University of Tartu

Faculty of Science and Technology

Institute of Technology

Aleksandrs Rebriks

Style transfer based artistic transformation using Albert Gulk's pencil drawings

Bachelor's Thesis (12 ECTS)

Curriculum Science and Technology

Supervisors:

Prof. Gholamreza Anbarjafari

Mr. Hans Hõrak

Tartu 2021

Abstract

Style transfer based artistic transformation using Albert Gulk's pencil drawings

Style transfer is the process of retouching images to appear in the artistic style of another image. Researchers have conducted intensive work using advanced machine learning methods and high level mathematical models in order to transfer style of artworks. One of the challenges, which still exists, is to perform style transfer which preserves the shapes of the content image as much as possible while transferring finer style patterns. In this thesis, we investigate how state-of-the-art style transfer models can be improved such that the shape of the content image is less disturbed during the process. For this purpose we introduce a discrete wavelet transform based neural style transfer pipeline. Experiments were conducted using the artworks of Albert Gulk, a famous Estonian artist of the Kursi school. A challenge in Albert Gulk's pencil drawings is that they are monotone works, hence using any basic style transfer will significantly disturb the shape of content image. Experimental results show that the proposed wavelet based neural style transfer approach can preserve the shape of content when monotone artworks are used as style images.

CERCS:

T111 Imaging, image processing

Keywords:

Style transfer, Deep learning, Wavelet transformation, Art convolution, Machine learning

Kokkuvõte

Kunstilise stiili ülekanne Albert Gulki pliiatsijoonistustega

Stiili ülekanne on pilditöötluse tehnika, mille eesmärgiks on kanda ühe pildi kunstiline stiil teisele. Teadlased on kasutanud kunstiteoste stiili ülekandmiseks masinõppe meetodeid ja kõrgetasemelisi matemaatilisi mudeleid. Üheks püsivaks väljakutseks on stiili ülekande tehnika, mis säilitaks võimalikult hästi sisupildi suuremad vormid, kandes üle peenemaid stiilimustreid. Lõputöös uurime, kuidas täiendada kaasaegseid stiiliülekandemudeleid nii, et sisupildi vormid oleks töötluse käigus vähem moondatud. Selleks töötame välja diskreetse lainikute teisenduse põhise neuro-stiiliülekande meetodi. Eksperimentides kasutati kuulsa Eesti Kursi koolkonna kunstniku, Albert Gulki, kunstiteoseid. Kuna Gulki pliiatsijoonistused on ühetoonilised, moonutavad lihtsad stiili ülekande meetodid oluliselt sisupildi vorme. Esmased tulemused näitavad, et väljaarendatud lainikute põhise neuro-stiiliülekanne säilitab edukalt sisupildi vorme, kui stiilipiltidena kasutada ühetoonilisi pilte.

CERCS:

T111 Pilditehnika

Märksõnad:

Masinõpe, pilditöötlus, kunstilise stiili ülekanne, diskreetne lainiktransformatsioon

Contents

Abstract **3**

Kokkuvõte **4**

List of Figures **7**

List of Tables **9**

Abbreviations, definitions **10**

Introduction **11**

1 Literature review **13**
- 1.1 Wavelet Transformation . . . 13
- 1.2 Machine Learning . . . 14
- 1.3 Deep Learning . . . 15
- 1.4 Image Style Transfer . . . 16
- 1.5 Image Style Transfer Without Neural Networks . . . 17
- 1.6 Image Style Transfer With Neural Networks . . . 18

2 Methods **20**
- 2.1 The Proposed Pipeline . . . 20
- 2.2 Neural Style Transfer . . . 22
- 2.3 Working setup . . . 25

3 Experimental results and discussion **27**
- 3.1 DWT based NST . . . 28
- 3.2 SWT based NST . . . 29

Conclusion and future work **32**

Acknowledgement **33**

References **34**

Non-Exclusive licence to reproduce thesis and make thesis public **38**

List of Figures

1 Style transfer takes a content image and style image and generate a new image by applying the style into the content [4]. 11

1.1 The flowchart of 1 level DWT and the generated four subbands [9]. 13
1.2 LL, LH, HL, and HH subbands images of a face image from the FERET face database achieved by using DWT. For better visibility, high frequency subbands have been passed through histogram equalisation. 14
1.3 A schematic representation of the convolution operation being applied to a single pixel [11]. 16
1.4 A few different art styles. 17
1.5 SBR process: (a) input image, (b) the first layer after applying circular brush, (c) the image after painting with a brush of radius 4, and (d) the final image, after painting with a brush of size 2 [21]. 18
1.6 Original images on top and the rendered images below [24]. 19

2.1 Albert Gulk and some of his art works. Photo: Leimann Eve 20
2.2 The proposed style transform pipelines utilising discrete wavelet transform and stationary wavelet transform. 21
2.3 The flow chart of neural style transformation using CNN. 22
2.4 Example of neural style transfer between the photo of the Great Wall of China and artwork ”Dwelling in the Fuchun Mountains” by Gongwang Huang [20]. . 23
2.5 Illustration of the architecture of VGG19 model, *conv* stands for convolutional layer and *FC* for fully-connected layer [34]. 25

3.1 Photo of Albert Gulk in his studio used for creating baseline. Photo: Urmas Vadi 27
3.2 Style transfer of four of Alber Gulk’s artworks. 28
3.3 The artwork which has been used for style transfer in DWT experiments. 28

3.4 Results of two DWT experiments 2: a)NST has been used to transfer the style in LL subband; b)NST has been used to transfer the style in HH subband. . . . 29
3.5 The content images that have been used for conducting the third experiment. . . 30
3.6 Four of Albert Gulk's artworks used for style transfer in the experiment 3. . . . 31
3.7 Results of experiment 3 utilizing SWT: a)content image is Fig. 3.5 (a) and style image is Fig. 3.6 (a); b)content image is Fig. 3.5 (b) and style image is Fig. 3.6 (b); c)content image is Fig. 3.5 (c) and style image is Fig. 3.6 (c); d)content image is Fig. 3.5 (d) and style image is Fig. 3.6 (d). 31

List of Tables

2.1 Weighting factors for content and style loss used in the model, mentioned in Eq. 2.1. 23
2.2 Weighting factor for content and style loss used in the model, mentioned in Eq. 2.5. 24
2.3 Number of parameters of VGG19 model neural network, for each layer. Total number of parameters is 144M. 26

Abbreviations, definitions

AI - artificial intelligence

ANN - artificial neural network

CNN - convolutional neural network

DWT - discrete wavelet transform

HH - high-high

HL - high-low

IB-AR - image-based artistic rendering

LH - low-high

LL - low-low

NN - neural network

NST - neural style transfer

SWT - stationary wavelet transform

Introduction

Style transfer was introduced in 2015 followed by the launching of an App called Prisma, which allows people to apply famous artists' painting styles to their own photos. Unlike Instagram filters, where some transformations are applied to the color domain of the picture. Basically, in Neural Style Transfer there are two images, namely, style and content. The algorithms copy the style from the style image and apply it to the content image, as can be seen in Fig. 1. Style is usually referred to the patterns and brushstrokes. Although, since then many scientists have worked in this problem, yet there are some open challenges, which are forming the objective of this thesis [1]–[3].

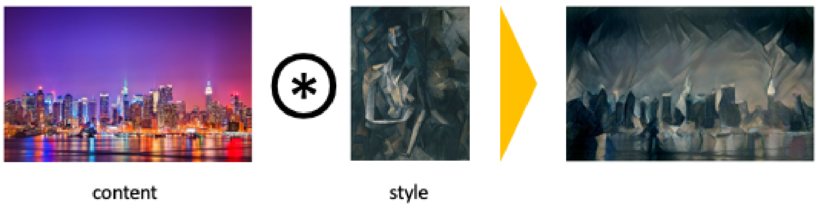


Figure 1: Style transfer takes a content image and style image and generate a new image by applying the style into the content [4].

The main objective of this thesis is to investigate how we can preserve the shapes in the content image while the new style is being transferred. Given that the style transfer is conducted by using a convolutional process, it is very difficult to preserve the shapes. To tackle this issue, we have introduced a new pipeline, in which we are utilizing wavelet transforms. Wavelet transforms enable us to separate an image into low and high-frequency subbands. Given that the shape of an image is defined by its edges which are preserved in high-frequency subbands, it is expected that our proposed pipeline will perfume well.

In this thesis, we are also aiming to focus on transferring the style of content images into the style of Albert Gulk's artworks. Albert has his unique styles which are defined on monotone images (greyscale). Throughout this thesis, we are showing some of his artworks and the results of the style transfer conducted on various content images. We are hoping that this thesis work will also contribute to the better recognition of this Estonian artist.

The thesis is structured into five main parts. Chapter 1 gives an overview of the literature on style transfer as well as the signal processing methodologies such as wavelet transformation which have been used in this work. Chapter 2 describes in detail the proposed pipelines and the reasoning behind all the proposed steps. The experimental results and discussions are presented in chapter 3. The last chapter is dedicated to the conclusion of this work and the possible direction of the future work.

1 Literature review

1.1 Wavelet Transformation

Discrete wavelet transforms (DWT) have been used in many image processing applications [5]–[8]. The decomposition of images into different frequency ranges results in isolating small changes in an image mainly in high-frequency subband images. The two-dimensional wavelet decomposition of an image is performed by applying the one-dimensional DWT along the rows of the image first, and then the results are decomposed along with the columns. This operation results in four decomposed subband images refer to Low-Low (LL), Low-High (LH), High-Low (HL), and High-High (HH), as shown in Fig. 1.1.

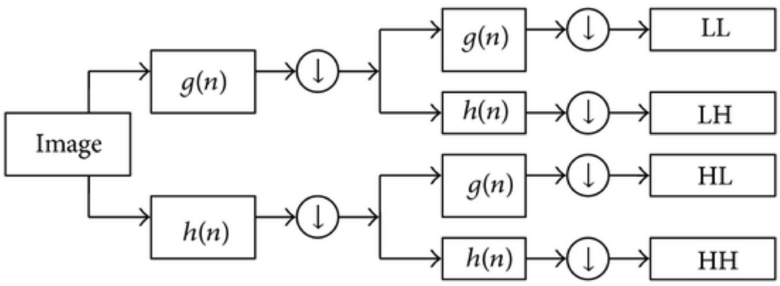


Figure 1.1: The flowchart of 1 level DWT and the generated four subbands [9].

Fig. 1.2 shows the four subband images of a face image from the FERET face database [10], where the 1-level DWT has Daubechies mother wavelet. If the scaling function which was shown in Fig. 1.1 is eliminated from the transformation process, the output subbands will have the same size as the input. This is referred to as stationary wavelet transformation (SWT). SWT has been used in many image processing applications and one of the reasons for its application is preserving the creation of any artifacts that can be created in scaling images in the transformation and inverse transformation process.

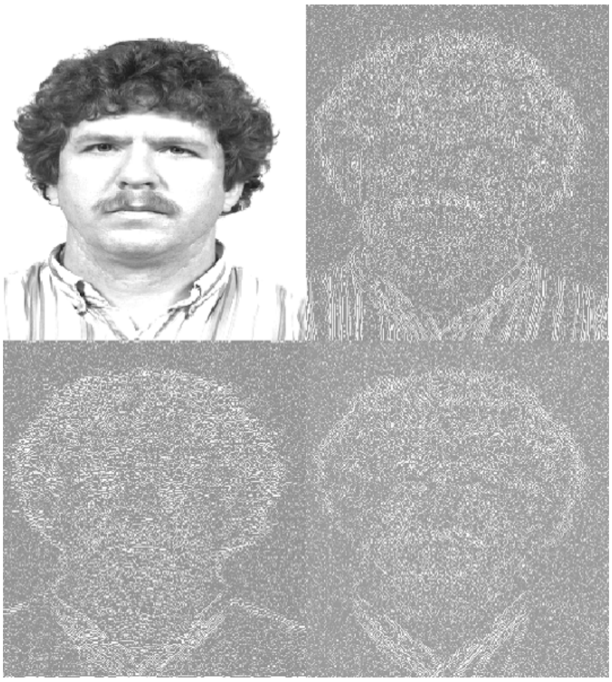


Figure 1.2: LL, LH, HL, and HH subbands images of a face image from the FERET face database achieved by using DWT. For better visibility, high frequency subbands have been passed through histogram equalisation.

1.2 Machine Learning

Machine learning is the study of computer algorithms that can autonomously improve their performance and result in a specific task over a period of time by using and processing provided data. Producing machine learning algorithms consists of creating a model, to which the sample data or training data (training set) is provided, and training the model on the sample data to build and improve its accuracy. Following this, it is possible to provide the model with additional data, for which the analysis is required, for it to make predictions.

There are three main approaches to machine learning:

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning

Neural network (NN) is a machine learning model, whose structure is inspired by the biological neural network of mammals. The structure of a typical NN consists of numbers of interconnected neurons, those are able to receive, process, and transmit data to further connected neurons. Connections between the neurons are called "edges" and have assigned a weight to them, which determines the strength or importance of the corresponding connection. Usually, neurons are packed into layers that can perform different types of processing of transmitted input signals and NNs. The signal is transmitted from the first layer (input layer) to the last layer (output layer) through the intermediate layers in between (hidden layers). Autonomous learning is the process of adjusting the weights after each iteration on the training set, by introducing the cost function, calculating the difference between the desired output (the one on the training set) compared to the actual output. This is followed by a backpropagation method, which adjusts the weights to compensate for the error. NN can be expanded into a more complicated architecture, which is introducing a new way of learning known as deep learning.

1.3 Deep Learning

An artificial neural network (ANN) consists of an input layer, one or more hidden layer(s), and an output layer, being connected to each other through weights which need to be adjusted based on the task at hand. The output, h_i, of a neuron i in the hidden layer is calculated as follows:

$$h_i = \sum_{j=1}^{N} W_j X j, \tag{1.1}$$

where X_j is the input value of a neuron j in the input layer, and W_j is the weight value of its connection to the neuron i in the hidden layer.

For the best possible performance, the number of neurons in the hidden layer needs to be configured based on the requirements of the problem. Increasing the latter quantity, as well as the number of hidden layers, which is required for constructing a convolutional neural network (CNN) and conducting deep learning, has recently become feasible as a result of the improvements made to the computational capabilities of GPUs. The main virtue of CNNs in image processing settings is their contributions to obviating the necessity of performing preprocessing tasks such as smoothing and sharpening, which is due to the convolutions carried out in the course of applying them, as illustrated in Fig. 1.3. A convolution operator finds the relationship between a given number of neighboring pixels. For each image enhancement technique, a specific convolution kernel needs to be used, which is not required by CNNs, as they are capable of learning and creating the required kernels, i.e. filters, on their own.

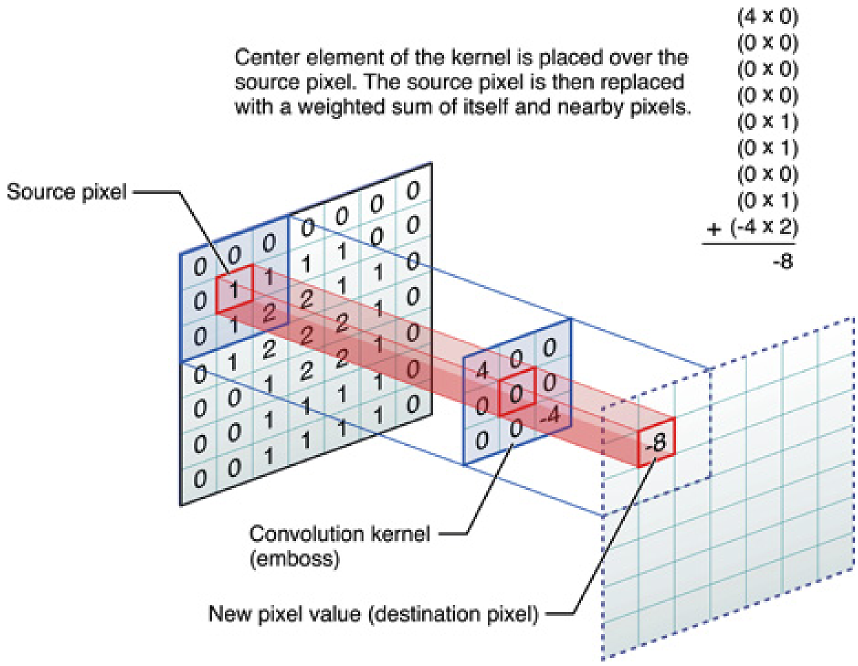


Figure 1.3: A schematic representation of the convolution operation being applied to a single pixel [11].

1.4 Image Style Transfer

Image style transfer aims to transfer the style of one image to another [12]. In this process, unlike many applications of ANN, image style transfer generates a new image based on the features of the input image known as the content image which now has the desired image style. From an artistic point of view, there are many different styles, which can be considered as style images, as can be seen in Fig. 1.4. In other words, a typical image style transfer algorithm transforms only the content image style, and the structure of the content image will be kept the same.

There have been many research works focusing on conducting style transfer. Initial works were focusing on texture synthesis [13]. Signal processing tools such as wavelet have also been used to achieve such transformation [14] without any depth investigation on such usages, which is being investigated within this thesis. Recently more ANN based style transformations have been introduced which are trying to preserve more structure of the image while transferring the styles [15]–[17].

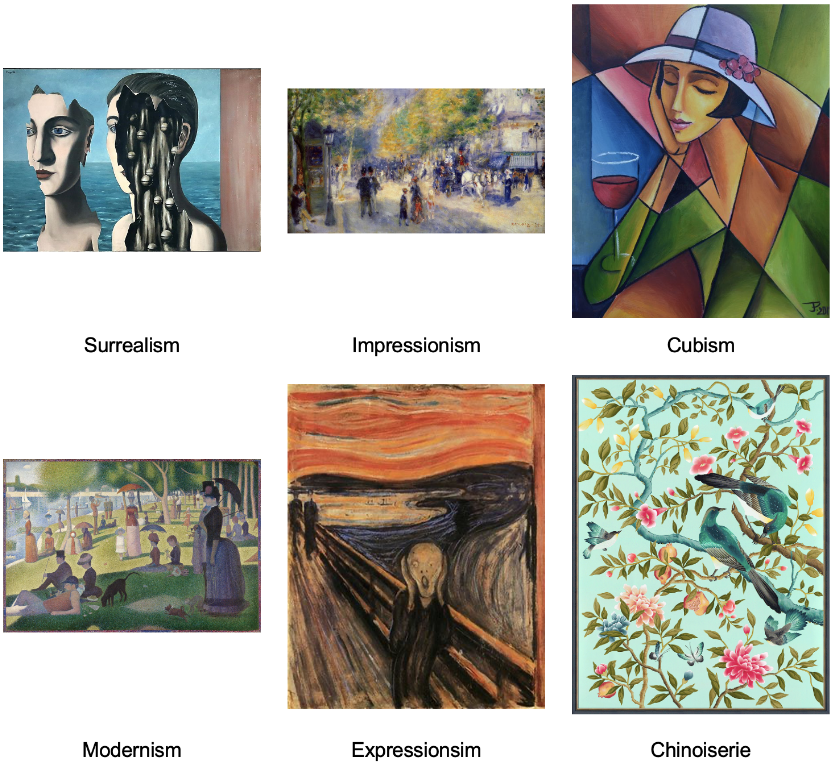


Figure 1.4: A few different art styles.

1.5 Image Style Transfer Without Neural Networks

As it was stated in the previous section, image style transfer was initially conducted without the use of ANN. Such techniques are usually referred to as image-based artistic rendering (IB-AR) [18]–[20]. One of the approaches is using the placement of virtual strokes upon a canvas to render an image with the desired style. This method is referred to as stroke-based rendering (SBR). [21], which can be seen in Fig. 1.5.

The problem with SBR is that each style requires its own algorithm and approach. Another commonly used IB-AR technique is known as region-based rendering, which is taking into account the region segmentation to enable the adaptation of style transferring based on the content within the region [22], [23]. For instance, Song et al. [24] introduced manipulating geometry for artistic styles, in which regions with several canonical shapes are replaced by simplified shapes, as shown in Fig. 1.6.

Another none-ANN based image style transfer is by using image processing filters [25]. For instance, bilateral [26] and difference of Gaussians filters [27], where used to produce cartoon-like effects. Image-filtering based rendering algorithms are generally straightforward to implement and efficient in practice, however, they are limited in style diversity.

Figure 1.5: SBR process: (a) input image, (b) the first layer after applying circular brush, (c) the image after painting with a brush of radius 4, and (d) the final image, after painting with a brush of size 2 [21].

Hence, it is obvious that ANN based image style transfer is a necessity in order to achieve a more holistic methodology for transferring various styles without introducing style-based limitations and preserving the structure of the content image in the best possible form.

1.6 Image Style Transfer With Neural Networks

ANN based style transfer utilizes Convolutional Neural Networks (CNN). Typical CNN based style transfer uses a variety of neural activations from different layers of a CNN in order to represent the artistic style of an image [28]. Different optimization processes are also used [29], [30] in order to generate a new image from white noise by matching the neural activations with the content image and the style image. The results can be improved by applying different complementary schemes such as spatial constraints [31] and semantic guidance [32]. Li et al. proposed a novel interpretation of neural style transfer by treating it as a domain adaptation problem in order to match the feature distributions between the style images and the generated images [28].

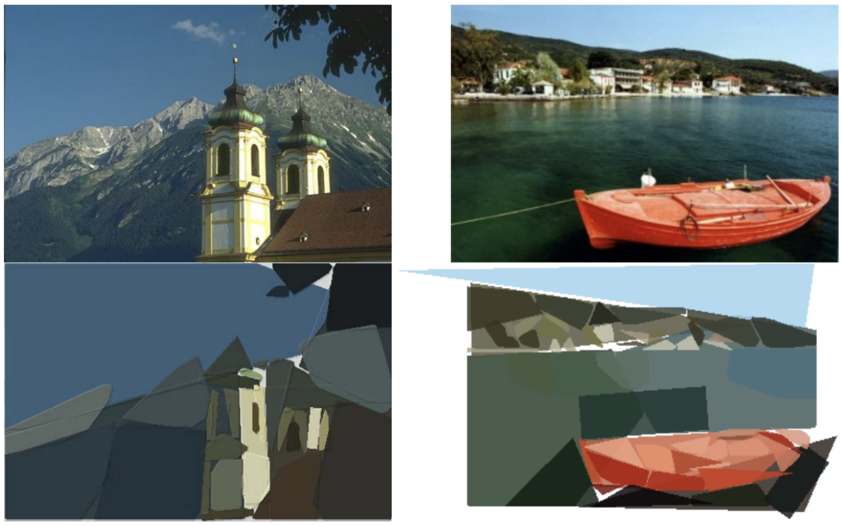


Figure 1.6: Original images on top and the rendered images below [24].

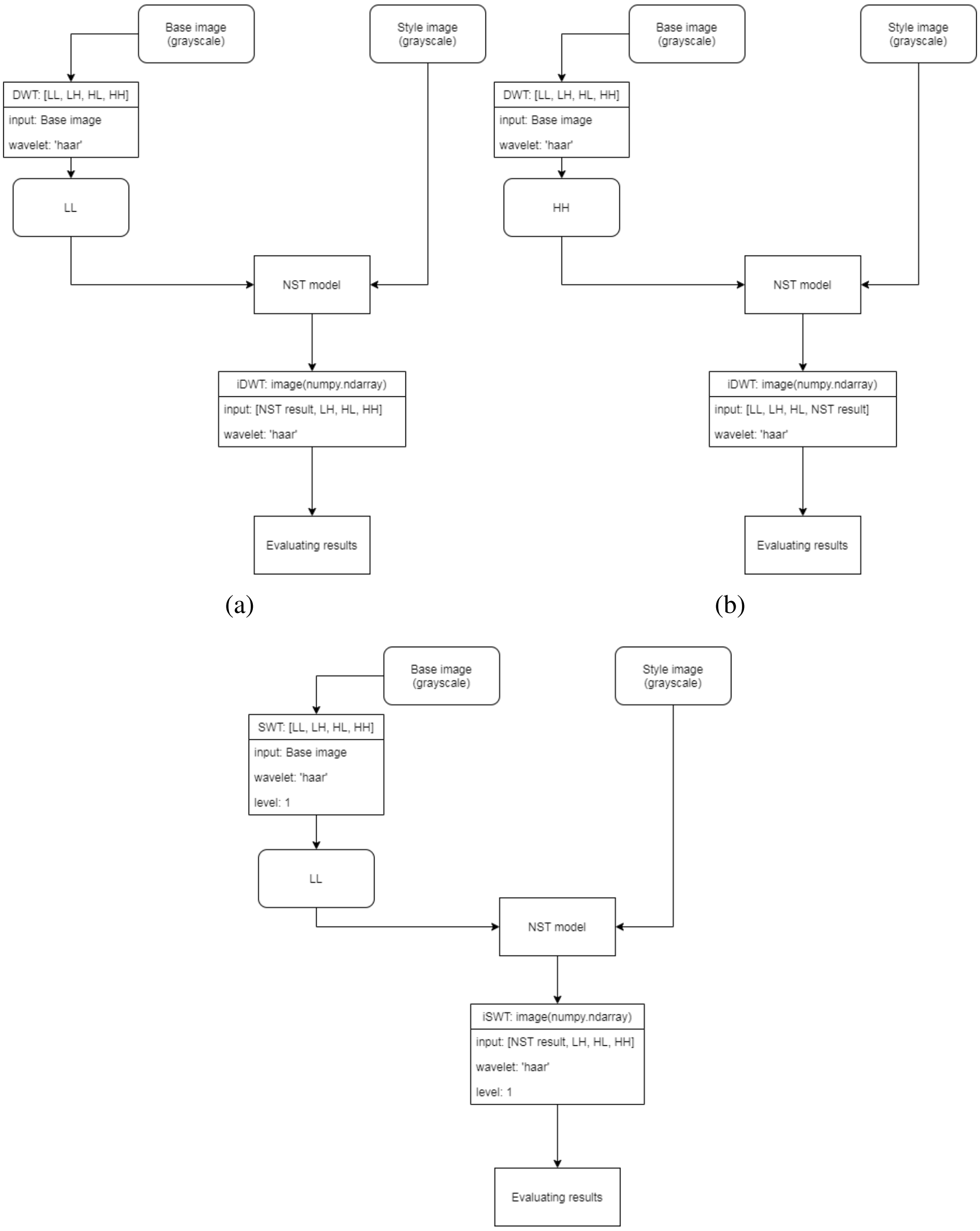
2 Methods

2.1 The Proposed Pipeline

Although, as it was shown in Chapter 1, there have been lots of studies in style transfer, there have not been many attempts in order to preserve the shape of content. One of the main reasons has been the psychovisual effect created by using color images in this process. This thesis is focusing on creating a new style transfer pipeline, in which the style of the content image will transform into Albert Gulk's artworks. Albert Gulk is a known Estonian artist who is living in Tartu and is well known for greyscale artworks. Fig. 2.1 shows his picture and some of his art works.

Figure 2.1: Albert Gulk and some of his art works. Photo: Leimann Eve

In this thesis, we are focusing on creating a new style transfer pipeline in which not only the style but the shape of the content image will preserve even more. For this purpose, we have introduced three setups which are taking into account the utilization of wavelet transformation aiming to preserve the shape while style has been transformed. These workflow diagrams of proposed experiments can be seen in Fig. 2.2.



(c)

Figure 2.2: The proposed style transform pipelines utilising discrete wavelet transform and stationary wavelet transform.

Discrete wavelet transform (DWT) and stationary wavelet transform (SWT), in which the input and output subbands have the same size (i.e., no scaling function is being applied) have been employed to improve the shape of the content image by applying style transformation into either high frequency subbands (Fig. 2.2 (b)) or main components of image in low frequency subbands (Fig. 2.2 (a) and (c)).

2.2 Neural Style Transfer

In general, separating a content from a style in natural images is extremely challenging. However, recent advancements in CNN allowed solving this problem. Gatys et al. [14] were first to study and test the usage of CNNs in reproducing appealing artworks on ordinary images. Their experimental results concluded that CNN feature extraction ability is capable of extracting content from arbitrary image and style of famous artwork. Following this, Gatys et al. [20] proposed to use CNN for recombining the content of any photo and the style of specific artwork, which led to opening a new field called Neural Style Transfer. Fig. 2.3 illustrates the general flow of the method proposed in [20].

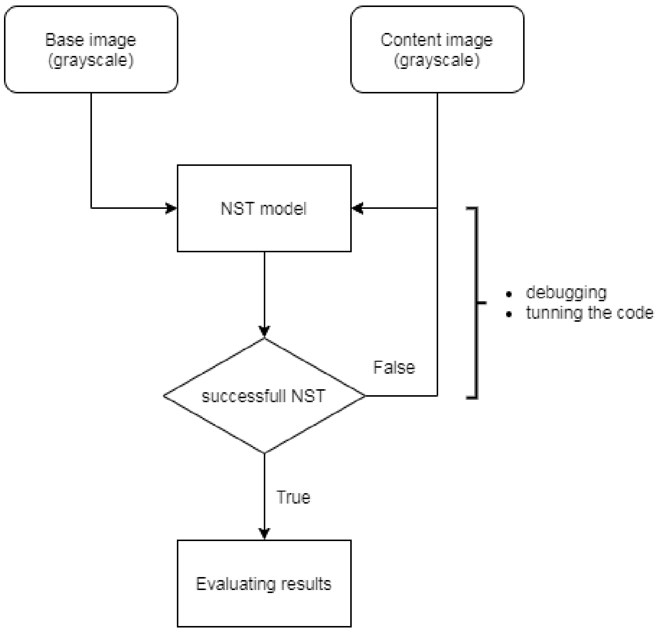


Figure 2.3: The flow chart of neural style transformation using CNN.

Neural Style Transfer (NST) is a technique of recombining images in the style of other images using CNNs, by extracting content from content image (base) and style from style image and merging them, as can be seen in Fig. 2.4. In this thesis, we will use NST as a backbone prior to our experiments and tests.

The main principle of generating a combinational image is to introduce two distance functions. First being, content distance, which indicates the distance of generated image content from the content image representation, and, second, style distance, indicating the distance of generated image style from the style image representation [14]. In neural style transfer the content and style distance functions are represented as a content loss and style loss. The output is

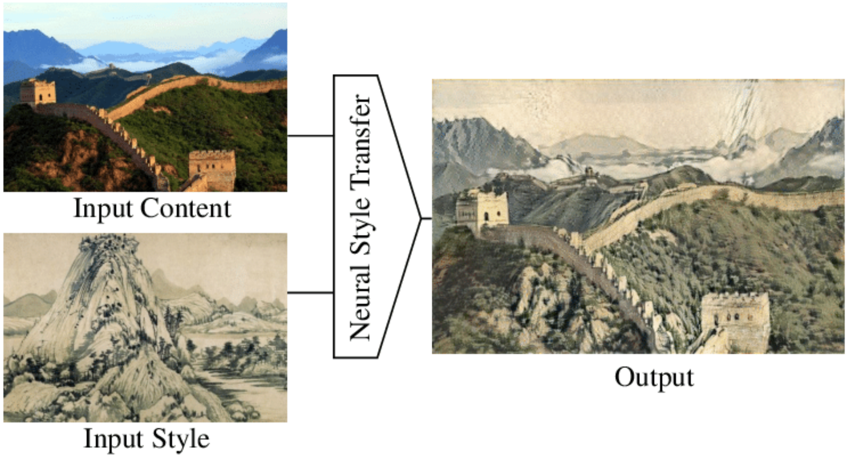


Figure 2.4: Example of neural style transfer between the photo of the Great Wall of China and artwork "Dwelling in the Fuchun Mountains" by Gongwang Huang [20].

generated iteratively by minizing the sum of those two [14]:

$$L_{total} = \alpha L_{content} + \beta L_{style}, \tag{2.1}$$

where α and β are the weights for content and style losses (see Table 2.1 for more specific values used in our experiments).

Table 2.1: Weighting factors for content and style loss used in the model, mentioned in Eq. 2.1.

Weight name	Value
Content weighting factor, α	$2.5 \cdot 10^{-8}$
Style weighting factor, β	$1 \cdot 10^{-6}$

In order to construct a combinational image with the same content as the base image it is necessary to match their feature maps. Feature map is the output of filters applied to the previous layer or, simply, the collection of all features extracted from given layer. Therefore, a layer with N_l distinct filters has N_l feature maps each with a size of M_l, where M_l is the size of height times the width of the feature map. Following this, the feature responses of the given filter l are stored in a matrix $F^l \in R^{N_l \times M_l}$, where F^l_{ij} is the activation of i^{th} filter at j position in the given layer l. Image information is encoded at different layers of CNN in hierarchy and the higher layers become significantly more sensitive to the actual content of the image. Consequently, the feature maps in higher layers of the CNN are referred as the image content representation [16].

Given a content layer l, the feature maps of content image and generated image are P^l and F^l consecutively. We then define the squared-error loss between the two feature maps and our content loss is described as follows:

$$L_{content} = \frac{1}{2} \sum_{ij} (F^l_{ij} - P^l_{ij})^2 \tag{2.2}$$

Matching the content of the generated image and content image means minimizing this content loss, in which case the *feature responses* on the given layer l of the generated image are very similar to the responses on the content image. Following the above-mentioned image information encoding in CNN layers, when reaching higher layers, one can transfer the content from the content image.

To obtain the representation of the style from the style image one needs to extract the *feature correlations* of multiple layers, which will capture the image texture information, but not the content [33]. These feature correlations can be obtained with the Gram matrix $G^l \in R^{N_l \times N_l}$ described as follows:

$$G^l_{ij} = \sum_k (F^l_{ik} F^l_{jk})^2 \tag{2.3}$$

where G^l_{ij} is the dot product of vectorized feature maps i and j in a given layer l.

It is important to note that we are using multiple levels of CNN for style extraction. Let $\vec{a}$ and $\vec{x}$ be the style image and generated image, and A^l and G^l are their style representations in a given layer l. Then the contribution of the given layer l to the *total style loss* can be described as a mean-squared distance between the Gram matrices of style image and Gram matrices from the generating image:

$$L^l_{style} = \frac{1}{4N_l^2 M_l^2} \sum_{ij} (G^l_{ij} - A^l_{ij})^2, \tag{2.4}$$

where N_l and M_l are the number of the feature maps and the size of of each feature map. Following from the Eq. 2.4, the total style loss is then

$$L_{style} = \sum_{l=0}^{L} w_1 L^l_{style}, \tag{2.5}$$

where w_1 are weighting factors assigned to each layer and L is the total number of layers (see Table 2.2 for more specific values used in our experiments).

Table 2.2: Weighting factor for content and style loss used in the model, mentioned in Eq. 2.5.

Total style loss weighting factor w_1	$1 \cdot 10^{-6}$

Finally, to perform Neural Style Transfer between an artwork $\vec{a}$ and arbitrary image $\vec{p}$, one need to generate a new image $\vec{x}$ that matches the content representation of $\vec{p}$ and the style representation of $\vec{a}$ simultaneously. We minimize the distance of the feature representation of a generated image from the content representation of the arbitrary image on *one layer*, while, at the same time minimizing the distance of the style representation of the generated image from the style representation of the artwork on *multiple layers.*. In that way, the total loss function we need to minimize corresponds to

$$L_{total}(\vec{p}, \vec{a}, \vec{x}) = \alpha L_{content}(\vec{p}, \vec{x}) + \beta L_{style}(\vec{a}, \vec{x}), \tag{2.6}$$

where α and β are, as mentioned before, the weights for content and style reconstruction.

2.3 Working setup

To run our experiments we have chosen the VGG model, convolutional neural network model proposed by Karen Simonyan and Andrew Zisserman [34], as a backbone and we have utilized one of its variations - VGG19 with pre-trained *ImageNet* weights. VGG19 consists of 19 layers (note that when counting the number of layers, only those layers that have tunable or trainable weights/parameters are taken into account. Pooling layers serve to downsample the detection of features in feature maps, by reducing the output dimensions. Hence, there are no parameters), containing 16 convolutional layers, 5 pooling layers, which follow some of the convolutional layers, but not all of them, 3 fully-connected layers and 1 soft-max layer (Fig 2.5).

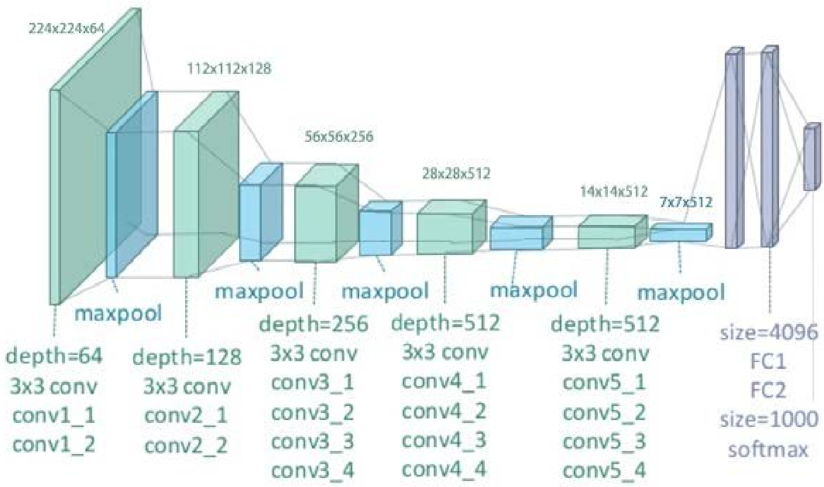


Figure 2.5: Illustration of the architecture of VGG19 model, *conv* stands for convolutional layer and *FC* for fully-connected layer [34].

The code for the model was followed from publicly available Keras GitHub repository, which is neural network API, recently adopted by TensorFlow - an open-source library machine learning. Wavelet transforms were conducted using MATLAB 2018 and the code for NST was written using Python 3.7.4. The tunned code, with experimental results as well as extra results and full list of required libraries can be found in the author of this thesis' Gitlab repository[1]. The weighting factors (mentioned in Section 2.2) and number of parameters can be found in Table 2.1, 2.2, and 2.3 respectively.

Table 2.3: Number of parameters of VGG19 model neural network, for each layer. Total number of parameters is 144M.

Layer name	Number of filters	Number of parameters	Number of activations
input			150K
conv1_1	64	1.7K	3.2M
conv1_2	64	36K	3.2M
max pooling			802K
conv2_1	128	73K	1.6M
conv2_2	128	147K	1.6M
max pooling			401K
conv3 1	256	600K	802K
conv3 2	256	600K	802K
conv3 3	256	600K	802K
conv3_4	256	600K	802K
max pooling			200K
conv4_1	512	1.1M	401K
conv4_2	512	2.3M	401K
conv4_3	512	2.3M	401K
conv4_4	512	2.3M	401K
max pooling			100K
conv5_1	512	2.3M	100K
conv5_2	512	2.3M	100K
conv5_3	512	2.3M	100K
conv5_4	512	2.3M	100K
max pooling			25K
fc6		103M	4K
fc7		17M	4K
softmax		4M	1K

[1]https://gitlab.com/EpicMaze/artistic-style.git

3 Experimental results and discussion

As discussed earlier, in this thesis we are following three experiments in which we are utilizing two types of wavelet transformations, namely, discrete wavelet transform and stationary wavelet transform. Additionally, we have compared the results with the neural style transfer algorithm in which no wavelet has been used. Given that Albert Gulk's pencil drawings are in greyscale, within this thesis we have only used the greyscale style transfer.

As the baseline results, in which only NST has been applied, we have used a photo of Albert Gulk taken in his studio, as shown in Fig. 3.1. The original image was in color, however, for our experimental results it has been converted into greyscale, as it can be seen in Fig. 3.1.

Figure 3.1: Photo of Albert Gulk in his studio used for creating baseline. Photo: Urmas Vadi

Fig. 3.2 illustrates the result of implemented NST algorithm without any wavelet transformations on Albert Gulk's original picture using four of his artworks. All artworks have been converted to the grey scale during implementation of NST.

Figure 3.2: Style transfer of four of Alber Gulk's artworks.

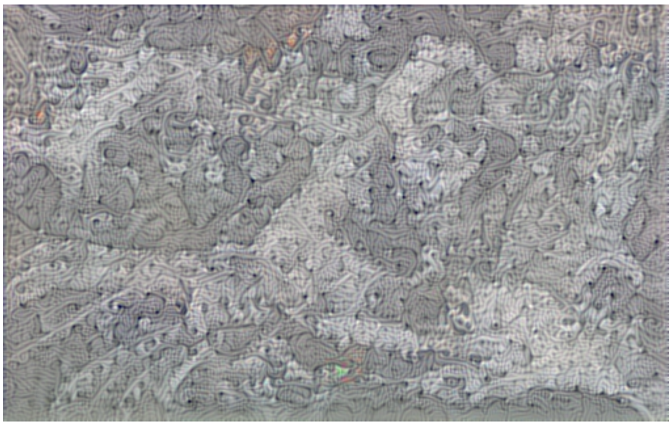
3.1 DWT based NST

In order to preserve shape in the first two experiments, where style transfer is conducted in DWT subbands one of Albert Gulk's artwork, as shown in Fig. 3.3, has been chosen. The embedding have been conducted in LL subband and HH subband and the results can be seen in the Fig. 3.4.

Figure 3.3: The artwork which has been used for style transfer in DWT experiments.

As it is visible in Fig. 3.4, the shape of the content image is preserved significantly when

(a)



(b)

Figure 3.4: Results of two DWT experiments 2: a)NST has been used to transfer the style in LL subband; b)NST has been used to transfer the style in HH subband.

the style transfer is conducted in the LL subband. This is because the shape is defined by details of an image, i.e., edges and high-frequency subbands. When the style transfer is conducted in the LL subband, high-frequency subbands will not be affected, hence there will be significant preservation of shape compared to the original NST and when the transfer is being conducted in HH.

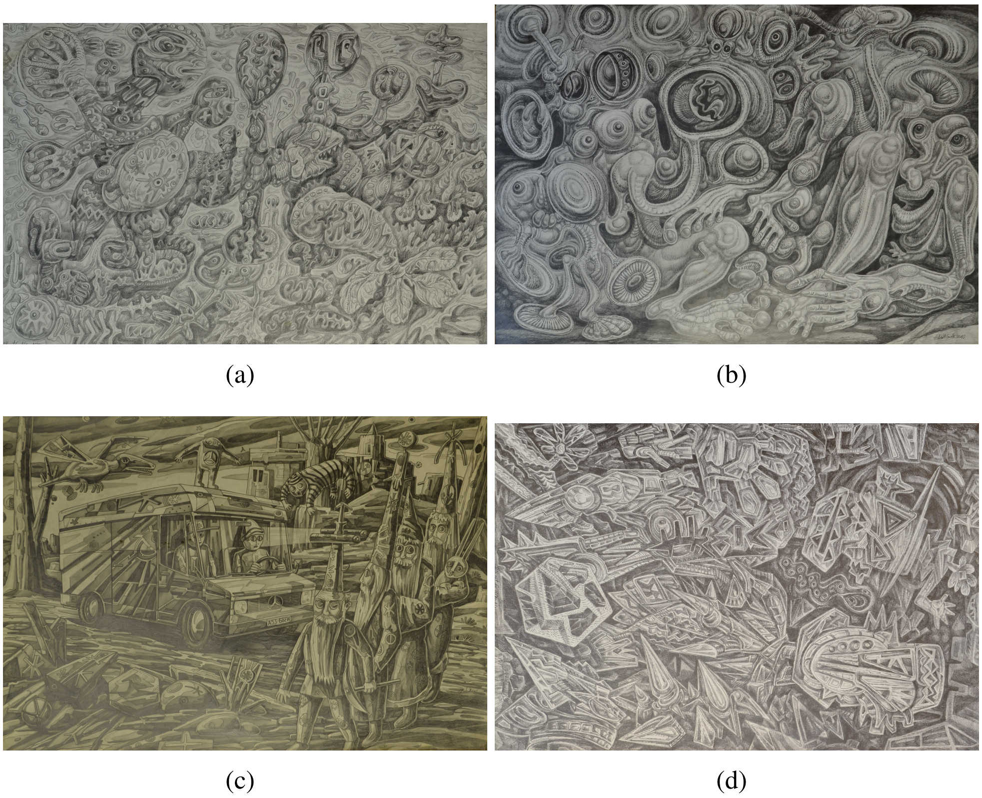
3.2 SWT based NST

DWT has an embedding scaling function and because of that, some artifacts might be created during downsampling and upsampling processes in the style transfer. To prevent these possible artifacts, we have utilized SWT as well. For this purpose, we have used four content images, as shown in Fig. 3.5. For the style transfer, four of Albert Gluk's artworks have been selected, as shown in Fig. 3.6. The result of embedding the content images and the artwork (in respective order) is shown in Fig. 3.7.

(a) (b)

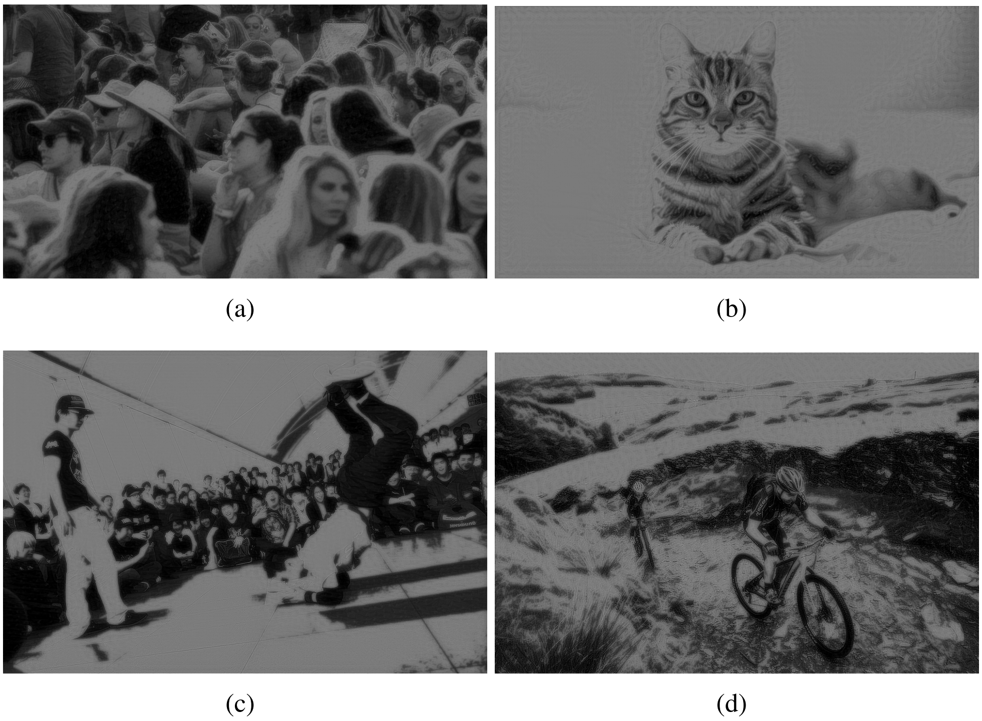
(c) (d)

Figure 3.5: The content images that have been used for conducting the third experiment.



(a) (b) (c) (d)

Figure 3.6: Four of Albert Gulk's artworks used for style transfer in the experiment 3.



(a) (b) (c) (d)

Figure 3.7: Results of experiment 3 utilizing SWT: a)content image is Fig. 3.5 (a) and style image is Fig. 3.6 (a); b)content image is Fig. 3.5 (b) and style image is Fig. 3.6 (b); c)content image is Fig. 3.5 (c) and style image is Fig. 3.6 (c); d)content image is Fig. 3.5 (d) and style image is Fig. 3.6 (d).

Conclusion and future work

This thesis aimed to introduce a new pipeline in which the shape of the content image can be preserved during the process of CNN based style transfer. For this, we introduced a DWT based NST and investigated the effect of transferring styles in LL and HH subbands. Moreover, we investigated the effect of using SWT in the NST process. We used Albert Gulk's pencil drawings for experiments, which showed that the proposed LL-DWT based NST can preserve the shape of content when monotone artworks are used as style images.

In future research, we can investigate in-depth the effect of using more sophisticated wavelet transformations such as complex wavelet transform. Furthermore, the NST algorithm can be changed with advanced generative networks. This can result in even better preservation of the shape of the content image.

Acknowledgement

I would like to thank my closest friends and family members for their unlimited support during my studies. In particular, I would like to thank, Assoc. Prof. Ilona Faustova for the greatest moral support, when it was needed. Additionally I would like to thank my supervisors Hans Hõrak and Shahab for their guidance and infinitive patience while I was conducting my investigation in this topic. Last but not least, I would like to thank Albert Gulk and Peeter Talvistu for providing us access to these artworks in digital form so that we could conduct all the experiments.

References

[1] K. Xu, L. Wen, G. Li, H. Qi, L. Bo, and Q. Huang, "Learning self-supervised space-time cnn for fast video style transfer", *IEEE Transactions on Image Processing*, vol. 30, pp. 2501–2512, 2021.

[2] W.-Y. Chen, J. J. M. Ople, M. J. Si, D. S. Tan, and K.-L. Hua, "Perspective preserving style transfer for interior portraits", *IEEE Access*, vol. 9, pp. 7033–7042, 2021.

[3] D. Egan, M. Alain, and A. Smolic, "Light field style transfer with local angular consistency", in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2021, pp. 2300–2304.

[4] L. A. Gatys, M. Bethge, A. Hertzmann, and E. Shechtman, "Preserving color in neural artistic style transfer", *arXiv preprint arXiv:1606.05897*, 2016.

[5] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis", *Bulletin of the American Meteorological society*, vol. 79, no. 1, pp. 61–78, 1998.

[6] S. Mallat, *A wavelet tour of signal processing*. Elsevier, 1999.

[7] H. Demirel and G. Anbarjafari, "Image resolution enhancement by using discrete and stationary wavelet decomposition", *IEEE transactions on image processing*, vol. 20, no. 5, pp. 1458–1460, 2010.

[8] B. Xu, H. Shen, Q. Cao, Y. Qiu, and X. Cheng, "Graph wavelet neural network", in *International Conference on Learning Representations*, 2018.

[9] H. Demirel, C. Ozcinar, and G. Anbarjafari, "Satellite image contrast enhancement using discrete wavelet transform and singular value decomposition", *IEEE Geoscience and remote sensing letters*, vol. 7, no. 2, pp. 333–337, 2009.

[10] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The feret evaluation methodology for face-recognition algorithms", *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[11] *Single bit convolution example*, Accessed: 20-05-2021. [Online]. Available: `https://tinyurl.com/p2ut4h22`.

[12] L. Liu, Z. Xi, R. Ji, and W. Ma, “Advanced deep learning techniques for image style transfer: A survey”, *Signal Processing: Image Communication*, vol. 78, pp. 465–470, 2019.

[13] L. Zhang, Y. Han, Y. Yang, M. Song, S. Yan, and Q. Tian, “Discovering discriminative graphlets for aerial image categories recognition”, *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5071–5084, 2013.

[14] L. Gatys, A. Ecker, and M. Bethge, “A neural algorithm of artistic style”, *Journal of Vision*, vol. 16, no. 12, pp. 326–326, 2016.

[15] Y. Shih, S. Paris, C. Barnes, W. T. Freeman, and F. Durand, “Style transfer for headshot portraits”, 2014.

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

[17] Y. Gao and J. Wu, “Gan-based unpaired chinese character image translation via skeleton transformation and stroke rendering”, in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 2020, pp. 646–653.

[18] J. E. Kyprianidis, J. Collomosse, T. Wang, and T. Isenberg, “State of the art: A taxonomy of artistic stylization techniques for images and video”, *IEEE transactions on visualization and computer graphics*, vol. 19, no. 5, pp. 866–885, 2012.

[19] A. Semmo, T. Isenberg, and J. Döllner, “Neural style transfer: A paradigm shift for image-based artistic rendering?”, in *Proceedings of the symposium on non-photorealistic animation and rendering*, 2017, pp. 1–13.

[20] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, and M. Song, “Neural style transfer: A review”, *IEEE transactions on visualization and computer graphics*, vol. 26, no. 11, pp. 3365–3385, 2019.

[21] A. Hertzmann, “Painterly rendering with curved brush strokes of multiple sizes”, in *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, 1998, pp. 453–460.

[22] A. Kolliopoulos, *Image segmentation for stylized non-photorealistic rendering and animation*. Citeseer, 2005.

[23] B. Gooch, G. Coombe, and P. Shirley, “Artistic vision: Painterly rendering using computer vision techniques”, in *Proceedings of the 2nd international symposium on Non-photorealistic animation and rendering*, 2002, 83–ff.

[24] Y.-Z. Song, P. L. Rosin, P. M. Hall, and J. P. Collomosse, “Arty shapes.”, in *Computational aesthetics*, Citeseer, 2008, pp. 65–72.

[25] H. Winnemöller, S. C. Olsen, and B. Gooch, “Real-time video abstraction”, *ACM Transactions On Graphics*, vol. 25, no. 3, pp. 1221–1226, 2006.

[26] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images”, in *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*, IEEE, 1998, pp. 839–846.

[27] B. Gooch, E. Reinhard, and A. Gooch, “Human facial illustrations: Creation and psychophysical evaluation”, *ACM Transactions On Graphics*, vol. 23, no. 1, pp. 27–44, 2004.

[28] Y. Li, N. Wang, J. Liu, and X. Hou, “Demystifying neural style transfer”, in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 2017, pp. 2230–2236.

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

[30] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. S. Lempitsky, “Texture networks: Feed-forward synthesis of textures and stylized images.”, in *ICML*, vol. 1, 2016, p. 4.

[31] A. Selim, M. Elgharib, and L. Doyle, “Painting style transfer for head portraits using convolutional neural networks”, *ACM Transactions on Graphics (ToG)*, vol. 35, no. 4, pp. 1–18, 2016.

[32] A. J. Champandard, “Semantic style transfer and turning two-bit doodles into fine artworks”, *arXiv preprint arXiv:1603.01768*, 2016.

[33] L. A. Gatys, A. S. Ecker, and M. Bethge, “Texture synthesis using convolutional neural networks”, *arXiv preprint arXiv:1505.07376*, 2015.

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

Non-Exclusive licence to reproduce thesis and make thesis public

I, **Aleksandrs Rebriks**,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

   Style transfer based artistic transformation using Albert Gulk's Artworks

   supervised by Prof. Gholamreza Anbarjafari and Mr. Hans Hõrak

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Aleksandrs Rebriks

Tartu, 24.05.2021