2021 International Conference on Artificial Intelligence, Big Data and Algorithms (CAIBDA) | 978-1-6654-2490-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/CAIBDA53561.2021.00060

XRDI: A Database of X-Ray Dangerous Items

Yu Wang, Li Ming Yang, Zhuo Qing
Tsinghua University
Dept. of Automation, School of Information Science and Technology
Beijing China
{ yu-w18, limy17}@mails.tsinghua.edu.cn
zhuoqing@tsinghua.edu.cn

***Abstract*—In order to provide a data set and basic detection indicators for the identification and detection of X-ray dangerous items, the study proposed a method to generate X-ray dangerous items data, which is to collect more than ten kinds of dangerous items and over 100 kinds of non-dangerous items. The original data under different photographing directions is processed by segmenting and denoising, random rotating and zooming, and gray-scale image overlay synthesis based on the principle of X-ray imaging. Finally, a data set containing 10,000 positive samples and 100,000 negative samples is generated, which contains three difficulty levels divided by photographing directions. The research uses several existing mainstream target detection algorithms to conduct algorithm experiments on different data sets, gives the mAP and Recall results of different kinds of dangerous items under different difficulty test sets, organizes and compares the results under the conditions of different training sets and test sets, explores the influence of the number of training samples on the results. These results verified the research significance of the data set in the field of dangerous items identification and detection, and laid the foundation for subsequent research.**

***Index Terms*—X-ray, Dangerous items, Database, Object Detection**

I. INTRODUCTION

Image detection plays an important role in the protection of public social spaces from security threats. In recent years, with the rapid development of deep learning[1], convolutional neural networks have brought great progress to the field of image processing. Some Related works have focused on tasks such as identifying and discovering targets in X-ray images[2][3][4]. For X-ray images, targets may be mixed among many other objects, and the difficulty of identification may even make a human miss such targets.

In order to provide datasets and basic detection metrics for research in this area, we present a simple method for generating dangerous items data, which overcomes the difficulty of collecting dangerous items data. We present a well-generated dataset called the X-ray Data Set of Dangerous Items (XRDI), where we get the x-ray scans at multiple angles of each original sample. We present a simple schematic of the angular acquisition and some sample acquisitions. At the same time, we generate a dataset that will be much larger than the currently existing datasets. Samples of the dataset are shown in Figure 1.

Dataset: https://pan.baidu.com/s/1-gYdq6P833GVF5dpL5u_qw

Password:XRDI

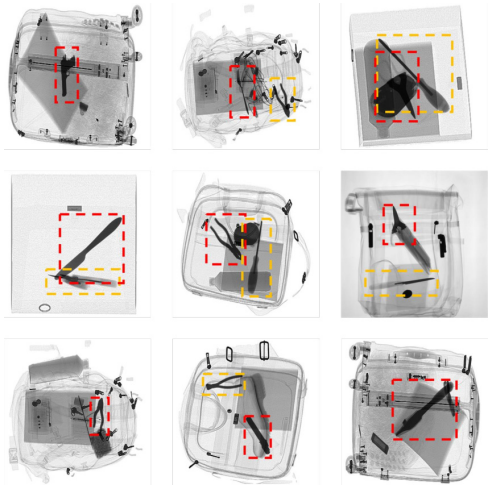

Fig.1. Samples in X-Ray Dangerous Items

In order to adapt the context of the safety field，we apply the dataset to target detection tasks to give a reference basis for subsequent research in the field of dangerous items identification.

In this work, our main contributions include: we propose and implement a method for generating a large database of dangerous items samples, which is based on the collection of data from more than 10 dangerous items from different angles and up to 100 non-hazardous items that can easily be seen in real-life scenarios as sample seeds, synthesized by the physical laws of X-ray imaging; we apply the dataset on different algorithms and the database of different difficulty levels is constructed through the distribution of different viewpoints, and the effect of viewpoint on detection is revealed through the analysis of identification results.

II. RELATED WORK

A. X-ray Imaging Principles and Datasets

Assume that the incident intensity of X-rays is I_0， The intensity of the transmission to depth x is I_x. Experiments show that the rate of change of the optical intensity of the X-ray penetration through the thin layer is proportional to the thickness of the layer：

$$\frac{dI_x}{I_x} = \mu_l dx \tag{1}$$

978-1-6654-2490-5/21/$31.00 ©2021 IEEE
DOI 10.1109/CAIBDA53561.2021.00060

Where μ_l is called the line attenuation coefficient. Assuming that the thickness of the object through which the X-rays pass is l, the integration of the above equation yields the X-ray intensity：

$$I_l = I_0 e^{-\mu_l l} = I_0 e^{-\frac{\mu_l}{\rho} l \rho} = I_0 e^{-\mu_m m} \quad (2)$$

Where μ_m is the mass attenuation coefficient and m is the mass per unit area of the object penetrated in the direction of X-ray penetration. Mass attenuation coefficients of the same substance are the same, and are not affected by the state of the object.

Consider the case of a mixture of substances through which X-rays penetrate. The total mass attenuation coefficient of the mixture is equal to the weighted average of the elemental mass attenuation coefficients of the constituent substances according to their mass fractions：

$$I_2 = I_0 e^{-(\mu_1 m_1 + \mu_2 m_2)} \quad (3)$$

Where μ_1， μ_2 are the mass attenuation coefficient of two substances， m_1，m_2 are the mass of two substances penetrated per unit area in the direction of travel. The same calculation can be made when passing through more than two substances.

Compared to traditional image datasets, X-ray datasets are characterized by more overlap and by different gray levels depending on the constituent material. Existing open source X-ray datasets include GDXray[5]、ChestXray8[6]、SIXray[7] etc. GDXray contains only three kinds of dangerous items, with few background types and overlaps, and the overall size of the dataset is relatively small; ChestXray8 is a dataset of medical chest radiographs, which has little connection with dangerous items; SIXray's data is taken from the security machines in subway stations, which contains six types of dangerous items, with a large amount of data, but mostly negative samples that do not contain dangerous items, and one of the samples is excluded from the experiment..

B. Identification and Targeting

This study will focus on target detection, a research area based on deep learning methods. In fact, there may be more than one target in the image of our dataset. Therefore, we have two types of methods to confirm the existence of a target. The first one works at the image level and will predict a score for each category to indicate whether the category exists or not [8]. The other one, acting at the target level, will generate a box with the same category information for each target, which will give both the location and presence of the target[9][10] . Both methods have their place in real tasks. In one case, we need to determine the presence or absence of a dangerous item in order to decide whether or not to alert the police, and in the other case, we need to identify the location of the dangerous item in more detail so that it can be eliminated in a timely manner. In this study, we use the target detection method for our task in order to provide a baseline for subsequent researchers.

In the case of object recognition in X-ray images, researchers realized that these images typically contain less texture information and more shape information. Therefore, the topic of designing effective and efficient manual features was explored in depth. As deep learning becomes a standard tool for optimizing complex functions, researchers are beginning to apply it to extracting compact visual features for X-ray image representation or to fine-tune the use of X-ray images in pre-trained models so that representations can be borrowed from natural images, as discussed in the experimental section of this paper.

III. DATASET GENERATION

Considering the fact that images containing dangerous items are few and difficult to obtain in reality and the principles of X-ray imaging, we propose the idea of using a synthesis method to construct the dataset. That is, we first take X-ray images of a single item, including both dangerous items and general articles and background containers, and then generate a dataset by superimposing the images.

A. Raw Data

We selected about 140 items commonly found in luggage or dangerous items, with two or three poses for each item, and took X-ray fluoroscopic images every 10° around 180°.

We took a total of 7,401 X-ray images, including 1,222 of dangerous items, 4,688 of general items, and 1,491 of containers.

B. Generating Methods

In this study, we propose a method to obtain dangerous items data individually and synthesize them to form large dangerous items datasets based on the assumption of physical laws through X-ray imaging. In order to overlay the synthesis of the captured individual items, the raw data first needs to be pre-processed. We used the Gray-level thresholding selection [11]、Canny edge detection[12] and Graph Cuts Method[13] to segment and denoise the image, so that an image that contains only the target object can be extracted. An example of the original image and the extracted image is shown in Figure 2.

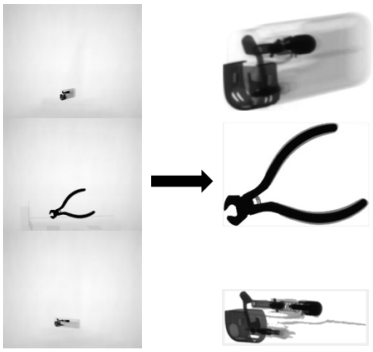

Fig.2. Examples of original and extracted images

After obtaining the segmented image, according to the X-ray imaging principle, when X-rays pass through different substances, their emission intensity is superimposed exponentially. Therefore, the grayscale image should be superimposed exponentially in the image synthesis. The formula used for the superposition is

$$I_2 = e^{I_0 + I_1} \quad (4)$$

Where I_0 is the grayscale of the original image, I_1 is the grayscale of the item to be superimposed, and I_2 is the grayscale of the image after superimposition.

In order to simulate the images taken in real scenes and to enrich the data set, we randomized the size, orientation, and position of each object within a certain range during compositing, and included as many types of objects and overlapping relationships as possible. When compositing, we considered the effect of occlusion on X-ray intensity and applied transparency to the objects to be overlaid. Figure 3 shows an example composite of an image in the dataset, which contains a background image, two dangerous items and one common object.

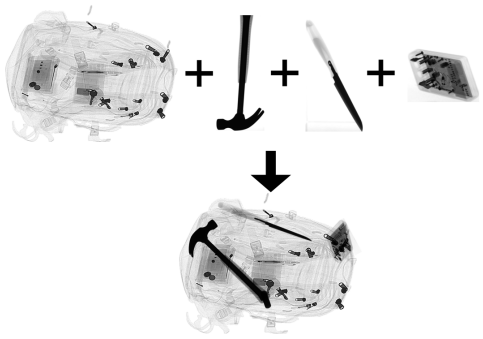

Fig.3. Example of data set image composition

We classify the synthesized database into three levels: easy, medium, and hard, according to the number of photo- graphing angles used. The easy level is composed of six angles for each dangerous items, the medium level is composed of twelve angles, and the hard level is composed of eighteen angles, as shown in Figure 4. In addition, we also synthesized a negative sample dataset without dangerous items, consisting only of background and general items. The positive sample contains 10,000 images of each difficulty and the negative sample contains 100,000 images.

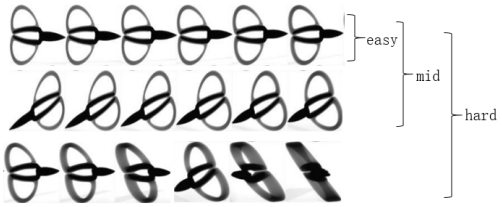

Fig.4. Example of data set difficulty levels

C. Overview of the Dataset

Our synthetic dataset has the following features: first, we have reproduced the X-ray images of real-life scenarios, with various items appearing randomly in various poses and with many overlapping cases; second, we have selected more than a dozen dangerous items as positive samples, which basically cover the various types of dangerous items that may appear in security checks, and these dangerous items appear either individually or in combination in the positive data set; third, in order to simulate the real situation, the number of negative samples is much larger than the number of positive samples, which poses a challenge for both classification and detection tasks; last but not least, our raw data contains the postures of various items rotated in a circle at various angles, and the dataset is divided into three levels according to the number of selected angles for subsequent experimental studies.

IV. ALGORITHM EXPERIMENTS

A. Experiment Setups

We used three difficulty levels, simple, mid, and hard, to evaluate the performance of different algorithms. On each sub-dataset, each model will use 90% of the dataset for training and the remaining 10% as a test dataset. These datasets are divided into two parts using random sampling, while ensuring that the relative proportions of the various target items remain constant. In addition, we also synthesize a dataset rotating around the object as a whole for further research, which contains continuous angular image information of the target objects.

In target detection algorithms, two main metrics are evaluated, the image-level mean accuracy (mAP) and the target-level location accuracy (IOU). Due to the specificity of the dangerous items detection task, we are mainly concerned with detecting the dangerous items as accurately as possible, without having to be very precise about its location. Therefore, we consider AP50 to represent AP, i.e., we set the IOU threshold to 0.5, and we only calculate APs with confidence thresholds higher than 0.3, which is relatively low, but more suitable for the actual dangerous items detection task. For calculating the average accuracy, we followed the PascalVOC image classification task, and for each category, each test image was ranked by the confidence level obtained from the model algorithm that contains a particular category, and then the average accuracy (mAP) was calculated.

In order to give a relatively comprehensive benchmark, we apply three popular one-stage detection algorithms, including YoLov3[14]、SSD[15] and RetinaNet[16]. The number of training steps is set to 100, and all parameters are set to common settings. When evaluating the model, the thresholds mentioned above were used, and each set of experiments was repeated 10 times to get the average result.

Finally, in order to show the advantages of multi-angle dataset, we train the model on different training sets with different difficulty levels and test it on different test sets to investigate the effect of angle on the accuracy of dangerous items detection.

B. mAP Results

First, we used the model obtained from training with the mid difficulty dataset to test it on all difficulties to obtain the mAP results, which are presented in Table 1 (note: M-E stands for training with the mid difficulty training set, testing with the simple difficulty test set, and so on).

TABLE I. MAP RESULTS

YoLoV3			SSD			RetinaNet		
M-E	M-M	M-H	M-E	M-M	M-H	M-E	M-M	M-H

1	87.15	84.75	78.94	88.57	84.35	79.60	92.03	83.11	79.47
2	92.74	85.44	68.11	87.49	88.60	68.91	92.59	87.28	67.29
3	83.15	85.97	69.78	81.71	83.64	64.93	83.12	82.98	67.53
4	92.61	92.45	79.20	93.27	93.57	85.14	93.16	97.77	79.00
5	79.17	80.00	26.27	76.13	83.81	22.62	76.23	82.63	26.86
6	82.32	91.97	74.81	81.12	91.94	71.66	80.19	91.41	79.12
7	82.52	76.95	68.18	82.38	72.78	62.72	83.69	77.77	68.26
8	96.12	91.08	88.03	94.35	89.60	88.00	96.16	92.18	91.31
9	92.77	86.10	77.75	93.99	83.08	79.14	89.70	91.40	83.14
10	91.92	91.53	85.18	94.82	87.46	85.16	91.18	93.59	85.83
11	91.57	90.05	84.15	88.65	88.46	81.26	96.77	90.40	90.27
12	82.82	87.76	79.34	78.43	87.44	83.11	87.86	86.61	81.56
mAP	87.90	87.00	73.31	86.75	86.23	71.86	88.56	88.09	74.97

1-armyknife 2-axe 3-chopper 4-cutterknife 5-dagger 6-fork 7-fruitknife 8-hammer 9-pistol 10-pilers 11-scissors 12-screwdriver

In general, RetinaNet gets the highest mAP in all three difficulty datasets, which is sufficient to show that the algorithm is better in terms of detection accuracy. This is due to the fact that the dataset we generated fits the real dangerous items detection scenario, where there are a large number of negative samples distributed in the collected images and may intersect or overlap with the positive samples, whereas RetinaNet uses Focal loss in an effort to overcome the problems caused by such a situation. It reduces the background loss and focuses the training on object detection. YoLov3, on the other hand, combines speed and accuracy, and ranks second in overall performance, as shown in the table. We find that some small items can still be detected with good performance, thanks to the multi-scale feature map detection of these algorithms, so that small targets are not lost by the convolution of multiple layers. In the dataset, the mAP is also relatively high for some very distinctive categories, both in terms of morphology and X-ray spectral imaging. The categories that would produce greater similarity in some angles tend to influence the judgments.

Further, we looked at the mAP for test sets of different difficulty and found that samples that contain more angles will be more difficult to detect, thus reducing the accuracy of detection. This illustrates that multiple perspectives and angles of access to the target would contribute to the robustness of the model. In later experiments, we will verify the relationship between the number of angles and the detection model.

C. Recall Results

The goal of the inspector is to detect all dangerous items and ensure safety, so a high recall rate is needed to ensure that the dangerous items are detected. However, a high recall rate leads to a low accuracy, i.e., a much higher false detection rate. In the course of our experiments, we found that a confidence threshold of 0.3 can have a good accuracy rate with a high recall rate.

Next, we show in Table 2 the recall performance of various algorithms on different difficulty test datasets with a threshold of 0.3. We find that in terms of recall performance, RetinaNet still has some advantages in the overall metrics. On individual categories, such as daggers, we find that on the hard test dataset, we get very poor results, as shown in Tables 1 and 2. This is because at some extreme angles the dagger appears as a long black bar in the X-ray image, which interferes with training and testing. How to make the algorithmic model better identify dangerous items at extreme angles is also a direction for future research.

TABLE II. RECALL RESULTS

	YoLoV3			SSD			RetinaNet		
	M-E	M-M	M-H	M-E	M-M	M-H	M-E	M-M	M-H
1	87.78	88.76	83.75	91.71	86.53	83.39	88.01	92.05	81.83
2	92.86	87.76	68.25	88.83	84.61	71.15	92.41	90.71	71.69
3	83.87	86.15	74.55	83.17	86.71	74.96	84.55	86.98	77.45
4	93.48	93.10	80.28	93.73	88.49	79.31	98.64	89.40	79.55
5	79.07	80.11	32.00	76.07	83.63	29.80	80.17	83.64	34.53
6	82.93	92.55	75.61	82.99	89.30	74.79	85.24	95.31	78.26
7	85.54	79.72	72.36	85.86	81.85	66.86	83.32	79.77	69.60
8	96.26	91.11	88.24	94.40	88.33	87.46	97.69	92.34	85.25
9	93.51	86.67	80.90	89.68	82.60	84.69	96.74	87.18	82.20
10	92.41	91.82	85.67	91.81	93.83	86.77	94.47	93.34	90.51
11	92.18	90.55	85.61	95.71	90.56	83.99	91.43	90.35	92.19
12	83.72	89.29	80.95	79.45	89.42	76.16	84.03	90.44	81.99
recall	88.64	88.12	75.68	87.78	87.15	74.94	89.72	89.29	77.09

1-armyknife 2-axe 3-chopper 4-cutterknife 5-dagger 6-fork 7-fruitknife 8-hammer 9-pistol 10-pilers 11-scissors 12-screwdriver

D. Multi-Perspective Experiments

The datasets of different difficulties contain dangerous items distributed in different angles. We did experiments using training sets and test sets of multiple difficulties on YoLov3. The detailed results are shown in Figure 5.

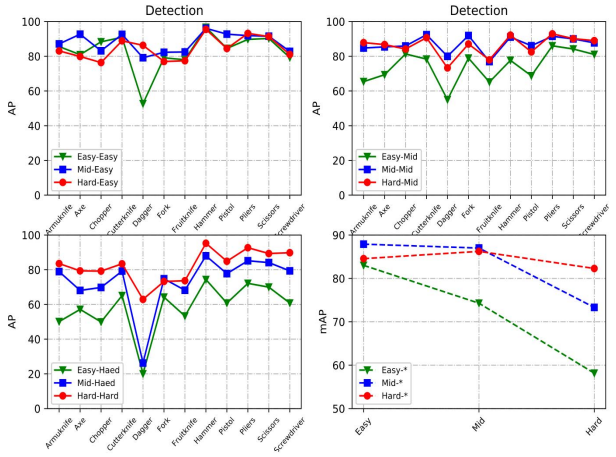

Fig.5. Demonstration of experimental results

The graph in the upper left corner shows that the algorithms have the worst performance using the simple training on the simple difficulty test set. And the network performance is better for the mid difficulty than for the hard difficulty, especially on the axe and pistol categories. Meanwhile, the results in the upper right corner also show this characteristic. When the test set is of mid difficulty, we find a significant lag in the performance of the networks trained with the simple training set, while the relative comparison between the mid and hard training acquisition models shows that the mid difficulty still has an advantage. The results illustrate that too much angular information in training sometimes does not lead to better model performance.

In the lower left corner of the graph, it can be seen that in the case where the test set is hard, the results for the three difficulty training sets are consistent with more angular information leading to better performance. The results show that when the model training data is much less than the data used for testing, it will lead to poorer performance. Finally, the lower-right graph

further shows the overall performance of the fixed-difficulty trained model on the full test set. We find that in the simple and mid training models, performance decreases as the test difficulty increases, while the hard do not show such a trend. This further suggests that, under certain conditions, increasing the sample angle too much can reduce the performance of the model on relatively simple data.

We attempted to investigate whether too much angular information can have a negative effect on the model by increasing the sample. We generated another 90000 images as the training set and the results of the experiments are presented in Table 3.

TABLE III. THE INFLUENCE OF TRAINING SAMPLE NUMBER ON MAP

Sample number	H-E	H-M	H-H
9000	84.56	**86.24**	82.29
45000	88.03	**88.47**	85.79
90000	**90.11**	89.72	88.74

The results show that as the number of training samples increases, the advantage of angular information becomes apparent， and when the sample size is limited, it is more important to choose the right angle information.

V. CONCLUSION

Dangerous items detection is a very valuable application that is still less focused on in computer vision research. In order to contribute to the development of this field, we propose XRDI, which is a large dataset of multi-angle X-ray images formed by the generation method proposed in this paper. All sample seeds are obtained from real collections. We also consider complex scenarios and add more than 100 common sample seeds to fit the actual formation of complex backgrounds. At the same time, based on the flexibility of our method, we can also generate a dataset on demand to meet the requirements of specific tasks, which is not bound by a fixed dataset and is more meaningful for practical applications. At the algorithmic level, we provide an exhaustive overview of the performance of mainstream target detection algorithms on XRDI to establish a baseline for subsequent research. Finally, we discuss the impact of multiple perspectives on the algorithm: as the sample size increases, more perspective information can help the model achieve better performance, while the selection of appropriate perspective information is more important in the case of limited sample size. There are two main directions for future research. First, the overlapping images from the penetration assumption may be inaccurate in some ways, and we expect better physical models to generate the data. Second, how to further improve the target detection performance using multi-view target data, so that these methods can be extended to the broader field of computer vision.

ACKNOWLEDGMENT

This work is supported by The Third Research Institute Of Ministry Of Public Security.

REFERENCES

[1] LeCun Y, Bengio Y, and Hinton G. Deep learning[J]. Nature, 521(7553):436, 2015.

[2] Mery D. X-ray testing by computer vision[C] // In CVPRW: 360–367, 2013.

[3] D. Mery and C. Arteta. Automatic defect recognition in x-ray testing using computer vision[C] // In WCCV, 1026–1035, 2017.

[4] Mery D, Svec E, Arias M, et al. Modern computer vision techniques for x-ray testing in baggage inspection[J]. Systems, Man, and Cybernetics: Systems, 47(4):682–692, 2017.

[5] GDXray: The Database of X-ray Images for Nondestructive Testing[J]. Journal of Nondestructive Evaluation, 2015, 34(4):1-12.

[6] Wang X , Peng Y , Lu L , et al. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases[C]// Computer Vision & Pattern Recognition. IEEE, 2017.

[7] Miao C , Xie L , Wan F , et al. SIXray: A Large-Scale Security Inspection X-Ray Benchmark for Prohibited Item Discovery in Overlapping Images[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020..

[8] Lu J , Liong V E , Zhou X , et al. Learning Compact Binary Face Descriptor for Face Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(10):1-1.

[9] Girshick R . Fast r-cnn[C]// 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, 2016.

[10] Ren S , He K , Girshick R , et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2017, 39(6):1137-1149.

[11] Otsu, N. A thresholding selection method from gray-level histogram[J]. IEEE Trans.syst.man. & Cybern, 1979, 9(1):62-66.

[12] Canny, John. A computational approach for edge detection[J]. 1986, PAMI-8(6):679-698.

[13] Boykov Y , Veksler O , Zabih R . Fast approximate energy minimization via graph cuts[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 23(11):1222-1239.

[14] Redmon J , Farhadi A . YOLOv3: An Incremental Improvement[J]. 2018.

[15] Liu W , Anguelov D , Erhan D , et al. SSD: Single Shot MultiBox Detector[C]// European Conference on Computer Vision. Springer International Publishing, 2016.

[16] Lin T Y , Goyal P , Girshick R , et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2017, PP(99):2999-3007.