Unsupervised Transformation Network Based on GANs for Target-domain Oriented Image Translation

HONGWEI GE\(^1\), YAO YAO\(^1\), ZHENG CHEN\(^2\), LIANG SUN\(^1\)

\(^1\)College of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China
\(^2\)Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA

Corresponding author: Hongwei Ge (e-mail: hwge@dlut.edu.cn).

ABSTRACT

Image-to-image translation usually refers to the task of translating an input image from the source domain to the target domain while preserving the structure in the source domain. Recently, generative adversarial networks (GANs) using paired images for this task have made great progress. However, paired training data will not be available for many tasks. In this paper, a GAN-based unsupervised transformation network (UTN-GAN) is proposed for image-to-image translation. Importantly, UTN-GAN employs hierarchical representations and weight-sharing mechanism to translate images from the source domain to the target domain without paired images. We employ two groups of unsupervised GANs to generate the images in different domains firstly, and then discriminate them. In UTN-GAN, an auto-encoder reconstruction network is designed to extract the hierarchical representations of the images in the source domain by minimizing the reconstruction loss. In particular, the high-level representations (semantics) are shared with a translation network to guarantee that the input image and the output image are paired up in the different domains. All network structures are trained together by using a joint loss function. The experimental studies in qualitative and quantitative aspects on several image translation tasks show that the proposed algorithm is effective and competitive compared with some state-of-the-art algorithms.

INDEX TERMS

Generative adversarial networks, unsupervised learning, style transfer, image-to-image translation

I. INTRODUCTION

Many classic problems in image processing, computer graphics, and computer vision can be posed as image-to-image translation tasks, mapping the input image in one domain to the desired output image in another domain. Although there are different distributions for images in different domains, they may exist a certain relationship. For example, colorization [1], [2] can be regarded as transforming a grayscale image to the corresponding color image; image super-resolution [3]–[5] can be regarded as generating a high-quality image from the low-resolution counterpart; style transfer [6] can be considered as the problem of rendering an image in the style of a given domain.

The idea of image-to-image translation can be tracked back at least to image analogies [7], which use a non-parametric texture model on an input-output image pair [8]. Recently, different convolutional neural networks (CNNs) have been constructed to solve image-to-image translation tasks. These CNNs are designed for distinct tasks and differ in loss function. By penalizing the difference between the output images and the real images, CNNs can be trained to discover the mapping from the input image to the translated image [9]. And using a pixel loss function, e.g., L1 or L2 norm, to measure the output images is one of the most straightforward methods [1], [2], [4], [10]. Although such approach can generate reasonable images, there are some undeniable defects that the pixel-wise losses often produce blurry results. As these losses can capture the low frequency information rather than the high frequency information in many cases.

Recent years have witnessed that the algorithms employing the idea of generative adversarial network (GAN) [11] have achieved impressive results in image-to-image translation tasks. GAN, proposed by Goodfellow et al., consists of a generator G and a discriminator D, aiming to model the natural image distribution by synthesizing realistic samples
that are indistinguishable from natural images. Some experimental results show that the models based on GANs can obtain the high-frequency information, generating sharper and more realistic images [12]. Therefore, the algorithms employing the idea of GANs or conditional GANs (cGANs) [13] have become a popular choice for many image-to-image translation tasks [14]–[19]. Based on cGAN, pix2pix [12] is a representative framework designed for the problems with available paired images, which aims to model the conditional distributions of the natural images conditioned on the input images [15], [18], [19]. However, it is difficult to provide a large number of paired images in different domains for the complex practical tasks, since manual annotation is expensive.

To solve these difficulties, several GAN-based unsupervised methods without employing any paired data have been proposed. CycleGAN [8], DiscoGAN [20] and DualGAN [21] couple two GANs together, each responsible for generating and discriminating images in respective domain. Further, additional reconstruction losses are introduced to make the two generators strongly coupled and also alleviate the problem of model collapse [22]. Besides, CoGAN [23] and UNIT [24] uses a weight-sharing constraint to learn a joint distribution without any corresponding images in different domains.

Inspired by the ability of hierarchical feature representation from deep neural models and the capacity of restraint from weight-sharing mechanism, this paper proposes a GAN-based unsupervised transformation network (UTN-GAN) for image translation. UTN-GAN employs hierarchical representations and weight-sharing mechanism to translate images from a source domain to the target without paired images. The hierarchical representations of the input images are learned by a reconstruction network and the high-level representations are shared with a translation network to realize the image-to-image translation.

The rest of this paper is organized as follows. Section II introduces the related work about image-to-image translation tasks and generative adversarial networks. Section III proposes an unsupervised transformation network for target-domain oriented image translation, and formulate the proposed model. Section IV presents the experimental studies and analyses on several image-to-image translation tasks. Section V concludes the paper.

II. RELATED WORK

The generative adversarial networks (GANs) are inspired by a two-player minimax game, which generate new samples by constructing a generative network and a discriminative network to capture the latent distribution of real data. GANs have shown great potential for image generation and representation learning. As an extension, the algorithms utilizing GANs or their variants, such as cGANs [13], LAPGAN [25], and InfoGAN [26], have been applied to image translation in recent years. The cGANs [13] use class labels as conditional information for generator and discriminator to generate class-specific images. The incorporated conditional information in the model makes the generator of the cGANs learn more effectively. Based on cGANs, some improved architectures have been proposed for image-to-image translation tasks when paired images are available. The common framework based on cGANs is developed to learn a loss adapted to different image translation problems that traditionally would require different loss formulations [12]. Further, Wang et. al [15] extend the framework by using a novel adversarial loss as well as a coarse-to-fine generator and a multi-scale discriminator architecture for translating semantic label maps to high-resolution photo-realistic images. In [18], an end-to-end sequence of two cGAN networks is proposed to synthesize cross-view images as well as maintain the true semantics of objects well in both source and target domains, in which the strategy of generating semantic segmentations together with images in target domain helps the networks learn better. With the supervision provided by paired data, these methods can generate impressive images, however, building a dataset with paired images sometimes can be a challenging task.

To address this challenge, some unsupervised learning methods are developed for image translation task without paired images. CycleGAN [8] directly uses two groups of generators and discriminators to train a mapping and an inverse mapping between source and target domains. Besides, for the cycle consistency loss, CycleGAN essentially trains the image reconstruction by the source-target-source and target-source-target transformation to reduce the space of possible mappings, generating high quality images and alleviating model collapse. Similar works include DiscoGAN [20] and DualGAN [21]. Considering the images in two domains have the same underlying characteristic, a weight-sharing strategy is presented in [23], with which a joint distribution in different domains can be learned, by sharing the weight corresponding to high-level semantic information in both generators and discriminators. However, on one hand, in CoGAN, the errors from the discriminators have a direct impact on the decoded representation and further lead to the inaccuracy of the shared high-level semantics. On the other hand, using a random vector z as inputs to generate high-resolution images may result in model collapse. In UNIT [24], this framework is extended with two domain image encoders based on variational autoencoders (VAEs) [27]. The weight-sharing strategy provides a novel perspective to treat the image-to-image translation problem and is quite useful for this problem.

Although GANs have achieved impressive results in image-to-image translation tasks, it is known that GANs are unstable to train and often result in artificial synthesized images. As proved in [28], if the supports of the generated density and the true data density are disjoint or lie in low dimensional manifold, the GAN could lead to vanishing gradient that makes the generator impossible to update. Many researchers have focus on improving the stability of training and the quality of generated images by applying deep learning techniques or optimizing the objective function.
stein GAN [29] and LSGAN [30] replace the cross-entropy loss to overcome the vanishing gradients problem, making GAN able to generate higher quality images and more stable during the training. DCGAN [31] uses deep convolutional networks with a set of architectural constraints for making them stable to train. In [32], a new training method that grows both the generator and discriminator progressively is introduced to improve the quality and stability. Although the performance of GANs has been greatly improved, there is still much to investigate for different application problems.

In this paper, we propose a GAN-based unsupervised transformation network (UTN-GAN) for image translation. Using an auto-encoder reconstruction network to extract the hierarchical representations, UTN-GAN avoids the extra errors caused by the discriminators in feature extraction. This strategy pushes the generator to be learned more effectively. The weight-sharing mechanism is also employed to share the high-level representations with the translation network, which can restrain the capacity of the translation network. Instead of using a random vector $z$ as inputs to generate high-resolution images, UTN-GAN rather employs the source domain images as the inputs of generators, hence help the model escape from collapsing to a certain extent. The strategies employed in UTN-GAN are beneficial to train the model more stably.

III. PROPOSED UNSUPERVISED UTN-GAN

In this section, we develop an unsupervised transformation network based on GANs for target-domain oriented image translation, which can learn to translate the input images from a source domain $A$ to a target domain $B$ without employing any paired data.

A. OBJECTIVE

GAN consists of a generator $G$ and a discriminator $D$ and they are trained simultaneously through a minimax game to seek an equilibrium. The goal of generator $G$ is to learn a distribution $p_G$ from noise distribution $p_z$ that matches the real data distribution $p_{data}$, while the discriminator $D$ is trained to distinguish real images $x \sim p_{data}$ from generated images. The objective for GAN can be formulated as follows:

$$
\begin{align*}
\min_G \max_D L(G,D) &= E_{x \sim p_{data}(x)}[\log D(x)] \\
&\quad + E_{z \sim p_z(z)}[\log (1 - D(G(z)))]
\end{align*}
$$

Our goal is to learn a network that translates images from domain $A$ to domain $B$ given training samples $\{x_i\}_{i=1}^N \in A$ and $\{y_j\}_{j=1}^M \in B$ without any paired information. The proposed algorithm considers the following two aspects. First, for image-to-image translation tasks, although the input images and output images are in different surface appearance, they could have the same underlying representation [24]. Second, the GAN with infinite model capacity could suffer from vanishing gradient and model collapse [33]. To restrain the translation network and train it more stably, we utilize a reconstruction network to learn the hierarchical representations by reconstructing the input image, and keep the consistency of the mutual high-level representations in the translation processing.

As shown in Figure 1, the proposed transformation network is composed of two GAN networks. The reconstruction network can obtain hierarchical representations of the input image $x$ by reconstructing the original input images. The translation network translates the input image $x$ to the target domain. The shared high-level representations can provide constraints through the weight-sharing strategy for the generator $G_t$ and $G_t$ translates the input image $x$ to the desired image under such constraints. The discriminator $D_r$ and $D_t$ are used to distinguish the real members from the generated images, where $D_r$ aims to distinguish between images $\{x\}$ and the reconstructed images $\{G_r(x)\}$; in the same way, $D_t$ aims to distinguish between $\{y\}$ and the translated $\{G_t(x)\}$.

Here, the GAN losses can be expressed as:

$$
\begin{align*}
\min_{G_t} \max_{D_t} L(G_t, D_t) &= E_{x \sim p_{data}(x)}[\log D_r(x)] + E_{y \sim p_{data}(y)}[\log (1 - D_t(G_t(x)))] \\
&= E_{x \sim p_{data}(x)}[\log (1 - D_r(G_t(x)))] \\
&+ E_{y \sim p_{data}(y)}[\log (1 - D_t(G_t(x)))]
\end{align*}
$$

To push the generator to be learned more effectively, a reconstruction loss is employed to extract accurate representation. Researches on conditional images translation have shown that it is beneficial to mix the adversarial loss with another loss, such as L1 distance and L2 distance. And it has also shown L1 distance often leads to less blurriness than L2 [12]. Moreover CycleGAN [8] argues that combining the adversarial losses with additional reconstruction losses can get better results and more stable generators. This motivates us to use the adversarial loss to model the high frequency information as well as use the pixel loss to force the low frequency structure for the image translation tasks.

![Figure 1. Proposed unsupervised transformation network architecture for image-to-image translation. $G_t$ is used to reconstruct the input images $x$ while $G_t$ is used to translate the input images to target domain. $D_r$ and $D_t$ are used to distinguish the real members from the generated images of each domain. The weight-sharing constraint is enforced in both the generators and discriminators.](image-url)
The first layers of discriminator extract detail features, while the last layers extract high-level semantic features. We employ the weight-sharing constraint in the latter layers of $D_r$ and $D_t$ as the input images have the similar high-level semantics in two domains. And sharing the weights in the discriminators also enables to reduce total numbers of parameters in the network.

IV. EXPERIMENTS

To validate and analyze the performance of the proposed unsupervised transformation network for target-domain oriented image translation, several sets of experiments are conducted on challenging translation tasks and the obtained qualitative and quantitative results are analyzed.

A. PARAMETER SETTINGS

In the experiments, the Adam solver [37] is adopted to train the proposed model, where the learning rate is set to 0.0002 and momentums are 0.5 and 0.999 as suggested in [29]. Leaky ReLU is with a slope of 0.2, and ReLU is used in the output. The discriminator is trained to judge whether an image is generated or from true data or from generated image. The discriminator is trained to judge whether an image is generated or from true data or from generated image. In this paper, the architectures of $D_r$ and $D_t$ are identical CNNs similar to the encoder of the generator. Specially, we employ the PatchGAN structure [12], [36] for the 256x256 images, which only penalizes structure at the scale of patch rather than the full image. This architecture is effective in capturing local high-frequency features such as texture and style. An extra advantage of the PatchGAN is that it requires fewer parameters, and run faster. The patch size in this paper is fixed at 70x70.

The first layers of discriminator extract detail features, while the last layers extract high-level semantic features. We employ the weight-sharing constraint in the latter layers of $D_r$ and $D_t$ as the input images have the similar high-level semantics in two domains. And sharing the weights in the discriminators also enables to reduce total numbers of parameters in the network.

2) Discriminators

The discriminator is trained to judge whether an image is from true data or from generated image. In this paper, the architectures of $D_r$ and $D_t$ are identical CNNs similar to the encoder of the generator. Specially, we employ the PatchGAN structure [12], [36] for the 256x256 images, which only penalizes structure at the scale of patch rather than the full image. This architecture is effective in capturing local high-frequency features such as texture and style. An extra advantage of the PatchGAN is that it requires fewer parameters, and run faster. The patch size in this paper is fixed at 70x70.

The first layers of discriminator extract detail features, while the last layers extract high-level semantic features. We employ the weight-sharing constraint in the latter layers of $D_r$ and $D_t$ as the input images have the similar high-level semantics in two domains. And sharing the weights in the discriminators also enables to reduce total numbers of parameters in the network.

IV. EXPERIMENTS

To validate and analyze the performance of the proposed unsupervised transformation network for target-domain oriented image translation, several sets of experiments are conducted on challenging translation tasks and the obtained qualitative and quantitative results are analyzed.

A. PARAMETER SETTINGS

In the experiments, the Adam solver [37] is adopted to train the proposed model, where the learning rate is set to 0.0002 and momentums are 0.5 and 0.999 as suggested in [29]. Leaky ReLU is with a slope of 0.2, and ReLU is used in the output. The discriminator is trained to judge whether an image is generated or from true data or from generated image. The discriminator is trained to judge whether an image is generated or from true data or from generated image. In this paper, the architectures of $D_r$ and $D_t$ are identical CNNs similar to the encoder of the generator. Specially, we employ the PatchGAN structure [12], [36] for the 256x256 images, which only penalizes structure at the scale of patch rather than the full image. This architecture is effective in capturing local high-frequency features such as texture and style. An extra advantage of the PatchGAN is that it requires fewer parameters, and run faster. The patch size in this paper is fixed at 70x70.

The first layers of discriminator extract detail features, while the last layers extract high-level semantic features. We employ the weight-sharing constraint in the latter layers of $D_r$ and $D_t$ as the input images have the similar high-level semantics in two domains. And sharing the weights in the discriminators also enables to reduce total numbers of parameters in the network.

IV. EXPERIMENTS

To validate and analyze the performance of the proposed unsupervised transformation network for target-domain oriented image translation, several sets of experiments are conducted on challenging translation tasks and the obtained qualitative and quantitative results are analyzed.

A. PARAMETER SETTINGS

In the experiments, the Adam solver [37] is adopted to train the proposed model, where the learning rate is set to 0.0002 and momentums are 0.5 and 0.999 as suggested in [29]. Leaky ReLU is with a slope of 0.2, and ReLU is used in the output. The discriminator is trained to judge whether an image is generated or from true data or from generated image. The discriminator is trained to judge whether an image is generated or from true data or from generated image. In this paper, the architectures of $D_r$ and $D_t$ are identical CNNs similar to the encoder of the generator. Specially, we employ the PatchGAN structure [12], [36] for the 256x256 images, which only penalizes structure at the scale of patch rather than the full image. This architecture is effective in capturing local high-frequency features such as texture and style. An extra advantage of the PatchGAN is that it requires fewer parameters, and run faster. The patch size in this paper is fixed at 70x70.
CycleGAN [8] CycleGAN can learn mapping functions $G$: A $\rightarrow$ B and $F$: B $\rightarrow$ A between two domains A and B, using cycle-consistent loss to enforce $F(G(A)) \approx A$ and $G(F(B)) \approx B$.

CoGAN [23] CoGAN can learn a joint distribution of multi-domain images. One generator is for domain A and another is for domain B, with weight-sharing on the first few layers. The generators of CoGAN produce images from a random vector $z$, which is different from our network.

DiscoGAN [20] DiscoGAN learns to discover relations between different domains. Then the discovered relations are used to transfer style from one domain to another.

Pix2pix [12] Pix2pix is based on conditional GAN structure and is trained with paired data.

2) Metrics
Besides the qualitative evaluation that the translated images obtained by using different methods are directly observed, the following quantitative metrics are also adopted to evaluate the performance of the proposed algorithm. As in some generative networks [3], [9], [18], we also employ Structural Similarity Index (SSIM) [38], Signal to Noise Ratio (PSNR) [39] and Universal Quality Index (UQI) [40] to evaluate our methods.

SSIM measures the similarity between the generated image $x'$ and the ground truth image $x$ based on their luminance, contrast and structural aspects. Higher SSIM means $x'$ is more similar to $x$.

PSNR is the average ratio of the power of the original signal to the power of the noise.

UQI models image distortion from loss of correlation, luminance, and contrast. A higher value means better quality of the generated image.

$$SSIM(x,x') = \frac{(2\mu_x \mu_{x'} + c_1)(2\sigma_{xx'} + c_2)}{\mu_x^2 + \mu_{x'}^2 + c_1(\sigma_x^2 + \sigma_{x'}^2 + c_2)}$$

$$MSE(x,x') = \frac{1}{n} \sum_{i=1}^{n} (x[i] - x'[i])^2$$

$$PSNR(x,x') = 10 \times \log_{10}\left(\frac{\text{max}^{2}_{x'}\text{max}^{2}_{x}}{\text{MSE}}\right)$$

$$UQI(x,x') = \frac{4\sigma_{xx'} \mu_x \mu_{x'}}{\mu_x^2 + \mu_{x'}^2} \left(\frac{\sigma_x^2 + \sigma_{x'}^2}{\sigma_x^2 + \sigma_{x'}^2 + \mu_{x'}^2}ight)$$

C. ANALYSIS OF THE LOSS FUNCTION
To analyze the components of the whole objective formulated in Equation 5, we conduct ablation studies of the loss function. We remove the reconstruction loss (Equation 4) and the adversarial loss (Equation 2) of the reconstruction network $G_r$, in turn.

Figure 3 shows the qualitative results of these variations on season translation task. The single translation network leads to colorless and blurry results. Combining the adversarial loss of reconstruction network $G_r$, the output can be colorful but loses some information. Adding the reconstruction loss to the objective, the better results of image translation are obtained. Clearly it is that every loss item in the objective plays an important role and the joint loss integrating reconstruction loss and the generative adversarial losses can synthesize more realistic translated images.

D. IMAGE RECONSTRUCTION QUALITY
The image reconstruction is one critical step for the proposed image translation network as it provides useful information with the underlying characteristic. We compare the reconstructed images in UTN-GAN with the ones obtained in the reconstruction process in DiscoGAN [20] and CycleGAN [8], which the reconstructed images are also very important. They translate the input images to the target domain and then map the translated images back to the original input images.

<table>
<thead>
<tr>
<th>Method</th>
<th>PSNR(db)</th>
<th>SSIM</th>
<th>UQI</th>
</tr>
</thead>
<tbody>
<tr>
<td>DiscoGAN</td>
<td>18.3383</td>
<td>0.6563</td>
<td>0.6003</td>
</tr>
<tr>
<td>CycleGAN</td>
<td>24.8148</td>
<td>0.9428</td>
<td>0.8768</td>
</tr>
<tr>
<td>UTN-GAN</td>
<td>26.5777</td>
<td>0.9869</td>
<td>0.9690</td>
</tr>
</tbody>
</table>
Figure 4 shows a few reconstructed images from our method. The first two rows are the instances from season transfer task and the latter two rows are the instances from collection style transfer task. It can be seen that the reconstructed images are very close to the input images. Table 1 and Table 2 presents the quantitative metrics for the obtained results using different methods on season transfer and collection style transfer tasks. The proposed algorithm obtained the best reconstruction results compared with DiscoGAN [20] and CycleGAN [8].

**TABLE 1.** Comparisons with CycleGAN and DiscoGAN in season transfer task.

<table>
<thead>
<tr>
<th>Method</th>
<th>PSNR (dB)</th>
<th>SSIM</th>
<th>UQI</th>
</tr>
</thead>
<tbody>
<tr>
<td>DiscoGAN</td>
<td>15.38</td>
<td>0.59</td>
<td>0.52</td>
</tr>
<tr>
<td>CycleGAN</td>
<td>21.13</td>
<td>0.93</td>
<td>0.94</td>
</tr>
<tr>
<td>UTN-GAN</td>
<td>24.07</td>
<td>0.97</td>
<td>0.95</td>
</tr>
</tbody>
</table>

**E. IMAGE-TO-IMAGE TRANSLATION**

1) Face conversion

We apply our method to face attributes conversion task on the CelebA dataset [41]. This dataset consists of 202599 celebrity images with 40 attribute annotations. We resize the images into 64x64 pixels and train our model on several attributes, including black hair, blond hair, mouth slightly open and arched eyebrow. As the input images are smaller scale, in this experiment, the architectures are modified by removing the last three layers of Convolution-BatchNorm-LeakyReLU from encoder and the first three layers of Deconvolution-BatchNorm-ReLU from decoder of generator.

Figure 5 shows the translation results on the hair, mouth and eyebrow attributes. For each pair, the first row is the input images and the second row is the corresponding output using proposed UTN-GAN. As visually demonstrated in Figure 5, UTN-GAN successfully translates the input images to the desired domain while preserving structures in the input images.

In contrast, Figure 6 shows the visual results obtained by UTN-GAN and those obtained by CoGAN [23] and DiscoGAN [20]. In Figure 6, the first row is the input images. The results demonstrate the proposed algorithm is competitive and efficient.

2) Season transfer

In this section, the season transfer that translates an image from summer to winter is used to validate the performance of the UTN-GAN. The model is trained on 1273 summer photos and 962 winter photos from [8]. We scale them to 256x256 pixels. The results obtained by UTN-GAN, CycleGAN [8], DiscoGAN [20], and CoGAN [23] are shown in Figure 7.

From Figure 7, we find CoGAN fails to generate realistic images and DiscoGAN tends to generate blurry images. One possible reason is the input of CoGAN is noise z, resulting in the difficulty of the generation of high-resolution images. Both CycleGAN and UTN-GAN can generate realistic images.

3) Style transfer

In this section, the input photos are translated to collection style images. The experiments are conducted on the dataset with 6853 photos and 1074 Monet’s paintings from [8], and
scale them to 256x256 pixels. In contrast, Figure 8 shows the visual results obtained by UTN-GAN and those obtained by CycleGAN [8], DiscoGAN [20] and CoGAN [23]. The results reflect that UTN-GAN can preserve the textures in the input images successfully and generate realistic painting images which are compelling with those produced by CycleGAN [8], CoGAN [23] and DiscoGAN [20].

4) Results on paired data
To validate the generalization ability of the proposed algorithm on paired images, we also conduct the test on the CMP Facade Database [42], and the images are scaled to 256x256 pixels. Figure 9 show some samples on architectural labels to photos compared with CycleGAN [8], CoGAN [23] and Pip2pix [12], among which, Pip2pix is designed for paired data and it is not suitable for unpaired case. The results show that our algorithm can preserve the structures in the label images well. The quality of our images is competitive with those produced by CycleGAN [8] and CoGAN [23] trained without pairwise supervised images and comparable with those produced by supervised Pip2pix [12].

V. CONCLUSIONS
To address the problems of unpaired image to image translation, this paper proposes a GAN-based unsupervised transformation network (UTN-GAN) with hierarchical representations learning and weight-sharing mechanism. The hierarchical representations of the input image are learned by the reconstruction network and the mutual high-level representations are shared with the translation network to realize the target-domain oriented image translation. The experimental results on some challenging translation tasks show the strategies employed in UTN-GAN are beneficial to train the model more stably and the obtained translation results are competitive compared with some state-of-the-art algorithms. In the future work, we aim to develop the models with better representation and generalization abilities for more effective knowledge transfer, and apply the models to more applications, such as camouflage face recognition and camouflage transfer task.
ACKNOWLEDGMENT

The authors are grateful for the support of the National Natural Science Foundation of China (61572104, 61402076, 61502072), the Fundamental Research Funds for the Central Universities (DUT17JC04), and the Project of the Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University (93K172017K03).

REFERENCES


Ge et al.: Unsupervised Transformation Network Based on GANs for Target-domain Oriented Image Translation


***