Hindawi
Complexity
Volume 2018, Article ID 4613935, 12 pages
https://doi.org/10.1155/2018/4613935

Research Article

GANs with Multiple Constraints for Image Translation

Yan Gan,[1] Junxin Gong,[1] Mao Ye,[1] Yang Qian,[1] Kedi Liu,[1] and Su Zhang[2]

[1] *School of Computer Science and Engineering, Key Laboratory for NeuroInformation of Ministry of Education, University of Electronic Science and Technology of China, Chengdu 611731, China*
[2] *School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China*

Correspondence should be addressed to Mao Ye; maoye@uestc.edu.cn

Received 30 July 2018; Accepted 22 November 2018; Published 9 December 2018

Academic Editor: Danilo Comminiello

Unpaired image translation is a challenging problem in computer vision, while existing generative adversarial networks (GANs) models mainly use the adversarial loss and other constraints to model. But the degree of constraint imposed on the generator and the discriminator is not enough, which results in bad image quality. In addition, we find that the current GANs-based models have not yet been implemented by adding an auxiliary domain, which is used to constrain the generator. To solve the problem mentioned above, we propose a multiscale and multilevel GANs (MMGANs) model for image translation. In this model, we add an auxiliary domain to constrain generator, which combines this auxiliary domain with the original domains for modelling and helps generator learn the detailed content of the image. Then we use multiscale and multilevel feature matching to constrain the discriminator. The purpose is to make the training process as stable as possible. Finally, we conduct experiments on six image translation tasks. The results verify the validity of the proposed model.

1. Introduction

Image translation [1] is similar to language translation, which converts the input image of source domain to a corresponding image of target domain. For example, inputting an image of the pear and turning it into an image of the apple. There are many methods [1–6] to solve image translation problem, but GANs based methods [1, 7–11] have gained increasing attention in the image translation. In the methods of GANs, they view the input image of source domain as the input of generator, which generates fake samples to deceive discriminator. And then the discriminator is responsible for judgment whether they are real samples of target domain or generated fake samples, in which, the deep convolution or deconvolution neural networks [12–17] are often used to construct the generator or discriminator.

According to whether the datasets are paired or unpaired, the image translation of GANs based methods can be roughly classified into two categories: paired and unpaired image translation. For paired methods [1, 18], they require paired datasets. It is very difficult to prepare paired training datasets in practical applications, whose cost is expensive [19]. To reduce the cost of obtaining paired training datasets, [20–22] propose the unpaired methods, which are unsupervised domain translation methods [23].

However, these paired or unpaired models are inadequate in generating the detailed information of images. To bring the better result of image translation, it is still a challenge task because of the following problems: (1) how to constrain the generator to generate the detailed content in the image and (2) how to stabilize the training process to obtain better performance of model, such as generated images and generalization performance.

On the one hand, there is the constraint generator. These GANs based methods gradually close to real data distribution by adjustment the generator parameters. In other words, this type of GANs based methods does not need to be pre-modelled, so they are too free for modelling. For the situation that there are many pixels in the image, this type of methods has an uncontrollable problem. To address this problem, [24] proposes constraining generator and discriminator by adding condition variable y. Inspired by adding constraint, many researchers have come up with models from different perspectives. The first is to add text or semantic information to model. Reference [25] adds text information into the cascaded GANs, which generates high definition of image from text. Reference [26] proposes structural GANs, which incorporates semantic information into a conditional generative model. The second is to add regularization term. Reference [27] combines mutual information to conduct adversarial modelling, which actually adds a regularization term of mutual information. Reference [7] brings in cyclic consistency constraint to achieve cross domain image translation. Reference [28] takes advantage of Wasserstein distance and the gradient penalty to conduct adversarial modelling. Then, due to the fact that Nash equilibrium points of original GANs are not asymptotically stable, in order to overcome this difficulty, [29] adds a regularization term in gradient update and proves that the equilibrium points of the original GANs are locally asymptotically stable after adding it.

By looking at the similar objects in real life, we find that they have a certain similarity in appearance or structure. In addition, there are few methods to constrain generator by adding an auxiliary domain [30]. Inspired by this, we add the auxiliary domain, which is used to help the generator learn the detailed information in the image during image translation.

On the other hand, there is the stabilization training process. To obtain better generated images, many researchers have put forward their ideas on stabilize training process. Firstly, there is the missing model perspective. Reference [31] analyses the problem of original GANs objective function, which leads to easily the gradient vanishing and missing model when training GANs. Then [28] achieves the improvement of training process. Since the convergence of original GANs training is not proved, [32] proposes a two time scale update rule to train GANs, which can make the training process converge to a local Nash equilibrium. Secondly, there is the probabilistic perspective. Reference [33] studies the distribution of the squared singular values of the input-output Jacobian of the generator. Thirdly, there is the multiscale discriminator. Reference [34] proposes a multiscale discriminator to stabilize training process and generates the high resolution image. Inspired by these discussions, in order to stabilize the training process and improve the discriminating ability of the discriminator. We make use of ideas of the multiscale and multilevel to constrain the discriminator, which is implemented by deep convolutional neural networks [15, 35–37].

In this paper, we focus on the unpaired image translation task based on the method of GANs. To try to solve two problems: (1) In GANs based methods [7–9, 20, 21], the generator lacks control over the detailed information in the process of image generation. We try to constrain generator to generate more detailed content during image translation. (2) To obtain better generated images or generalization performance of model, we stabilize the process of GANs training as much as possible.

To solve the aforementioned concerns, we propose a novel unpaired image translation framework from the perspective of simultaneous constraint generator and discriminator. On the one hand, in order to constrain generator, we add the auxiliary domain to model, which can help generator learn the detailed information in the image. Then we combine the additional auxiliary domain with the domains to be learned to model and design multiple generators and discriminators for image translation. On the other hand, to improve the ability of discriminator or stabilization training process, we use multiscale and multilevel feature matching to constrain discriminator. Finally, we use multiple generative losses, multiscale discriminator losses, multilevel feature matching losses, and full cycle consistency losses to constrain the proposed model.

Our main contributions are

(1) We propose an unpaired, multiscale and multilevel feature matching generative adversarial networks (MMGANs) by adding auxiliary domain to achieve cross domain image translation.

(2) We modify the original GANs model from the perspective of simultaneous constraint generator and discriminator. In our model, we add an additional auxiliary domain as auxiliary information to help generator learn the details information during generative images. And we constrain the discriminator by multiscale and multilevel feature matching losses.

(3) Finally, we conduct experiments on the six tasks of image translation. According to the proposed evaluation method and the generated images, the experimental results show that our model has better performance.

The rest of this paper is organized as follows. Section 2 describes the proposed method and the detailed model. Section 3 provides the results and discussion. Section 4 concludes this paper.

2. Materials and Methods

In this paper, inspired by the constraint generator, cycle consistency of CycleGAN, multiscale discriminator and multilevel feature matching, we design the MMGANs model. On the one hand, we constrain generator by adding an additional domain [30] as ancillary information. On the other hand, we make use of multiscale and multilevel feature matching to constrain discriminator. To achieve this model, we specifically design multiple generator losses and multiscale discriminator losses, full cycle constraint losses, and multilevel feature matching losses.

2.1. Formulation Description. We focus on unpaired image translation problem. For the convenience of the following description, we suppose that image translation is implemented in the domains X and Y. The added auxiliary domain is Z. We have n_x, n_y, and n_z samples from the X, Y, and Z domains. In order to implement the proposed model, we design six generators (namely, G1: $X \longrightarrow Y$, G2: $Y \longrightarrow Z$, G3: $Z \longrightarrow X$, F1: $Y \longrightarrow X$, F2: $Z \longrightarrow Y$, and F3: $X \longrightarrow Z$) and three discriminators (namely, D_{Xi}, D_{Yi}, and D_{Zi}). In particular, to stabilize the training process, our discriminator D_{Xi} has multiscale ($i \in [1, ms]$). Moreover, to improve the quality of generated images, we add the multilevel feature matching in the discriminator.

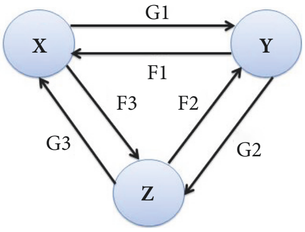


Figure 1: The framework of MMGANs model. In this model, there are six generators (G1, F1, G2, F2, G3, and F3) and the letters 'X', 'Y', and 'Z' represent different domains.

For example, the generator G3 generates samples of the target domain X. Its input is Z domain. The generator F1 generates samples of the target domain X, whose input is Y domain. Then the discriminator D_{Xi} determines whether the generated samples $x \in F1(Y)$, $x \in G3(Z)$, and the real samples $x \in X$ are real or fake. The discriminators D_{Yi} and D_{Zi} are similar to D_{Xi}. The framework of MMGANs model and its discriminator model are shown in Figures 1 and 2.

2.2. Adversarial Loss with Multiscale and Multilevel. The proposed MMGANs model is inspired by [30], which consists of multiple generators and discriminators. In the proposed model, we use multiscale and multilevel to constrain the discriminator. In this way, the training process can be stabilized and the discriminator pays attention to the identification of the detailed content of the image.

Concretely, there are three inputs x, $F1(y)$, and $G3(z)$. It needs to distinguish real or fake sample in the three inputs for the discriminator. To stabilize the training process, we design multiscale discriminator losses. The adversarial losses with multiscale discriminator are as follows:

$$\begin{aligned} &L^1_{Dis}(G3, F1, D_{Xi}, Z, Y, X) \\ &\quad = \sum_{i=1}^{ms} \left(E_{x\sim pdata(x)}[\log D_{Xi}(x)] \right. \\ &\quad + E_{z\sim pdata(z)}[\log(1 - D_{Xi}(G3(z)))] \\ &\quad \left. + E_{y\sim pdata(y)}[\log(1 - D_{Xi}(F1(y)))]\right) \end{aligned} \quad (1)$$

$$\begin{aligned} &L^2_{Dis}(G1, F2, D_{Yi}, X, Z, Y) \\ &\quad = \sum_{i=1}^{ms} \left(E_{y\sim pdata(y)}[\log D_{Yi}(y)] \right. \\ &\quad + E_{x\sim pdata(x)}[\log(1 - D_{Yi}(G1(x)))] \\ &\quad \left. + E_{z\sim pdata(z)}[\log(1 - D_{Yi}(F2(z)))]\right) \end{aligned} \quad (2)$$

$$\begin{aligned} &L^3_{Dis}(G2, F3, D_{Zi}, Y, X, Z) \\ &\quad = \sum_{i=1}^{ms} \left(E_{z\sim pdata(z)}[logD_{Zi}(z)] \right. \\ &\quad + E_{y\sim pdata(y)}[\log(1 - D_{Zi}(G2(y)))] \\ &\quad \left. + E_{x\sim pdata(x)}[\log(1 - D_{Zi}(F3(x)))]\right) \end{aligned} \quad (3)$$

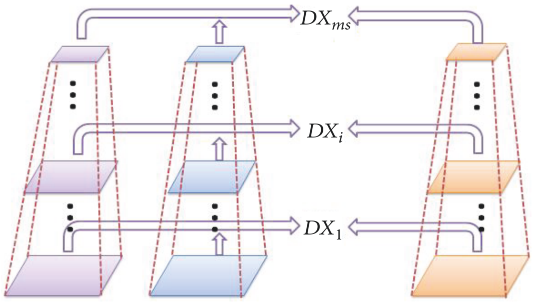


Figure 2: The example of multiscale and multilevel discriminator. There are three columns, which, respectively, represent the real samples X domain, generated samples from Y domain by F1, and generated samples from Z domain by G3. D_{Xi} represents multiscale and multilevel discriminator, $i \in [1, ms]$.

where the subscript Dis means the cases of multiscale discriminator.

To improve further the quality of generated images, we add the constraint on the discriminators, which focus on the detailed content judgment of the generated image. So, we specifically add the multilevel feature matching losses of discriminator. They are as follows:

$$\begin{aligned} &L^1_{feature}(G3, F1, D_{Xi}, Z, Y, X) \\ &\quad = \sum_{i=1}^{ms}\sum_{j=1}^{ml} \left(\left\|D^j_{Xi}(x) - D^j_{Xi}(G3(z))\right\|_F \right. \\ &\quad \left. + \left\|D^j_{Xi}(x) - D^j_{Xi}(F1(y))\right\|_F\right) \end{aligned} \quad (4)$$

$$\begin{aligned} &L^2_{feature}(G1, F2, D_{Yi}, X, Z, Y) \\ &\quad = \sum_{i=1}^{ms}\sum_{j=1}^{ml} \left(\left\|D^j_{Yi}(y) - D^j_{Yi}(G1(x))\right\|_F \right. \\ &\quad \left. + \left\|D^j_{Yi}(y) - D^j_{Yi}(F2(z))\right\|_F\right) \end{aligned} \quad (5)$$

$$\begin{aligned} &L^3_{feature}(G2, F3, D_{Zi}, Y, X, Z) \\ &\quad = \sum_{i=1}^{ms}\sum_{j=1}^{ml} \left(\left\|D^j_{Zi}(z) - D^j_{Zi}(G2(y))\right\|_F \right. \\ &\quad \left. + \left\|D^j_{Zi}(z) - D^j_{Zi}(F3(x))\right\|_F\right) \end{aligned} \quad (6)$$

where the subscript F means the norm in the multilevel feature matching losses.

2.3. Objective Function of MMGANs. Our entire objective function L includes multiple losses of the generators and multiscale discriminators, cycle losses, full cycle losses and multilevel feature losses. The MMGANs model optimal aim is

$$G^*, F^* = \arg \min_{G1,F1,G2,F2,G3,F3} \max_{D_X,D_Y,D_Z} L \quad (7)$$

```
Initialization $setp = 1$, $maxstep = 200000$, generator parameters $\theta_{G1}, \theta_{F1}, \theta_{G2}, \theta_{F2}, \theta_{G3}, \theta_{F3}$,
  discriminator parameters $\theta_{DX}, \theta_{DY}, \theta_{DZ}$. Suppose $Temp = (\theta_{G1}, \theta_{F1}, \theta_{G2}, \theta_{F2}, \theta_{G3}, \theta_{F3}, \theta_{DX}, \theta_{DY}, \theta_{DZ})$.
while $step \leq maxstep$ do
    Sample one sample at a time $x \sim P_X$, $y \sim P_Y$, $z \sim P_Z$.
    The losses G1_loss, F1_loss, G2_loss, F2_loss, G3_loss and F3_loss of generators
    are calculated according to formula (8).
    for i =1 to $ms$ do
        for j =1 to $ml$ do
            The losses DX_loss, DY_loss and DZ_loss of discriminators are respectively
            calculated according to formula (1), (2), (3), (4), (5), (6).
        end for
    end for
    $Temp \longleftarrow$ Adam( $\nabla_{Temp}$ (G1_loss, F1_loss, G2_loss, F2_loss, G3_loss, F3_loss, DX_loss,
    DY_loss, DZ_loss), $Temp$, $\alpha$, $\beta$)
    $step$ ++
end while
```

Algorithm 1: MMGANs. We use default values of $\lambda_{11} = \lambda_{12} = \lambda_{13} = 10$, $\lambda_{21} = 5$, $\lambda_{22} = \lambda_{23} = 0$, $\lambda_{31} = \lambda_{32} = \lambda_{33} = 1$, initial learning rate $\alpha = 0.0002$, and $\beta = 0.5$. Requirement is as follows: There are n_x, n_y, and n_z samples in the X, Y, and Z domains, the batch size is 1, and Adam optimizer hyperparameters are α and β. The multiscale parameter $ms = 1$ and the multilevel parameter $ml = 1$.

And L is

$$L = L_{Dis} + \lambda_1 * L_{cycle} + \lambda_2 * L_{fullcycle} + \lambda_3 * L_{feature} \quad (8)$$

where λ_1, λ_2, and λ_3 are parameters of vectors which are rows (dimension $1*3$). L_{cycle} and $L_{fullcycle}$, respectively, mean cycle loss and full cycle loss [30].

$$L_{feature} = \lambda_{31} * L^1_{feature} + \lambda_{32} * L^2_{feature} + \lambda_{33} * L^3_{feature} \quad (9)$$

We use the Adam optimizer to solve this model and we give a description of the algorithm. See Algorithm 1.

In addition, this article is a further version of our previous work [30], which has been greatly improved in the following places.

Firstly, we modify the discriminator by multiscale and multilevel constraint. In this way, the training process is more stable and the discriminator pays more attention to the matching of features in the generated image. The goal is to make the details of the generated image more realistic. Secondly, we also add the UNIT method [9] as baseline in the experiment. Furthermore, in the performance evaluation part, the AMT evaluation index is further modified to make the evaluation methods more objective and comprehensive.

3. Results and Discussion

Our experimental environment is ubuntu16.04, tensorflow, python3, GPU: GeForce GTX 1080 Ti, and memory: 64GB, CPU: Intel XeonR@2.30GHz $*$ 36.

3.1. Training and Testing Datasets. We use fruits and seasons datasets for experiment in this paper. In which, the fruits dataset includes images of apple, orange, and pear and the seasons dataset covers the images of summer, autumn, and winter. Finally, we adopt the datasets of [30] and resize all images as 128 $*$ 128. The training and testing sets are shown in the Table 1.

Table 1: Training and testing sets. We list the number of image in different data sets.

Name	Training set	Testing set
Apple	995	100
Orange	1019	100
Pear	1036	100
Summer	1048	100
Autumn	1001	100
Winter	1145	100

3.2. Baseline. To evaluate the performance of the MMGANs, we compare our method with CycleGAN [7], DualGAN [8], and UNIT [9].

CycleGAN: It makes use of the adversarial loss and cycle constraint to achieve the unpaired and unsupervised image translation from X domain to Y domain or Y domain to X domain.

DualGAN: It takes advantage of dual learning to constrain generator and discriminator. There are two generators and two discriminators to achieve the unpaired and unsupervised image translation in the DualGAN. This way expands the basic GANs into two coupled GANs.

UNIT: It encodes the images of two domains into a shared latent space through the shared weighted encoder and then realizes unsupervised cross-domain image translation by GANs.

3.3. Performance Evaluation. Amazon Mechanical Turk (AMT) [8] is one of the methods for evaluating generated images by GANs. It requires that the observers only pick out

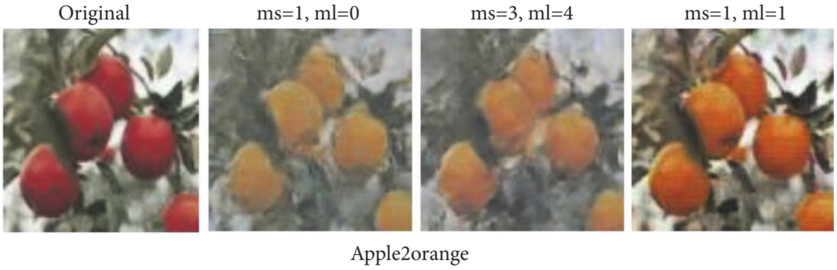


Figure 3: The results of values of different multiscale and multilevel from apple to orange. We show the original image, generated images with different multiscale and multilevel.

Table 2: Experiment settings. Image translation between apple and orange, we add images of the pear domain, namely, apple2orange or orange2apple. Others are similar.

Task	Augmented domain
Apple ⟷ orange	Pear
Apple ⟷ pear	Orange
Orange ⟷ pear	Apple
Summer ⟷ autumn	Winter
Summer ⟷ winter	Autumn
Autumn ⟷ winter	Summer

Table 3: Multiscale and multilevel parameters in the different datasets.

Task	Multi-scale	Multi-level
Apple ⟷ orange	1	1
Orange ⟷ pear	1	0
Pear ⟷ apple	1	0
Summer ⟷ autumn	2	0
Autumn ⟷ winter	2	0
Winter ⟷ summer	2	0

the real or fake image among the testing images. For selecting only real or fake samples, the performance of the model is not well characterized. The generated images can be classified into four categories by the observers: better, worse, both of are bad or good.

Based on the above discussion, we propose a comprehensive way to describe the model. Firstly, let several observers distinguish worse, better, and both of good or bad images and count the numbers of four types of picked images. Then we obtain the mean number of them and respectively calculate their proportions.

Supposing that n_{total} is the sum of the testing images. Compared with CycleGAN, DualGAN, and UNIT models, we, respectively, calculate the values of n_{worse}, n_{better}, $n_{bothgood}$, and $n_{bothbad}$, which represent the mean number of worse, better, and both of good or bad generated images using MMGANs model. Next, the percentages can be calculated, respectively:

$$\begin{aligned} n1 &= \frac{n_{worse}}{n_{total}} \times 100\%, \\ n2 &= \frac{n_{better}}{n_{total}} \times 100\% \\ n3 &= \frac{n_{bothgood}}{n_{total}} \times 100\%, \\ n4 &= \frac{n_{bothbad}}{n_{total}} \times 100\% \end{aligned} \tag{10}$$

Finally, we analyze the model by quantifying indicators $n1$, $n2$, $n3$, and $n4$.

3.4. Results of Testing. To be fair, all experiments are trained 200,000 times and testing by the data sets mentioned in this paper. The experiment focuses on two points: (1) compared with CycleGAN, DualGAN, UNIT, and MMGANs, we show generated images in different datasets and (2) we use our comprehensive performance evaluation method to calculate a quantitative ratio and analyze it.

For the convenience of narration, we set, respectively, six experimental cases on image translation. As shown in the Table 2.

3.4.1. Setting Training Parameters. In this section, we discuss parameter settings about the experiment, which includes λ_{11}, λ_{12}, λ_{13}, λ_{21}, λ_{22}, λ_{23}, λ_{31}, λ_{32}, and λ_{33}, and select the multiscale of discriminator. In addition, we set the norm $F = 1$ when training. In the specific implementation, we adopt the network structure of the CycleGAN model.

We choose the parameters by considering the training time and the quality of the generated images. Parameter λ: set $\lambda_{11} = \lambda_{12} = \lambda_{13} = 10$, $\lambda_{21} = 5$, $\lambda_{22} = \lambda_{23} = 0$, and $\lambda_{31} = \lambda_{32} = \lambda_{33} = 1$ for all cases. The multiscale and multilevel parameters are shown in the Table 3. For example, we show that the results of values of different multiscale and multilevel from apple to orange in Figure 3. It explains that the different values of multiscale and multilevel affect the generated images. Therefore, we use different multiscale and multilevel parameters for different image translation tasks.

3.4.2. Generated Images. To fairly compare the models, we present generated images by CycleGAN, DualGAN, UNIT, and MMGANs. And the auxiliary domains are, respectively, pear, orange, apple, winter, autumn, and summer. The testing results are shown as in the Figures 4, 5, and 6.

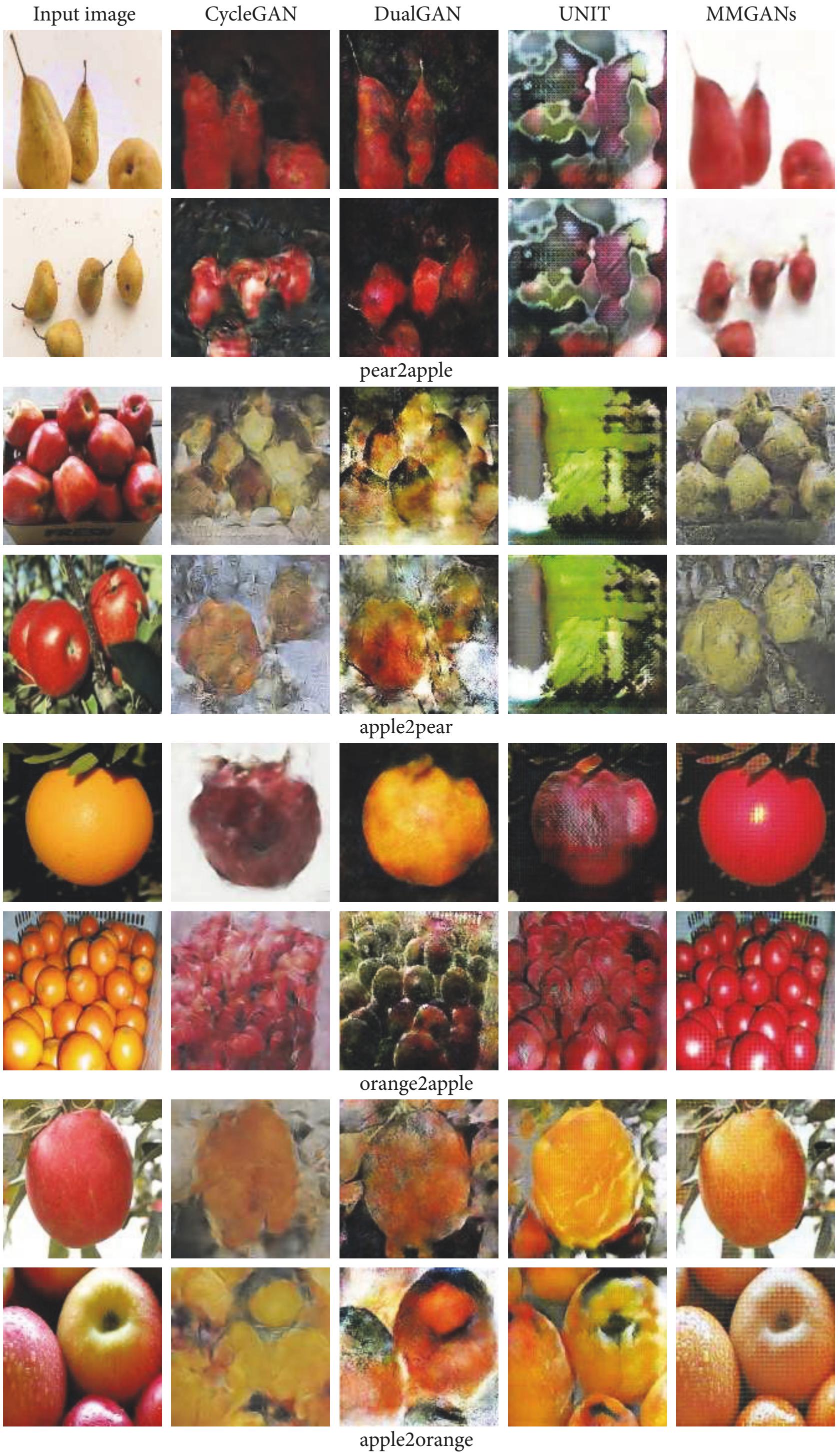


Figure 4: Generated images from test datasets. The left column is input image, then generated images with CycleGAN, DualGAN, UNIT, and MMGANs, respectively.

In Figures 4, 5, and 6, the results show that the generated images by MMGANs are more realistic. The outline, color, background, and foreground are better than the generated images by CycleGAN, DualGAN, and UNIT. What is more, for generated images with multiple target entities, our model works also better. In contrast, the fidelity of generated images by CycleGAN, DualGAN and UNIT is even worse. These results tell that the augmented auxiliary domain information helps MMGANs model focus on the generation of image details. The generator is constrained by the augmented auxiliary domain, which makes generated images more realistic.

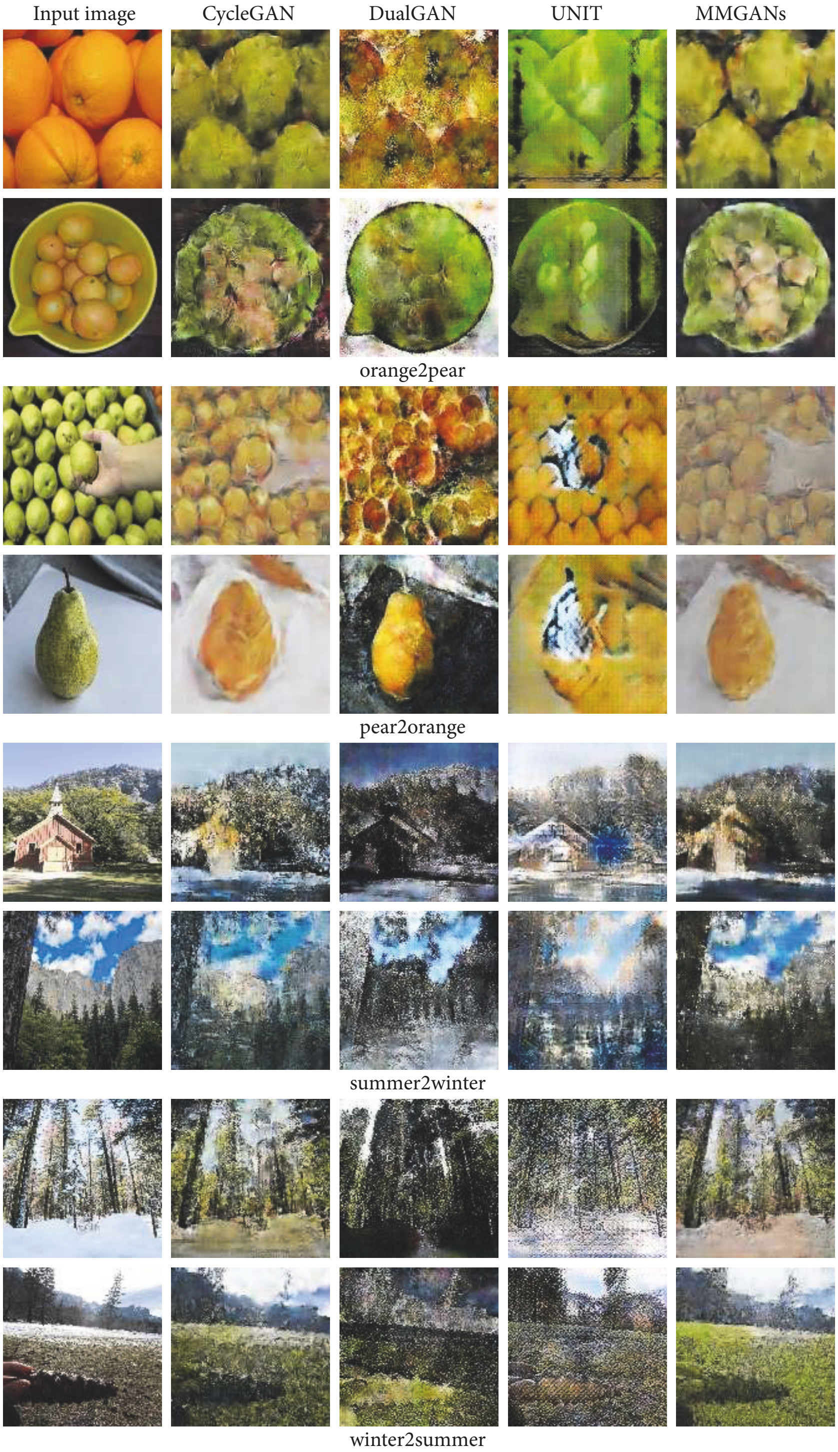
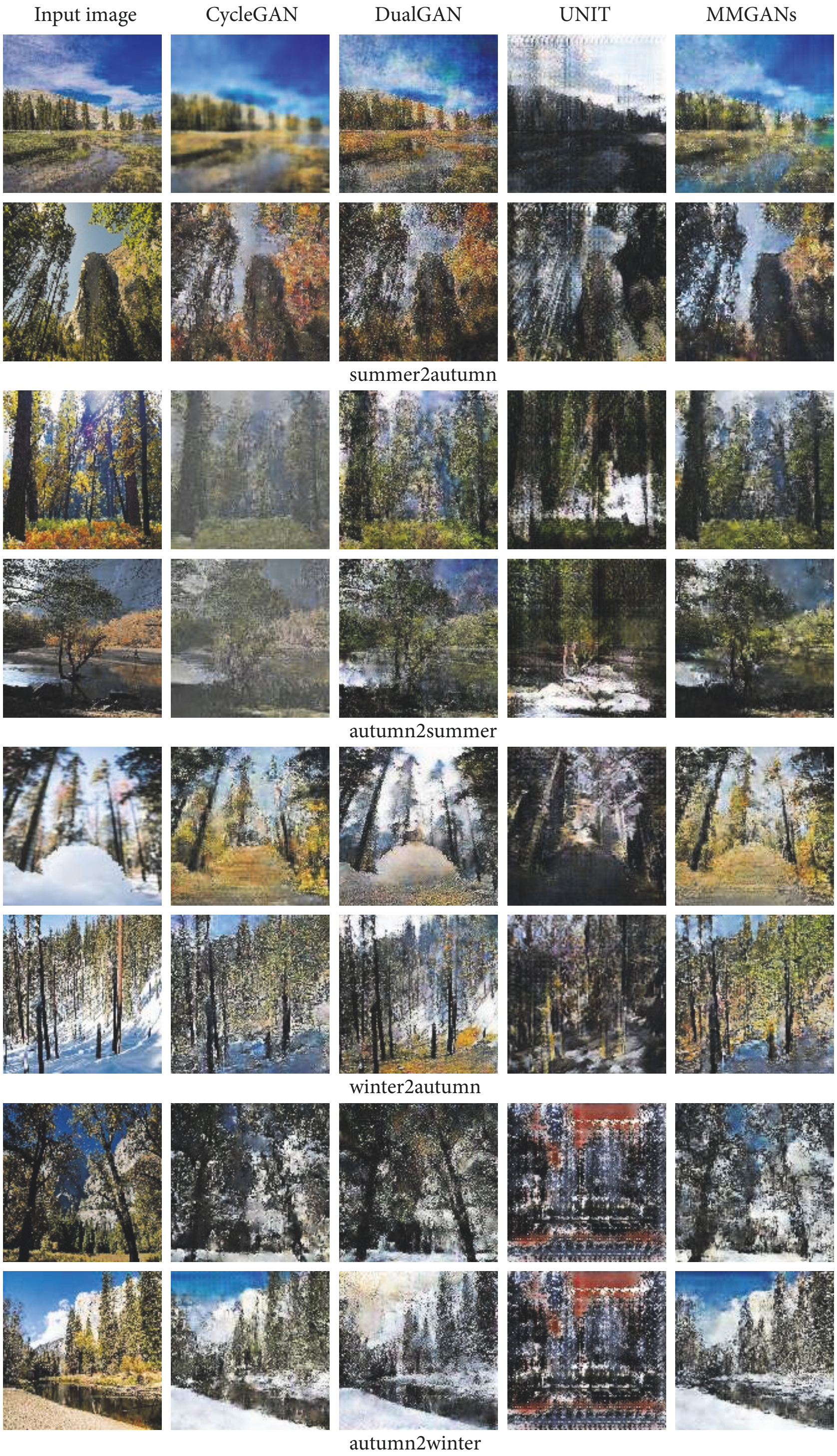


Figure 5: Generated images from test datasets (continuous). The left column is input image, then generated images with CycleGAN, DualGAN, UNIT, and MMGANs, respectively.

3.4.3. Performance Results. According to our proposed evaluation method, we only tell a few observers to sort the images from the testing results. For example, when testing apple2orange, based on the observers first impression, let observers judge which image in the testing results is worse, better, both of are good or bad.

Compared with CycleGAN, DualGAN, and UNIT, the testing results show that our model can obtain better performance by comparing the values of $n1$, $n2$, $n3$, and $n4$. In addition, they also show that the probability of better quality of generated images is higher than the CycleGAN, DualGAN, and UNIT. Concretely, the value of $n2$ is higher



summer2autumn

autumn2summer

winter2autumn

autumn2winter

Figure 6: Generated images from test datasets (continuous). The left column is input image, then generated images with CycleGAN, DualGAN, UNIT, and MMGANs, respectively.

than *n*1 from Tables 4, 5, and 6, except for summer2autumn of DualGAN. However, in terms of the average ratio of all image translation tasks, our MMGANs are better than CycleGAN, DualGAN, and UNIT. And the value of *n*3 means that the compared models have a higher probability of better generated images. On the contrary, the value of *n*4 represents the compared models being the probability of failure during image translation.

At last, through experiments on the six tasks of image translation, the results prove that the constraint generator and the losses of multiscale and multilevel discriminator make our model better generalization performance. In comparison,

Table 4: This table shows that the results of CycleGAN VS MMGANs, in which the third and the fourth column, respectively, represent a percentage of better quality of generated images by CycleGAN and MMGANs. The last two columns mean the generated images by the two models generated good images and bad images in the testing sets.

Augmented domain	Task	*n*1	*n*2	*n*3	*n*4
Pear	Apple2orange	2.2	61.4	8.6	27.8
Pear	Orange2apple	2.4	39.6	4	54
Orange	Apple2pear	0.6	31.6	0.2	67.6
Orange	Pear2apple	2.4	43.2	0	54.4
Apple	Orange2pear	2.4	17.4	4.6	75.6
Apple	Pear2orange	3	36.4	11.8	48.8
Winter	Summer2autumn	6.4	11.4	32.8	49.4
Winter	Autumn2summer	0.4	54.2	2.6	42.8
Autumn	Summer2winter	14	14.6	27.4	44
Autumn	Winter2summer	7.2	11.6	34.8	46.4
Summer	Autumn2winter	4	28.2	31.2	36.6
Summer	Winter2autumn	4.4	16.4	37.4	41.8
	Average	4.12	30.5	16.28	49.1

Table 5: This table shows that the results of DualGAN VS MMGANs, in which the third and the fourth column, respectively, represent a percentage of better quality of generated images by DualGAN and MMGANs. The last two columns mean the generated images by the two models generated good images and bad images in the testing sets.

Augmented domain	Task	*n*1	*n*2	*n*3	*n*4
Pear	Apple2orange	1.8	61.6	2.8	33.8
Pear	Orange2apple	1.6	44.4	1.4	52.6
Orange	Apple2pear	1	39.2	0.4	59.4
Orange	Pear2apple	2.4	49	1.4	47.2
Apple	Orange2pear	0.4	22.8	0.4	76.4
Apple	Pear2orange	2.6	51.8	3.6	42
Winter	Summer2autumn	21.6	13.6	32	32.8
Winter	Autumn2summer	8.4	21	36.2	34.4
Autumn	Summer2winter	17.6	42.6	1.4	38.4
Autumn	Winter2summer	10.8	35.2	2.8	51.2
Summer	Autumn2winter	8.2	26	33	32.8
Summer	Winter2autumn	10.8	33.2	13	43
	Average	7.27	36.7	10.7	45.33

Table 6: This table shows that the results of UNIT VS MMGANs, in which the third and the fourth column, respectively, represent a percentage of better quality of generated images by UNIT and MMGANs. The last two columns mean the generated images by the two models generated good images and bad images in the testing sets.

Augmented domain	Task	*n*1	*n*2	*n*3	*n*4
Pear	Apple2orange	8.8	47	17.6	26.6
Pear	Orange2apple	8.2	28.8	13.8	49.2
Orange	Apple2pear	0	37.2	0	62.8
Orange	Pear2apple	0	48	0	52
Apple	Orange2pear	2.2	20	0	77.8
Apple	Pear2orange	0	47.2	0.2	52.6
Winter	Summer2autumn	2.4	51.8	0.8	45
Winter	Autumn2summer	0	59	0	41
Autumn	Summer2winter	3.2	49.2	6.6	41
Autumn	Winter2summer	0.4	38	3.8	57.8
Summer	Autumn2winter	0	53.4	0	46.6
Summer	Winter2autumn	0.4	45.2	3.8	50.6
	Average	2.133333	43.73333	3.883333	50.25

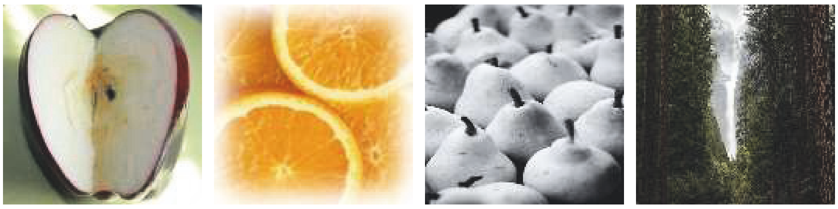
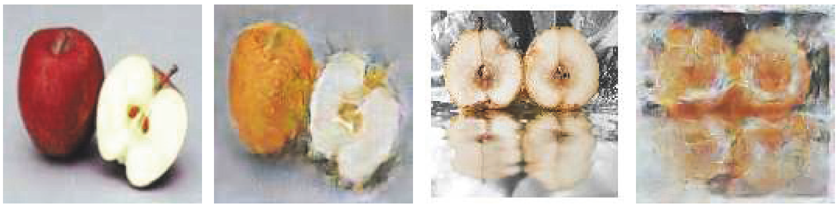


Figure 7: Training samples. These bad training samples may cause the failure of image translation.



apple2orange pear2orange

Figure 8: The failure of image translation in the testing sets.

our MMGANs has a higher probability of generated better images.

3.5. Discussion. In this paper, we propose an approach to enhance the ability of GANs on image translation using augmented auxiliary domain to constrain generator and using multiscale and multilevel losses to constrain discriminator. In this regard, we have studied the controllability of GANs. Through the study of controllability of GANs, our MMGANs can make the generated images more realistic in image translation.

However, our model also failed in the image translation task. Possible reasons include the following. (1) There are bad samples in our training and testing sets, which may generate bad samples in testing sets. It is shown as in Figures 7 and 8. (2) The performance of model needs further improvement. For example, introducing attention mechanism [38–41]. In addition, to make the generated image diversified, and to make the generated process interactive, we will consider using semantic to control image generation in future work.

4. Conclusion

We presented a MMGANs model for image translation, which constrains the generator and discriminator by adding the auxiliary domain, full cycle consistency loss, multiscale, and multilevel loss. In particular, the constraint on the generator allows the generator to learn the detailed content in the image. The constraint of discriminator makes the whole training process more stable and the produced image more lifelike. Through experiments on the six tasks of image translation, our model achieves better performance than CycleGAN, DualGAN, and UNIT.

Data Availability

The apple, orange, pear, summer, autumn, and winter datasets used to support the findings of this study are downloaded from ImageNet and Flickr. This paper has cited [7] and the researchers can access the data easily from ImageNet and Flickr.

Conflicts of Interest

There are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (61773093), Key R&D Program (Intelligent Processing Technology of Multi-source Litigation Letters and Visits National 2018YFC0831800), Important Science and Technology Innovation Projects in Chengdu (2018-YF08-00039-GX), and Research Programs of Sichuan Science and Technology Department (2016JY0088, 17ZDYF3184).

References

[1] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pp. 5967–5976, USA, July 2017.

[2] R. Rosales, K. Achan, and B. Frey, “Unsupervised image translation,” in *Proceedings of the 9th IEEE International Conference on Computer Vision*, pp. 472–478, France, October 2003.

[3] Q. Luan, F. Wen, D. Cohen-Or, L. Liang, Y.-Q. Xu, and H.-Y. Shum, “Natural image colorization in,” in *Proceedings of the 18th Eurographics conference on Rendering Techniques*, pp. 309–320, 2007.

[4] R. Zhang, P. Isola, and A. A. Efros, “Colorful Image Colorization,” in *Computer Vision – ECCV 2016*, vol. 9907 of *Lecture Notes in Computer Science*, pp. 649–666, Springer International Publishing, Cham, 2016.

[5] A. Van Den Oord, N. Kalchbrenner, O. Vinyals, L. Espeholt, A. Graves, and K. Kavukcuoglu, “Conditional image generation

with PixelCNN decoders," in *Proceedings of the 30th Annual Conference on Neural Information Processing Systems, NIPS 2016*, pp. 4797–4805, Spain, December 2016.

[6] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 1395–1403, 2015.

[7] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *Proceedings of the 16th IEEE International Conference on Computer Vision, ICCV 2017*, pp. 2242–2251, Italy, October 2017.

[8] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised Dual Learning for Image-to-Image Translation," in *Proceedings of the 16th IEEE International Conference on Computer Vision, ICCV 2017*, pp. 2868–2876, Italy, October 2017.

[9] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Advances in Neural Information Processing Systems*, pp. 700–708, 2017.

[10] H. Dong, P. Neekhara, C. Wu, and Y. Guo, "Unsupervised image-to-image translation with generative adversarial networks," https://arxiv.org/abs/1701.02676.

[11] S. Ma, J. Fu, C. W. Chen, and T. Mei, "Da-gan: Instance-level image translation by deep attention generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5657–5666, 2018.

[12] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, 2017.

[13] C. Xu, Q. Liu, and M. Ye, "Age invariant face recognition and retrieval by coupled auto-encoder networks," *Neurocomputing*, vol. 222, pp. 62–71, 2017.

[14] X. Li, M. Ye, Y. Liu, F. Zhang, D. Liu, and S. Tang, "Accurate object detection using memory-based models in surveillance scenes," *Pattern Recognition*, vol. 67, pp. 73–84, 2017.

[15] X. Li, M. Ye, Y. Liu, and C. Zhu, "Adaptive Deep Convolutional Neural Networks for Scene-Specific Object Detection," *IEEE Transactions on Circuits and Systems for Video Technology*, 2018.

[16] Y. Zhang, H. Chen, Y. He, M. Ye, X. Cai, and D. Zhang, "Road segmentation for all-day outdoor robot navigation," *Neurocomputing*, vol. 314, pp. 316–325, 2018.

[17] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pp. 5987–5995, USA, July 2017.

[18] C. Ledig, L. Theis, F. Huszár et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pp. 105–114, USA, July 2017.

[19] J. Li, Y. Wu, and K. Lu, "Structured domain adaptation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 8, pp. 1700–1713, 2017.

[20] Y. Taigman, A. Polyak, and L. Wolf, "Unsupervised cross-domain image generation," https://arxiv.org/abs/1611.02200.

[21] M.-Y. Liu and O. Tuzel, "Coupled generative adversarial networks," in *Proceedings of the 30th Annual Conference on Neural Information Processing Systems, NIPS 2016*, pp. 469–477, Spain, December 2016.

[22] M. Li, H. Huang, L. Ma, W. Liu, T. Zhang, and Y. Jiang, "Unsupervised Image-to-Image Translation with Stacked Cycle-Consistent Adversarial Networks," in *Computer Vision – ECCV 2018*, vol. 11213 of *Lecture Notes in Computer Science*, pp. 186–201, Springer International Publishing, Cham, 2018.

[23] J. Li, K. Lu, Z. Huang, L. Zhu, and H. T. Shen, "Transfer Independently Together: A Generalized Framework for Domain Adaptation," *IEEE Transactions on Cybernetics*, pp. 1–12, 2018.

[24] M. Mirza and S. Osindero, "Conditional generative adversarial nets," https://arxiv.org/abs/1411.1784.

[25] H. Zhang, T. Xu, and H. Li, "StackGAN: Text to Photo-Realistic Image Synthesis with Stacked Generative Adversarial Networks," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 5908–5916, Venice, October 2017.

[26] Z. Deng, H. Zhang, X. Liang et al., "Structured generative adversarial networks," in *Advances in Neural Information Processing Systems*, pp. 3902–3912, 2017.

[27] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets," in *Proceedings of the 30th Annual Conference on Neural Information Processing Systems, NIPS 2016*, pp. 2180–2188, Spain, December 2016.

[28] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in *Advances in Neural Information Processing Systems*, pp. 5769–5779, 2017.

[29] V. Nagarajan and J. Z. Kolter, "Gradient descent gan optimization is locally stable," in *Advances in Neural Information Processing Systems*, pp. 5591–5600, 2017.

[30] Y. Gan, J. Gong, M. Ye, Y. Qian, and K. Liu, "Unpaired cross domain image translation with augmented auxiliary domain information," *Neurocomputing*, vol. 316, pp. 112–123, 2018.

[31] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," https://arxiv.org/abs/1701.04862.

[32] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, G. Klambauer, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a nash equilibrium," *Advances in Neural Information Processing Systems*, vol. 1, no. 2, article 4, 2017.

[33] A. Odena, J. Buckman, C. Olsson et al., "Is generator conditioning causally related to gan performance?" https://arxiv.org/abs/1802.08768.

[34] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," https://arxiv.org/abs/1711.11585.

[35] C. S. Chin, J. Si, A. S. Clare, and Maode Ma, "Intelligent Image Recognition System for Marine Fouling Using Softmax Transfer Learning and Deep Convolutional Neural Networks," *Complexity*, vol. 2017, Article ID 5730419, 9 pages, 2017.

[36] Á. Morera, Á. Sánchez, J. F. Vélez, and A. B. Moreno, "Gender and Handedness Prediction from Offline Handwriting Using Convolutional Neural Networks," *Complexity*, vol. 2018, Article ID 3891624, 14 pages, 2018.

[37] Z. Tu, W. Xie, J. Dauwels, B. Li, and J. Yuan, "Semantic Cues Enhanced Multi-modality Multi-Stream CNN for Action Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, 2018.

[38] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," in *Advances in Neural Information Processing Systems*, pp. 5998–6008, 2017.

[39] K. Xu, J. L. Ba, R. Kiros et al., "Show, attend and tell: Neural image caption generation with visual attention," in *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015*, pp. 2048–2057, France, July 2015.

[40] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," https://arxiv.org/abs/1805.08318.

[41] A. Pumarola, A. Agudo, A. M. Martinez, A. Sanfeliu, and F. Moreno-Noguer, "GANimation: Anatomically-Aware Facial Animation from a Single Image," in *Computer Vision – ECCV 2018*, vol. 11214 of *Lecture Notes in Computer Science*, pp. 835–851, Springer International Publishing, Cham, 2018.