A Two-dimensional Gaussian Distribution and Rotation Equivariant for Rotationally Dense Obejct Detection

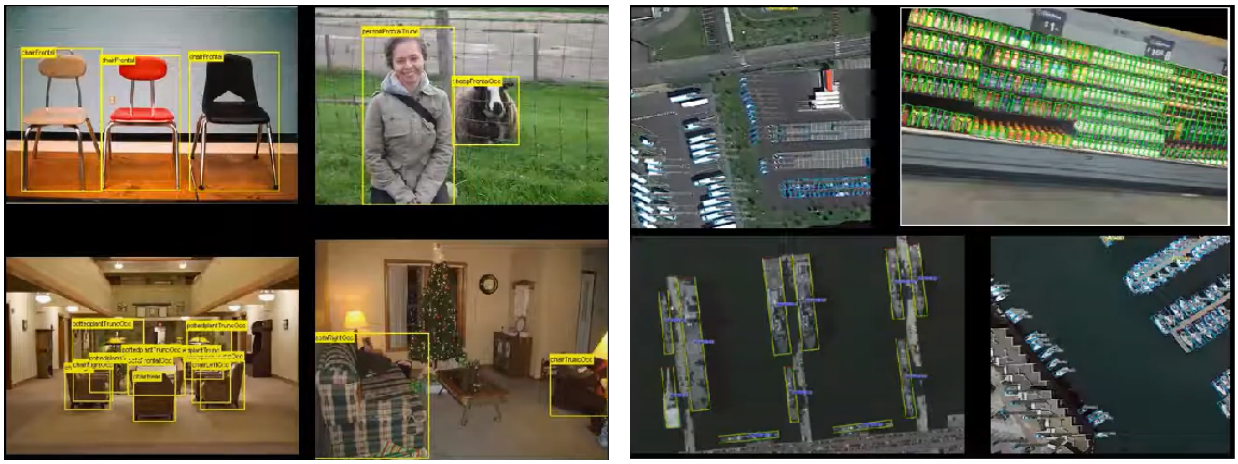
Abstract

In the past few years, object detection has developed rapidly, and common object detection has achieved significant breakthroughs. However, there are still many challenges for detecting rotating dense objects such as aerial images and dense goods in supermarkets. The objects to be detected by the detector are not always horizontal or vertical, and the objects to be detected usually have different orientations. When the object rotates, the heat map of the feature of the current common deep neural network will also change with the rotation of the object, and the heat map of concentric circles is used to represent it. In order to solve these two problems, this paper uses a rotation equivariant network to solve the problem that Angle regression and the current deep network do not have rotation equivariant, and the modified two-dimensional Gaussian distribution makes the heat map change from concentric circles to oblique ellipals to better fit the detection target. In this paper, we conduct experiments on the challenging public dataset HRSC 2016. The experimental results show that compared with the DR-Net network before improvement, the mAP is improved by 1.4.

Keywords: Object Detection, rotation equivariant, Gaussian Distribution.

1 Introduction

Object detection has achieved remarkable achievements on some routine tasks. For example, two-stage algorithms are: RCNN [7], Fast RCNN [6], Faster RCNN [21], FPN [15], Mask RCNN [10]. The one-stage algorithms are: YOLO [22], SSD [17], Retinanet [16]. Transformer-based algorithms are Relation Net [13] and DETR [1]. However, when these object detection algorithms are applied to the task of rotating dense object detection, such as aerial images and supermarket product images, which are densely distributed in any direction, there will be problems, resulting in that the size and shape of the detected object cannot be accurately identified, or even the object cannot be well fitted. The difference between traditional object detection and rotating object detection tasks is shown in Figure 1. Moreover, almost all detectors optimize the model parameters on the training set and keep them fixed afterwards. This static paradigm uses general knowledge and may not be flexible enough to detect specific samples during testing.



(a) Traditional object detection (b) Rotating Dense Object detection

Figure 1. Comparison of traditional object detection(a) and rotating dense object detection(b) tasks.The figure on the left shows a traditional object detection task, where the detected objects are usually large, sparse objects with horizontal or vertical bounding boxes. On the right of the figure is the rotating dense object detection task. Most of the detected objects are small objects in any direction and densely stacked.

The paper referred to in this paper, DRNet [20], specifically explores the following questions. Current detectors optimize parameters during training and keep fixed parameters unchanged after training; such static data is not flexible enough in the testing phase and may not detect specific samples. Most of the current object detection is based on the RCNN network, which first generates a large number of horizontal bounding boxes as RoIs, and then performs the prediction of position regression and classification based on rois. This method of using horizontal RoIs leads to misalignment between bounding boxes and oriented targets at specific targets. For example, an aerial image of a target will produce multiple instances covered by a single instance. The directional bounding box is used as anchor to deal with the rotating target, because a large number of anchors have different angles, different sizes, and different aspect ratios, which leads to a large amount of calculation. Dynamic filters are a simple yet effective way to make the model vary from sample to sample.

In summary, in the DRNet paper, it is mainly discussed that the current rotating dense object detection still has the following problems:

- The receptive fields of neurons are all axially aligned with the same shape, while the targets are usually of different shapes and arranged along different directions.
- The detection model is trained with general knowledge and may not generalize well to deal with specific objects when in the testing phase.
- Limited datasets hinder the development of such tasks.

To solve the above three problems, DRNet proposes three corresponding solutions:

- To solve the first and second problems, the authors propose a dynamic refinement network, including FSM (feature selection module) and DRH (Dynamic Optimization head). FSM is able to adjust the receptive fields of neurons according to the shape and orientation position of the target object. DRH enables modules to dynamically optimize predictions in a target-aware manner.
- To solve the third problem, we extend the dataset with complete annotations: SKU110K-R, which is based on SKU110K [8] directional bounding box recalibution dataset.
- The quantitative evaluation is carried out on multiple public baselines of DOTA, HRSC 2016, SKU110K and SKU110K-R datasets.

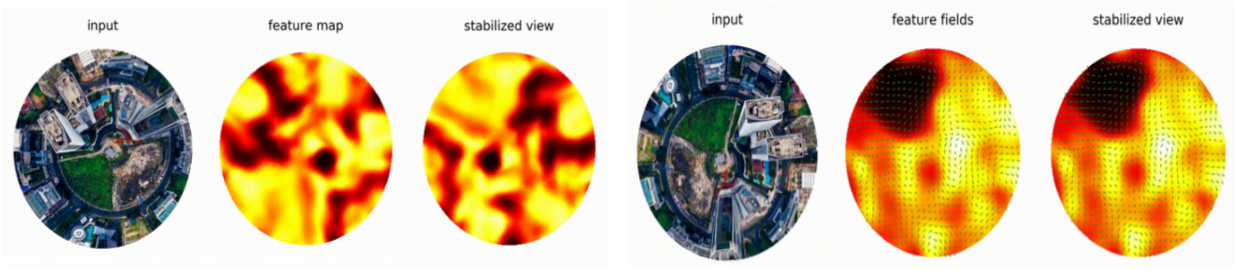
Although DRNet uses a module that adaptively adjusts the neuron receptive field based on the target shape and direction to effectively alleviate the imbalance between the receptive field and the target, and DRHs is used to model according to the uniqueness and particularity of each sample and perform detection in an object-oriented method. The object detection box can better fit the rotated object and has strong generalization, but it still fails to solve the problem that Angle regression does not have rotation homology with the current deep network. That is, when the detected object is rotated, the rotated feature map is different from the rotated feature map of the original image. If we visualize it in the heat map, we can see that the rotated heat map is not a rotation of the original heat map of the image, but a distortion.And the Gaussian heat map used in DRNet is obtained by using the target center point as the center of the circle and then calculating the radius of the Gaussian circle according to the target box to fill the calculated value of the Gaussian function. The Gaussian circle in the Gaussian heat map obtained in this way is a concentric circle with the center point as the center, which cannot fit those slender targets well. Therefore, starting from DRNet, this paper modifies the network model by adding rotation equivariant network and expanding two-dimensional Gaussian distribution, and proposes the following improvements:

- In this paper, we re-implement all layers of the rotation equivariant network based on e2cnn [23], including convolution, pooling, normalization, and nonlinearity, and re-implement the fully convolutional encoder-decoder network.
- In the Gaussian heat map, this paper uses a two-dimensional Gaussian distribution and modifies the covariance matrix in it to an off-diagonal matrix, so that the Gaussian circle in the heat map becomes an ellipse with an oblique Angle to better fit the detection target.

2 Related works

2.1 Rotating Dense Object detection

Most popular object detection methods focus on axis-aligned or vertical objects, and may encounter problems when objects have arbitrary orientations or present dense distributions. For rotating dense object detection, in order to detect objects in arbitrary directions, some methods [12] employ many rotating anchor boxes with different angles, proportions, and aspect ratios for better regression, while



(a) regular CNN (b) rotational equivariation

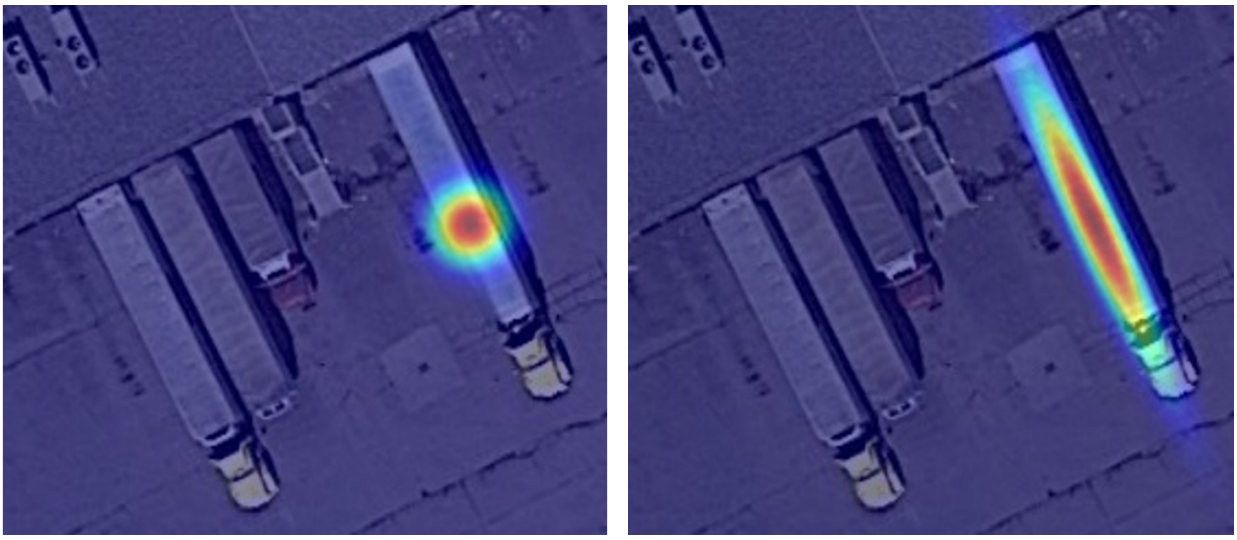
Figure 2. Comparison of a regular CNN (left image) and one with rotational equivariation (right image).

increasing the computational complexity. DRNet detects oriented objects by dynamic feature selection and refinement. Recently, some CenterNet-based methods have shown their advantages in detecting small objects. The above methods focus on using FSM to adaptively adjust the receptive fields of neurons to reassemble appropriate features for various objects with different angles, shapes, and scales and introduce dynamic filters for DRH to improve detection results in a content-aware manner. In this paper, by using the rotation equivariant network, the e2cnn layer is introduced to reconstruct the rotation equivariant feature extraction network, which greatly reduces the complexity of direction change modeling and makes the network with rotation equivariant. In addition, a two-dimensional Gaussian distribution with an off-diagonal matrix of covariance matrix is used to regenerate the heat map, so that it can better fit the detection target.

2.2 Rotational equivariant networks

Group convolution [3] was first proposed by Cohen et al. by adding rotation equivariant of 4x fold to CNNs. Hexaconv [11] extends the group convolutions by a factor of 6 on a hexagonal lattice. To achieve rotational equivariation in more directions, some methods [29] filter the filter by resampling through the difference, while others [24] use harmonics as filters to produce equivariant features in the continuous domain. These methods gradually extend rotation equivariant to larger groups and achieve good results on classification tasks [9]. The method in this paper adds rotation equivariant network to the detector, reimplements the fully convolutional encoder-decoder network, and achieves great improvement in rotation dense detection tasks. Rotation invariant features are very important for detecting objects in arbitrary directions. However, CNNS show poor performance in modeling rotation variations, which means that more parameters are needed to encode orientation information. STN [14] and DCN [4] model rotation directly in the network and have been widely used for object-oriented detection. Cheng et al. [2] proposed a rotation invariant layer that imposes an explicit regularization constraint on the objective function. Although the above methods can achieve an approximate effect of rotation invariance in Imagelevel, they require a large number of training samples and parameters. In addition, object detection requires instance-level rotation invariant features. Therefore, some methods [19] extend Rol warping to RRol warping, for example, Rol Transformer [5] learns how to transform HRoIs into RRols, and then utilizes a rotation position sensitive Rol Align operation to warp the region features.

However, as shown in Figure 2, regular CNNS do not have rotation equivariability, so even with



(a) DRN (b) Improved method

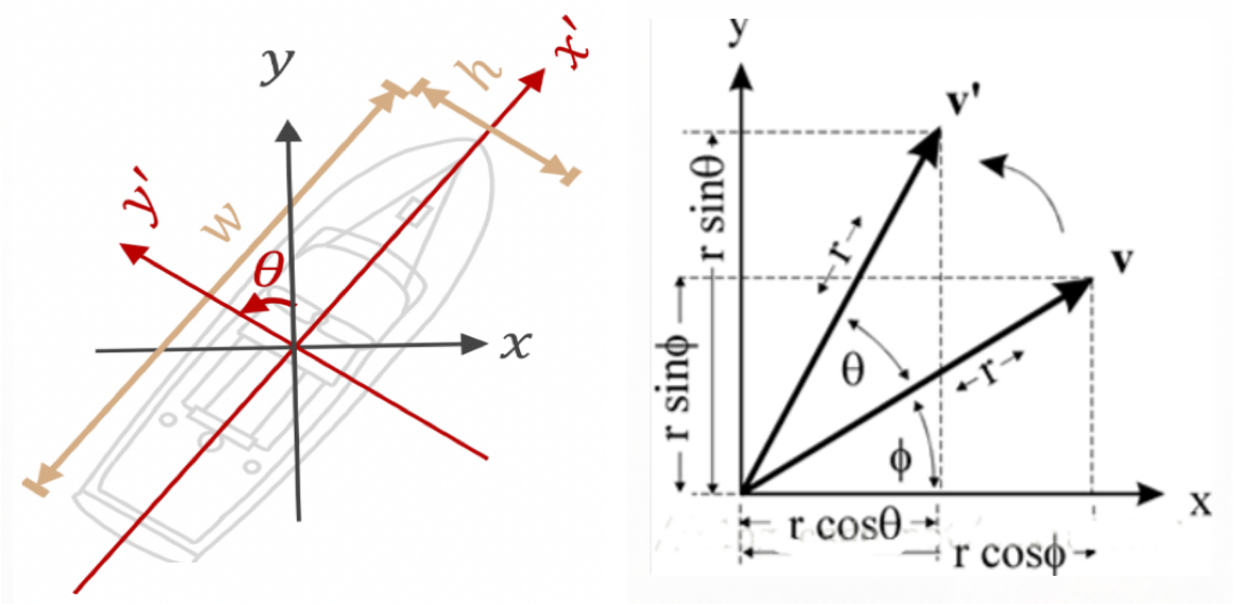
Figure 3. Comparison of heat maps. Heat map of the original DRN(left image), Improved heat map(right image)

RRO alignment, we still cannot extract rotation invariant features. CNN features are not equivariant with respect to rotated image. That is, a rotated image fed into the CNN is not the same as the rotated feature map of the original image. In the heat map, after the object in the left picture is rotated, it can be seen that as the input rotates, the stabilized view which is used to observe the features of the same part with the fixed feature map is distorted. The stabilized view on the right is quite stable after the rotation equivariant property. Different from the above methods, our method uses rotation-invariant Rol Align (RiRol Align) to extract rotation invariant features from rotation equivariant features. Specifically, a rotation equivariant network is added to the backbone network to generate rotation equivariant features, and then RiRol Align is used to completely extract rotation invariant features from rotation equivariant features in both spatial and directional dimensions.

2.3 Two-dimensional Gaussian distribution

In general object detection, such as CenterNet [28], the object is detected as a point, that is, the center point of the object box is used to represent the object. The offset of the target center and the width and height size are predicted to get the actual box of the object, while the heatmap represents the classification information. There is a heatmap for each category. On each heatmap, if there is a center point of the object target at a certain coordinate, a keypoint(represented by a Gaussian circle) is generated at that coordinate. The Gaussian circle is usually a concentric circle with radius R. In this paper, the covariance matrix of the off-diagonal matrix is used to replace the original covariance matrix, so that the generated heat map becomes an ellipse with an oblique Angle to better fit the direction and size of the target.

As can be seen from Figure 3, the thermal map of the original DRN (as shown in the left figure) is Gaussian positive circle, which is unable to fit the rotating elongated detection target well. After improving the method in this paper, as shown in the right figure, the generated thermal map can fit the



(a) (b)

Figure 4. Equation diagram of rotational isovariant network. Five-parameter definition method for rotating object detection(a). The coordinate change of a point in polar coordinates before and after rotation(b).

direction and size of the detection target well, and the detected object can be basically covered by the heat map.

3 Method

3.1 Preliminaries

The rotating target detection uses the five-parameter definition+6+ method to define the rotating box, which has more information about the rotation Angle compared with the horizontal box. The bounding box parameters can be determined by the 5d vector, as shown in Figure 4(a). Where x and y are the coordinates of the center point of the object, w and h are the real length and width of the object, respectively, and θ is the Angle between the width and the X-axis or the Angle between the height and the Y-axis of the object. We can convert the five arguments to the four vertices of the target bounding box using the following formula.

$$\begin{aligned}
A_x &= x + (\frac{h}{2} \times sin(\theta) - \frac{w}{2} \times cos(\theta)) \quad A_y = y + (\frac{h}{2} \times cos(\theta) + \frac{w}{2} \times sin(\theta)) \\
B_x &= x + (\frac{h}{2} \times sin(\theta) + \frac{w}{2} \times cos(\theta)) \quad B_y = y + (\frac{h}{2} \times cos(\theta) - \frac{w}{2} \times sin(\theta)) \\
C_x &= x - (\frac{h}{2} \times sin(\theta) - \frac{w}{2} \times cos(\theta)) \quad C_y = y - (\frac{h}{2} \times cos(\theta) + \frac{w}{2} \times sin(\theta)) \\
D_x &= x - (\frac{h}{2} \times sin(\theta) + \frac{w}{2} \times cos(\theta)) \quad D_y = y - (\frac{h}{2} \times cos(\theta) - \frac{w}{2} \times sin(\theta))
\end{aligned} \tag{1}$$

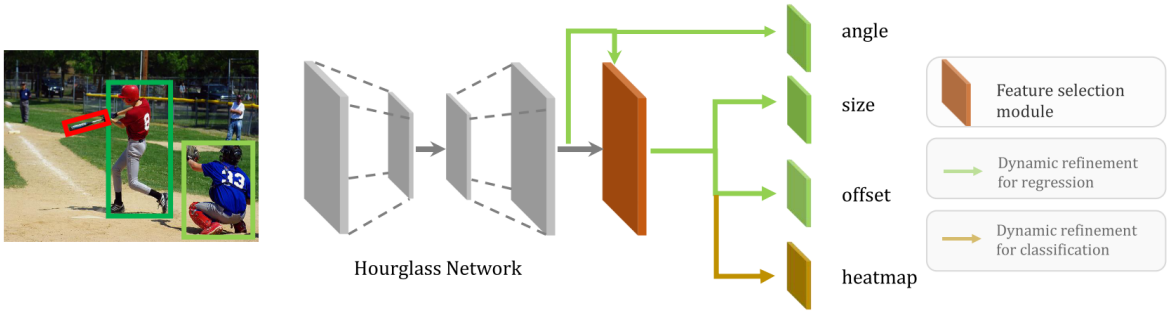


Figure 5. Overall network architecture. In this paper, the overall Network architecture of DRNet is used, where the backbone is Hourglass Network, and the backbone network is followed by two modules, Feature Selection Module (FSM) and Dynamic Refinement Head (DRHs). FSM selects the most appropriate features by adaptively adjusting the receptive field. DRH dynamically refines predictions in an object-aware manner.

3.2 Network Architecture

The overall network architecture is shown in Figure 5. We use DRNet as a baseline, which models the object as a single point (the center of the bounding box) and regress the object size, offset, and angle of the bounding box. The loss for the overall training is as follows.

$$L_{det} = L_k + \lambda_{size} L_{size} + \lambda_{off} L_{off} + \lambda_{ang} L_{ang} \tag{2}$$

Where L_k, L_{size}, L_{off} and L_{ang} are the losses of center point identification, scale regression,offset regression and Angle regression,which are the same as DRNet.λ_{size}, λ_{off}, and λ_{ang} are constant factors and are all set to 0.1 in our experiments.

3.3 Rotational equivariant networks

We consider the scene in two-dimensional rotation, as shown in Figure 4(b), point v' is rotated around the origin of coordinates, we set the coordinates of v as (x, y), and the coordinates of v' as (x', y'), The Angle between v' and v is θ, and the Angle between v and axis x is ϕ. By converting to polar coordinates, we can obtain:

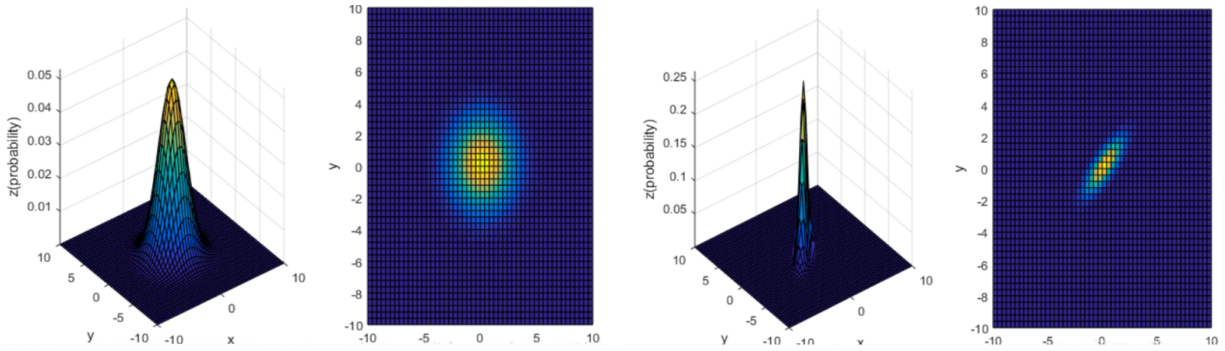
$$\begin{aligned} x &= rcos\phi \quad y = rsin\phi \\ x' &= rcos(\theta + \phi) \quad y' = rsin(\theta + \phi) \end{aligned} \tag{3}$$

The following is obtained by trigonometric expansion:

$$\begin{aligned} x' &= rcos\theta cos\phi - rsin\theta sin\phi \\ y' &= rsin\theta cos\phi + rcos\theta sin\phi \end{aligned} \tag{4}$$

Substituting x and y in 3 yields:

$$\begin{aligned} x' &= xcos\theta - ysin\theta \\ y' &= xsin\theta + ycos\theta \end{aligned} \tag{5}$$



(a) (b)

Figure 6. 2D normal distribution plots with different covariance matrices and their projection on the XOY plane.Left panel (a) has expectation $u = (0, 0)$ with variance $\sum = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$. Right panel (b) has expectation $u = (0, 0)$ with variance $\sum = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$.

Replacing this with a matrix representation gives this:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} * \begin{bmatrix} x \\ y \end{bmatrix} \tag{6}$$

With the above transformation, we can map the original pixel to the rotated pixel, so as to achieve the effect of rotation equivariant.

3.4 Two-dimensional Gaussian distribution

In Gaussian normal distribution, the joint probability density function of multidimensional variable $x = (x_1, x_2 \dots x_n)$ is as follows:

$$f(X) = \frac{1}{(2\pi)^{d/2} |\sum|^{1/2}} exp[-\frac{1}{2}(X-u)^T \sum^{-1} (X-u)], X = (x_1, x_2 ... x_n) \tag{7}$$

Where d: dimension of the variable. For a two-dimensional Gaussian distribution, we have d=2; $u = (u_1, u_2 \dots u_n)$: the mean of each variable $\sum$: The covariance Q matrix, which describes the correlation between the variables in each dimension. For the two-dimensional Gaussian distribution, we have:$\sum = \begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{pmatrix}$. As shown in Figure 6, the projection of the 2D normal distribution on the XOY plane will present different angles under different covariance matrices.We can see that when the covariance is an off-diagonal matrix, the projection of XOY is a slanted ellipse. Instead, we use the following covariance matrix:

$$\sum = \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix} \tag{8}$$

Where θ is the rotation Angle of each corresponding pixel before and after the rotation of the object.

4 Reproduction detail

4.1 Comparing with the released source codes

The original paper [20] reproduced in this paper is: Dynamic Refinement Network for Oriented and Densely Packed Object Detection. The original author did not completely open source code, so this paper is based on the details of the original paper to reproduce the original paper. I will describe the details of this in the following paragraphs.

For the original CenterNet, this paper uses the model pre-trained on DOTA by DRNet. For DRNet, this paper implements it based on Hourglass-104. And to implement the rotational covariant network, we re-implement the fully convolutional encoder-decoder network using e2cnn [23]. In this paper, the polar covariance matrix with Angle is used to replace the diagonal covariance matrix in the Gaussian kernel, and then the Gaussian heat map with Angle and stretch is generated.

After training our network on HRSC 2016, the model is trained for a total of 150 epochs. After the 90th and 120th epochs, the learning rate is finally reduced from the initial value of $1.25e-4$ to $1.25e-6$.This paper uses 2 3090 Gpus for training, using Adam optimizer and batch size set to 8.

4.2 Datasets

DOTA [25] is the largest dataset for oriented object detection in aerial images and is available in two versions: DOTAv1.0 and DOTA-v1.5. DOTA-v1.0 contains 2806 large aerial images ranging from 800 × 800 to 4000×4000, with 188,282 instances in 15 common categories: Aircraft (PL), baseball field (BD), Bridge (BR), track field (GTF), small car (SV), large car (LV), boat (SH), tennis court (TC), basketball court (BC), storage tank (ST), football field (SBF), circular intersection (RA), harbor (HA), swimming pool (SP), and helicopter (HC). DOTA-v1.5 was released for the DOAI Challenge 2019, which includes a new category, Container Cranes (CC) and more extremely small instances (less than 10 pixels). DOTA-v1.5 contains 402,089 instances. Compared to DOTA-v1.0, DOTA-v1.5 is more challenging but stable during training. Following the setup in the previous method, we used the training and validation sets for training and the test set for testing.

HRSC 2016 [18] is a challenging OBB-annotated ship detection dataset containing 1061 aerial images ranging in size from 300 × 300 to 1500 × 900. It contains 436, 181 and 444 images in the training, validation and test sets, respectively. We use the training and validation sets for training and the test set for testing. All images were resized to (800,512) without changing the aspect ratio. Random horizontal flips are applied during training.

5 Results and analysis

Table 1 shows quantitative results comparing the results of our method with DRNet's method on HRSC 2016 for the Oriented Bounding Box (OBB) task. Compared with DRNet, our method achieves significant gain in mAP_{50}. In addition, this paper also uses the MMRotate framework to compare two advanced methods, oriented-rcnn [26] and rotated-faster-rcnn [27], on the HRSC 2016 dataset using our method. It can be seen that our improved model still has advantages.

Method	mAP50	Recall	mIoU
DRNet	86.8	82.3	57.3
oriented-rcnn	87.4	80.3	57.5
rotated-faster-rcnn	87.6	81.9	56.1
Ours	88.2	82.6	58.6

Table 1. Evaluation results of the oriented bounding box task on the HRSC 2016 dataset.

Method	Map(VOC2007)	Map(VOC2012)
Baseline	86.54	92.3
e2cnn	87.27	93.04
e2cnn + Two-dimensional Gaussian distribution	88.03	93.38

Table 2. Results of ablation experiments on the HRSC 2016 dataset.

6 Ablation Study

In this section, we perform a series of ablation experiments on the HRSC 2016 test set and report quantitative results in VOC 2007 and VOC 2012 to evaluate the validity of our proposed approach. Note that we use hourglass-104 as the backbone of our section and DRNet as our baseline method.

Table 2 shows the results of different ablation experiments. The first line is the baseline, the reference paper DRNet, where you can see the original accuracy data. It can be seen that after adding the rotational isovariable network e2cnn, the average accuracy of 0.77 and 0.74 has been improved in VOC2007 and VOC2012 respectively. Then, we continued to add Two-dimensional Gaussian distribution on this basis, so that the detection effect was better improved, which were 1.49 and 1.08 respectively. This ablation experiment further reveals the effectiveness of the two improved methods proposed in this paper.

7 Conclusion

In this work, we improve the rotation dense object detector of DRNet by introducing a rotation equivariant network and modifying the covariance matrix in the two-dimensional Gaussian distribution. Rotation equivariant network can solve the problem that Angle regression does not have rotation equivariant with the current deep network, and the covariance matrix in the improved two-dimensional Gaussian distribution makes the heat map better fit the detected target. In this paper, we conduct

experiments with DRNet and other two models on HRSC 2016 to prove the effectiveness of the two methods.

References

[1] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In European conference on computer vision, pages 213–229. Springer, 2020.

[2] Gong Cheng, Peicheng Zhou, and Junwei Han. Learning rotation-invariant convolutional neural networks for object detection in vhr optical remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 54(12):7405–7415, 2016.

[3] Taco Cohen and Max Welling. Group equivariant convolutional networks. In International conference on machine learning, pages 2990–2999. PMLR, 2016.

[4] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In Proceedings of the IEEE international conference on computer vision, pages 764–773, 2017.

[5] Jian Ding, Nan Xue, Yang Long, Gui-Song Xia, and Qikai Lu. Learning roi transformer for oriented object detection in aerial images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2849–2858, 2019.

[6] Ross Girshick. Fast r-cnn. In Proceedings of the IEEE international conference on computer vision, pages 1440–1448, 2015.

[7] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 580–587, 2014.

[8] Eran Goldman, Roei Herzig, Aviv Eisenschtat, Jacob Goldberger, and Tal Hassner. Precise detection in densely packed scenes. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 5227–5236, 2019.

[9] Jiaming Han, Jian Ding, Nan Xue, and Gui-Song Xia. Redet: A rotation-equivariant detector for aerial object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2786–2795, 2021.

[10] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision, pages 2961–2969, 2017.

[11] Emiel Hoogeboom, Jorn WT Peters, Taco S Cohen, and Max Welling. Hexaconv. arXiv preprint arXiv:1803.02108, 2018.

[12] Meng-Ru Hsieh, Yen-Liang Lin, and Winston H Hsu. Drone-based object counting by spatially regularized regional proposal network. In Proceedings of the IEEE international conference on computer vision, pages 4145–4153, 2017.

[13] Han Hu, Jiayuan Gu, Zheng Zhang, Jifeng Dai, and Yichen Wei. Relation networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3588–3597, 2018.

[14] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. Advances in neural information processing systems, 28, 2015.

[15] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2117–2125, 2017.

[16] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision, pages 2980–2988, 2017.

[17] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, pages 21–37. Springer, 2016.

[18] Zikun Liu, Hongzhen Wang, Lubin Weng, and Yiping Yang. Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds. IEEE geoscience and remote sensing letters, 13(8):1074–1078, 2016.

[19] Jianqi Ma, Weiyuan Shao, Hao Ye, Li Wang, Hong Wang, Yingbin Zheng, and Xiangyang Xue. Arbitrary-oriented scene text detection via rotation proposals. IEEE transactions on multimedia, 20(11):3111–3122, 2018.

[20] Xingjia Pan, Yuqiang Ren, Kekai Sheng, Weiming Dong, Haolei Yuan, Xiaowei Guo, Chongyang Ma, and Changsheng Xu. Dynamic refinement network for oriented and densely packed object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 11207–11216, 2020.

[21] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28, 2015.

[22] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 7464–7475, 2023.

[23] Maurice Weiler and Gabriele Cesa. General e (2)-equivariant steerable cnns. Advances in neural information processing systems, 32, 2019.

[24] Maurice Weiler, Fred A Hamprecht, and Martin Storath. Learning steerable filters for rotation equivariant cnns. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 849–858, 2018.

[25] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. Dota: A large-scale dataset for object detection in aerial images. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3974–3983, 2018.

[26] Xingxing Xie, Gong Cheng, Jiabao Wang, Xiwen Yao, and Junwei Han. Oriented r-cnn for object detection. In Proceedings of the IEEE/CVF international conference on computer vision, pages 3520–3529, 2021.

[27] Sheng Yang, Ziqiang Pei, Feng Zhou, and Guoyou Wang. Rotated faster r-cnn for oriented object detection in aerial images. In Proceedings of the 2020 3rd International Conference on Robot Systems and Applications, pages 35–39, 2020.

[28] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl. Objects as points. arXiv preprint arXiv:1904.07850, 2019.

[29] Yanzhao Zhou, Qixiang Ye, Qiang Qiu, and Jianbin Jiao. Oriented response networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 519–528, 2017.