Pattern Recognition Letters

Authorship Confirmation

Please save a copy of this file, complete and upload as the "Confirmation of Authorship" file.

As corresponding author I, Elissavet Batziou, hereby confirm on behalf of all authors that:

1. This manuscript, or a large part of it, has not been published, was not, and is not being submitted to any other journal.
2. If presented at or submitted to or published at a conference(s), the conference(s) is (are) identified and substantial justification for re-publication is presented below. A copy of conference paper(s) is(are) uploaded with the manuscript.
3. If the manuscript appears as a preprint anywhere on the web, e.g. arXiv, etc., it is identified below. The preprint should include a statement that the paper is under consideration at Pattern Recognition Letters.
4. All text and graphics, except for those marked with sources, are original works of the authors, and all necessary permissions for publication were secured prior to submission of the manuscript.
5. All authors each made a significant contribution to the research reported and have read and approved the submitted manuscript.

Signature Elissavet Batziou Date 11/10/2021

List any pre-prints:

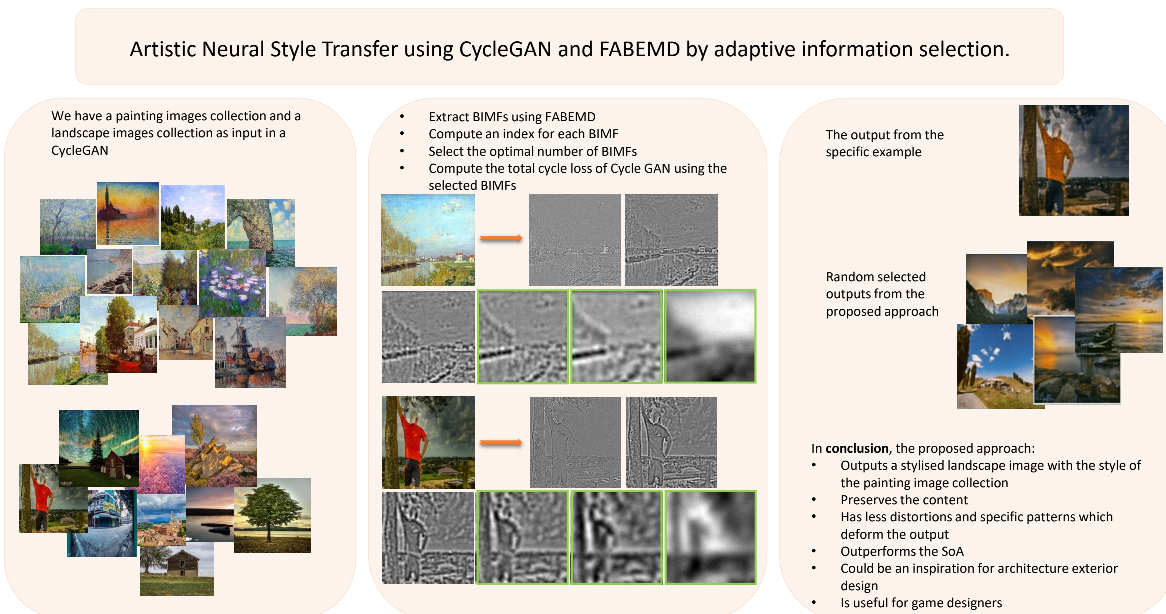
Relevant Conference publication(s) (submitted, accepted, or published):

Justification for re-publication:

Graphical Abstract (Optional)

Artistic Neural Style Transfer using CycleGAN and FABEMD by adaptive information selection.

Elissavet Batziou, Konstantinos Ioannidis, Ioannis Patras, Stefanos Vrochidis, Ioannis Kompatsiaris



Graphical Abstract

Research Highlights (Required)

- Neural Style Transfer from a painting images collection using CycleGAN and adaptive selection of spectral information.
- Analytic framework creation for the CycleGAN using FABEMD and the selection of the optimal number of BIMFs.
- Evalution of the proposed method in a qualitative and a quantitative way.

Pattern Recognition Letters
journal homepage: www.elsevier.com

Artistic Neural Style Transfer using CycleGAN and FABEMD by adaptive information selection

Elissavet Batziou[a,b,**], Konstantinos Ioannidis[b], Ioannis Patras[a], Stefanos Vrochidis[b], Ioannis Kompatsiaris[b]

[a] *School of Electronic Engineering and Computer Science, Queen Mary University of London (QMUL), Mile End Road, London, E1 4NS, United Kingdom*
[b] *Information Technologies Institute, Centre for Research and Technology (CERTH), 6th Km Charilaou-Thermi Road, Thessaloniki, 57001, Greece*

ABSTRACT

Neural Style Transfer (NST) comprises a class of computer vision methods that manipulate digital images to reformulate the visual content of one input image adopting visual features of an another image set. Artistic NST is the particular case of NST where the visual features are extracted from painting images. The combination of Cycle-Consistent Adversarial Networks (CycleGANs) with Fast and Adaptive Bidimensional Empirical Mode Decomposition (FABEMD) is proposed to adopt the specific artist's style on images effectively, where the cycle-consistency loss is modified to incorporate texture information by estimating the corresponding Bidimensional Intrinsic Mode Functions (BIMFs). An adaptive approach for identifying the optimal BIMF number that must be considered in order to manipulate the required amount of the involved texture, is proposed. For this purpose, the computation of a metric is considered for each BIMF to characterise the texture of each image or major intensity alterations at local scale. Experimental results reveal that adaptive comixture of texture features comprises an efficient approach in such artistic applications. Qualitative and quantitative results demonstrate that the proposed framework outperforms state-of-the-art (SoA) methods.
Keywords: Neural style transfer, FABEMD, cycle-consistency loss, CycleGAN.

1. Introduction

Artistic Neural Style Transfer (NST) has gained a lot of attraction by the research community the last decade in applications such as in game industry where new graphics and gaming interfaces can be designed using specific styles and artworks (e.g. Stadia[1]); in mobile application development to create artificial artwork from mobile photos (e.g. Maestro [1]); and in architecture to recommend new and automatic ways for designing interior or exterior environments, buildings etc [2]. Artistic style transfer is known as the process of generating an image that modifies the content of an input image (content) using the style of another or the collection of other related images (style). At early times, artistic style transfer relied on statistical methods to extract and recreate textures, patterns and local characteristics of an image [3, 4, 5]. Nowadays, Artistic NST adopts deep learning architectures to produce a stylised image by separating and recombining image content and style. Furthermore, transferring artistic style (e.g. color features, patterns, brushes) from a collection of images instead of a single image is a more complex challenge since it involves the representation of images with proper features which could describe the style of an art school (impressionism, cubism, etc.) or of a creator. However, in style transfer from a collection of images there might be some special techniques that the artist used rarely, that are not transferred adequately. A collection of images that represents the artist's style is more generic. Van Gogh, for example, has created paintings which belong in several schools of art such as Expressionism, Impressionism, Cubism, etc. These schools of art vary a lot among each other, but each artist has its own characteristics like preferred colors, or paintbrushes. The aim of the proposed model is to find common characteristics from a given collection that represents the artist's style.

Research community has proposed multiple NST approaches

**Corresponding author: Tel.: +30 6936616202
e-mail: batziou.el@iti.gr , e.batziou@qmul.ac.uk (Elissavet Batziou), kioannid@iti.gr (Konstantinos Ioannidis), i.patras@qmul.ac.uk (Ioannis Patras), stefanos@iti.gr (Stefanos Vrochidis), ikom@iti.gr (Ioannis Kompatsiaris)

[1] https://stadia.dev/blog/behind-the-scenes-with-stadias-style-transfer-ml/

from painting images, where the main focus of researchers over the last years was to improve the quality of the generated images [6]. Towards this objective, a number of painting collections were used to improve the adaptation of a painter's style in NST. The imposed limitations are related to the lack of ground truth data for a direct comparison of the generated images with the expected images, for the validation and training process. Additionally, in the case of style transfer from a single style image, only one input style image is required, which cannot necessarily be representative of the art school or the creator. Moreover, neural style transfer in painting images is a highly-subjective task, as it is affected by the viewer's perception for style. The generated images are mainly evaluated in a qualitative way, and only a few metrics can quantitatively evaluate the performance. Furthermore, the generation of a new stylised image usually deforms the patterns of the provided content image, lacking in the ability to preserve local patterns and features.

Aiming at reducing the impact of the aforementioned limitations, the presented novel framework estimates adaptively the level of the spectral information that is multiplexed for a CycleGAN towards improving the quality of the stylised images while preserving the essential patterns of the input content image. Cycle-consistency loss comprises one of the key elements in a CycleGAN architecture and requires the consistency between the original and the reconstructed image which is measured bidirectionally [7]. In addition, for a quantitative evaluation, a performance measure based on the trace norm is introduced, exploiting saliency maps in the context of style transfer. The usage of saliency maps in style transfer aims to quantify the content preservation.

Motivated by the CycleGAN framework in [8] and [9], the use of textural information has been proposed in [10], where only 3 experimentally defined sub-signals were incorporated. Current work expands [10], inserting a novel mechanism that defines the optimal BIMF number for involving the most essential spectral information to effectively multiplex texture information from the context image with low-level features of the style image(s). Experimental results and a qualitative comparison through a questionnaire, reveal that the presented approach comprises an efficient and robust alternative in the artistic NST as the estimated quantitative and qualitative results are considered comparable and beyond SoA methods. The core contributions can be summarized as follows:

- An adaptive selection mechanism is proposed to involve different amounts of spectral information by exploiting FABEMD.
- Two information metrics, entropy and edgeness, are considered following either a backward or a forward process for different amounts and types of information.
- The quantitative experimental comparison using the mean Trace norm scores is calculated over the entire dataset. The two-sample significance test on the trace norm values shows the superiority of the proposed method.
- The proposed method is also evaluated in a qualitative way using a questionnaire to incorporate the opinion of experts.

The remainder of the paper is structured as follows. Section 2 provides an overview of SoA works while Section 3 presents the proposed framework. Experimental results along with the configurations and the exploited datasets are presented in Section 4. Finally, Section 5 concludes the manuscript providing most major insights of the method and valuable outcomes.

2. Related Work

The first attempts in style transfer relied on parametric and non-parametric methods for texture synthesis and transfer [4, 11]. The transition between statistical and NST approaches occurred by the influential work of Gatys et al. [12], in which the authors utilise Deep Neural Networks (DNN) to encode both content and style of an image. The Gram matrices of the style image are calculated to capture the linear dependencies among several feature vectors, and subsequently, they are connected as convolutional layers of a DNN. It has also been demonstrated in [13] that the matching process using Gram matrices is equivalent to minimising the Maximum Mean Discrepancy with the second order polynomial kernel which captures more fine grained textures.

In [14] the authors presented an instance normalisation technique to improve the effectiveness of the generator in artistic NST. Comparing to batch normalisation, the latter can normalise each individual content image. In addition, Adaptive Instance Normalisation (AdaIN) [15] receives a content and a style image and aligns the channel-wise mean and variance of content to match the corresponding mean and variance of the style image, following an encoder-decoder architecture. A generalisation of AdaIN and whitening and coloring transformation [16] uses Zero-phase Component Analysis to project features into the same space, in order to extract the features of the stylised image and use them as input to a decoder network. Furthermore, an adversarial learning feed-forward network [17] uses a generator and a discriminator is conditioned on the domain categories, to distinguish the generated images from the same style class.

In [18], a style loss function is based on the combination of local and global measures. Local style retains the style while global loss function preserves the structural details. In a similar approach, the Patch Permutation GAN (P^2-GAN) [19] network is trained for one stroke style while patch permutation randomly divides the style image into multiple patches and a patch discriminator processes both patch-wise and natural images. In addition, cross distribution feature matching was adopted in [20] to perform the desire style transfer from one single style image. The authors adapt the high-order Central Moment Discrepancy (CMD) to minimize the difference between the target style and the feature distribution of the output image. However, computing high-order statistics explicitly introduces computational overhead. Similarly, another NST method from one single style image is introduced in [21]. The authors propose reversible neural flows and an unbiased feature transfer module that supports both forward and backward inferences and operates to formulate a projection-transfer-reversion pipeline.

Contrary to above approaches, transferring the style from an image collection comprises a more challenging task as the framework is trained to discriminate common features of one image collection. The work proposed in [22] relies on an adversarial network that is consisted of three modules: an encoder, a gated transformer, and a decoder. Gated-GAN is trained for multiple styles to generate new stylised images through weighted connections among the branches of the gated transformer. On the contrary, CycleGAN [8] not only generates stylised images from an image collection, but it also learns an inverse transformation that generates an identical-to-the-original content image. Sanakoyeu et al. [23] adopted an encoder-decoder network architecture and a fixed-point loss that ensures stylisation has reached a fixed-point after one feed-forward pass. The cycle-consistency loss has been replaced by the perception loss of [24], introducing a new objective function for diversity in image-to-image translation [25]. The exploitation of a perception loss benefits the training process in measuring image similarities more robustly than per-pixel losses [24].

The aforementioned techniques do not consider the spectral information or patterns during the styling process which comprises a significant aspect in improving the artistic NST performance. Recently, in [10], the cycle-consistency loss function of CycleGAN has been modified by involving spectral and texture information. A fixed number of BIMFs are extracted for each content and style image, which sequentially are introduced in the computation of the cycle consistency loss of CycleGAN. Here, aiming at identifying adaptively the amount of the required contextual information, the proposed framework introduces a novel cycle-consistency loss computation and an adaptive mechanism for defining the essential texture information that should be considered in the NST process.

3. Methodology

The proposed approach modifies the cycle-consistency loss of a CycleGAN to involve texture information and patterns in the training and the validation process, relying on FABEMD. Hence, the necessary background is presented for each component of the proposed framework (Fig. 1). In brief, the latter consists of two core elements, an original CycleGan architecture with updated loss function calculation and the FABEMD-based module that estimates the required texture information that should be processed by the first component.

The network involves two generators F and G and two discriminators D_X and D_Y, as in [9]. The first generator G receives as input a landscape image and creates a painting image visually relevant with the defined artistic style. The second generator F creates landscape photos given painting images. As illustrated in Fig. 1, the representation size is reduced in the encoder phase, remains constant in the transformer phase, and expands in the decoder stage. There are two bilateral comparisons in which cycle-consistency loss is calculated and update the generator models in each training iteration. The discriminator architecture learns the relationship between one output of the model to the number of pixels in the input image, named as PatchGAN model (Fig. 1).

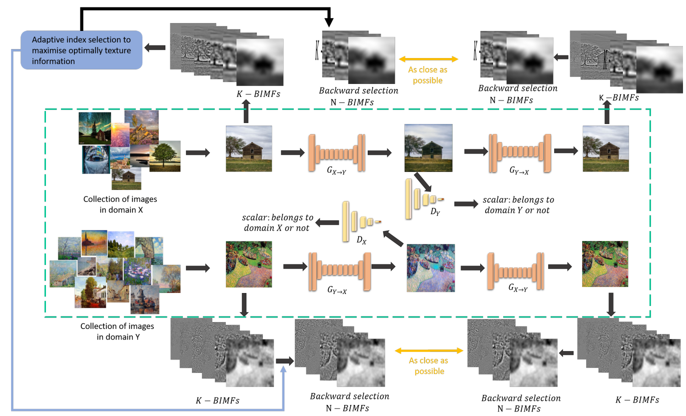


Fig. 1. Proposed style transfer framework with CycleGAN and optimal number of BIMFs.

3.1. The core CycleGAN

The following section presents the most essential fundamentals for the basic CycleGAN framework. CycleGAN defines two mapping functions $G : X \to Y$ and $F : Y \to X$, 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 model considers two adversarial discriminators D_X and D_Y, where D_X discriminates images $\{x\}$ from stylised representations $\{F(y)\}$, while D_Y aims to discriminate $\{y\}$ from $\{G(x)\}$, respectively.

Adversarial losses are applied to both mapping functions [26]. For function $G : X \to Y$ and its discriminator D_Y, the objective is to minimise the following loss:

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

where, G generates images $G(x)$ that are similar to images from Y , while D_Y discriminates stylised samples from $G(x)$ and real samples y. Accordingly, the mapping function $F : Y \to X$ and its discriminator D_X, denoted by $\mathcal{L}_{GAN}(F, D_X, Y, X)$, discriminates the reconstructed images from $F(y)$ and the real samples.

Cycle consistency loss introduces a two-way loss computation between the generated image and the original image, measuring the "consistency" with its inverse mapping from the generated to the original image. Forward and backward cycle consistency are defined as follows: for each image x of domain X, the image stylisation cycle should recover x to the original representation, i.e. $x \to G(x) \to F(G(x)) \approx x$. Similarly, for each image y from domain Y, G and F should also satisfy backward cycle consistency, as it is defined by: $y \to F(y) \to G(F(y)) \approx y$. The forward and backward cycle consistency define the cycle consistency loss, as follows in Eq. (2):

$$\begin{aligned}\mathcal{L}_{cyc}(G, F) = E_{x\sim P_{data}(x)}[||F(G(x)) - x||_1]+ \\ +E_{y\sim P_{data}(y)}[||G(F(y)) - y||_1]\end{aligned} \tag{2}$$

The total loss of the style transfer model with cycle-consistent adversarial networks is defined as the sum 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:

$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y)+ \\ +\mathcal{L}_{GAN}(F, D_X, Y, X) + \lambda\mathcal{L}_{cyc}(G, F)\end{aligned} \tag{3}$$

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

$$G^*, F^* = argmin_{G,F} max_{D_X,D_Y} \mathcal{L}(G, F, D_X, D_Y) \quad (4)$$

where G^*, F^* are the optimal mapping functions.

3.2. Texture extraction using FABEMD

Empirical Mode Decomposition (EMD) analyses one-dimensional signals based on their local characteristics (extrema) and provides a spectral representation of nonlinear and non-stationary data into a finite number of Intrinsic Mode Functions (IMFs) [27]. Each IMF is determined by the total number of extrema and the mean signal as it is extracted by the upper and the lower envelope, while the iteration process of IMF definition (sifting process) is terminated when predefined stopping criteria are fulfilled. Thus, the characterisation of an IMF strictly relies on its local features and components rendering the approach as an adaptive method of texture extraction.

The extension of EMD to bidimensional arrays is known as bidimensional EMD (BEMD) and applies the same processing rationale. One of the limitations of BEMD is the calculation time, as it involves the use of 2D bicubic spline interpolation for the calculation of the two envelopes. To overcome this limitation, BEMD has been updated to a fast and adaptive algorithm (FABEMD) [28] to reduce the computational complexity.

Texture features are commonly refer to luminosity values, hence, the optimum BIMF amount is defined via luminance adopting a neighboring method for the extrema identification [29]. A pixel is considered as a local extrema, if its luminosity displays a superior or inferior value compared to its adjacent values within a square-shaped window of specific size.

The images are decomposed into l number of sub-components (BIMFs) and a residual sub-component which process is completely reversible and the initial image can be recovered as: $A = B_1 + B_2 + ... + B_K + R$ where A is the original image, R is the residual and $B_k, k = 1, 2, \ldots, K$ represent the extracted BIMFs. As k increases, each BIMF includes specific spectral components from higher to lower frequencies. High frequency components that corresponds to texture features are involved in the lower-indexed BIMFs while the residual corresponds to the the data trend [30]. The texture information that each BIMF depicts varies from one input image to another, therefore it is necessary to define a process that does not get an arbitrary selection of a fixed number of BIMFs [10].

3.3. Cycle consistent loss function with spectral features

The following sub-section presents and analyses an adaptive approach to define the required number of BIMFs for the loss function definition, as shown in Fig. 1, where the optimal selection $\mathcal{K}_{x_i}$ from the spectral decomposition of content and style images is used to define the latter. The estimated texture features are considered in the loss function definition, for the optimisation of the Generators' G and F and the Discriminators' D_X and D_Y performance. $x_i^B(k)$ is denoted as the k-th BIMF corresponding to the content image x_i and by $r_i^B(k)$ the k-th BIMF of the reconstructed image $r_i = F(G(x_i))$. Similarly, let $y_j^B(k)$ be the k-th BIMF corresponding to the style image y_j and by $s_j^B(k)$ the k-th BIMF of the stylised image $s_j = G(F(y_j))$. In addition, let $\mathcal{I}(x_i^B(k))$ be the metric which quantifies the amount of information in each BIMF $x_i^B(k)$. It has been stated in [31] that the function value of Shannon's entropy indicates how much information on the texture from the original image remains in each sub-signal (e.g. BIMFs). Entropy is a popular concept in information theory [32], previously used to quantify the amount of information of a transmitted message within a discrete set of probabilities p_i and in our context it is defined as:

$$H(A^B(k)) = -\sum_{i=1}^{g} p_i log(p_i) \quad (5)$$

where g corresponds to the number of luminosity levels and p_i is obtained from the frequency of each grayscale level in the k-th BIMF of image A, whether it refers to content image x_i or style image y_j.

Similarly, edgeness quantifies the arrangement of intensities in a region, and serves as an image texture metric to determine patterns [33]. Texture complexity has an impact on the amount of information in the corresponding region of each extracted BIMF. Given a block of $M \times N$ pixels, the gradient-based edge detector is applied to calculate the gradient magnitude $\nabla(B_k(a, b))$ which then defines the edgeness of BIMF B_k per unit area (a, b):

$$\mathcal{E}(B_k) = |\{B_k(a, b) : \nabla B_k(a, b) > t\}|/(M \times N) \quad (6)$$

for some parameter t, $\mathcal{E}(B_k)$ is the edgeness of the k-th BIMF of image A. Once a measure of information is defined, to maintain the number of components that preserve a certain fraction of the content image, the minimum number d of BIMFs $k = 1, 2, \ldots, d, \ldots, K$ is obtained as follows:

$$\sum_{k=d}^{K} \mathcal{I}(B_k))/\sum_{k=1}^{K} \mathcal{I}(B_k)) > w \quad (7)$$

where K is the total number of extracted BIMFs and $w \in \{50.0\%, 68.2\%, 95.4\%, 99.7\%\}$ motivated from the Empirical Rule. The value of 0.997 results to a special case where all available BIMFs are used from the spectral decomposition, and, in both cases, the model coincides with the original CycleGAN. However, the goal of the proposed approach is to translate patterns into styles considering frequency components, for a better quality in the generated stylised images. This optimal selection of information index and corresponding threshold is extensively studied in Section 4, since the use of the original content and style image results to the original CycleGAN model [8]. Eq. (7) refers to the backward selection of BIMFs from $K, K-1, \ldots$ until d. Similarly, forward selection is defined as follows:

$$\sum_{k=1}^{d} \mathcal{I}(B_k))/\sum_{k=1}^{K} \mathcal{I}(B_k)) > w \quad (8)$$

Due to the imbalanced information distribution of an image among spectral components, the proposed approach is bidirectional. After the definition of d, the selected set of BIMFs for the image A is denoted as $\mathcal{K}_A = \{B_k, k \in \mathcal{S}\}$ where, $\mathcal{S}$ is obtained either by Eq. (7) for the backward selection or Eq. (8) for the forward selection.

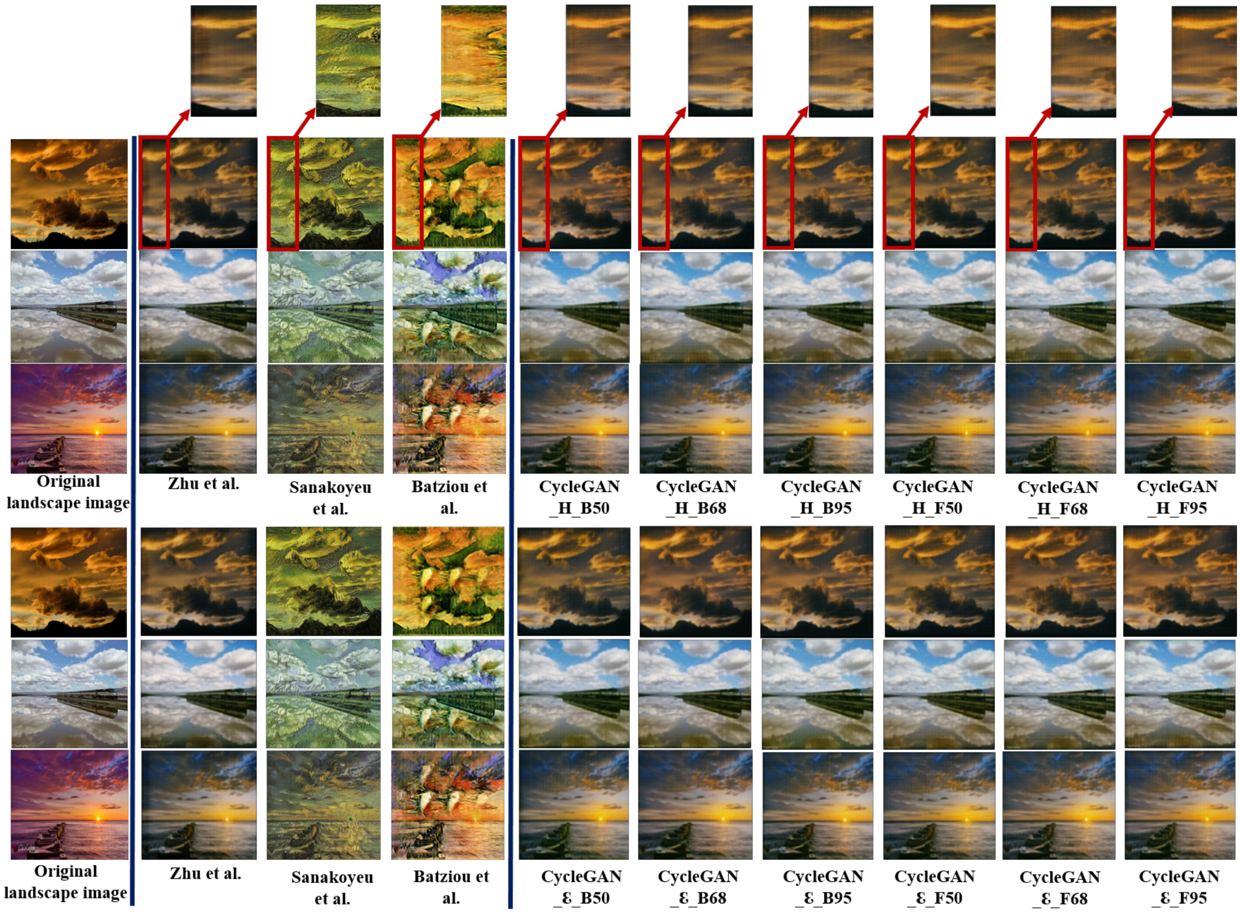


Fig. 2. Comparison of the proposed approach with the CycleGAN model[8], Sanakoyeu et al., [23] and Batziou et al., [10] on the "vangogh2photo" dataset.

The cycle consistency loss is determined as follows:

$$\begin{aligned}\mathcal{L}_{cyc}(G,F) &= E_{x\sim P_{data}(x)} \textstyle\sum_{k\in\mathcal{K}}[\|r^B(k)-x^B(k)\|_1] \\ &+E_{y\sim P_{data}(y)} \textstyle\sum_{k\in\mathcal{K}}[\|s^B(k)-y^B(k)\|_1]\end{aligned} \tag{9}$$

where the backward and forward cycle consistency is computed over $k \in \mathcal{K}$ meaning adaptively defined contrary to its definition in [10] where the value was experimentally defined.

4. Experimental Results

The following section provides insights on the exploited datasets, the configurations adopted for the conducted experiments as well as the qualitative and quantitative validation of the corresponding results in comparison with SoA methods. For the conducted experiments, each image was divided into 5 BIMFs plus the residual, using a 3×3 window [28] for the extrema definition. Each time a set of BIMFs is defined using one of the two metrics, entropy and edgeness by following either a backward or a forward process. For the experiments the NVIDIA GeForce RTX 2060 SUPER was used.

4.1. Datasets

The appropriate image collection from the TensorFlow catalogue[2] was utilised to accomplish an effective performance comparison on the Van Gogh and Monet styles. Monet is an artist who has left a mark in the landscaping art and Van Gogh's artworks are included in diversified movements such as Impressionism and Expressionism, so the datasets selection covers a wide range of different painting styles. The "monet2photo" dataset is comprised by 1074 Monet's painting images and 6853 landscape photos. Likewise, the "vangogh2photo" dataset contains 401 painting images of Vincent Van Gogh and the same landscape photos as the "monet2photo" dataset.

4.2. Training process

The training process was initiated by feeding the model with the style images, content images and the selected BIMFs for each input image. For both datasets, the weight for cycle consistency loss λ was set equal to 10 and the number of epochs is set to 200 as in [8]. Generators and discriminators are all optimised with an Adam solver (default parameters), the same learning rate (2e-4) and the default instance normalisation. The BIMF index was estimated to identify the involved sub-signals d based on Eq. (7) and Eq. (8). Then, the cycle consistency loss of Eq. (9) is computed using the optimal BIMF number.

4.3. Results

The baseline methods of the presented comparison are the CycleGAN [8], the approach of Sanakoyeu et al. [23] and the

[2] `https://www.tensorflow.org/datasets/catalog`

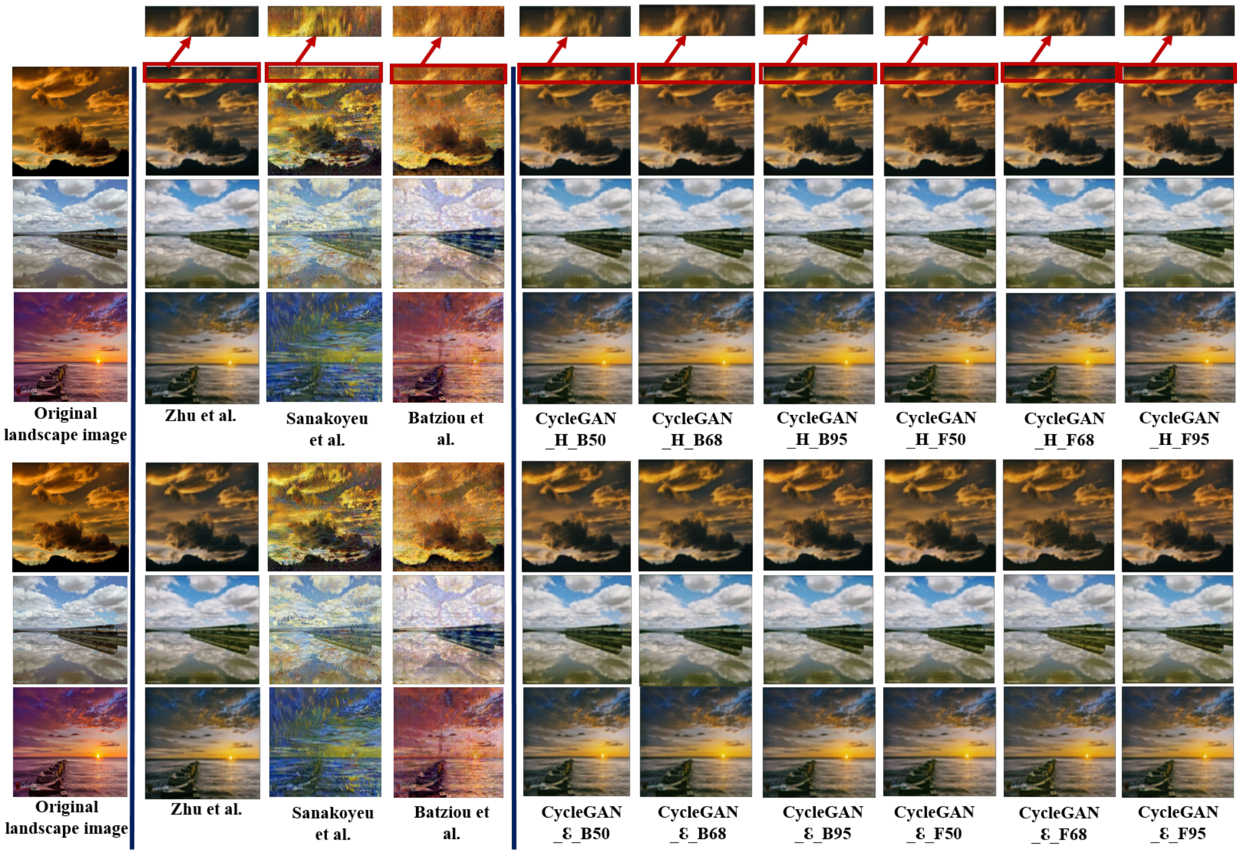


Fig. 3. Comparison of the proposed approach with the CycleGAN model[8], Sanakoyeu et al., [23] and Batziou et al., [10] on the "monet2photo" dataset.

Table 1. Mean Trace norm scores of the saliency map differences.

mean	[8]	[23]	[10]	[21]	[20]	CG_$\mathcal{H}$_B50	CG_$\mathcal{H}$_B68	CG_$\mathcal{H}$_B95	CG_$\mathcal{H}$_F50	CG_$\mathcal{H}$_F68	CG_$\mathcal{H}$_F95	CG_$\mathcal{E}$_B50	CG_$\mathcal{E}$_B68	CG_$\mathcal{E}$_B95	CG_$\mathcal{E}$_F50	CG_$\mathcal{E}$_F68	CG_$\mathcal{E}$_F95
Monet	65905	79754	76119	522719	231021	65972	**65735**	66275	66331	66108	66184	66062	65912	66094	65856	65763	66167
VanGogh	66334	91063	91642	368815	301465	66405	**65669**	66323	66095	66333	66681	66458	66120	66029	65700	66004	66157

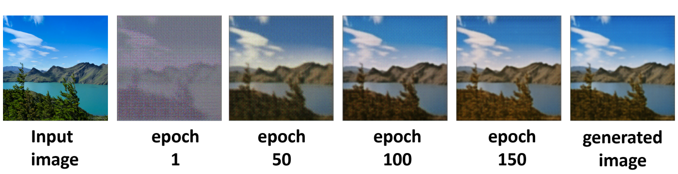


Fig. 4. The progress of the stylisation process using the proposed method.

recently introduced approach of [10], due to their relevance to artistic NST from a painting image collection. The proposed approach denoted in Fig. 3 and 2 as 'CycleGAN_$\mathcal{I}$_$\mathcal{S}$ w' where, in the $\mathcal{I}$ field, H is noted for entropy or $\mathcal{E}$ for edgeness. In the $\mathcal{S}$ field "B" and "F" correspond to the backward and forward BIMF selection, respectively. We denote by w the upper information bound (50.0%, 68.2%, 95.4%).

From the Figs. 2, 3 and Table 1, it is derived that the best-performing proposed method is the one with backward ($\mathcal{S} = B$) Entropy selection ($\mathcal{I} = H$) at $w = 68.2\%$. Thus, the best performance was obtained by the setup that corresponds to the "CycleGAN_H_B68" or simply to "CG_H_B68". From the first rows of Fig. 2 and 3, it can be observed that this approach does not generate noisy patterns in the sky regions, an effect that appears in the resulted images of the other SoA methods. In addition, the proposed method eliminates blurry artifacts which indicatively appear in the image regions of other SoA methods. Moreover, the images in the last row depict either shadow effects or noisy patterns at their left side. By contrast, in the images generated by the proposed method, these artifacts are minimised. Concerning a qualitative comparison, the the conducted results of each version of the proposed method are similar meaning that the general method can adequately transfer the artist's style and retain the content details. A qualitative analysis (Figs. 2 and 3) reveals that the proposed method transfers the style with no deformations. By contrast, [8] generates aesthetically appealing images and well preserving content structures but distortions are observed in each image (Fig. 2). Involving spectral information via the extracted BIMFs, the focus is concentrated on pure texture elements and patterns of high intensity from the input source images. The inclusion of noisy signals in the consistency loss computation is avoided by the use of spectral information. The methods of [23, 10] generate images which have noisy patterns and deform the content. Although we observe that color and texture information in the proposed method is similarly transferred, as in the CycleGAN [8], the content is well-preserved in our proposed method compared to [8]. Qualitatively, as we observe in Fig. 2 and Fig. 3, the CycleGAN [8] and the other methods produce blurry artifacts in comparison with the different versions of the proposed approach. We have provided a zoomed patch for all methods on the same image so as to be able to directly compare the artifacts from one method to another. Additional images for a qualitative comparison are included in Appendix A of the supplementary material.

In Fig. 4 the progress of the stylisation process is presented using the proposed method for an indicative "monet2photo" image. During the first epochs the style is not sufficiently transferred. In the middle of the process, some low level style features have already been transported. Close to the predefined number of epochs, the stylised image is obtained to minimise the noisy parts due to a proper texture transfer. The inference time for the stylisation of a single image of size 768x768 pixels is similar for all considered methods (0.7 sec), so the modification in the consistency loss does not affect the inference process. For the proposed method, the FABEMD algorithm was used in order to produce more effective stylised images at the expense of computational complexity. Based on [34] the computational complexity of EMD is $O(n \log n)$ where n is the data length. BEMD [29] is applied on two-dimensional signals n^2 so the complexity becomes $O(n^2 \log n^2) = O(n^2 \log n)$, where n is the image width. FABEMD [35] has been proposed to decrease the complexity levels compared to the basic BEMD. More specific, FABEMD envelope estimation process using order-statistics filters is almost independent of the image and the texture patterns in terms of complexity. Moreover, FABEMD substitutes the data interpolation step of BEMD by a direct envelope estimation method and limits them into one iteration. Spatial domain sliding order-statistics filters are utilized to obtain the maxima and minima envelopes. The application of a smoothing operator results in the final upper and lower envelopes respectively. The order-statistics filters size is adaptively derived from the available information of maxima and minima maps meaning it is depended on the input image. Although the FABEMD algorithm decrease the complexity compared to the BEMD, it is hard to estimate the computational complexity of this process phase accurately since each image has different local extrema.

Table 2. Results of the questionnaire

Criterion	[8]	[23]	[10]	CG_H_B68
More aesthetically appealing	6.7%	**46.7%**	13.3%	33.3%
Artist's style	13.3%	**40%**	6.7%	**40%**
Content preservation	33.3%	0.0%	6.7%	**60%**
Less distortions	33.3%	0.0%	6.7%	**60%**
Game appropriateness	26.7%	33.3%	0.0%	**40%**
Exterior design	13.3%	6.7%	0.0%	**80%**

Within the scope of a qualitative evaluation, a questionnaire was developed to directly validate the performance of the methods. It has been completed by 15 experts (architects and game designers) over questions relative to aesthetic appealing and content preservation. The ultimate goal of the 15 experts is to utilize the proposed outputs as inputs of either 3D reconstruction in architecture or game environment creation as close to the expected style, but without compromising the content. For the qualitative evaluation we have used the CG_H_B68 version of the proposed method since this gave the best quantitative results (Table 1). Higher average scores correspond to less artifacts, best content preservation and better representation of the artist's artistic style following the scale defined by the experts. As it is presented in Table 2 the proposed approach displays the highest score in 5 of 6 criteria while it result the second higher for the 6th criterion.

The results are presented in Table 2 and it can be observed that the experts confirmed that the proposed method generates images which retain the artist's style, preserve the content sufficiently and are more suitable for both architects and game designers. Based on the conducted experiments and the corresponding results, the proposed method outperforms the other under evaluation methods. More results and questionnaire details are shown in Appendix C of the supplementary material.

Table 3. Two sample t-test on mean scores of trace norm

p-value	Monet	VanGogh
[8]	.812	.322
[23]	< .001	< .001
[10]	< .001	< .001
[21]	< .001	< .001
[20]	< .001	< .001

In NST evaluation, one of the main characteristics to be validated is content preservation. Saliency detection relies on image features and statistics to localise the most interesting regions of an image. Saliency maps function as suitable features for evaluating the content preservation of style transfer outputs. A small value of saliency map differences between the stylised image and the input content one indicates that the most interesting regions coincide. The method in [36] was adopted to implement saliency map detection. The saliency maps of Fig. 2 and Fig. 3 are extracted towards a direct comparison between the saliency map of the original landscape image (SM^{or}) and of the corresponding stylised image (SM^{st}), the Trace norm of the matrix difference is used:

$$\sum_i \sigma_i = \|(\sqrt{(SM^{or} - SM^{st})^*(SM^{or} - SM^{st})})\|_* \tag{10}$$

where, σ_i is the i-th singular value of the matrix $SM^{or} - SM^{st}$.

In addition to the qualitative comparison of Fig. 2 and Fig. 3, a quantitative evaluation has also been performed and is provided in Table 1. The mean scores correspond to the stylised images of the whole test set in Monet and Van Gogh datasets. The minimum trace norm value for each image corresponds to the output of the proposed framework, which implies that our generated images are closer to the input landscape image than the ones generated by other methods in terms of content preservation. Furthermore, statistical significance t-tests were performed to examine the hypothesis of the average decrease in mean trace norm between the baselines and the best proposed approach (Table 1). In Table 3 the p-values of the two sample t-test are provided, and as it can be observed, the decrease in mean trace norm between the best proposed approach and [23], [10], [21] and [20] is statistically significant. Although, the decrease between the best proposed approach and [8] is not statistically significant, our approach extracts more qualitative results. Additionally to the aforementioned comparison, the proposed approach is also compared with the two state-of-the-art methods of An et al., (2021) [21] and Kalischek et al., (2021) [20] that address style transfer problem in a different way. The results of the quantitative evaluation are presented in Table 1 and Table 3 and the results of the qualitative evaluation are presented in Appendix B of the supplementary material.

5. Conclusion

In this work a novel framework for artistic NST was proposed utilising an image collection. The presented style transfer framework goes beyond a color transfer, finding common characteristics from a given collection that represents the artist's style. The style is transferred effectively while it preserves the content sufficiently. Any modification in cycle-consistency loss of CycleGAN architecture includes spectral information about the input images (content and style) in an adaptive way, where the number of spectral components (BIMFs) is selected in a bidirectional way. The spectral components refer to the texture information contained in a source image, as they are quantified by FABEMD. The cycle consistency loss is updated accordingly towards preserving the content characteristics of the input image while transferring the style effectively. A comparison with other artistic SoA style transfer methods have been conducted to showcase the effectiveness and robustness of the proposed method. Experiments have shown that the proposed approach generates less distortions in the created stylised image and is closer to the original landscape image.

References

[1] M. Reimann, M. Klingbeil, S. Pasewaldt, A. Semmo, M. Trapp, J. Döllner, Maestro: A mobile app for style transfer orchestration using neural networks, in: Int. Conf. on Cyberworlds, 2018, pp. 9–16. doi:10.1109/CW.2018.00016.

[2] H. Zhang, E. Blasetti, 3d architectural form style transfer through machine learning (2020).

[3] J. S. De Bonet, Multiresolution sampling procedure for analysis and synthesis of texture images, in: Proc. of 24th conf. Comput. graphics & interact techn, 1997, pp. 361–368.

[4] L.-Y. Wei, M. Levoy, Fast texture synthesis using tree-structured vector quantization, in: Proc. 27th conf. Comput graphics & interact techn, 2000, pp. 479–488.

[5] M. Bertalmio, L. Vese, G. Sapiro, S. Osher, Simultaneous structure and texture image inpainting, IEEE trans on imag proc 12 (8) (2003) 882–889.

[6] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, M. Song, Neural style transfer: A review, IEEE trans on visualiz and comput graph 26 (11) (2019) 3365.

[7] Y. Zhao, R. Wu, H. Dong, Unpaired image-to-image translation using adversarial consistency loss, in: Proc. ECCV, Spr., 2020, pp. 800–815.

[8] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proc. CVPR, 2017, pp. 2223–2232.

[9] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros, Image-to-image translation with conditional adversarial networks, in: Proc. CVPR, 2017, pp. 1125–1134.

[10] E. Batziou, P. Alvanitopoulos, K. Ioannidis, I. Patras, S. Vrochidis, I. Kompatsiaris, Cycle-consistent adversarial networks and fast adaptive bi-dimensional empirical mode decomposition for style transfer, in: Proc. 25th ICPR, IEEE, 2021, pp. 2360–2367.

[11] H. Zhao, X. Jin, J. Shen, F. Wei, Real-time photo style transfer, in: 2009 11th IEEE Int Conf Computer-Aided Design and Comput Graphics, IEEE, 2009, pp. 140–145.

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

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

[14] D. Ulyanov, A. Vedaldi, V. Lempitsky, Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis, in: Proc. IEEE Conf. CVPR, 2017, pp. 6924–6932.

[15] X. Huang, S. Belongie, Arbitrary style transfer in real-time with adaptive instance normalization, in: Proc. CVPR, 2017, pp. 1501–1510.

[16] L. Sheng, Z. Lin, J. Shao, X. Wang, Avatar-net: Multi-scale zero-shot style transfer by feature decoration, in: Proc. CVPR, 2018, p. 8242.

[17] Z. Xu, M. Wilber, C. Fang, A. Hertzmann, H. Jin, Learning from multi-domain artistic images for arbitrary style transfer, arXiv preprint arXiv:1805.09987 (2018).

[18] H.-H. Zhao, P. L. Rosin, Y.-K. Lai, M.-G. Lin, Q.-Y. Liu, Image neural style transfer with global and local optimization fusion, IEEE Access 7 (2019) 85573–85580.

[19] Z. Zheng, J. Liu, P2-gan: Efficient style transfer using single style image, arXiv preprint arXiv:2001.07466 (2020).

[20] N. Kalischek, J. D. Wegner, K. Schindler, In the light of feature distributions: moment matching for neural style transfer, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 9382–9391.

[21] J. An, S. Huang, Y. Song, D. Dou, W. Liu, J. Luo, Artflow: Unbiased image style transfer via reversible neural flows, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 862–871.

[22] X. Chen, C. Xu, X. Yang, L. Song, D. Tao, Gated-gan: Adversarial gated networks for multi-collection style transfer, IEEE Transactions on Image Processing 28 (2) (2018) 546–560.

[23] A. Sanakoyeu, D. Kotovenko, S. Lang, B. Ommer, A style-aware content loss for real-time hd style transfer, in: Proc. ECCV, 2018, pp. 698–714.

[24] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in: Proc. ECCV, Springer, 2016, p. 694.

[25] H. Liu, P. N. Michelini, D. Zhu, Artsy-gan: A style transfer system with improved quality, diversity and performance, in: Proc. 24th ICPR, IEEE, 2018, pp. 79–84.

[26] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in neural information processing systems, 2014, pp. 2672–2680.

[27] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, H. H. Liu, The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. Royal Soc. London A: math., phys & eng sci 454 (1998) 903–995.

[28] S. M. Bhuiyan, R. R. Adhami, J. F. Khan, A novel approach of fast and adaptive bidimensional empirical mode decomposition, in: 2008 IEEE Int Conf Acoust, Speech & Signal Process, IEEE, 2008, pp. 1313–1316.

[29] J. C. Nunes, Y. Bouaoune, E. Delechelle, O. Niang, P. Bunel, Image analysis by bidimensional empirical mode decomposition, Image and vision computing 21 (12) (2003) 1019–1026.

[30] J. Pan, Y. Tang, Texture classification based on bidimensional empirical mode decomposition and local binary pattern, Int. J. Adv. Comput. Sci. Appl. 4 (9) (2013).

[31] P. J. Navarro, C. Fernández-Isla, P. M. Alcover, J. Suardíaz, Defect detection in textures through the use of entropy as a means for automatically selecting the wavelet decomposition level, Sensors 16 (8) (2016) 1178.

[32] T. M. Cover, Elements of information theory, John Wiley & Sons, 1999.

[33] T. Butz, J.-P. Thiran, Affine registration with feature space mutual information, in: Int. Conf. on Medical Imag Comput and Computer-Assisted Intervention, Springer, 2001, pp. 549–556.

[34] Y.-H. Wang, C.-H. Yeh, H.-W. V. Young, K. Hu, M.-T. Lo, On the computational complexity of the empirical mode decomposition algorithm, Physica A: Statistical Mechanics and its Applications 400 (2014) 159–167.

[35] S. M. Bhuiyan, R. R. Adhami, J. F. Khan, Fast and adaptive bidimensional empirical mode decomposition using order-statistics filter based envelope estimation, EURASIP Journal on Advances in Signal Processing 2008 (2008) 1–18.

[36] X. Hou, L. Zhang, Saliency detection: A spectral residual approach, in: Proc. CVPR, Ieee, 2007, pp. 1–8.