2017 IEEE Intelligent Vehicles Symposium (IV)
June 11-14, 2017, Redondo Beach, CA, USA

Self-Adapting Part-Based Pedestrian Detection Using a Fish-eye Camera

Yeqiang Qian[1], Ming Yang[1], Chunxiang Wang[2] and Bing Wang[1]

***Abstract*— Nowadays, fish-eye cameras play an increasingly important role in intelligent vehicles because of its wide field of view. Using fish-eye camera, pedestrians around the vehicles could be monitored expediently, but the problem of pedestrian distortion has always existed. This paper creates a new warping pedestrian benchmark using imaging principle of the fish-eye camera based on ETH pedestrian benchmark. With this practical benchmark, warping pedestrians are trained differently according to the position in fish-eye images. A self-adapting part-based algorithm is proposed to detect pedestrian with different degrees of deformation. Moreover, GPU is used to accelerate the whole algorithm to guarantee the real-time performance. Experiments show that the algorithm has competitive accuracy.**

I. INTRODUCTION

In advanced driver assistance systems, fish-eye cameras [1] play an important role because of its wide field of view for almost 180 degrees. We could monitor the environment around vehicles [2] with only 2 cameras theoretically [3]. However, because of the nonlinear mapping, pedestrians show warping body in images, which leaves a intractable problem for the normal algorithm.

Fortunately, many papers are proposed to handle the problem. Most papers focus on image un-warping as the first step, so that they can apply the original detecting algorithm conveniently such as Histograms of Oriented Gradients (HOG) [4], the Local Binary Pattern (LBP) [5], Aggregated Channel Features (ACF) [6] and so on. In [7], reprojection is used, followed by a Soft-Cascade +ACF classifier, and pedestrians in the cylindrical image can be upright. Similarly, the cylindrical model is used and Hybrid Cascade Framework plays an important role in [14]. A little different with [7], pinhole models are utilized to correct the fish-eye image in [8] and [9], moreover, raw images produced by the front fish-eye camera are divided into 3 parts for more accurate detection in [8]. All these algorithms have shown competitive results but rely on the parameters of cameras seriously [8],[9],[10]. Distinguishing parameters set are required for different cameras, which barely guarantee the sound effect in all cases.

Making all the training samples that the detecting algorithm needed manually is another solution. In [11], authors focus on the evaluation of the deformation part model (DPM) proposed by Felzenszwalb et al. [12], whose dataset is captured from one fish-eye camera and labeled aboratively. However, the dataset can not contain all the scenarios compared with standard benchmark such as ETH and KITTI. What is more, parameters are constant for all pedestrians with different degrees of deformation, which is unreasonable. Nothing is done with the raw image in [13], and different methods are adopted according to the position of the original image, which has a good result on vehicle detection but pedestrians have larger deformation than vehicles in fish-eye images.

Inspired by all the methods mentioned before, this paper proposes a new warping pedestrian benchmark based on ETH pedestrian benchmark. Deformation is quantified in it, which is important for self-adapting detection algorithm. What is more, this benchmark could contribute to other methods. Thanks to the work for P.F. Felzenszwalb et al. [12], deformable part model (DPM) and DPM-based methods could be acquired conveniently. Part-based algorithm is an effective solution for pedestrian detection. However, templates DPM uses are constant and can not handle all the situation. Therefore, deformation is added in the self-adapting algorithm with the help of new benchmark. Besides, GPU is used to solve the speed bottleneck of DPM.

This paper is organized as follows: a new warping pedestrian benchmark is introduced in Section II; the self-adapting part-based detection algorithm is detailed in Section III; experimental results are demonstrated in Section IV, followed by a conclusion in Section V.

II. WARPING PEDESTRIAN BENCHMARK

A. The Imaging Principle of Fish-eye Camera

Spherical perspective imaging model is usually used in simulating the imaging principle of fish-eye camera. As it is shown in the Fig.1 [14], the surface of fish-eye camera is regarded as sphere. The light of outside world refracts into the photosensitive components of the camera, which forms a fish-eye image. Similarly, Light could map into the image directly through the pinhole models, which form the usual image as equation (1).

$$\rho = f \tan\theta \tag{1}$$

Normally, fish-eye camera lenses are designed according to one of the following 4 projections:

Equidistance projection:

$$\rho = f\theta \tag{2}$$

Stereographic projection:

$$\rho = 2f \tan(\theta/2) \tag{3}$$

This work was supported by the National Natural Science Foundation of China (91420101) and International Chair on automated driving of ground.

[1]Yeqiang Qian, Ming Yang and Bing Wang are with the Department of Automation, Shanghai Jiao Tong University, Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai, 200240, CN (phone: +86-21-34204533; email: MingYang@sjtu.edu.cn).

[2]Chunxiang Wang is with Research Institute of Robotics, Shanghai Jiao Tong University, Shanghai 200240, China.

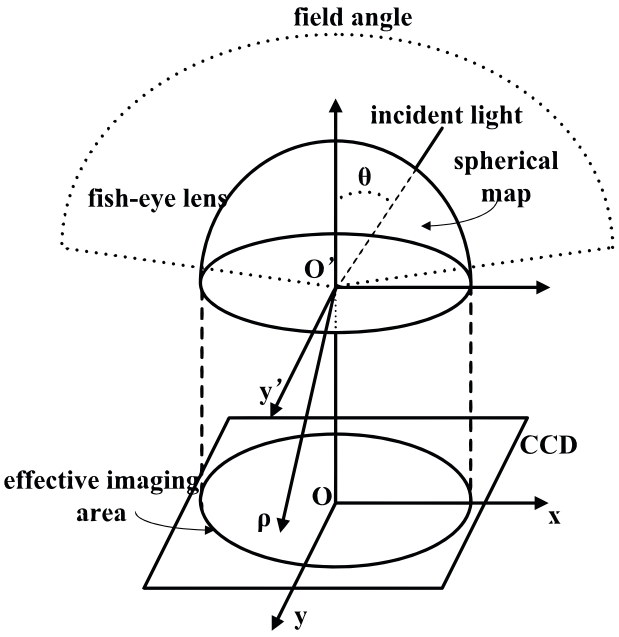

Fig. 1. Imaging principle diagram of fisheye camera. O is the center of imaging area, ρ is the projected point radial coordinate, f is the focal length and θ is the angle between the lens principal axis and the incoming light ray

Equisolid angle projection:

$$\rho = 2f\sin(\theta/2) \tag{4}$$

Orthogonal projection:

$$\rho = f\sin\theta \tag{5}$$

Here, ρ is the projected point radial coordinate, f is the focal length and θ is the angle between the lens principal axis and the incoming light ray. We assume that the main point of flat image is exactly the center of fish-eye image, ignoring the little error of the camera.

The lens distortion is relatively small when using projection (3) and projection (4), but the models are complex and information is difficult to handle, which have not been widely used. When using orthogonal projection, even in a small area, lens distortion is more obvious than any other models, which will lose almost all the information in nearly 180 degrees. Without loss of generality, the equidistance projection (2) is considered in this paper because it takes the advantage of high precision and is widely used in the market.

B. A New Warping Pedestrian Benchmark

The process of benchmark fabrication is shown in the Fig.2.

Firstly, flat image is transformed to the sphere by the equation (2), every point K in the sphere correspond to a single point $Q(x, y)$ in the raw image. Then K is remapped into imaging plane as $P(u, v)$. So, the relevance between $P(u, v)$ and $Q(x, y)$ can be shown using coordinate transformation as in the equation (6). For convenient, radius of

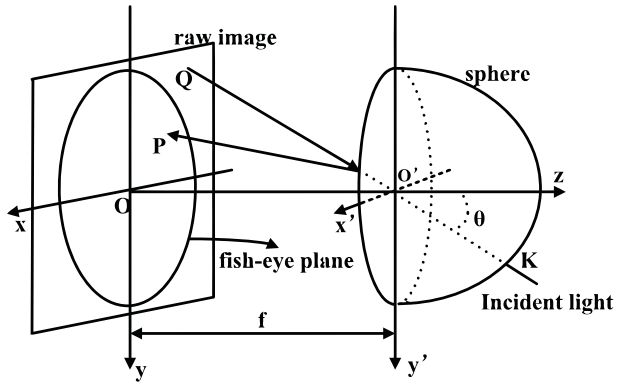

Fig. 2. The changes of the light path. Every point K in the sphere correspond to a single point $Q(x, y)$ in the raw image. Then K is remapped into imaging plane as $P(u, v)$

fish-eye imaging plane is set 250 pixels constantly in this paper, which corresponds to 90 degrees in the sphere.

$$x^2 + y^2 = f^2\tan^2(\sqrt{u^2 + v^2}/f) \tag{6}$$

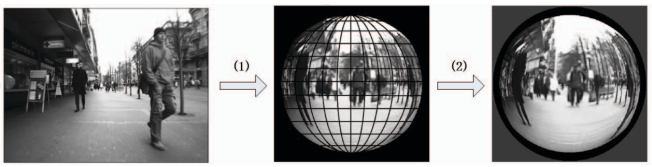

Fig. 3. The process of image manufacture. The flat image is transformed to the sphere, then remapped into fish-eye imaging plane

In order to cover all kinds of weather, backgrounds, pedestrian postures and to guarantee sufficient amount of data, the new warping pedestrian benchmark is designed utilizing the ETH benchmark [14] instead of making a new image collection. Besides, pedestrians in the ETH benchmark have more pixels than any others and it has lots of street scenes. The ETH benchmark contains 8 subsets and 5361 images in total. Fig.3 shows the process of image manufacture.

Fig.4 shows some examples of fish-eye images and raw images. It is worth noting that, pedestrian has different degrees of deformation, which is quantized in this paper. Specially, the largest deformation occurs in the edge of image, where pedestrian become a circular arc and the radius is the same as fish-eye image. Oppositely, the pedestrian who stands in the middle of fish-eye image keeps upright, and radius could assume infinity. The radius is recorded as an index to describe the deformation degree of pedestrian according to the position of fish-eye image, which plays an important role in the following self-adapting part-based detection algorithm.

On the whole, every image in the new fish-eye benchmark includes pedestrian with mark of position and radius. It is worth mentioning that, the radius of pedestrian is continuous and it is the basic of self-adapting.

Fig. 4. Some examples of fish-eye images and raw images. The new benchmark based on ETH contains 8 subsets and 5361 images in total, which covers all kinds of weather, backgrounds and pedestrian postures

III. SELF-ADAPTING PART-BASED DETECTION ALGORITHM

A. Part-based Theory

Since pedestrian detection is a challenging problem because of its complexity, part-based theory solves the problem perfectly and Deformable Parts Models (DPM) [12] is the typical representation. DPM and its variants have gained a lot of attention in object detection and recognition and it is the winner of some recent Pascal-VOC detection challenges [15]. The DPM can represent an object model with different parts floating around their reference locations and find the optimal part-configuration at every root position, which brings lots of benefits in detecting a person in different postures. However, the DPM has not shown its value when detecting pedestrian in a fish-eye image, where pedestrian has larger deformation. Moreover, the template designed should adapt the actual circumstances of the deformation instead of using the constant templates. This paper proposes a self-adapting part-based detection algorithm based on DPM.

B. Self-adapting DPM

The score of DPM at position $X = (x_0, y_0)$ is calculated by

$$score(P, X) = F_0 \cdot \phi_0(X) + \Sigma \max_{\Delta X} \{F_i \cdot \phi_i(X + v_i + \Delta X) - d_i \cdot \theta_d(\Delta X)\} + b \quad (7)$$

Where F_0 is a root filter representing the tough shape of an object, and $F_i(i = 1, ...n)$ are part filters representing the local structures of the object. A part filter is composed of a part position v_i relative to the root filter and a deformation cost d_i. In equation (7), the first term is the response of the root filter, and the next one is the sum of the response of part filters subtracted by the deformation cost.

The deformation cost d_i is a weight vector for $\theta_d(\Delta X)$, which can be calculated by the following equation.

$$\theta_d = \{\Delta x, \Delta y, (\Delta x)^2, (\Delta y)^2\}^T \quad (8)$$

Here, three main modifications are proposed.

- The response of the root-filter F_0 is lowered but high response of part-filters F_i are reserved. Besides, the response of root-filter F_0 is according to the radius of pedestrian, which is mentioned in the section I. In other words, the pedestrian whose radius is infinite will keep the original response of F_0.
- The score of part-filters $\varphi_i(i = 1, ..n)$ is continuously changed according to the radius of pedestrian too, because the distortion in fish-eye images is particularly strong at the image boundaries.
- The radius of pedestrian r can be acquired when making the new benchmark. For convenience, $R = e^{125-r}$ is set to guarantee that the maximum of variable R is 1.

Then, the equation (7) is changed to

$$newscore(P, X) = \varphi_0(R) \cdot F_0 \cdot \phi_0(X) + \Sigma \max_{\Delta X} \{F_i \cdot \phi_i(X + v_i + \Delta X + \varphi_i(R)) - d_i \cdot \theta_d(\Delta X)\} + b \quad (9)$$

Where φ_0 is the coefficient of the response of the root-filter, and $\varphi_i(i = 1, ...n)$ are the variables compared with the standard template. $\varphi_i(i = 0, ...n)$ can be acquired using following equation.

$$\varphi_i = f(defs_k.anchor, defs_k.w) \quad i = 0, ...n; k = 0, 1, 2 \quad (10)$$

Here, def is the anchor point set of part models, $anchor$ is the anchor point coordinate of part models and w is deformation parameters of part models.

Pedestrians are divided into left and right according to positions. It is worth noting that 3 independent templates in left side are trained using the new benchmark through Latent SVM (LSVM) according to

$$f_\beta(x) = \max_{z \in Z(x)} \beta \cdot \phi(x, z) \quad (11)$$

Fig.5 shows templates trained by the new benchmark, $\varphi_i(R)(i = 1, ...n)$ can be fitted through $\phi_i(X)(i = 1, ...n)$ in 3 templates, as well as the right half. Since the variable R changes continuously, we call it self-adapting algorithm.

IV. EXPERIMENT RESULT AND ANALYSIS

Here is some pedestrian detection results detected by our self-adapting part-based algorithm, which is trained using the new benchmark. Then 3 main experiments are conducted to prove the reliability of the new benchmark and the superiority of our algorithm.

As it is shown in the Fig.6, our algorithm performs well in most of scenarios, but it has shortage in detecting pedestrians who stand nearby or far away from the shot. These scales of pedestrians change largely, so image pyramid used in our method can not handle them, which will be solved in the future work.

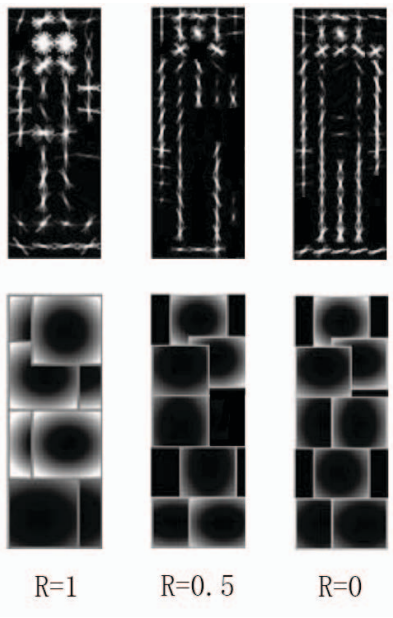

Fig. 5. 3 templates trained by the new benchmark

Fig. 6. Pedestrian detection results detected by our self-adapting part-based algorithm. (a) The new benchmark; (b) The images in life; (c) Error detection and leakage detection, the red bounding boxed are error detection and white ones are leakage detection

A. The accuracy of the new benchmark

Firstly, a large number of real scene pictures are taken, which contain pedestrians in different locations. The same scene is taken by a regular camera and a fish-eye camera. To compare with the real fish-eye images, the regular images are remade according to the second section. Moreover, enough images are taken for all positions and surroundings to avoid contingency. Fig.7 shows some samples of different positions and surroundings.

In the end, experiments show that images of production have almost the same sound effect as the real ones. Moreover,

Fig. 7. Comparison of real fish-eye images and ones made by our methods (a) Real scene pictures; (b) The new benchmark; (c) Real fish-eye images taken at the same location

the new benchmark contains more information in the edge, which benefits to detect pedestrians in the larger range.

B. The effectiveness of the new benchmark

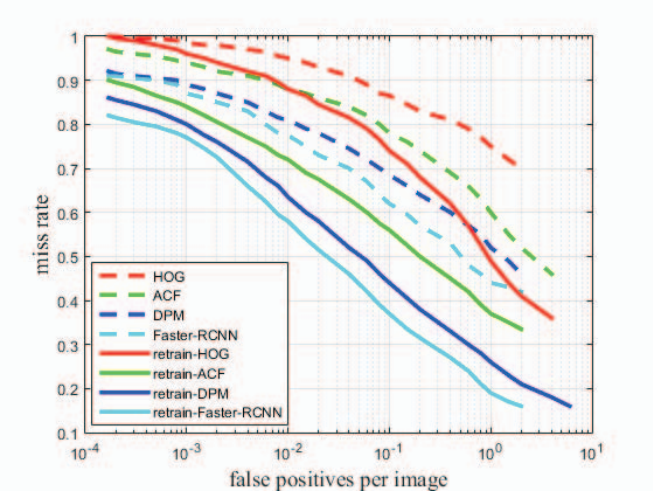

Fig. 8. ROC curves of several state-of-the-art detectors and their retrained results

As is said in the introduction, the normal algorithms have uncompetitive result on the fish-eye images. Here, 4 main target detection algorithms including HOG [4], ACF [6], DPM [12] and Faster-RCNN [17] are retrained according to the new benchmark and ETH benchmark.

4000 images is set as training set and the rest as testing set. The detailed usage of training set is shown in the Tab.1 and result is shown in the Tab.2. The ZF model (Zeiler and Fergus model) [18] is used when training Faster-RCNN, which limited to our hardware facility. Since these algorithm put all the pedestrians as single template, the variable R does not work.

As it is shown in the Fig.8, after applying the new benchmark, performance of all algorithms have been promoted,

TABLE I

THE USAGE OF SAMPLES

Positive Samples	R=1	R=0.5	R=0	total
Numbers	4527	3807	4231	12565
Negative Samples	total			
Numbers	8000			

especially the raw DPM and Faster-RCNN. However, low detection accuracy is achieved even the retrained HOG. In general, DPM and Faster-RCNN have better ability to comprehend the deformation of pedestrians, which benefit from the new benchmark.

In the view of the fact that there is no available standard database for fish-eye images, the new benchmark will play a greater role in the future.

C. The performance of our algorithm

In the end, using the new benchmark equally, the self-adapting part-based algorithm is compared with the 4 methods mentioned before. ROC curve is generated and mean average precision (mAP) is calculated to show the result. The usage of samples is the same as the second experiment.

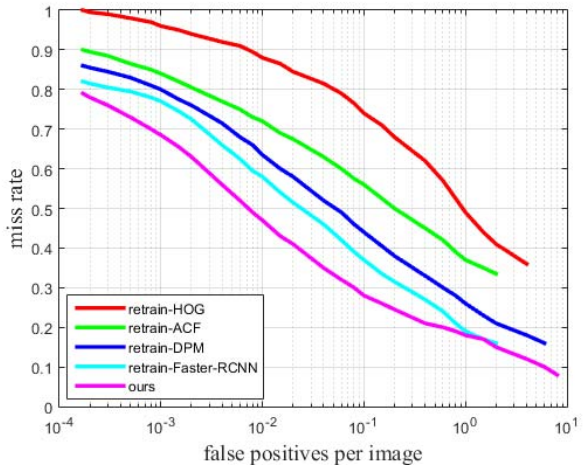

Fig. 9. ROC curves of several state-of-the-art detectors and our approach

TABLE II

THE PERFORMANCE OF ALL ALGORITHMS

Algorithms(retrained)	HOG	ACF	DPM	Faster-RCNN	ours
mAP(%)	42.6	53.4	56.8	61.8	**70.5**
Time per frame(ms)	80	**68**	93	105	103

Obviously, our self-adapting part-based algorithm acquire the highest accuracy and ACF has the fastest detection speed. In addition, retrained Faster-RCNN also have sound effect, and it is believed that a better performance will be achieved when applying a deeper level of neural network.

Thanks to previous work on accelerating DPM using GPU, the detection speed of our self-adapting algorithm could reach to 10HZ, which satisfies most application scenarios (our cpu is Intel(R) Core(TM) i7-4790 @ 3.6GHz and our GPU is Nvidia gtx 960).

Of course, more efforts should be taken to optimize the algorithm, because even highest mAP is 70.5%, which has not meet the security requirement for unmanned vehicles.

V. CONCLUSIONS

In this paper, we propose a new warping pedestrian benchmark based on ETH pedestrian benchmark. Besides the position mark of pedestrian, the new benchmark contains radius to estimate the deformation quantity of pedestrian, which is continuous from image boundary to the center. The new benchmark can be used to design or test other algorithm in the future. Furthermore, a self-adapting algorithm based on DPM is proposed to detect pedestrian in the fish-eye image. With the assist of new benchmark, experiments show the superiority of this algorithm compared with original DPM or other methods.

Future work includes optimization of detection for the pedestrians who stand nearby or far away from the shot. Besides, estimation of security for the surrounding pedestrians is also in considering.

REFERENCES

[1] Zhang B, Appia V, Pekkucuksen I, et al. A surround view camera solution for embedded systems[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2014: 662-667.

[2] Castangia L, Grisleri P, Medici P, et al. A coarse-to-fine vehicle detector running in real-time[C]//17th International IEEE Conference on Intelligent Transportation Systems (ITSC). IEEE, 2014: 691-696.

[3] Gressmann M, Palm G, Lhlein O. Surround view pedestrian detection using heterogeneous classifier cascades[C]//2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC). IEEE, 2011: 1317-1324.

[4] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE, 2005, 1: 886-893.

[5] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on pattern analysis and machine intelligence, 2002, 24(7): 971-987.

[6] Dollr P, Appel R, Belongie S, et al. Fast feature pyramids for object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(8): 1532-1545.

[7] Bertozzi M, Castangia L, Cattani S, et al. 360 Detection and tracking algorithm of both pedestrian and vehicle using fisheye images[C]//2015 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2015: 132-137.

[8] Broggi A, Cardarelli E, Cattani S, et al. Vehicle detection for autonomous parking using a soft-cascade AdaBoost classifier[C]//2014 IEEE Intelligent Vehicles Symposium Proceedings. IEEE, 2014: 912-917.

[9] Silberstein S, Levi D, Kogan V, et al. Vision-based pedestrian detection for rear-view cameras[C]//2014 IEEE Intelligent Vehicles Symposium Proceedings. IEEE, 2014: 853-860.

[10] Kubo Y, Kitaguchi T, Yamaguchi J. Human tracking using fisheye images[C]//SICE, 2007 Annual Conference. IEEE, 2007: 2013-2017.

[11] Bui M T, Frmont V, Boukerroui D, et al. Deformable parts model for people detection in heavy machines applications[C]//Control Automation Robotics and Vision (ICARCV), 2014 13th International Conference on. IEEE, 2014: 389-394.

[12] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]//Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008: 1-8.

[13] Dooley D, McGinley B, Hughes C, et al. A Blind-Zone Detection Method Using a Rear-Mounted Fisheye Camera With Combination of Vehicle Detection Methods[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(1): 264-278.
[14] Schulz W, Enzweiler M, Ehlgen T. Pedestrian recognition from a moving catadioptric camera[C]//Joint Pattern Recognition Symposium. Springer Berlin Heidelberg, 2007: 456-465.
[15] Everingham M, Van Gool L, Williams C K I, et al. The pascal visual object classes (voc) challenge[J]. International journal of computer vision, 2010, 88(2): 303-338.
[16] Jun Zhang, Zhizhou Wang, Zhengling Wang, et al. Correction of single circular fisheye image[J]. Journal of Computer Applications, 2015, 35(5): 1444-1448.
[17] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91-99.
[18] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[C]//European Conference on Computer Vision. Springer International Publishing, 2014: 818-833.