TEXTURE SIMILARITY METRICS COMPARISON

Xinyu Guo⋆ *Jinghan Zhou*⋆ *Xiaoyi Zheng*⋆ *Shiyun Qin*⋆

⋆ Electrical and Computer Engineering
University of Waterloo

ABSTRACT

We carried out a novel evaluation using Maximum differentiation (MAD) method [1] between texture similarity metrics on visual textures. In total, three models are chosen to be compared, containing metrics from image quality assessment domain and model from texture synthesis domain. We use MAD method to make comparison between all the chosen models using a constrained optimization process. A subjective test is conducted to evaluate the results from the MAD competition, the aggressive matrix, resistance matrix and global scores for all metric are computed based on the subjective test.

***Index Terms*—** metric comparison, texture similarity metrics, MAD competition

1. INTRODUCTION

We aim to accomplish a novel evaluation based research project by using MAD method to make comparison between texture similarity metrics on visual textures. Here, we mainly pick up models which are good at synthesising textures and use them as texture similarity metrics.

Texture images are important subclass of all possible images which can be usually seen. This kind of images usually consist of repeated elements which have randomized location, size, color, orientation, etc. As a result, visual textures are often characterized using statistical models. The pioneering work for characterizing visual textures is done by Julesz who describes textures using stationary random fields [2]. Based on this, many parameter-based statistical models have appeared in the field of texture synthesis. The parameters used by these methods aim to match the what human visual system (HVS) can do. In our project, we consider three models, two of them coming from the image quality assessment field, and the other one is from the texture synthesis field. We conduct MAD experiments on these models, aiming to using one model to falsify another one. Many texture image pairs are synthesised from this step which later are evaluated in the subjective test period. The aggressive matrix and resistance matrix are computed based on the results of the subjective test. Meanwhile, the global ranking score based on aggressiveness and resistance of each model are computed.

2. RELATED WORKS

Many kinds of methods are used to synthesis textures, here, we divided them into two groups. One group contains linear filter methods, and the other contains deep learning based methods.

linear filter methods were proposed motivated by human visual perception, such as Gabor filters [3] and orthonormal wavelet representation [4]. The Gabor filters are not invertible, and the orthonormal wavelet representation is not translation-invariance. Pyramid-based method [5] is a kind of method that is good at dealing with these problems. Simoncelli proposed a model consisting of steerable pyramid and complex "analytic" filters which is one of the state of the art texture synthesis models till now [6].

On the other hand, deep learning methods are becoming more and more popular recent years, there are many researchers combining deep learning models into the texture synthesis methods. One representative texture synthesis model has been proposed[7] .This method combines a deep learning model with the Gram Matrix which is calculating correlation between feature maps. After that, several fast implementations of this method have been proposed [8][9].In these works, a complex loss function proposed by in [7] has been used to supervise a feed forward convolutional neural network; then, the trained network can be used to synthesis textures. Later, improvement on the gram matrix has been made. One shifts feature maps before correlation calculation, by doing so, global arrangement of objects are considered[10].The other one adds spectrum constraint in addition to the gram matrix which takes large scale structure in to consideration [11].There exists another branch of texture synthesis models which mainly based on Deep Convolutional Generative Adversarial Networks (DCGANs)[12].These methods try to let generators learn to map noise distribution to texture distribution.Some representatives are spatial generative adversarial networks(SGANs)[13] which extend DCGANs by allowing networks synthesis arbitary size of textures, and Markovian Generative Adversarial Networks (MGANs)[14].

Fig. 1. reference textures

3. METHODS

Three models are chosen as texture similarity metrics, and we use Maximum Differentiation Competition (MAD) to compare between them.

3.1. Texture Similarity Metrics

We have chosen three metrics for comparison. The first metric we have chosen is the is Mean Square Error(MSE) which is traditional used for image quality assessment(IQA).One the one hand, MSE has good mathematics properties; for example, we can express the gradient of MSE in an simple analytical form. On the other hand, this metric has been used frequently in the field of image quality assessment.

The second one is structural similarity(SSIM) which is one of the state of the art IQA models, SSIM is proposed for solving the problem that the MSE is not good at match what human visual system(HVS) can get. SSIM is computed by combining the luminance similarity, contrast similarity, and structure similarity together. These components are decided under the hypothesis that our HVS are using these information explaining what we see.

The last model we have used is a deep learning model which composed of a deep learning model and the gram matrix.The deep learning model used here is VGGnet[15] which is 19 layers convolutional neural network which has good performance on ImageNet Large Scale Visual Recognition Challenge (ILSVRC). The gram matrix here is proposed to make use of the structure information in the texture by doing correlation computation. After this computation, the exact spatial location information is cut.

Suppose we have one reference texture image denoted as X, and one distorted texture image denoted as Y, MSE model compute similarity between them by calculating the pixel-wise error which can also be written as

$$MSE(X,Y) = \frac{1}{N}(X-Y)^T(X-Y) \tag{1}$$

where N is the number of pixels in the images

SSIM is a metric design to capture the luminance, contrast and structure of reference images and the distorted images, and the compare between them to yield a overall image quality score. In our setting, the comparison is then conducted between X and Y. The luminance, contrast and structure comparison between two patches(**x** and **y**) extracted from X and Y can be written as

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{2}$$

Where

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i, \quad \sigma_x = (\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2)^{\frac{1}{2}},$$

$$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y).$$

The overall SSIM index is computed by averaging the SSIM index for all patches in two images, which can be written as

$$MSSIM(X,Y) = \frac{1}{M}\sum_{j=1}^{M} SSIM(x_j, y_j) \tag{3}$$

The last model we adopted for comparison is a deep learning based texture synthesis model composed of a deep learning model and the Gram Matrix[7]. The VGGnet[15] is used to extract features from two input texture images detoted as X and Y. Then the Gram Matrix is used the capture the local structure of the texture images.This is actually a correlation computation which can be written as

$$G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l \tag{4}$$

Where F^l denotes feature maps from the layer l of a VGGnet, and i, j are feature map indexes.

3.2. MAD Competition

The MAD proposed by Z. Wang et al.[1] can be used to make comparisons between the texture similarity metrics mentioned above. MAD method can accelerate the process of metric comparison by searching for the stimuli which have

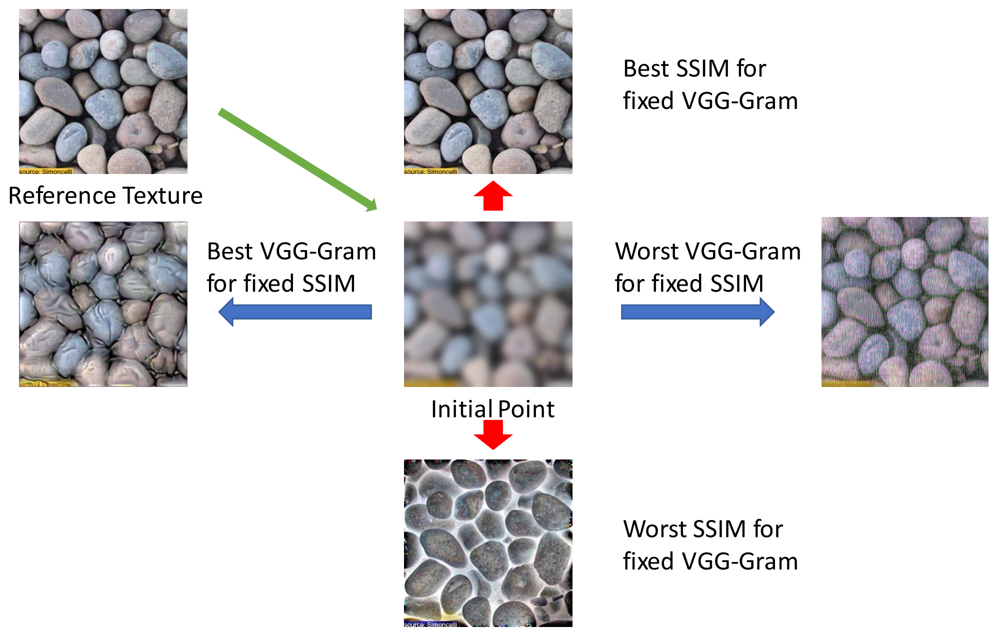

Fig. 2. MAD experiment result

the greatest possibility to discriminate between metrics. This method works as following. At the beginning, an initial distortion image is created by adding some kinds of distortions to the reference image. Then, two metrics are chosen from the three metrics mentioned above; for example, MSE and SSIM. We aim to find the worst and the best image (a image pair) by maximizing and minimizing MSE while keeping the output of SSIM fixed, subjects are then asked to rate this image pair. The same process is also applied when keeping the output of MSE fixed. This whole process is shown in Fig.3

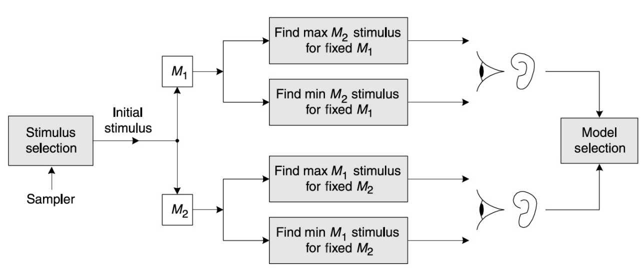

Fig. 3. MAD competition method for metric selection. [1]

4. EXPERIMENTS

We carry out our experiment on Simoncelli's example color textures. Firstly we use MAD to compare between two of the three metrics; as a result, three separate smaller experiments are done. After that, a small subjective experiment is carried out. Finally, we use a scoring strategy in [16] to calculate scores for three texture similarity metrics

4.1. Database and MAD experiment

The reference images we have used to carry out the experiment come from Simoncelli's color example textures[6]. There are totally 34 textures in this database which has covered a wide range of natural textures. Here we can see 5 examples shown in Fig.1

For the MAD experiment, we have conducted comparison between 3 texture similarity metrics. We need to compare between 3 models on 34 reference images with one given initial point. Under the one MAD experiment, we need to alternatively keep the loss of a model fixed, and then find out the worst and the best texture image by maximizing and minimizing the loss of another model. Here, we call the model whose loss needed to be kept unchanged as the defender, and the other one is called attacker. One MAD competition can be described as following.

Firstly, losses and gradients with respect to two models are being calculated. After that, the gradient ascent or descent strategy is used with regard to the attacker. The gradient g used here is calculated by projecting attacker's gradient denoted as g_a to the defender's gradient denoted as g_d

$$g = g_a - \frac{g_a^T g_d}{g_d^T g_d} g_d \tag{5}$$

Secondly, to keep the loss of the defender, we use a different strategy compared with the one use in the original MAD work. We combined the line search strategy used in the original work with a optimizer. The optimizer takes the initial loss and the current defender loss as input, and the goal is to make two loss as close as possible. Then, the roles of defender and

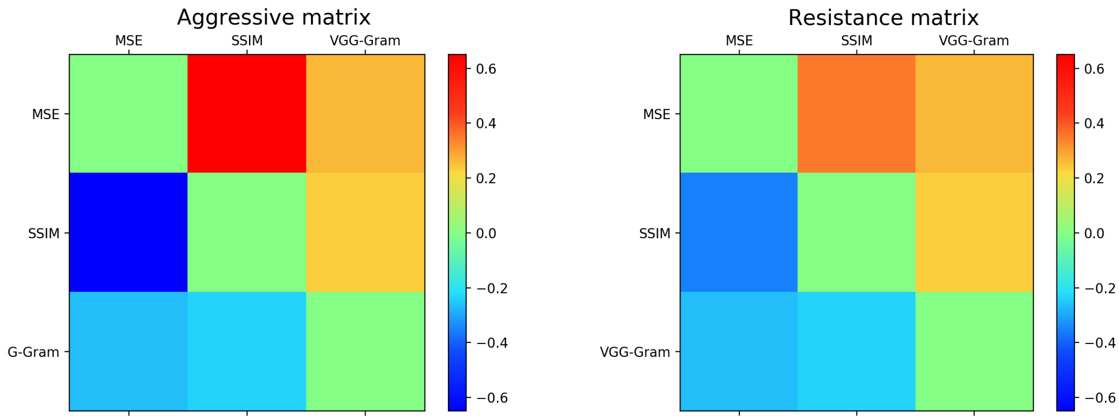

Fig. 4. Pairwise competition matrices: Each entry indicates the Aggressiveness or the Resistance of the row model against the column model. $a - a^T$ and $r - r^T$ are drawn here for better visibility

attacker are changed and a similar process is carried out. We can get two pairs of texture images from one single MAD competition.

A example result from one single experiment is shown in Fig.2

4.2. Subjective Experiment and Scoring

In total, we got 204 pairs of texture images from the whole MAD experiments. The subjective experiment is divided into six parts in which every two model would be compared to each other. For each part of the experiment, subjects are asked to evaluate 34 pairs of texture images. Each pair of texture images are shown to subjects along with the corresponding reference texture. The reference one is placed in the middle; while the left one and right one are the ones generated by holding one model and searching for minimal and maximal in the other model. Subjects are asked to choose which one is more similar to the reference texture and the results are collected.

In each part of the experiment, the aggressiveness of the attacker and resistance of the defender are evaluated by

$$a_{ij} = \frac{n3}{n1 + n2 + n3} \quad r_{ji} = \frac{n2}{n1 + n2 + n3} \tag{6}$$

where $n1 + n2 + n3$ denotes all decisions given by subjects, while $n1$, $n3$ and $n2$ are the amount of choices made by subjects. Both a_{ij} and r_{ji} range between $[0, 1]$. The value of a_{ij} denoted the frequency a model falsify another model. This means how aggressive the model would be as an attacker while the value of r_{ji} is the frequency a model is not falsified by another model, This indicates the resistent of a defender to be defeated by attacker. The matrics A and R are the all-pairs comparisons of aggressiveness and resistance and the results are shown in Fig.4

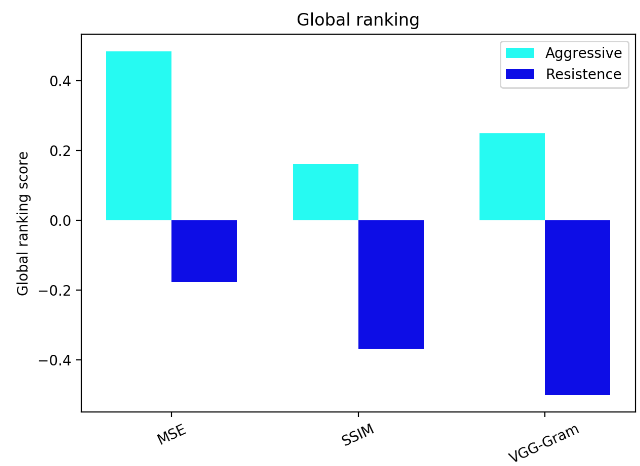

Fig. 5. Global ranking

In this experiment, each model has played a attacker role twice, and also a defender role twice. The global aggressive score and the global resistance score is thus computed by averaging the results obtained from these parts of experiment respectively. The global ranking results are shown in Fig.5

5. CONCLUSION

From the results of our experiment, we found that in our experiment conditions, MSE performs best both in aggressiveness and resistance. While SSIM and VGG-Gram does not perform well in both areas. However, obviously VGG-Gram which is specially designed for Texture Synthesis should performs better than MSE. This might because the limitation of our experiments. We just used one type of distortion and only one distortion level for initial points. In addition, a better subjective test may also help us get more accurate results.

6. REFERENCES

[1] Z. Wang and E.P. Simoncelli, "Maximum differentiation (mad) competition: A methodology for comparing computational models of perceptual quantities," *Journal of Vision*, vol. 8, no. 12, pp. 8–8, 2008.

[2] Bela Julesz, "Visual pattern discrimination," *IRE transactions on Information Theory*, vol. 8, no. 2, pp. 84–92, 1962.

[3] John G Daugman and Daniel M Kammen, "Pure orientation filtering: A scale-invariant image-processing tool for perception research and data compression," *Behavior Research Methods, Instruments, & Computers*, vol. 18, no. 6, pp. 559–564, 1986.

[4] Ingrid Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications on pure and applied mathematics*, vol. 41, no. 7, pp. 909–996, 1988.

[5] David J Heeger and James R Bergen, "Pyramid-based texture analysis/synthesis," in *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. Citeseer, 1995, pp. 229–238.

[6] Javier Portilla and Eero P Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *International journal of computer vision*, vol. 40, no. 1, pp. 49–70, 2000.

[7] Leon Gatys, Alexander S Ecker, and Matthias Bethge, "Texture synthesis using convolutional neural networks," in *Advances in neural information processing systems*, 2015, pp. 262–270.

[8] Dmitry Ulyanov, Vadim Lebedev, Andrea Vedaldi, and Victor S Lempitsky, "Texture networks: Feed-forward synthesis of textures and stylized images.," in *ICML*, 2016, vol. 1, p. 4.

[9] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711.

[10] Guillaume Berger and Roland Memisevic, "Incorporating long-range consistency in cnn-based texture generation," *arXiv preprint arXiv:1606.01286*, 2016.

[11] Gang Liu, Yann Gousseau, and Gui-Song Xia, "Texture synthesis through convolutional neural networks and spectrum constraints," in *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016, pp. 3234–3239.

[12] Alec Radford, Luke Metz, and Soumith Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[13] Nikolay Jetchev, Urs Bergmann, and Roland Vollgraf, "Texture synthesis with spatial generative adversarial networks," *arXiv preprint arXiv:1611.08207*, 2016.

[14] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in *Computer Vision – ECCV 2016*, Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, Eds., Cham, 2016, pp. 702–716, Springer International Publishing.

[15] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[16] Kede Ma, Qingbo Wu, Zhou Wang, Zhengfang Duanmu, Hongwei Yong, Hongliang Li, and Lei Zhang, "Group mad competition-a new methodology to compare objective image quality models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1664–1673.