OA-MSTFTM Over large number of iterations and can cause significant differences in different applications.

H. Fast Signal Reconstruction From Magnitude Stft Spectrogram Based On Spectrogram Consistency (2010)

Authored by: *Jonathan Le Roux, Hirokazu Kameoka, Nobutaka Ono,Shigeki Sagayama*

The Research paper presents the latest advancements on the applications of Short-term Fourier Transform in Signal Reconstruction from an altered Spectrogram.

Approach

This article reviews the concept of spectrogram consistency and the method to derive the numerical consistency criterion and proceeds to assess and investigate the performance of the fast phase estimation algorithms.

The algorithms investigated are the Griffin and Lim's iterative Short-term Fourier transform and the Fast minimization of (I)

Griffin and Lim proposed the iterative STFT algorithm, which consists in iteratively updating the phase at step k by replacing it with the phase of the STFT of its inverse STFT, while keeping the A(magnitude) fixed.

The Fast Minimization of I algorithm aims to cut the cost of computing the STFT and the Inverse STFT of a signal by taking advantage of sparseness of the magnitude array(A) and truncating A accordingly and proceeding with the phase estimation.

The experimental analysis was conducted on 100 speech samples of both male and female candidates, taken from Bagshaw's database for the length of 5 minutes and 8 portions of musical pieces of piano and guitar solos from the RWC music database for the length of 3 minutes

Inference

The authors introduced a framework to estimate the phase of the signals by studying the relation between Griffin and Lims's method and their own method the Fast minimization of (I) and developed the framework to take advantage of the sparseness of A(Magnitude array) and perform the estimation in a more efficient manner.

I. Adam: A Method For Stochastic Optimization (2015)

Authored by: *Diederik P. Kingma, Jimmy Lei Ba*

The Authors of the Research paper introduced a Stochastic Gradient optimization algorithm that is computationally efficient and has minimum memory requirement and which can be used for problems which are large in terms of parameters and data and that can be applied to methods with noise in the data or has sparse gradients.

Approach

Proposed Algorithm:

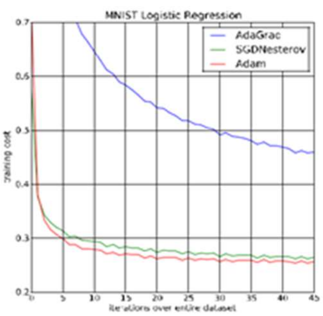
Let f be a noisy stochastic scalar function that is differentiable. The goal of the algorithm is to minimize the expected value of this function with respect to its parameters.

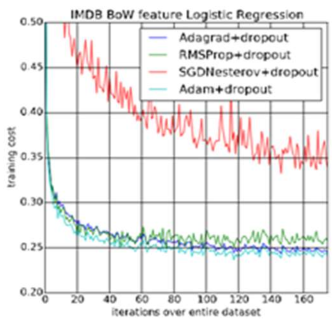
Pseudocode for the algorithm:

Algorithm 1: *Adam*, our proposed algorithm for stochastic optimization. See section 2 for details, and for a slightly more efficient (but less clear) order of computation. g_t^2 indicates the elementwise square $g_t \odot g_t$. Good default settings for the tested machine learning problems are $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. All operations on vectors are element-wise. With β_1^t and β_2^t we denote β_1 and β_2 to the power t.

Require: α: Stepsize
Require: $\beta_1, \beta_2 \in [0, 1)$: Exponential decay rates for the moment estimates
Require: $f(\theta)$: Stochastic objective function with parameters θ
Require: θ_0: Initial parameter vector
  $m_0 \leftarrow 0$ (Initialize 1st moment vector)
  $v_0 \leftarrow 0$ (Initialize 2nd moment vector)
  $t \leftarrow 0$ (Initialize timestep)
  while θ_t not converged **do**
    $t \leftarrow t + 1$
    $g_t \leftarrow \nabla_\theta f_t(\theta_{t-1})$ (Get gradients w.r.t. stochastic objective at timestep t)
    $m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$ (Update biased first moment estimate)
    $v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$ (Update biased second raw moment estimate)
    $\widehat{m}_t \leftarrow m_t / (1 - \beta_1^t)$ (Compute bias-corrected first moment estimate)
    $\widehat{v}_t \leftarrow v_t / (1 - \beta_2^t)$ (Compute bias-corrected second raw moment estimate)
    $\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \widehat{m}_t / (\sqrt{\widehat{v}_t} + \epsilon)$ (Update parameters)
  end while
  return θ_t (Resulting parameters)

Results:
Performance of Adam Optimizer on MNIST and IMDB data





Inference

The Authors have successfully created a Stochastic Gradient Optimizer that is more efficient in regression tasks and out performs the existing optimizing methods such as AdaGrad and SGDNesterov.

J. Arbitrary Style Transfer with Style-Attentional Networks (2019)

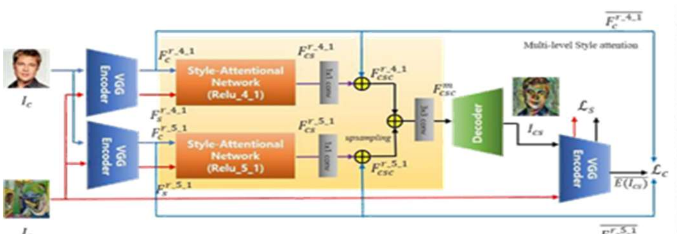
Authored by: *Dae Young Park, Kwang Hee Lee*

The Research Paper introduces and proposes the use of a novel style attention network called the (SANet) and a new Identity loss function and multi-level feature embeddings to tackle the issue of maintaining local and global style patterns and the content structure.

Approach

The Style transfer operation proposed in the paper is composed on an encoder-decoder module and the Style-attentional module. The SANet Architecture is as follows, the encoder module and the symmetric decoder module is a pre-trained VGG-19 network.
To capture the global and the local style patterns two SANet's have been used the inputs take are the feature maps encoded from arbitrary layers of the VGG, the output is generated by combining the output feature maps. Adam optimizer was used with the earning rate of 0.0001 during the training of the model.

Network Architecture:



(a) Network Architecture

The Model was trained using MS-COCO for the content images and WikiArt for the style images,the datasets contained roughly 80,000 images.

Method	Time (256 px)	Time (512 px)
Gatys et al. [5]	15.863	50.804
WCT [13]	0.689	0.997
Avatar-Net [20]	0.248	0.356
AdaIN [7]	0.011	0.039
ours (Relu_4_1)	0.012	0.042
ours (multi-level)	0.017	0.055

Table 1: Execution time comparison (in seconds).

Inference

The Authors of the Paper have proposed a new algorithm to perform Style transfer tasks effectively and efficiently by using their proposed SANet architecture in the form of an encoder-decoder module and has achieved better results and quicker runtimes than the existing methods except the AdaIN method which has shown faster run-times.

III. RESULTS AND DISCUSSIONS

The research papers discussed in section [II] have provided us with information and the statistical performance data of different approaches and machine learning models which aided us in gaining inspiration and the necessary knowledge to successfully find an innovative approach to transfer learning on audio signals.

IV. CONCLUSION

We have studied and interpreted the research papers in the section [II] and converged on an approach to implement Transfer Learning on an audio signal by stitching together different concepts implemented in the papers.

Our Approach is set to employ the preprocessing techniques form section [G], and the development of human auditory system inspired CNN and the pipeline architecture used in section [E] in the aim to achieve an enhanced performance and a viable contribution to the field of Transfer Learning

V. REFERENCES

[1] Eric Grinstein, Ngoc Q. K. Duong Alexey Ozerov, Patrick P´erez, “Audio style transfer,” presented in IEEE International Conference on Acoustics, Speech and Signal Processing, 2018.

[2] Prateek Verma, Julius O. Smith " Neural Style Transfer for Audio Spectrograms" Appeared in 31st Conference on Neural Information Processing Systems, 2018.

[3] Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, Ming-Hsuan Yang, " Universal Style Transfer via Feature Transforms " Accepted by NIPS 2017.

[4] Leon A. Gatys, Alexander S. Ecker, Matthias Bethge " Image Style Transfer Using Convolutional Neural Networks" presented in IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[5] Yongcheng Jing, YezhouYang, Zunlei Feng, JingwenYe, Mingli Song, " Neural Style Transfer: A Review" presented in 2018.

[6] Fujun Luan, Sylvain Paris, Eli Shechtman, Kavita Bala, " Deep Photo Style Transfer" presented in IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[7] Daniel W. Griffin, Jae S. Lim" Signal Estimation from Modified Short-Time Fourier Transform" Published in IEEE Transactions on Acoustics, Speech, and Signal Processing, Apr 1984

[8] Jonathan Le Roux, Hirokazu Kameoka, Nobutaka Ono, Shigeki Sagayama "Fast Signal Reconstruction From Magnitude Stft Spectrogram Based On Spectrogram Consistency" Proc. of International Conference on Digital Audio Effects, 2010.

[9] Diederik P. Kingma, Jimmy Lei Ba "Adam: A Method For Stochastic Optimization" Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015

[10] Dae Young Park, Kwang Hee Lee “Arbitrary Style Transfer with Style-Attentional Networks”; presented in IEEE Conference on Computer Vision and Pattern Recognition, 2017.