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

CERTH
CENTRE FOR RESEARCH & TECHNOLOGY HELLAS

Queen Mary
University of London

Information
Technologies
Institute

MKLab
Multimedia Knowledge and Social Media Analytics Laboratory

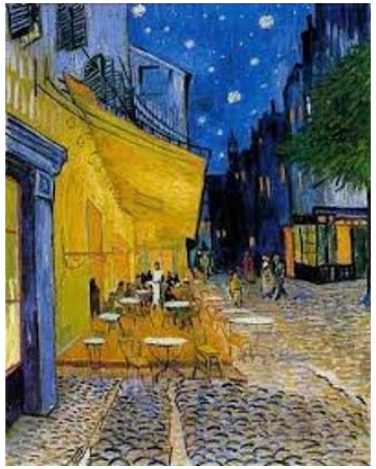
MAD
MultiMoDal Data Fusion and Analytics Group

OUTLINE

- Introduction
  - Style transfer
  - Problem statement
  - Aims and Objectives
- Background in CycleGAN and Bidimensional EMD
- Proposed methodology
- Evaluation
  - Qualitative Evaluation
  - Quantitative Evaluation
- Conclusion

INTRODUCTION – STYLE TRANSFER

- Style transfer is the process of generating an image that combines the style of an input image and the content from a different one.
- Application domains:
  - Gaming industry: design in an automatic way new graphics and gaming interfaces
  - Mobile application development: create artificial artwork from mobile photos
  - Architecture: recommend new and automatic ways for designing interior or exteriors objects and buildings.

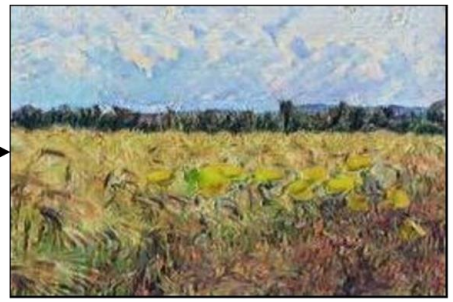


PROBLEM STATEMENT – DEFINITION

- **Goal**: Transform images based on the features of specific paintings using style transfer from a collection of paintings
- Challenges
  - Complex challenge when style transfer is not based on a single image
  - Extraction of common features that characterise the whole collection of images and not the style of a single image.
- Aims and Objectives
  - Preserve the content and style characteristics by removing the noise of the style and the content image.
  - Involve BEMD to preserve the more informative parts of the content image and style elements.
  - Quantify the deformations between the original content image and the generated one, and in parallel to preserve the style from a collection of style images.

Photograph

Vincent Van Gogh collection



Photograph with Vincent Van Gogh style

RELATED WORK

- Zhu et al., (2017). CycleGAN unpaired image-to-image translation
  - extend the architecture of a Generative Adversarial Network (GAN) to their Cycle-Consistent Adversarial Network architecture
  - learn an inverse generator that is able to create an image identical to the original content image, starting from the stylised one.
- Chen et al., (2018). 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.
  - Gated-GANs are trained for multiple styles in order to generate new stylised images through weighted connections between the branches of the gated transformer.
- Liu et al. (2018) Artsy-GAN
  - replace the Cycle-Consistency loss with the so-called "perception loss" (Johnson et al., 2016)
  - 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 colour channels.
- Sanakoyeu et al. (2018) encoder-decoder network architecture
  - fixed point loss that ensures stylisation has converged and reached a fixed point after one feed-forward pass.
  - This style-aware content loss forces the stylisation to take place in the decoder.

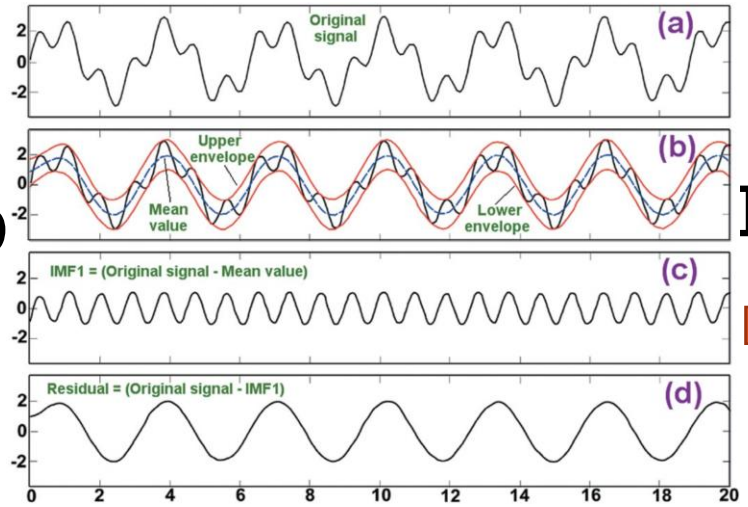
OUR CONTRIBUTION

- We extend the CycleGAN architecture (Zhu et al.,2017) by integrating the FABEMD component. The representation of content and style images through a set of frequency elements is introduced and examined in the cycle-consistency loss
- We include a Cycle consistency loss function which penalizes incorrect reconstruction of the input image from the stylised one (content-to-content and style-to-style generation) (Chen et al., 2018). .
- The perception loss (Liu et al., 2018) is calculated based mainly on colour 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.
- We keep the cycle consistency loss and combine it with frequency elements so as to achieve high-quality stylized images that also preserve the content patterns, contrary to the style-aware content loss (Sanakoyeu et al., 2018).

BACKGROUND – FAST AND ADAPTIVE BIDIMENSIONAL EMPIRICAL MODE DECOMPOSITION (FABEMD)

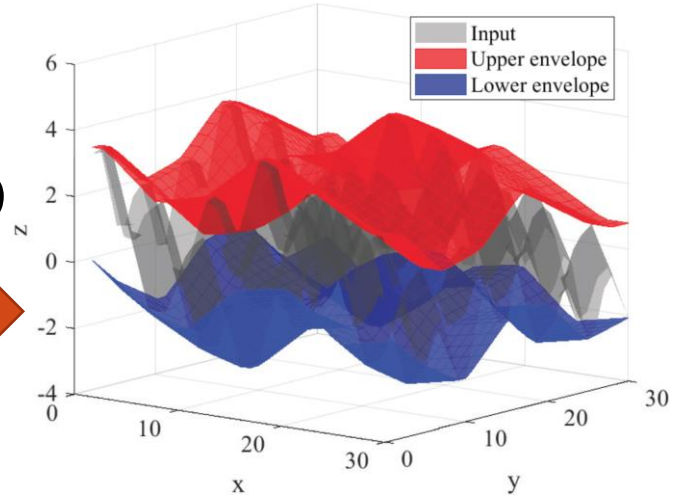
- Empirical mode decomposition (EMD) decomposes signals into a set of spectral components, known as Intrinsic Mode Functions (IMFs).
- BEMD decomposes a 2D array into Bidimensional IMFs (BIMFs).
- FABEMD is a fast method to extract BIMFs, using a sliding window when searching for local minima and maxima, in the creation of an envelope.

EMD



Source: https://towardsdatascience.com/

BEMD

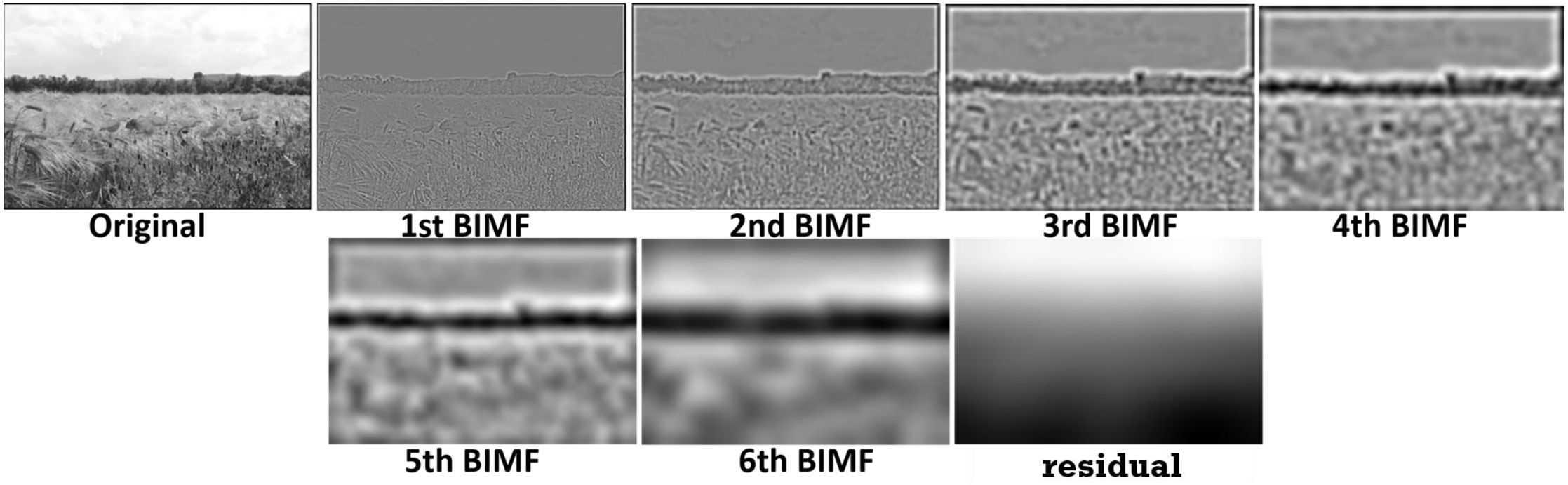


Source: https://www.mathworks.com/matlabcentral/fileexchange/71270-fast-and-adaptive-multivariate-and-multidimensional-emd

Sliding window for fast and adaptive computation of local maxima and minima in the envelope extraction

FABEMD

A LANDSCAPE IMAGE AND ITS EXTRACTED BIMFS



- The FABEMD output using as input a landscape image.
- From higher frequencies (strong edges) to lower ones (smooth edges)
- Higher enumerated BIMFs and the residual trend do not depict any image like content
- BIMFs are not images but distinct sub-signals linearly transformed for presentation purposes

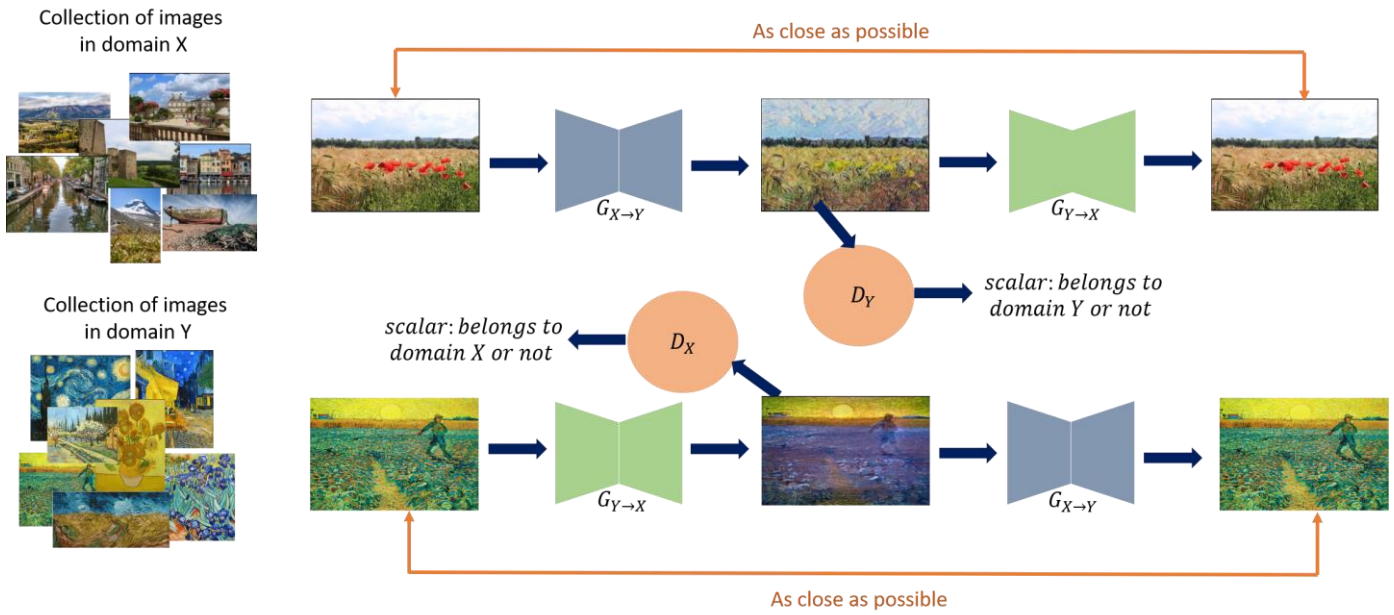
BACKGROUND – CYCLEGAN

- Cycle-Consistent Adversarial Networks learn mapping functions between two domains X and Y
- The model introduces two adversarial discriminators D_X and D_Y, where D_X distinguishes between images $\{x\}$ and stylized images $\{F(y)\}$ and D_Y aims to discriminate between $\{y\}$ and $\{G(x)\}$.
- **Adversarial Loss** minimisation for matching the distribution of generated images to the data distribution in the target domain (style)

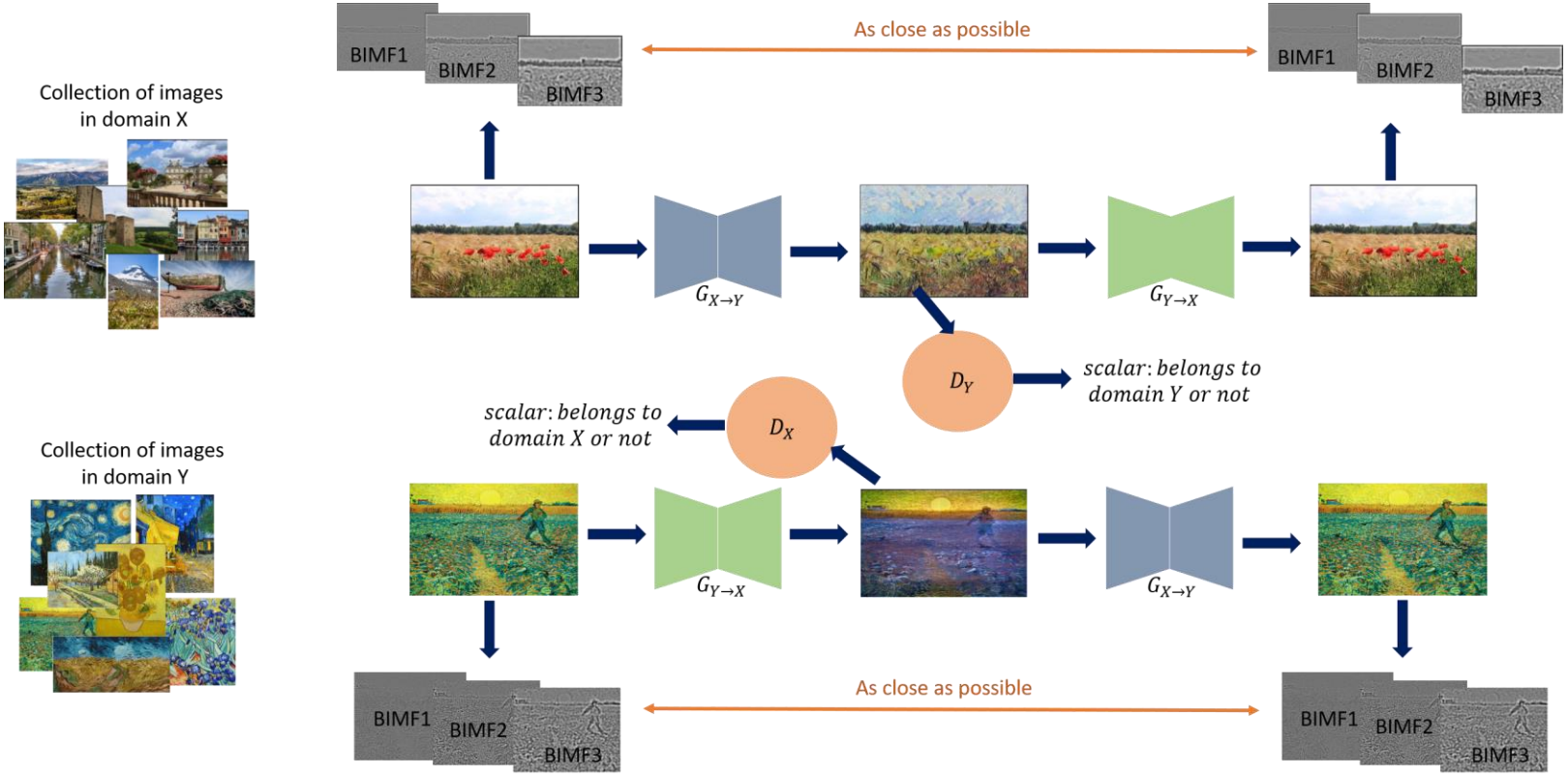
$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = E_{y \sim P_{data}(y)}[log D_Y(y)] + E_{x \sim P_{data}(x)}[log(1 - D_Y(G(x))]$$

- **Cycle Consistency Loss** minimisation to prevent the learned mappings G and F from contradicting each other

$$\mathcal{L}_{\mathrm{cyc}}(\mathrm{G}, \mathrm{F}) = \mathrm{E}_{\mathrm{x} \sim P_{data}(x)}\left[\left\|\mathrm{F}(\mathrm{G}(\mathrm{x})) - \mathrm{x}\right\|_1\right] + \mathrm{E}_{\mathrm{y} \sim P_{data}(y)}\left[\left\|\mathrm{G}(\mathrm{F}(\mathrm{y})) - \mathrm{y}\right\|_1\right]$$



PROPOSED FRAMEWORK



Summation over the top-k BIMF-to-BIMF comparisons

- The network contains two generators F and G and two discriminators D_X and D_Y
- 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.
- Cycle Consistency loss is computed between sets of BIMFs instead of the corresponding original images

PROPOSED APPROACH

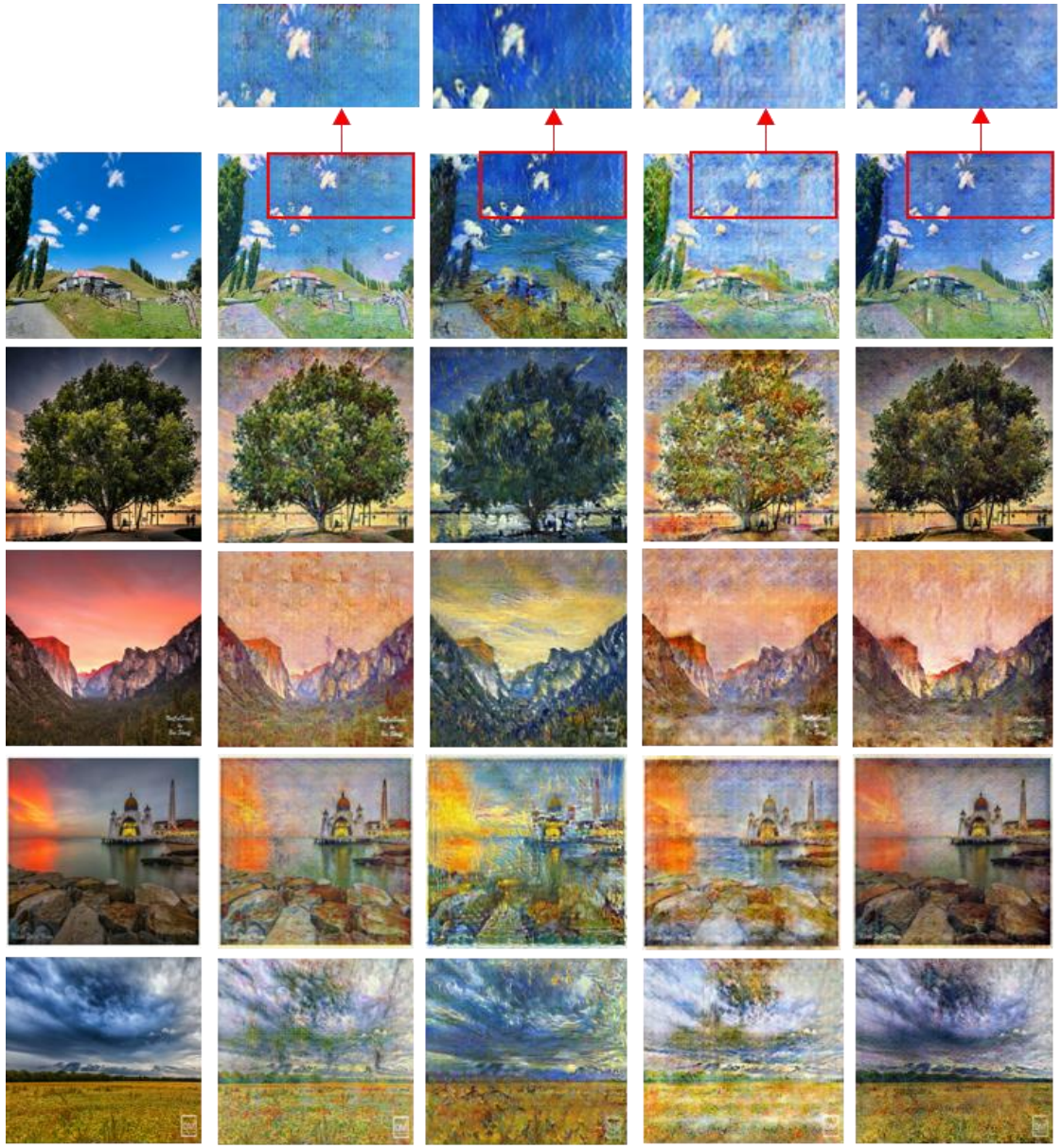
- $x_i^{bimf}(k)$: the k-th BIMF corresponding to the content image x_i
- $r_i^{bimf}(k)$: the k-th BIMF of the reconstructed image $r_i = \mathrm{F}(\mathrm{G}(x_i))$
- $y_j^{bimf}(k)$: the k-th BIMF corresponding to the style image y_j
- $s_j^{bimf}(k)$: the k-th BIMF of the stylized image $s_j = G\left(F(y_j)\right)$
- **Adversarial Loss** same as in CycleGAN
- **Cycle Consistency Loss** is defined as:

$$\mathcal{L}_{\mathrm{cyc}}(\mathrm{G},\mathrm{F}) =$$

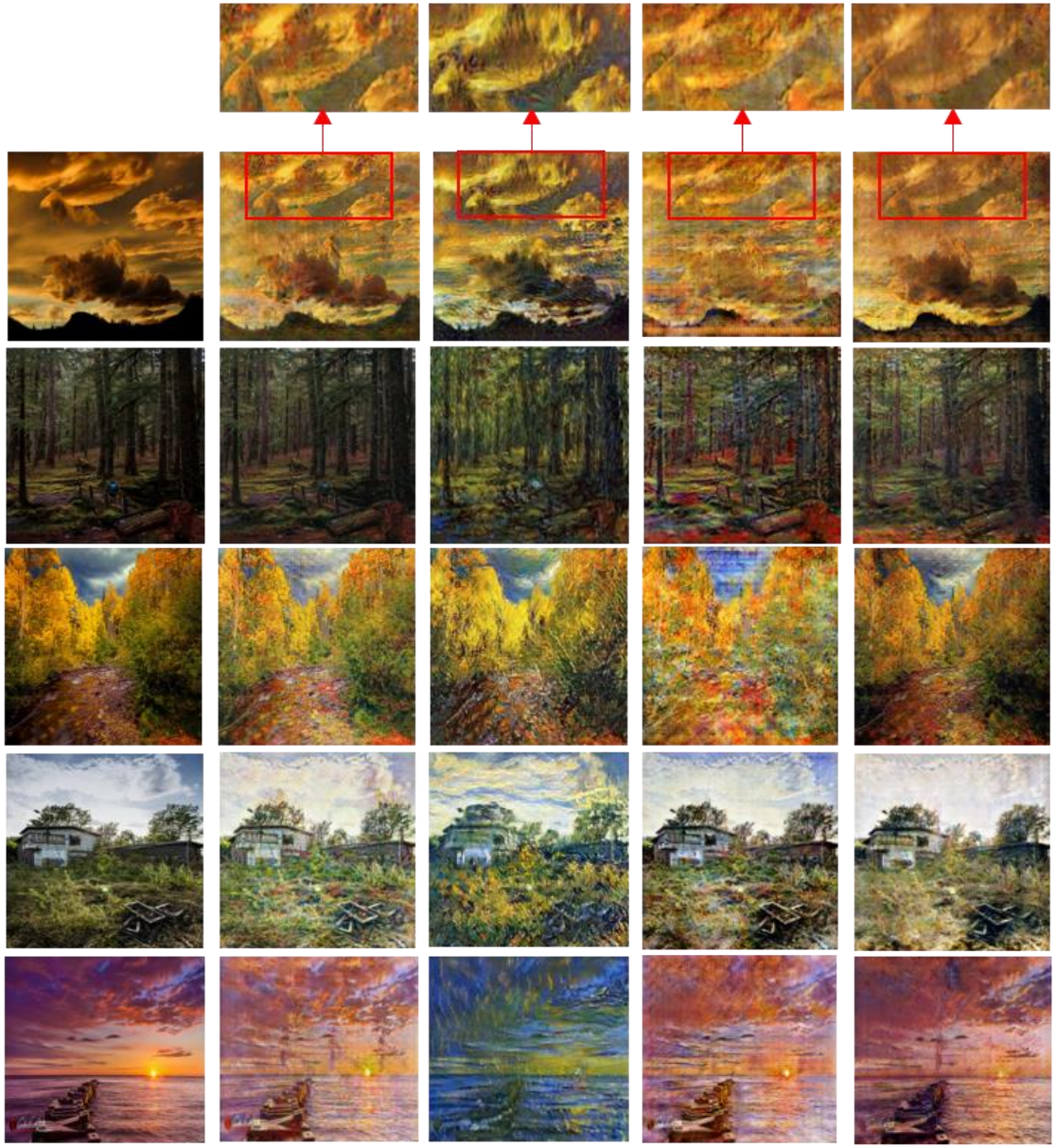
$$= \mathrm{E}_{x\sim P_{data}(x)} \sum_{k=1}^{K} \left[\underbrace{\left\| r^{bimf}(k) - x^{bimf}(k) \right\|_1}_{\text{difference between the BIMFs of the content image and the BIMFs of the reconstructed content}} \right] + \mathrm{E}_{y\sim P_{data}(y)} \sum_{k=1}^{K} \left[\underbrace{\left\| s^{bimf}(k) - y^{bimf}(k) \right\|_1}_{\text{difference between the BIMFs of the style image and the BIMFs of the reconstructed style}} \right]$$

EVALUATION

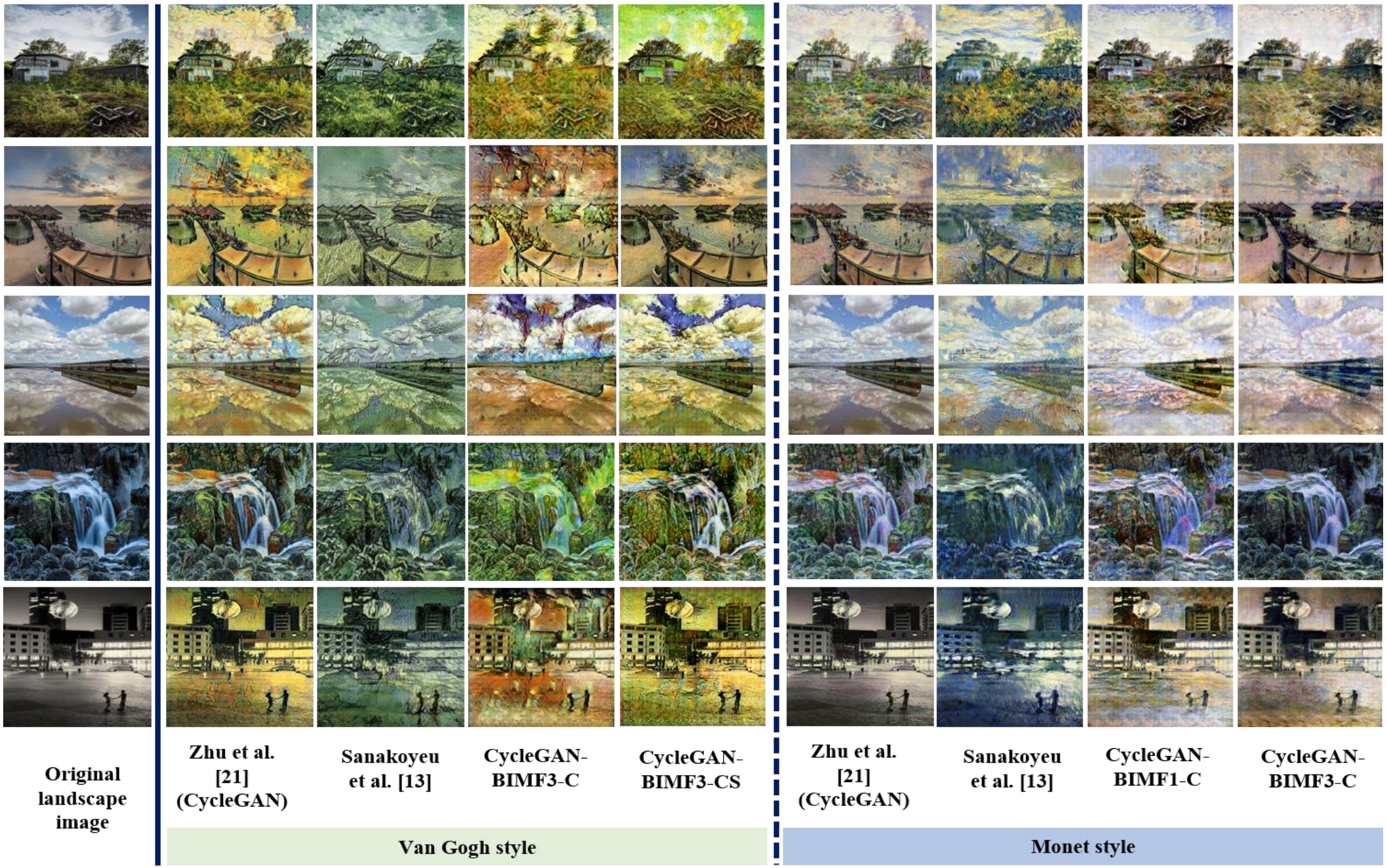
- Two benchmark datasets from the TensorFlow catalogue (https://www.tensorflow.org/datasets/catalog/cycle_gan) are considered
  - "monet2photo": It comprises 1074 painting images of Monet and 6853 photos.
  - "vangogh2photo": It contains 401 painting images of Vincent Van Gogh and the same landscape photos as in the first dataset.
- Comparison with the following approaches that use collection of style images:
  - **Zhu et al., (2018) – CycleGAN:** extend the architecture of a GAN to their Cycle-Consistent Adversarial Network architecture. It learns an inverse generator that is able to create an image identical to the original content image, starting from the stylised one.
  - **Sanakoyeu et al., (2018)** encoder-decoder network architecture: fixed point loss that ensures stylisation has converged and reached a fixed point after one feed-forward pass. This style-aware content loss forces the stylisation to take place in the decoder.
- Qualitative comparison in two different styles, and saliency maps
- Quantitative comparison:
  - deception rate
  - time performance
  - trace distance between saliency maps of the original content images and the stylised ones



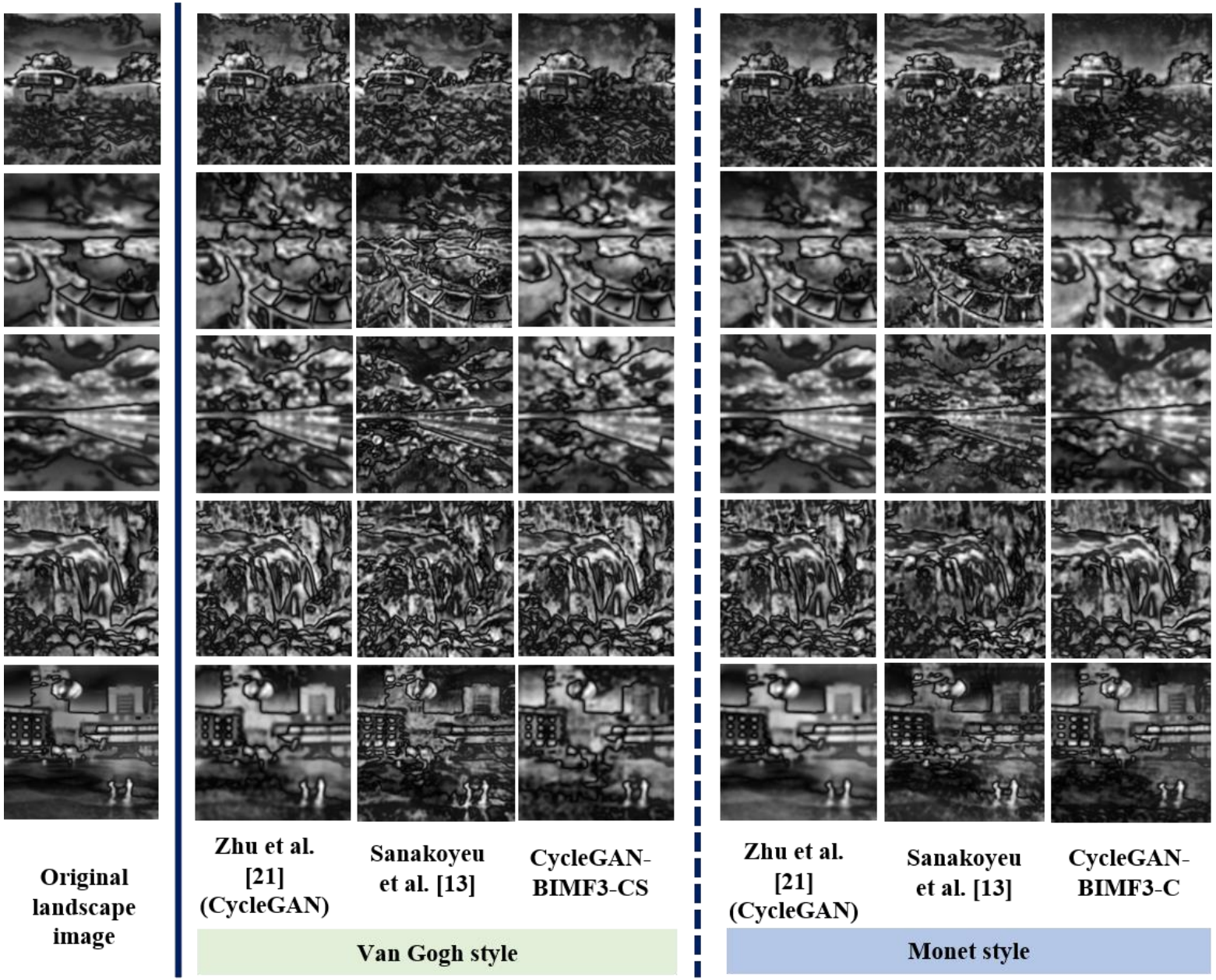
Original landscape image | Zhu et al. [21] (CycleGAN) | Sanakoyeu et al. [1] | CycleGAN -BIMF1-C | CycleGAN-BIMF3-C



Original landscape image | Zhu et al. [21] (CycleGAN) | Sanakoyeu et al. [1] | CycleGAN -BIMF1-C | CycleGAN-BIMF3-C


Original landscape image
Zhu et al. [21] (CycleGAN)
Sanakoyeu et al. [13]
CycleGAN-BIMF3-C
CycleGAN-BIMF3-CS
Zhu et al. [21] (CycleGAN)
Sanakoyeu et al. [13]
CycleGAN-BIMF1-C
CycleGAN-BIMF3-C
Van Gogh style
Monet style

SALIENCY MAPS AND THEIR TRACE DISTANCES



Zhu et al., 2018	Sanakoyeu et al., 2018	CycleGAN-BIMF3-CS
90171	91674	**80024**
76903	81565	**76531**
80210	74566	**74264**
95222	87048	**79269**
82599	87572	**76088**
88090	95426	**80963**
74492	85466	**73918**
97914	81509	**77637**
88359	**79749**	97649
71731	86364	82098

TIME PERFORMANCE

Method	Time	Deception rate
Zhu et al. (2017)	0.7 sec	0.49
Sanakoyeu et al. (2018)	0.7 sec	0.77
CycleGAN-BIMF3-CS (ours)	0.7 sec	0.51

Input image | Stylised image from [13]

Original image | 15 epochs | 20 epochs | 25 epochs | 30 epochs | 35 epochs | 40 epochs

CONCLUSIONS

- The estimation of cycle consistency loss through the BIMFs from the decomposition of content-to-content and style-to-style images generates less distorted outputs.
- The experiments reveal that the proposed method produces better qualitative and quantitative results than the State-of-the-art models.
- Saliency maps and deception rates proved the effectiveness of the presented method compared to other similar approaches.
- Distances between saliency maps as a new evaluation measure in style transfer.

THANK YOU FOR YOUR ATTENTION!

- V4Design - Visual and textual content re-purposing FOR(4) architecture, Design and video virtual reality games - H2020-779962
- https://v4design.eu/

MindSpaces

- MindSpaces - Art-driven adaptive outdoors and indoors design - H2020-825079
- https://mindspaces.eu/