STUDIA UNIV. BABEŞ–BOLYAI, INFORMATICA, Volume **LXVI**, Number 1, 2021
DOI: 10.24193/subbi.2021.1.05

RETINAL BLOOD VESSEL SEGMENTATION ON STYLE-AUGMENTED IMAGES

MARCELL DÁVID TÓTH AND ATTILA KISS

Abstract. The average human lifespan increased dramatically in the second half of 20th century. It was mainly due to technological improvements, which were driven by the continuous war preparations, and while humans have got another 20 years to live, unfortunately there are some sad side effects added to the elderly life. Various diseases can attack the eye, our major organ responsible for receiving information, therefore many researches were devoted to examine these diseases, their early signs, and how could they be stopped. From the start of 21th century, methods aided by computer were more and more involved in these processes, up to the current trend of using Convolutional Neural Networks (CNNs). While supervised methods, CNNs do achieve accuracy which can be compared to a skilled ophtalmologist, they require a tremendous amount of labeled data which is sparse in medical fields because the amount of time and resources needed to create them. One natural solution is to augment the data present, that is, copying the distribution while adding a small variety, like coloring an image differently. That is, what our paper aims to explore, whether a texturing algorithm, the Neural Style Transfery can be used to make a data set richer, and therefore helping a classifier CNN to achieve better results.

1. Introduction

The eye is one of the most important parts of the human body as the majority amount of information gained by perceiving the world around us. Sadly, there are numerous diseases could threaten this organ, and many of them do not result in immediate loss of eyesight, but damage the tissues and other parts of the eye over a long period of time. These diseases are more common among the elderly, and as average lifespan increased in the 20th century, they

Received by the editors: 4 December 2020.
2010 *Mathematics Subject Classification.* 68U10, 68T01, 92B20.
1998 *CR Categories and Descriptors.* I.4.6 [**I.2 Image Processing and Computer Vision**]: Segmentation – *Region, partitioning*; I.5.1 [**Pattern Recognition**]: Models – *Neural Nets*.
Key words and phrases. Medical Image Processing, Retinal Blood Vessel Segmentation, Neural Style Transfer.

became more relevant to combat. Such disease is *diabetic retinopathy* (DR), which are the leading cause of blindness in developed countries, according to the World Health Organization (WHO).

On the bright side, there are number of indicators of DR, which can suggest early treatments, that slow down or even halt vision loss. These signs can detected during *dilated fundus examination* by screening the retina with a fundus camera, gaining information about the retina, blood vessels, etc.

This is done with *fluorescein angiography*: a small amount of sodium fluorescent dye is given to the patient orally or via injection. Then the retina gets illuminated with blue light, and green light reflected back by the dye reveals the blood vessels. Although the method is known to be reliable and accurate, it can cause side effects, such as nausea or anaphylaxis.

To offer an alternative solution which is more comfortable for the patients, researches began in search of *blood vessel segmentation methods* on retinal images. The task is to create a classifier algorithm, which, upon seeing a fundus image, accurately labels each pixel as vessel or non-vessel. To be reliable, such algorithms require a huge amount of training examples with ground truth labelling, as the classifier's parameters need to be set. Unfortunately, segmented fundus images are expensive and time consuming to produce, because the process to create one involves trained ophthalmologists.

The area of data augmentation deals with the *problem of lacking data* in quantity. A natural solution is to create new images using certain transformations on the original ones: color enhancing, whitening, adding noise, etc. These transformations preserve spatial attributes, therefore the new image has the same segmentation as its origin. Also, there are some other techniques, such as cropping, rotation or flipping the images, executing the same on the segmented image.

After this introduction, our paper will go as follows: in the next two sections, we will give a short description of the already known supervised methods for segmenting retinal images, and also, we will introduce Neural Style Transfer (NST) algorithm, emphasizing it as a data augmentation technique. In section 4, we will go through our experiment of using NST to create fundus images, the results and conclusions will be presented in section 5 afterwards.

2. Related works

2.1. Segmentation methods. The area of retinal blood vessel segmentation is rich in researches, from the late 80's until today. The very first methods were *rule-based*, which means that certain areas were marked as vessel, if pixels in it had a common property, followed a shape, etc. Later on *superwised methods* took a major role, with trained classifiers achieving accuracy over 90% and more.

Such one rule based method was *matched filtering*, dating back to, where researchers discovered that the cross section of a blood vessel follows the shape of a Gauss-curve, speaking of grayscale pixel intensity value. While being easy to interpret, this method achieved good results in 1989 [2], and was improved later, see papers [1, 19]. An other rule based method features *Multi-Scale Line Detection*, which builds upon an even more simple observation: blood vessels are made of linear segments piecewise [12].

With *Neural Networks* (NN) became popular in the last decade, more and more works were devoted to explore their abilities in the topic of discussion. Marín et al. [11] used a small, 3-layer deep NN feeded with 7-dimensional feature vectors to measure the probability of a pixel being blood vessel in the image. The 7 features consist the local average, minimal, maximal and center intensities, as well as variance and Hu-moments.

We follow the work of [18], details explained along our paper. For a thorough review, look up [4].

2.2. **Neural Style Transfer.** While seemingly confusing to mention here, NST has its relevance in our topic. The idea to create artistic images with the aid of a computer is not new, but the first truly successful attempt was only in 2016 by Gatys et al. [5, 6]. Since then, more than a hundred research papers were dedicated to explore the capabilities of the method. For a thorough review until the near end of 2018, look up paper [8].

The goal of NST is to create an image, given a content image and a style image, with the restriction that the result has to have the similar semantic information (what we actually see on the image) as the content and also similar textures (colors, shapes, etc.) as the style. The original NST, that we used, is an image optimisation technique, which means that the algorithm starts from an initial white-noise or the content image, and in each iteration small adjustments are done to match content and style respectively. Without delaying further, the original NST can be described the following way, picture [1] attached:

(1) The core is made of a pretrained deep neural network. Newer frameworks exploit the capabilities of the VGG19 because its success on image recognition tasks [17], that it can already classify images into a vast number of categories.

    As an image is passing through the network, the responses of the kernels are accumulated in *feature maps* for each layer. The deeper the layer we are currently examining, its feature maps contain the more and more complex information about the image. The user must

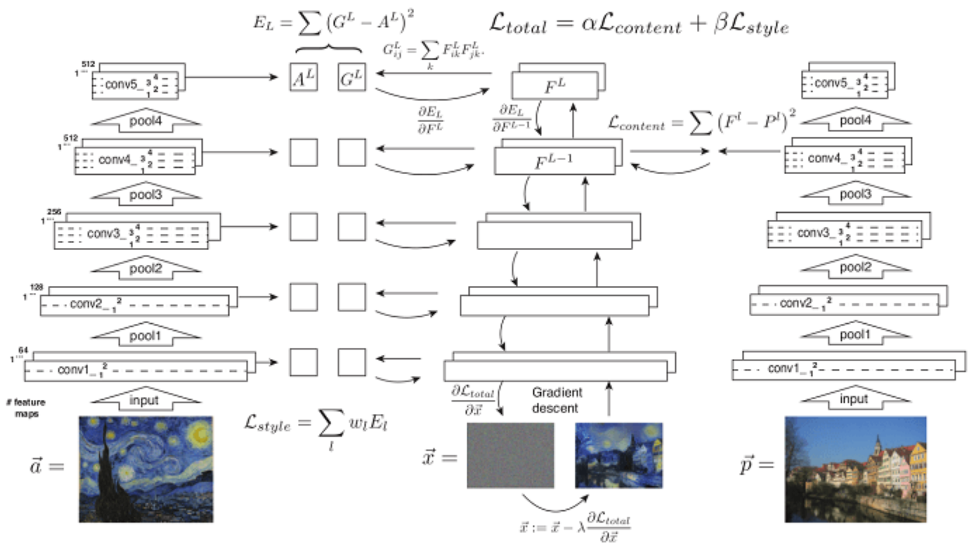

FIGURE 1. An illustration of the Neural Style Transfer algorithm framework with the VGG16 network, originally used by Gatys et al.

choose a few layers to get these inner representations as outputs, with many guidelines had already been made to choose certain groups for perceptually pleasant results: one content representation layer is chosen in deeper sections, and multiple layers are chosen across the network for representing style.

(2) Let us denote the content representation feature maps by $P^1, \ldots, P^l$, with the same indexing on the result, let it be $F^1, \ldots, F^l$. The difference between actual contents is expressed as the L_2 distance of the feature maps, the sum of the element-wise difference squared. This is simply written here as:

$$\mathcal{L}_{content} = \sum_{i=1}^{l} (F^i - P^i)^2$$

In our work, we chose only the *block5_conv2* convolutional layer for content representation in the VGG19 network.

(3) Style matching is done in a slightly different way. For both style and result image, their *Gram-matrices* are calculated from features maps, denoted by $A^1, \ldots, A^l$ and $G^1, \ldots, G^l$. These matrices represent correlation between features on an arbitrary image, therefore the task

is to pull these correlations closer feature-wise:

$$\mathcal{L}_{style} = \sum_{i=1}^{l} (G^i - A^i)^2$$

It is worth noting that since its 2016, this way of representing style came through many refinements. Consider reading paper [10] for a better understanding on style matching, and papers [15, 7] for additional techniques like color-histogram matching and total variation loss. In our work, we chose the first convolutional layer from each block: *block1_conv1*, ..., *block5_conv1*.

(4) The total loss is weighted sum of the two previously defined losses:

$$\mathcal{L}_{total} = \alpha\mathcal{L}_{content} + \beta\mathcal{L}_{style}$$

With respect to the pixel values of the result image as variables, $\mathcal{L}_{total}$ is differentiable, therefore optimal intensities can be calculated via back-propagation.

* We must also mention, that improvements were also done to speed up the work of NST, see paper [9].

Surprisingly, NST was not a subject of researches related to direct data augmentation until the end of 2019. The first experiment was to measure, if NST can improve a classifier's performance by creating stylised images, therefore the training dataset will have more variance, see paper [20] for further details. In 2020, NST was also used in medical fields for dermatological data augmentation, with the same motivations there as mentioned in the introduction, see paper [13].

3. Proposed Method

Our hypothesis is that a sufficiently used NST can be used to synthesize retina images with style to gain more variance, and therefore a classifier trained on the augmented data would be more robust to outliers, creating less false positives. To test this, we executed the following plan:

(1) In the beginning, we use the original 20 DRIVE images [3] to measure the performance of the classifying Convolutional Neural Network (CNN). The results we are mainly looking to improve is specifity and training time, see section 5 for explaining evaluation metrics.

(2) We create stylized images with the NST algorithm. For getting desirable results, one must address many properties of the algorithm: losses and respective weights, the way of representing style, etc. In the end, we will have 20 times the number of styles images.

(3) The CNN is re-trained, but now on the augmented dataset, and we compare the two classifier's performances.

4. Experiment

4.1. **Used frameworks.** The CNN we used in our experiment, called *Retina U-net*, was implemented by Orobix, resources can be found at [14]. The idea of the U-network was first presented in paper [16], with the motivation to create encode-decoder framework for medical image processing. The network in subject could be marked as a tiny U-net, as its size does not even approach the one presented in the original, having only ≈ 470.000 parameters.

The network processes retinal images in patches with size 48×48 pixel, encodes them in convolution-dropout-convolution-maxpooling manner, and decodes with upsampling in the end of the same structure, convolution-dropout-convolution-upsampling. All convolutions are 3×3, dropout layers were used with 0.2 probability. In the beginning we start with 32 kernels and double the numbers after each maxpooling, up to 128, and halving after upsampling, back to 32. We used the *Adam* optimizer to train the network, with no final activation and Binary Cross-Entropy loss functions.

The predicted images were thresholded to gain binary images, this cut-off was set in the interval $[0.15; 0.25]$ after multiple trials.

Retinal images were obtained from the DRIVE database. The database consists of 40 fundus images, 20 for training and 20 for testing purposes, each given with a manually segmented blood vessel map, as well as an image mask that separates the background. Each image has the size of 584×565 pixel.

In the beginning, we measured the stand-alone performance of the Retina U-net by training it on 10.000 patches (500 extracted randomly from each image, see pictures 2) and evaluating it later on the test patches. Test patches were obtained by first expanding the test images to the size of 624×576, and then making 13×12 regular patch cuts.

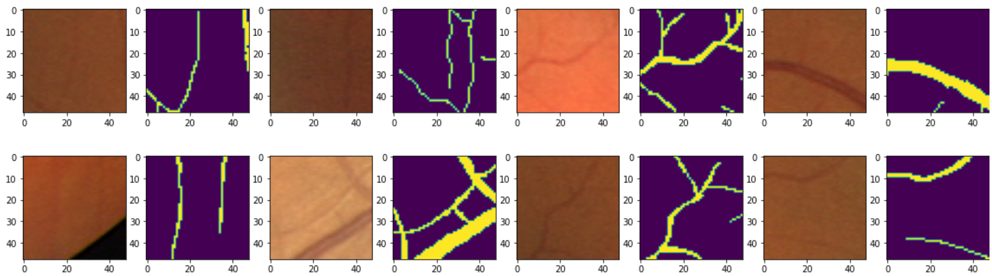

Figure 2. Patches extracted from retinal images

After initial result, we now turn to use the NST algorithm. We chose 3 style images: *Composition VII* from Wassily Kandinsky, *Starry Night* from Vincent van Gogh and *The Great Wave Off Kanagawa* from Hokusai, see pictures 3. The content and style loss weights were chosen 10^{-6} and 10^{-3} respectively and optimization process were executed also with *Adam* on the VGG19 network.

FIGURE 3. Used styles

After stylization, the same patch extraction-training-evaluation procedure was executed, see the results in the next section and examples of stylized retina in the appendix.

Upon applying the transformation to the training set, the difference shift in colors was measured in the fundamental way: calculating the mean and standard deviation for each color channel. This is due to get insights, expectations before going through the classification process again:

Looking at the values, what first thing that meets the eye is that for each transformed image, the mean values are much more regular, are closer to each

TABLE 1. Average value and standard deviation of colors among the training images, compared with the augmented ones (scaling from 0 to 255)

	R	**G**	**B**
Original	181.89 45.19	97.78 28.57	57.37 17.26
Composition VII	153.99 23.00	119.57 19.12	98.93 15.97
Great Wave	145.10 21.15	118.12 17.25	104.28 15.92
Starry Night	154.81 23.41	115.62 19.00	97.14 17.48

other in contrast to the original training set: the red values decreased by about $30/255 = 11.76\%$, while the green and blue channel values increased by roughly $20/255 = 7.84\%$ and $40/255 = 15.68\%$. While this means shift occurs, each channel's standard deviation interval shrank to $[-25, 25]$. This will be presented with images in the appendix, but this means, that the same content, the blood vessels, must be extracted from the new, augmented images, which are more homogeneous in color, therefore the new CNN has to be more sensitive to small changes in order to perform well.

5. Results and discussion

5.1. **Statistical measures used.** The final results we calculated follow the traditional evaluation metrics and statistics used in image segmentation. Upon getting the model predictions, a usual thresholding scheme is used to obtain binarized images. These images were matched with their corresponding ground truth segmentation, and all four class scores are calculated: *True positives* (TP), *True Negatives* (TN), *False Positives* (FP) and *False Negatives* (FN). After this, the following measures were used:

(1) *Accuracy* (ACC): the overall performance of classifying a seen pixel correctly. While being a widely used measure, ACC has its limitations and can be misleading in cases, for example when the label set imbalanced.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

(2) *Sensitivity* (true positive rate, SENS) and *Specificity* (true negative rate, SPEC): both measures the percentage of misclassification in vessel and non-vessel cases respectively. High SENS can be interpreted that the classifier pays attention to small details, that is, it can find tiny vessels, while high SPEC means that it is robust enough not to be distracted with noise.

$$SENS = \frac{TP}{TP + FN} \quad SPEC = \frac{TN}{TN + FP}$$

(3) *Balanced Accuracy* (BACC): makes up for the imbalanced cases, where ACC is misleading, commonly said as average true predictive power.

$$BACC = \frac{SENS + SPEC}{2}$$

(4) *Precision* (PREC): another measure for positive prediction ratio besides SENS. In this case we compare true vessels pixels to those that

are marked as vessel, and get a view about how well the algorithm separates noise from blood vessels.

$$PREC = \frac{TP}{TP + FP}$$

(5) *Matthew's Correlation Coefficient* (MCC): while no single number can capture a classifier's performance, MCC is regarded to be one of the best so far. It ranges between $[-1; 1]$ with 1 perfect predictive power, 0 meaning randomness and -1 meaning failure.

$$\begin{aligned} MCC = & \\ & = \sqrt{PREC \cdot SENS \cdot SPEC \cdot NPREC} \\ & - \sqrt{(1 - PREC) \cdot (1 - SENS) \cdot (1 - SPEC) \cdot (1 - NPREC)} \end{aligned}$$

where NPREC is the negative class precision.

5.2. **Evaluation and Discussion.** The aformentioned final scores can be seen in table 2, where we took the average in both cases.

We can see that the augmentation technique indeed helped to improve our CNN's classification power, with returning positive difference in almost each category, except in SPEC. Although these results look promising with respect to employing NST in retinal image augmentation, we want to mention that this method seems far from done. We list some number of parameters that need to be set correctly:

- The network used is an unexplored part of NST, but represention s do depend on that the underlying network previously learnt. Could it be, that a network trained on retinal images could perform NST better than the VGG-networks?
- Related to the previous, but a different network might need other layers to be chosen to represent content and style weights $\mathcal{L}_{content}, \mathcal{L}_{style}$ should be chosen accordingly as well.

Table 2. Scores on the original and the augmented dataset, and respective differences (positive means improvement on the augmented data)

	ACC	**PREC**	**SENS**	**SPEC**	**BACC**	**MCC**
Original Data	95.95	84.31	66.53	98.78	82.66	72.64
Augmented Data	96.49	84.44	74.17	98.65	86.41	77.09
Difference	0.53	0.13	7.63	-0.12	3.75	4.44

- The concept of style loss are examined thoroughly in previous works (e.g. *histogram losses*), and that is one thing could enhance a retinal image augmenting NST.
- If we follow the patch-based strategy, then there might be a number of patches, where creating synthesized images no longer affects the classifier's performance. Therefore, this image augmentation technique should be performed only when limited data is available.
- Additional metrics for measuring differences can be applied before retraining the CNN, to get more insights. One can mention the normalized cross-correlation, or measuring the euclidean distance of feature maps on a normally trained U-net.

This concludes our work. We have seen that there are ways to deploy NST in medical image processing and we are eager to see and continue with further improvements.

6. Acknowledgments

The project has been supported by the European Union, co-financed by the European Social Fund (EFOP-3.6.3-VEKOP-16-2017-00002).

References

[1] Al-Rawi, M., Qutaishat, M., and Arrar, M. An improved matched filter for blood vessel detection of digital retinal images. *Computers in biology and medicine 37*, 2 (2007), 262–267.

[2] Chaudhuri, S., Chatterjee, S., Katz, N., Nelson, M., and Goldbaum, M. Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Transactions on medical imaging 8*, 3 (1989), 263–269.

[3] Drive: Digital retinal images for vessel extraction. https://drive.grand-challenge.org/.

[4] Fraz, M. M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A. R., Owen, C. G., and Barman, S. A. Blood vessel segmentation methodologies in retinal images–a survey. *Computer methods and programs in biomedicine 108*, 1 (2012), 407–433.

[5] Gatys, L. A., Ecker, A. S., and Bethge, M. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576* (2015).

[6] Gatys, L. A., Ecker, A. S., and Bethge, M. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 2414–2423.

[7] Gatys, L. A., Ecker, A. S., Bethge, M., Hertzmann, A., and Shechtman, E. Controlling perceptual factors in neural style transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 3985–3993.

[8] Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., and Song, M. Neural style transfer: A review. *IEEE transactions on visualization and computer graphics* (2019).

[9] Johnson, J., Alahi, A., and Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision* (2016), Springer, pp. 694–711.

[10] Li, Y., Wang, N., Liu, J., and Hou, X. Demystifying neural style transfer. *arXiv preprint arXiv:1701.01036* (2017).

[11] Marín, D., Aquino, A., Gegúndez-Arias, M. E., and Bravo, J. M. A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. *IEEE Transactions on medical imaging 30*, 1 (2010), 146–158.

[12] Nguyen, U. T., Bhuiyan, A., Park, L. A., and Ramamohanarao, K. An effective retinal blood vessel segmentation method using multi-scale line detection. *Pattern recognition 46*, 3 (2013), 703–715.

[13] Nyíri, T., and Kiss, A. Style transfer for dermatological data augmentation. In *Proceedings of SAI Intelligent Systems Conference* (2019), Springer, pp. 915–923.

[14] Orobix: Retina u-net. https://github.com/orobix/retina-unet.

[15] Risser, E., Wilmot, P., and Barnes, C. Stable and controllable neural texture synthesis and style transfer using histogram losses. *arXiv preprint arXiv:1701.08893* (2017).

[16] Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (2015), Springer, pp. 234–241.

[17] Simonyan, K., and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[18] Wang, C., Zhao, Z., Ren, Q., Xu, Y., and Yu, Y. Dense u-net based on patch-based learning for retinal vessel segmentation. *Entropy 21*, 2 (2019), 168.

[19] Zhang, B., Zhang, L., Zhang, L., and Karray, F. Retinal vessel extraction by matched filter with first-order derivative of gaussian. *Computers in biology and medicine 40*, 4 (2010), 438–445.

[20] Zheng, X., Chalasani, T., Ghosal, K., Lutz, S., and Smolic, A. Stada: Style transfer as data augmentation. *arXiv preprint arXiv:1909.01056* (2019).

Appendix A

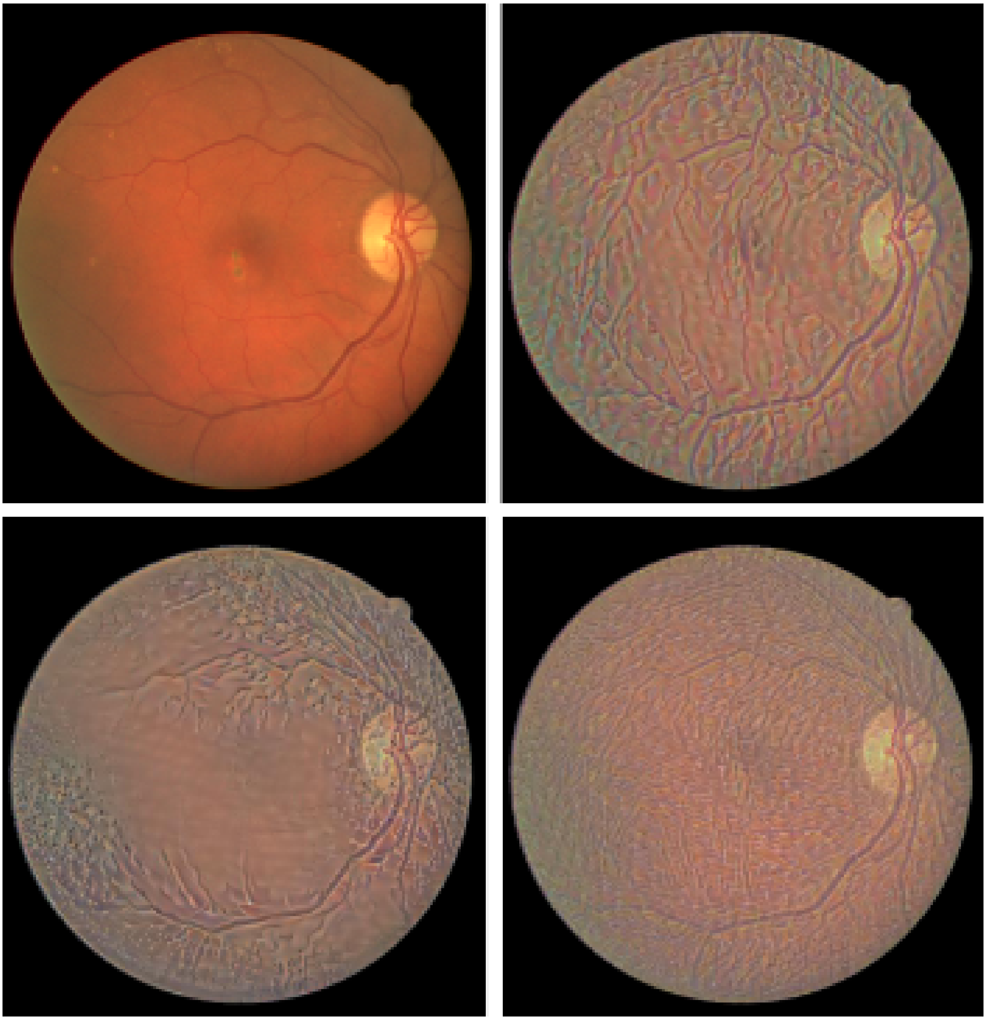

Figure 4. Retina images with corresponding stylization, from left to right row-wise: original, Composition VII, Great Wave, Starry Night

ELTE Eötvös Loránd University, Budapest, Hungary
Email address: p3kxga@inf.elte.hu

J. Selye University, Komárno, Slovakia
Email address: kissae@ujs.sk