Tampere University

COMP.SGN.300. Advanced Image Processing

Free-Form Image Inpainting with Gated Convolution

Project Presentation

Long Nguyen

Contents

Motivation

- Advancements in computer technology have led to the development of complex models in various fields
- Following these advancements, a significant challenge continues: how to approach with inpainting images.
- This task is crucial as it can enhance the quality of images, aid in data recovery, and reduce the need for manual editing and time consumption.

Objective

- The first objective is to devise a model that can learn the relationship or mapping between original images and the inpainting images without the need for paired data.
- The second objective is to test the effectiveness and accuracy of this approach through extensive evaluation using a dataset that includes both types of images, with 2 different masking options: Random/Box masking and Line/Circle masking.

Figure: Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, & Thomas Huang. (2019). Free-Form Image Inpainting with Gated Convolution

Introduction:

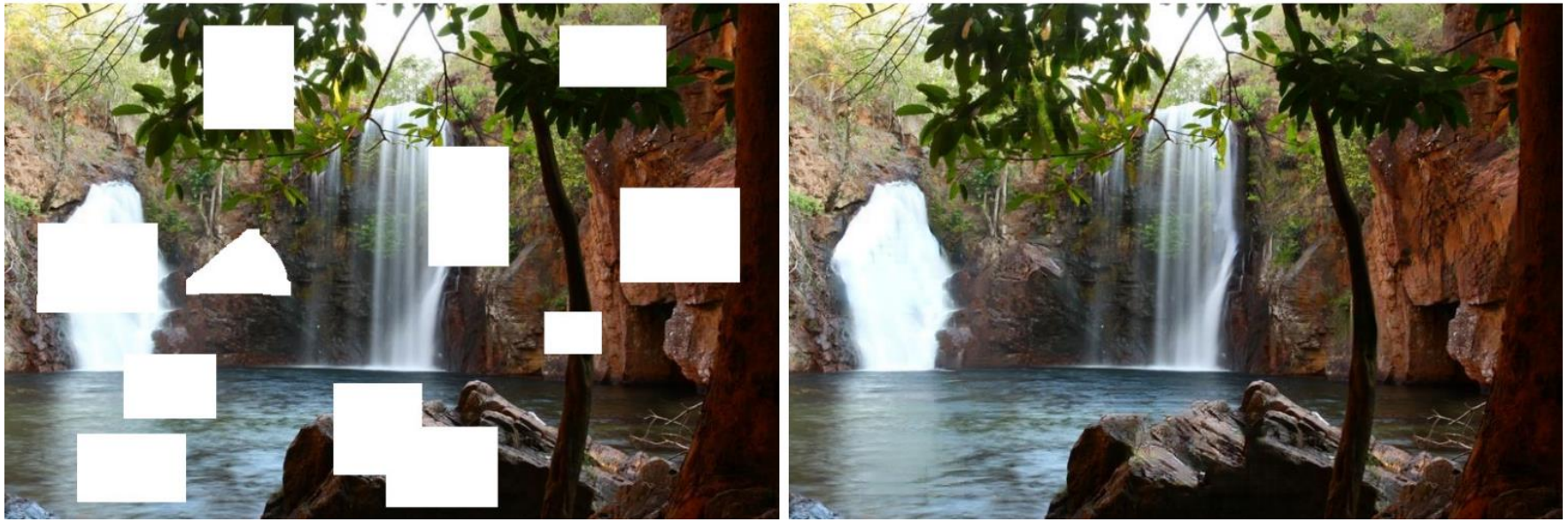
Inpainting is a technique in image processing.

It’s used to fill in missing or corrupted parts of images.

The goal is to make the filled parts seamlessly blend with the rest of the image.

It’s like digital restoration, fixing damages in old photos or removing unwanted objects.

Inpainting can be used in various fields like art conservation, photography, and computer vision.



Model Architecture

[1] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, & Thomas Huang. (2019). Free-Form Image Inpainting with Gated Convolution

2D Convolutional Layer (Conv2D) vs Gated Convolutional Layer (Genv2D)

- Conv2D

$$y[m,n] = x[m,n] * h[m,n] = \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} x[i,j] \cdot h[m-i, n-j]$$

- Genv2D[1]

$$O_{y,x} = \sum_{i=-k'_h}^{k'_h} \sum_{j=-k'_w}^{k'_w} W_{k'_h+i,k'_w+j} \cdot I_{y+i,x+j},$$

$$O_{y,x} = \begin{cases} \sum\sum W \cdot (I \odot \frac{M}{sum(M)}), & \text{if sum(M)} > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$Gating_{y,x} = \sum\sum W_g \cdot I$$

$$Feature_{y,x} = \sum\sum W_f \cdot I$$

$$O_{y,x} = \phi(Feature_{y,x}) \odot \sigma(Gating_{y,x})$$

Figure: Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, & Thomas Huang. (2019). Free-Form Image Inpainting with Gated Convolution

Gated Convolutional Layer

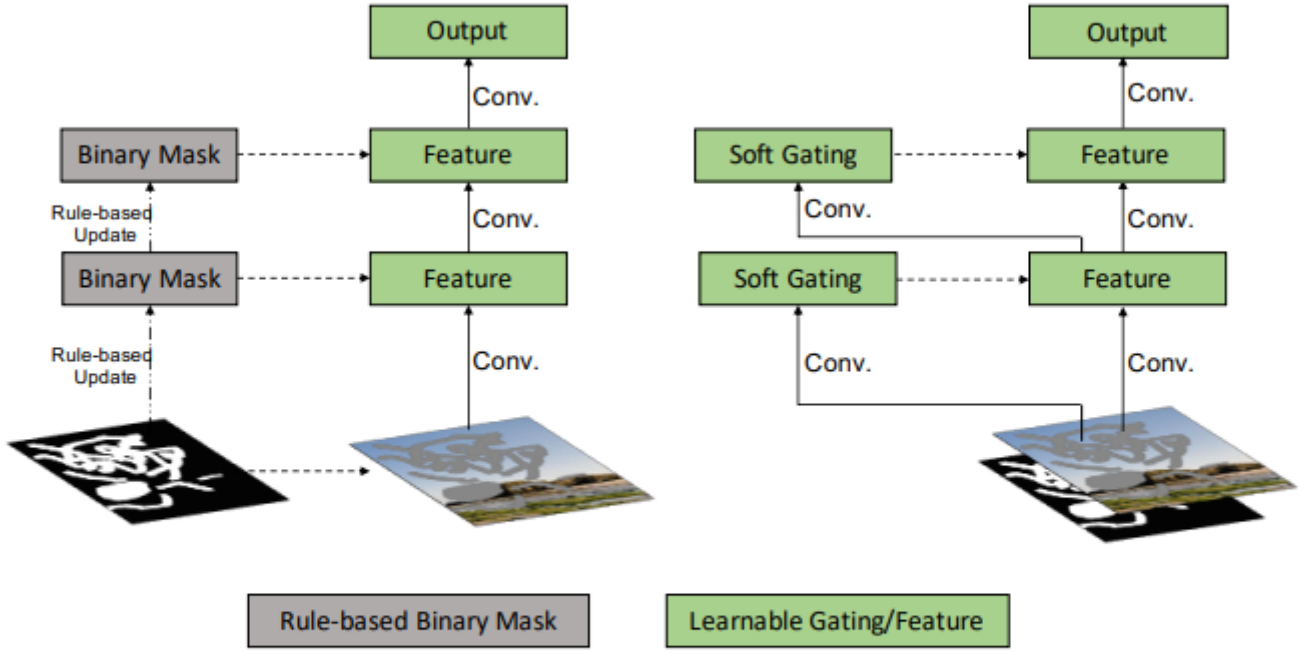
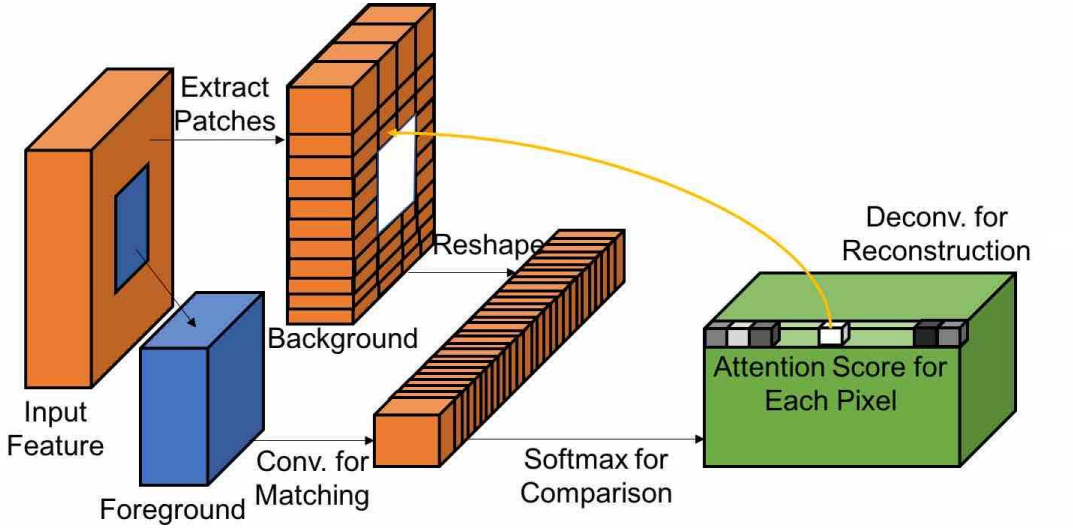
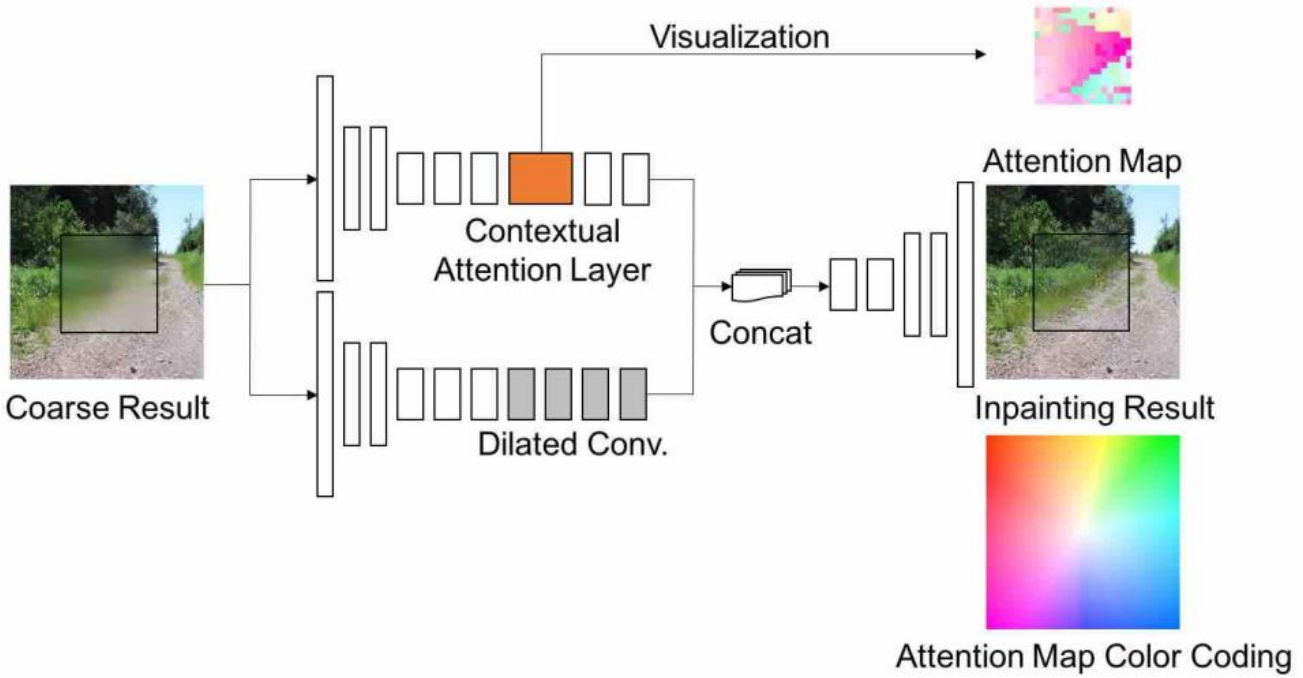


Figure 2: Illustration of partial convolution (left) and gated convolution (right).

[2] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, & Thomas S. Huang. (2018). Generative Image Inpainting with Contextual Attention.

Contextual Attention Layer[2]





[3] Fisher Yu, & Vladlen Koltun. (2016). Multi-Scale Context Aggregation by Dilated Convolutions.

Dilated Gated Convolutional [3]

$$(F *_l k)(\mathbf{p}) = \sum_{\mathbf{s}+l\mathbf{t}=\mathbf{p}} F(\mathbf{s})\, k(\mathbf{t}).$$

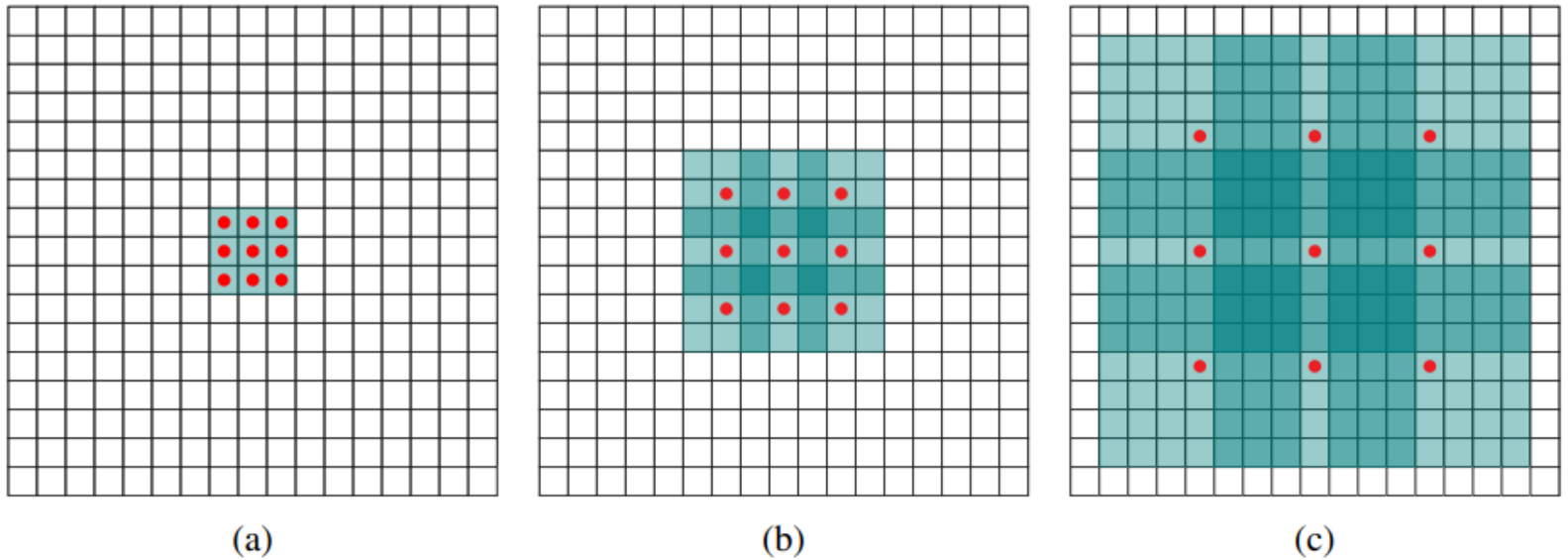
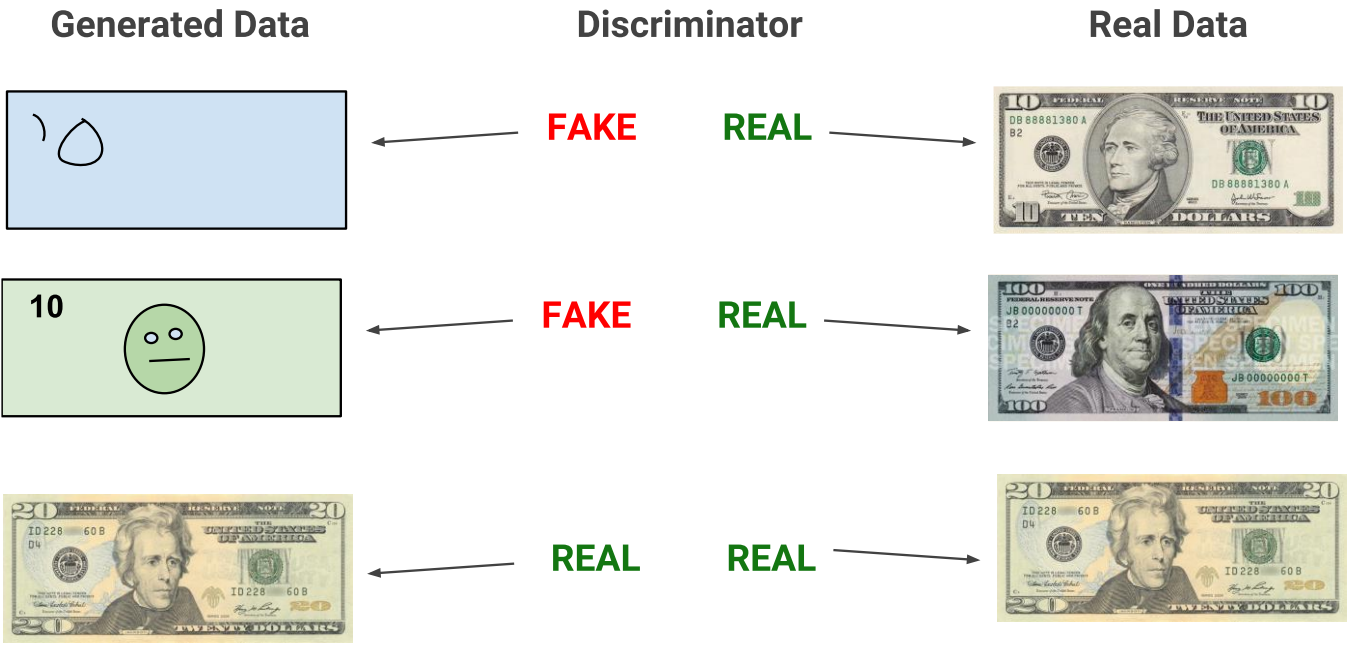


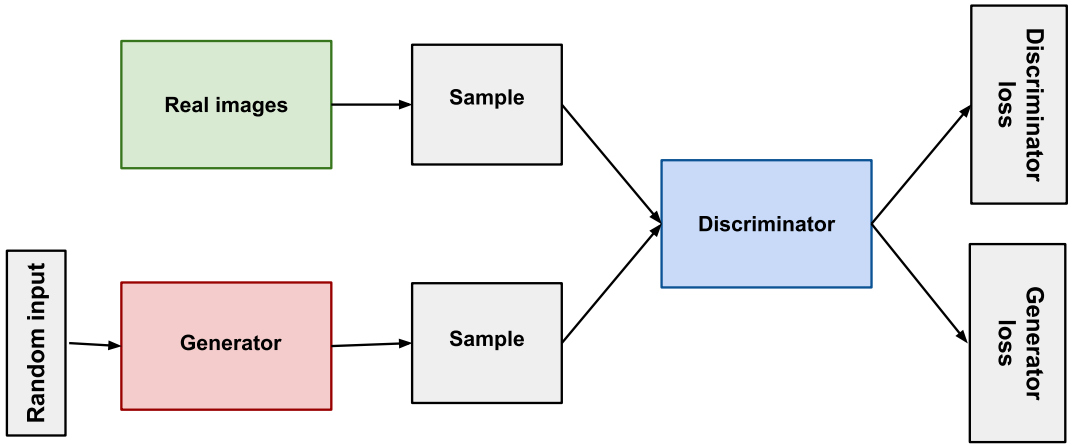
Figure 1: Systematic dilation supports exponential expansion of the receptive field without loss of resolution or coverage. (a) F_1 is produced from F_0 by a 1-dilated convolution; each element in F_1 has a receptive field of 3×3. (b) F_2 is produced from F_1 by a 2-dilated convolution; each element in F_2 has a receptive field of 7×7. (c) F_3 is produced from F_2 by a 4-dilated convolution; each element in F_3 has a receptive field of 15×15. The number of parameters associated with each layer is identical. The receptive field grows exponentially while the number of parameters grows linearly.

Tampere University

[4] Overview of GAN structure. (n.d.). Google for Developers. https://developers.google.com/machine-learning/gan/gan_structure

GAN Architecture[4]





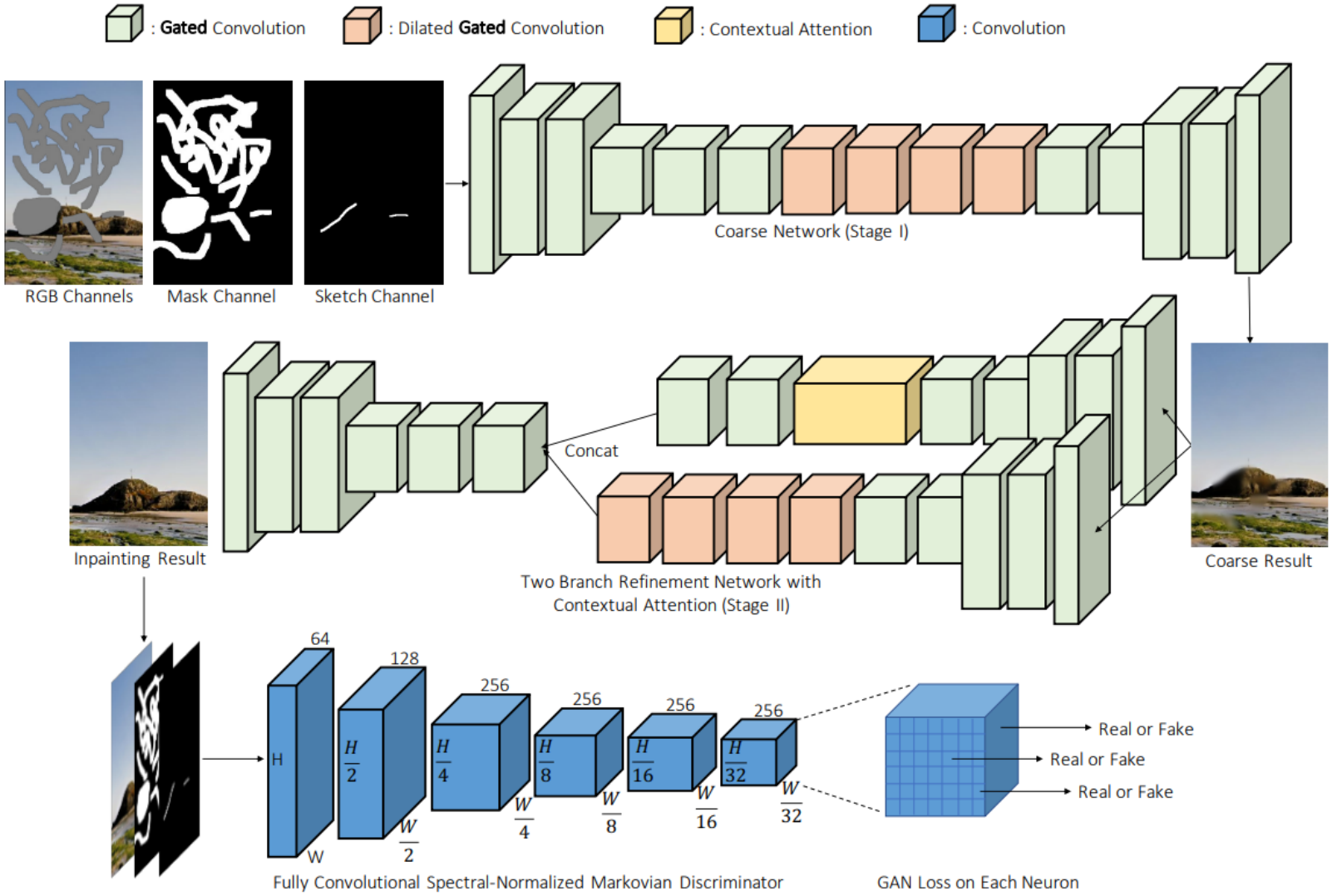
[1] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, & Thomas Huang. (2019). Free-Form Image Inpainting with Gated Convolution

Losses Function[1]

- Generator: $\mathcal{L}_G = -\mathbb{E}_{z\sim\mathbb{P}_z(z)}[D^{sn}(G(z))]$
- Discriminator: $\mathcal{L}_{D^{sn}} = \mathbb{E}_{x\sim\mathbb{P}_{data}(x)}[ReLU(\mathbb{1} - D^{sn}(x))] + \mathbb{E}_{z\sim\mathbb{P}_z(z)}[ReLU(\mathbb{1} + D^{sn}(G(z)))]$
- Gradient penalty for Discriminator: WGAN-GP

[1] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, & Thomas Huang. (2019). Free-Form Image Inpainting with Gated Convolution

Framework with gated convolution and SN-PatchGAN [1]



Evaluation metrics

[5] Wang, Z., Simoncelli, E. and Bovik, A. Multiscale structural similarity for image quality assessment. Vol. 2. Dec. 2003, 1398–1402 Vol.2. ISBN: 0-7803-8104-1. DOI: 10.1109/ACSSC.2003.1292216.

Structural Similarity Index (SSIM)[5]

- The SSIM is a method for comparing similarities between two images.
- It is a full reference quality metric, where the result of image quality requires an original image as a reference.
- The SSIM index is calculated using a specific formula that takes into account the average, variance, and covariance of the pixel intensities in the images, as well as two small constants.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2.12)$$

where in the image calculation, x is the matrix values of original image, μ_x is the average of x; μ_y is the average of y; σ_x^2 is the variance of x; σ_y^2 is the variance of y; σ_{xy} is the covariance of x and y; $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$ are two small constants; L is the dynamic range of image pixel-values [15] ; $k_1 = 0.01$ and $k_2 = 0.03$ in the implementation [16].

[6] Horé, A. and Ziou, D. Is there a relationship between peak-signal-to-noise ratio and structural similarity index measure?: Image Processing, IET 7 (Feb. 2013), pp. 12–24. DOI: 10.1049/iet-ipr.2012.0489

Peak Signal-to-Noise Ratio (PSNR)[6]

- The PSNR is the ratio between the maximum power of a signal and the power of corrupting noise that is represented in the signal.
- It is popular in image analysis because images can be understood as a distribution of pixel values with different intensities in the matrix area.
- PSNR is defined via the Mean Squared Error (MSE), which measures the difference between the original image and the transformed image.

$$PSNR(f,g) = 10 \cdot \log_{10}\left(\frac{255^2}{MSE(f,g)}\right)$$

$$MSE(f,g) = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[f(i,j) - g(i,j)]^2 \qquad (2.14)$$

Here, m and n are the dimensions of the images, and $f(i,j)$ and $g(i,j)$ are the pixel intensities at location (i,j) in images f and g.

Specification

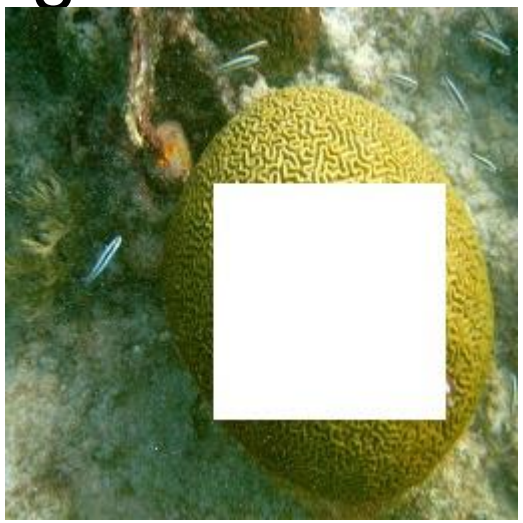
- The training procedure could not be processed, due to the limitation of CNN model where the necessity of computing resources needs to be satisfied. Therefore, I will choose the pretrained model from the author
- The testing procedure has been done in GPU Tesla V100 32GB/16GB, 8 CPU cores, 64 GB Ram per CPU
- Each Image testing is processed with 2.7 seconds, Testing with 5000 images.

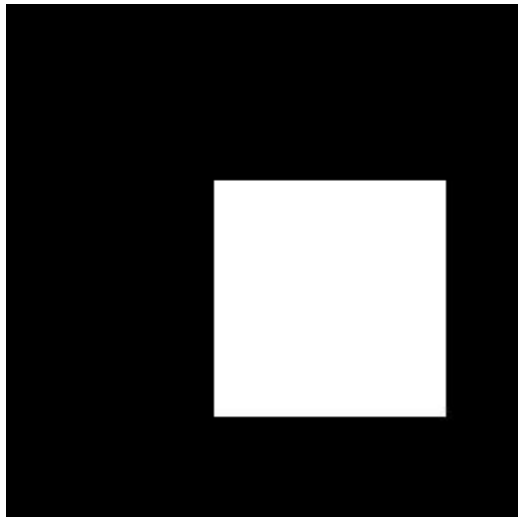
Parameters for the pretrained model

Parameters	Value
Adversarial loss	WGAN_GP, with λ = 0,0005
WGAN_GP Gradient Penalty Loss	λ = 10
Random Crop	True, 10
Padding	Same
Iteration	1000000
Batch Size	16
Image Shapes	[256, 256, 3]
Box Shaping (Randomly distributed) (Height, Width)	[128±32, 128±32]
Autoencoder Loss	True, the multiplication of L1_LOSS · $\|\|$Batch Relation$\|\|_1$

Dataset testing procedure

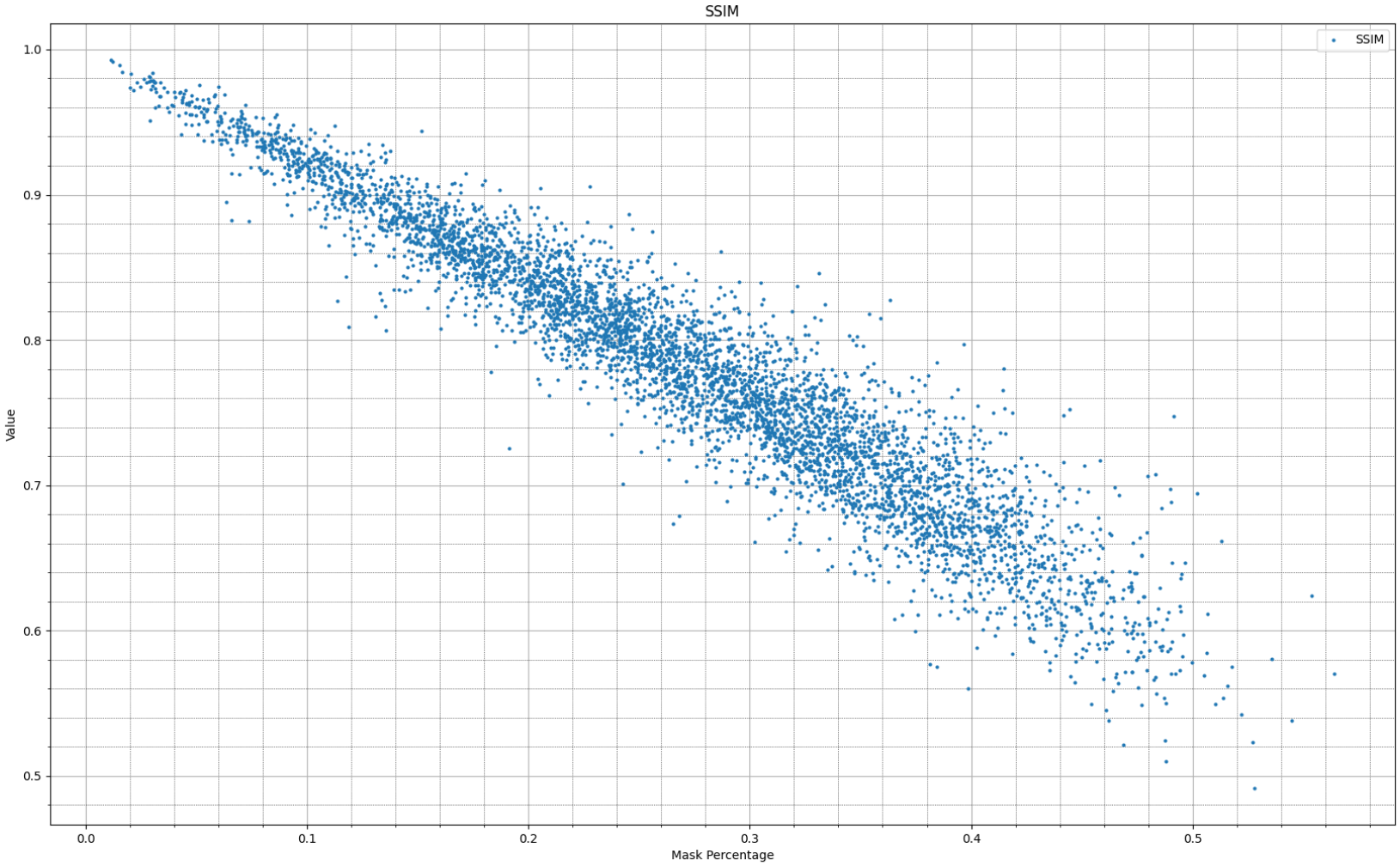
- 5000 ImageNet images have been selected as the dataset for testing the project
- The original image from dataset will go through a Mask to produce an necessary inpainted image, which could be seen in figures below





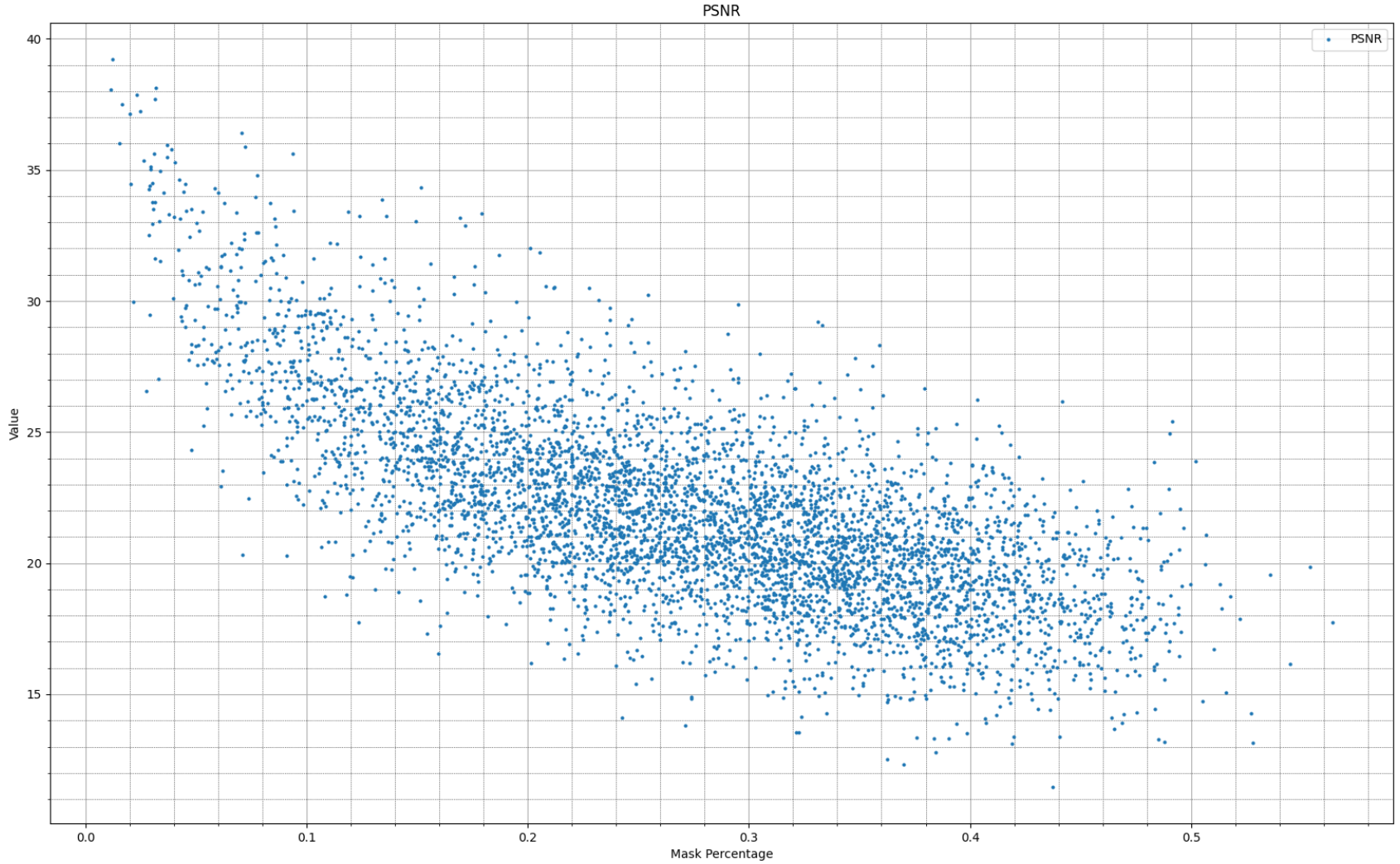
Tampere University

Output Results for Circle/Line Mask



Tampere University

Output Results for Circle/Line Mask

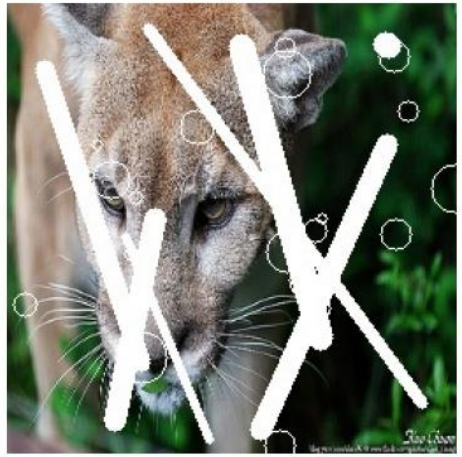


Tampere University

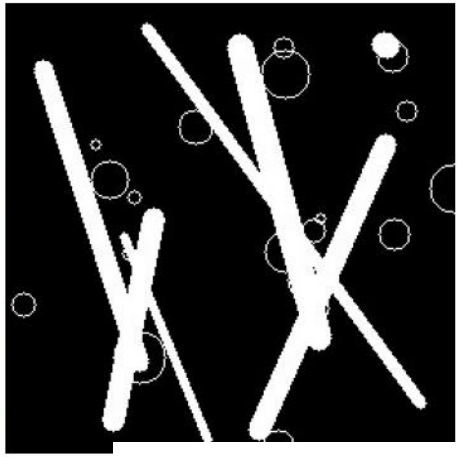
Image Output Results for Circle/Line/Curve Mask

PSNR: 23.0185, SSIM: 0.8422

Original



Mask

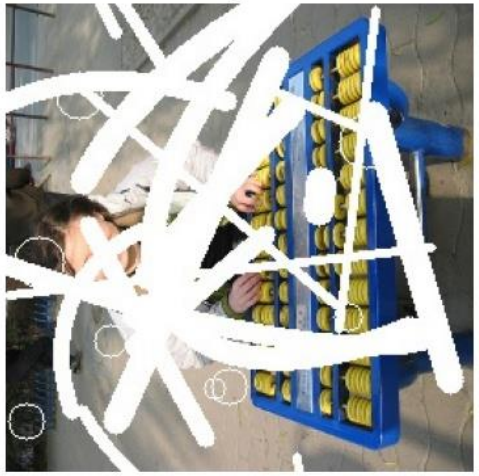


Output

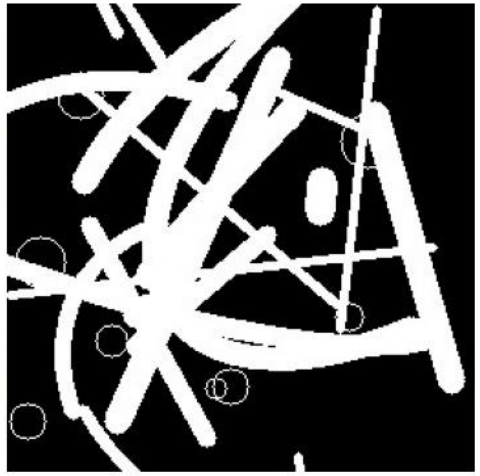
Grouth Truth

PSNR: 20.6855, SSIM: 0.7483

Original



Mask



Output

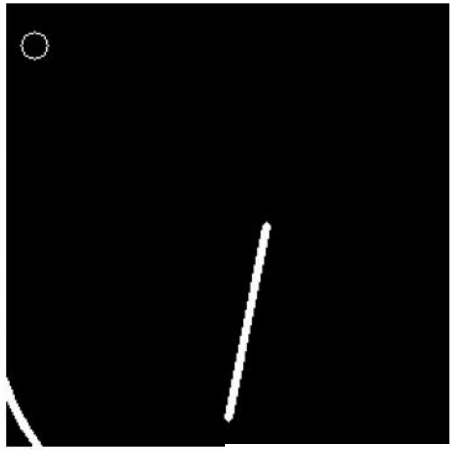
Grouth Truth

Image Output Results for Circle/Line/Curve Mask – Best PSNR and SSIM results

PSNR: 39.2191, SSIM: 0.9914

Original

Mask



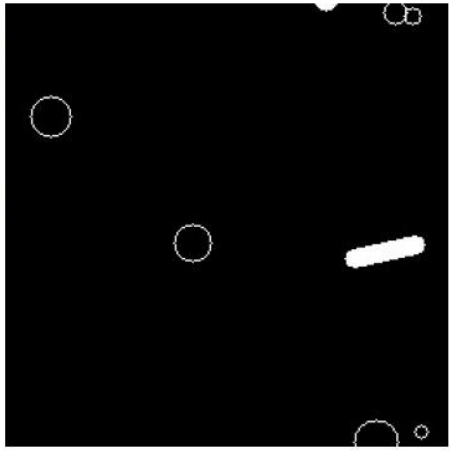
Output

Grouth Truth

PSNR: 38.0546, SSIM: 0.9928

Original

Mask

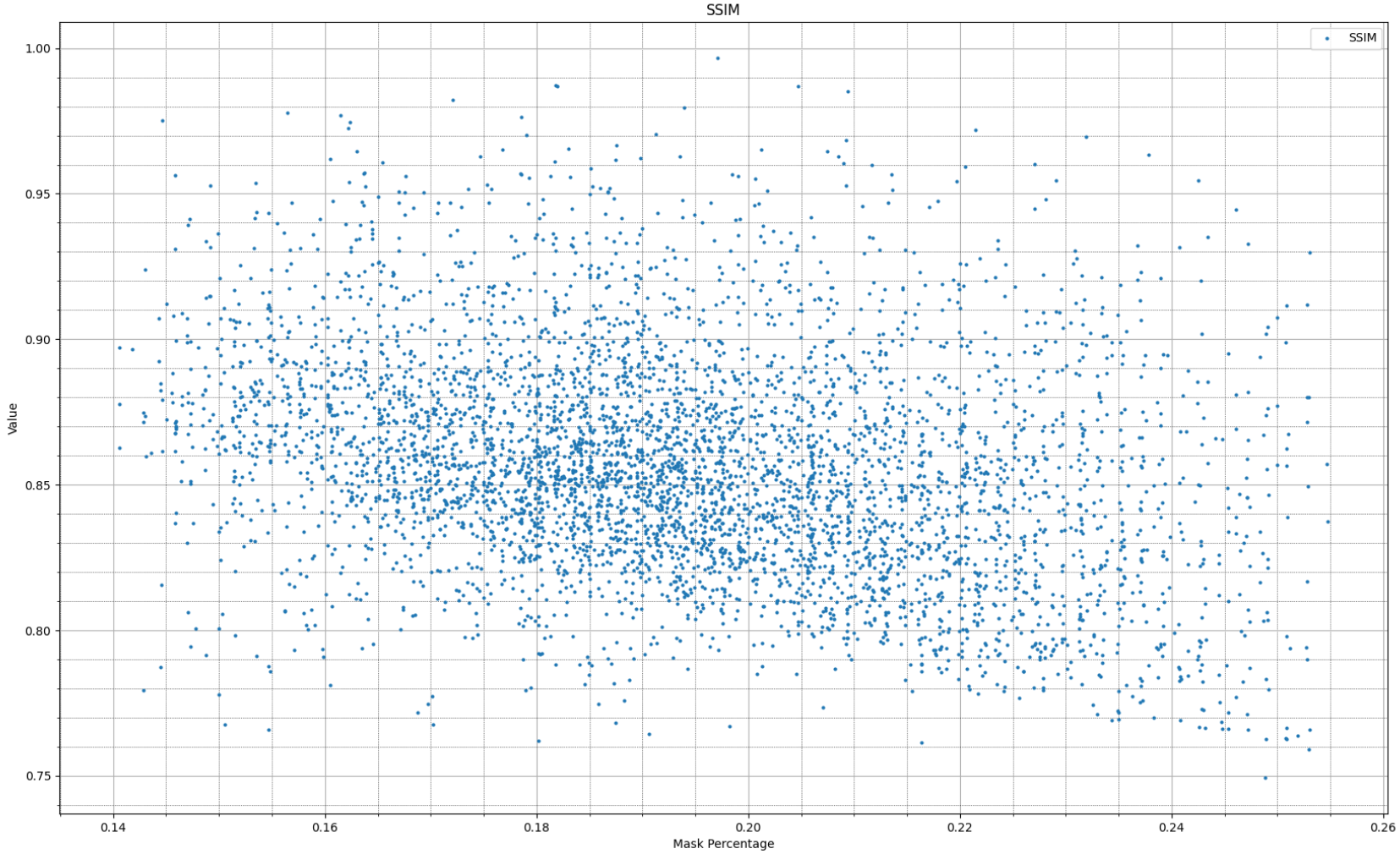


Output

Grouth Truth

Tampere University

Output Results for Random Box Mask



Tampere University

Output Results for Random Box Mask

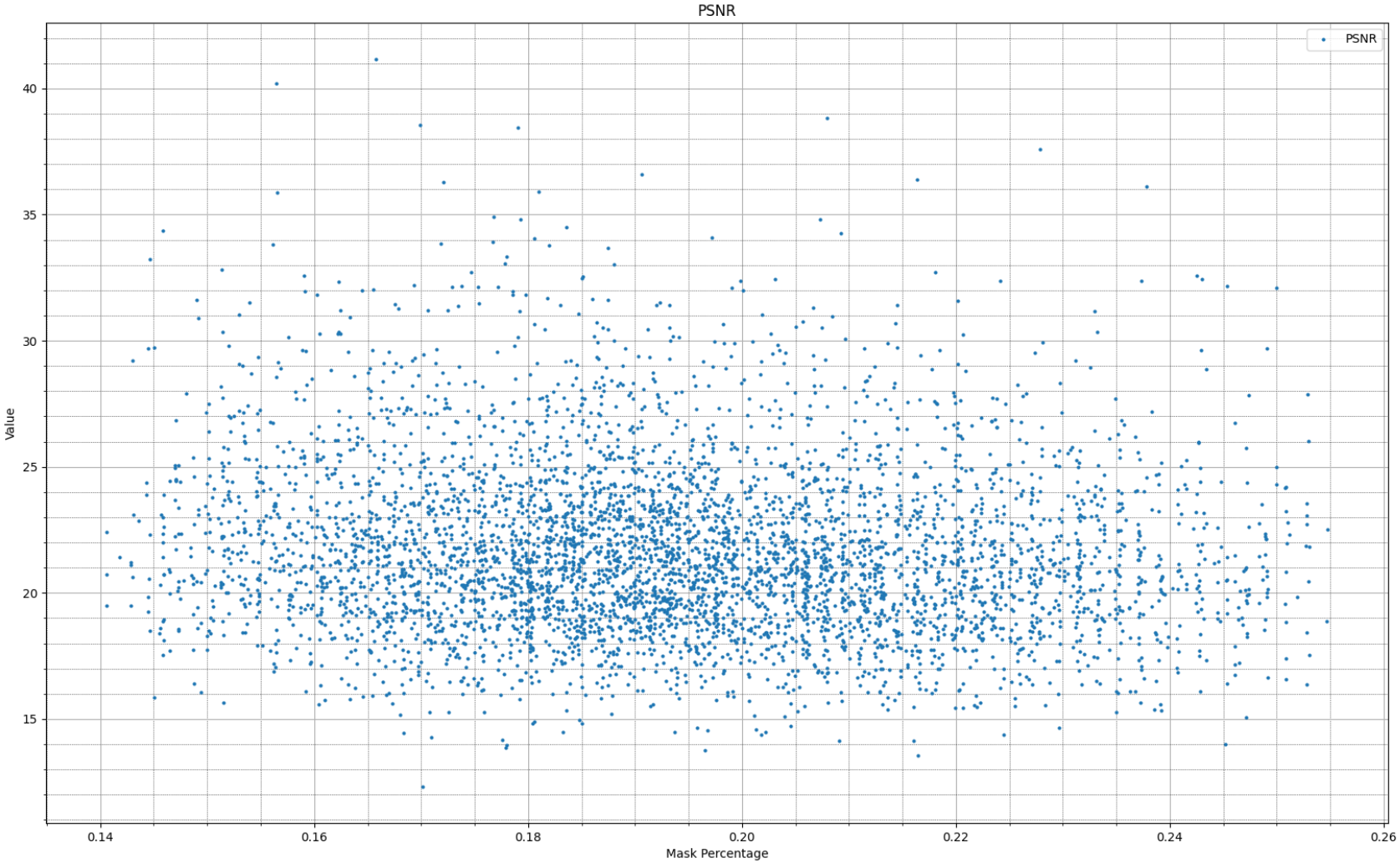
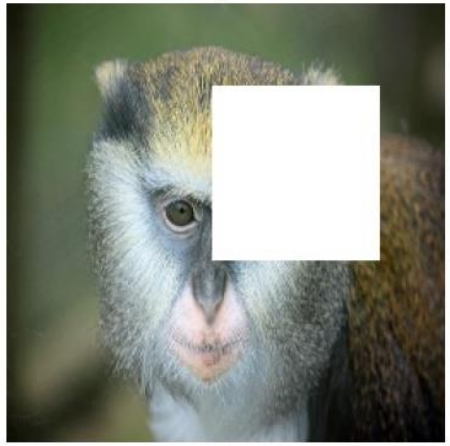


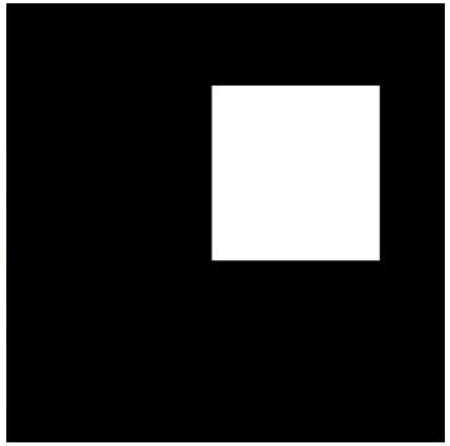
Image Output Results for Random Box Mask

PSNR: 21.2633, SSIM: 0.8774

Original



Mask

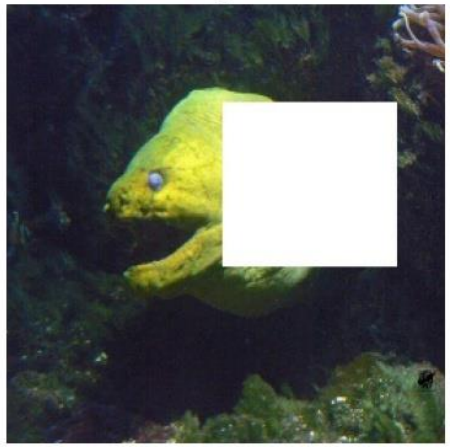


Output

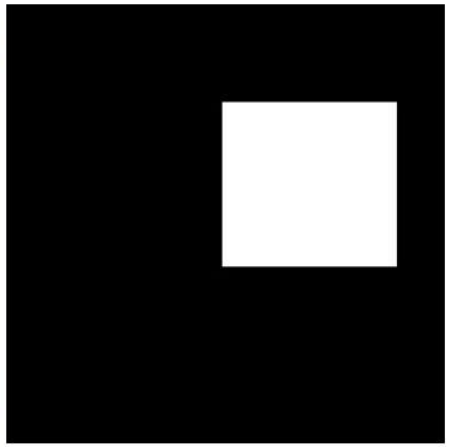
Grouth Truth

PSNR: 25.7222, SSIM: 0.9059

Original



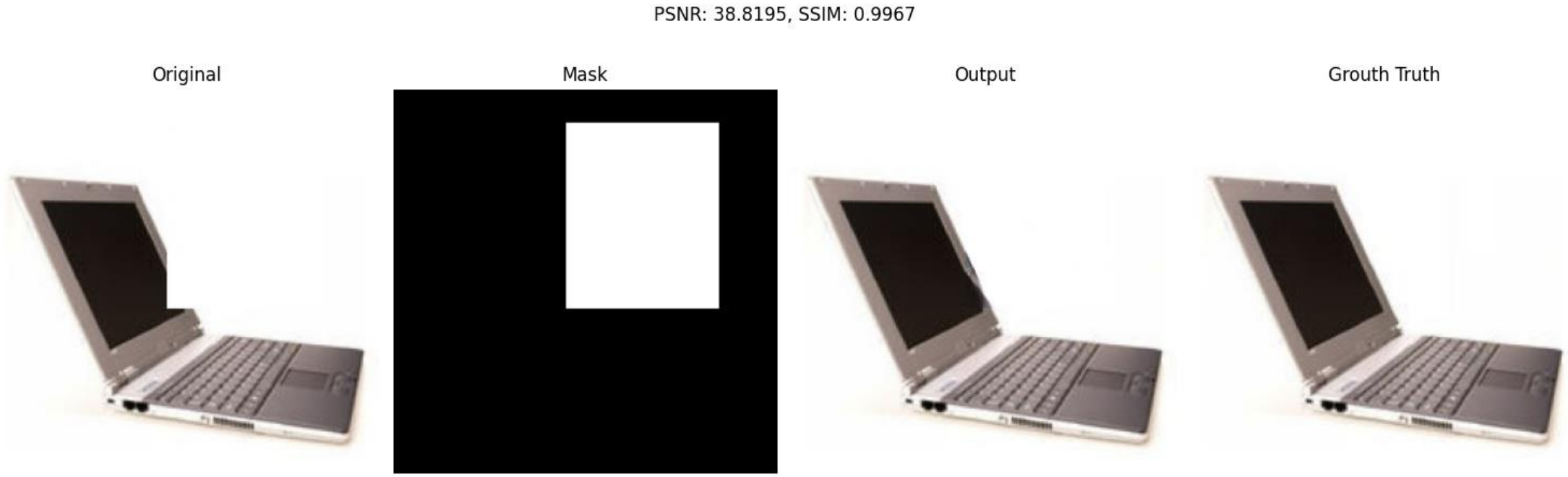
Mask



Output

Grouth Truth

Image Output Results for Random Box Mask – Best PSNR and SSIM results



Conclusions

- While framework seems to produce the results with a good general idea, even though there are still some limitations in the results.
- Evaluation metrics are suitable for the progress. And in general term, the lower the mask percentage is, the lower PSNR and SSIM values are.
- The necessity of applying with different models architecture for comparison in the future.

[7] Lin, X., Ma, L., Liu, W., Chang, SF. (2020). Context-Gated Convolution. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science(), vol 12363. Springer, Cham. https://doi.org/10.1007/978-3-030-58523-5_41

Limitation

- Complexity: Gated Convolutional Networks combine convolutional networks with a gating mechanism, which increases the complexity of the model.
- Lack of Global Context: As a basic building block of Convolutional Neural Networks (CNNs), the convolutional layer is designed to extract local patterns and lacks the ability to model global context. [7]
- Batch size limitation: due to the model is complicated, batch size needs to be 16 for satisfying the time requirement. However, 1 is always the best batch size

[8] Serkan Kiranyaz, Junaid Malik, Habib Ben Abdallah, Turker Ince, Alexandros Iosifidis, & Moncef Gabbouj. (2020). Self-Organized Operational Neural Networks with Generative Neurons.
[9] Leon A. Gatys, Alexander S. Ecker, & Matthias Bethge. (2015). A Neural Algorithm of Artistic Style.

Further Improvement

- New models with new hyperparameter, new evaluation metrics
- Applying different model architecture, which is producing the better result.
- An example of that is worth trying is applying Self-Organized Operational Layer[8]
- Applying neural style transfer for model architecture[9]

References

[1] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, & Thomas Huang. (2019). Free-Form Image Inpainting with Gated Convolution

[2] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, & Thomas S. Huang. (2018). Generative Image Inpainting with Contextual Attention.

[3] Fisher Yu, & Vladlen Koltun. (2016). Multi-Scale Context Aggregation by Dilated Convolutions.

[4] Overview of GAN structure. (n.d.). Google for Developers. https://developers.google.com/machine-learning/gan/gan_structure

[5] Wang, Z., Simoncelli, E. and Bovik, A. Multiscale structural similarity for image quality assessment. Vol. 2. Dec. 2003, 1398–1402 Vol.2. ISBN: 0-7803-8104-1. DOI: 10.1109/ACSSC.2003.1292216.

[6] Horé, A. and Ziou, D. Is there a relationship between peak-signal-to-noise ratio and structural similarity index measure?: Image Processing, IET 7 (Feb. 2013), pp. 12–24. DOI: 10.1049/iet-ipr.2012.0489

[7] Lin, X., Ma, L., Liu, W., Chang, SF. (2020). Context-Gated Convolution. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science(), vol 12363. Springer, Cham. https://doi.org/10.1007/978-3-030-58523-5_41

[8] Serkan Kiranyaz, Junaid Malik, Habib Ben Abdallah, Turker Ince, Alexandros Iosifidis, & Moncef Gabbouj. (2020). Self-Organized Operational Neural Networks with Generative Neurons.

[9] Leon A. Gatys, Alexander S. Ecker, & Matthias Bethge. (2015). A Neural Algorithm of Artistic Style

Thank you!
Any questions???