

目标检测的损失函数研究进展

杨倩会¹, 张长伦^{1,2}, 何强¹, 王恒友¹

¹北京建筑大学理学院, 北京

²北京建筑大学, 北京未来城市设计高精尖创新中心, 北京

收稿日期: 2021年10月25日; 录用日期: 2021年11月23日; 发布日期: 2021年11月30日

摘要

损失函数是目标检测中的一个研究热点。在深度学习飞速发展的过程中, 目标检测算法取得众多研究成果, 并广泛应用于人脸检测、门禁识别、自动驾驶等领域。为了进一步了解目标检测的发展细节, 本文从目标检测的损失函数优化出发, 梳理概括近年来目标检测中分类损失和回归损失的发展进程, 并对其归类、分析。首先介绍分类损失的发展历程, 然后介绍回归损失的两大发展方向: 基于 L_n 范数与基于交并比(IoU)的损失函数, 并分析这些损失函数的优缺点以及相互之间的关联性。最后, 对基于目标检测的损失函数的未来发展方向进行展望。

关键词

目标检测, 损失函数, 交并比, L_n 范数

Research Progress of Loss Function in Object Detection

Qianhui Yang¹, Changlun Zhang^{1,2}, Qiang He¹, Hengyou Wang¹

¹Science School, Beijing University of Civil Engineering and Architecture, Beijing

²Beijing Advanced Innovation Center for Future Urban Design, Beijing University of Civil Engineering and Architecture, Beijing

Received: Oct. 25th, 2021; accepted: Nov. 23rd, 2021; published: Nov. 30th, 2021

Abstract

Loss function is a research hotspot in object detection. With the rapid development of deep learning, object detection algorithms have achieved many research results, and are widely used in face detection, access control recognition, and automatic driving and so forth. In order to further understand the development details of object detection, this article starts from the optimization of the loss function of object detection, combs and summarizes the development process of classifi-

cation loss and regression loss in object detection in recent years, and classifies and analyzes them. First introduce the development history of classification loss, and then introduce the two major development directions of regression loss: loss function based on L_n norm and based on Intersection over Union (IoU), and analyze the advantages and disadvantages of these loss functions and the correlation between them. Finally, the future development direction of the loss function based on object detection is prospected.

Keywords

Object Detection, Loss Function, Intersection over Union, L_n Norm

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

目标检测是计算机视觉的三大任务之一，同时也是其他视觉任务的基石，因此目标检测的发展对计算机视觉具有重要意义。近年来，随着深度学习的快速发展，目标检测在数据预处理、网络结构、损失函数以及后处理等方面取得了丰硕的成果。

损失函数用于指导网络的训练，决定网络的优化方向，因此在目标检测算法中具有重要的地位。目标检测的损失函数分为分类损失和回归损失，其中分类损失用于指导网络对目标预测正确的类别，回归损失用于指导网络生成定位准确的回归框。在检测网络如 R-CNN 系列[1]-[7]、Yolo 系列[8] [9] [10] [11]、SSD 系列[12] [13]、以及 Transformer 系列[14]中，常用的分类损失有 Cross Entropy Loss [15]、Focal Loss [16] 等，常用的回归损失有 L_1 Loss [7]、 L_2 Loss [1]、Smooth L_1 Loss [2]、Balanced L_1 Loss [6]、IoU 系列损失函数[17] [18] [19] [20]等。

本文将分别从分类损失与回归损失介绍目标检测损失函数的研究进展。

2. 分类损失的研究现状

2.1. Cross Entropy Loss

网络预测的分类结果并不是 0~1 之间的概率值，因此在计算样本分类损失之前，需要通过 Softmax 或 Sigmoid 函数得到样本属于各个类别的概率，此处以 Softmax 为例进行介绍。

令向量 V 表示一个样本预测的多个类别的预测结果， V_i 表示该样本属于第 i 类的预测结果，为了将预测结果映射到 0~1 之间保证概率的非负性与归一化，同时保证样本属于所有类别的概率之和为 1。通过 Softmax 对预测结果进行映射，则样本属于第 i 类的概率为：

$$p = S_i = \frac{e^{V_i}}{\sum_j e^{V_j}}$$

交叉熵是用来衡量预测结果与标签概率分布的参数的差异，交叉熵越大，说明预测结果越差，此时交叉熵损失对网络的惩罚也就越大。以二分类任务为例，类别 y 的取值为 $\{+1, -1\}$ ，其中 +1 代表正类，-1 代表负类。假设 p 是该样本属于正类的概率，则交叉熵损失函数如公式(1)所示：

$$L_{CE} = -\log(p_t) \quad (1)$$

其中：

$$p_i = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{otherwise} \end{cases}$$

2.2. Focal Loss

一阶段算法中，难易样本分类难度差异与正负样本数量不平衡是导致分类效果差的原因：难分类样本与正样本对损失函数贡献少且所占比例低。因此为了提高模型的精度，平衡一阶段算法中难易样本以及正负样本对损失函数的贡献，Focal Loss 在交叉熵损失函数的基础上进行改进：通过引入调制系数 $(1 - p_i)^\gamma$ ，赋予难分类样本较大的权重，易分类样本较小的权重；通过引入权重系数 $\alpha \in [0, 1]$ 增加正样本的权重，减小负样本的权重，具体计算如公式(2)所示：

$$L_{FL} = -\alpha_i (1 - p_i)^\gamma \log(p_i) \tag{2}$$

其中 α_i 的定义与 p_i 类似：

$$\alpha_i = \begin{cases} \alpha, & \text{if } y = 1 \\ 1 - \alpha, & \text{otherwise} \end{cases}$$

其中 $\alpha = 0.25$ ， $\gamma = 2$ 。

3. 回归损失研究现状

3.1. L_n Loss

L_n Loss 是计算候选框与真实框相应参数的 L_n 范数。真实框的坐标为 (G_x, G_y, G_w, G_h) ，候选框的坐标为 (P_x, P_y, P_w, P_h) ，网络要学习一个线性映射 φ ，使得对候选框进行映射之后的坐标与真实框的坐标无限接近，即 $\varphi(P_x, P_y, P_w, P_h) = (\widehat{G}_x, \widehat{G}_y, \widehat{G}_w, \widehat{G}_h) \approx (G_x, G_y, G_w, G_h)$ 。对于线性映射的训练，优化项为：

$$\begin{array}{ll} t_x = (P_x - \widehat{G}_x) / \widehat{G}_w & t_x^* = (G_x - \widehat{G}_x) / \widehat{G}_w \\ t_y = (P_y - \widehat{G}_y) / \widehat{G}_h & t_y^* = (G_y - \widehat{G}_y) / \widehat{G}_h \\ t_w = \log(P_w / \widehat{G}_w) & t_w^* = \log(G_w / \widehat{G}_w) \\ t_h = \log(P_h / \widehat{G}_h) & t_h^* = \log(G_h / \widehat{G}_h) \end{array}$$

则此时对应的 L_n Loss 为 $L_n = \sqrt[n]{\sum (t)^n}$ ，其中 $t = t_i - t_i^*$ ， $i \in \{x, y, w, h\}$ 。

3.1.1. L_1 Loss

L_1 Loss 又被称为最小绝对误差函数，可以用来度量两个向量之间的差异，损失函数形式如公式(3)所示。

$$L_1(t) = \sum |t| \tag{3}$$

$$\frac{dL_1(t)}{dt} = \begin{cases} 1, & t \geq 0 \\ -1, & t < 0 \end{cases} \tag{4}$$

L_1 Loss 图像如图 1 蓝色曲线所示，其导数形式如公式(4)所示。由于其梯度是常数，因此在训练过程中可以保持稳定的梯度使得模型得以收敛。也正因为梯度为常数，对于收敛之后的模型，较小的误差也采用同样的梯度，因此在极值点附近容易发生震荡，导致模型难以达到更高的精度。

3.1.2. L_2 Loss

L_2 Loss 又被称为均方误差函数，也是用来度量两个向量之间的差异，与 L_1 Loss 略有不同，损失函数形式如公式(5)所示。

$$L_2(t) = \sum(t)^2 \quad (5)$$

$$\frac{dL_2(t)}{dt} = 2t \quad (6)$$

L_2 范数函数图像如图 1 橙色曲线所示，其导数形式如公式(6)所示。函数连续且处处可导，并且随着误差值的减小，梯度也减小，有利于收敛到最小值，但同时训练初期误差值较大，梯度也大，容易产生梯度爆炸导致模型无法收敛，因此使用 L_2 范数作为损失函数时，往往使用较小的学习率从而避免训练初期梯度爆炸。

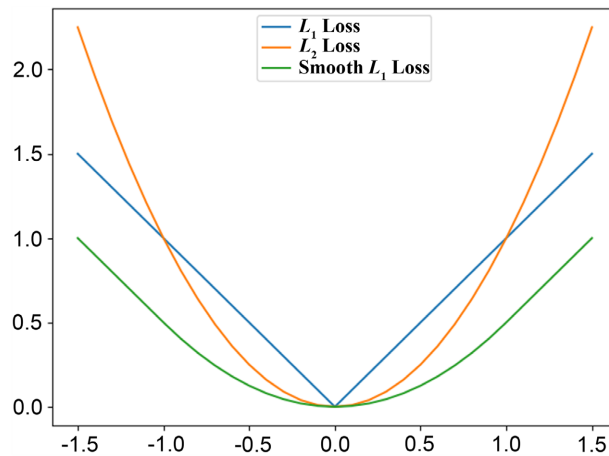


Figure 1. L_n Loss function curve

图 1. L_n Loss 函数曲线

3.1.3. Smooth L_1 Loss

Smooth L_1 Loss 函数形式如公式(7)所示：

$$\text{Smooth } L_1(t) = \begin{cases} 0.5t^2 & |t| < 1 \\ |t| - 0.5 & \text{otherwise} \end{cases} \quad (7)$$

$$\frac{d\text{Smooth } L_1(t)}{dt} = \begin{cases} t & |t| < 1 \\ \pm 1 & \text{otherwise} \end{cases} \quad (8)$$

Smooth L_1 Loss 函数图像如图 1 绿色曲线所示，其导数形式如公式(8)所示。由求导过程可知，Smooth L_1 Loss 在损失较大时，按照一个恒定的速率梯度下降，损失较小时，不再按照一个恒定的梯度下降，而是按照变量自身进行动态调整。Smooth L_1 Loss 结合了 L_1 Loss 与 L_2 Loss 的优点：模型开始训练时，损失较大，采用 L_1 Loss 的形式，使网络保持稳健的梯度；模型收敛时，当损失较小时采用 L_2 Loss 的形式，使模型可以收敛到更高的精度。

3.1.4. Balanced L_1 Loss

受 Smooth L_1 Loss 的启发，为了平衡难易样本对梯度的影响，Balanced L_1 Loss 在 Smooth L_1 Loss 的梯度公式上进行改进，设计 L_b 的梯度如公式(9)所示：

$$\frac{dL_b(t)}{dt} = \begin{cases} \alpha \ln(b|t|+1) & |t| < 1 \\ \gamma & \text{otherwise} \end{cases} \quad (9)$$

其中， α 控制简单样本的梯度， γ 控制调整回归误差的上界，能够使不同任务间更加平衡。 α ， γ 分别从样本和任务层面控制平衡，使网络得到更加平衡的训练。

对梯度公式(9)进行积分，得到 L_b 的表达式如公式(10)所示：

$$L_b(t) = \begin{cases} \frac{\alpha}{b}(b|t|+1)\ln(b|t|+1) - \alpha|t| & \text{if } |t| < 1 \\ \gamma|t| + C & \text{otherwise} \end{cases} \quad (10)$$

其中 $\alpha = 0.5$ ， $\gamma = 1.5$ ， $\alpha \ln(b+1) = \gamma$ 。

3.2. IoU-Based Loss

由于 L_n Loss 是将候选框的中心点坐标以及长宽四个参数视为互不相关的变量分别进行优化，但实际上这些变量是相关的，因此会导致回归框定位不准确；此外， L_n Loss 对尺度敏感，无法做到尺度归一化，并且 L_n Loss 并不能很好地反映候选框的回归情况。由此开启了 IoU (Intersection over Union)-based Loss 的研究，IoU 是评价候选框回归情况质量的度量函数，优化度量函数本身要比优化代理损失函数更加直观，并且 IoU 将候选框作为整体进行优化，同时具有尺度不变性，避开了 L_n Loss 的缺点。

3.2.1. IoU Loss

在目标检测中，IoU 是两个框的交并比，定义如公式(11)所示：

$$\text{IoU}(P, G) = \frac{P \cap G}{P \cup G} \quad (11)$$

其中 P 、 G 分别表示预测框与真实框，IoU 的直观表示如图 2 所示：

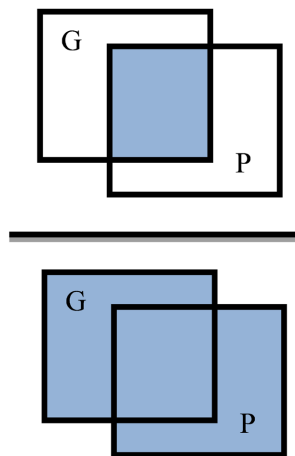


Figure 2. IoU schematic diagram
图 2. IoU 示意图

IoU Loss 的损失函数如公式(12)所示，易知 IoU 的取值范围是[0, 1]，IoU 的值越接近 1，说明候选框与真实框越接近，定位越准确，此时损失函数值越小，惩罚也就越小，反之惩罚越大。但同时也可看出 IoU Loss 的缺点是当候选框与真实框不相交时，IoU 为 0， L_{IoU} 为常数 1，梯度无法回传。

$$L_{\text{IoU}} = 1 - \text{IoU}(P, G) \quad (12)$$

3.2.2. GIoU Loss

由上述分析可知, IoU Loss 只关注重叠部分, 无法回归候选框与真实框不相交的情况。为了克服这一缺点, GIoU Loss 在 IoU Loss 的基础上引入惩罚项 $R = \frac{|C - P \cup G|}{|C|}$, 其中 C 是两框的最小闭包, GIoU Loss 由公式(13)定义:

$$L_{\text{GIoU}} = 1 - \text{IoU} + \frac{|C - P \cup G|}{|C|} \quad (13)$$

惩罚项 R 如图 3 所示, 左图是两框相交的情况, 右图是两框不相交的情况, 可以直观看出 GIoU Loss 不仅关注两框重叠部分, 还关注非重叠部分, 对于非重叠部分大的样本加以更大的权重。

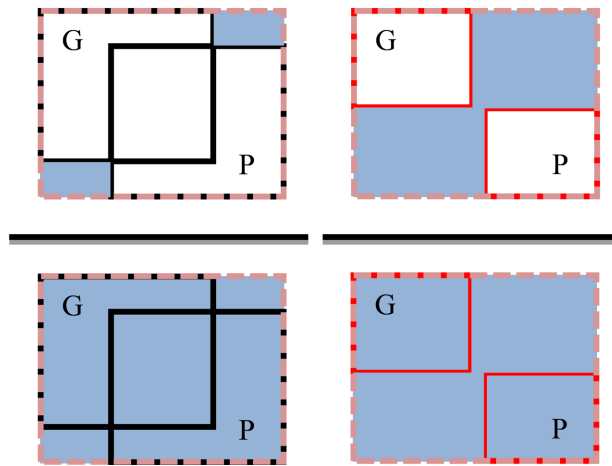


Figure 3. GIoU schematic diagram of penalty items
图 3. GIoU 惩罚项示意图

3.2.3. DIoU Loss

虽然 GIoU Loss 解决了两框不相交无法回归的情况, 但是当两框处于包含关系时, GIoU Loss 退化为 IoU Loss, 无法反映不同位置下回归情况的好坏, 如图 4 所示, 被包含的框处于不同的位置时, 得到的 GIoU Loss 与 IoU Loss 是一样的, 无法进一步反映回归情况。且 GIoU Loss 在优化过程中, 先减小两框之间的闭包, 再增大两框之间的 IoU, 收敛过程较慢。

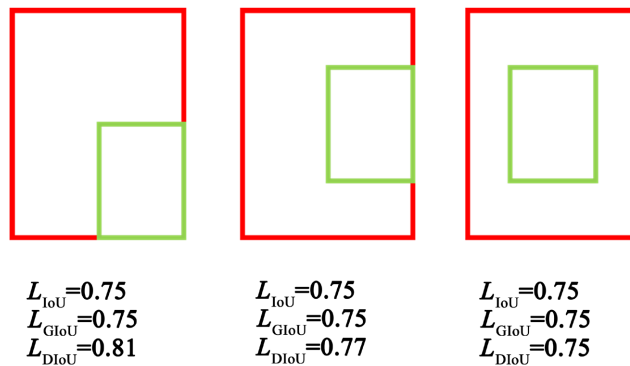


Figure 4. The instances of GIoU unable to regression
图 4. GIoU 无法回归示例

为了解决上述问题，DIoU Loss 提出将 GIoU Loss 中的惩罚项改为 $R = \frac{\rho^2}{c^2}$ ，如图 5 所示：

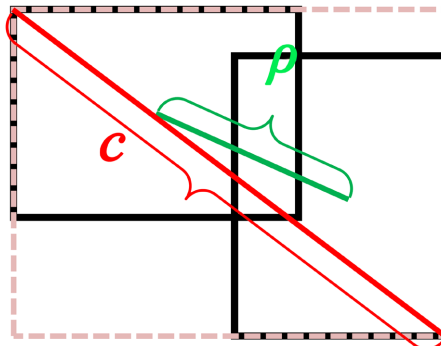


Figure 5. DIoU schematic diagram of penalty items
图 5. DIoU 惩罚项示意图

其中 ρ^2 表示两个矩形框的中心点的欧氏距离， c^2 表示两个矩形框的最小闭包的对角线平方。DIoU Loss 由公式(14)定义：

$$L_{DIoU} = 1 - IoU + \frac{\rho^2}{c^2} \tag{14}$$

由图 3 可知，DIoU 能够反映两框之间的位置关系，并且 DIoU loss 可以直接拉进两个目标框的距离，同时增大两框之间的 IoU，加快收敛速度。

3.2.4. CIoU Loss

CIoU Loss 认为候选框与真实框进行匹配的几何要素包括面积、中心点距离、长宽比例，这三部分决定了候选框的定位精确度。DIoU Loss 已经包含了面积与中心点距离，因此 CIoU Loss 将两框的长宽比也作为惩罚项加入损失函数中，如公式(15)所示：

$$L_{CIoU} = 1 - IoU + \frac{\rho^2}{c^2} + \alpha v \tag{15}$$

其中权衡参数 $\alpha = \frac{v}{1 - IoU + v}$ 用于控制长宽比的系数，而 $v = \frac{4}{\pi^2} \left(\tan^{-1} \frac{w^G}{h^G} - \tan^{-1} \frac{w^P}{h^P} \right)^2$ 则用于描述两框的比例一致性。

3.2.5. Focal EIoU Loss

当两个矩形框的长度和宽度呈倍数关系时，长宽比相等但大小不等，此时 CIoU Loss 中 αv 失效。并且通过求导发现 $\frac{\partial v}{\partial w} = -\frac{h}{w} \cdot \frac{\partial v}{\partial h}$ ，宽和高的梯度方向相反，即当宽和高其中一个变量增大时，另一个变量必然减小，不能做到同增同减。因此设计 EIoU Loss，将长和宽分别进行优化，函数定义如公式(16)所示：

$$L_{EIoU} = 1 - IoU + \frac{\rho^2}{c^2} + \frac{\rho^2(w, w^{gt})}{C_w^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \tag{16}$$

其中 C_w^2 和 C_h^2 分别为两个矩形框最小闭包的宽和高。

分类任务提高难分类样本有助于模型收敛，回归任务则是提高易分类样本有助于模型收敛。因此为了提高候选框回归的准确度，EIoU Loss 借鉴 Focal Loss 的思想，提高易分类样本在损失函数中所占比重，

提出 Focal EIoU Loss, 函数定义如公式(17)所示:

$$L_{\text{FocalEIoU}} = \text{IoU}^\gamma L_{\text{EIoU}} \quad (17)$$

4. 总结与展望

本文对目标检测中的损失函数从分类损失与回归损失两方面进行了详尽的介绍, 并分析了各个损失函数的优缺点。尽管目标检测已经得到了快速发展发展地较为成熟, 但是仍旧不能满足现实中的应用需求, 因此仍需从各个模块进行改进。其中损失函数的优化是必不可少的, 现有 IoU 损失函数仍有可优化的内容。此外, 在目标遮挡问题、小目标检测等领域中, 损失函数亟需更大的优化。

基金项目

国家自然科学基金(No. 62072024); 北京建筑大学北京未来城市设计高精尖创新中心资助项目(UDC2017033322, UDC2019033324); 北京建筑大学市属高校基本科研业务费专项资金资助(NO. X20084, ZF17061)。

参考文献

- [1] Girshick, R., Donahue, J., Darrell, T., *et al.* (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. <https://doi.org/10.1109/CVPR.2014.81>
- [2] Girshick, R. (2015) Fast R-CNN. *IEEE International Conference on Computer Vision (ICCV)*, Santiago, 7-13 December 2015, 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>
- [3] Ren, S., He, K., Girshick, R., *et al.* (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- [4] Cai, Z.W. and Vasconcelos, N. (2017) Cascade R-CNN: Delving into High Quality Object Detection. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 6154-6162. <https://doi.org/10.1109/CVPR.2018.00644>
- [5] He, K.M., Georgia, G., Piotr, D., *et al.* (2018) Mask R-CNN. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **42**, 386-397. <https://doi.org/10.1109/TPAMI.2018.2844175>
- [6] Pang, J.M., Chen, K., Shi, J.P., *et al.* (2019) Libra R-CNN: Towards Balanced Learning for Object Detection. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 821-830. <https://doi.org/10.1109/CVPR.2019.00091>
- [7] Sun, P., Zhang, R.F., Jiang, Y., *et al.* (2021) Sparse R-CNN: End-to-End Object Detection with Learnable Proposals. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 14449-14458. <https://doi.org/10.1109/CVPR46437.2021.01422>
- [8] Redmon, J., Divvala, S., Girshick, R., *et al.* (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [9] Redmon, J. and Farhadi, A. (2016) YOLO9000: Better, Faster, Stronger. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 6517-6525. <https://doi.org/10.1109/CVPR.2017.690>
- [10] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement. <https://arxiv.org/abs/1804.02767>
- [11] Bochkovskiy, A., *et al.* (2020) YOLOv4: Optimal Speed and Accuracy of Object Detection. <https://arxiv.org/abs/2004.10934>
- [12] Liu, W., Anguelov, D., Erhan, D., *et al.* (2016) SSD: Single Shot Multi-Box Detector. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
- [13] Fu, C.Y., Liu, W., Ranga, A., *et al.* (2017) DSSD: Deconvolutional Single Shot Detector. <https://arxiv.org/abs/1701.06659>
- [14] Carion, N., Massa, F., Synnaeve, G., *et al.* End-to-End Object Detection with Transformers. <https://arxiv.org/abs/2005.12872>

- [15] 鲁晨光. Shannon 公式改造[J]. 通信学报, 1991, 12(2): 95-96.
- [16] Lin, T.Y., Goyal, P., Girshick, R., He, K.M., *et al.* (2018) Focal Loss for Dense Object Detection. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 2999-3007. <https://doi.org/10.1109/ICCV.2017.324>
- [17] Yu, J.H., Jiang, Y., Wang, Z.Y., *et al.* (2016) Unitbox: An Advanced Object Detection Network. *Proceedings of the 24th ACM International Conference on Multimedia*, Amsterdam, 15-19 October 2016, 516-520. <https://doi.org/10.1145/2964284.2967274>
- [18] Rezatofighi, H., Tsoi, N., Gwak, J.Y., *et al.* (2019) Generalized Intersection over Union: A Metric and a Loss for Bounding Box Regression. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 658-666. <https://doi.org/10.1109/CVPR.2019.00075>
- [19] Zheng, Z., Wang, P., Liu, W., *et al.* (2020) Distance-IOU Loss: Faster and Better Learning for Bounding Box Regression. *The Association for the Advance of Artificial Intelligence (AAAI)*, New York, 7-12 February 2020, 12993-13000.
- [20] Zhang, Y.F., Ren, W.Q., Zhang, Z., *et al.* (2021) Focal and Efficient IOU Loss for Accurate Bounding Box Regression. <https://arxiv.org/abs/2101.08158>