Towards compact reversible image representations for neural style transfer

Xiyao Liu[1], Siyu Yang[1,5], Jian Zhang[1(✉)], Gerald Schaefer[2], Jiya Li[3], Xunli Fan[4], Songtao Wu[5], and Hui Fang[2(✉)]

[1] School of Computer Science and Engineering, Central South University, Changsha, China
[2] Department of Computer Science, Loughborough University, U.K.
[3] Hunan Embroidery Research Institute Co.,Ltd
[4] School of Information Science and Technology, Northwest University, Xi'an, China
[5] Sony R&D Center China

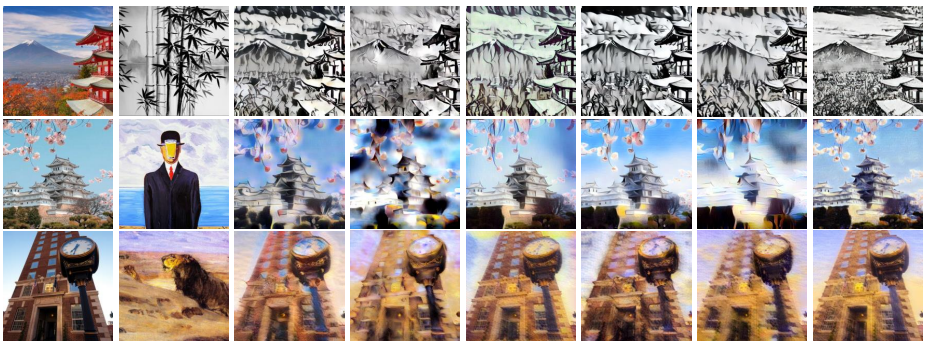


Fig. 1: Visual comparison of our proposed method to various state-of-the-art models. We can observe that our method effectively eliminates artefacts while adequately preserving structural content details, resulting in visually more pleasing results.

Abstract. Arbitrary neural style transfer aims to stylise a content image by referencing a provided style image. Despite various efforts to achieve both content preservation and style transferability, learning effective representations for this task remains challenging since the redundancy of content and style features leads to unpleasant image artefacts. In this paper, we learn compact neural representations for style transfer motivated from an information theoretical perspective. In particular, we enforce compressive representations across sequential modules of a reversible flow network in order to reduce feature redundancy without losing content preservation capability. We use a Barlow twins loss to reduce channel dependency and thus to provide better content expressiveness, and optimise the Jensen-Shannon divergence of style representations between reference and target images to avoid under- and

✉ Corresponding Authors: J.Zhang(jianzhang@csu.edu.cn) and H.Fang(H.Fang@lboro.ac.uk)

over-stylisation. We comprehensively demonstrate the effectiveness of our proposed method in comparison to other state-of-the-art style transfer approaches.

Keywords: neural style transfer · information theory · reversible generative model · feature redundancy

1 Introduction

Arbitrary neural style transfer (NST) aims to render a content image using a specific style from a reference image. In contrast to traditional optimisation-based NST methods such as [13, 14, 31], encoder-transfer-decoder architectures have gained popularity due to their efficiency for the task [18, 28, 29]. Classical methods based on this structure, e.g., AdaIN [17], WCT [24], or SANet [29], yield impressive performance by extracting effective representations to meet two criteria: content preservation and style transferability.

Various approaches try to improve these criteria for enhanced NST. For instance, in [9], a transformer model is used to explore long-range dependencies of image features to preserve content information with contexts. IEContraAST [7] introduces contrastive learning to capture more representative features by enforcing consistency between same content or style pairs. To address the content leakage problem, which refers to the unrecoverable content loss caused by the decoder, [2] designs a reversible neural flow model to ensure that the image content can be well preserved after multiple transfers. Meanwhile, other approaches focus on enhancing style transferability. LapStyle [26] presents a revision network to improve the transferability of local details, while [39] matches higher-order statistics to generalise the transfer across various domains.

Despite significant progress, it remains challenging to produce visually pleasing stylised images due to (*i*) over-rendering artefacts, causing the loss of content details, and (*ii*) under-transfer of low-level texture styles. Fig. 2 illustrates these effects using three state-of-the-art methods, IEContraAST [7], CAP-VSTNet [36] and ArtFlow [2]. As highlighted in [36], over-rendering artefacts are mainly caused by feature redundancy in the latent space. Theoretically, the NST encoder learns non-linear bases to extract image representations that enable the transfer between various styles via a linear projection. However, the intertwining representations across channels lead to artefacts when interpolating features of unseen contents and styles. Although CAP-VSTNet introduces a channel refinement module to reduce feature redundancy, it is still difficult to achieve compact and decoupled representations across channels without an explicit driven force. Under-stylisation is mainly due to the significant differences in style loss for different types of styles when training with traditional mean-variance style loss, which may be dominated by certain styles while neglecting others.

It is well known that information theory be can effectively used to reduce information redundancy [1] and to align data distributions [3]. Motivated by this, we seek a novel style transfer model, where, by minimising the mutual

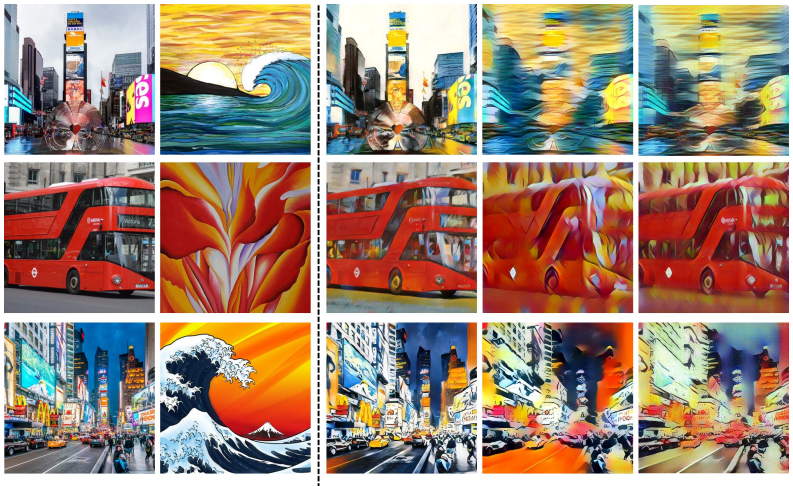


Fig. 2: Example results of IEContraAST, CAP-VSTNet and ArtFlow. We can observe that in (c) the style textures of the resulting images differ significantly from those of the style image, and that in (d,e) the images contain a significant amount of artefacts while fine image details are compromised.

information between input and output of network blocks as well as the correlation between feature channels, we can effectively reduce feature redundancy and avoid over-stylisation, whereas, by aligning style feature distributions in an information theoretic manner, we can overcome the under-stylisation issues caused by traditional statistical measures.

In this paper, we thus propose our InfoMinRev projection flow network to remove feature redundancy of the encoder without losing the expressive ability guaranteed by a reversible framework. We further reduce the channel dependency by applying a Barlow twins loss as a specific instantiating information theoretic objective, and employ a Jensen-Shannon divergence loss to ensure our learned features have expressive capability to align transferred style distributions to deal with under- and over-transfer effects. From an information theoretical perspective, these objectives can be unified to learn better representations for well-balanced style transfer. Our main contributions in this paper are:

- We introduce a novel reversible InfoMinRev module to a flow-based style transfer method, which employs mutual information minimisation losses between consecutive bijective transformation modules, enabling the extraction of compressed yet effective feature representations during the forward inference process.
- We use a Barlow Twins loss to alleviate channel-wise coupling issues, learning more compact and decoupled content-style representations so as to enhance content expressiveness while maintaining transferability, thus achieving a more robust content structure.

- We propose a new Jensen-Shannon divergent style loss with a more elegant distribution alignment objective to address the problem of insufficient style transfer.
- We demonstrate that each of the introduced components leads to a positive impact on style transfer. When combined, a superior balance between content and style is achieved, resulting in remarkable style transfer quality as illustrated in Fig. 1.

2 Related Work

2.1 Image Style Transfer

Since the seminal work in [13], neural style transfer has been an active research topic. Early work focusses on using feed-forward neural networks to replace the time consuming optimisation process for improved efficiency [6, 11, 23, 25, 34, 35]. However, these methods typically only support transfer across a fixed number of styles. For enhanced task generalisability, various methods for arbitrary style transfer have been proposed [16, 17, 24, 29, 32]. AdaIN [17] adjusts the feature statistics of the content image to align with the counterparts of the style image. Going one step further, WCT [24] employs whitening and colouring transforms to transfer styles on uncorrelated feature maps. More recently, feature enhancement strategies have been applied to improve the content and style representations for the task. SANet [29] introduces an attention mechanism to adjust local style patterns based on the representation of the content image. IEContraAST [7] adopts contrastive learning terms to enhance feature consistency of positive content/style pairs while increasing the feature distances of negative pairs. In [9], a transformer model is proposed to exploit image context to enhance its features.

2.2 Reversible Flow Networks

A reversible flow model was initially introduced in NICE [10] to learn a non-linear bijective transformation for tractability of exact latent inference in an unsupervised manner. Since then, its advantages, including efficiency, inversion capability, and tractable density, have lead to its application in a variety of tasks. For instance, RevNet [15] exploits memory efficiency and low computational overhead to train an image classifier, while Reformer [21] uses a reversible network to build a transformer model in order to significantly improve model efficiency without sacrificing model performance. In [20], a generative model, Glow, is proposed for efficient photo-realistic synthesis and semantic manipulation by adapting a flow model. Recently, reversible flow models have also been applied to the style transfer task. Motivated by its inversion capability, ArtFlow [2] designs a flow model to solve the content leakage problem, which is common in an encoder-transform-decoder architecture due to the reconstruction error caused by the decoder. Inspired by ArtFlow, CAP-VSTNet [36] introduce an extra channel refinement module and a matting Laplacian loss, while Cartoon-Flow [22] adapts a flow architecture in a more specialised cartoon stylisation task.

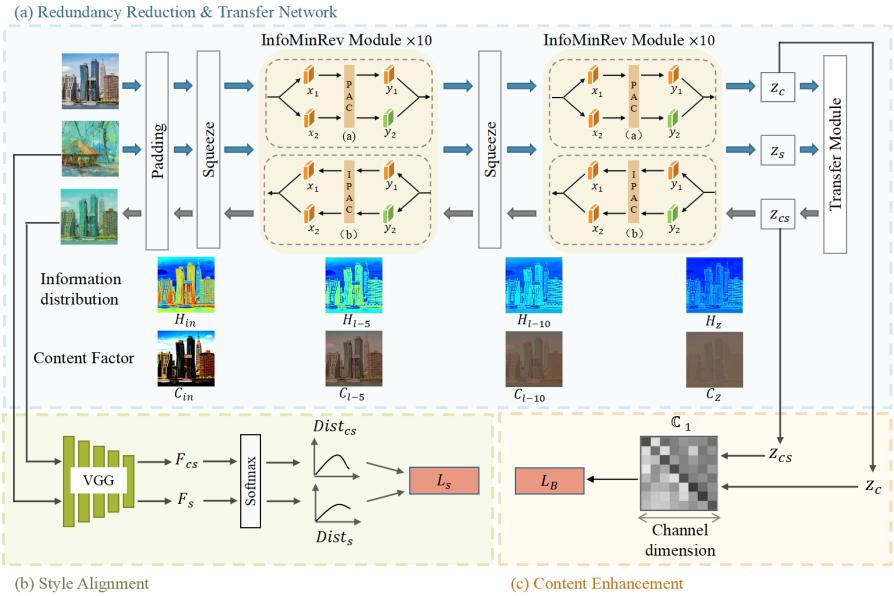


Fig. 3: Overview of our proposed structure, which comprises three parts: (a) style transfer and layer-wise redundancy reduction, (b) Jensen-Shannon divergence loss for style alignment, and (c) channel-wise redundancy reduction loss for content enhancement.

3 Proposed Method

An overview of our proposed flow model is illustrated in Fig. 3. In the following, after laying out our motivation and rationale, we introduce its three innovative components, a redundancy reduction-based flow network consisting of our proposed InfoMinRev modules, a channel-wise redundancy reduction loss for content enhancement, and a Jensen-Shannon divergence loss for style alignment. These designs improve our model by gaining concise and expressive feature representations to achieve an optimal balance of content preservation and stylisation.

3.1 Motivation and Rationale

Feature redundancy, a major barrier to effective style transfer, leads to image artefacts when an NST model is deployed on unseen content and style image pairs. To alleviate this problem, reduction of feature redundancy is an intuitive solution to achieve better style transfer. Information theory, commonly used to quantify the correlation and dependency between random variables, provides a powerful theoretical foundation to accomplish the task. Consequently, in our approach, we enforce mutual information minimisation losses across sequential

bijective transformation modules in order to obtain compressive image representations. However, without a reversible network architecture, this can weaken the expressiveness of the features and thus compromise the content preservation capability. We therefore design a novel reversible flow model, described in Section 3.2, by minimising the mutual information between its modules so as to learn compact yet expressive image representations.

Channel correlation reduction also contributes to compressive and expressive representations. When training reversible flow models in an unsupervised manner, e.g., NICE [10] and Glow [20], channel independence is enforced when learning the non-linear reversible transformation. However, this constraint has been relaxed in flow models for NST due to the unavoidable interactions between content and style features. In our model, we apply a soft constraint, explained in Section 3.3, to minimise the correlations across channels so as to enhance content expressiveness without damaging transferability via the linear projection based on feature statistics.

A typical style loss is designed to align low-level feature statistics, i.e., the mean and variance of feature maps, to transfer image styles. However, simply minimising these statistics differences leads to under- or over-stylisation due to heterogeneous styles from reference images. In our model, we therefore employ another information theoretical metric, namely the Jensen-Shannon divergence, to encourage feature learning which aligns style feature distributions to address this issue.

As illustrated in Fig. 3, our model is able to extract features with minimal information distribution in the latent space, while containing sufficient expressive capability to preserve the content structure.

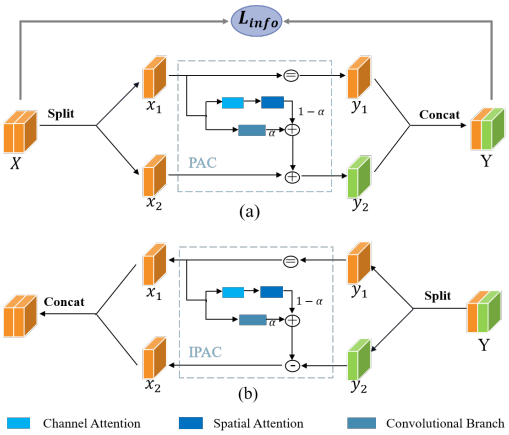


Fig. 4: Structure of the proposed InfoMinRev module, which consists of a forward and a backward process.

3.2 InfoMinRev Module

Similar to other flow models, such as ArtFlow and CAP-VSTNet, our model, illustrated in Fig. 3, comprises padding, squeeze, and cascaded reversible residual modules. Different from the classical flow model, as illustrated in Fig. 4, we introduce a parallel attention-convolution (PAC)-based coupling layer and its inverse operation (IPAC) to enhance the original convolution operation. Each reversible residual module performs a function by splitting the input X to a pair $x_1, x_2 \rightarrow y_1, y_2$ as

$$\begin{aligned} y_1 &= x_1 \\ y_2 &= x_2 + \alpha F\left(x_1\right) + (1-\alpha)\, Att\left(x_1\right), \end{aligned} \tag{1}$$

where $Att(\cdot)$ denotes the attention module, and $\alpha \in [0,1]$ is a learnable adaptive parameter.

Meanwhile, each module contains a reversible block, which obtains $y_1, y_2 \rightarrow x_1, x_2$ as

$$\begin{aligned} x_1 &= y_1 \\ x_2 &= y_2 - \alpha F\left(y_1\right) - (1-\alpha)\, Att\left(y_1\right). \end{aligned} \tag{2}$$

We name our proposed reversible residual module InfoMinRev (IMR) since mutual information minimisation loss is enforced between the input and output of the module. Assume that Y^i and Y^{i+1} represent the input and the output of the i-th module, respectively, then feature redundancy can be reduced by minimising

$$L_{info} = \sum_{i=1}^{n} NMI\left(Y^i, Y^{i+1}\right), \tag{3}$$

where $NMI\left(\cdot\right)$ denotes the normalised mutual information [30] in range $[0,1]$, calculated as

$$NMI\left(Y^i, Y^{i+1}\right) = 2\frac{MI\left(Y^i, Y^{i+1}\right)}{H\left(Y^i\right) + H\left(Y^{i+1}\right)}, \tag{4}$$

where $H\left(\cdot\right)$ denotes the information entropy and

$$MI\left(Y^i, Y^{i+1}\right) = D_{KL}\left(P\left(Y^i, Y^{i+1}\right) \| P\left(Y^i\right) P\left(Y^{i+1}\right)\right). \tag{5}$$

We use kernel density estimation (KDE) [12] to estimate the probability distributions, which can be expressed as

$$\widehat{p}_n\left(z\right) = \frac{1}{Nh^d} \sum_{i=1}^{N} k_o\left(\frac{z - z_i}{h}\right), \tag{6}$$

where d represents the dimension of z, h is the bandwidth, N denotes the number of data points, and $k_o\left(\cdot\right)$ is the chosen kernel function. We employ a Gaussian kernel as it can effectively capture high-dimensional distributions while preserving their smoothness.

3.3 Channel-wise Redundancy Reduction

Neuroscientist H. Barlow postulated [5] that the primary objective of sensory processing is to convert highly redundant sensory inputs into a factorial code. This hypothesis has been proven insightful in elucidating the organisation of the human visual system, from the retina to cortical areas [4]. Inspired by [7, 38], we reduce the channel coupling by enforcing a soft constraint based on Barlow twins loss but allowing channel interdependency to ensure style transferability. In particular, the diagonal coefficients of the cross correlation matrix between content features Z_c and stylised features Z_{cs} are maximised, while the off-diagonal coefficients are minimised. This objective can be formulated as

$$L_B = \sum_i \left(1 - \mathbb{C}_{ii}\right)^2 + \lambda \sum_i \sum_{j \neq i} \left(\mathbb{C}_{ij}\right)^2, \tag{7}$$

where λ is a positive constant trading off the importance of the first and second terms of the loss, and $\mathbb{C}$ is the cross-correlation matrix computed using Z_c and Z_{cs} along the batch dimension, whose elements are calculated as

$$\mathbb{C}_{i,j} = \frac{\sum_b z^C_{b,i} z^{CS}_{b,j}}{\sqrt{\sum_b \left(z^C_{b,i}\right)^2} \sqrt{\sum_b \left(z^{CS}_{b,j}\right)^2}}, \tag{8}$$

where $z^C_{b,i}$ denotes the i-th channel of the b-th sample of Z_c (and analogous for $z^{CS}_{b,j}$).

3.4 Jensen-Shannon Divergence Style Loss

Reduction of statistical value discrepancies between style and transferred images across feature maps is usually used to define the style loss. However, discrepancy gaps vary when setting different reference images, which can lead to over-rendering effects when transferring texture-rich styles or to under-stylisation when using images with simple structures. To achieve a more balanced transfer effect, we replace the mean-variance L2 loss and design a Jensen-Shannon divergence loss for style alignment as

$$\begin{aligned} L_s = \frac{1}{N_l} \sum_{i=1}^{N_l} [&JSD\left(\mu\left(\phi_i\left(I_s\right)\right) \| \mu\left(\phi_i\left(I_{cs}\right)\right)\right) \\ &+ JSD\left(\sigma\left(\phi_i\left(I_s\right)\right) \| \sigma\left(\phi_i\left(I_{cs}\right)\right)\right)], \end{aligned} \tag{9}$$

where $JSD\left(\cdot\right)$ denotes the Jensen–Shannon divergence, $\phi_i\left(\cdot\right)$ are the features extracted from the i-th layer in a pre-trained VGG19 network with N_l layers, and $\mu\left(\cdot\right)$ and $\sigma\left(\cdot\right)$ are the mean and variance of the extracted features. In its implementation, we apply a warm-up stage using the original mean-variance style loss so as to mitigate the vanishing gradient issue in non-overlapping regions of JS divergence. Finally, the Jensen-Shannon divergence aligns the distribution of style features, thus solving the imbalanced style transfer problem.

3.5 Training Loss

We combine all loss terms to arrive at the overall loss function as

$$L_{total} = L_c + \lambda_s L_s + \lambda_B L_B + \lambda_{info} L_{info}, \tag{10}$$

where L_c is the content loss, L_s is the style loss from Eq. (9), L_B is the channel-wise Barlow twins loss from Eq. (7), and L_{info} is the mutual information loss from Eq. (3). The content loss is obtained as

$$L_c = \left\| nm\left(\phi_{con4-2}\left(I_{cs}\right)\right) - nm\left(\phi_{con4-2}\left(I_c\right)\right)\right\|_2, \tag{11}$$

where $nm(\cdot)$ is the normalisation operation used in e.g. [7, 9].

4 Experimental Results

4.1 Implementation Details

We implement an architecture with 20 InfoMinRev modules and 2 squeeze modules, with the transfer module used being identical to that in AdaIN [17]. Following established style transfer methods [2, 24], we use the MS-COCO dataset [27] for content images and the WikiArt dataset [19] for style images. We randomly select 40 style images and 30 content images to generate 1200 stylised images for both quantitative and qualitative testing. The ablation experiments are conducted using the same images. Except for the images used in Section 4.5, which have a resolution of 768×768, all experiments use test images with a resolution of 512×512.

4.2 Style Transfer Results

To demonstrate the superiority of our proposed method, we compare it with seven state-of-the-art methods, namely AdaIN [17], WCT [24], SANet [8], ArtFlow [2], IEContraAST [7], StyTr2 [9], and CAP-VSTNet [36]. For a fair comparisons, we retrain the above methods using the content loss from Eq. (11).

Fig. 5 shows some visual results of stylised images obtained by all methods. As we can see, the results of AdaIN contain cluttered noises and unpleasing texture that compromise the integrity of content structure. Although WCT conceptually excels in transferring styles, it often yields messy stylised results. SANet employs attention-based style transfer modules for transformations in similar semantic regions, however, it has a tendency to improperly apply repeating style patterns to content images. IEContraAST achieves superior visual quality owing to the use of contrastive learning and generative adversarial learning. However, there exist various inconsistencies in compared to the input style reference. Due to the presence of redundant information, ArtFlow suffers from a substantial loss of content details and undesirable artefacts which diminish the overall visual quality. StyTr2 also tends to incorrectly integrate repetitive style patterns into content images, while it may also erroneously transfer the structure of the style

image into the final result. In contrast, our model enables the preservation of more content details and substantially improves the visual quality of stylised images, thus achieving improved performance.

For a quantitative evaluation, following [37], we use the structural similarity index (SSIM) and the content loss between the original content images and the stylised images as objective metrics for assessing content preservation. To gauge the algorithm's capacity to produce artistic effects, following [9], we calculate the mean and variance errors of the feature maps extracted by a VGG [33] model pre-trained on ImageNet between the style and style-transferred images. The obtained results for all methods as well as the computational times are reported in Table 1.

Table 1: Quantitative evaluation results for all methods. Best results for each measure are **bolded**. The computation time is based on 512×512 images.

	WCT	AdaIN	SANet	IEContraAST	StyTr2	ArtFlow	CAP-VSTNet	Ours
SSIM ↑	0.322	0.351	0.407	0.445	0.426	0.365	0.408	**0.496**
Content Loss ↓	0.992	0.914	0.896	0.605	0.712	0.803	0.754	**0.491**
Style Loss ↓	0.912	0.836	**0.587**	1.853	0.634	0.861	0.608	0.652
Time (s) ↓	0.761	**0.003**	0.009	0.042	0.212	0.134	0.096	0.081

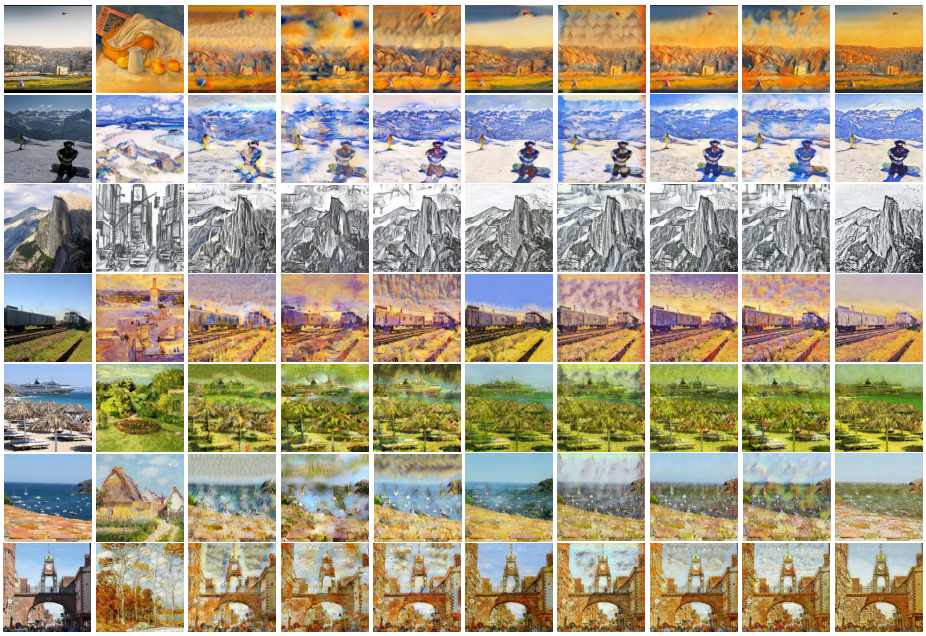


Fig. 5: Style transfer examples of all methods. Our method significantly reduces overlapping style patterns and artefacts, achieving a better balance of content and style, thus yielding superior visual effects

As is evident from there, our proposed method yields the best results in terms of both SSIM and the content loss metric, followed by IEContraAST. However, IEContraAST clearly performs worst in terms of the style loss metric, with the generated images generally poorly aligning with the input style reference as is also noticeable in Fig. 5. In contrast, our proposed approach also performs well in terms of style loss while excelling in terms of content loss. This impressive balance is further illustrated in Fig. 6 which shows all methods in a scatter plot of content loss vs. style loss. As is clearly noticeable, our method outperforms all other techniques when considering both metrics.

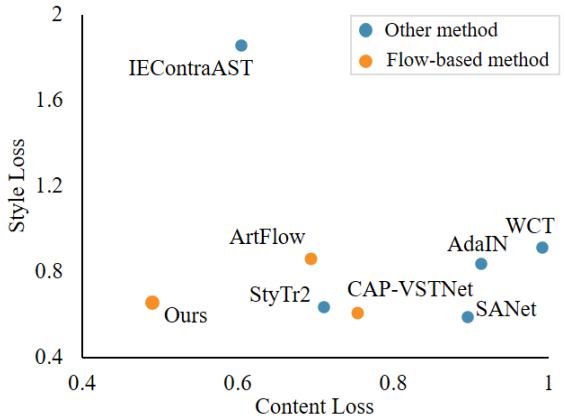


Fig. 6: Scatter plot of content loss vs. style loss.

In terms of runtime, AdaIN is fastest as expected, while our approach outperforms the other flow-based models (ArtFlow and CAP-VSTNet).

4.3 User Study

We conduct a user study to further compare our method with the other approaches. For this, we randomly select 25 pairs of content and style images, performing style transfer using all eight methods. We show the stylised results to a cohort 20 participants (age 20-52; 13 males, 7 females) with normal vision and ask them to rank the images based on three metrics: content preservation, style consistency, and overall performance. We thus collect a total of 1500 sets of rankings and report the average ranks for all methods and all metrics in Table 2. It is evident from there that our method is the most preferred method with average rankings significantly lower compared to the other methods in terms of all three measures, thus further underlining the performance of our proposed method.

Table 2: User study results. Each number represents the average ranking, using the criteria of overall quality, preservation of content and style, respectively.

	AdaIN	WCT	SANet	IEContraAST	ArtFlow	StyTr2	CAP-VSTNet	Ours
Content Preservation ↓	5.63	7.13	7.79	2.14	4.38	4.63	2.88	**1.43**
Style Consistency ↓	5.25	6.50	6.63	3.25	4.07	3.88	4.63	**1.63**
Preference ↓	5.63	7.88	7.00	2.38	4.50	4.88	2.50	**1.50**

4.4 Ablation Study

We conduct an ablation study to prove the usefulness of each introduced component, i.e., the InfoMinRev module, the channel-wise Barlow twins loss, and the Jensen-Shannon divergence style loss. The results are given in Table 3 which reports the various metrics for models where the introduced components are removed in comparison to our full model.

Table 3: Results of ablation study. IMR=InfoMinRev module.

	baseline	w/o IMR&L_S	w/o IMR&L_B	w/o L_B&L_S	w/o IMR	w/o L_B	w/o L_S	full model
SSIM ↑	0.372	0.430	0.361	0.424	0.477	0.459	0.462	**0.496**
Content Loss ↓	0.725	0.627	0.674	0.625	0.542	0.538	0.507	**0.491**
Style Loss ↓	0.912	0.823	**0.566**	0.907	0.719	0.740	0.836	0.652

As we can see from Table 3, all proposed components contribute to the overall performance of our model. Both the InfoMinRev module and the channel-wise Barlow twins loss term support preserving a more robust content structure, while the Jensen-Shannon loss term allows to maintain style consistency. The full model hence strikes an optimal balance between content and style to give the overall best results, while removing the various components leads to diminished performance.

This is further illustrated in Fig. 7, which shows some example ablation results for when the InfoMinRev module and/or the Barlow twins loss are omitted. Both L_B and InfoMinRev can effectively remove redundant information to enhance visual quality. The best results are achieved when both are employed simultaneously, enabling preservation of content details and removal of artefacts.

Fig. 8 further illustrates the importance of the employed Jensen-Shannon loss and shows some examples obtained by replacing this term with the widely used mean-variance L2 loss compared to the result of the full model. As we can observe, the divergence style loss alleviates the issue of stylisation imbalance.

4.5 Higher-resolution Transfer Results

To further validate the performance of our proposed method, we generate style transfer results using higher-resolution images of resolution 768×768 in Fig. 9. As we can notice, when handling these high-resolution images with more content and texture details, our method still successfully captures the content structures

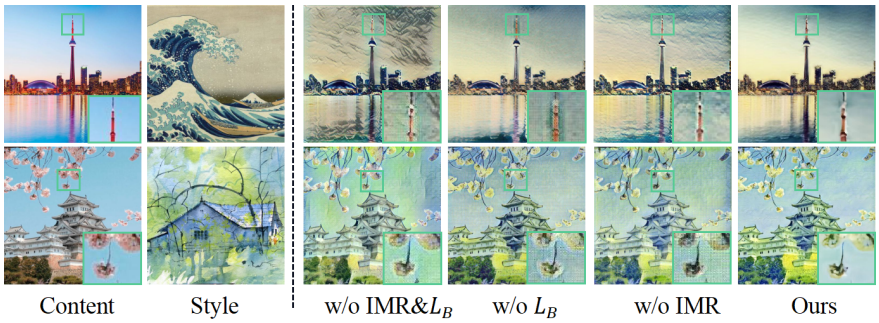


Fig. 7: Example ablation results, comparing models with removed InfoMinRev module and/or Barlow twins loss to the full model.

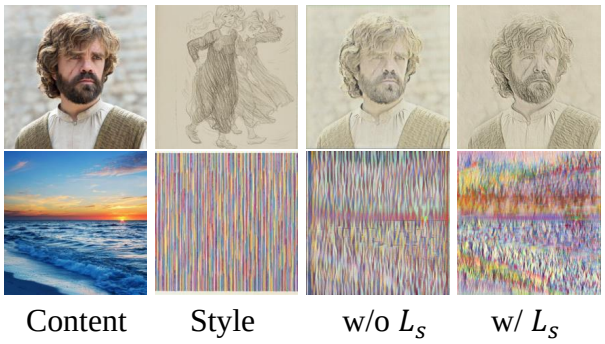


Fig. 8: Example ablation results, comparing a model with omitted Jensen-Shannon loss term to the full model.

(e.g. (going from left to right) the bird, the train, the horses, the building, the horse and cart), and stylises them with consistent style patterns.

4.6 Information Compression Analysis

Our proposed InfoMinRev module is designed to guide the model to extract a more meaningful feature representation with less redundant information. To further investigate its role, we visualise, in Figure 10, the information distributions in intermediate layers of ArtFlow, our proposed model and our model without the L_{info} loss term.

We can observe that for ArtFlow the information distribution is fairly constant in all regions without discrimination. In contrast, our model, whether using L_{info} or not, better aggregates information in the relevant regions. This demonstrates the effectiveness of the attention mechanism we have introduced. Additionally, the inclusion of L_{info} significantly reduces the amount of information contained in the feature maps, allowing our model to focus on the most task-relevant parts.

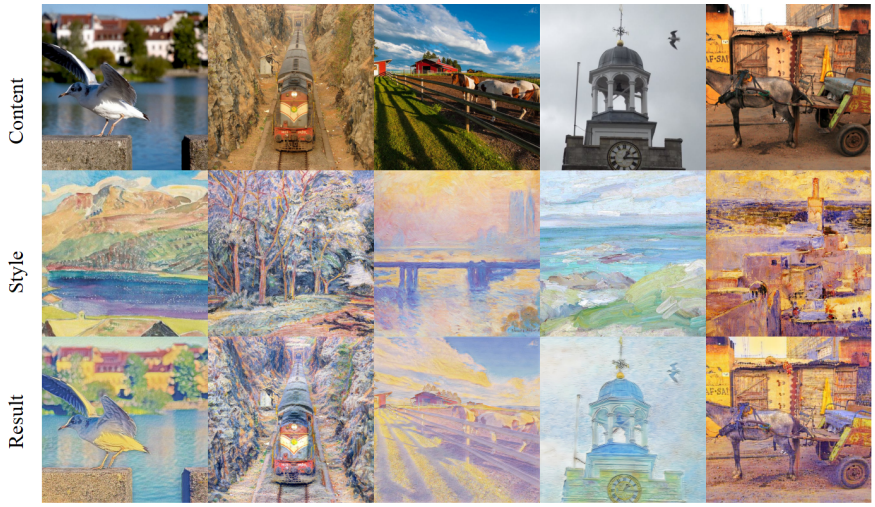


Fig. 9: Examples of higher-resolution (768 × 768) transfer results.

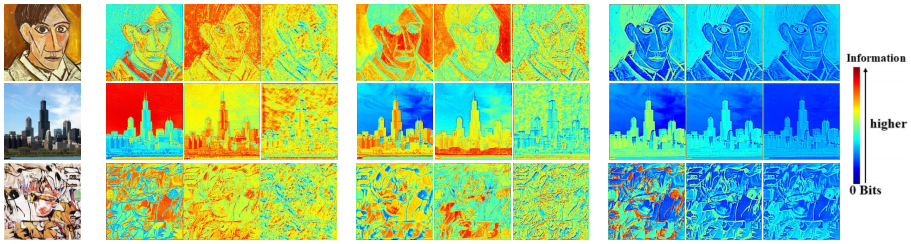


Fig. 10: Visualisation examples of the information distribution in the output of the (from left to right) 5-th, 10-th and 20-th block of (a) ArtFlow, (b) our model without L_{info}, and (c) our full model. The colours represent how many bits are used to encode features at each position of an image.

5 Conclusions

In this paper, we have proposed a novel flow-based approach for neural style transfer, motivated by an information theoretic perspective. In particular, we introduce a reversible InfoMinRev module designed to extract effective feature representations, employ a Barlow twins loss to address the channel-wise coupling issue, and a Jensen-Shannon divergence style loss to balance the style transfer. Our method reduces feature redundancy generated during the encoding process, extracts features that possess both excellent content preservation and consistent style transformation capabilities, and effectively addresses the issue of unwanted artefacts present in previous methods. Extensive experiments demonstrate the effectiveness and superiority of our method in comparison with other state-of-the-art approaches.

Acknowledgements

This research is supported by the Natural Science Foundation of Hunan Province, China (2022GK5002,2024JK2015,2024JJ5440), the National Natural Science Foundation of China (61602527), the Special Foundation for Distinguished Young Scientists of Changsha (kq2209003), the Changsha Municipal Natural Science Foundation (kq2202109), the National Foreign Expert Project (G2023041039L), the 111 Project (D23006), and the High Performance Computing Center of Central South University.

References

1. Alemi, A.A., Fischer, I., Dillon, J.V., Murphy, K.: Deep variational information bottleneck. arXiv preprint arXiv:1612.00410 (2016)
2. An, J., Huang, S., Song, Y., Dou, D., Liu, W., Luo, J.: ArtFlow: Unbiased image style transfer via reversible neural flows. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 862–871 (2021)
3. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: International Conference on Machine Learning. pp. 214–223 (2017)
4. Barlow, H.: Redundancy reduction revisited. Network: Computation in Neural Systems **12**(3), 241 (2001)
5. Barlow, H.B.: Possible principles underlying the transformation of sensory messages. Sensory Communication **1**(01), 217–233 (1961)
6. Chen, D., Yuan, L., Liao, J., Yu, N., Hua, G.: Stylebank: An explicit representation for neural image style transfer. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1897–1906 (2017)
7. Chen, H., Zhao, L., Wang, Z., Zhang, H., Zuo, Z., Li, A., Xing, W., Lu, D.: Artistic style transfer with internal-external learning and contrastive learning. In: Advances in Neural Information Processing Systems. vol. 34, pp. 26561–26573 (2021)
8. Cheng, J., Jaiswal, A., Wu, Y., Natarajan, P., Natarajan, P.: Style-aware normalized loss for improving arbitrary style transfer. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 134–143 (2021)
9. Deng, Y., Tang, F., Dong, W., Ma, C., Pan, X., Wang, L., Xu, C.: StyTr2: Image style transfer with transformers. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11326–11336 (2022)
10. Dinh, L., Krueger, D., Bengio, Y.: Nice: Non-linear independent components estimation. arXiv preprint arXiv:1410.8516 (2014)
11. Dumoulin, V., Shlens, J., Kudlur, M.: A learned representation for artistic style. arXiv preprint arXiv:1610.07629 (2016)
12. Fisher, R.A.: On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society of London. Series A **222**(594-604), 309–368 (1922)
13. Gatys, L., Ecker, A.S., Bethge, M.: Texture synthesis using convolutional neural networks. Advances in Neural Information Processing Systems **28** (2015)
14. Gatys, L.A., Bethge, M., Hertzmann, A., Shechtman, E.: Preserving color in neural artistic style transfer. arXiv preprint arXiv:1606.05897 (2016)
15. Gomez, A.N., Ren, M., Urtasun, R., Grosse, R.B.: The reversible residual network: Backpropagation without storing activations. Advances in Neural Information Processing Systems **30** (2017)

16. Huang, S., Xiong, H., Wang, T., Wen, B., Wang, Q., Chen, Z., Huan, J., Dou, D.: Parameter-free style projection for arbitrary image style transfer. In: IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 2070–2074 (2022)
17. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: IEEE/CVF International Conference on Computer Vision. pp. 1501–1510 (2017)
18. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision. pp. 694–711 (2016)
19. Karayev, S., Trentacoste, M., Han, H., Agarwala, A., Darrell, T., Hertzmann, A., Winnemoeller, H.: Recognizing image style. arXiv preprint arXiv:1311.3715 (2013)
20. Kingma, D.P., Dhariwal, P.: Glow: Generative flow with invertible 1x1 convolutions. Advances in Neural Information Processing Systems **31** (2018)
21. Kitaev, N., Kaiser, Ł., Levskaya, A.: Reformer: The efficient transformer. arXiv preprint arXiv:2001.04451 (2020)
22. Lee, J., Kim, H., Shim, J., Hwang, E.: Cartoon-flow: A flow-based generative adversarial network for arbitrary-style photo cartoonization. In: ACM International Conference on Multimedia. pp. 1241–1251 (2022)
23. Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., Yang, M.H.: Diversified texture synthesis with feed-forward networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3920–3928 (2017)
24. Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., Yang, M.H.: Universal style transfer via feature transforms. Advances in Neural Information Processing Systems **30** (2017)
25. Li, Z., Meng, Q., Liu, Q., Zhang, Z., Zhang, L., Zhang, S.: Face shape transfer via semantic warping. Available at SSRN 3997660
26. Lin, T., Ma, Z., Li, F., He, D., Li, X., Ding, E., Wang, N., Li, J., Gao, X.: Drafting and revision: Laplacian pyramid network for fast high-quality artistic style transfer. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5141–5150 (2021)
27. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: Common objects in context. In: European Conference on Computer Vision. pp. 740–755 (2014)
28. Liu, S., Lin, T., He, D., Li, F., Wang, M., Li, X., Sun, Z., Li, Q., Ding, E.: AdaAttN: Revisit attention mechanism in arbitrary neural style transfer. In: IEEE/CVF International Conference on Computer Vision. pp. 6649–6658 (2021)
29. Park, D.Y., Lee, K.H.: Arbitrary style transfer with style-attentional networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5880–5888 (2019)
30. Pluim, J.P., Maintz, J.A., Viergever, M.A.: Mutual-information-based registration of medical images: a survey. IEEE Transactions on Medical Imaging **22**(8), 986–1004 (2003)
31. Risser, E., Wilmot, P., Barnes, C.: Stable and controllable neural texture synthesis and style transfer using histogram losses. arXiv preprint arXiv:1701.08893 (2017)
32. Sheng, L., Lin, Z., Shao, J., Wang, X.: Avatar-net: Multi-scale zero-shot style transfer by feature decoration. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8242–8250 (2018)
33. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

34. Ulyanov, D., Lebedev, V., Vedaldi, A., Lempitsky, V.: Texture networks: Feed-forward synthesis of textures and stylized images. arXiv preprint arXiv:1603.03417 (2016)
35. Wang, X., Oxholm, G., Zhang, D., Wang, Y.F.: Multimodal transfer: A hierarchical deep convolutional neural network for fast artistic style transfer. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5239–5247 (2017)
36. Wen, L., Gao, C., Zou, C.: Cap-vstnet: Content affinity preserved versatile style transfer. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18300–18309 (2023)
37. Yoo, J., Uh, Y., Chun, S., Kang, B., Ha, J.W.: Photorealistic style transfer via wavelet transforms. In: IEEE/CVF International Conference on Computer Vision. pp. 9036–9045 (2019)
38. Zbontar, J., Jing, L., Misra, I., LeCun, Y., Deny, S.: Barlow twins: Self-supervised learning via redundancy reduction. In: International Conference on Machine Learning. pp. 12310–12320 (2021)
39. Zhang, Y., Li, M., Li, R., Jia, K., Zhang, L.: Exact feature distribution matching for arbitrary style transfer and domain generalization. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8035–8045 (2022)