**EIoU: An Improved Vehicle Detection Algorithm Based on VehicleNet Neural Network**

Zuomin Yang*, Xianlun Wang and Jianguang Li

College of Electromechanical Engineering, Qingdao University of Science and Technology, Qingdao, China

*QUSTYANG@mails.qust.edu.cn

**Abstract.** The paper’s primary purpose is to optimize the performance (speed/accuracy) of vehicle detection. The vehicle dataset Vehicle2020 used in this paper is divided into ten different vehicle classes. Intersection over Union (IoU) is usually used as a standard to evaluate the accuracy of vehicle detection in a specific dataset. However, IoU as a performance of the object detection algorithm is still shortcomings. IoU is further improved and called a new algorithm EIoU. Finally, the neural network structure was redesigned, which was called VehicleNet. The experimental results show that EIoU as a performance evaluation algorithm used the vehicle detection framework can improve the performance of vehicle detection. Using the algorithm of this paper shows the performance superiority of vehicle detection.

1. **Introduction**

In 2014, Ross B. Girshick used region proposal + CNN [1], and designed the R-CNN framework, which made a huge breakthrough in object detection. The detection framework used by NIPS2015 version of Faster R-CNN [2] is the RPN network + Fast R-CNN network. The RPN and Fast R-CNN used for target detection accuracy will be further improved, but the speed does not meet the real-time requirements. The YOLO [3][4][5] series of methods slowly showed their importance. Given the input image, it directly returns the target of this position at multiple positions of the image border and target category. Besides, the researchers have proposed a series of methods to improve the performance of target detection based on these frameworks from other aspects. For example, SNIPER [6], CornerNet-Lite [7], etc. which are relatively strong target detection algorithms recently.

Intersection over Union (IoU) is a standard for measuring the accuracy of detecting vehicles in a specific data set. However, IoU as a metric and loss has a significant problem. In this article, it is an excellent choice to solve the weaknesses of IoU by extending the IoU algorithm to non-overlapping situations. We continue any two shapes of IoU to any three forms and call it IoUt. We define the generalized version of IoUt as EIoU, and use EIoU as a new metric for comparing any three shapes. Incorporating EIoU loss into the most advanced object detection algorithm, based on the standard of IoUt and the performance measure based on EIoU, improves the performance of vehicle detection.

2. **The dataset**

We made a simple dataset called it Vehicle2020 for vehicle detection. The dataset is divided into ten classes, Truck, Racing car, Minibus, Taxi, Family sedan, Bus, SUV, Fire engine, Jeep and Motorbike.
As shown in Figure 1. The distribution of the number of training sets and test sets for each class in Table 1.

![Figure 1. Classes of data sets.](image1)

### Table 1. Number of the vehicle per classes.

| Classes            | Train | Test | Classes    | Train | Test |
|--------------------|-------|------|------------|-------|------|
| Truck              | 200   | 20   | Bus        | 200   | 20   |
| Racing car         | 200   | 20   | SUV        | 200   | 20   |
| Minibus            | 200   | 20   | Fire engine| 200   | 20   |
| Taxi               | 200   | 20   | Jeep       | 200   | 20   |
| Family sedan       | 200   | 20   | Motorbike  | 200   | 20   |

**Figure 1.** Classes of data sets.

### 3. Methodology

#### 3.1. Intersection over Union (IoU) improvements

Bounding boxes regression detection is an excellent method to improve the accuracy of object detection. IoU as the performance evaluation of target detection is usually adopted. IoU calculates the ratio of the intersection and union of "predicted border" and "true border," as shown in Figure 2.

![Figure 2. Intersection over Union (IoU).](image2)

Since the regression loss of the bounding box does not directly represent the accuracy evaluation of the target detection. This paper promotes it IoU and names it IoU\_i as shown in Fig. 3. However, IoU\_i has a big disadvantage. If |A ∩ B ∩ C| = 0, IoU\_i(A, B, C) = 0. In this case, IoU\_i cannot reflect whether the three shapes are near each other or far from each other. Based on the above problem, IoU\_i is further improved, as shown in Table 2. A, B, C and D respectively represent different regions of target detection, and each region has four different coordinates. The algorithm in Table 2 correspond to the schematics in Figure 2 and Figure 3.

![Figure 3. Intersection over Union-Improved (IoU\_i).](image3)

### Table 2. Intersection over Union-Improved (EIoU).

**Algorithm1: Intersection over Union-Improved**

| Step | Description                                                                                           |
|------|--------------------------------------------------------------------------------------------------------|
| 1    | Input :Three arbitrary shapes: A, B, C ⊆ S ⊆ \( \mathbb{R}^n \)                                      |
| 2    | Output: EIoU                                                                                           |
| 3    | 1 For A, B and C, find the smallest enclosing object D, where D ⊆ S ⊆ \( \mathbb{R}^n \)           |
| 4    | \( \text{EIoU} = \text{IoU}_i - \frac{|D \setminus (A \cup B \cup C)|}{|B|} \)                   |
| 5    | \( \mathcal{L}_{\text{IoU}_i} = 1 - \text{IoU}_i, \mathcal{L}_{\text{EIoU}} = 1 - \text{EIoU} \)   |
As shown in Table 2, to improve the performance of target detection. First, find an arbitrary convex shape (body) $D$, making $D \subseteq \mathbb{S} \subseteq \mathbb{R}^n$. $D$ is the smallest closed convex shape between $A$, $B$, and $C$. Then, $EIoU$ is obtained by subtracting $|\Delta(\text{Area})|$ from the $IoU_i$ value. See Alg.1 for the algorithm of $EIoU$. $EIoU$ as a new target detection performance evaluation, $EIoU$ is better than $IoU_i$, $\forall A, B, C \subseteq \mathbb{S}$, $EIoU(A, B, C) \leq IoU_i(A, B, C)$, $\lim_{A \rightarrow B \rightarrow C} IoU_i(A, B, C) = 0$. $\forall A, B, C \subseteq \mathbb{S}$, $0 \leq IoU_i(A, B, C) \leq 1$, but $EIoU$ has symmetry, when $\forall A, B, C \subseteq \mathbb{S}$, $-1 \leq EIoU \leq 1$. When the three objectives completely overlap, $|A \cup B \cup C| = |A \cap B \cap C|$, then $EIoU(A, B, C) = IoU_i(A, B, C) = 1$. When the ratio of the areas occupied by the three target shapes $|A \cup B \cup C|$, the volume (area) of the closed form $|D|$ approaches 0, and the $EIoU$ value gradually converges to -1.

3.2. Bounding box regression losses

| Table 3. The analytical solution of $IoU_i$ and $EIoU$ as bounding box losses. |
|-------------------------------------------------|
| Algorithm 2: $IoU_i$ and $EIoU$ as bounding box losses |
|-------------------------------------------------|
| **Input:** Predicted $G^p$, $G^q$ and ground truth $G^g$ coordinates |
| $G^p = (x_1^p, y_1^p, x_2^p, y_2^p)$, $G^q = (x_1^q, y_1^q, x_2^q, y_2^q)$, $G^g = (x_1^g, y_1^g, x_2^g, y_2^g)$ |
| **Output:** $L_{IoU_i}$, $L_{EIoU}$ |
| 1 For the predicted box $G^p$, ensuring $x_2^p > x_1^p$, $y_2^p > y_1^p$ & $x_2^q > x_1^q$, $y_2^q > y_1^q$ |
| $\hat{x}_1^p = \min(x_1^p, x_2^p)$, $\hat{x}_2^p = \max(x_1^p, x_2^p)$ |
| $\hat{y}_1^p = \min(y_1^p, y_2^p)$, $\hat{y}_2^p = \max(y_1^p, y_2^p)$ |
| $\hat{x}_1^q = \min(x_1^q, x_2^q)$, $\hat{x}_2^q = \max(x_1^q, x_2^q)$ |
| $\hat{y}_1^q = \min(y_1^q, y_2^q)$, $\hat{y}_2^q = \max(y_1^q, y_2^q)$ |
| 2 Calculating area of $G^g$: $A^g = (x_2^g - x_1^g) \times (y_2^g - y_1^g)$ |
| 3 Calculating area of $G^p$: $A^p = (\hat{x}_2^p - \hat{x}_1^p) \times (\hat{y}_2^p - \hat{y}_1^p)$ |
| 4 Calculating area of $G^q$: $A^q = (\hat{x}_2^q - \hat{x}_1^q) \times (\hat{y}_2^q - \hat{y}_1^q)$ |
| 5 Calculating intersection $\mathcal{K}$ between $G^p$, $G^q$ and $G^g$: |
| $x_1^\mathcal{K} = \max(\hat{x}_1^p, \hat{x}_1^q, x_1^g)$, $x_2^\mathcal{K} = \min(\hat{x}_2^p, \hat{x}_2^q, x_2^g)$ |
| $y_1^\mathcal{K} = \max(\hat{y}_1^p, \hat{y}_1^q, y_1^g)$, $y_2^\mathcal{K} = \min(\hat{y}_2^p, \hat{y}_2^q, y_2^g)$ |
| $\mathcal{K} = \{(x_2^\mathcal{K} - x_1^\mathcal{K}) \times (y_2^\mathcal{K} - y_1^\mathcal{K}) \text{ if } x_2^\mathcal{K} > x_1^\mathcal{K}, y_2^\mathcal{K} > y_1^\mathcal{K} \}$ |
| otherwise |
| 6 Finding the coordinates of smallest enclosing box $G^d$: |
| $x_1^d = \min(\hat{x}_1^p, \hat{x}_1^q, x_1^g)$, $x_2^d = \max(\hat{x}_2^p, \hat{x}_2^q, x_2^g)$ |
| $y_1^d = \min(\hat{y}_1^p, \hat{y}_1^q, y_1^g)$, $y_2^d = \max(\hat{y}_2^p, \hat{y}_2^q, y_2^g)$ |
| 7 Calculating area of $G^d$: $A^d = (x_2^d - x_1^d) \times (y_2^d - y_1^d)$ |
| 8 $IoU_i = \frac{\mathcal{K}}{U}$ where $U = A^p + A^q + A^g - 2\mathcal{K}$ |
| 9 $EIoU = 1 - IoU_i - \frac{A^d - U}{A^d}$ |
| 10 $L_{IoU_i} = 1 - IoU_i$, $L_{EIoU} = 1 - EIoU$ |
In this paper, the intersection and the smallest closed target are represented by rectangular bounding boxes. It can be seen that the coordinates of their vertices are only the coordinates of one of the three bounding boxes compared, and the values of each vertex compare by the min and max value-coordinate realization. Therefore, $EIoU$ can directly use as a loss of $L_{EIoU}$ to optimize target detection performance based on deep neural networks. Table 3 is the bounding box loss algorithm of $IoU_i$ and $EIoU$, the area of $G^g = (x_1^g, y_1^g, x_2^g, y_2^g)$ is greater than 0, $G^p = (x_1^p, y_1^p, x_2^p, y_2^p), G^q = (x_1^q, y_1^q, x_2^q, y_2^q)$ and $EIoU = (x_1, y_1, x_2, y_2) \in \mathbb{R}^4, \mathcal{K} \subseteq \mathcal{U}, 0 \leq L_{IoU_i} \leq 1, 0 \leq L_{EIoU} \leq 2$.

In conclusion, the above algorithm retains the primary function of $IoU$, but also maintains the role of $IoU_i$, while making up for the lack of $IoU_i$. Therefore, in all performance indicators of the two-dimensional target detection task, $EIoU$ can be used as an appropriate replacement for $IoU_i$. To improve the performance of vehicle detection and make it possible to improve the accuracy and speed of vehicle detection, and we can easily derive the analytical solution of $EIoU$, and apply it to the metric and loss of target detection.

3.3. Neural network structure

The improved network structure is named VehicleNet, and the VehicleNet network structure uses for training. VehicleNet's network structure uses convolution kernel size $3 \times 3, 1 \times 1$. The enhanced network structure as shown in Figure 4.

Figure 4. VehicleNet neural network structure.

As shown in Fig. 4, DBL stands for VehicleNet_conv2D_BN_Leaky, which is the essential component of VehicleNet. DBL consists of three parts, namely convolution layer, batch normalization layer, and activation function layer. Res_unit is a part of the ResNet network structure, which consists of two DBL and layer jumper connections. Res_n indicates that there are $n$ Res_units, for example, Res1, Res2, Res3, ... Res8, etc. Concat represents tensor stitching, stitching the middle layer of VehicleNet, and the upsampling of the next layer. The reason for using upsampling in the network is that the more profound the network, the better the feature expression effect. Here, upsampling with a step size of 2 is used, and finally, three tests performed, which are downsampling at 32 times, downsampling at 16 times, and eight times. Downsampling to detect, the predicted size of the three feature layers are $y_1 = 13, y_2 = 26, y_3 = 52$, respectively. The in-depth feature extracted through
upsampling and its dimension merged with the feature layer dimension. The purpose is to obtain detailed image features.

4. Experimental results

4.1. Training results

Figure 5 shows the training results on the Vehicle2020 dataset. The abscissa indicates the number of training steps, and the ordinate indicates the value of training loss. (a) is the change in the loss value of $IoU_i$. As the number of training steps increases, the loss value becomes smaller and smaller. (b) is the change in the loss value of $EIoU$. It's roughly the same as the trend of (a), the loss value becomes smaller and smaller. The loss change of the whole training process shown in (c). (d) is the change curve of the learning rate in the entire training process.

4.2. Comparison of vehicle detection algorithm performance

By training the VehicleNet algorithm model and using the Vehicle2020 for vehicle detection, performance results can be obtained. Since $L_{IoU_i} = 1 - IoU_i$ and $L_{EIoU} = 1 - EIoU$, it can be seen from the analysis in Table 3 that only the larger $L_{IoU_i}$ and $L_{EIoU}$ can indicate that $IoU_i$ and $EIoU$ are smaller, the better the vehicle detection performance. The results in Table 4 show that the method is feasible and meets the expectations of this article.
Table 4. Comparison between the evaluation of VehicleNet trained using its own loss (MSE) as well as $\text{IoU}$, $\text{IoU}_i$ and $\text{EIoU}$ losses.

| Evaluation | AP | AP75 |
|------------|----|------|
| MSE        | 0.324 | 0.317 | 0.302 | 0.329 | 0.315 | 0.305 |
| $\mathcal{L}_{\text{IoU}}$ | 0.328 | 0.321 | 0.308 | 0.339 | 0.325 | 0.312 |
| Improv %   | 1.23% | 1.26% | 1.98% | 3.04% | 3.17% | 2.29% |
| $\mathcal{L}_{\text{EIoU}}$ | 0.332 | 0.32 | 0.315 | 0.343 | 0.334 | 0.325 |
| Improv %   | 2.47% | 3.47% | 4.30% | 4.26% | 6.03% | 6.56% |

Detection framework on Vehicle2020. In the experiment, to prove that this paper’s detection framework is superior to other network model frameworks, we selected the previous classic model to compare with the network model of this paper. mAP and FPS are the performance of vehicle detection. As can be seen from Table 5, the results of using VehicleNet detection framework in this paper are significantly higher than other neural network frameworks. Prove that the method in this paper is feasible.

Table 5. Detection frameworks on Vehicle2020. VehicleNet is faster and more accurate than prior detection methods.

| Detection Frameworks | mAP | FPS |
|----------------------|-----|-----|
| Faster R-CNN Res[2]  | 74.1 | 8   |
| YOLOv3 544 × 544[5]  | 81.4 | 45  |
| SNIPER[6]            | 77.6 | 42  |
| CornerNet-Squeeze[7] | 85.6 | 55  |
| VehicleNet 544 × 544 | 90.8 | 52  |

4.3. The comparison results of the accuracy of different types of vehicles.

Table 6. Based on the target detection algorithm model proposed in recent years, by testing different neural network structures and algorithm models, get the comparison results of the accuracy of different types of vehicles.

| Method                  | Truck | Race car | Minibus | Taxi | Bus | Fire eng. | Fam. sed. |
|-------------------------|-------|----------|---------|------|-----|-----------|-----------|
| Faster R-CNN ResNet[2]  | 71.5  | 76.4     | 74.2    | 73.1 | 72.4| 78.7      | 69.8      |
| YOLOv3 544 × 544[5]    | 78.3  | 85.4     | 84.2    | 79.7 | 76.8| 83.9      | 80.2      |
| SNIPER[6]               | 75.4  | 76.5     | 74.3    | 77.1 | 79.9| 80.3      | 75.8      |
| CornerNet-Squeeze[7]    | 85.4  | 87.8     | 85.5    | 86.6 | 86.3| 84.7      | 86.2      |
| VehicleNet 544 × 544    | 91.3  | 90.2     | 93.5    | 92.7 | 93.6| 90.1      | 93.8      |

This article is an experiment on the Vehicle2020 data set. A total of ten models were tested, including Truck, Racing car, Minibus, Taxi, Family sedan, Bus, Fire engine, the number of each model in the test set equal. Use 544×544 as the size test data set of the picture and compare with other neural network models. The results in the table show that the VehicleNet neural network model is superior to other neural network models.
4.4. Visualization of vehicle detection results

![Vehicle detection results](data/bus.jpg)

Fig. 6. Vehicle detection verification results.

4.4. Visualization of vehicle detection results

Fig. 6 shows the verification results of this experiment. Different classes of vehicles used to obtain detection accuracy. The research shows that this experiment has achieved good results.

5. Conclusion

The primary research of this paper is vehicle detection. Using a simple dataset with ten vehicle classes. The IoU algorithm was further optimized to improve the performance of vehicle detection, the improved algorithm was called EIoU, and the neural network structure redesigned. This results show that the detection accuracy of different classes of vehicles has achieved good results. Such as: bus, motorcycle, truck and so on, the recognition power is relatively high, which can reach 100% respectively. For other vehicles, the effect is also very good through recognition detection. This paper only lists the results of three types of vehicles. Compared with other algorithm models, our neural network algorithm model is superior to other neural network models in vehicle detection. The application has played an important role.

References

[1] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C] // CVPR. IEEE, 2014:580-587
[2] 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 and Machine Intelligence, 2017, 39(6):1137-1149.
[3] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016:779-788
[4] J. Redmon and A. Farhadi,” YOLO 9000: Better, Faster, Stronger,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 6517-6525.
[5] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. 2018. arXiv:1804.02767
[6] Singh B, Najibi M, Davis L S. SNIPER: Efficient Multi-Scale Training[J]. NIPS2018:9310-9320
[7] Hei Law, Yun Teng, Olga Russakovksy, Jia Deng. CornerNet-Lite: Efficient Keypoint Based Object Detection.2019 arXiv:1904.08900