OAIJSE

OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

DEEP PHOTO STYLE TRANSFER USING WAVELET TRANSFORM

Praveen Shinde[1], Prajwal Sakle[2], Abhishek Gupta[3], Sandip Vidhate[4] Namrata Pagare[5]

Student, *Department of Computer Engineering, KKWIEER, Nashik, India*[1,2,3,4]
Professor, Department of Computer Engineering, KKWIEER, Nashik, India[5]
prvshinde777@gmail.com [1] , *sakleprajwal@gmail.com* [2] , *abhishekguptaab567@gmail.com* [3] ,
sandipvidhate@gmail.com [4] *nmpagare@kkwagh.edu.in*[5]

Abstract: Photorealistic Style transfer models have specified promising artistic results. This system introduces a deep-learning approach to photorealistic style transfer that handles an outsized sort of image content while transferring the reference style. The approach used in this paper is build upon the recent work on painterly transfer that separates style from the content of a picture by considering different layers of a neural network. A neural algorithm of artistic style proposes an iterative algorithm for neural style transfer. The neural style transfer algorithm involves taking a content image C and a style image P and combine them to produce a new image that has the content of C and the style of P. Photorealistic image stylization involves transferring style of a styled photo to a content photo with the limitation that the stylized photo should remain photorealistic. However, if given a photograph as a reference style, present methods are limited by spatial biases or improbable artefacts, which should not happen in real photographs. The proposed network architecture enhances photorealism and transfers the style. The key factor of our method is wavelet transforms that naturally fits in deep networks. To preserve the structural information and features we have proposed a wavelet corrected transfer method in this paper that support whitening and colouring transform.

Keywords: *Image stylization, Photorealism, Haar wavelet, VGG*

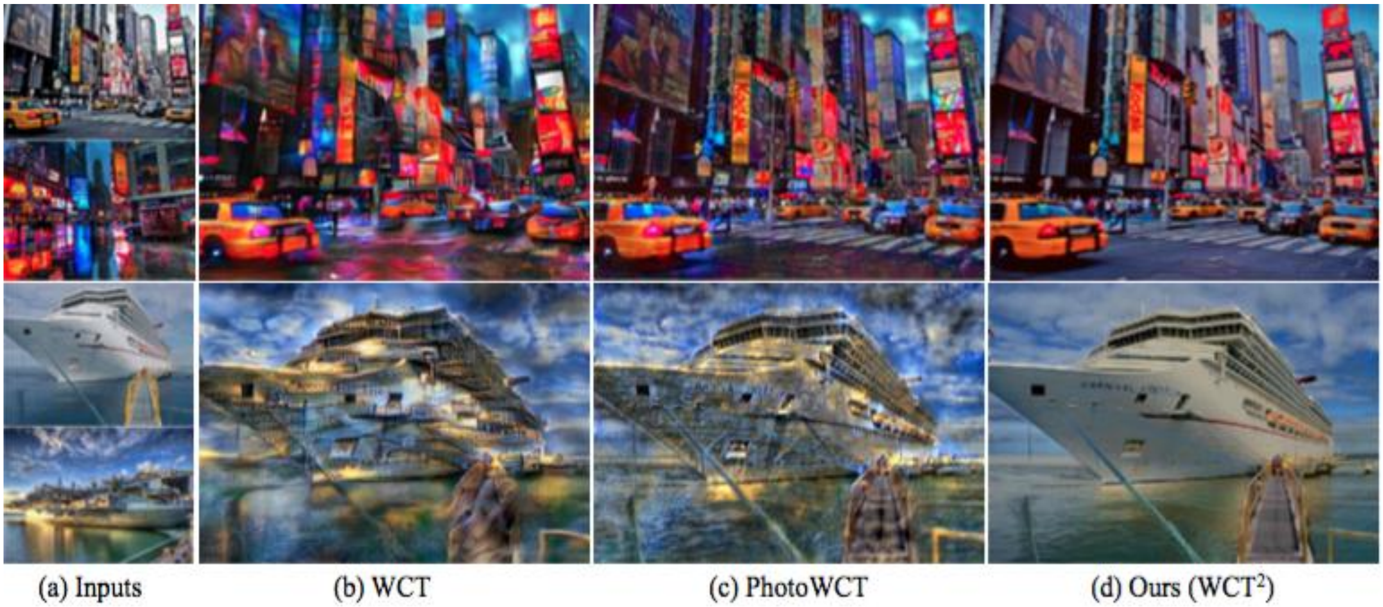

Figure 1 Photorealistic Stylization Results

I INTRODUCTION

Photorealistic style transfer has to satisfy various objectives. To be photorealistic image, a model should apply the reference style on the content image i.e scene without distorting the details of an image. In Figure 1, for example, the general style (color and tone) of sky and sea should change, while the fine structures of the ship and the bridge remain intact. However, solving the optimization problem requires heavy computational costs, which limits their use in practice. To overcome this issue, Li et al[12] recently proposed a photorealistic variant of WCT (PhotoWCT) by replacing the upsampling components of the VGG decoder with unpooling. By providing a max-pooling mask, PhotoWCT is designed to compensate for information loss during the encoding step and suppress the spatial distortion. Although their approach was effective, the summary of the mask was not able to determine the information loss that comes from the max-pooling of VGG network. To fix the remaining artifacts, they had to perform a series of post-processing steps, which require the original image to patch up the result. Not only do these post-processing steps require cumbersome computation and time but they entail another unfavorable blurry artifact and hyper-parameters to manually set. Instead of providing partial amendments, we address the fundamental problem by introducing a theoretically sound correction on the downsampling and upsampling operations.

We propose a wavelet corrected transfer based on whitening and coloring transforms that substitutes the pooling and unpooling operations in the VGG encoder and decoder with wavelet pooling and unpooling. Our motivation is that the learned function by the network should have its inverse operation to enable exact signal recovery, and accordingly, photorealistic stylization. It allows WCT2 to fully reconstruct the signal without any processing steps, due to the favorable properties of wavelets providing less information loss. The disintegrated wavelet features provide interesting interpretations on the feature space as well, such as component-wise stylization and why average pooling is known to give better stylization than max-pooling. In addition, we propose progressive stylization instead of following the multi-level strategy that is used in WCT and PhotoWCT. To maximize the stylization effect, WCT and PhotoWCT recursively transformed features in a multi-level manner from coarse to fine. In contrast, we gradually transform features during a single pass. This allows two significant advantages over the others. First, our model is simple and efficient since we only have a single decoder during training as well as in the inference time. On the other hand, the multi-level strategy requires to train a decoder for each level without sharing parameters, which is inefficient in terms of the number of parameters and training procedure. This overhead remains in the inference time as well because the model requires to pass multiple encoder and decoder pairs to stylize an image. Second, by recursively encoding and decoding the signal with the lossy VGG networks, artifacts are amplified during the multi-level stylization. Because of wavelet operations and progressive stylization, our model does not have such a problem, and even more, it shows little error amplification when the multi-level strategy is employed. Our contributions are summarized as follows. We first show that the spatial distortions come from the network operations that cannot satisfy the reconstruction condition .By employing the wavelet corrected transfer and progressive stylization, we propose the first end-to-end photorealistic style transfer model that allows to remove the additional post-processing steps. Our model can process a high resolution image (1024×1024) in 4.7 seconds, which is 830 times faster than the state-of-the-art models, where PhotoWCT fails due to an out-of-memory issue and Deep Photo Style Transfer (DPST) takes 3887.8 seconds. Our experimental results show quantitatively enhanced visual quality in both SSIM and Gram loss and qualitatively being preferred by 62.21% in the user study. In addition, our model has three times fewer parameters than PhotoWCT and provides temporally stable stylization enabling video applications without additional constraints, such as optical flow**.**

II LITERATURE SURVEY

Mr. Youngjung Uh [1] proposed a wavelet corrected transfer method based on coloring and whitening transform which allows features to preserve their properties and structural information of VGG feature space during stylization. It can stylize a 1024*1024 resolution image in seconds.

Mr. Yijun Li [2] proposed a paper in which they created an algorithm that consists of a stylization step and a smoothing step. While the stylization step transfers the style of the style photo to the content photo, the smoothing step ensures spatially consistent styles.

Kamal Hasan Talukderl [3] proposed method to replace max pooling layers with wavelet pooling where the high frequency components(LH, HL, HH) are skipped to the decoder directly. Thus, only the low frequency component LL is passed to next encoding layer.In Wavelet Pooling Image signals are divided into four sub bands which are LL, LH, HL, HH as already discussed. LL Sub-band contains approximation of input signal or image i.e. it contains smooth features of input signal. HH, HL, LH Sub-band contains structural information such as horizontal, vertical, diagonal edge

Chen Fang [4] insinuates an algorithm. The key factor of algorithm is a pair of feature transforms, whitening and coloring, that is embedded to an image reconstruction network. The whitening and coloring transform method reflects a direct matching of feature covariance of the content image to a given reference image.

Karen Simonyan[5] proposed a method to investigate the effect of the convolutional neural network in depth based on its accuracy in the large-scale image recognition. Convolutional Neural Networks (CNNs) are a category of Neural Network that has proven very effective in areas such as image recognition and classification.

III METHODOLOGY

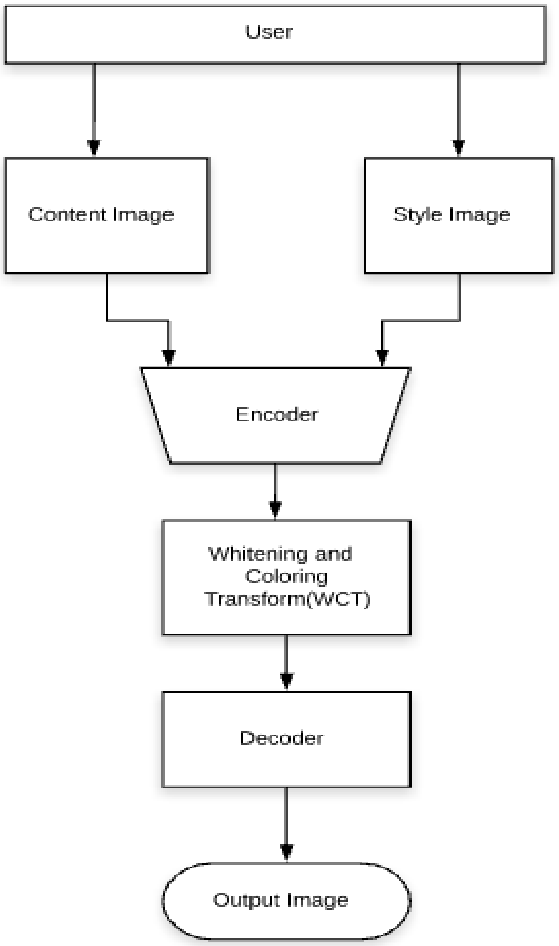

Figure 2 Block Diagram

Figure 2 shows a diagram of the proposed system. In this method, the user selects two images of content image and style image. These images are then fed to the encoder. The encoder then encodes the image and feed to Whitening and Coloring model for transformation. The whitening step helps extract the style from an input image while preserving the global content structure. The outcome of this operation is ready to be transformed with the target style. First, the multi-level strategy requires to train a decoder for each level without sharing parameters, which is inefficient. On the other hand, our training procedure is simple because we only have a single pair of encoder and decoder, which is advantageous in the inference time as well. Second, recursively encoding and decoding the signal with VGG network architecture amplifies errors causing unrealistic artifacts in the output. To avoid distortions we skip high-frequency subband and directly pass to the corresponding decoder and low pass filters only pass to next encoder. At the decoder, the components are aggregated by the wavelet unpooling. Thus by undergoing various layers finally output image is formed which is photorealistic.

IV SYSTEM ARCHITECTURE

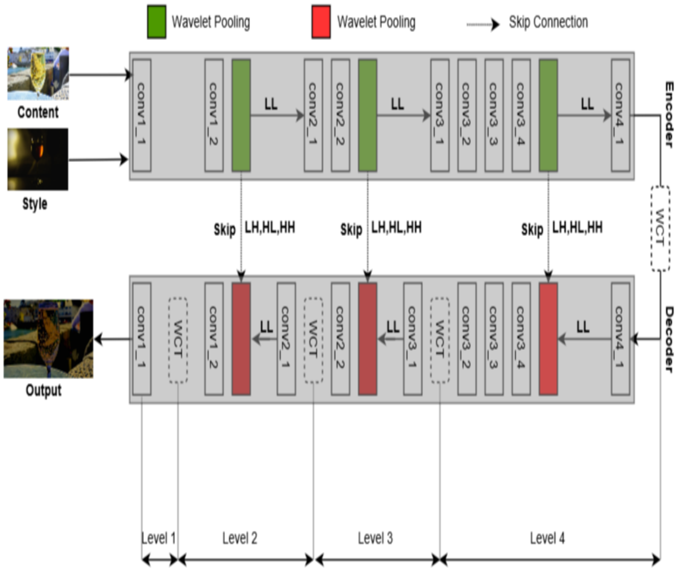

Figure 3 Overview of the proposed progressive stylization

As shown in Figure 3, the content and style image is passed to the encoder. The images are then passed to the first two convolutional layers and then the output of these layers is given to wavelet pooling operation. In wavelet pooling the image signals are divided into four sub-bands which are Low(L) and High(H) pass filters which are LL, LH, HL, HH. But here only low pass filters is passed to next convolutional layer while the high pass filters are skipped to wavelet unpooling operation in the decoder. If the high-frequency sub-bands are passed to the next encoder then it may cause distortions. To avoid distortions, high-frequency sub-bands are skipped and directly passed to the corresponding decoder. Thus simultaneously the output is passed until the final layer of convolution where feature maps are extracted from the content and style image. The features extracted in the form of output from the encoder is given to Whitening and Coloring Transform(WCT). The image thus obtained is fed to decoder. The first layer of decoder receives the output of WCT and is fed to wavelet unpooling operation which also receives the output of previous layer as LL(Low pass filter) which also receives input from previously skipped high-frequency bands(LL, LH, HH) and now the output thus obtained is passed to next convolution layer till the first layer. Thus after passing through the final convolution layer, we obtain final styled image i.e. the output image.

V WAVELET CORRECTED TRANSFER

3.1 Haar Wavelet Pooling and Unpooling

- Haar wavelet pooling has four kernels:-

{$LL^\top$, $LH^\top$, $HL^\top$,$HH^\top$}, where the low (L) and high (H) pass filters are

$$L^\top = \frac{1}{\sqrt{2}}\left[\, 1 \quad 1 \,\right], \quad H^\top = \frac{1}{\sqrt{2}}\left[\, -1 \quad 1 \,\right]$$

- The low-pass filter captures smooth surface and texture.
- High-pass filters extract vertical, horizontal, and diagonal edge like information.
- One important property of our wavelet pooling is that the original signal can be exactly reconstructed by mirroring its operation; i.e., wavelet unpooling.
- Maxpooling does not have its exact inverse so that the encoder-decoder structured networks used in the WCT and PhotoWCT cannot fully restore the signal
- We choose Haar wavelet because it splits the original signal into channels that capture different components, which leads to better stylization

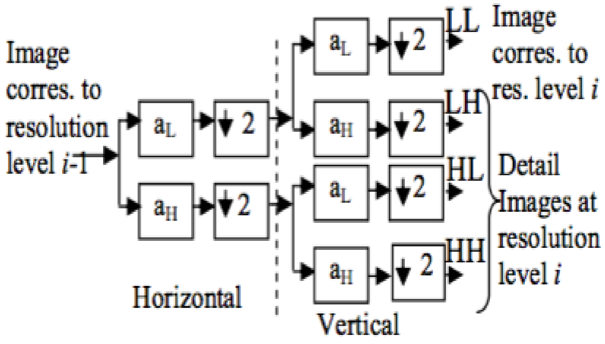

Figure 4. One Filter Stage in 2D DWT

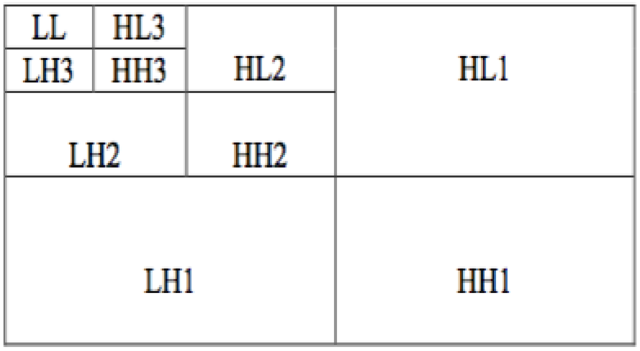

Figure 5. Structure of wavelet decomposition

As shown in Figure 4. starting image is passed through low pass filter aL and High pass filter aH. We apply subsampling by 2 for output of low pass filters and high pass filters. Note that we apply low pass filters and high pass filters and subsampling along the row i.e. horizontal. After this output, we have obtained we apply low pass filtering operation and high pass filtering and subsampling by two on the output of aL filter. In the same way, we apply low pass filtering operation and high pass filtering and subsampling by two on output of aH filter.In Second phase we apply filters along the column. i.e. Vertically. As shown in Figure 4 we got four output in the form of sub-bands as LL, LH, HL, HH. High pass filters extract the edges and low pass filter does approximation**.**

As shown in figure we output LL is passed through 2 low pass filters hence output LL will give an approximation to an input image. Output LH is passed through 1 low pass filter and 1 high pass filter and output LH will extract horizontal features input image. Output HL is passed through 1 low pass filter and 1 high pass filter and output HL will extract Vertical features input image. Output HH is passed through 1 low pass filter and 1 high pass filter and output HH will extract diagonal features input image.

As shown in Figure 5. subsampling is applied by a factor of 2 along the row as well as the column that means that we have done subsampling by a factor of 4.Therefore each of the subbands i.e. HH subband, HL subband, LL subband and LH subband contains ¼ of the total number of samples in our image. Anyone of these subbands or all subbands can be analysed or partitioned further. In our paper, we do not apply any further operations on HH, HL, and LH subband as it contains information about sharp edges and can cause distraction. Hence we directly pass these subbands to decoder and LL subband is passed to next encoder that we will see in architecture.

3.2 Whitening and Coloring Transform

We give a pair of content image Ic and style image Is, we first extract their vectorized VGG feature maps fc i.e features of content image ∈ RC ×Hc Wc and fs i.e. features of style image ∈ RC ×HsWs at a certain layer (e.g., Relu_4_1), where Hc , Wc (Hs , Ws) are the height and width of the content (style) feature, and C is the number of channels. The decoder will reconstruct the original image if features of the content image are directly fed into it. We also propose to use a whitening and coloring transform to adjust features of the content image with respect to the statistics of features of content image. The goal of WCT is to directly transform the features of the content image to match the covariance matrix of features of style image. It involves two steps, i.e., whitening and coloring transform.

Whitening Transform

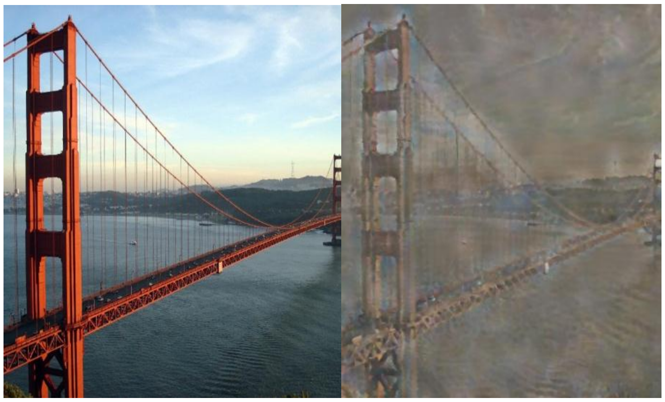

Figure 6. Effect of Whitening Transform

As we see in the Figure 6. which indicates that the whitened features still maintain global structures of the image contents, but greatly help remove other information related to styles. In other words, the whitening step helps peel off the style from an input image while preserving the content of an image. The output of this method is ready to be transformed with the target style.

Coloring Transform

It is inverse of the whitening method. In this method, we transfer the style of the style image to the image that obtains from previous whitening method, in this it finds the correlation between style image and content image to transfer the style i.e. to directly transform the feature map of the content image to match the covariance matrix of feature map of style image. For example, if there is the sky in both images then it finds the correlation such that the style of the sky is transferred to the sky of the content image.

3.3 Experimental Results

In this section, we show that our simple modification can be enhanced the performance of photorealistic style transfer. Here, every result is reported based on the concatenated version of our model. For a fair comparison and time-efficiency, we only perform whitening and coloring on LL (Low pass filter)components progressively.

	DPST	PhotoWCT	WCT
Fewest Artifacts	21.34%	9.33%	69.33%
Best Stylization	30.49%	12.74%	56.77%
Most Preferred	24.63%	11.16%	62.71%

Table 1. Comparison with different algorithms

3.4 Loss Function

It is a method of evaluating how well your model works on a given input data. If the output is not up to expectation then the loss is higher but if the output is pretty good the loss will be less. So in our system, we use a loss function to calculate the structural or content loss between the input image and final output. If the loss function gives a large number then there is more distortion in output image or we can say that structure of input image is distorted. Below is the Mean square error loss function.

$$L^l_{content}(p,x) = \sum_{i,j}(F^l_{ij}(x) - P^l_{ij}(p))^2$$

Where,

- Fij(x) is the pixel value of output image
- Pij(p) is the pixel value of input image.
- Lcontent(p,x) is a content loss between output and input image

When compared with PhotoWCT and DPST, our algorithm i.e.Whitening and Coloring Transform(WCT) yeilds less loss which is between 4.5% to 5.5%.

Image size	**PhotoWCT(%)**	**DPST(%)**	**WCT (%)**
256 x 256	17.3	8.4	4.3
512 x 512	17.7	8.7	4.7
896 x896	18.3	8.83	4.9
1024 x 1024	18.9	9.1	4.98

Table 2. Content Loss

V CONCLUSION

In photorealistic style transfer method, WCT2 is based on the conjectural analysis, we specifically designed our model to satisfy the reconstruction condition. The exact recovery of the wavelet transforms allows our model to preserve structural information while providing stable stylization without any constraints. By employing progressive stylization, we achieved better results with less noise amplification. Compared to the other state-of-the-arts, our analysis and experimental results showed that WCT2 is scalable, lighter, faster and achieves better photorealism qualitatively. Future study will include removing the necessity of semantic labels, which should be accurate for a flawless result.

REFERENCES

[1] Jaejun Yoo, Youngjung Uh, Sanghyuk Chun, Byeongkyu Kang, Jung-Woo Ha*" Photorealistic Style Transfer via Wavelet Transforms"*, IEEE-2019

[2] Yijun Li, Ming-Yu Liu2, Xueting Li, Ming-Hsuan Yang1, Jan Kautz *"A Closed-form Solution to Photorealistic Image Stylization"*, IEEE-2019

[3] Kamrul Hasan TalukderI and Koichi Harada *"Haar Wavelet Based Approach for Image Compression and Quality Assessment of Compressed Image"*, Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213, U.S.A, IJRTC-2018

[4] Yijun Li, Chen Fang, jimel Yang, Zhaowen Wang, Xin Lu, MingHsuan Yang *"Universal Style Transfer via Feature Transforms"*, 2017 International Conference on Computational Intelligence and Communication Networks

[5] Piotr Porwik, Agnieszka Lisowska *"The Haar-wavelet transform in digital image processing"*, 2017 ICLR

[6] Karen Simonyan, Andrew Zisserma *"Very Deep Convolutional Network for Large scale Image recognition"*,2015

[7] K. V. Arya, Abhinav Adarsh *"Image Style Transfer Using Convolutional Neural Networks"*, 2018 International Conference on Computational Intelligence and Communication Networks

[8] L. Yuille.M. Cimpoi, S. Maji, A. Vedaldi *"Color transfer between images. IEEE Computer Graphics and Applications"*, 2017 ACM Transactions on Graphics

[9] Aniruddha Dey, *"Deep Photo Style Transfer "*, 3rd Int'l Conf. on Recent Advances in Information Technology IRAIT-2018

[10] A. C. Kokaram, and R. Dahyot, *"Automated Deep Photo Style Transfer "*, ACM Transactions on Graphics-2017

[11] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, *"Color transfer between images. IEEE Computer Graphics and Applications "*,IEEE- 2017

[12] L.A.Gatys, A.S.Ecker, and M.Bethge, *"Texture Synthesis using CNN"*,ACM, 2017

[13] Jaejun Yoo, Abdul Wahab, and Jong Chul Ye, *"A mathematical frame- work for deep learning in elastic source imaging"*,SIAM Journal on Applied Mathematics-2017

[14] Jong Chul Ye, Yoseob Han, and Eunju Cha, *"Deep convolutional framelets: A general Deep learning framework for inverse problems"*,SIAM Journal on Imaging Science, 2018.