Generative Adversarial Networks for Parallel Vision*

WANG Kunfeng[1,2], LI Xuan[3], YAN Lan[1,4], WANG Fei-Yue[1,5]

1(The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190)

2(Parallel Vision Innovation Technology Center, Qingdao Academy of Intelligent Industries, Qingdao 266000)

3(Key Laboratory of Intelligent Control and Decision of Complex Systems, School of Automation, Beijing Institute of Technology, Beijing 100081)

4(School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100049)

5(Research Center of Computational Experiments and Parallel Systems Technology, National University of Defense Technology, Changsha 410073)

***Abstract*—Video image dataset is playing an essential role in design and evaluation of traffic vision methods. However, there is a longstanding difficulty that manually collecting and annotating large-scale diversified dataset from real scenes is time-consuming and prone to error. In 2016, we proposed the parallel vision methodology to tackle the issues of conventional vision computing approach in data collection, model learning and evaluation. We built the ParallelEye dataset with virtual reality and the scene-specific virtual pedestrian dataset with augmented reality. In the dataset compiling process, the graphics rendering engine was used to render the artificial scenes and generate virtual images. However, the fidelity of virtual images is not satisfactory due to limitation of rendering engine, so that there is a distribution gap between virtual data and real data. In our opinion, Generative Adversarial Networks (GANs) can generate more realistic images for parallel vision research. We introduce some GANs and explain their utility in parallel vision.**

Keywords—Computer vision; parallel vision; parallel imaging; Generative Adversarial Networks; parallel system theory

I. INTRODUCTION

Computer vision is an interdisciplinary research field that deals with how computers can be used to fulfill high-level understanding from digital images or videos. In particular, computer vision can realize the perception and understanding of the physical world [1]–[8]. As we know, dataset plays a very important role in training and evaluation of computer vision algorithms. Unfortunately, it is time-consuming to collect large-scale diversified video image dataset, especially from complex traffic scenes. Moreover, manually annotating large-scale dataset is expensive and prone to error.

Traditional computer vision methods face some challenges and difficulties in data acquisition, model learning and evaluation. First, the actual application environment is very complex, such as day and night changes, adverse weather, shadows, target occlusion, and scene clutter. Second, it is time-consuming to acquire and annotate large-scale dataset collected from real scenes. Third, there is subjectivity in manually labeled dataset. Fourth, the real scene is uncontrollable and cannot be repeated in accordance with our intention. Therefore, the effect of each component factor of the scene on vision algorithms cannot be analyzed separately. To address the above issues, Wang *et al.* [9][10] proposed the theoretical framework of parallel vision. They use the computer graphics and virtual reality technology to build photorealistic artificial scenes. Afterwards, they obtain large-scale diversified virtual dataset from the artificial scene, and generate precise annotation information automatically. The virtual dataset is combined with real dataset to learn and evaluate the computer vision algorithms [11].

In 2007, Bainbridge [12] used Second Life and World of Warcraft games for research of sociology, behavioral and financial sciences. In 2013, Prendinger *et al.* [13] used virtual laboratory to simulate the traffic environment, and then analyzed the driving behavior. With the development of data-driven computer vision, people begin to train visual algorithms with virtual data. In 2016, Ros *et al.* [14] set up a virtual city scene to study semantic segmentation. They added buildings, roads, bicycles and pedestrians to the virtual scene to make the scene realistic. In addition, they configured different weather for the virtual scenes, which increased the diversity of virtual data and improved the accuracy of semantic segmentation by using deep neural networks. Gaidon *et al.* [15] cloned the KITTI dataset and generated a virtual KITTI dataset for multi-object tracking (MOT). However, the idea of using virtual dataset to train and test algorithms is still in a relatively preliminary exploration stage . Some problems need to be solved urgently, such as domain-shift. This problem is due to the fact that there is a distribution gap between real data and virtual data. In 2014, Goodfellow *et al.* [16] proposed the Generative Adversarial Networks(GANs), which was used mainly for image generation at that time. After three years of development, researchers proposed EBGAN (Energy-based GAN) [17], Wasserstein GAN [18], BEGAN (Boundary Equilibrium GAN) [19], and many other GAN variants [20][21]. The training stability of the model and the fidelity of the generated image samples get improved significantly. In order to solve the domain shift problem and get better training effect, we can use the virtual dataset as input of GANs to generate new artificial image data that are closer to the actual image distribution.

The rest of this paper is organized as follows. Section II introduces the research status of parallel vision. Section III

*Supported by National Natural Science Foundation of China (61533019, 71232006, 91520301)

describes the development and evolution of GANs. Finally, concluding remarks are given in section IV.

II. RESEARCH STATUS OF PARALLEL VISION

Parallel vision [9][10] is an extension of the ACP (Artificial societies, Computational experiments, and Parallel execution) approach [22]–[26] into the computer vision field.

ACP = Artificial societies + Computational experiments + Parallel execution

For parallel vision, artificial scenes are used to simulate and represent complex challenges in actual scenes, computational experiments are conducted to design and evaluation of various visual algorithms, and parallel execution is achieved to optimize the vision system in both the actual scene and the artificial scene. The basic framework and architecture for parallel vision is shown in Fig. 1.

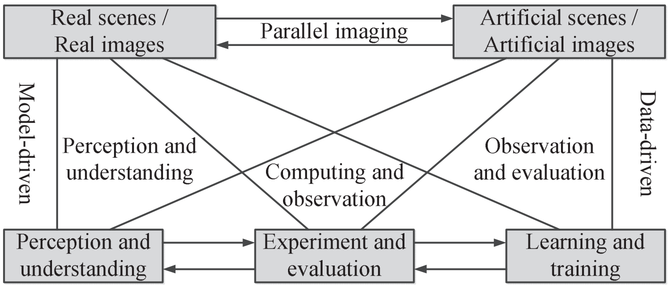

Fig. 1. Basic framework and structure of parallel vision, in which parallel imaging provides the real and artificial images.

The core unit of the parallel imaging is the software-defined artificial imaging system. From the actual scenes, we obtain specific image "small data", which are input into the artificial imaging system. The properties of real images "small data" are analyzed and a large amount of new artificial image data are generated. Fig. 2 shows the technical flowchart of parallel imaging: real images "small data" → parallel imaging "big data" → specific "small knowledge".

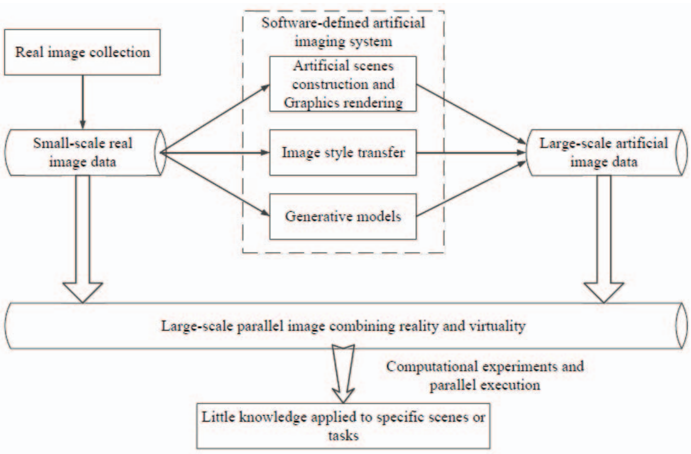

Fig. 2. Technical flowchart of parallel imaging.

The relationship between parallel vision [9][10], parallel imaging [11], and parallel learning [25] is illustrated in Fig 3.

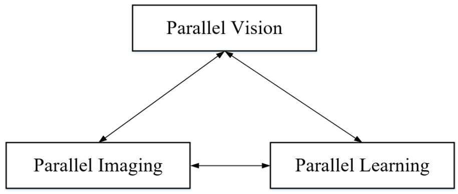

Fig. 3. Relationship between parallel vision, parallel imaging, and parallel learning.

We have designed the ParallelEye dataset (see Fig. 4). ParallelEye [27] is synthesized by referring to the real urban network of Zhongguancun Area, Beijing. We use the computer animation-like techniques to model the artificial scenes. Large-scale virtual images and their annotation information as shown in the following figure, including semantic/instance segmentation, object bounding box, object tracking, optical flow, and depth. These annotation information can be used for intelligent traffic surveillance and intelligent vehicle research.

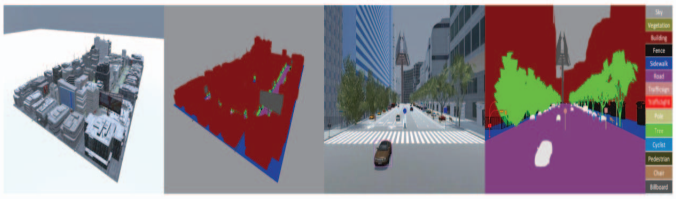

Fig. 4. Examples of our ParallelEye dataset. From left to right: a general view of the constructed artificial scenes, its semantic labels, a sample frame with tracking bounding boxes generated automatically, and its semantic labels. Best viewed with zooming.

As shown in Fig. 5, we can control the target appearance and movement, lighting and weather conditions in the artificial scene to increase the diversity of virtual images. The experimental results show that the parallel vision approach can improve the accuracy and robustness of object detection algorithms (see Table I).

We also constructed specific scenes (see Fig. 6). The virtual pedestrians are superimposed on the real scene background, and the virtual pedestrians naturally walk to generate annotation data [28]. The experimental results verified the utility of virtual data in scene-specific object detection (see Table II).

III. DEVELOPMENT AND EVOLUTION OF GANS

As an important approach to parallel imaging, GANs can generate image samples from random noises, and transform the style of original images while keeping the contents. In this section, we focus on the development and evolution of GANs.

TABLE I. EVALUATION RESULTS OF CAR CLASSES BASED ON KITTI AND PARALLELEYE DATASETS

Algorithm	Train dataset	MAP
DPM	KITTI	0.516
	KITTI&ParallelEye	0.527
Faster R-CNN	KITTI	0.741
	KITTI&ParallelEye	0.757

TABLE II. AVERAGE PRECISION OF OBJECT DETECTION IN THE EVALUATED SCENES

	TownCenter			Atrium			PETS 2009		
	syn	***voc***	***incre***	***syn***	***voc***	***incre***	***syn***	***voc***	***Incre***
Faster R-CNN	79.2%	45.2%	34%	67%	47.5%	12.5%	90.4%	89.5%	0.9%
DPM	34.2%	30.8%	3.6%	56.2%	45.4%	10.8%	85.3%	81%	4.3%

$$\min_G \max_D V(D,G) = E_{\mathbf{x}\sim p_{data}(\mathbf{x})}[\log D(\mathbf{x})] + E_{\mathbf{z}\sim p_{data}(\mathbf{z})}[\log(1-D(G(\mathbf{z})))] \quad (1)$$

$$\min_G \max_D V(D,G) = E_{\mathbf{x}\sim p_{data}(\mathbf{x})}[\log D(\mathbf{x}|\mathbf{y})] + E_{\mathbf{z}\sim p_{data}(\mathbf{z})}[\log(1-D(G(\mathbf{z}\mid\mathbf{y})))] \quad (2)$$

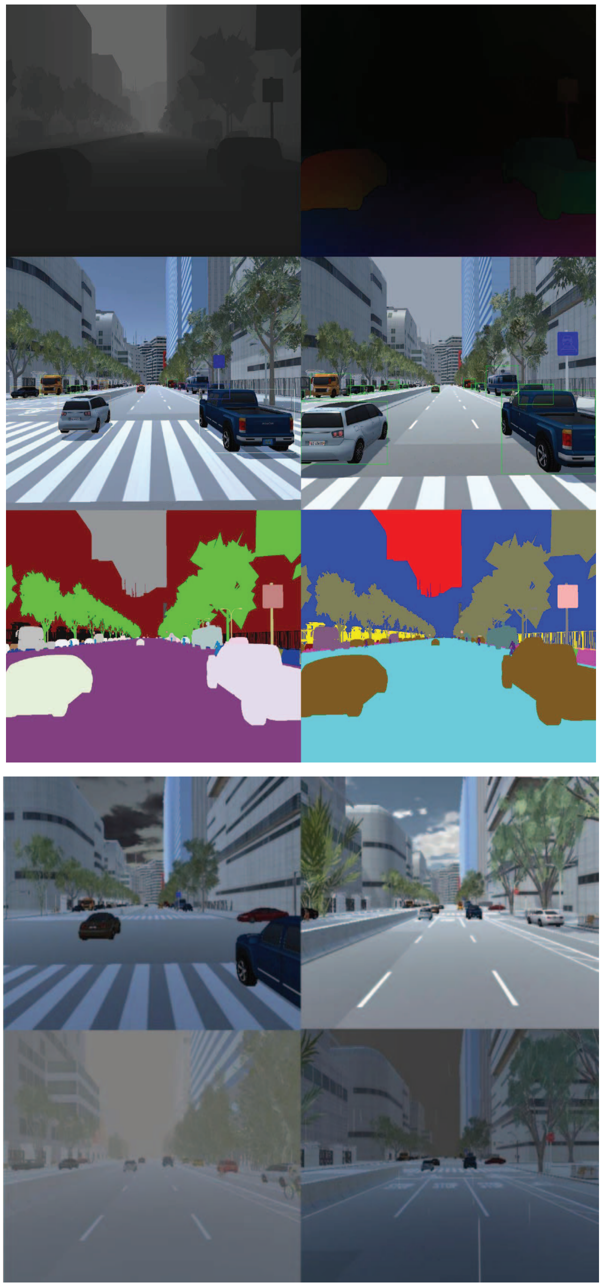

Fig. 5. Examples of ground-truth annotations generated automatically by Unity3D (top). Illustration of the diversity of artificial scenes (down).

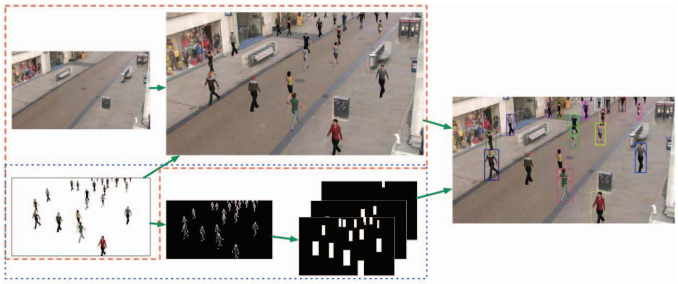

Fig. 6. The virtual pedestrians in a specific scene.

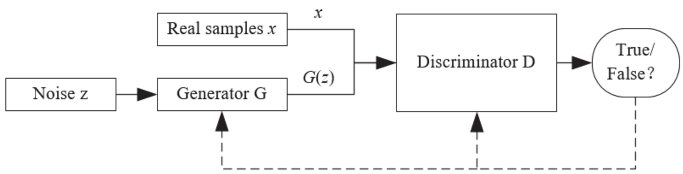

Fig. 7. The architecture of GAN.

A. *Image generation: from GAN to BEGAN*

1) *GAN [16]*

The basic principle of GAN (see Fig. 7): a generator and a discriminator make up the whole model, and the target of generator and discriminator is opposite. The target of the discriminator is to correctly distinguish the real data X and the pseudo data $G(z)$ generated by the generator G. The goal of generator G is to generate image $G(z)$ from random noise z, so that $G(z)$ can mislead the discriminator to the maximum extent. These two goals are mutually antagonistic, and therefore D and G play a two-player minimax game with value function $V(D, G)$, as shown in equation (1).

2) *CGAN [29]*

Conditional GAN (CGAN) is an extension of the original GAN, and both generator and discriminator add additional information y as a condition. y can be any information, such as the category information, or other modal data. CGAN is realized by sending additional information y to the discrimination model and the generation model as part of the input layer. In the generation model, the a prior input noise $p(z)$ and the conditional information y are combined to form the joint hidden layer representation. The antagonistic training framework is quite flexible in terms of the composition of the implicit representation. Similarly, the objective function of

CGAN is a two-player minimax game with conditional probability, as shown in equation (2).

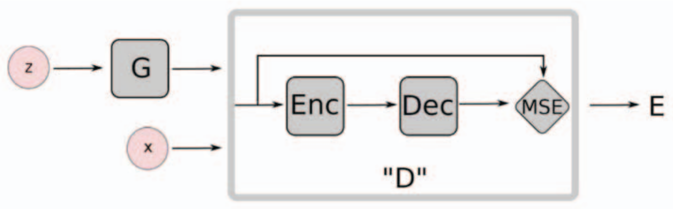

Fig. 8. EBGAN architecture with an auto-encoder discriminator.

3) EBGAN [17]

Energy-based GAN (EBGAN) uses a broader structure and loss function to design GAN. EBGAN uses a self-encoder to represent the discriminator. The discriminator is regarded from the point of view of energy . Zhao *et al.* [17] think that if the true image X corresponds to a space, the discriminator outputs a small energy, but for the pseudo-image $G(z)$, the discriminator should output a relatively large energy. The structure of EBGAN is shown in Fig. 8.

4) WGAN [18]

Arjovsky *et al.* think that if the real samples and pseudo samples are distributed with very small space overlapping, the discriminator can distinguish them easily. Using J-S divergence to measure the distance between the real distribution p_{data} and generator distribution p_g, may result in the gradient disappearance of the generator and unsteady training. WGAN does not change the model structure, but uses Earth-Mover distance to measure the distance between p_{data} and p_g, making the training much more stable. In WGAN, the loss function after each iteration is an estimate of the EM distance, or Wasserstein estimates, which is a good indicator of the training convergence, and this indicator is closely related to the visual quality of generated images.

Then, on this basis, Gulrajani *et al.* use gradient penalty method and propose improved version of WGAN, namely WGAN-GP [30]. Gradient penalty method only function in the true and false sample concentration area and its middle of the transition zone, but because it directly limit the discrimination of gradient norm near one, so the gradient control is very strong, making it easy to adjust the size to appropriate scale.

5) BEGAN [19]

Boundary Equilibrium GAN (BEGAN) presents an improved method for address the difficulty in generating various samples, enhancing the sample diversity, and balancing convergence of the discriminator and the generator. Berthelot *et al.* absorb the advantages of EBGAN and WGAN, and use simple model structure to achieve amazing results under standard training steps. BEGAN can also make very natural transition between actual images. For example, it achieves smooth transition from one face to another. The main contributions of BEGAN are as follows: 1) A simple and robust GAN structure is proposed, and the standard training process of fast and stable convergence is proposed. 2) A balance concept that balances the discriminator and the generator is presented. 3) A new method to balance image diversity and visual quality is presented. 4) An approximate method for measuring convergence is proposed. High quality face images (see Fig. 9) with 128×128 resolution can be generated with BEGAN, including a variety of gestures, facial expressions, gender, skin color, light exposure, and facial hair. Compared with previous GAN models, the visual quality of the generated images improves significantly.

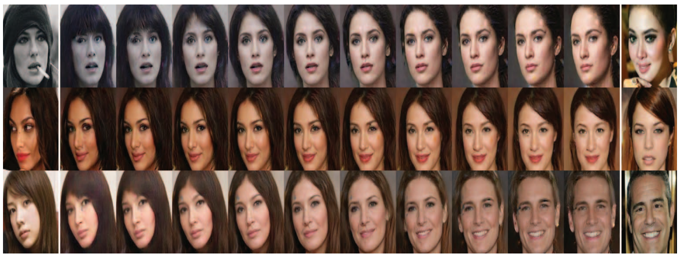

Fig. 9. Interpolations of real images in latent space (128x128 with 128 filters)

B. Image refinement: SimGAN [31]

We think SimGAN is in agreement with the parallel vision ideology. Annotating large-scale dataset is very expensive and time-consuming, with the rapid development of computer graphics, the simulated images can simulate the actual images and automatically provide labeling information for the study of computer vision models (see Fig. 10). In spite of the rapid development of computer graphics, the generated simulated images are still not realistic enough, and there is a gap between the distribution of simulated images and that of real images (called dataset shift).

However, SimGAN can improve the fidelity of the simulated images. SimGAN uses a similar idea to GAN, but the input values are simulated images, rather than random noise. SimGAN trains a refiner with the adversarial loss and improves the simulated image. The real image with unlabeled information and the improved simulated image are used as the input of the discriminator to learn the discriminator, and the purpose of discriminator is to correctly judge source data as real images or refined simulated image.

In order to keep the annotation information of the simulated image unchanged, in the adversarial loss, a regularization is added to punish the change of the original simulated image and the refined simulated image. The loss function of the Refiner model and the objective function of the discriminator form the opponent, which is reflected by the positive and negative of the symbol in the loss function. In order to increase the fidelity of the simulated image and maintain the annotation information of the original image, the following loss function is minimized, as shown in equation (3).

$$\mathcal{L}_{\mathrm{R}}(\theta) = \sum \ell_{real}(\theta; \tilde{\mathbf{x}}_i, \mathrm{y}) + \lambda \ell_{reg}(\theta; \tilde{\mathbf{x}}_i, \mathbf{x}_i) \tag{3}$$

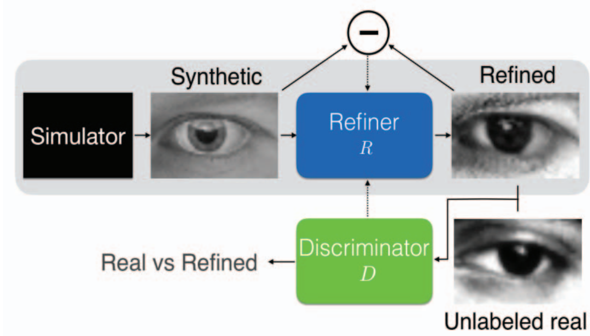

Fig. 10. Overview of SimGAN.

$\tilde{\mathbf{x}}$ is the refined simulated image, $\mathbf{x}$ is the original image, and y is the real data. The loss function of network R includes the fidelity cost and the difference cost. The former is closer to the real picture in order to make the refined image more realistic. The latter is to make the refined images not change much after the refinement.

Using the counter discrimination network D_ϕ, the discriminator parameters are learned by minimizing the following formula, as shown in equation (4):

$$\mathcal{L}_{\mathrm{D}}(\phi) = -\sum_i \log(D_\phi(\tilde{\mathbf{x}}_i)) - \sum_j \log(1 - D_\phi(\mathbf{y}_j)) \tag{4}$$

SimGAN uses two training tricks. One trick is local adversarial loss: training discriminator with the whole image may overemphasize certain image characteristics and introduce noise. Therefore, the method that train the discriminator with local image blocks is adopted to encourage each local image block to be realistic. Another trick is to use historical refined images to update the discriminator to make the training process more stable.

C. *Image translation: CycleGAN [32]*

CycleGAN is a universal, unpaired image-image translation method (see Fig. 11). Given a source domain image set and a target domain image set, CycleGAN can learn the potential relationships between the two domains and transform the source domain image into the target domain image. There's a difference between CycleGAN and several previous GANs: CycleGAN has two generators and two discriminators, which take advantage of the cyclic consistency constraint, which is a regularization constraint. The consistency against losses of CycleGAN has circulation loss function. Zhu *et al.* argue that simply adversarial loss can't guarantee that the learned generator can map the source image to the structure invariable to target image, and could lead to mode collapse. The cycle consistency constraint is used for ensuring the generator to generate images corresponding to the matched image of the target image domain, making the reconstruction and the source image different as little as possible, by limiting the source image in turn after the generator G and F, the resulting image and the difference of source image, namely the differences in the source image and reconstruction of the image. The overall objective of the loss function of CycleGAN is a minimax game problem. CycleGAN has a wide range of applications. For example, it can change the style of traffic images, but not as accurate as pix2pix (which adopts the paired training method). The Google satellite map also can be converted. CycleGAN is a universal image translation method, which can be applied to style transfer, object translation, season transfer, generating realistic images from painting, image enhancement, and so on.

However, CycleGAN has some limitations, especially it is incompetent to adapt to large geometric transformation, e.g., switching from a dog to a cat, because this change in appearance structure is really big. But in parallel vision research, we don't want the dog to be translated into a cat, we hope the image to have light, weather, and season style transfer, and improve the image fidelity and diversity. In addition, it is not a translation in terms of video, but only works according to each image frame, without taking into account the correlation between adjacent frames. These questions need further study.

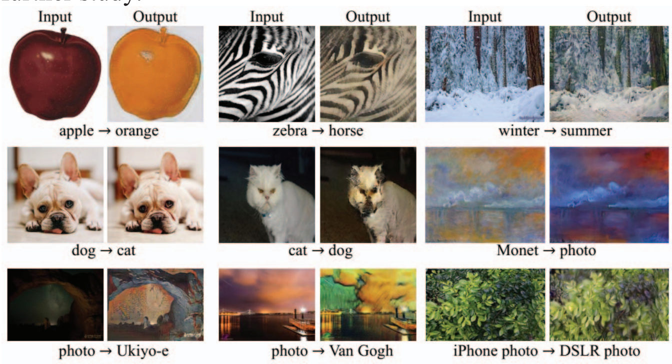

Fig. 11. Image translation effects of CycleGAN.

IV. Concluding Remarks

Fei-Yue Wang proposed the parallel system theory, which is an important achievement of modeling, analysis, and control of complex systems and can be applied to research of various complex systems. The essence of computer vision system is complex system, and it can be naturally combined with parallel system theory, leading to parallel vision.

For parallel vision, photo-realistic artificial scenes are used to model and represent the complex real scenes. Computational experiments are utilized to learn and evaluate a variety of visual models, and parallel execution is conducted to online optimize the visual system and realize smart perception and understanding of complex scenes.

On the basis of parallel vision, we proposed parallel imaging. Parallel imaging is a branch of parallel vision, and its core is to use artificial images to expand the real images. In addition, GANs is an important artificial image generation method, and has broad application prospects in image generation, image refinement, and image translation. Combining GANs and parallel vision can solve the lack of diversified image samples, thereby promoting the development of computer vision.

In the end, it should be pointed out that the Chinese Academic Journal "Acta Automatica Sinica" is calling for papers to organize a special issue "Technology and

Application of Generative Adversarial Networks". Kunfeng Wang and Fei-Yue Wang are two of the guest editors. See http://www.aas.net.cn/CN/column/item435.shtml for more details. We sincerely welcome researchers who are interested in GANs to submit their papers to the special issue.

References

[1] K. Wang, Y. Liu, C. Gou, F.-Y. Wang. A multi-view learning approach to foreground detection for traffic surveillance applications [J]. IEEE Transactions on Vehicular Technology, 2016, 65(6): 4144−4158.

[2] K. Wang and Y. Yao Video-based vehicle detection approach with data-driven adaptive neuro-fuzzy networks [J]. International Journal of Pattern Recognition and Artificial Intelligence, 2015, 29(7), [32 pages], DOI: 10.1142/S0218001415550150.

[3] C. Gou, K. Wang, Y. Yao, and Z. Li. Vehicle license plate recognition based on extremal regions and restricted Boltzmann machines [J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(4): 1096−1107.

[4] C. Gou, Y. Wu, K. Wang, K. Wang, F.-Y. Wang, and Q. Ji. A joint cascaded framework for simultaneous eye detection and eye state estimation [J]. Pattern Recognition, vol. 67, pp. 23−31, Jul. 2017.

[5] Y. Liu, K. Wang, and D. Shen. Visual tracking based on dynamic coupled conditional random field model [J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(3): 822−833.

[6] K. Wang, W. Huang, B. Tian, and D. Wen. Measuring Driving Behaviors from Live Video [J]. IEEE Intelligent Systems, 2012, 27(5): 75−80.

[7] K. Wang, Z. Li, and S. Tang. Visual Traffic Data Collection Approach Based on Multi-features Fusion [J]. Acta Automatica Sinica, 2011, 37(3): 322−330.

[8] H. Zhang, K. Wang, and F.-Y. Wang. Advances and perspectives on applications of deep learning in visual object detection [J]. Acta Automatica Sinica, 2017, 43(8): 1289−1305.

[9] K. Wang, C. Gou, and F.-Y. Wang. Parallel vision: An ACP-based approach to intelligent vision computing [J]. Acta Automatica Sinica, vol. 42, no. 10, pp. 1490-1500, 2016.

[10] K. Wang, C. Gou, N. Zheng, J. M. Rehg, and F.-Y. Wang. Parallel vision for perception and understanding of complex scenes: methods, framework, and perspectives [J]. Artificial Intelligence Review, vol. 48, no. 3, pp. 298−328, Oct. 2017.

[11] K. Wang, Y. Lu, Y. Wang, Z. Xiong, and F.-Y. Wang. Parallel Imaging: A New Theoretical Framework for Image Generation [J]. Pattern Recognition and Artificial Intelligence, 2017, 30(7): 577−587.

[12] Bainbridge W S. The scientific research potential of virtual worlds[J]. science, 2007, 317(5837): 472-476.

[13] Prendinger H, Gajananan K, Zaki A B, et al. Tokyo virtual living lab: Designing smart cities based on the 3D Internet [J]. IEEE Internet Computing, 2013, 17(6): 30-38.

[14] Ros G, Sellart L, Materzynska J, et al. The SYNTHIA dataset: A large collection of synthetic images for semantic segmentation of urban scenes[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 3234-3243.

[15] Gaidon A, Wang Q, Cabon Y, et al. Virtual worlds as proxy for multi-object tracking analysis[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 4340-4349.

[16] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//International Conference on Neural Information Processing Systems. MIT Press, 2014:2672-2680.

[17] J. Zhao, M. Mathieu, Y. LeCun. Energy-based Generative Adversarial Network [C/OL]. [2017-08-15]. https://arxiv.org/abs/1609.03126

[18] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein GAN [C/OL]. [2017-08-15]. https://arxiv.org/abs/1701.07875

[19] D. Berthelot, T. Schumm, and L. Metz. BEGAN: Boundary Equilibrium Generative Adversarial Networks [C/OL]. [2017-08-15]. https://arxiv.org/abs/1703.10717v2

[20] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, and F.-Y. Wang. Generative adversarial networks: The state of the art and beyond. Acta Automatica Sinica, 2017, 43(3): 321−332.

[21] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, and F.-Y. Wang. Generative adversarial networks: Introduction and outlook. IEEE/CAA Journal of Automatica Sinica, 2017, 4(4): 588−598.

[22] F.-Y. Wang. Parallel control and management for intelligent transportation systems: Concepts, architectures, and applications [J]. IEEE Transactions on Intelligent Transportation Systems, vol. 11, no. 3, pp. 630-638, 2010.

[23] F.-Y. Wang. Parallel control: A method for data-driven and computational control [J]. Acta Automatica Sinica, vol. 39, no. 4, pp. 293-302, 2014.

[24] F.-Y. Wang, X. Wang, L. Li, and L. Li. Steps toward parallel intelligence [J]. IEEE/CAA Journal of Automatica Sinica, vol. 3, no. 4, pp.345-348, 2016.

[25] L. Li, Y. Lin, D. Cao, N. Zheng, and F.-Y. Wang. Parallel learning — A new framework for machine learning [J]. Acta Automatica Sinica, vol. 43, no. 1, pp. 1-8, 2017.

[26] X. Liu, X. Wang, W. Zhang, J. Wang, and F.-Y. Wang. Parallel data: From big data to data intelligence [J]. Pattern Recognition and Artificial Intelligence, vol. 30, no. 8, pp. 673-682, 2017.

[27] X. Li, K. Wang, Y Tian, L Yan, and F.-Y. Wang. The ParallelEye Dataset: Constructing Large- Scale Artificial Scenes for Traffic Vision Research [C]// 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), to be published.

[28] W. Zhang, K. Wang, H. Qu, J. Zhao, and F.-Y. Wang. Scene-Specific Pedestrian Detection Based on Parallel Vision [C]// 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), to be published.

[29] Mirza M, Osindero S. Conditional Generative Adversarial Nets[J]. Computer Science, 2014: 2672-2680.

[30] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved Training of Wasserstein GANs. arXiv preprint arXiv: 1704.00028, 2017.

[31] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Josh Susskind, et al. Learning from Simulated and Unsupervised Images through Adversarial Training[C]. CVPR 2017.

[32] Zhu J Y, Park T, Isola P, et al. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. arXivpreprint arXiv:1703.10593, 2017.