

Cycle-Consistent Adversarial Networks and Fast Adaptive Bi-dimensional Empirical Mode Decomposition for Style Transfer

Elissavet Batziou^{*†}, Petros Alvanitopoulos[†], Konstantinos Ioannidis^{*}, Ioannis Patras^{*},
Stefanos Vrochidis[†] and Ioannis Kompatsiaris[†]

^{*}School of Electronic Engineering and Computer Science

Queen Mary University of London, Mile End Road London E1 4NS

Email: {e.batziou,ipatras}@qmul.ac.uk

[†]Information Technologies Institute

Centre for Research and Technology Hellas

6th Km Charilaou-Thermi Road, Thessaloniki, Greece

Email: {batziou.el,palvanitopoulos,kioannid,stefanos,ikom}@iti.gr

Abstract—Recently, research endeavors have shown the potentiality of Cycle-Consistent Adversarial Networks (CycleGAN) in style transfer. In Cycle-Consistent Adversarial Networks, the consistency loss is introduced to measure the difference between the original images and the reconstructed in both directions, forward and backward. In this work, the combination of Cycle-Consistent Adversarial Networks with Fast and Adaptive Bidimensional Empirical Mode Decomposition (FABEMD) is proposed to perform style transfer on images. In the proposed approach the cycle-consistency loss is modified to include the differences between the extracted Intrinsic Mode Functions (BIMFs) images. Instead of an estimation of pixel-to-pixel difference between the produced and input images, the FABEMD is applied and the extracted BIMFs are involved in the computation of the total cycle loss. This method enriches the computation of the total loss in a content-to-content and style-to-style comparison by connecting the spatial information to the frequency components. The experimental results reveal that the proposed method is efficient and produces qualitative results comparable to state-of-the-art methods.

I. INTRODUCTION

Neural Style transfer is a deep learning technique that renders a content image in different styles. It is based on Convolutional Neural Networks (CNNs) and creates a stylised image by separating and recombining image content and style. Style transfer is popular for the creation of artificial artwork from photos with extensive usage in mobile application development and the game industry. Fig. 1 illustrates a style transfer example. The content image (left image) is an image of Gendarmenmarkt square in Berlin and the style image (middle image) is the famous painting “Cafe Terrace at night” of Van Gogh. The third image is the new stylised image created by the style transfer method [1], which fuses the content image of Gendarmenmarkt square with the style image of Van Gogh.

Before the Deep Learning era, the style transfer problem was mainly based on parametric and non-parametric methods for texture synthesis and transfer. The parametric approach of [2] presents a method for synthesis of a novel image from an input texture by generating and sampling randomly.

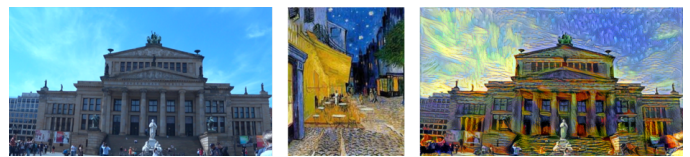


Fig. 1. Style transfer from a content image (Gendarmenmarkt square in Berlin) and a style image (painting of Vincent Van Gogh “Cafe Terrace at night”) for the generation of a new stylised image [1]

Wei and Levoy [3] presented a different method based on Markov Random Field texture models. It generates textures through a deterministic searching process. Bertalmio et al. [4] decompose an image into the sum of two functions, one that can be efficiently reconstructed via inpainting and one that can be efficiently reconstructed via texture synthesis. Efros and Freeman [5] synthesise a new texture by taking patches of an existing texture and append them together in a consistent way. However, the recent advances in Neural Style Transfer have shown significant improvement of these parametric and non-parametric methods.

The method proposed by Gatys et al [1] has been a seminal work in style transfer. This technique uses Deep Neural Networks (DNN) to encode not only the content but also the style information of an image, using the Gram matrices of the input style image to capture the linear dependencies among several feature vectors. A similar approach with DNNs [6] trains a DNN which is able to learn different characteristics that specify each style, such as colors, shapes and edges. Huang et al. [6] adjust channel-wise statistics of the content features by adaptive instance normalisation (AdaIN) and train a feature decoder by utilizing combinational scale-adapted content and style losses. In [7] the authors introduce an instance normalisation module to replace batch normalisation in order to improve the performance of the deep neural generators in style transfer from painting images. Recently, Sheng et al. [8] proposed feature decoration that generalises AdaIN

and whitening and coloring transformation (WCT). They use Zero-phase Component Analysis (ZCA) in the proposed style transfer method. In [9] the authors propose an adversarial feed-forward network for arbitrary style transfer. They introduce techniques to tackle the problem of adversarial training from multi-domain data and train a discriminator to rank and find the representative generated images in each style category.

The style transfer becomes more challenging when the proposed style is not based on a single image but from a set of paintings. The style transfer from a collection of paintings of a creator or a style is a more complex challenge since it involves the representation of images with proper features which could describe the style of a school of art (impressionism, cubism, etc.) or of a creator. In this direction several works have been proposed to tackle style transfer from a collection of images such as CycleGAN [10], Gated-GAN [11], Artsy-GAN [12] and the work of Sanakoyeu et al. [13].

In this work, we extend the recent Cycle-Consistent Adversarial Network architecture involving also a spectral decomposition in the optimisation problem of loss minimisation, so as to improve the quality of the stylised images. This modification allows the analysis of images into its key texture elements through the Bidimensional Intrinsic Mode Functions (BIMFs) [14] and provides a more qualitative definition of the loss function. A modification in the estimation of the cycle loss function is proposed with the computation of the BIMFs on the given content and style images. Instead of the original BEMD [15], the FABEMD [16] is examined since it is faster and provides an adaptive size of window for the detection of extrema during the BIMF extraction process.

II. RELATED WORK

Gated-GANs have been proposed by Chen et al. [11], as a novel adversarial network, in which it is possible to transfer multiple styles in one single model. Gated-GANs have three modules: an encoder, a gated transformer, and a decoder. The gated transformer allows the user to select style by switching gate. Once the gated transformer is skipped, then the encoder and decoder modules are trained as an auto-encoder to preserve semantic consistency between content images and their stylised ones. Gated-GANs are trained for multiple styles in order to generate new stylised images through weighted connections between the branches of the gated transformer. However, our proposed method includes a Cycle consistency loss function which penalizes incorrect reconstruction of the input image from the translated one (content-to-content and style-to-style generation).

Zhu et al. [10] propose an unpaired image-to-image translation, where the style of a given collection of images can be transferred to the content of a photo. The proposed architecture uses Generators and Discriminators to extend the architecture of a Generative Adversarial Network (GAN) [17] to their Cycle-Consistent Adversarial Network architecture, known also as CycleGAN, which also learn an inverse generator that is able to create an image identically to the original content image from the stylised one. An example is presented in Fig.



Fig. 2. Style transfer to a content photograph from a collection of Van Gogh's painting images [10]

2. However, our proposed approach extends the CycleGAN architecture by integrating the FABEMD component. The representation of content and style images through a set of BIMFs is introduced and examined in the cycle-consistency loss.

Sanakoyeu et al. [13] adopt an encoder-decoder network architecture. They also introduce a fix point loss that ensures stylisation has converged and reached a fix-point after one feed-forward pass. This style-aware content loss forces the stylisation to take place in the decoder. Contrary to this style-aware content loss we keep the cycle consistent loss and combine it with frequency elements so as to achieve high-quality stylized images that also preserve the content patterns.

Liu et al. [18] propose a variation of CycleGAN, namely the Artsy-GAN, by replacing the Cycle-Consistency loss with the so-called perception loss [12]. The purpose of this variation is to introduce a new objective function for diversity in image-to-image translation. The generator of Artsy-GAN model consists of three branches. Each branch receives the same input and produces three different channels of the output images: one luminance channel and two color channels. The output image is reconstructed in its final RGB format. However, the perception loss is calculated based mainly on color and luminance values, and not on the content and style patterns and details that we are tackling through the integration of frequency components inserted by the decomposition process via FABEMD.

In the present study a similar approach is examined having as a goal to transfer the style of a collection of images to the content of an image. One of the main challenges is to pass the style of the collection to the content image, without significant deformations on the objects included in content image. Moreover, we examine the application of Empirical Mode Decomposition in the loss function computation of the CycleGANs model, and how the use of a subset of BIMFs affects the quality of the stylised image. In general, the empirical mode decomposition (EMD) method decomposes signals into a set of spectral components, known as Intrinsic Mode Functions (IMFs). The EMD is a nonlinear non-stationary and an adaptive decomposition method and analyses a signal into a finite and often small number of components. These components form a complete and nearly orthogonal basis for the original signal [19]. The higher index value of the IMF is, the lower spatial frequency components of the original image are involved. Fig. 3 and 4 present the bidimensional empirical

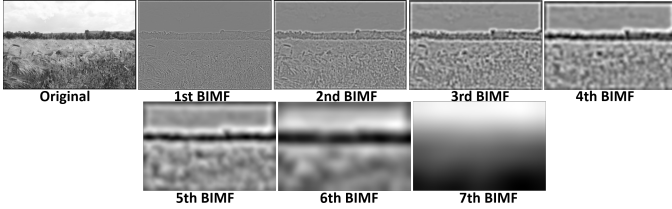


Fig. 3. The original landscape image and the extracted BIMFs.

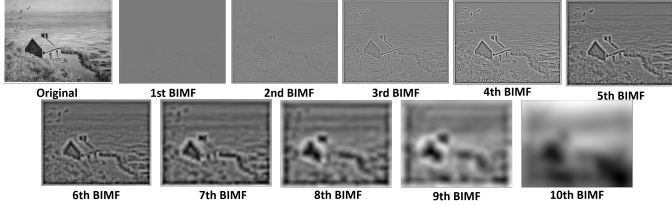


Fig. 4. The image of the painting named “Customs House” by Monet Claude and the extracted BIMFs.

mode decomposition of images into finite number of BIMFs.

III. BACKGROUND

A. Empirical Mode Decomposition

EMD analyses signals based on the local characteristics (extrema) and serves as a spectral representation of nonlinear and non-stationary data. EMD differs from Fourier or Fourier based wavelet analysis [19]. It is based on local characteristics in the original signal and are determined by two successive minima and the maximum which necessarily belongs between the two minima. EMD analyzes nonstationary and nonlinear signals into a finite number of IMFs. An essential requirement for non-stationary and nonlinear data analysis is to have an adaptive basis. The condition of adaptivity means that the decomposition of the original signal is not based on predetermined functions and the definition of basis is data-dependent. In EMD this is achieved through the detection of extrema.

We map visual content (images) into $[-1, 1]$ so as to use bidimensional EMD (BEMD).

BEMD decomposes an image into its bidimensional IMFs (BIMFs) and a Residue, using the local spatial information, i.e. group of pixels. The BIMFs have the following properties:

- A. at any pixel, the mean value of the upper and lower envelopes, defined by the local maxima and minima, approximates zero.
- B. they are locally orthogonal to each other.

In our approach the computation of BIMFs is based on a modified version of BEMD, known also as Fast and Adaptive Bidimensional Empirical Mode Decomposition (FABEMD). FABEMD differs from the original BEMD algorithm, basically in the process of estimating the upper and lower envelopes and in limiting the number of iterations per BIMF to one. The neighboring window method of Nunes et al. [15] is employed to find the local maxima and minima points, in the following

steps, where a data point is considered as a local maximum (minimum), if its value is strictly higher (lower) than all of its neighbors within a given window. The FABEMD is described below:

- 1) Set $i = 1$ and $S_i = I$
- 2) Set $j = 1$ and $W_{T_j} = S_i$
- 3) Obtain the local maxima map of W_{T_j} , denoted as P_j and local minima map of W_{T_j} , denoted as Q_j
- 4) Generate the upper envelope (UE) U_{E_j} and the lower envelope (LE) L_{T_j} of W_{T_j} , by interpolating the maxima points in P_j and the minima points in Q_j , respectively.
- 5) Find the mean envelope (ME) as $M_{E_j} = (U_{E_j} + L_{E_j})/2$
- 6) Calculate $W_{T_{j+1}}$ as $W_{T_{j+1}} = W_{T_j} - M_{E_j}$
- 7) Check whether $W_{T_{j+1}}$ follows the BIMF properties and find the standard deviation (SD), between $W_{T_{j+1}}$ and W_{T_j} defined below so as to compare it to a desired (low) threshold (0.01).

$$SD = \frac{\sum_{x=1}^M \sum_{y=1}^N |W_{T_{j+1}}(x, y) - W_{T_j}(x, y)|^2}{\sum_{x=1}^M \sum_{y=1}^N |W_{T_j}(x, y)|^2} \quad (1)$$

where (x, y) is the pixel position, M is the total number of rows (image height) and N is the total number of columns (image width) of the image array.

- 8) If $W_{T_{j+1}}$ meets the criteria as per step 7, then take $W_i = W_{T_{j+1}}$ set $S_{i+1} = S_i$ and $i = i + 1$, go to step 9. Otherwise, set $j = j + 1$, go to step 3 and continue up to step 8.
- 9) Determine if S_i has less than three extrema points, and if so, the Residue, $R = S_i$, and the decomposition is complete. Otherwise, go to step 2 and iterate again up to step 9.

In this work, we leverage the FABEMD to extract the BIMFs of content and style images for the computation of Cycle-Consistent loss function of CycleGANs.

B. Generative Adversarial Networks

Generative Adversarial Networks (GANs) train a generative model by formulating the problem as a supervised learning problem with two sub-models: the generator model, which is trained to generate new images, and the discriminator model that classifies images as either real (from the domain) or fake (generated). GANs have been applied in several domains of application. Zhang, et al. [20] demonstrate the use of GANs to generate realistic looking photos from textual descriptions of simple objects like birds and flowers. Zhou et al., [21] use ground-truth synthetic-to-synthetic correspondences, to train a convolutional neural network, aiming to predict synthetic-to-real, real-to-real and real-to-synthetic correspondences that are cycle-consistent with the ground-truth. Godard et al. [22] presented an unsupervised deep neural network for single image depth estimation. Based on the concept of GANs, a novel Neural Network architecture has been recently introduced and used in style transfer.

1) *Cycle-Consistent Adversarial Networks*: Cycle-Consistent Adversarial Networks learn mapping functions between two domains X and Y given training samples $\{x_i\}_{i=1}^N \in X$ and $\{y_j\}_{j=1}^N \in Y$. The Cycle-GAN model includes two mappings $G : X \rightarrow Y$ and $F : Y \rightarrow X$. In addition, the model introduces two adversarial discriminators D_X and D_Y , where D_X distinguishes between images $\{x\}$ and stylised images $\{F(y)\}$. Similarly, D_Y aims to discriminate between $\{y\}$ and $\{G(x)\}$.

The total (full) objective contains two kinds of sub-objectives to minimize:

- 1) adversarial losses [17] for matching the distribution of generated images to the data distribution in the target domain (style)
- 2) a cycle consistency loss to prevent the learned mappings G and F from contradicting each other [10].

Adversarial Loss: Adversarial losses are applied [17] to both mapping functions. For the mapping function $G : X \rightarrow Y$ and its discriminator D_Y , the objective is expressed as:

$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = E_{y \sim P_{data}(y)}[\log D_Y(y)] + E_{x \sim P_{data}(x)}[\log(1 - D_Y(G(x)))] \quad (2)$$

where, G tries to generate images $G(x)$ that look similar to images from domain Y , while D_Y aims to distinguish between stylised samples $G(x)$ and real samples y .

A similar adversarial loss is learned for the mapping function $F : Y \rightarrow X$ and its discriminator D_X , denoted by $\mathcal{L}_{GAN}(F, D_X, Y, X)$.

Cycle Consistency Loss: Cycle Consistency loss aims to introduce a two-way direction between the generated image and the original image. In other words, the learned mapping from the original image to the generated one should be ‘‘consistent’’ with its inverse mapping from the generated image to the original one.

Forward and backward cycle consistency are defined as follows: For each image x from domain X , the image stylisation cycle should be able to bring x back to the original image, i.e. $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$, and this is known as forward cycle consistency. In the same manner, for each image y from domain Y , G and F should also satisfy backward cycle consistency, defined as: $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$.

The forward and backward cycle consistency determine the following cycle consistency loss:

$$\mathcal{L}_{cyc}(G, F) = E_{x \sim P_{data}(x)}[\|F(G(x)) - x\|_1] + E_{y \sim P_{data}(y)}[\|G(F(y)) - y\|_1] \quad (3)$$

Total Loss (Full Objective): The total loss of the style transfer model with cycle-consistent adversarial networks is define as a combination of the adversarial losses $\mathcal{L}_{GAN}(G, D_Y, X, Y)$ and $\mathcal{L}_{GAN}(F, D_X, Y, X)$ and the cycle-consistency loss as follows:

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, Y, X) + \lambda \mathcal{L}_{cyc}(G, F) \quad (4)$$

where, λ controls the relative importance of the two objectives. The total loss \mathcal{L} is minimised by solving:

$$G^*, F^* = \operatorname{argmin}_{G, F} \max_{D_X, D_Y} \mathcal{L}(G, F, D_X, D_Y) \quad (5)$$

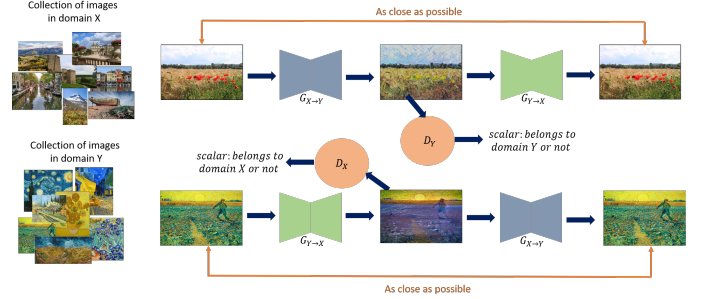


Fig. 5. The style transfer framework with Cycle-Consistent Adversarial Networks

In our approach we combine the CycleGAN architecture with the FABEMD decomposition. We extract BIMFs for both content and style images in order to compute the total cycle consistency loss function. In the following section a thorough description of the proposed methodology is presented.

IV. METHODOLOGY

Based on the aforementioned Cycle-Consistent Adversarial Networks and FABEMD, we present our proposed framework, where the spectral decomposition of content and style images is used to optimise the Generators G and F , and the Discriminators D_X and D_Y .

We denote by $x_i^{bimf}(k)$ the k -th BIMF corresponding to the content image x_i and by $r_i^{bimf}(k)$ the k -th BIMF of the reconstructed image $r_i = F(G(x_i))$. Similarly, let $y_j^{bimf}(k)$ be the k -th BIMF corresponding to the style image y_j and by $s_j^{bimf}(k)$ the k -th BIMF of the stylised image $s_j = G(F(y_j))$. The total number K of extracted BIMFs is examined in the experiments to find the optimal value of K . The forward and

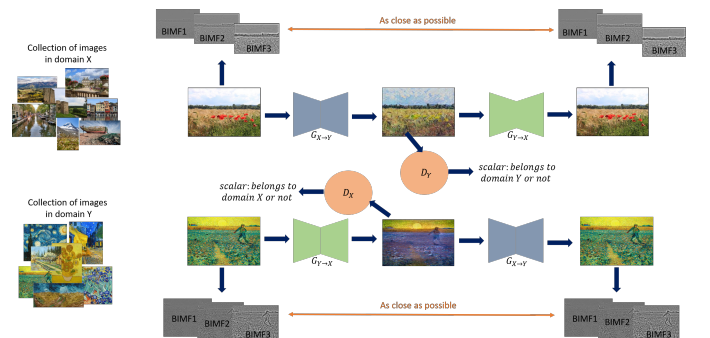


Fig. 6. The proposed Cycle-Consistent Adversarial Networks with FABEMD framework

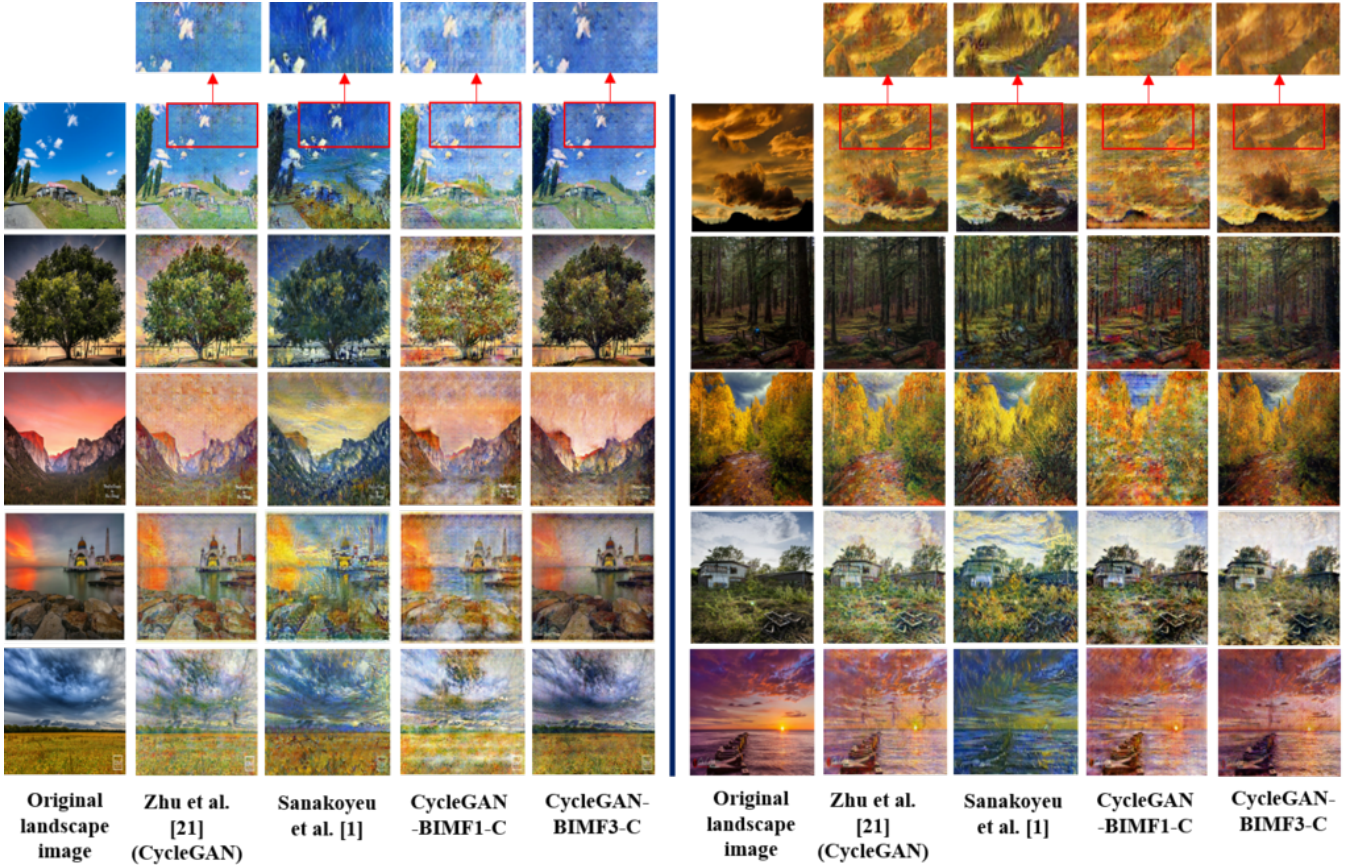


Fig. 7. Comparison of our approach with the CycleGAN model and Sanakoyeu et al., [13] on the “monet2photo” dataset

backward cycle consistency determine the cycle consistency loss as follows:

$$\begin{aligned} \mathcal{L}_{cyc}(G, F) = & E_{x \sim P_{data}(x)} \sum_{k=1}^K [\|r^{bimf}(k) - x^{bimf}(k)\|_1] + \\ & + E_{y \sim P_{data}(y)} \sum_{k=1}^K [\|s^{bimf}(k) - y^{bimf}(k)\|_1] \end{aligned} \quad (6)$$

while the total loss function remains the same.

In Fig. 6 we present our framework, in which the network contains two generators F and G and two discriminators D_X and D_Y . The model architecture examined in this work is similar to Isola et al. [23]. The first generator G takes an image of a landscape and generates painting images of the given style. The second generator F generates photos of landscapes, given photos of paintings. There are two ways in which cycle-consistency loss is calculated and used to update the generator models in each training iteration.

Firstly, the cycle-consistency loss computes the difference between the BIMFs of the input of G and the BIMFs of the output of F , i.e. $\|r^{bimf}(k) - x^{bimf}(k)\|_1$. The generated models are updated accordingly to reduce the differences in the images. Secondly, the cycle-consistency loss computes the difference between the BIMFs of the real image in domain Y

and the BIMFs of the stylised image (fake image in domain Y), i.e. $\|s^{bimf}(k) - y^{bimf}(k)\|_1$. The comparison is performed and the differences are summed for the K selected image decomposition layers.

We adopt the use of instance normalisation instead of batch normalisation. The selected encoder-decoder of the generator has a “U-Net” [24] architecture with skip connections between mirrored layers. The generator loss function is a cross-entropy loss.

V. EXPERIMENTS

In this section we present the datasets that we use for our experimental comparison, the settings of our implementation and the results of the qualitative evaluation. The baseline methods of the presented comparison are the CycleGAN approach [10] and the approach of Sanakoyeu et al. [13]. These methods have shown better performance than other Neural Style Transfer methods in painting image collections.

A. Datasets

In the present experiments two datasets from the TensorFlow catalogue¹ are exploited. The first dataset is the “monet2photo” dataset. It comprises 1074 painting images of

¹https://www.tensorflow.org/datasets/catalog/cycle_gan

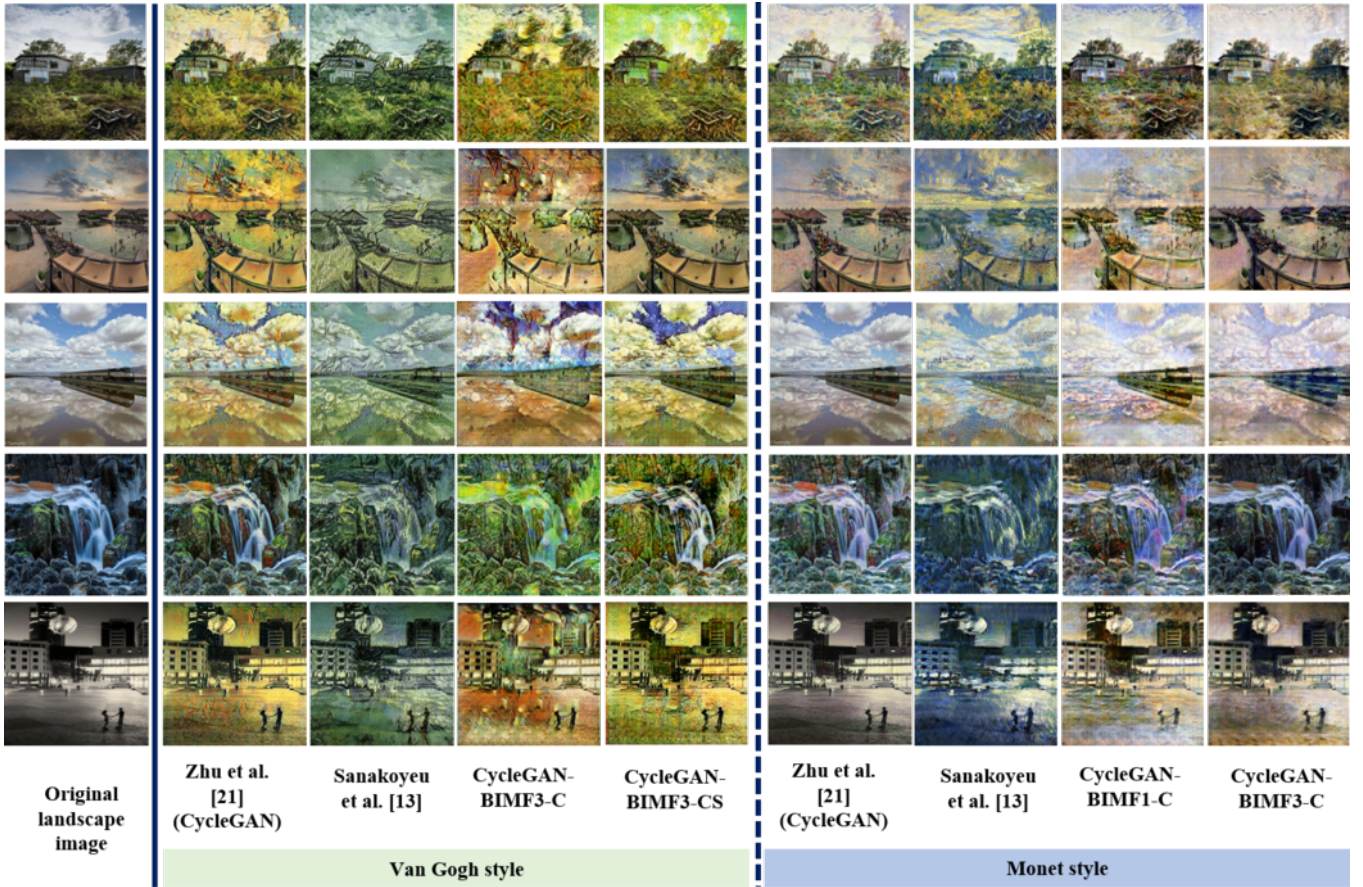


Fig. 8. Comparison of our approach with the CycleGAN model and Sanakoyeu et al., [13] on both "vangogh2photo" and "monet2photo" datasets

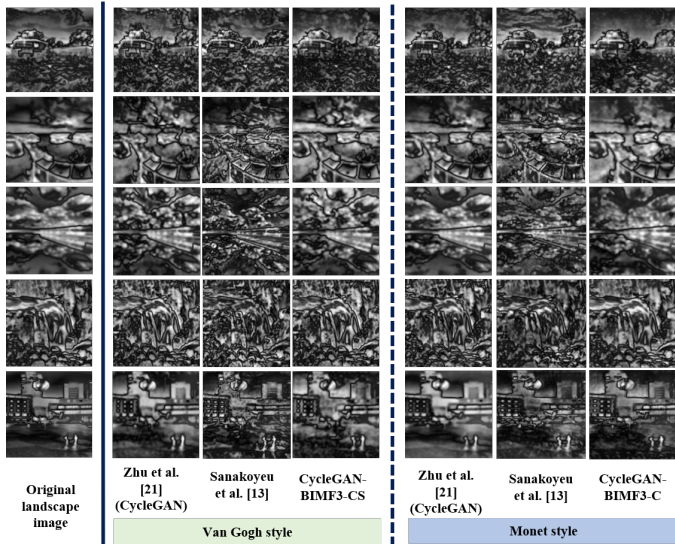


Fig. 9. Saliency detection results corresponding to the images of Fig. 8

Monet and 6853 photos. These images are downloaded from Flickr using the tags "Landscape" and "Landscapephotography". The second dataset is the "vangogh2photo". It contains 401 painting images of Vincent Van Gogh and the same

landscape photos as in the first dataset (monet2photo).

B. Settings

For both datasets, the parameter λ is equal to 10. The implementation of the CycleGAN is found online². The number of epochs is set to 40 for all methods and experiments. In the Ahmed and Mandic [25], it is observed that the image edge information is contained in high frequency scales. In the proposed method the number of BIMFs is set to 3. A number higher to 3 does not provide significant improvement and increases the time for training. An example with two different settings $K \in 1, 3$ concerning the "monet2photo" dataset is presented. Regarding the "vangogh2photo" experiments the number of BIMFs is set to 3. Two different models are examined. The first named CycleGAN-BIMF3-C applies the FABEMD only in the content-to-content loss. The second named CycleGAN-BIMF3-CS includes the application of the FABEMD in the content-to-content and style-to-style loss to perform diverse experiments. The latter is a pre-trained model from the CycleGAN-BIMF3-C model, and it was trained for additional 40 epochs analysing both content and style images via FABEMD.

²<https://www.tensorflow.org/tutorials/generative/cyclegan>



Fig. 10. The progress of the stylisation process using CycleGAN-BIMF3-CS method for Van Gogh style

C. Results

As shown in Fig. 7, CycleGAN-BIMF3-C outperforms the original CycleGAN method and the Sanakoyeu et al. [13] method, and the corresponding method where only the first BIMF of each content image is used. From this figure, in the first row, it can be seen that CycleGAN-BIMF3-C does not generate noise patterns in the sky region, an effect that we can observe in the outputs of the rest three models.

Regarding the “monet2photo” experiment it can be seen that the the CycleGAN model produces noise especially on the sky areas. In comparison to the Sanakoyeu et al. and CycleGAN-BIMF1-C model the CycleGAN-BIMF3-C generates images of better quality.

Fig. 8 illustrates some output results of CycleGAN, CycleGAN-BIMF3-C and CycleGAN-BIMF3-CS methods using also the “vangogh2photo” dataset. Two different styles (Van Gogh-Monet) are presented for the same content images and the CycleGAN-BIMF3-CS model produces more qualitative results.

The saliency maps of Fig. 8 are presented in Fig. 9. Saliency detection relies on image features and statistics to localize the most interesting regions of an image. From the saliency maps we observe that our approach preserves the structures of content images. For the implementation of the saliency detection we use the OpenCV library and the class Static saliency based on [26]. In addition to the qualitative comparison of Fig. 8, a quantitative evaluation is also presented in Table I. Each row of Table I corresponds to a stylised image of Fig. 8. For a direct comparison between the saliency map of original landscape image (SM^{or}) and of the corresponding stylised image (SM^{st}), the Trace (or nuclear) norm [27] of the matrix difference is used:

$$Tr(\sqrt{(SM^{or} - SM^{st})^*(SM^{or} - SM^{st})}) = \sum_i \sigma_i \quad (7)$$

where, σ_i is the i -th singular value of the matrix $SM^{or} - SM^{st}$. The computation of the trace norm is based on SciPy³. In this Table it is shown that for the stylised images of our method the trace distance is minimised for the majority of cases.

Fig. 10 presents the progress of the stylisation process for a number of epochs for a given content image using CycleGAN-BIMF3-CS method and Van Gogh style. It can be seen that in the first epochs the style is not sufficiently transferred and

³<https://docs.scipy.org/doc/numpy/reference/generated/numpy.linalg.norm.html>

there is noise in the stylised images. However after 40 epochs the artifacts are reduced and the quality of the stylised image is better.

TABLE I
TRACE NORM OF THE SALIENCY MAP DIFFERENCES OF FIG. 9

Zhou et al. [10]	Sanakoyeu et al. [13]	CycleGAN-BIMF3-CS	Zhou et al. [10]	Sanakoyeu et al. [13]	CycleGAN-BIMF3-C
90171	91674	80024	88090	95426	80963
76903	81565	76531	74492	85466	73918
80210	74566	74264	97914	81509	77637
95222	87048	79269	88359	79749	97649
82599	87572	76088	71731	86364	82098
Van Gogh			Monet		

TABLE II
AVERAGE INFERENCE TIME FOR DIFFERENT METHODS WITH BATCH SIZE 1 AND SAME IMAGE SIZE OF 768 768 PIX. AND DECEPTION RATE

Method	Time	Deception rate
CycleGAN [10]	0.7sec	0.49
Sanakoyeu [13]	0.7sec	0.77
(Ours)	0.7sec	0.51

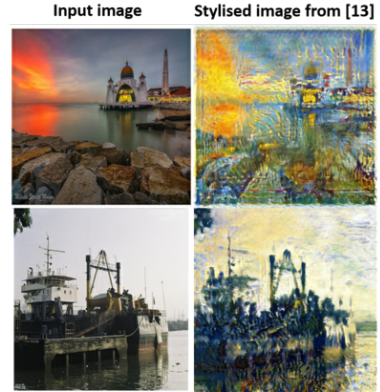


Fig. 11. Examples of stylised images which deform the content, generated by method of [13]

Besides the qualitative results, quantitative results are also presented in Table II. Table II shows the execution time required for the stylisation of a single image. The execution times are for images of size 768x768 pixels and are measured on the same NVIDIA GeForce RTX 2080TI. It is observed that all methods have approximately identical execution time. In the case of original CycleGAN and the proposed method the only difference is on the training of the model and since we have the same architecture it is obvious that the time will be the same. Style transfer deception rate [13] is calculated as the fraction of generated images which were classified by the VGG16 network pre-trained to Places dataset, as the artworks of an artist for which the stylisation was produced. For every method we generated 1502 stylisations (2 styles, 751 per style). Although in Table II the Sanakoyeu et al. [13] approach significantly outperforms both CycleGAN and our proposed methodology, in terms of deception rate, it is evident

that this method in its attempt to transfer style, deforms the content, as it is clearly illustrated in Fig. 11

VI. CONCLUSION AND FUTURE WORK

In the present study, a method for style transfer based on CycleGANs and FABEMD is proposed. A new formulation of the cycle consistency loss is introduced by estimating the loss through the BIMFs from the decomposition of content-to-content and style-to-style images. The experiments reveal that the proposed method produces better qualitative and quantitative results than the original CycleGAN method and the method of Sanakoyeu et al. [13]. Saliency maps and Deception rates are included in order to prove the effectiveness of the proposed method.

Although our method can achieve competitive results in style transfer, it is important to note that the introduction of frequency components (BIMFs) in the training process of a CycleGAN, in the estimation of the cycle consistency loss is a more qualitative expression that could be further examined, in other applications, beyond style transfer.

ACKNOWLEDGMENT

This work was partially supported by the European Commission under contracts H2020-779962 V4Design and H2020-825079 MindSpaces.

REFERENCES

- [1] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," *arXiv preprint arXiv:1508.06576*, 2015.
- [2] J. S. De Bonet, "Multiresolution sampling procedure for analysis and synthesis of texture images," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, 1997, pp. 361–368.
- [3] L.-Y. Wei and M. Levoy, "Fast texture synthesis using tree-structured vector quantization," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 2000, pp. 479–488.
- [4] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, "Simultaneous structure and texture image inpainting," *IEEE transactions on image processing*, vol. 12, no. 8, pp. 882–889, 2003.
- [5] A. A. Efros and W. T. Freeman, "Image quilting for texture synthesis and transfer," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 2001, pp. 341–346.
- [6] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1501–1510.
- [7] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6924–6932.
- [8] L. Sheng, Z. Lin, J. Shao, and X. Wang, "Avatar-net: Multi-scale zero-shot style transfer by feature decoration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8242–8250.
- [9] Z. Xu, M. Wilber, C. Fang, A. Hertzmann, and H. Jin, "Learning from multi-domain artistic images for arbitrary style transfer," *arXiv preprint arXiv:1805.09987*, 2018.
- [10] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.
- [11] X. Chen, C. Xu, X. Yang, L. Song, and D. Tao, "Gated-gan: Adversarial gated networks for multi-collection style transfer," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 546–560, 2018.
- [12] 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.
- [13] A. Sanakoyeu, D. Kotovenko, S. Lang, and B. Ommer, "A style-aware content loss for real-time hd style transfer," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 698–714.
- [14] J. C. Nunes, S. Guyot, and E. Deléchelle, "Texture analysis based on local analysis of the bidimensional empirical mode decomposition," *Machine Vision and applications*, vol. 16, no. 3, pp. 177–188, 2005.
- [15] J. C. Nunes, Y. Bouaoune, E. Delechelle, O. Niang, and P. Bunel, "Image analysis by bidimensional empirical mode decomposition," *Image and vision computing*, vol. 21, no. 12, pp. 1019–1026, 2003.
- [16] S. M. Bhuiyan, R. R. Adhami, and J. F. Khan, "A novel approach of fast and adaptive bidimensional empirical mode decomposition," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008, pp. 1313–1316.
- [17] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.
- [18] H. Liu, P. N. Michelini, and D. Zhu, "Artsy-gan: A style transfer system with improved quality, diversity and performance," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 79–84.
- [19] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
- [20] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 5907–5915.
- [21] T. Zhou, P. Krahenbuhl, M. Aubry, Q. Huang, and A. A. Efros, "Learning dense correspondence via 3d-guided cycle consistency," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 117–126.
- [22] C. Godard, O. Mac Aodha, and G. J. Brostow, "Unsupervised monocular depth estimation with left-right consistency," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 270–279.
- [23] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- [24] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [25] M. U. Ahmed and D. P. Mandic, "Image fusion based on fast and adaptive bidimensional empirical mode decomposition," in *2010 13th International Conference on Information Fusion*. IEEE, 2010, pp. 1–6.
- [26] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *2007 IEEE Conference on computer vision and pattern recognition*. Ieee, 2007, pp. 1–8.
- [27] K. Fan, "Maximum properties and inequalities for the eigenvalues of completely continuous operators," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 37, no. 11, p. 760, 1951.