IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 20, 2021, accepted March 13, 2021, date of publication March 23, 2021, date of current version March 30, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3068094

ArCycleGAN: Improved CycleGAN for Style Transferring of Fruit Images

HONGQIAN CHEN[1,3], MENGXI GUAN[1,3], AND HUI LI[2]
[1]School of Computer Science and Engineering, Beijing Technology and Business University, Beijing 100048, China
[2]Management College, Beijing Union University, Beijing 100101, China
[3]Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China

Corresponding author: Hui Li (lihui@buu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 31701517, in part by the Beijing Philosophy and Social Science Foundation under Grant 17GLC060 and Grant 20GLB032, and in part by the Academic Research Projects of Beijing Union University under Grant ZB10202005.

ABSTRACT CycleGAN can realize image translation and style transferring among unpaired images. However, it will easily generate inappropriate image results when the number and shapes of objects in the style offering image and the source image are greatly different. The paper proposed an improved network, named arCycleGAN, which introduced the mechanism of attribute registration into CycleGAN to solve the problem. The arCycleGAN can transfer the freshness styles from the style offering images to the unpaired input source images. The generated target images will have the freshness attributes of the style offering images, while maintaining the shapes and key features of the input source images. The realization of mechanism of attribute registration consists of three modules. The first module is attribute recognition module, which can identify and label the attributes of objects in images. The second module is image pre-screening module, which selects appropriate image subset as screened training set from raw image set according to the attributes of the input source images. The third module is similarity matching module, which matches the images in screened training set based on the similarity. The generator and discriminator in the new network are similar to that in the CycleGAN network. Experimental results demonstrate the effectiveness and better performance of the arCycleGAN. Compared with the CycleGAN, the new network can generate more convincing images. It can generate the target images of similar quality based on a smaller training set and less training time than the original CycleGAN. For generating images of similar quality, the number of images in the required training set can be reduced by 50%, while training time is reduced by 5.8%.

INDEX TERMS Image generation, style transferring, attribute registration, image registration, improved CycleGAN.

I. INTRODUCTION

Image generation with specified style attributes is an important problem in the field of computer vision. It has been applied in many scenes, such as advertising, photography, scene beautification. Researchers have brought up many applications based on CycleGAN because of its great features and performances. For example, Tian *et al.* [1] used CycleGAN to expand the image data set of apple diseases to solve the problem of insufficient image data caused by the random occurrences of apple diseases. Cap *et al.* [2] proposed the LeafGAN network, which can generate images of apples with various diseases based on images of healthy apples. These generated images can be used as data enhancements to improve the diagnostic performance of plant diseases. Wu *et al.* [3] proposed the adjusted DCGAN (Deep Convolutional Generative Adversarial Networks) model to expand the tomato leaves image set.

Freshness is an important attribute for plants and fruits. But in many application scenes like daily photo shooting, the qualities of fruit images cannot meet our expectations due to various reasons. Through CycleGAN, the images of non-fresh fruits can be translated into attractive images of fresh fruits, vice versa. However, in some applications, we found that there are a lot of unsatisfactory results in the generated images based on CycleGAN.

The associate editor coordinating the review of this manuscript and approving it for publication was Li He.

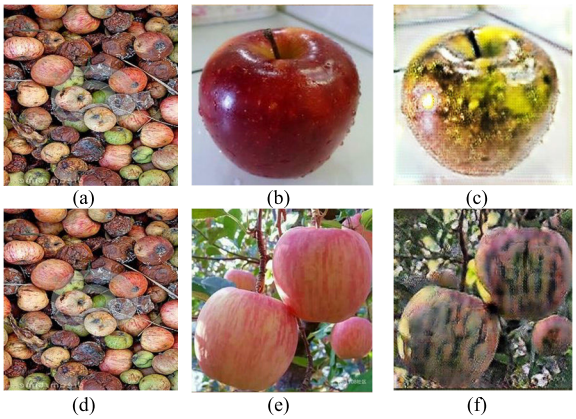


FIGURE 1. Examples of unsatisfactory results in generated images using CycleGAN: (a) (d) style offering image; (b) (e) input source image; (c) (f) generated target image.

In addition to the inappropriate results that we found, Zhang *et al.* [4] pointed out that CycleGAN lacks effective constraints, resulting in some unnecessary changes or/and losses of information in generated images. For example, the image with tumor often loses the information of tumor in the medical image conversion process.

Based on a large number of experiments, we found that the degrees of differences among the images in the training set have a great influence on the qualities of the generated images. For example, in the process of generating fruit images with our specified freshness, some unsatisfactory images and even some weird image results will be generated when the number of objects in the image set are greatly different, as shown in Fig. 1. The surfaces of apple(s) in the generated target images, Fig. 1 (c) or Fig. 1 (f), are covered with messy smears due to the multi-object feature of the style offering images. The attributes that have great influences on the results include the number of objects, the category of objects and the freshness of objects.

Based on the conclusions of the above-mentioned experimental analysis, we boldly put forward a hypothesis. The unsatisfactory results will be greatly reduced through avoiding or decreasing the attributes mismatching situations between the style offering images and the input source images during the training processes. The attribute registration mechanism is introduced to arrange the unordered training set into the matched training set, which will hopefully solve the problem of unsatisfactory results. Preliminary experimental results verified our above hypothesis.

We have implemented a new network which can effectively avoid unsatisfactory results by introducing an attribute registration mechanism to CycleGAN. Image classification and image recognition technology are adopted to extract the attributes such as the number of main objects, freshness and other characteristics in the images. The screening weights are assigned to each image in raw image set based on the extracted attributes. The best matching of the images are achieved based on similarity detection. Stable and efficient fruit freshness attribute transferring can be realized through this new network.

This paper focuses on the quality improvement and quality stability in the process of freshness attribute transferring of fruit images. We built an improved image generation network, named arCycleGAN, which introduced the mechanism of attribute registration into CycleGAN. The new network can generate stable and excellent fruit image sequences with different freshness levels for an input image. It can generate more convincing images on the same image set. Moreover, it can generate the images of similar quality based on a smaller training set in less training time.

The innovations of new network are listed as follows. First, an automatic recognition method combining depth target detection and image classification is used to screen the training sets. Second, similarity detection method is used to pair the images in the screened training set for the generator module. Third, the new network can be used to train multiple generators to build image sequence generators, where each generator can generate the target image with different freshness attributes. CycleGAN is still used as the technical basis to enable image style transferring among unpaired images.

The rest of this paper is organized as follows. The related work will be discussed in the second part. The architecture of the proposed network will be described in the third part. The fourth part will discuss the experimental results of the new network. The fifth part will give the comparisons and analysis between our new network and the original CycleGAN. Finally, we will present the conclusion of our work and the direction of further researches.

II. RELATED WORK

A. IMAGE-TO-IMAGE CONVERSION

Image-to-image conversion is a hot research field in the image field. GANs (Generative Adversarial Networks) is the framework for achieving significant results in addressing this problem. Goodfellow *et al.* [5] proposed the GAN framework for generating models through the adversarial process estimation. GAN framework can train the generator network and the discriminator network at the same time.

Isola *et al.* [6] conducted image-to-image conversion based on CGANs (Conditional Generative Adversarial Nets). Image-to-image experiments based on paired data sets show that CGANs is a promising conversion method, especially for images that involve highly structured graphic output. Image-to-image conversion usually requires the paired images to learn the mapping between the input images and the output images. But paired training images are difficult to acquire in many tasks.

The adjusted DCGAN model [3] can be improved its performance by adjusting the super parameters such as learning rate, momentum and batch size. Zhu *et al.* [7] proposed a CycleGAN model for bidirectional style transferring, which converts unpaired images from source domain X to target domain Y through two generators and two discriminators.

B. IMPROVEMENT OF CYCLEGAN

The improvement measures for CycleGAN include two aspects, one is improving the quality of generated images, and the other is reducing the complexity and parameters of the model training.

In terms of improving the quality of generated images, module introducing, constraint adding and process decomposing are three common strategies.

Nazki *et al.* [8] proposed the AR-GAN (Activation Reconstruction GAN) by introducing an activation reconstruction module into CycleGAN. The activation reconstruction module can help enhance the sense of perception between the original image and the generated image as well as improve the stability of the generated image. The AR-GAN can be used to generate disease images based on health images to solve the problem of the unbalanced distribution in a tomato leaves image set. DiscoGAN [9] and DualGAN [10] use similar ideas to achieve bidirectional conversion between two image domains. But the reconstructions in both methods have problems like image blurring and information loss.

LeafGAN [2] uses a new background similarity loss function, which makes the algorithm pay more attention to ROIs (Region of Interest), resulting in making the background of generated images as close as possible to the original input images. CSGAN (Cyclic-Synthesized Generative Adversarial Networks) [11] increases the consistency losses between reconstructed images and synthesized images in CycleGAN to improve the qualities of generated images. Similarly, CDGAN (Cyclic Discriminative Generative Adversarial Networks) [12] increases the discriminating losses on the reconstructed images in CycleGAN structure. Lin *et al.* [13] added an auxiliary domain between the source domain and the target domain to CycleGAN network. The constraint of the source domain images and the auxiliary domain images can help to reduce the randomness of the converted target images and to improve the qualities of the generated images. However, the consistency constraints of the auxiliary and target domains increase the number of parameters in generators and training difficulties of generators. OT-CycleGAN (Optimal Transport-CycleGAN) [14] uses the Optimal Transport to add extra constraints to achieve a controllable one-to-one mapping to implement the attribute transformation in some specific tasks.

SCAN (Stacked Cycle-Consistent Adversarial Networks) [15] decomposes the image generation process into multiple stages, which are combined by CycleGAN. In each stage of generation process, the model generates images with different resolutions. The multi-stage strategy of SCAN not only improves the quality of the generated images during the same resolution conversion process, but also enables higher resolution generated images to be obtained. But there are still problems with SCAN like random mapping and information loss.

In terms of reducing the complexity and parameters of the model training, CycleGAN++ [16] removes the bidirectional consistency constraint and cascades the prior information of the target domain and the source domain in the image generation stage. This method can improve the training rate and reduce the computational complexity by cancelling the annular structure in CycleGAN.

C. IMAGE CLASSIFICATION AND IDENTIFICATION

There are many classification networks that can be used as references in fruit image attributes recognition. The typical networks include deep convolution neural network, lightweight neural network and deep separable network.

As the cornerstone of modern convolution neural network, LeNet5 [17] is one of the earliest deep convolution neural networks. It uses convolution, parameter sharing, pooling and other steps to extract features and uses fully connected neural network for classification and recognition. AlexNet [18] successfully uses ReLU, Dropout and LRN on the basis of LeNet5. VGGNet [19] constructs a convolution neural network with 16-19 layers by repeatedly stacking 3∗3 convolution kernels and 2∗2 maximum pooling layers. ResNet [20] proposed a residual learning unit to solve the information loss problem that may occur in convolutional layer or fully connected layer. It has 152 layers, but the network parameters are lower than VGGNet.

Although these networks have good accuracy rate in image recognition tasks, the huge storage and computing cost affect their application. Lightweight neural network model is proposed to solve this problem. SqueezeNet [21] reduces the number of parameters by replacing 3∗3 convolution with 1∗1 convolution and reducing the number of channels of 3∗3 convolution. It takes only 1/50 of the AlexNet parameters, but it can achieve the approximate effect and accuracy of AlexNet on the IMAGENET.

MobileNetV3 [22] proposed hard Sigmoid instead of traditional Sigmoid in order to reduce the cost of Swish nonlinear activation. MicroNet [23] proposed micro-factorized convolutional layer which decomposes the depth convolution into a low-rank matrix. The layer can help to achieve a good balance between the number of channels and input/output connectivity.

D. ATTRIBUTE MATCHING AND REGISTRATION

Attribute matching is an important technology in the attribute registration stage of the network proposed in this article. Tang *et al.* [24] proposed a novel heterogeneous network based on the attribute dependence. The attribute dependence can help to improve the performance of access selection by reducing blocking and handoff dropping rate.

Linlin *et al.* [25] proposed an Attribute-GAN to generate clothing-match pairs automatically. The Attribute-GAN builds a large-scale outfit dataset and annotates manually clothing attributes, such as color, texture, shape of dressing. The Attributed-GAN can be applied to match the dressing styles according to the similarity of semantic attributes. Fengfeng *et al.* [26] proposed a cohesion-based metric to improve the matching performance among distinct attribute values. The cohesion-based metric can effectively avoid the

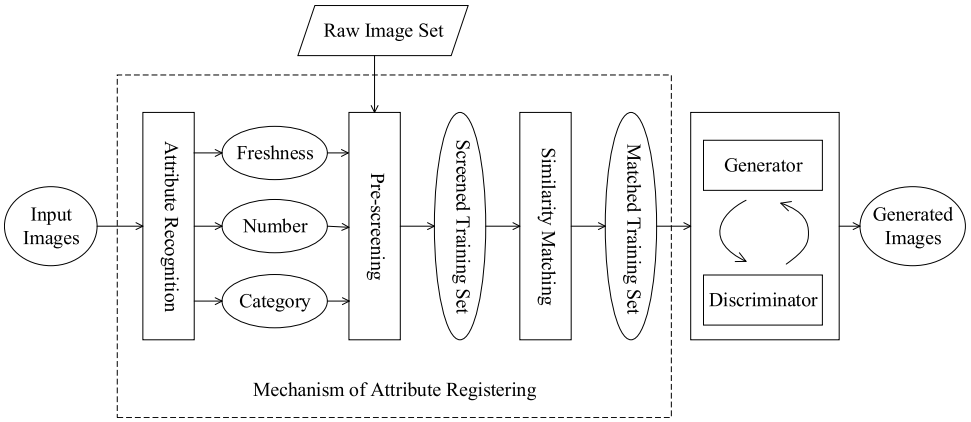


FIGURE 2. Architecture and operational flow of the training stage in arCycleGAN network.

limitations such as similarity metric, threshold and domain knowledge.

Wenhao *et al.* [27] proposed DEXIN (Dynamic EXclusive and INclusive), a content-based multi-attribute event matching algorithm. The multi-attribute matching algorithm can support efficient and stable event matching in multiple scenes.

HeeChun [28] proposed an algorithm of matching potential partners based on players' multi-attribute preferences. The matching algorithm attempts to achieve enough stimuli for players with incomplete preferences and unrevealed attributes in each round of the two-sided games. Ronald *et al.* [29] refined the similarity measure by considering importance weights of attributes of objects and events. The method deemed certainly not all attribute combinations are typically of the same interest. The hyper-matching of objects or events is provided by the certain combination of the relevant attributes, while the soft-matching is provided by the importance of unusual attribute values. Han-Mu and Kuk-Jin [30] proposed a multi-attribute matching method for multi-layer graphs. The method addresses the ambiguity and uncertainty arisen from the attribute integration of the multi-layer structure.

III. ARCHITECTURE OF ARCYCLEGAN

The proposed arCycleGAN network can generate target fruit images with specified freshness attributes based on input source fruit images. It introduces the mechanism of attribute registration into CycleGAN to improve the effectiveness of the training process. The arCycleGAN consists of four parts. The first part is the attribute recognition module. The module obtains correct attributes of objects in images. The second part is the pre-screening module. The module filters out appropriate image subsets from raw image set according to the attributes of the input source images. The third part is the similarity matching module. The module sets the matching rules for the screened training set of CycleGAN. The fourth part includes generators and discriminators similar to CycleGAN.

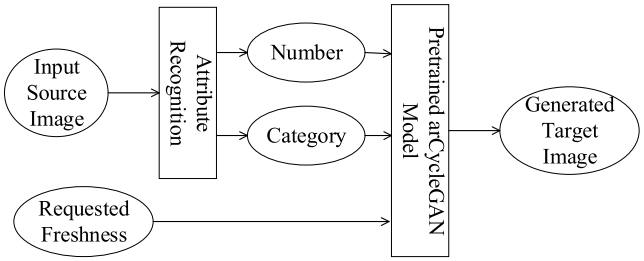


FIGURE 3. Architecture and operational flow of the testing stage in arCycleGAN network.

Figure 2 shows the structure and main operational flow of training stage in the arCycleGAN network. The pre-screening can filter the images from the raw image set. The similarity matching is used to pair the images in the screened training set. The image generation model is trained based on the matched training set. Figure 3 shows the structure and main operational flow of the testing stage in the arCycleGAN network. For an input source image, the corresponding target image can be obtained by selecting a suitable pre-trained model according to the number and category of objects in the source image as well as the requested freshness.

A. ATTRIBUTE RECOGNITION BASED ON ALEX-YOLO

The purpose of this module is to identify the attributes of fruit images. The module recognizes three attributes, freshness, number and category of objects.

AlexNet is selected to identify the freshness attribute of fruits in images. AlexNet has a simpler network structure and lower computing cost, which can better meet our requirements. YOLO V2 is chosen as the target detection model to recognize the number and category of objects in images. YOLO V2 can achieve a high detection accuracy on the premise of maintaining an acceptable detection speed. Tiny-YOLO-VOC is selected as the basic network of YOLO V2 because of its simple structure and low computational complexity.

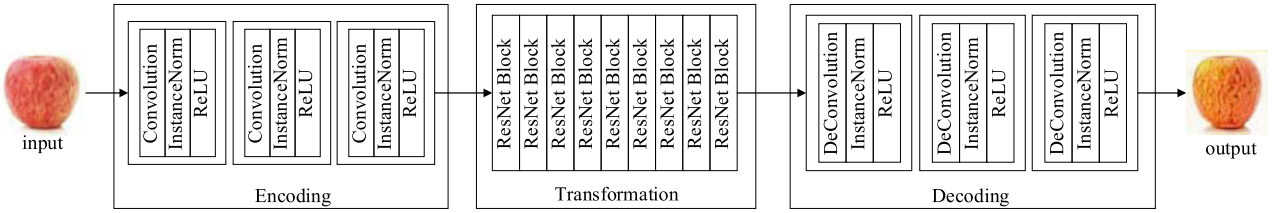


FIGURE 4. The structure of generator in ArCycleGAN network.

For images in raw image set, the number N_i, the category C_i and the freshness F_i of objects can be obtained offline through the above two models. When there are multiple objects in an image, the category of objects with the largest quantities is taken as the category of the image, while the freshness of objects with the largest quantities is taken as the freshness of the image.

Based on the above recognition results, the attribute vector $NCF_i = (N_i, C_i, F_i)$ is established for each image in raw image set.

For an input source image, the number N_s, the category C_s and the freshness F_s of objects are also obtained by the above method. The attribute vector of input source image is established as $NCF_s = (N_s, C_s, F_s)$. Meanwhile, the attribute vector of the generated target image is established as $NCF_t = (N_s, C_s, F_t)$ according to the attributes of the input source image and the requested freshness F_t.

B. IMAGE PRE-SCREENING BASED ON ATTRIBUTE HOMOGENEITY

The screening weight of each image in raw image set is calculated based on its attribute vector NCF_i. The calculation method of screening weight is shown as Formula (1) and Formula (2).

$$\begin{aligned} W_{is} &= \|A \cdot (NCF_i - NCF_s)\| \\ &= a_1 * (N_i - N_s)^2 + a_2 * (C_i - C_s)^2 \\ &\quad + a_3 * (F_i - F_s)^2 \end{aligned} \tag{1}$$

$$\begin{aligned} W_{it} &= \|A \cdot (NCF_i - NCF_t)\| \\ &= a_1 * (N_i - N_s)^2 + a_2 * (C_i - C_s)^2 \\ &\quad + a_3 * (F_i - F_t)^2 \end{aligned} \tag{2}$$

In Formula (1), W_{is} is the screening weight between the i^{th} image and the input source image. $A = (a1, a2, a3)$ is the setting vector, which is usually set by users. For example, it can be set as $A = (0.3, 0.3, 0.4)$. NCF_s is the attribute vector of the input source image. Similarly, in Formula (2), W_{it} is the screening weight between the i^{th} image and the generated target image. NCF_t is the attribute vector of the generated target images.

The screening weight of each image in the raw image set will vary with different input source images and requested freshness. Appropriate images are screened from the raw image set according to the screening weight of each image and some randomness. The combination of screening weight and randomness is to maintain a high quality and the diversity of generated target images. Finally, the screened images are jointly formed into a screened training set.

Algorithm 1 Similarity Matching for Images

Input: input source image, generated target image and raw image set

Output: similarity among images

1. Reduce the size of the image to 8×8 pixels;
2. Convert the reduced image to a 64-level grayscale image;
3. Calculate the grayscale average of all pixels in the image;
4. Compare the gray scale of each pixel with the average value;
5. Mark each pixel as either 1 or 0. Marked as 1 if the pixel gray is greater than or equal to the average, otherwise marked as 0;
6. Calculate the hash value and combine it into a 64-bit binary integer, denoted as a finger;
7. Calculate the similarity according to the finger of the images.

C. TRAINING SET PAIRING BASED ON SIMILARITY DETECTION

The screened images will be matched based on similarity detection. Perceptual Hash Algorithm is adopted to implement the image similarity detection. The implementation process is shown as Algorithm 1.

D. GENERATOR MODULE FOR GENERATED TARGET IMAGES

The network structure of the generator used in this paper is based on CycleGAN network. The network structure of the generator is shown as Fig. 4.

The network structure of the generator can be described in detail as below. Layer 0 is the input layer, whose input are the fruit images. The size of fruit images is 256∗256. Next, there are three Convolution-InstanceNorm-ReLU layers, named from Layer 1 to Layer 3. The step length of these three convolution layers are 1, 2, 2, respectively. The purpose of these three layers is to conduct dimensional-reduction sampling in the feature space. After the dimensional-reduction sampling, there are nine continuous ResNet residual blocks, numbered sequentially from Layer 4 to Layer 12. At last, there are two

TABLE 1. The structure of generator in ArCycleGAN network.

Layer	Layer Name	Convolution Kernel Size	Number of Convolution Kernels	Step Length
Layer 0	Input	-	-	-
Layer 1	Convolution-InstanceNorm-ReLU	7*7	32	1
Layer 2	Convolution-InstanceNorm-ReLU	3*3	64	2
Layer 3	Convolution-InstanceNorm-ReLU	3*3	128	2
Layer 4, 5, 6, 7, 8, 9, 10, 11, 12	ResNet Block	3*3	128	1
Layer 13	Fractional-strided-Convolution-InstanceNorm-ReLU	3*3	64	12
Layer 14	Fractional-strided-Convolution-InstanceNorm-ReLU	3*3	32	12
Layer 15	Convolution-InstanceNorm-ReLU	7*7	3	1

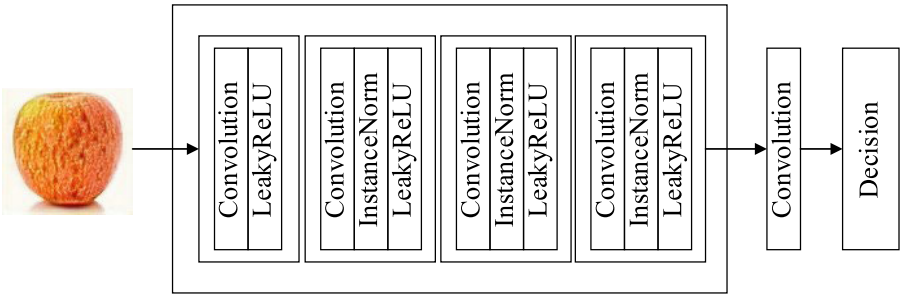


FIGURE 5. The structure of discriminator in ArCycleGAN network.

TABLE 2. The structure of discriminator in ArCycleGAN network.

Layer	Layer Name	Convolution Kernel Size	Number of Convolution Kernels	Step Length
Layer 1	Convolution-LeakyReLU	4*4	64	2
Layer 2	Convolution-InstanceNorm-LeakyReLU	4*4	128	2
Layer 3	Convolution-InstanceNorm-LeakyReLU	4*4	256	2
Layer 4	Convolution-InstanceNorm-LeakyReLU	4*4	512	1
Layer 5	Convolution	4*4	1	1

up sampling modules, that is, transpose convolution. Their purpose is to bring the space and channel back to full size. The parameters, such as convolution kernel size, number and step length of the generator are shown in Table 1.

E. DISCRIMINATOR MODULE FOR IMAGES

PatchGANs is adopted as the network structure of the discriminator. The discriminator has five layers and its network structure is shown as Fig. 5. A Convolution-LeakyReLU layer is taken as the first layer. The following are three Convolution-InstanceNorm-ReLU layers, named from Layer 2 to Layer 4. The fifth layer is a convolutional layer used to generate the one-dimension output. The parameters, such as convolution kernel size, number and step length of each layer in the discriminator are shown in Table 2.

LeakyReLU is selected as the activation function of the discriminator. LeakyReLU is shown as Formula (3), where the value of a is 0.2.

$$LeakyReLU(X)\begin{cases} x, & x \geq 0 \\ ax, & x < 0 \end{cases} \in R \tag{3}$$

In the training process, the goal of generator G is to generate images as real as possible to deceive the discriminator D, while the goal of D is to distinguish the images generated by G from the real images as far as possible. The objective function is shown as Formula (4),

$$\min_G \max_D (D, G) = \mathbb{E}_{x\sim p_{data}(x)}\left[\log D(x)\right] + \mathbb{E}_{z\sim p_z(z)}\left[\log(1 - D(G(z)))\right] \tag{4}$$

FIGURE 6. Two groups of matched training sets: (a) images of a fresh apple at one end of Group 1; (b) images of a wrinkled apple at the other end of Group 1; (c) images of the fresh apples at one end of Group 2; (d) images of the rotten apples at the other end of Group 2.

TABLE 3. Running environments of our experiments.

Item	Detailed Information
CPU	Intel Xeon CPU E5-2609 v3
GPU	NVIDIA Titan RTX 24GB
Operating System	Windows 10(64-bit)
Python	Python 3.7
Tensorflow	Tensorflow 2.1.0

where x represents the real image. z represents the noise of the input G network. $G(z)$ represents the image generated by G network, $D()$ represents the probability of Network D to determine whether the image is authentic.

IV. EXPERIMENTAL RESULTS

We have carried out detailed experiments to verify the arCycleGAN proposed in this paper. The experimental results show that arCycleGAN can generate better results and have better performance than the original CycleGAN in terms of required training set, training rounds and training time.

We adopted the GPU version of TensorFlow as the running environment. The detailed hardware configurations and software versions are shown in Table 3.

A. RAW IMAGE SET

We collected fruit images with different freshness attributes as our raw image set. The raw image set mainly contains three parts which are the fruit subset in IMAGENET, the fruit images crawled from the network and the images taken by ourselves. Images were classified into four categories including Fresh, Dull, Wrinkled and Rotten according to the freshness of the fruits. The numbers of images with different freshness attributes are shown in Table 4.

TABLE 4. Number of images with different freshness attributes in raw image set.

Category	IMAGENET	Ineternet	Ourselves	Total
Fresh	159	263	69	491
Dull	31	269	60	360
Wrinkled	94	220	46	360
Rotten	152	206	33	390

B. B RAW IMAGE SET AND MATCHED TRAINING SET

In the training phase of CycleGAN and arCycleGAN, the screened training set is selected from the raw image set. The number of images in the screened training set can usually be preset from 60 to 200. If it is less than 60, the quality of the generated images could be inacceptable. If it is greater than 200, the quality of the generated images will be difficult to improve.

There are three main differences between the training set used in CycleGAN and the matched training set used in arCycleGAN. Firstly, the training set of CycleGAN is randomly selected, while the matched training set of arCycleGAN is

FIGURE 7. Experimental results of freshness attribute transferring: (a) input source images of the fresh apple(s); (b) generated target images of the dull apple(s); (c) generated target images of the wrinkled apple(s); (d) generated target images of the rotten apple(s).

composed of the closer images selected according to the category, quantity, and freshness attributes of objects. Secondly, the images at both ends in CycleGAN are randomly matched, while the images at both ends in arCycleGAN are optimally matched based on the image similarity. At last, the importance of the images in the training set of CycleGAN is the same, while the images in the matched training set of arCycleGAN are valued as different weights according to the image contents.

Figure 6 shows an example of a matched training set. The numbers and categories of objects in all images are close. The freshness attributes of the fruits in images at the same end are similar, and the freshness attributes at the other end are different. Compared with CycleGAN, the better performance of arCycleGAN is reflected in a smaller training set and fewer training rounds.

C. GENERATED TARGET IMAGE WITH THE SPECIFIED FRESHNESS

We use arCycleGAN to change the freshness attributes for apple images. Examples of generated target images are shown in Fig. 7. The fresh fruit images in Fig. 7 (a) are the input source images in our arCycleGAN. Images in Fig. 7 (b), (c) and (d) are the corresponding generated target images with Dull, Wrinkled and Rotten freshness attributes respectively.

From Fig. 7, we can denoted that arCycleGAN can generate more convincing image results. The generated target images have the specified freshness attributes on premise of maintaining the same shapes and details as the input source images.

D. APPEARANCE TRANSFERRING BETWEEN DIFFERENT FRUIT CATEGORIES

The arCycleGAN can not only change freshness attributes of fruit images, but also transfer appearances between different categories of fruit images. Figure 8 shows the appearance transferring results between Apple and Orange. As can be seen in Fig. 8, arCycleGAN can achieve attractive performances during appearance transferring.

V. PERFORMANCE EVALUATION AND ANALYSIS

Compared with CycleGAN, the mechanism of attributes registration, similar to the attention mechanism, are introduced into the preparation process of the training set of arCycleGAN. As a result, arCycleGAN acquires two aspects of improvement effects. In the first aspect, arCycleGAN can produce more stable style transferring results on the training image set of the same size, or only needs a smaller training image set to obtain the results of similar quality. In the second aspect, the convergence speed in the generator module is improved, resulting in arCycleGAN only requiring fewer training rounds and less training time.

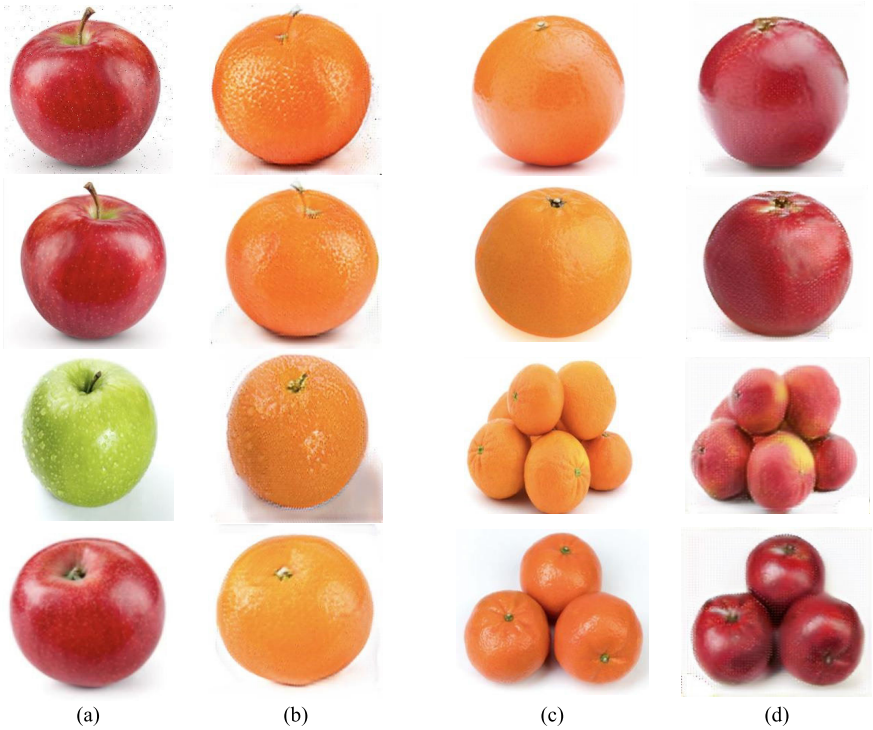


FIGURE 8. Experimental results of appearance transferring between Apple and Orange: (a) input source images of the apple; (b) generated target images of the orange; (c) input source images of the orange(s); (d) generated target images of the apple(s).

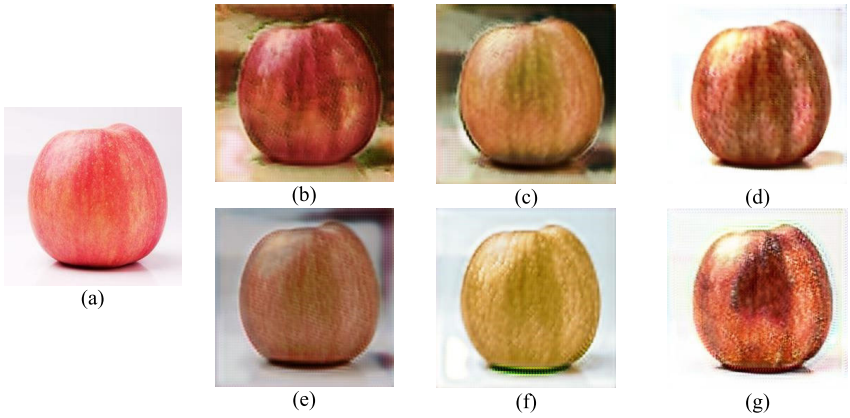


FIGURE 9. Comparisons of the results of CycleGAN and arCycleGAN when the number of images in the training set is 40: (a) the input source image of the fresh apple; (b) the dull apple image generated by CycleGAN; (c) the wrinkled apple image generated by CycleGAN; (d) the rotten apple image generated by CycleGAN; (e) the dull apple image generated by arCycleGAN; (f) the wrinkled apple image generated by arCycleGAN; (g) the rotten apple image generated by arCycleGAN.

A. IMPROVEMENT EFFECTS OF ARCYCLEGAN

ArCycleGAN can generate the superior results than the original CycleGAN based on a training set containing the same number of images.

Figure 9 shows the comparison of the generated target images between the original CycleGAN and arCycleGAN when the number of images in the training set is 40. Figure 10 shows another comparison of the target images generated by the CycleGAN and arCycleGAN when there are 100 fruit images in the training set.

It can be found from Fig. 9 and Fig. 10 that arCycleGAN can better maintain the shapes of the fruits and show more obvious indications of freshness under the same size of training image set.

B. REDUCING THE SIZE OF THE TRAINING SET

Compared with the original CycleGAN, arCycleGAN only needs smaller training set to generate images of the same quality.

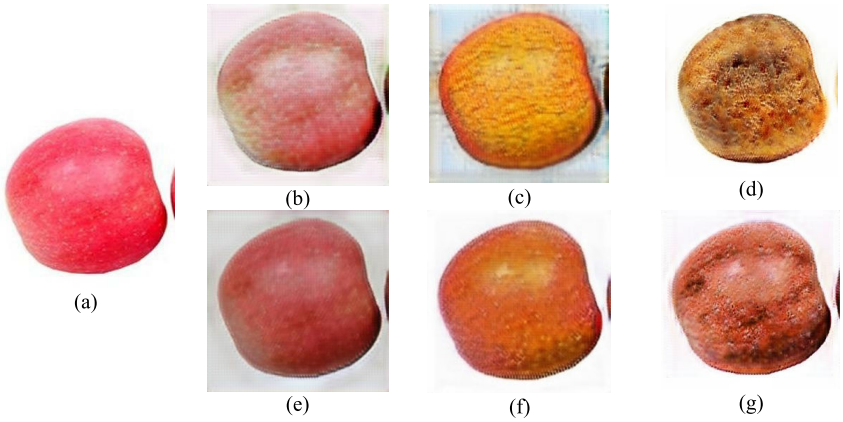


FIGURE 10. Comparisons of the results of CycleGAN and arCycleGAN when the number of images in the training set is 100: (a) the input source image of the fresh apple; (b) the dull apple image generated by CycleGAN; (c) the wrinkled apple image generated by CycleGAN; (d) the rotten apple image generated by CycleGAN; (e) the dull apple image generated by arCycleGAN; (f) the wrinkled apple image generated by arCycleGAN; (g) the rotten apple image generated by arCycleGAN.

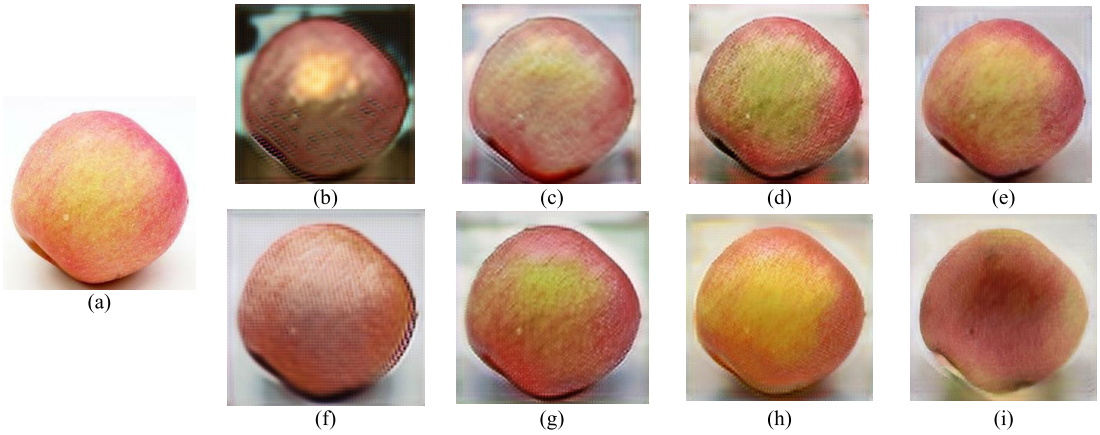


FIGURE 11. Comparisons of the dull apple images generated by CycleGAN and arCycleGAN on training sets of different sizes: (a) the input source image of the fresh apple; (b) generated by CycleGAN on the training sets of size 40; (c) generated by CycleGAN on the training sets of size 60; (d) generated by CycleGAN on the training sets of size 100; (e) generated by CycleGAN on the training sets of size 200; (f) generated by arCycleGAN on the training sets of size 40; (g) generated by arCycleGAN on the training sets of size 60; (h) generated by arCycleGAN on the training sets of size 100; (i) generated by arCycleGAN on the training sets of size 200.

Figure 11 shows comparisons of the dull apple images generated by CycleGAN and arCycleGAN on training sets of different sizes. The sizes of the training sets used in the experiments in Fig. 11 are 40, 60, 100 and 200, respectively.

Three preliminary conclusions can be drawn from Fig. 11. Firstly, through the horizontal comparison, it can be found that as the size of the training set are becoming larger, the images generated by both networks are gradually getting better. Secondly, through the longitudinal comparison, it can be found that on the training sets of the same size, arCycleGAN can produce better images than CycleGAN. Finally, arCycleGAN can generate images of similar quality using a smaller training set than CycleGAN. The conclusion can be illustrated by Fig. 11 (d) being the closest to Fig. 11 (g) and Fig. 11 (e) being the closest to Fig. 11 (h).

Figure 12 shows comparisons of the wrinkled apple images generated by CycleGAN and arCycleGAN on training sets of different sizes. The sizes of the training sets used in the experiments in Fig. 12 are also 40, 60, 100 and 200, respectively. The three above preliminary conclusions can also be drawn from Fig. 12, similar to Fig. 11. In Fig. 12, these phenomenon also can be seen that the horizontal images are getting better and better, and the longitudinal results are also better, and smaller training sets are required to generate images of similar quality.

Through a large number of experiments, compared with CycleGAN, arCycleGAN only needs about half the size of the training set to obtain images of the same quality.

C. REDUCING TRAINING ROUNDS AND TRAINING TIME

Compared with the CycleGAN, arCycleGAN requires fewer training rounds to generate images of similar quality. Table 5 describes in detail the reduction ratio of training rounds required by arCycleGAN compared to CycleGAN.

The reduction of the training rounds of arCycleGAN is due to its faster convergence speed caused by the introduced

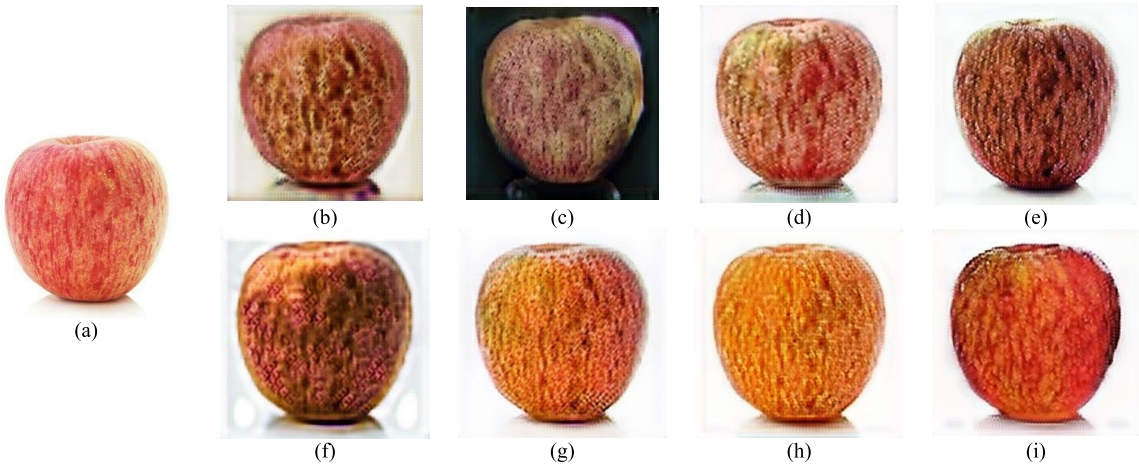


FIGURE 12. Comparisons of the wrinkled apple images generated by CycleGAN and arCycleGAN on training sets of different sizes: (a) the input source image of the fresh apple; (b) generated by CycleGAN on the training sets of size 40; (c) generated by CycleGAN on the training sets of size 60; (d) generated by CycleGAN on the training sets of size 100; (e) generated by CycleGAN on the training sets of size 200; (f) generated by arCycleGAN on the training sets of size 40; (g) generated by arCycleGAN on the training sets of size 60; (h) generated by arCycleGAN on the training sets of size 100; (i) generated by arCycleGAN on the training sets of size 200.

TABLE 5. Reduction ratio of training rounds of the new model.

Number of images in the training set	Generator for dull fruits	Generator for wrinkled fruits	Generator for rotten fruits
40	11.11%	3.85%	2.40%
60	3.03%	5.17%	2.40%
100	3.74%	3.23%	5.35%
200	5.00%	6.47%	7.77%

TABLE 6. Reducing ratio of training time of ArCycleGAN.

Number of images in the training set	Generator for dull fruits	Generator for wrinkled fruits	Generator for rotten fruits
40	5.71%	3.03%	6.06%
60	5.56%	1.89%	3.70%
100	4.76%	2.50%	3.61%
200	8.78%	10.60%	5.93%

mechanism of attribute registration. It can be concluded from Table 5 that, compared with the original CycleGAN, the training rounds needed by arCycleGAN can be reduced by 5.0% on average and the highest can be reduced by 11.1%.

Since arCycleGAN only requires a smaller training set and fewer training rounds, the training time is significantly reduced. Table 6 describes in detail the reduction ratio of training time required by arCycleGAN compared to CycleGAN.

It can be concluded from Table 6 that, compared with the original CycleGAN, the training time needed by arCycleGAN can be reduced by 5.2% on average and the highest can be reduced by 10.6%.

VI. CONCLUSION AND FUTURE WORK

We proposed a novel style transferring network, named arCycleGAN, which can generate fruit images with specified freshness while maintaining the shapes and main features of the fruit in input source images. The new network also enables appearance transferring between different categories of fruits images. On the basis of CycleGAN, arCycleGAN introduces attribute registration mechanism to improve the stability and the qualities of the generated target images. The attribute registration mechanism, implemented by the three modules of attribute recognition, pre-screening and similarity matching, can effectively improve the performance of style transferring and reduce the computational cost of the training process. Detailed experimental results show that, compared with CycleGAN, arCycleGAN can generate images of similar quality on only about half the size of the training set. Meanwhile, the training rounds and training time have also been reduced. The training rounds required by arCycleGAN to generate images of similar quality are reduced by 5.0% on average, and the training time is reduced by 5.8% on average.

The proposed network, arCycleGAN, can be widely used to expand the data set of fruit images with different freshness attributes, or to improve the qualities of fruit images taken by cameras. The next of this paper is to further improve the adaptive adjustment capabilities of network parameters.

REFERENCES

[1] Y. Tian, G. Yang, Z. Wang, E. Li, and Z. Liang, "Detection of apple lesions in orchards based on deep learning methods of CycleGAN and YOLOV3-dense," *J. Sensors*, vol. 2019, Apr. 2019, Art. no. 7630926, doi: 10.1155/2019/7630926.

[2] Q. H. Cap, H. Uga, S. Kagiwada, and H. Iyatomi, "LeafGAN: An effective data augmentation method for practical plant disease diagnosis," presented at the CVPR, 2020. [Online]. Available: https://arxiv.org/abs/2002.10100

[3] Q. Wu, Y. Chen, and J. Meng, "DCGAN-based data augmentation for tomato leaf disease identification," *IEEE Access*, vol. 8, pp. 98716–98728, 2020, doi: 10.1109/ACCESS.2020.2997001.

[4] R. Zhang, T. Pfister, and J. Li, "Harmonic unpaired image-to-image translation," presented at the ICLR, 2019. [Online]. Available: https://arxiv.org/abs/ 1909.13028

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, X. Bing, and Y. Bengio, "Generative adversarial nets," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, vol. 2, Dec. 2014, pp. 2672–2680.

[6] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," Presented at the Comput. Vis. Pattern Recognit. (CVPR), 2017. [Online]. Available: https://arxiv.org/abs/1611.07004

[7] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," presented at the CVPR, 2018. [Online]. Available: https://arxiv.org/abs/1703.10593v2, doi: 10.1109/ICCV.2017.244.

[8] H. Nazki, S. Yoon, A. Fuentes, and D. S. Parkcd, "Unsupervised image translation using adversarial networks for improved plant disease recognition," *Comput. Electron. Agricult.*, vol. 168, Jan. 2020, Art. no. 105117. [Online]. Available:https://arxiv.org/abs/1909.11915, doi: 10.1016/j.compag.2019.105117.

[9] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. 34th Int. Conf. Mach. Learn. (ICML)*, vol. 70, 2017, pp. 1857–1865. [Online]. Available: https://arxiv.org/abs/1703.05192

[10] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised dual learning for image-to-image translation," presented at the IEEE Int. Conf. Comput. Vis. (ICCV) 2017. [Online]. Available: https://arxiv.org/abs/1704. 02510

[11] K. K. Babu and S. R. Dubey, "CSGAN: Cyclic-synthesized generative adversarial networks for image-to-image transformation," *Expert Syst. Appl.*, vol. 169, May 2021, Art. no. 114431. [Online]. Available: http://arxiv.org/abs/1901.03554

[12] K. K. Babu and S. R. Dubey, "CDGAN: Cyclic Discriminative Generative Adversarial Networks for Image-to-Image Transformation," 2020, *arXiv:2001.05489*. [Online]. Available: https://arxiv.org/abs/2001.05489

[13] J. Lin, Y. Xia, Y. Wang, T. Qin, and Z. Chen, "Image-to-image translation with multi-path consistency regularization," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Macao, China, Aug. 2019, pp. 2980–2986. [Online]. Available: http://arxiv.org/abs/1905.12498, doi: 10.24963/ijcai.2019/413.

[14] G. Lu, Z. Zhou, Y. Song, K. Ren, and Y. Yu, "Guiding the one-to-one mapping in CycleGAN via optimal transport," presented at the AAAI Conf. Artif. Intell., 2019, vol. 33, no. 1. [Online]. Available: https://arxiv.org/abs/1811.06284

[15] M. Li, H. Huang, L. Ma, W. Liu, T. Zhang, and Y. Jiang, "Unsupervised image-to-image translation with stacked cycle-consistent adversarial networks," presented at the Eur. Conf. Comput. Vis. (ECCV), 2018. [Online]. Available: https://arxiv.org/pdf/1807.08536, doi: 10.1007/978-3-030-01240-3_12.

[16] J. Zhang and Y. Hou, "Image-to-image translation based on improved cycle-consistent generative adversarial network," *J. Electron. Inf.*, vol. 42, no. 5, pp. 1216–1222, 2020.

[17] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst. (NPIS)*, vol. 1, Dec. 2012, pp. 1097–1105.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," presented at the ICLR, San Diego, CA, USA, 2015. [Online]. Available: https://arxiv.org/abs/ 1409.1556

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," presented at the IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2015, Las Vegas, NV, USA, doi: 10.1109/CVPR.2016.90.

[21] N. F. Iandola, S. Han, W. M. Moskewicz, K. Ashraf, J. W. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size," presented at the ICLR, 2016. [Online]. Available: https://arxiv.org/abs/ 1602.07360

[22] A. Howard, M. Sandler, B. Chen, W. Wang, L. Chen, M. Tan, G. Chu, V. Vasudevan, and Y. Zhu, "Searching for MobileNetV3," presented at the IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2019. [Online]. Available: https://arxiv.org/abs/1905.02244

[23] Y. Li, Y. Chen, X. Dai, D. Chen, M. Liu, L. Yuan, Z. Liu, L. Zhang, and N. Vasconcelos, "MicroNet: Towards image recognition with extremely low FLOPs," 2020, *arXiv:2011.12289*. [Online]. Available: https://arxiv.org/abs/2011.12289

[24] L. Tang, S. Ji, and J. Yan, "A heterogeneous network access selection algorithm based on attribute dependence," *Wireless Pers. Commun.*, vol. 92, no. 3, pp. 1163–1176, Feb. 2017. [Online]. Available: https://link.springer.com/article/10.1007/s11277-016-3600-6, doi: 10.1007/s11277-016-3600-6.

[25] L. Liu, H. Zhang, Y. Ji, and Q. M. J. Wu, "Toward AI fashion design: An attribute-GAN model for clothing match," *Neurocomputing*, vol. 341, pp. 156–167, May 2019. [Online]. Available: https://www.sciencedirect.com/ science/article/pii/S0925231219303133, doi: 10.1016/j.neucom.2019.03.011.

[26] F. Fan, Z. Li, and Y. Wang, "Cohesion based attribute value matching," in *Proc. 10th Int. Congr. Image Signal Process., Biomed. Eng. Inform. (CISP-BMEI)*, Shanghai, China, 2017. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8302312

[27] W. Fan, Y. Liu, and B. Tang, "DEXIN: A fast content-based multi-attribute event matching algorithm using dynamic exclusive and inclusive methods," *Future Gener. Comput. Syst.*, vol. 68, pp. 289–303, Mar. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X1630437X, doi: 10.1016/j.future.2016.10.020.

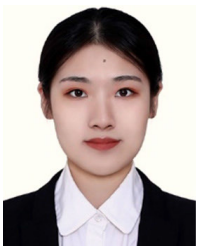
[28] K. HeeChun, "Repeated two-sided matching with multi-attribute preference," *Korean J. Ind. Org.*, vol. 27, no. 4, pp. 87–121, 2019. [Online]. Available: http://kiss.kstudy.com/thesis/thesis-view.asp?key=3747041, doi: 10.36354/KJIO.27.4.3.

[29] Y. Ronald, P. Fred, and E. Paul, "Multiple attribute similarity hypermatching," *Soft Comput.*, vol. 22, no. 8, pp. 2463–2469, 2018. [Online]. Available: https://link.springer.com/article/10.1007/s00500-017-2721-5, doi: 10.1007/s00500-017-2721-5.

[30] P. Han-Mu and Y. Kuk-Jin, "Exploiting multi-layer graph factorization for multi-attributed graph matching," *Pattern Recognit. Lett.*, vol. 127, pp. 85–93, Nov. 2019. [Online]. Available: https://arxiv.org/pdf/1704.07077.pdf, doi: 10.1016/j.patrec.2018.09.024.

HONGQIAN CHEN was born in Shandong, China, in 1982. He received the Ph.D. degree in computer application technology from the Beijing Institute of Technology, in 2009.

From 2009 to 2013, he was a Lecturer with the School of Computer Science and Engineering, Beijing Technology and Business University, where he has been an Assistant Professor, since 2013. He is the author of two books and more than 40 articles. He is also the inventor of more than 20 inventions. His research interests include computer vision, computer graphics, image generation, and data mining.



MENGXI GUAN was born in Beijing, China, in 1997. She received the B.S. degree in software engineering from Beijing Technology and Business University, in 2019, where she is currently pursuing the master's degree with the School of Computer Science and Engineering.

Her research interests include data visualization, computer vision, and image generation.

HUI LI was born in Shandong, China, in 1983. She received the Ph.D. degree in computer application technology from the University of Science and Technology, Beijing, in 2014.

From 2014 to 2018, she was a Lecturer with the College of Management, Beijing Union University, where she has been an Assistant Professor, since 2018. She is the author of one book and more than 20 articles. She is also the inventor of one invention. Her research interests include data analysis, recommended system, and data mining.