HIGH TECHNOLOGY LETTERS | Vol. 30 No. 1 | Mar. 2024 | pp. 23-30

doi:10.3772/j.issn.1006-6748.2024.01.003

Unsupervised multi-modal image translation based on the squeeze-and-excitation mechanism and feature attention module①

HU Zhentao(胡振涛)*, HU Chonghao②*, YANG Haoran*, SHUAI Weiwei**

(*School of Artificial Intelligence, Henan University, Zhengzhou 450046, P. R. China)

(**95795 Troops of the PLA, Guilin 541003, P. R. China)

Abstract

The unsupervised multi-modal image translation is an emerging domain of computer vision whose goal is to transform an image from the source domain into many diverse styles in the target domain. However, the multi-generator mechanism is employed among the advanced approaches available to model different domain mappings, which results in inefficient training of neural networks and pattern collapse, leading to inefficient generation of image diversity. To address this issue, this paper introduces a multi-modal unsupervised image translation framework that uses a generator to perform multi-modal image translation. Specifically, firstly, the domain code is introduced in this paper to explicitly control the different generation tasks. Secondly, this paper brings in the squeeze-and-excitation (SE) mechanism and feature attention (FA) module. Finally, the model integrates multiple optimization objectives to ensure efficient multi-modal translation. This paper performs qualitative and quantitative experiments on multiple non-paired benchmark image translation datasets while demonstrating the benefits of the proposed method over existing technologies. Overall, experimental results have shown that the proposed method is versatile and scalable.

Key words: multi-modal image translation, generative adversarial network (GAN), squeeze-and-excitation(SE) mechanism, feature attention (FA) module

0 Introduction

With the rapid development of deep learning, image translation[1] has evolved considerably over the past few years. Image translation aims to translate images from one domain to another. Many computer vision and image processing problems can be handled within this framework, such as super-resolution[2], image coloring[3], image drawing[4], image restoration, style transfer[5]. Previous work has given impressive results on tasks via deterministic one-to-one mapping, but pattern collapse occurs when the output corresponds to multiple possibilities. To address this problem, recent research has focused on one-to-many translations and explores this problem: multi-modal translation[6].

Recently, generative adversarial network (GAN)[7] has become a promising field of research. Following pioneering work on Pix2Pix[8], BicycleGAN[9] is developed to support multiple styles of image translation. Later, researchers propose unsupervised image-to-image translation networks (UNIT)[10], CycleGAN[11], multimodal unsupervised image-to-image translation (MUNIT)[12], diverse image-to-image translation (DRIT)[13], and DRIT ++ for unsupervised image translation. Among them, MUNIT, DRIT, and DRIT + + allow multimodal translation. Nevertheless, multimodal image translation is still challenging.

A great deal of work has been done on image translation to address the diversity of images generated in multi-modal image translation. Invertible conditional GAN (ICGAN)[14] and Fader networks brought together encoder-decoder architecture with the GAN, enabling the transformation of multiple attributes. Later, StarGAN[15] added domain labels to control transitions between multiple domains. BicycleGAN achieved a one-to-many mapping between source and target domains by combining the goals of conditional variational autoencoder GAN (CVAE-GAN)[16] and conditional latent regressor GAN (CLRGAN).

In conclusion, unsupervised multimodal image translation remains a pressing problem. Given the lack of output image diversity, an innovative model is proposed. In the proposed algorithm, domain codes are introduced as auxiliary inputs to the network, explicitly

① Supported by the National Natural Science Foundation of China (No. 61976080), the Academic Degrees & Graduate Education Reform Project of Henan Province (No. 2021SJGLX195Y), the Teaching Reform Research and Practice Project of Henan Undergraduate Universities (No. 2022SYJXLX008), and the Key Project on Research and Practice of Henan University Graduate Education and Teaching Reform (No. YJSJG2023XJ006).

② To whom correspondence should be addressed. E-mail: chonghaohu@henu.edu.cn.

Received on Mar. 31, 2023

controlling the different generating tasks. At the same time, the generator network has been modified. The proposed model produces high-quality and diverse images while retaining the content of the source images, and the main contributions of this paper are as follows.

(1) This paper presents an unsupervised multi-modal image translation based on the squeeze-and-excitation mechanism and feature attention GAN module (SEFAGAN). It utilizes a generator and a set of discriminators to perform the transformation between unpaired multimodal images.

(2) In this paper, a new network structure is designed. The network structure in the generator introduces feature attention (FA) blocks after the convolution blocks in the second and third layers, and the FA module is used to capture the interconnections of various features and can improve their context-aware translation capabilities. At the same time, squeeze-and-excitation (SE) blocks are inserted within the residual block module layer. The SE module establishes dependencies between feature maps and assigns weight values to different feature maps to increase the importance of useful features.

(3) This paper performs qualitative and quantitative evaluations on a variety of datasets. The experimental results show that the proposed method has excellent output and good results in experiments compared with the advanced methods.

1 Related work

1.1 Generative adversarial network

GAN is described as a framework for learning data distributions in an unsupervised manner. Generator is used to map random noise onto an image. A discriminator is used to determine whether the image generated by the generator is true or false. Training of GAN always involves a min-max game between two individuals. To improve the model structure, researchers have proposed various GAN-based derivative models. This will further broaden the theory and applications of GAN.

1.2 Multi-modal image translation

Several researchers have attempted to increase the versatility and elasticity of the model by creating one-to-many mappings between the source and target domains. BicycleGAN addressed this constraint by limiting the double-tap conformance of the output and potential layer encoding. DRIT and MUNIT separated images into a domain-invariant content space and a domain-specific attribute space to augment the diversity of the generated images. Scalable and diverse cross-domain image translation (SDIT)[17] proposed a model that enables multi-modal and multi-domain image translation. In contrast to these approaches, this paper proposes a new model that realizes multi-modal outputs between multiple domains in the lack of pairwise datasets.

2 The proposed methodology

This paper aims to explore a new multi-modal model that can realize image translation. The proposed model applies only one generator to perform instance-perception mapping in multi-modal image translation. This not only simplifies the model structure but also allows instances and images to share some common features. This makes it easier to incorporate translated instances into translated images. As shown in Fig. 1, the proposed model consists of five parts: namely image preparation, domain code acquisition, encoder, feature latent code acquisition, and generator.

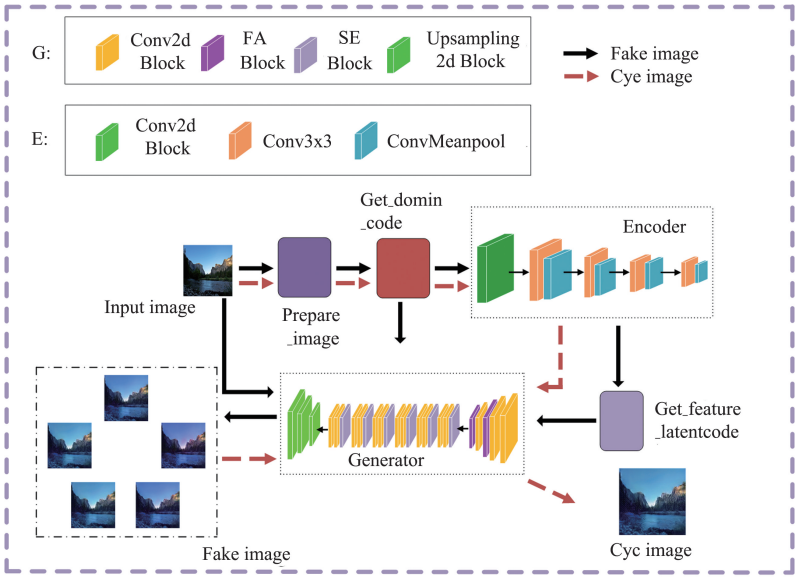

Fig. 1 An overview of SEFAGAN

2.1 SEFAGAN architecture

In this section, this paper focuses on the proposed framework and then designs it in detail to consider different factors. Suppose X and Y are collections of the image source domain and target domain, respectively. Provided with an image $x \in X$ and the other $y \in Y$, then the task of the proposed network is to train a single generator to convert the images from X to Y while obtaining multi-modal images. The function of the encoder E is to go through convolution and finally obtain the mean of the image obedience distribution, covariance, and feature latent code. G is then trained to reflect the style information for these specific domains. The architecture of the entire network framework is shown in Fig. 1.

To indicate specific mappings, this paper introduces domain codes to act as auxiliary inputs to the network, explicitly controlling the different generative tasks. Therefore, this paper improves the generator by adding FA blocks after the second and third layers of convolution blocks in down-sampling while introducing SE blocks within the residual block layer. This paper then integrates multiple optimization goals to learn specific translations. The structure of the feature attention module and the squeeze and excitation mechanism are shown in Fig. 2 and Fig. 3, respectively.

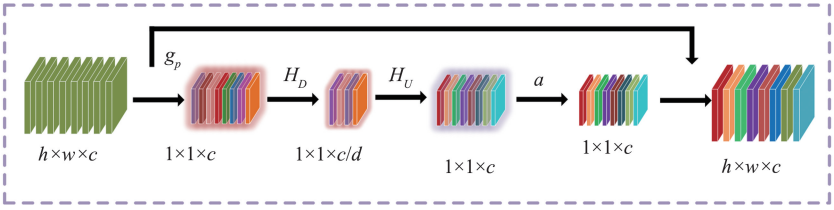

Fig. 2 Feature attention module

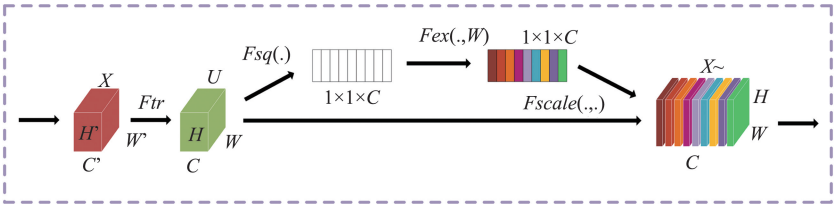

Fig. 3 Squeeze-and-excitation mechanism

2.2 Loss function

2.2.1 GAN loss

As a single generator has multiple domain outputs, this paper has matching targets for each target domain and uses a set of discriminators. Corresponding discriminators are used to identify images generated in one domain. The adversarial loss is applied to both mapping functions. With this adversarial loss, the translated image and the distribution of the target image are matched exactly. In the mapping functions $G_{AB}: A \times E_{XA} \rightarrow B \times E_{XA}$ and $G_{BA}: B \times E_{XB} \rightarrow A \times E_{XA}$, as well as their discriminators, the entire adversarial losses are defined as shown in Eq. (1) and Eq. (2).

$$L_{\text{adv}}(G, D_A) = E_{XA}[\log(D_A(x_A))] + E_{XB}[\log(1 - D_A(G(x_B, z_A)))] \quad (1)$$

$$L_{\text{adv}}(G, D_B) = E_{XB}[\log(D_B(x_B))] + E_{XA}[\log(1 - D_B(G(x_A, z_B)))] \quad (2)$$

By optimizing multiple generative adversarial objects, the generator restores the different domain distributions indicated by the domain code z, where x_A, x_B, z_A are the real image of the input target domain, the source domain image, and the target domain index, respectively. Similarly, the opposite x_B, x_A, z_B are also the real image of the input target domain, the source domain image, and the target domain index, respectively.

2.2.2 Cyclic consistency loss

While the above GAN loss can be accomplished with domain transformations, highly under-constrained mappings usually lead to pattern collapse. There are many possible mappings that can be inferred without using pairwise information.

To reduce the space for possible mapping, this paper uses a cyclic consistency constraint during the training phase. Cyclic consistency loss is defined as $x_A \approx G(G(x_A, z_B), z_A)$ and $x_B \approx G(G(x_B, z_A), z_B)$. The formula for cyclic consistency loss is defined as shown in Eq. (3).

$$L_{\text{cyc}}(G) = E_{XA}[\| x_A - G(G(x_A, z_B), z_A) \|_1] + E_{XB}[\| x_B - G(G(x_B, z_A), z_B) \|_1] \quad (3)$$

where $\| \cdot \|_1$ represents the 1 norm; x_A, z_B, z_A are the real image of the input source domain, the target domain index, and the source domain index, respectively; E_{XA} is a mapping of $A \rightarrow B \rightarrow A$. The reverse is true, x_B, z_A, z_B are also the real image of the input source domain, the target domain index, and the source domain index, respectively; E_{XB} is a mapping of $B \rightarrow A \rightarrow B$.

2.2.3 Overall loss target

The final objective function is defined as shown in Eq. (4). Training loss consists of an adversarial loss L_{adv} and a cyclic consistency loss L_{cyc}, where γ_{cyc} controls the relative importance of the two targets.

$$G^* = \arg \min_G \max_{D_A, D_{Bi} \in \{A,B\}} \sum L_{\text{adv}}(G, D_i) + \lambda_{\text{cyc}} L_{\text{cyc}}(G) \quad (4)$$

3 Experiments and results

3.1 Datasets and the baseline

The effectiveness of the proposed method is demonstrated by three datasets, which are Horse2Zebra, Summer2Winter, and Cat2Dog.

For multi-modal image translation tasks, this paper selects MUNIT[12], DRIT[13], ComboGAN[18], and SingleGAN[19] as the baseline.

3.2 Evaluation metrics and training details

To evaluate the results of the features in different metric learning, the Fréchet inception distance (FID)[20] and learned perceptual image patch similarity (LPIPS)[21] are used as the evaluation protocol.

Using the PyTorch experimental platform, the proposed model is trained on an RTX3090 GPU. The batch size is set to 16. The model iterates 400 epochs. The proposed model is trained and endorsed end-to-end by using Adam with momentum terms $\beta_1 = 0.500$ and $\beta_2 = 0.999$. KL-divergence and latent regression have weights of 0.100 and 0.500, respectively.

3.3 Experimental results and analysis

3.3.1 Qualitative evaluation

In this section, this paper experiments with the proposed approach and the advanced model. In order to verify the effectiveness of the proposed algorithm, the model is trained with each dataset separately.

Summer2Winter dataset: Fig. 4 shows the results of SEFAGAN and several advanced models. The experimental results look very realistic, which shows that SEFAGAN has good image transformation capabilities. The diversity of images generated by SingleGAN is not very good. The images generated by MUNIT achieve diversity, but compared with SEFAGAN, they did not perform as well as the images generated by SEFAGAN. The images generated by DRIT do achieve diversity, but they do not perform as well compared with SEFAGAN. ComboGAN generates less diverse images than the proposed method.

As shown in Fig. 5, the best results are obtained with SEFAGAN. Compared with the results from SingleGAN, the experimental results from SEFAGAN are more reasonable and transparent, outperforming the other models. The images generated by DRIT perform the worst. MUNIT performs better than DRIT, because it achieves the translation task better at lower levels of feature information. ComboGAN is less diverse although the quality of the generated images is slightly better than the proposed model.

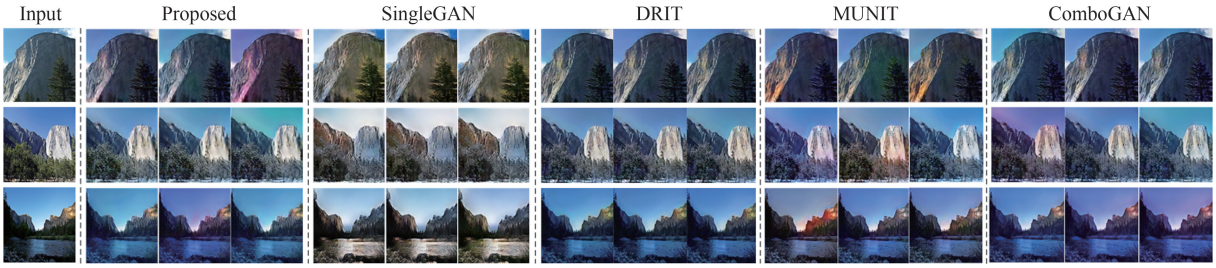

Fig. 4 Qualitative results on Summer→Winter task

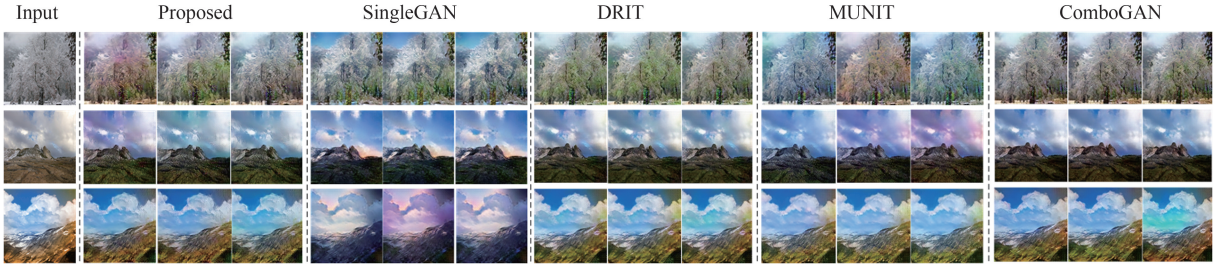

Fig. 5 Qualitative results on Winter→Summer task

Horse2Zebra dataset: as shown in Fig. 6, SEFAGAN gives the best results, with clear artifacts in those images generated by the existing models. At the same time, its results are still unsatisfactory, with a lack of diversity in the different styles. In most cases, the transformed zebra images are almost indistinguishable, and in the last row of SingleGAN, all three translated horse images are almost identical. Compared with the baseline approach, SEFAGAN produces more realistic and diverse high-quality images. In terms of realism, the real input images are no better than the three images generated by SEFAGAN. In terms of diversity, the three images generated can easily be divided into different styles of the corresponding zebra images.

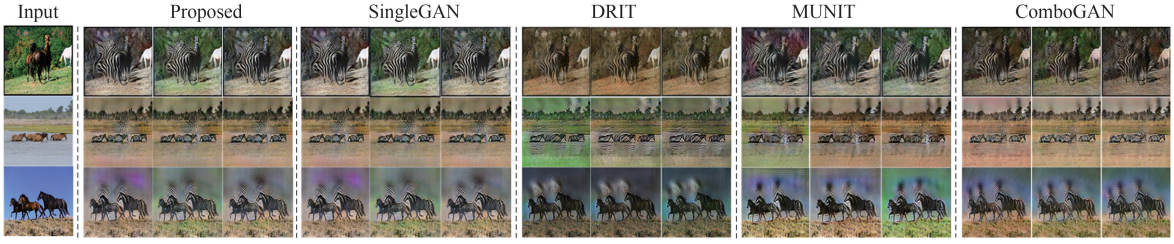

Fig. 6 Qualitative results on the Horse→Zebra task

As can be seen in Fig. 7, SEFAGAN gives the best results, while those images generated by the existing model not only have obvious artefacts. Also, their images are not of high quality and are not clear. In terms of the diversity of the images generated, SEFAGAN works the best. SEFAGAN can clearly generate images of different styles of horses. SingleGAN works second best. MUNIT works worse than SingleGAN. DRIT works the worst. The effect of ComboGAN is similar to SingleGAN.

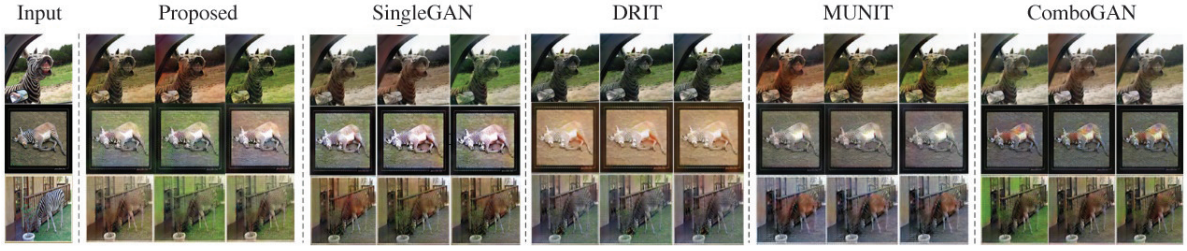

Fig. 7 Qualitative results on the Zebra→Horse task

Cat2Dog dataset: this paper also conducts experiments on the Cat2Dog dataset to illustrate the effectiveness of SEFAGAN, as shown in Fig. 8. On the Cat→Dog task, SEFAGAN is second only to MUNIT in terms of image quality and diversity, while DRIT is the worst. ComboGAN is worse in terms of image quality and SEFAGAN achieves the experimental results expectedly compared with the SingleGAN. Fig. 9 shows that the same is true for the Dog→Cat task.

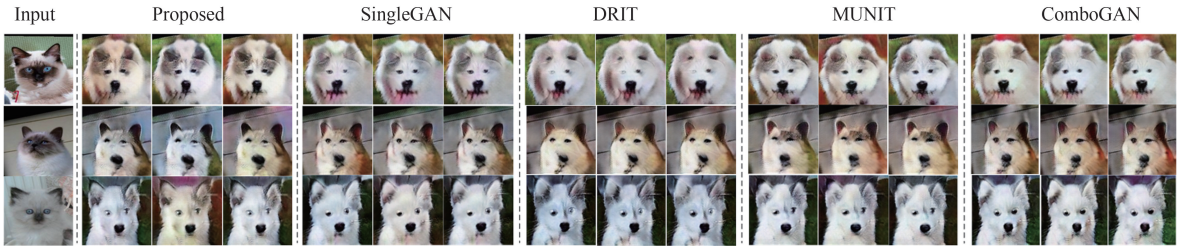

Fig. 8 Qualitative results on Cat→Dog task

3.3.2 Quantitative evaluation

This paper assesses quality and diversity based on the experimentally generated images, as shown in Tables 1, 2 and 3. For diversity, similar to BicycleGAN, this paper uses the LPIPS measure to measure similarity between images. In addition, this paper uses FID to obtain perceptual scores.

Summer2Winter dataset: Table 1 shows the quantitative results of SEFAGAN and the latest model. SE-

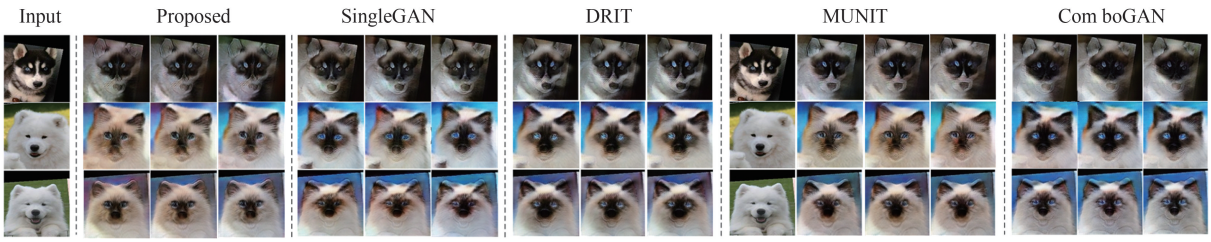

Fig. 9 Qualitative results on Dog→Cat task

FAGAN achieves the best FID and LPIPS scores in the Summer→Winter and Winter→Summer translation directions. Compared with other models, DRIT has the worst diversity effect at 0.147, and the FID indicator is also not good, at 133.535. Compared with SEFAGAN, its LPIPS score differs by 0.028, its FID score differs by 5.200, while MUNIT and ComboGAN are not as effective as the proposed model. In the direction of Summer→Winter translation, the LPIPS and FID of SingleGAN are second only to SEFAGAN. In the direction of Winter→Summer translation, the LPIPS of SingleGAN is second only to SEFAGAN and its FID score is the worst. On the whole, SEFAGAN has good multi-modal translation capabilities for these two tasks.

Table 1 Performance of different models on Summer2Winter dataset

Method	Summer→Winter		Winter→Summer	
	LPIPS ↑	FID ↓	LPIPS ↑	FID ↓
MUNIT[12]	0.152 ±0.016	136.298	0.150 ±0.016	113.747
DRIT[13]	0.147 ±0.017	133.535	0.142 ±0.016	112.925
ComboGAN[18]	0.157 ±0.016	130.525	0.154 ±0.016	112.257
SingleGAN[19]	0.160 ±0.015	128.505	0.160 ±0.015	115.604
Proposed	0.175 ±0.015	128.380	0.163 ±0.015	114.066

Horse2Zebra dataset: Table 2 shows the results of SEFAGAN and contrasting models. SEFAGAN has better experimental results. In the direction of Horse→Zebra, the LIPIPS and FID indicators of SEFAGAN experimental effect are the best. DRIT has the worst diversity indicator. Compared with SEFAGAN, the LPIPS score for MUNIT differs by 0.018, its FID score differs by 41.100. The LPIPS score of SingleGAN differs from that of SEFAGAN by 0.005, its FID score differs from that of SEFAGAN by 50.000. In the Zebra→Horse direction, The LPIPS score of SEFAGAN is the best. It has a poor FID score. In terms of DRIT, both its LPIPS and FID scores are the worst. Compared with SEFAGAN, the LPIPS and FID scores of MUNIT differ by 0.022 and 1.100 respectively. The LPIPS and FID scores of SingleGAN differed by 0.006 and 2.800 respectively. And ComboGAN is not as good as the proposed algorithm in any direction.

Table 2 Performance of different models on Horse2Zebra dataset

Method	Horse→Zebra		Zebra→Horse	
	LPIPS ↑	FID ↓	LPIPS ↑	FID ↓
MUNIT[12]	0.180 ±0.016	210.563	0.160 ±0.014	200.099
DRIT[13]	0.166 ±0.016	213.971	0.151 ±0.015	211.504
ComboGAN[18]	0.189 ±0.016	213.333	0.174 ±0.014	205.599
SingleGAN[19]	0.193 ±0.016	219.505	0.176 ±0.014	198.318
Proposed	0.198 ±0.016	169.472	0.182 ±0.014	201.114

Cat2Dog dataset: Table 3 is the results of SEFAGAN and existing models. Table 3 shows that the LIPIPS and FID indicators of SEFAGAN are not the best in the direction of Cat→Dog and Dog→Cat translation, but LIPIPS and FID indicators are second only to MUNIT. In contrast, although DRIT and SingleGAN can successfully translate the semantic information of objects, their translation effect is not as good as the proposed model. ComboGAN is similarly effective to the proposed model, but not as effective as the proposed approach.

Table 3 Performance of different models on Cat2Dog dataset

Method	Cat→Dog		Dog→Cat	
	LPIPS ↑	FID ↓	LPIPS ↑	FID ↓
MUNIT[12]	0.138 ±0.010	125.498	0.124 ±0.010	93.570
DRIT[13]	0.113 ±0.011	130.087	0.102 ±0.010	116.087
ComboGAN[18]	0.131 ±0.010	130.468	0.120 ±0.010	276.856
SingleGAN[19]	0.127 ±0.011	143.018	0.118 ±0.010	105.250
Proposed	0.136 ±0.011	128.313	0.120 ±0.009	95.524

3.3.3 Ablation study

In this section, an ablation study is conducted to investigate the contribution of the squeeze-and-excitation mechanism and feature attention module (SEFA) embedded in the proposed model. Include: (1) baseline: does not use any module; (2) baseline + SE: uses squeeze-and-excitation mechanism; (3) base-

line + FA: uses feature attention module; (4) proposed: uses SEFA module.

Table 4 shows the quantitative results of ablation experiments. As can be seen from Table 4, the SE module and the FA module each have their own effects and performance. Compared with the original model, the LPIPS index of the SE module decreased in the direction of Summer→Winter and Winter→Summer. The FA module rises in the direction of Summer→Winter and Winter→Summer. Compared with the original model, the SEFA module achieves better results with the LPIPS indicator.

Table 4 Performance of Ablation study

Method	Summer→Winter		Winter→Summer	
	LPIPS ↑	FID ↓	LPIPS ↑	FID ↓
Baseline	0.161 ± 0.016	128.505	0.160 ± 0.015	115.604
Baseline + SE	0.150 ± 0.016	135.375	0.147 ± 0.015	117.835
Baseline + FA	0.166 ± 0.016	130.335	0.169 ± 0.015	111.553
Proposed	0.175 ± 0.015	128.380	0.163 ± 0.015	114.066

4 Conclusions

To address the low diversity problem of unpaired image translation, this paper proposes a new multimodal unsupervised image translation model, named SEFAGAN. In the proposed method, the domain code is introduced as an auxiliary input to the network, explicitly controlling the different generative tasks. Then, the generator is improved by inserting FA blocks after the second and third layers of convolution blocks while introducing SE blocks within the residual block layer. This paper then integrates multiple optimization goals to learn specific translations. Compared with models that already exist, the proposed model works better.

References

[1] EMAMI H, ALIABADI M M, DONG M, et al. Spatial attention GAN for image-to-image translation [J]. IEEE Transactions on Multimedia, 2020, 23: 391-401.

[2] ZHANG W, LIU Y, DONG C, et al. Generative adversarial networks with ranker for image super-resolution [C] // IEEE International Conference on Computer Vision. Seoul, Korea: IEEE, 2019: 3096-3105.

[3] YAN J, LIN S, BING KANG S, et al. A learning-to-rank approach for image color enhancement [C] // IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014: 2987-2994.

[4] SHAO M, ZHANG W, ZUO W, et al. Multi cale generative adversarial inpainting network based on cross-layer attention transfer mechanism [J]. Knowledge-Based Systems, 2020, 196: 105778.

[5] YI Z, ZHANG H, TAN P, et al. Unsupervised dual learning for image-to-image translation [C] // IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2849-2857.

[6] SUN Y, LU Y, LU H, et al. Multimodal unsupervised image-to-image translation without independent style encoder [C] // International Conference on Multi-Media Modeling. Phu Quoc, Vietnam: MMM, 2022: 624-636.

[7] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C] // Proceedings of the International Conference Neural Information Processing Systems. Montreal, Canada: NIPS, 2014: 2672-2680.

[8] JIAO L, HU C, HUO L, et al. Guided-Pix2Pix: End-to-end inference and refinement network for image dehazing [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 3052-3069.

[9] ZHU J Y, ZHANG R, PATHAK D, et al. Toward multimodal image-to-image translation [C] // Annual Conference on Neural Information Processing Systems. Long Beach, USA: NeurIPS, 2017: 465-476.

[10] LIU M Y, BREUEL T, KAUTZ J M, et al. Unsupervised image-to-image translation networks [C] // Annual Conference on Neural Information Processing Systems. Long Beach, USA: NIPS, 2017: 700-708.

[11] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C] // IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2242-2251.

[12] HUANG X, LIU M Y, BELONGIE S, et al. Multimodal unsupervised image-to-image translation [C] // European Conference on Computer Vision. Munich, Germany: ECCV, 2018: 179-196.

[13] LEE H Y, TSENG H Y, HUANG J B, et al. Diverse image-to-image translation via disentangled representations [C] // European Conference on Computer Vision. Munich, Germany: ECCV, 2018: 35-51.

[14] SHRIVASTAVA U, THADA V, KUMAR A, et al. Deep learning approach to face conditioning using invertible conditional generative adversarial networks [J]. International Journal of Innovative Research in Computer Science & Technology, 2020, 8(3): 164-170.

[15] CHOI Y, CHOI M, KIM M, et al. Unified generative adversarial networks for multi-domain image-to-image translation [C] // IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 8789-8797.

[16] BAO J, CHEN D, WEN F, et al. Fine-grained image generation through asymmetric training [C] // IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2745-2754.

[17] WANG Y, GONZALEZ-GARCIA A, VAN DE WEIJER J, et al. SDIT: scalable and diverse cross-domain image translation [C] // ACM International Conference on Multimedia. Nice, France: Association for Computing Machinery, 2019: 1267-1276.

[18] ANOOSHEH A, AGUSTSSON E, TIMOFTE R, et al. Unrestrained scalability for image domain translation [C] // IEEE Conference on Computer Vision and Pattern Rec-

ognition. Salt Lake City, USA: IEEE, 2018: 783-790.
[19] YU X, CAI X, YING Z, et al. Image-to-image translation by a single-generator network using multiple generative adversarial learning[C]// Asian Conference on Computer Vision. Perth, Australia: Springer, 2019: 341-356.
[20] HEUSEL M, RAMSAUER H, UNTERTHINER T, et al. Gans trained by a two time-scale update rule converge to a local Nash equilibrium[J]. Advances in Neural Information Processing Systems, 2017, 30:6629-6640.
[21] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric [C]// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 586-595.

HU Zhentao, born in 1979. He received his Ph. D degree in Control Science and Engineering from Northwestern Polytechnical University in 2010. He also received his B. S. and M. S. degrees from Henan University in 2003 and 2006 respectively. Now, he is a professor of School of Artificial Intelligence, Henan University. His research interests include complex system modeling and estimation, multi-source information fusion and image translation, etc.