损失函数

YOLOv5损失函数包括：

- **classification loss,** 分类损失
- **localization loss,** 定位损失（预测边界框与GT之间的误差）
- **confidence loss,** 置信度损失（框的目标性；objectness of the box）

总的损失函数：

classification loss + localization loss + confidence loss

YOLOv5使用二元交叉熵损失函数计算类别概率和目标置信度得分的损失。

YOLOv5使用 CIOU Loss作为bounding box回归的损失。

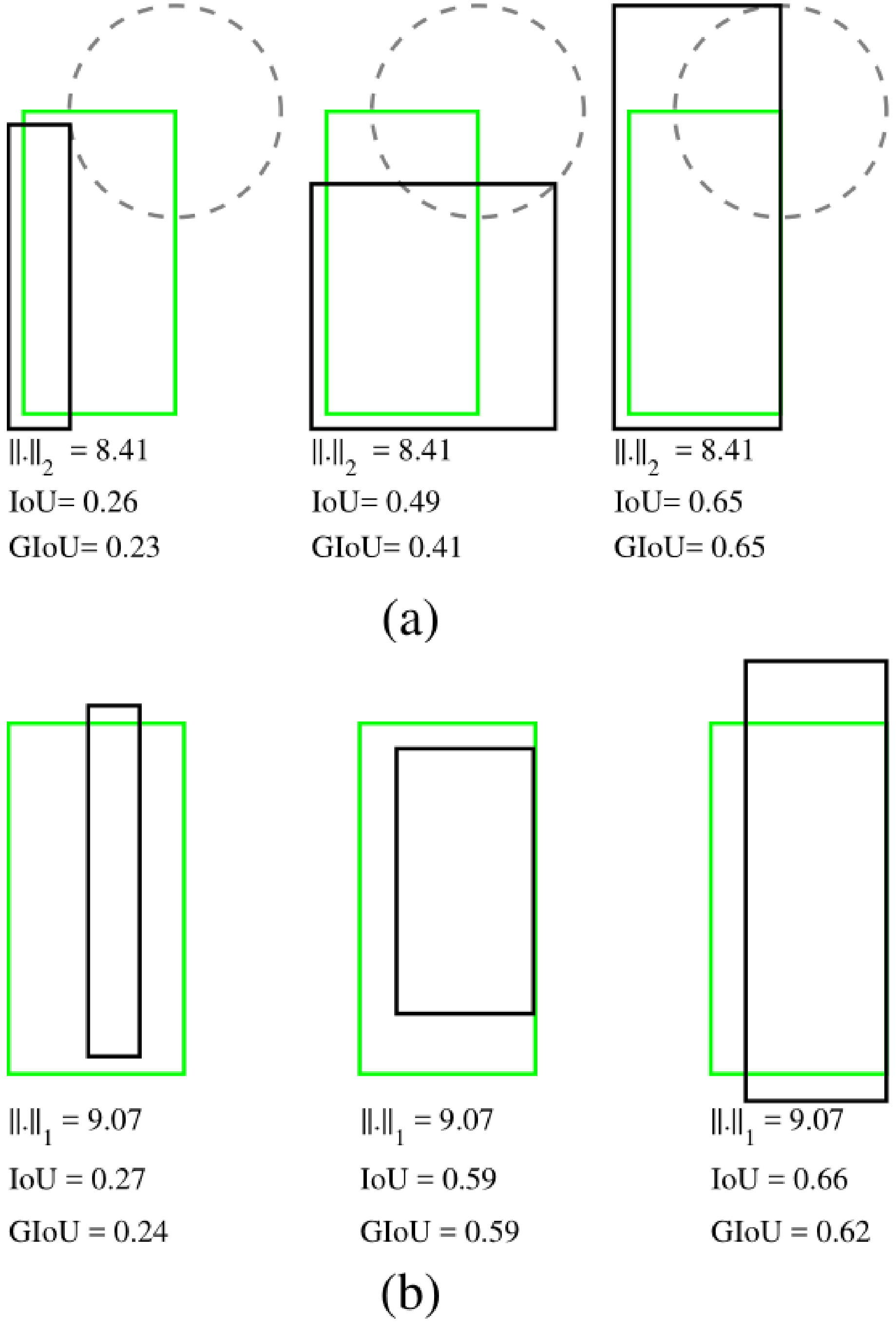
类别预测 (Class Prediction)

- 大多数分类器假设输出标签是互斥的。 如果输出是互斥的目标类别，则确实如此。 因此，YOLO应用softmax函数将得分转换为总和为1的概率。 而YOLOv3/v4/v5使用多标签分类。 例如，输出标签可以是“行人”和“儿童”，它们不是非排他性的。 （现在输出的总和可以大于1）
- YOLOv3/v4/v5用多个独立的逻辑（logistic）分类器替换softmax函数，以计算输入属于特定标签的可能性。
- 在计算分类损失进行训练时，YOLOv3/v4/v5对每个标签使用二元交叉熵损失。 这也避免使用softmax函数而降低了计算复杂度。

- 边界框回归是许多2D / 3D计算机视觉任务中最基本的组件之一。

- 一个改进机会是使用根据IoU计算的度量损失取代替代回归损失（例如l_1和 $l_2-norms$）

Given the choice between optimizing a metric itself vs. a surrogate loss function, the optimal choice is the metric itself.



predicted bounding box (black rectangle)

ground truth box (green rectangle)

$$IoU = \frac{|A \cap B|}{|A \cup B|}$$

IoU用作性能度量和损失函数的问题：

- 如果两个物体不重叠，则IoU值将为零，并且不会反映两个形状彼此之间的距离。
- 在物体不重叠的情况下，如果将IoU用作损失，则其梯度将为零并且无法进行优化。

major weakness:

$$\text{If } |A \cap B| = 0, IoU(A, B) = 0.$$

↓

想法：推广IoU到非重叠情形，并且确保：

（a）遵循与IoU相同的定义，即将比较对象的形状属性编码为区域(region)属性；

（b）维持IoU的尺寸不变性；

（c）在重叠对象的情况下确保与IoU的强相关性。

GIoU (Generalized Intersection over Union)

Algorithm 1: Generalized Intersection over Union

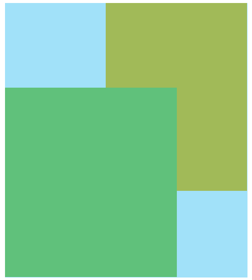
input : Two arbitrary convex shapes: $A, B \subseteq \mathbb{S} \in \mathbb{R}^n$

output: $GIoU$

For A and B, find the smallest enclosing convex object C, where $C \subseteq \mathbb{S} \in \mathbb{R}^n$

$$IoU = \frac{|A \cap B|}{|A \cup B|}$$

$$GIoU = IoU - \frac{|C \backslash (A \cup B)|}{|C|}$$



the smallest convex shapes C enclosing both A and B

Algorithm 2: IoU and $GIoU$ as bounding box losses

input : Predicted B^p and ground truth B^g bounding box coordinates:
$B^p = (x_1^p, y_1^p, x_2^p, y_2^p), \quad B^g = (x_1^g, y_1^g, x_2^g, y_2^g).$

output: $\mathcal{L}_{IoU}, \mathcal{L}_{GIoU}$.

1 For the predicted box B^p, ensuring $x_2^p > x_1^p$ and $y_2^p > y_1^p$:
$\hat{x}_1^p = \min(x_1^p, x_2^p), \quad \hat{x}_2^p = \max(x_1^p, x_2^p),$
$\hat{y}_1^p = \min(y_1^p, y_2^p), \quad \hat{y}_2^p = \max(y_1^p, y_2^p).$

2 Calculating area of B^g: $A^g = (x_2^g - x_1^g) \times (y_2^g - y_1^g)$.

3 Calculating area of B^p: $A^p = (\hat{x}_2^p - \hat{x}_1^p) \times (\hat{y}_2^p - \hat{y}_1^p)$.

4 Calculating intersection $\mathcal{I}$ between B^p and B^g:
$x_1^{\mathcal{I}} = \max(\hat{x}_1^p, x_1^g), \quad x_2^{\mathcal{I}} = \min(\hat{x}_2^p, x_2^g),$
$y_1^{\mathcal{I}} = \max(\hat{y}_1^p, y_1^g), \quad y_2^{\mathcal{I}} = \min(\hat{y}_2^p, y_2^g),$

$$\mathcal{I} = \begin{cases} (x_2^{\mathcal{I}} - x_1^{\mathcal{I}}) \times (y_2^{\mathcal{I}} - y_1^{\mathcal{I}}) & \text{if} \quad x_2^{\mathcal{I}} > x_1^{\mathcal{I}}, y_2^{\mathcal{I}} > y_1^{\mathcal{I}} \\ 0 & \text{otherwise.} \end{cases}$$

5 Finding the coordinate of smallest enclosing box B^c:
$x_1^c = \min(\hat{x}_1^p, x_1^g), \quad x_2^c = \max(\hat{x}_2^p, x_2^g),$
$y_1^c = \min(\hat{y}_1^p, y_1^g), \quad y_2^c = \max(\hat{y}_2^p, y_2^g).$

6 Calculating area of B^c: $A^c = (x_2^c - x_1^c) \times (y_2^c - y_1^c)$.

7 $IoU = \frac{\mathcal{I}}{\mathcal{U}}$, where $\mathcal{U} = A^p + A^g - \mathcal{I}$.

8 $GIoU = IoU - \frac{A^c - \mathcal{U}}{A^c}$.

9 $\mathcal{L}_{IoU} = 1 - IoU, \quad \mathcal{L}_{GIoU} = 1 - GIoU$.

DIoU loss

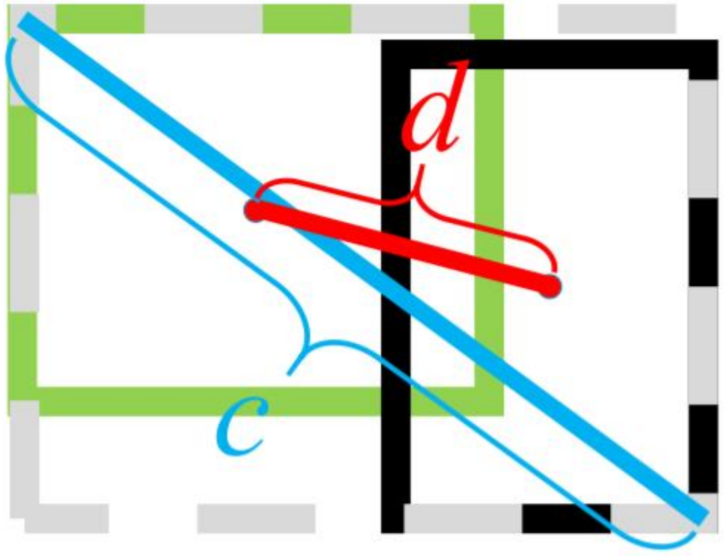
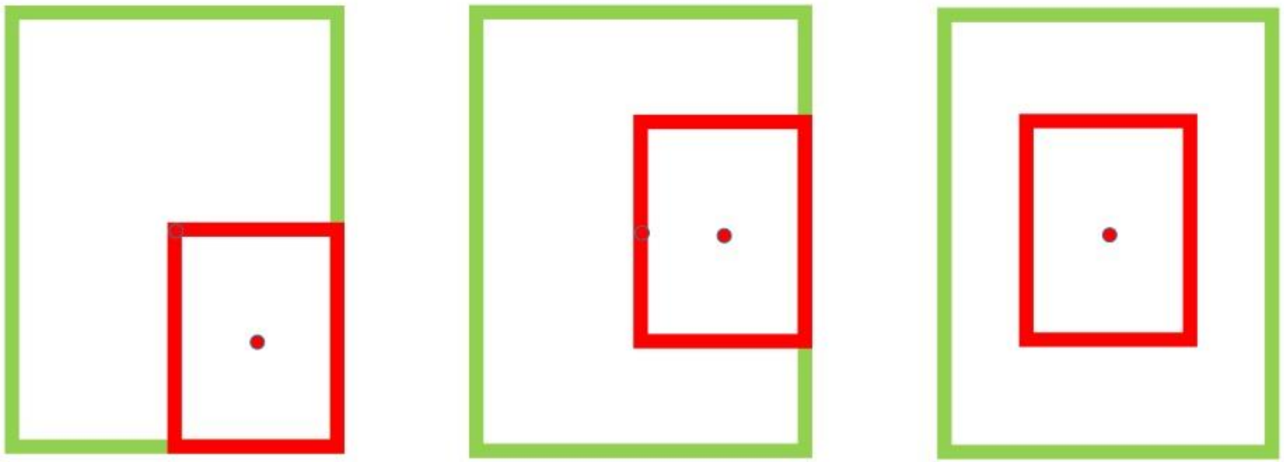


Figure 5: DIoU loss for bounding box regression, where the normalized distance between central points can be directly minimized. c is the diagonal length of the smallest enclosing box covering two boxes, and $d = \rho(\mathbf{b}, \mathbf{b}^{gt})$ is the distance of central points of two boxes.

As shown in Fig. 5, the penalty term of DIoU loss directly minimizes the distance between two central points, while GIoU loss aims to reduce the area of $C - B \cup B^{gt}$.

DIoU loss



$\mathcal{L}_{IoU} = 0.75$ $\mathcal{L}_{IoU} = 0.75$ $\mathcal{L}_{IoU} = 0.75$

$\mathcal{L}_{GIoU} = 0.75$ $\mathcal{L}_{GIoU} = 0.75$ $\mathcal{L}_{GIoU} = 0.75$

$\mathcal{L}_{DIoU} = 0.81$ $\mathcal{L}_{DIoU} = 0.77$ $\mathcal{L}_{DIoU} = 0.75$

Figure 2: GIoU loss degrades to IoU loss for these cases, while our DIoU loss is still distinguishable. Green and red denote target box and predicted box respectively.

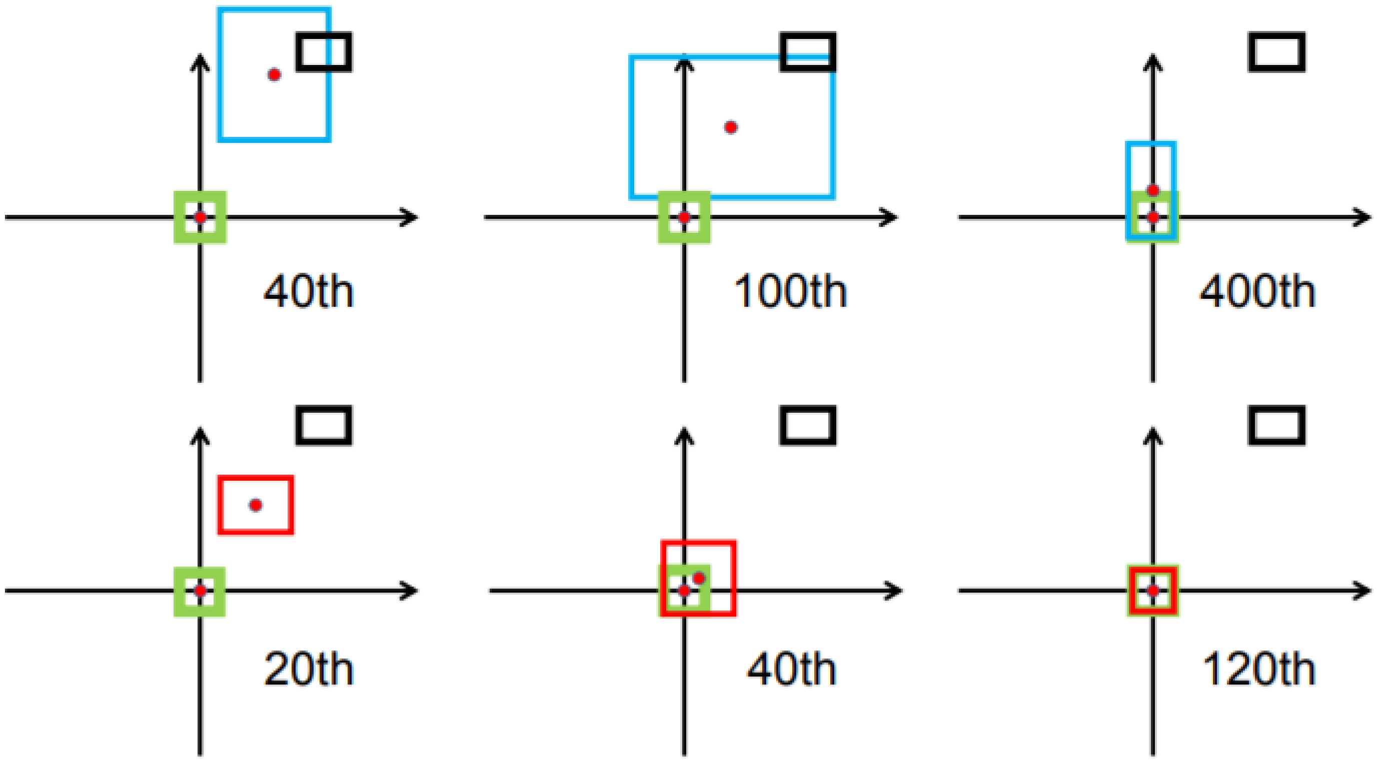


Figure 1: Bounding box regression steps by GIoU loss (first row) and DIoU loss (second row). Green and **black** denote target box and **anchor** box, respectively. Blue and red denote predicted boxes for GIoU loss and DIoU loss, respectively. GIoU loss generally increases the size of predicted box to overlap with target box, while DIoU loss directly minimizes normalized distance of central points.

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|},$$

$$\mathcal{L}_{IoU} = 1 - \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|}.$$

$$\mathcal{L}_{GIoU} = 1 - IoU + \frac{|C - B \cup B^{gt}|}{|C|}$$

Generally, the IoU-based loss can be defined as

$$\mathcal{L} = 1 - IoU + \mathcal{R}(B, B^{gt}),$$

where $\mathcal{R}(B, B^{gt})$ is the penalty term for predicted box B and target box B^{gt}.

Generally, the IoU-based loss can be defined as

$$\mathcal{L} = 1 - IoU + \mathcal{R}(B, B^{gt}),$$

where $\mathcal{R}(B, B^{gt})$ is the penalty term for predicted box B and target box B^{gt}.

$$\mathcal{R}_{DIoU} = \frac{\rho^2(\mathbf{b}, \mathbf{b}^{gt})}{c^2}, \tag{6}$$

where $\mathbf{b}$ and $\mathbf{b}^{gt}$ denote the central points of B and B^{gt}, $\rho(\cdot)$ is the Euclidean distance, and c is the diagonal length of the smallest enclosing box covering the two boxes. And then the DIoU loss function can be defined as

$$\mathcal{L}_{DIoU} = 1 - IoU + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{gt})}{c^2}. \tag{7}$$

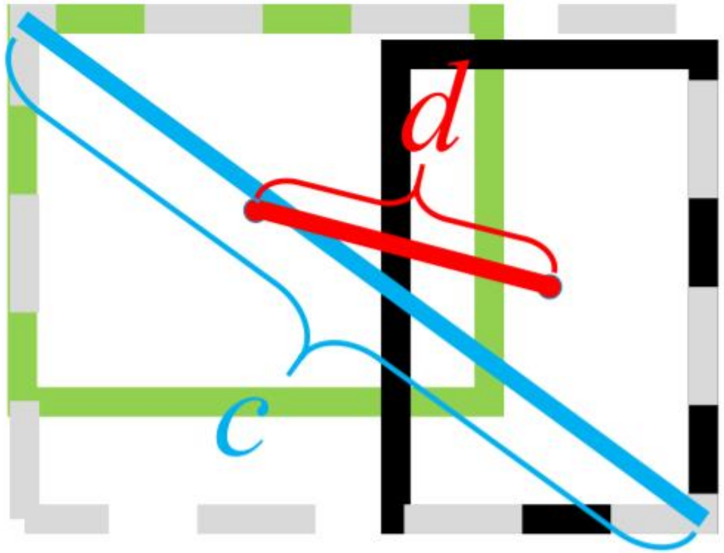


Figure 5: DIoU loss for bounding box regression, where the normalized distance between central points can be directly minimized. c is the diagonal length of the smallest enclosing box covering two boxes, and $d = \rho(\mathbf{b}, \mathbf{b}^{gt})$ is the distance of central points of two boxes.

As shown in Fig. 5, the penalty term of DIoU loss directly minimizes the distance between two central points, while GIoU loss aims to reduce the area of $C - B \cup B^{gt}$.

CIoU（Complete IoU） Loss

Therefore, based on DIoU loss, the CIoU loss is proposed by imposing the consistency of aspect ratio,

$$\mathcal{R}_{CIoU} = \frac{\rho^2(\mathbf{b}, \mathbf{b}^{gt})}{c^2} + \alpha v, \tag{8}$$

where α is a positive trade-off parameter, and v measures the consistency of aspect ratio,

$$v = \frac{4}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h})^2. \tag{9}$$

Then the loss function can be defined as

$$\mathcal{L}_{CIoU} = 1 - IoU + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{gt})}{c^2} + \alpha v. \tag{10}$$

And the trade-off parameter α is defined as

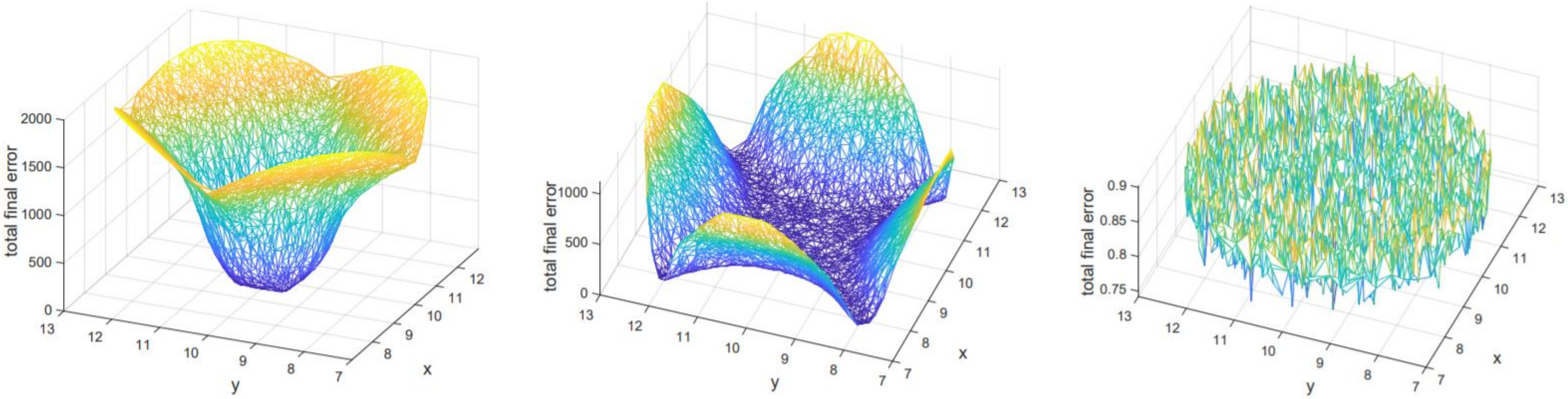
$$\alpha = \frac{v}{(1 - IoU) + v}, \tag{11}$$

by which the overlap area factor is given higher priority for regression, especially for non-overlapping cases.

Finally, the optimization of CIoU loss is same with that of DIoU loss, except that the gradient of v w.r.t. w and h should be specified,

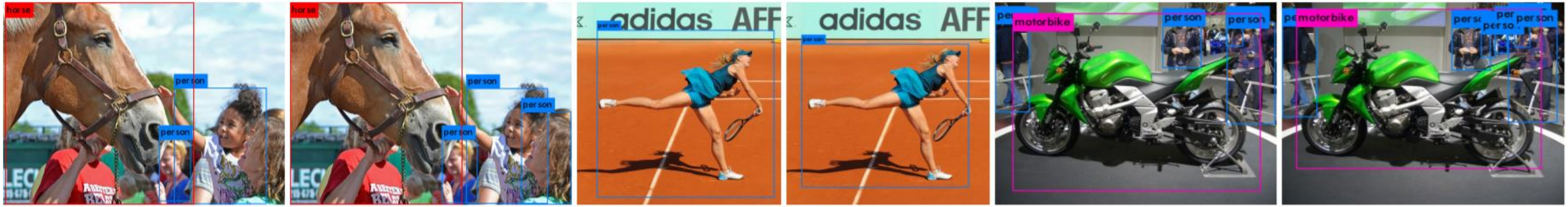
$$\begin{aligned}\frac{\partial v}{\partial w} &= \frac{8}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h}) \times \frac{h}{w^2+h^2},\\ \frac{\partial v}{\partial h} &= -\frac{8}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h}) \times \frac{w}{w^2+h^2}.\end{aligned} \tag{12}$$

The dominator w^2+h^2 is usually a small value for the cases h and w ranging in $[0,1]$, which is likely to yield gradient explosion. And thus in our implementation, the dominator w^2+h^2 is simply removed for stable convergence, by which the step size $\frac{1}{w^2+h^2}$ is replaced by 1 and the gradient direction is still consistent with Eqn. (12).



(a) $\mathcal{L}_{IoU}$ (b) $\mathcal{L}_{GIoU}$ (c) $\mathcal{L}_{DIoU}$

Figure 4: Visualization of regression errors of IoU, GIoU and DIoU losses at the final iteration T, i.e.. $\mathbf{E}(T, n)$ for every coordinate n. We note that the basins in (a) and (b) correspond to good regression cases. One can see that IoU loss has large errors for non-overlapping cases, GIoU loss has large errors for horizontal and vertical cases, and our DIoU loss leads to very small regression errors everywhere.



$\mathcal{L}_{GIoU}$ $\mathcal{L}_{CIoU}$ $\mathcal{L}_{GIoU}$ $\mathcal{L}_{CIoU}$ $\mathcal{L}_{GIoU}$ $\mathcal{L}_{CIoU}$

Figure 6: Detection examples using YOLO v3 (Redmon and Farhadi 2018) trained on PASCAL VOC 07+12.

- A Distance-IoU loss, i.e., DIoU loss, is proposed for bounding box regression, which has faster convergence than IoU and GIoU losses.
- A Complete IoU loss, i.e., CIoU loss, is further proposed by considering three geometric measures, i.e., overlap area, central point distance and aspect ratio, which better describes the regression of rectangular boxes.