SingleGAN: Image-to-Image Translation by a Single-Generator Network using Multiple Generative Adversarial Learning

Xiaoming Yu⋆, Xing Cai⋆, Zhenqiang Ying, Thomas Li, and Ge Li

School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University, Shenzhen, China

Abstract. Image translation is a burgeoning field in computer vision where the goal is to learn the mapping between an input image and an output image. However, most recent methods require multiple generators for modeling different domain mappings, which are inefficient and ineffective on some multi-domain image translation tasks. In this paper, we propose a novel method, SingleGAN, to perform multi-domain image-to-image translations with a single generator. We introduce the domain code to explicitly control the different generative tasks and integrate multiple optimization goals to ensure the translation. Experimental results on several unpaired datasets show superior performance of our model in translation between two domains. Besides, we explore variants of SingleGAN for different tasks, including one-to-many domain translation, many-to-many domain translation and one-to-one domain translation with multimodality. The extended experiments show the universality and extensibility of our model.

1 Introduction

Recently, more and more attention has been paid to image-to-image translation due to its exciting potential in a variety of image processing applications [1]. Although existing methods show impressive results on one-to-one mapping problems, they need to build multiple generators for modeling multiple mappings, which are inefficient and ineffective in some multi-domain and multi-model image translation tasks. Intuitively, many multi-mapping translation tasks are not independent and share some common features such as scene contents in transformations between different seasons. By sharing a network between related tasks, we can enable our model to generalize better on each separated task. In this paper, we propose a single-generator generative adversarial network (GAN), called SingleGAN, to solve multi-mapping translation tasks effectively and efficiently. To indicate a specific mapping, we introduce the domain code as an auxiliary input to the network. Then we integrate multiple optimization goals to learn each specific translation.

⋆ indicates equal contribution

As illustrated in Fig. 1, the base SingleGAN model is utilized to learn the bijection between two domains. Since each domain dataset is not required to have the label of other domains, SingleGAN can make full use of the existing different datasets to learn the multi-domain translation.

To explore the potential and generality of SingleGAN, we also extend it to three cross domains translation tasks, which are more complex and practical. The first variant model tries to address the one-to-many domain translation task that processes a source domain input to a different target domains, such as the image style transfer. The Second model explore the many-to-many domain translation task. Unlike the recent method [2] requires detailed annotation of category information to training the auxiliary classifier, we use multiple adversarial objects to help network captures different domain distribution separately. It means that SingleGAN is capable of learning multi-domain mappings by weakly supervised learning since we do not need to label all the training data with detailed annotation. The third variant model attempts to increase the generative diversity by introducing attribute latent code. A similar idea is used in BicycleGAN [3] to address the multimodal translation problem. Our third model can be considered a generalization of BicycleGAN towards unpaired image-to-image translation.

To summarize, our contributions are as follows:

- We propose SingleGAN, a novel GAN that utilizes a single generator and a group of discriminators to accomplish the unpaired image-to-image translation.
- We show the generality and flexibility of SingleGAN by extending it to achieve three different kinds of translation tasks.
- Experimental results demonstrate that our approach is more effective and general-purpose than several state-of-art methods.

2   Related work

2.1   Generative Adversarial Networks

Influenced by a zero-sum game, a typical GAN model consists of two modules: a generator and a discriminator. While the discriminator learns to distinguish between real and fake samples, the generator learns to generate fake samples that are indistinguishable from real samples. The GANs have shown impressive results in various computer vision tasks such as image generation, image editing [4] and representation learning [5]. Recently, GAN-based conditional image generation has also been actively studied. To be specific, the various of extension GANs have achieved good results in many generation tasks such as image inpainting [6], super-resolution [7], text2image[8], as well as to other domains such as videos [9] and 3D data[10]. In this paper, we propose a scalable GAN framework to achieve image translation based on conditional image generation.

2.2   Image-to-Image Translation

The idea of image-to-image translation goes back to Image Analogies [11], in which Hertzmann *et al.* proposed a network to transfer the texture information from a source modality space onto a target modality space. Image-to-image translation has received more attention since the flourishing growth of GANs. The pioneering work, Pix2pix [1] uses cGAN[12] to perform supervised image translation from paired data. As those methods adopt supervised learning, sufficient paired data are required to train the network. However, preparing paired images can be time-consuming and laborious (*e.g.* artistic stylization) and even impossible for some applications (*e.g.* male to female face transfiguration). To address this issue, For example, CycleGAN [13], DiscoGAN [14] and DualGAN [15] introduce a cycle-consistency constraint, which widely used in visual tracking [16] and language domain [17], to learn convincing mappings across image domains from unpaired images. Based on a shared-latent space assumption, UNIT [18] extends the Coupled GAN [19] to learn a joint distribution of different domains without paired images. FaderNet [20] is also successful in the controlling of attributes by adding the discriminator to the latent space. Even though, these methods have promoted the development of one-to-one mapping image translation, they have limitations in scalability for multi-mapping translation. By introducing an auxiliary classifier in the discriminator, StarGAN [2] achieved translation among different facial attributes with a single generator. However, this method may learn an inefficient domain mapping when the attribute labels are not sufficient for training the auxiliary classifier even if it introduces a mask vector.

3   Base Model

The main architecture is shown in Fig. 1. In order to take advantage of the correlation between two related tasks, SingleGAN adopts a single generator to achieve bi-direction translation.

The goal of the model is to learn a mapping $G : A \leftrightarrow B$. By adding the domain code, G is redefined as

$$
\begin{aligned}
x_{\mathrm{B}}^{fake} &= G(x_{\mathrm{A}}, z_{\mathrm{B}}), \\
x_{\mathrm{A}}^{fake} &= G(x_{\mathrm{B}}, z_{\mathrm{A}}),
\end{aligned}
\tag{1}
$$

where x_*^{fake} is the fake sample generated by the generator, sample x_* belongs to the set of domain χ_* and $z_{\mathrm{A}}, z_{\mathrm{B}}$ are domain code for domain A and domain B respectively.

3.1   Domain Code Injection

For capturing the distribution of different domains with a single generator, it is necessary to indicate the mapping with auxiliary information. Therefore, we

introduce the domain code z to label the different mapping in the generator. The domain code is constructed as a one-hot vector and similar to the latent code that is widely used to indicated the attributes of generated image [21, 3].

Recent work [22] shows that different injection methods of latent code will effect the performance of generation model. So we adopt the central biasing instance normalization (CBIN) proposed in [22] to inject the domain code in our SingleGAN model. CBIN is defined as

$$\text{CBIN}(x_i) = \frac{x_i - \text{E}[x_i]}{\sqrt{\text{Var}[x_i]}} + \tanh(f_i(z)), \tag{2}$$

where i is the index of feature maps, f_i is affine transformation applied on the domain code z and its parameters are learned for each feature map x_i in one layer. The CBIN aims to adjust the different distributions of input feature maps adaptively with learnable parameters, which makes the domain code able to manage the different tasks. Meanwhile, the distance between the different distributions of input data is also trainable, which means that the coupling degree of different tasks is determined by the model itself. This advantage enables different tasks to share parameters better, so as to promote each other better.

3.2   Loss Functions

GAN Loss. Since our single generator has multi-domain outputs, we set up adversarial objectives for each target domain and employ a group of discriminators. The corresponding discriminator is used to identify the generated images in one domain. The adversarial loss is defined as

$$\begin{aligned}
\mathcal{L}_{adv}(G, D_\text{A}) &= \mathbb{E}_{\chi_\text{A}}[\log(D_\text{A}(x_\text{A}))] + \mathbb{E}_{\chi_\text{B}, z_\text{A}}[\log(1 - D_\text{A}(G(x_\text{B}, z_\text{A})))], \\
\mathcal{L}_{adv}(G, D_\text{B}) &= \mathbb{E}_{\chi_\text{B}}[\log(D_\text{B}(x_\text{B}))] + \mathbb{E}_{\chi_\text{A}, z_\text{B}}[\log(1 - D_\text{B}(G(x_\text{A}, z_\text{B})))].
\end{aligned} \tag{3}$$

By optimizing multiple generative adversarial objectives, the generator recovers different domain distributions that indicated by domain code z.
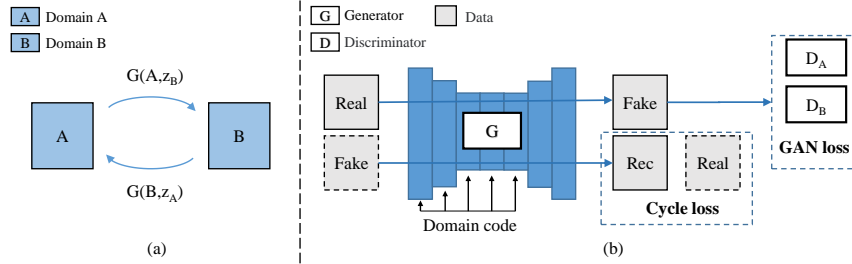

Fig. 1. (a) The base model contains two mapping direction: A → B and B → A. (b) Our base model architecture, which consists of a generator and a group of discriminators.

Cycle Consistency Loss. Although the above GAN loss can complete domain translation, highly under-constrained mapping often leads to a mode collapse. There are many possible mappings that can be inferred without the use of pairing information.

To reduce the space of the possible mappings, we use the cycle-consistency constraint [13, 2] in the training stage. The cycle consistency loss is defined as $x_A \approx G(G(x_A, z_B), z_A)$ and $x_B \approx G(G(x_B, z_A), z_B)$ in our model. The formula of the cycle consistency loss is defined as

$$\mathcal{L}_{cyc}(G) = \mathbb{E}_{\chi_A}\big[\, \|x_A - G(G(x_A, z_B), z_A)\|_1 \,\big] + \\ \mathbb{E}_{\chi_B}\big[\, \|x_B - G(G(x_B, z_A), z_B)\|_1 \,\big], \tag{4}$$

where $\|\cdot\|_1$ denotes ℓ_1 norm.

Full Objective. The final objective function is defined as

$$G^* = arg \min_{G} \max_{D_A, D_B} \sum_{i \in \{A,B\}} \mathcal{L}_{adv}(G, D_i) + \lambda_{cyc} \cdot \mathcal{L}_{cyc}(G) \tag{5}$$

where λ_{cyc} controls the relative importance of the two objectives.

4   Extended models

To explore the potential and generality of SingleGAN, based on the above model, we extend three variants of our model to different tasks: one-to-many domain translation, many-to-many domain translation and one-to-one domain translation with multi-modal mapping.

4.1   One-to-Many Domain Translation

The first trial in Fig.2(a) applies to unidirectional tasks, for example multi-task detection and image multi-style transfer. As far as image style transfer is concerned, different style transfer from a single input image is a representative
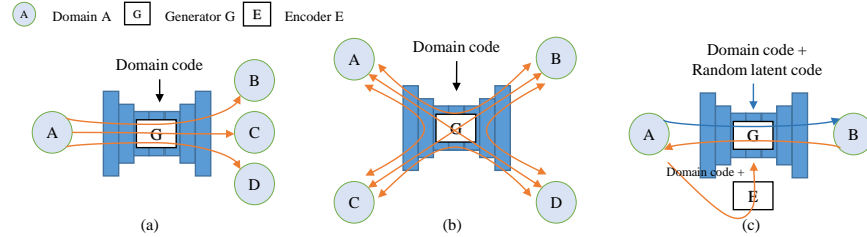

Fig. 2. Extended models: (a) one-to-many domain translation, (b) many-to-many domain translation, (c) one-to-one domain translation with multi-modal mapping.

task of sharing semantics. Our model shares the same texture information of the input image and apply different styles on it. Compared with traditional image style transfer methods, which learn mapping between one content image and one style image, our model learn different mappings between image collections. Such one-to-three translation task are shown in Fig.2(a), the $\mathcal{L}_{adv}$ is redefined as

$$\mathcal{L}_{adv}(G, D_{\{B,C,D\}}) = \sum_{i \in \{B,C,D\}} (\mathbb{E}_{\chi_i}[\log(D_i(x_i))] + \mathbb{E}_{\chi_A, z_i}[\log(1 - D_i(G(x_A, z_i)))]),$$

(6)

where A is the source domain and B, C, D are target domains. In the meantime, the cycle consistency loss is modified to

$$\mathcal{L}_{cyc}(G) = \sum_{i \in \{B,C,D\}} (\mathbb{E}_{\chi_A} \big[\|G(G(x_A, z_i), z_A)\|_1 \big] + \mathbb{E}_{\chi_i} \big[\|G(G(x_i, z_A), z_i)\|_1 \big]).$$

(7)

4.2   Many-to-Many Domain Translation

As illustrated in Fig.2(b), the second variation shows images in multi-domain translating to each other. In this model, our goal is to train a single generator that can learns mappings among multiple domains and realize the mutual transformation of multiple domains. For a four-domain transfer instance, the $\mathcal{L}_{adv}$ is redefined as

$$\mathcal{L}_{adv}(G, D_{\{A,B,C,D\}}) = \sum_{i,j \in \{A,B,C,D\}} (\mathbb{E}_{\chi_i}[\log(D_i(x_i))] +$$
$$\mathbb{E}_{\chi_j, z_i}[\log(1 - D_i(G(x_j, z_i)))]),$$

(8)

and the $\mathcal{L}_{cyc}$ also needs to be modified like the extended model (a).

4.3   One-to-One Domain Translation with Multi-Modal Mapping

To address the multi-modal image-to-image translation problem with unpaired data, we introduce the third variant as show in Fig.2(c). Inspired by Bicycle-GAN [3], we introduce the VAE-like encoder to extract feature latent code c for indicating the translation mapping. Although there is no paired data for supervised learning of the encoder, we utilize the cycle consistency to relax the constraint. During training time, we random sample latent code from a standard Gaussian distribution to indicate the multimodality. Then we concatenate the latent code c into the domain code z to indicate the final mapping. To constraint the image content and encourage the mapping from the latent code, we use the latent code encoded from the source image and the generated image to reconstruct the source image. Due to the introduction of a VAE-like encoder, the latent distribution encoded by the encoder is encouraged to be close to random Gaussian

$$\mathcal{L}_{KL}(E) = \mathbb{E}_{\chi_B}[D_{KL}(E(x_B)||\mathcal{N}(0, I))],$$

(9)

where $D_{KL}(p||q) = -\int p(c) \log \frac{p(c)}{q(c)} dc$. To enforce the generator utilizing the latent code, the reconstruction latent code loss is also used

$$\mathcal{L}_1^{VAE}(G, E) = \mathbb{E}_{\chi_A, c}||c - E(G(x_A, c))||_1. \tag{10}$$

Combining these two losses with the loss of base model, our model can solve the problem of the lack of diversity in unpaired image translation. Notice that we only discuss the translation of A-to-B, as the mapping of B-to-A is similar and concurrent during training time.

5 Implementation

5.1 Network Architecture

As in $[2, 13, 23, 22]$, our generator G uses the ResNet[24] structure with an encoder-decoder framework, which contains two stride-2 convolution layers for downsampling, six residual blocks and two stride-2 transposed convolution layers for upsampling. We replace all normalization layers except upsampling layers with CBIN layers. For the discriminators D, we use two discriminators [1] to discriminate the real and fake images in different scales. For the experiment of multi-modal SingleGAN, the encoder model E adopts the ResNet structure [3]. We equip the encoder with CBIN, so it can also extract the latent information from different domain images. Code and model are available at https://github.com/Xiaoming-Yu/SingleGAN.

5.2 Training Details

For all experiments, we train all models with Adam optimizer [25], setting $\beta_1 = 0.5$, $\lambda_{cyc} = 10$, learning rate of 0.001. In the extended multi-modal networks as shown in Fig.2(c), the weights for $\mathcal{L}_{KL}(E)$ and $\mathcal{L}_1^{VAE}(G, E)$ are 0.1 and 0.5 respectively. To generate higher quality results with stable training, we replace the negative log likelihood objective by a least-squares loss [26].

5.3 Structure of Domain Code

As mentioned in Sect. 3.1, we use the one-hot vector to present the domain code z. To the base model, we use the 2 dimensional domain code for indicating the mapping between domain A $\leftrightarrow$ B. For the one-to-many and many-to-many translation instances illustrated in Fig. 2, the domain code is 4 dimension and represents 4 different domains. In the third variant, the 8 dimensional latent code c is also used for multimodal image translation in the specific domain that indicated by the 2 dimensional domain code.
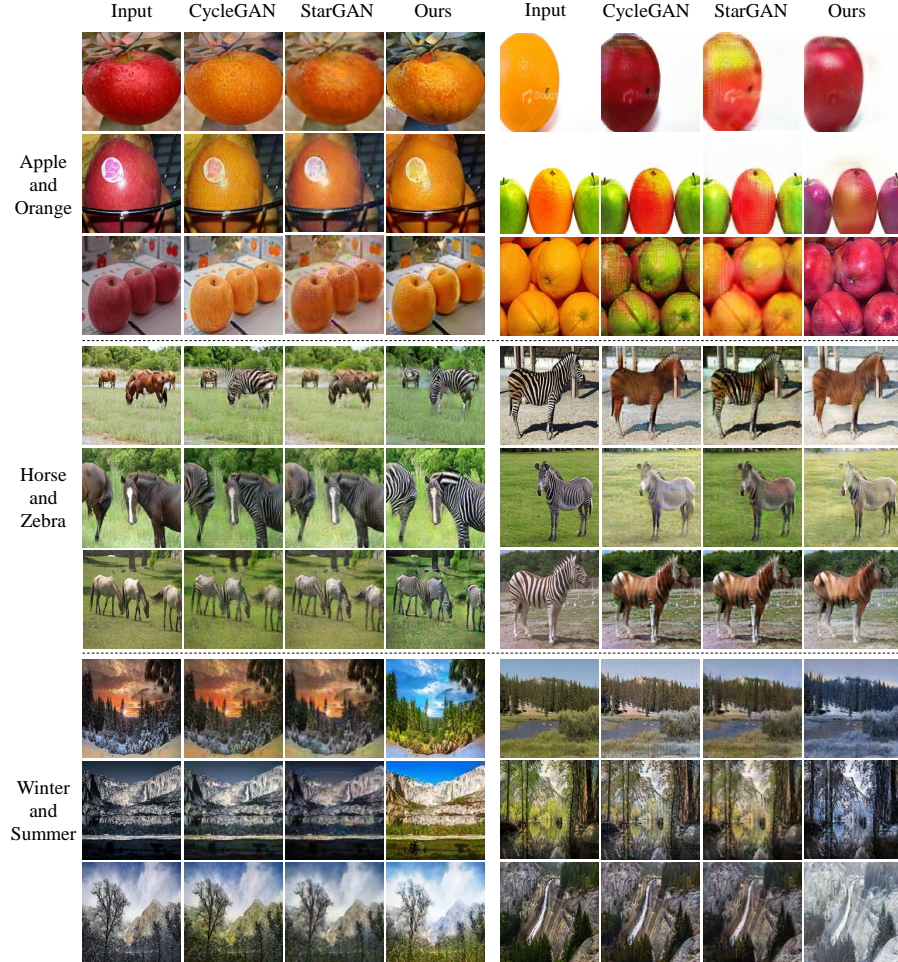
Fig. 3. Visualization and comparison on three unpaired datasets. The first four columns show mapping from domain A to domain B while the next four columns show mapping from domain B to domain A.

6 Experiments

6.1 Datasets

To evaluate the base model, we use three unpaired datasets: Apple↔Orange, Horse↔Zebra, and Summer↔Winter [13]. As for the three extended models, we use Photo↔Art [13] for one-to-many translation, Transient-Attributes [27] for many-to-many translation, and Edges↔Photos [1] for one-to-one multi-model translation. All of the images are scale to 128×128 resolution.

	horse&zebra	apple&orange	summer&winter
image number	240	480	500
real image	0.985	0.978	0.827
CycleGAN	0.850	0.935	0.644
StarGAN	0.858	**0.970**	0.689
SingleGAN	**0.859**	0.966	**0.742**

Table 1. The classification accuracy for three datasets. Best results are in boldface.

	real image		CycleGAN		StarGAN		SingleGAN	
	A	B	A	B	A	B	A	B
horse&zebra	1.198	1.177	**1.141**	**1.133**	1.083	1.081	1.112	1.128
apple&orange	1.205	1.499	1.106	1.144	1.128	1.132	**1.152**	**1.164**
summer&winter	1.272	1.824	1.223	1.189	1.209	1.173	**1.258**	**1.208**

Table 2. The perceptual distance for three datasets. Best results are in boldface. Here A represents horse, apple or summer, and B represents zebra, orange or winter.

6.2   Baselines

To compare the performance of our SingleGAN model, we adopt the Cycle-GAN [13] and StarGAN [2] as our baseline models. CycleGAN uses cycle loss to learn the mapping between two different domains. To achieve cycle consistency, CycleGAN requires two generators and discriminators for two different domains. To unify multi-domain translation with single generator, StarGAN introduces an auxiliary classifier trained on image-label pairs in its discriminator to assist the generator to learn the mapping cross multiple domains. We compare our method with CycleGAN and StarGAN on two domains translation tasks.

6.3   Base Model Comparison

In this section, we evaluate the performance of different models. It should be noted that both SingleGAN and StarGAN use a single generator for two domain image translation and CycleGAN uses two generators to achieve the similar mappings.

The qualitative comparison is shown in Fig. 3. We can observe that all these models present pleasant results in the simple case such as the apple to orange transformation. In the translation with complex scene, the performance of these models are degraded especially StarGAN. The possible reason is that the generator of StarGAN introduces the adversarial noise to fool the auxiliary classifier and fails to learn the effective mapping. Meanwhile, we can observer that SingleGAN presents the best results in most cases.

To judge the quality of the generated image quantitatively, we evaluate the classification accuracy of the images generated by these three models at first. We train three Xception [28] based binary classifiers for each image datasets.

The baseline is the classification accuracy in real images. Higher classification accuracy means that the generated images may more easy to distinguish. Second, we compare the domain consistency between real images and generated images by computing average distance in feature space. A similar idea is used for calculating the diversity of multi-modal generation task [3, 22]. we use the cosine similarity to evaluate the perceptual distance in the feature space of the VGG-16 network [29] that pre-trained in ImageNet [30]. We sum across the five convolution layers preceding the pool layers. The larger the value, the more similar between two images. In the test stage, we randomly sample the real image and the generated image from same domain to make up the data pair. Then we compute the average distance between 2,000 pairs. The baseline is computed by sampling from 2,000 pairs of real images.

The quantitative results are shown in Table 1 and Table 2. Both SingleGAN and CycleGAN produce the quantitative results that comply with qualitative

Fig. 4. The one-to-many translation results of multi-style image generation. The first column is the real images and the rest of columns are the translation results that represent different artistic styles.

performance. In contrast, StarGAN gets a higher classification accuracy but the poor performance in domain consistency. It validates our conjecture that the generated image of StarGAN may have the adversarial noise of fooling the classifier in some complex scenes. In StarGAN, the discriminator learns to tell the image is real or fake without considering the classification result of the image while the generator learns to fool the discriminator with an image that can be corrected classified by the auxiliary classifier. So the generator may not get enough encouragement if it generates the adversarial noise to the image. For example, on the task of Summer↔Winter, although the input summer image is expected to translate into winter, the generator of StarGAN tends to just add a tiny adversarial noise to the input image so that the discriminator still tell it is real while the classifier classifies it as the winter. As a result, the generated images will look unchanged to human but win high classification scores. This issue does not exist in SingleGAN and CycleGAN since these models optimizes different mappings with different discriminators. The main difference between SingleGAN and CycleGAN is the number of generators. As shown in Fig. 3 and Table 1, 2, we can observer that SingleGAN has the capacity to learn multiple mapping
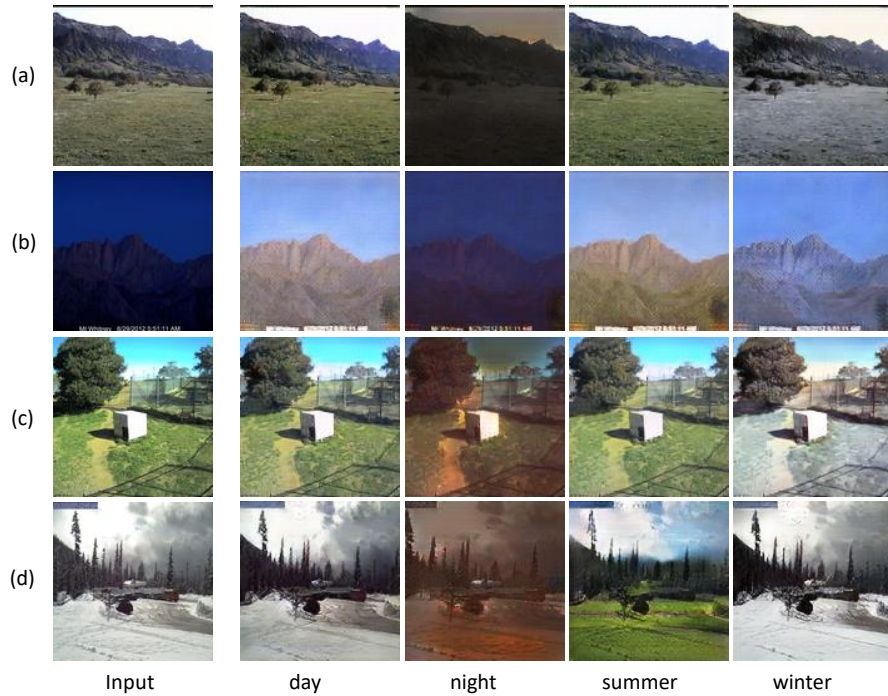

Fig. 5. The many-to-many domain translation results. The first column is the input images from different domains: (a) day, (b) night, (c) summer, (d) winter. The remaining columns are the transfer results.

Fig. 6. The multi-modal translation results. The first column shows the input and the other columns show randomly generated samples.

without performance degradation. By sharing the generator for different domain translation, SingleGAN can see more training data form different domains to learn the shared semantics and improve the performance of the generator.

6.4 Extended Model Evaluation

To explore the potential of SingleGAN, we test the extended models on three different translation tasks.

For one-to-many image translation, we perform the multi-style transfer to evaluate the model performance. Photo↔Art [13] dataset contains three artistic styles (500 images of Monet, 584 images of Cezanne and 401 images of Van Gogh) and 1000 real photos. The results are shown in Fig. 4. We can observe that the generate images have similar artistic styles when we perform same mapping while different styles are distinguishable.

For multi-domain translation, we choose four outdoor scenes in Transient-Attributes dataset [27] to evaluate the model: 'day', 'night', 'summer' and 'winter'. It should be note that the multiple domains do not have to be independent, *e.g.* the subset 'day' contains summer and winter. The training data for each domain do not need to consider other domain information. As shown in Fig. 5, SingleGAN is competent at the transformation from all domains, though the dataset has incomplete labels.

The final experiment is to verify the multi-modal performance of SingleGAN after introducing the attribute latent code. The dataset adopted is edge2shoes [31]. Please Note that this experiment is performed under the settings of unpaired data. The experimental result in Fig. 6 shows that SingleGAN has the ability to learn multimodal mapping under the unsupervised learning.
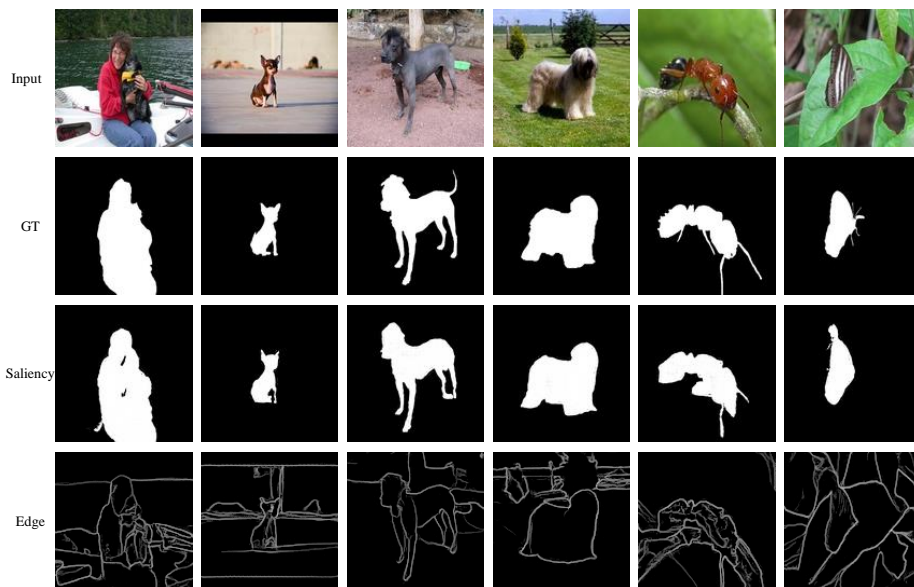
Fig. 7. Saliency and edge detection results of SingleGAN under paired data setting.

6.5   Translation under Paired Data Setting

Although the above experiments have the unpaired data assumption, SingleGAN can also perform multi-domain image translation with paired data by replacing the cycle consistency loss $\mathcal{L}_{cyc}$ with ℓ_1 reconstruction loss $\mathcal{L}_{rec}$.

Here we use the salient object dataset DUTS-TR [32] and BSDS500 edge dataset [33] to perform one-to-many image translation. Specify the real image as domain A, salient images as domain B and edge images as domain C. Then the $\mathcal{L}_{rec}$ can define as

$$
\begin{aligned}
\mathcal{L}_{rec}(G) =&\mathbb{E}_{\chi_{\mathrm{A}},\chi_{\mathrm{B}}}\big[\,\|x_{\mathrm{B}} - G(x_{\mathrm{A}}, z_{\mathrm{B}})\|_1\,\big]+ \\
&\mathbb{E}_{\chi_{\mathrm{A}},\chi_{\mathrm{C}}}\big[\,\|x_{\mathrm{C}} - G(x_{\mathrm{A}}, z_{\mathrm{C}})\|_1\,\big].
\end{aligned}
\tag{11}
$$

The results in Fig. 7 demonstrate the effectiveness of SingleGAN.

6.6   Limitations and Discussion

Although SingleGAN can achieve multi-domain image translations, multiple adversarial learning needs to be done simultaneously. This constraint makes Single-GAN only be able to learn limited domain translation at a time since our storage is limited. So it is valuable to explore the transfer learning for the existing models. Besides, the capacity of the network to learning different mappings is also an important problem. We also observe that integrate suitable tasks for one single

model may improve the performance of the generator. But what kind of tasks can promote each other remains to be explored in the future work. Nonetheless, we think the method proposed in this paper is valuable for exploring the multi-domain generation works.

7   Conclusion

In this paper we introduce a single generator based model, SingleGAN, for learning multi-mapping image-to-image translation. By introducing multiple adversarial learning for the generator, SingleGAN is able to learn a variety of mappings effectively and efficiently. Contrastive experimental results show quantitatively and qualitatively that our approach is effective in many image translation tasks. Furthermore, to improve the versatility and generality of the model, we present three variants of SingleGAN for different tasks: one-to-many domain transfer, many-to-many domain transfer and one-to-one domain transfer with varying attributes. The experiment results demonstrate these variants improve the corresponding translation effectively.

References

1. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on. (2017)
2. Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J.: Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. Computer Vision and Pattern Recognition (CVPR), 2018 IEEE Conference on (2018)
3. Zhu, J.Y., Zhang, R., Pathak, D., Darrell, T., Efros, A.A., Wang, O., Shechtman, E.: Toward multimodal image-to-image translation. In: Advances in Neural Information Processing Systems 30. (2017)
4. Brock, A., Lim, T., Ritchie, J.M., Weston, N.: Neural photo editing with introspective adversarial networks. ICLR (2017)
5. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. ICLR (2016)
6. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A.: Context encoders: Feature learning by inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 2536–2544

7. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on. (2017)

8. Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., Lee, H.: Generative adversarial text to image synthesis. In: International Conference on Machine Learning. (2016) 1060–1069

9. Vondrick, C., Pirsiavash, H., Torralba, A.: Generating videos with scene dynamics. In: Advances In Neural Information Processing Systems. (2016) 613–621

10. Wu, J., Zhang, C., Xue, T., Freeman, B., Tenenbaum, J.: Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In: Advances in Neural Information Processing Systems. (2016) 82–90

11. Hertzmann, A., Jacobs, C.E., Oliver, N., Curless, B., Salesin, D.H.: Image analogies. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, ACM (2001) 327–340

12. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)

13. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networkss. In: Computer Vision (ICCV), 2017 IEEE International Conference on. (2017)

14. Kim, T., Cha, M., Kim, H., Lee, J.K., Kim, J.: Learning to discover cross-domain relations with generative adversarial networks. In: International Conference on Machine Learning. (2017) 1857–1865

15. Yi, Z., Zhang, H., Tan, P., Gong, M.: Dualgan: Unsupervised dual learning for image-to-image translation. In: IEEE International Conference on Computer Vision. (2017) 2868–2876

16. Sundaram, N., Brox, T., Keutzer, K.: Dense point trajectories by gpu-accelerated large displacement optical flow. In: European conference on computer vision, Springer (2010) 438–451

17. Brislin, R.W.: Back-translation for cross-cultural research. Journal of cross-cultural psychology **1** (1970) 185–216

18. Liu, M.Y., Breuel, T., Kautz, J.: Unsupervised image-to-image translation networks. In: Advances in Neural Information Processing Systems. (2017) 700–708

19. Liu, M.Y., Tuzel, O.: Coupled generative adversarial networks. In: Advances in neural information processing systems. (2016) 469–477

20. Lample, G., Zeghidour, N., Usunier, N., Bordes, A., Denoyer, L., et al.: Fader networks: Manipulating images by sliding attributes. In: Advances in Neural Information Processing Systems. (2017) 5969–5978

21. Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., Abbeel, P.: Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In: Advances in Neural Information Processing Systems. (2016) 2172–2180

22. Yu, X., Ying, Z., Li, T., Liu, S., Li, G.: Multi-mapping image-to-image translation with central biasing normalization. arXiv preprint arXiv:1806.10050 (2018)

23. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision, Springer (2016) 694–711

24. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2016) 770–778

25. Kinga, D., Adam, J.B.: A method for stochastic optimization. In: International Conference on Learning Representations (ICLR). (2015)
26. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Smolley, S.P.: Least squares generative adversarial networks. In: 2017 IEEE International Conference on Computer Vision (ICCV), IEEE (2017) 2813–2821
27. Laffont, P.Y., Ren, Z., Tao, X., Qian, C., Hays, J.: Transient attributes for high-level understanding and editing of outdoor scenes. ACM Transactions on Graphics (proceedings of SIGGRAPH) **33** (2014)
28. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. (2016) 1800–1807
29. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. Computer Science (2014)
30. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International Journal of Computer Vision **115** (2015) 211–252
31. Zhu, J.Y., Krähenbühl, P., Shechtman, E., Efros, A.A.: Generative visual manipulation on the natural image manifold. In: European Conference on Computer Vision, Springer (2016) 597–613
32. Zhao, R., Ouyang, W., Li, H., Wang, X.: Saliency detection by multi-context deep learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2015) 1265–1274
33. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. IEEE transactions on pattern analysis and machine intelligence **33** (2011) 898–916