M2IoU: A Min-Max Distance-based Loss Function for Bounding Box Regression in Medical Imaging

Anurag Kumar Shandilya*
Kalash Shah*
anurag.k.shandilya@gmail.com
kalashsshah@gmail.com
IIT Bombay
Mumbai, India

Bhavik Kanekar
IIT Bombay
Mumbai, India
kanekar895@gmail.com

Akshat Gautam
IIT Bombay
Mumbai, India

Pavni Tandon
IIT Bombay
Mumbai, India
iampavnitandon@gmail.com

Ganesh Ramakrishnan
IIT Bombay
Mumbai, India
ganesh@cse.iitb.ac.in

Kshitij Jadhav
IIT Bombay
Mumbai, India
kshitij.jadhav@iitb.ac.in

Abstract

Computer vision applications such as object detection have increased manifolds in the medical domain for diagnosis and treatment purposes. Generally, object detection models such as YOLO (You Only Look Once) involve identifying the correct bounding box and classifying the objects inside the bounding box. However, medical imaging object detection is a challenging endeavor, requiring models that are both efficient and extremely accurate in the face of limited data and expensive annotations. In this paper, we propose **Min-Max IoU** (M2IoU) loss function by introducing a new min-max-based penalty term in the loss equation, between the predicted box and the ground truth coordinates. We further compare the results of several loss functions on the YOLOv8 model trained on multiple medical datasets and demonstrate that the M2IoU loss function leads to faster learning and outperforms other existing loss functions like CIoU and GIoU.

CCS Concepts

• **Computing methodologies** → **Object detection**.

Keywords

Object detection, loss function, IoU, bounding box regression

ACM Reference Format:
Anurag Kumar Shandilya, Kalash Shah, Bhavik Kanekar, Akshat Gautam, Pavni Tandon, Ganesh Ramakrishnan, and Kshitij Jadhav. 2024. M2IoU: A Min-Max Distance-based Loss Function for Bounding Box Regression in Medical Imaging. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24), October 21–25, 2024, Boise, ID, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3627673.3679958

*Both authors contributed equally to this research.

CIKM '24, October 21–25, 2024, Boise, ID, USA

ACM ISBN 979-8-4007-0436-9/24/10
https://doi.org/10.1145/3627673.3679958

1 Introduction

Medical imaging forms the foundation of contemporary diagnostic medicine, facilitating the early identification and precise diagnosis of various health conditions [8]. Medical imaging differs from typical object detection tasks by posing distinct challenges, including significant variability in images, subtle feature distinctions, and the critical implications of misdiagnosis. These factors demand not only greater accuracy but also enhanced robustness in models specifically crafted for this application [7]. Standard object detection algorithms frequently underperform in medical settings, where the consequences of false negatives can be particularly severe. This highlights the necessity for creating specialized methods that are specifically designed to meet the intricate demands of medical imaging [4]. The scarcity of annotated medical images, driven by privacy issues and the expensive nature of expert annotations, adds complexity to the creation of effective diagnostic tools. This situation emphasizes the need for efficient and adaptable learning algorithms in this domain [16]. With the growing complexity of medical datasets, the accuracy of object detection models is vital for successful diagnosis and effective treatment planning.

Object Detection models utilize a Bounding Box Regression (BBR) module to obtain a precise position of the object of interest. In terms of evaluation for the bounding box regression, Intersection-Over-Union (IoU) is the most popular metric. IoU is calculated as the ratio of the intersection of the predicted bounding box and the ground truth bounding box to the union of the two bounding boxes. Often, the performance of a fully-trained model depends on the loss function of the BBR module[2]. These loss functions aim to bring the predicted bounding box and the ground truth as close to each other as possible and maximize their overlap.

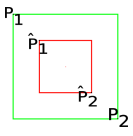
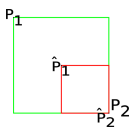
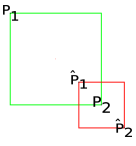
2 Our Contribution

The anchor points of a bounding box are the two points that identify a box uniquely. They could be the two diagonally opposite corners, or a corner and the center of that bounding box. Current loss functions like Complete Intersection-over-Union (CIoU)[13] do

not differentiate between the relative position of the two predicted anchor points and their corresponding ground truth. To accelerate the learning process, we introduce our min-max distance-based loss function (M2IoU loss) (*c.f.* Sec. 4). Here, we treat both anchor points separately by imposing a higher penalty on the predicted anchor point that is farther away from its corresponding ground truth by using a hyperparameter α. This leads to faster learning of the model when trained using our loss function, which ultimately performs better than the other existing loss functions on medical datasets. Table 1 shows loss values are consistently higher for the M2IoU loss function for the different misalignments of the ground truth and predicted bounding boxes.

The contribution of this paper can be summarized as follows:

Table 1: Illustration of loss values for various loss functions. Here Ground truth is the green box and Prediction is the red box

Image Case			
$\mathcal{L}_{IoU}$	0.75	0.75	0.95
$\mathcal{L}_{GIoU}$	0.75	0.75	1.19
$\mathcal{L}_{DIoU}$	0.75	0.81	1.1
$\mathcal{L}_{CIoU}$	0.75	0.81	1.1
$\mathcal{L}_{M2IoU}$	0.81	0.94	1.23

(1) We propose a new loss function M2IoU (*c.f.*section 4 loss), by introducing a novel hyper-parameter-controlled min-max penalty term. The primary idea is to assign a higher weight to the point that is farther away from the corresponding ground truth to enable quicker learning.
(2) The new penalty term yields better results than existing loss functions and beats the CIoU loss function with lesser computations.

3 Related Work

Bounding box regression (BBR) traditionally employed l_n norm loss which did not synergize with the IoU metric and is sensitive to scaling issues [15]. To incorporate the distance invariance issue in the IoU loss function, the generalized intersection over union (GIoU) loss function was proposed which considers three factors namely overlap area, union area, and enclosing area [11]. However, it does not take into account object orientation, which can be a significant limitation in detecting objects with varied orientations [10]. To overcome these drawbacks Distance IoU (DIoU) was proposed and used in YOLOv4 and YOLOv5 [13]. DIoU considers overlap area, central point distance, and normalized distance between the predicted and ground truth bounding box. However, it is sensitive to the aspect ratio, size, and location of the object [14]. Simultaneously in the same paper as an improvement, CIoU[13] was proposed with consideration of both, normalized distance and the aspect ratio. However, CIoU, models aspect ratio in relative terms and not absolute value. Hence it loses its effectiveness when the predicted bounding box and the ground truth have the same aspect ratio with different width and height values, which limits the convergence speed and accuracy [12]. Similarly, [5]introduced a new family of power IoU losses that have a power IoU term and an additional power regularization term with a single power parameter α.

pagestylefancy fancyhead

4 M2IoU Loss Function

4.1 Formulation

Let the ground truth bounding box B^{gt} be represented by (P_1, P_2) where P_1 and P_2 represent the two anchor points that can uniquely identify the box. Similarly, let $(\hat{P}_1, \hat{P}_2)$ be the prediction B given by the model. Let $\mathcal{D}(P, Q)$ denote the Euclidean distance between the two points P and Q. We define D^2_{min} and D^2_{max} in (1a) and (1b).

$$D^2_{min} = min(\mathcal{D}^2(P_1, \hat{P}_1), \mathcal{D}^2(P_2, \hat{P}_2)) \tag{1a}$$

$$D^2_{max} = max(\mathcal{D}^2(P_1, \hat{P}_1), \mathcal{D}^2(P_2, \hat{P}_2)) \tag{1b}$$

Let α be a hyper-parameter such that $\alpha \in [0, 1]$. The M2IoU loss is described in the (2)

$$\mathcal{L}_{M2IoU} = 1 - IoU + \frac{\alpha D^2_{min} + (1-\alpha) D^2_{max}}{C^2} \tag{2}$$

C is the diagonal length of the smallest enclosing box covering the two boxes (or the diagonal length of the convex hull of the two boxes). When the two boxes are almost aligned, $IoU \approx 1$, and $D^2_{min}, D^2_{max} \approx 0$. Thus, $\mathcal{L}_{M2IoU} >= 0$. When the two boxes are on the opposite corners of the image (farthest away and $IoU = 0$), $D^2_{min}, D^2_{max} \approx C^2$, and $\mathcal{L}_{M2IoU} <= 2$. Thus, the M2IoU loss function is both upper and lower-bounded:

$$0 <= \mathcal{L}_{M2IoU} <= 2 \tag{3}$$

4.2 Convergence Simulation Experiment

We perform an experiment outlined in Algorithm **??** to track the behavior and convergence in case of different loss functions. We denote $\nabla\mathcal{L}(B, B^{gt})$ as the derivative of the loss function with the predicted coordinates. We represent $\mathcal{L}(B, B^{gt})$ as $\mathcal{L}$ for brevity and, (x_1, y_1) and (x_2, y_2) represent the 2-D coordinates of the two anchor points respectively. Thus, in simulation, $\nabla\mathcal{L}$ is calculated as shown in eq. 4.

$$\nabla\mathcal{L} = [\frac{\partial\mathcal{L}}{\partial x_1}, \frac{\partial\mathcal{L}}{\partial y_1}, \frac{\partial\mathcal{L}}{\partial x_2}, \frac{\partial\mathcal{L}}{\partial y_2}] \tag{4}$$

Using Gradient Descent, prediction *pred* is updated with an adaptive learning rate η (to prevent significant overshoot). The IoU value (with the ground truth) is plotted at each iteration for each loss function. Fig. 1 demonstrates that M2IoU loss converges the fastest and at a higher IoU value amongst all other loss functions.

4.3 Experiment for α Value

Hyper-parameter α varies from $[0, 1]$. To impose a higher penalty on the coordinate that is farther away from the corresponding ground truth, we take $\alpha < 0.5$ such that the coefficient of D_{max}, $1 - \alpha$ is more than the coefficient of D_{min}, α. We choose the BCCD dataset[3] with the same configuration as mentioned in Table 2 and train the YOLOv8-x [6] model from scratch for three distinct values of α at 0.05, 0.25, and 0.45 respectively. The training performance

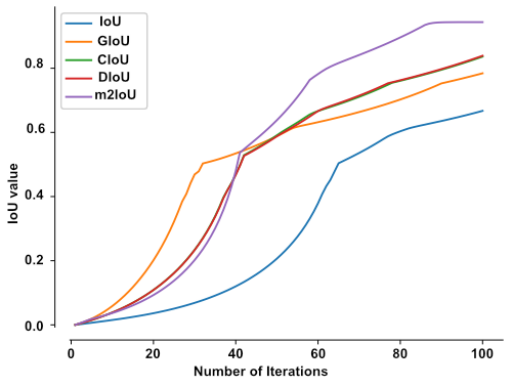


Figure 1: IoU vs Iterations

is monitored using mean Average Precision-50, mAP50 (IoU at threshold 50%). The result of $\alpha = 0.25$ shows faster convergence and best performances as shown in Fig 2. Based on these results we fix the value of α as 0.25 as it produces the best results.

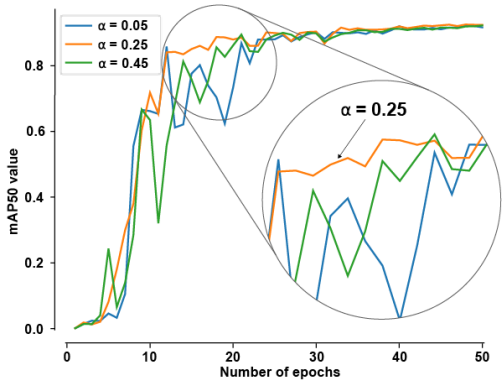


Figure 2: mAP50 metric for α simulations

5 Experimental Results

5.1 Datasets

We used three different medical datasets as outlined in Table 2. The **Dental-j1vge** is a dental object detection dataset used for the detection of different classes of teeth like molar, pre-molar, crown, etc. [1]. The **BCCD** Dataset contains three different classes, white blood cells, red blood cells, and platelets, mainly for blood cell detection [3]. **Kvasir** Dataset is a Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection from the Vestre Viken Health Trust (Norway)[9].

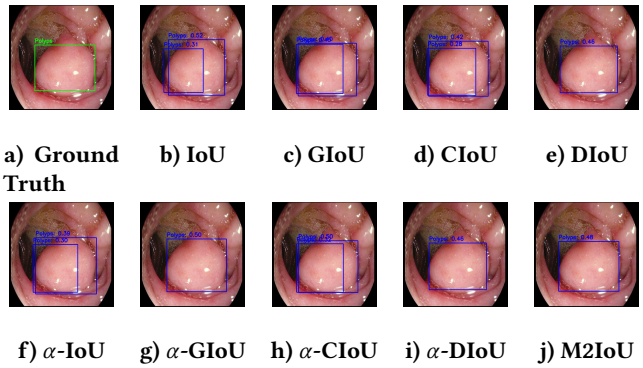
5.2 Methodology

We train the YOLOv8-x [6] model (for 50 epochs) from scratch on five different loss functions namely, IoU, GIoU, DIoU, CIoU, α-IoU, α-GIoU, α-DIoU, α-CIoU, and M2IoU loss function for each of the three datasets. We choose the best model for each loss function obtained till epochs 10, 20, 30, 40, and 50 to track the learning ability

Table 2: Dataset Configuration

Dataset	Train Size	Test Size	Classes
BCCD	765	109	3
Kvasir	800	200	8
Dental-j1vge	9926	1026	10

of the loss functions. We evaluate these models on the mAP50 and mAP50-95 (average of 10 mAP across different IoU thresholds = (AP50 + AP55+ . . . + AP95) / 10) on a randomly sampled subset of the test data which is not used to update the parameters of the model in any way. Finally, we compare all the trained models (after 50 epochs) on the Dice Coefficient Score metric defined as twice the size of the intersection divided by the sum of the sizes of the two boxes.



a) Ground Truth **b) IoU** **c) GIoU** **d) CIoU** **e) DIoU**

f) α-IoU **g) α-GIoU** **h) α-CIoU** **i) α-DIoU** **j) M2IoU**

Figure 3: Predictions of different loss functions.

5.3 Results and Discussions

The results of mAP50 and mAP50-95 values for the training phase are shown in Tables 3 and 4 respectively.[1] M2IoU achieves the highest metric scores at more than **43%** of the landmark epochs for all the datasets alone, demonstrating its faster learning at each stage.

The Dice Coefficient Score of the fully trained models, as reported in Table 5, reveals that M2IoU demonstrates competitive performance. For the Dental-j1vge dataset, M2IoU achieves a dice score of 0.891, surpassing IoU by **6.07%** and the CIoU loss function by an impressive **6.96%**. Although M2IoU scores less than α-DIoU on the Kvasir and BCCD datasets, it still performs better than the remaining loss functions. It is worth noting that M2IoU outperforms the CIoU loss function for all three datasets.

The M2IoU loss function's higher dice coefficient score makes it suitable for training in settings where high accuracy is required such as medical object detection. In Fig. 3, for the given ground truth, we observe that M2IoU performs well by identifying the precise bounding box over the polyp (polycyst). Most other loss functions produce incorrect and duplicate predictions, which could

[1]We embolden the maximum value at each landmark epoch and underline the maximum value achieved by the M2IoU loss function

Table 3: mAP50 values on the test set for different loss functions on multiple epochs

Datasets	Dental-j1vge					BCCD					Kvasir				
Epochs	10	20	30	40	50	10	20	30	40	50	10	20	30	40	50
$\mathcal{L}_{IoU}$	0.638	0.739	0.765	0.837	0.842	**0.784**	0.882	0.902	0.916	0.930	0.298	0.501	0.633	0.747	0.824
$\mathcal{L}_{GIoU}$	0.686	**0.745**	0.851	0.866	0.886	0.697	0.876	0.903	0.918	0.926	0.204	0.502	0.649	0.750	0.836
$\mathcal{L}_{CIoU}$	0.691	0.731	0.831	0.870	0.891	0.547	0.883	0.900	0.916	0.928	0.102	0.588	0.655	0.773	0.816
$\mathcal{L}_{DIoU}$	0.687	0.740	0.827	0.873	0.915	0.635	0.857	0.904	**0.921**	**0.931**	0.134	0.544	0.691	0.784	**0.857**
$\mathcal{L}_{\alpha-IoU}$	**0.699**	0.725	0.793	0.813	0.876	0.755	0.890	0.907	0.919	**0.907**	0.239	0.644	0.705	0.748	0.825
$\mathcal{L}_{\alpha-GIoU}$	0.635	0.692	0.843	0.857	0.902	0.752	0.888	0.907	0.913	0.915	0.287	0.640	0.714	0.790	0.822
$\mathcal{L}_{\alpha-CIoU}$	0.699	0.702	0.851	0.880	0.916	0.775	0.884	0.903	0.920	0.914	0.253	0.657	**0.755**	0.796	0.832
$\mathcal{L}_{\alpha-DIoU}$	0.693	0.745	0.819	0.843	0.904	0.774	0.890	0.904	0.909	0.910	0.146	**0.662**	0.725	0.796	0.831
$\mathcal{L}_{M2IoU}$	0.643	0.728	**0.852**	**0.890**	**0.927**	0.718	**0.887**	0.904	0.919	0.925	**0.305**	0.539	0.691	**0.799**	0.826

Table 4: mAP50-95 values on the test set for different loss functions on multiple epochs

Datasets	Dental-j1vge					BCCD					Kvasir				
Epochs	10	20	30	40	50	10	20	30	40	50	10	20	30	40	50
$\mathcal{L}_{IoU}$	0.302	0.393	0.416	0.468	0.510	0.420	0.523	**0.589**	0.616	0.645	0.120	0.276	0.414	0.515	0.590
$\mathcal{L}_{GIoU}$	0.323	0.386	0.461	0.519	0.543	0.344	0.498	0.578	0.627	0.641	0.093	0.293	0.443	0.505	0.581
$\mathcal{L}_{CIoU}$	0.324	0.377	0.458	0.514	0.554	0.248	0.509	0.580	0.616	0.648	0.035	0.327	0.424	0.550	0.593
$\mathcal{L}_{DIoU}$	0.330	0.392	0.447	0.480	0.574	0.276	0.531	0.588	0.616	0.644	0.053	0.324	0.444	0.539	0.621
$\mathcal{L}_{\alpha-IoU}$	0.376	0.442	0.408	0.534	0.538	0.489	0.489	0.528	**0.639**	**0.655**	0.090	0.375	0.425	0.496	0.543
$\mathcal{L}_{\alpha-GIoU}$	0.326	0.350	0.471	0.505	**0.521**	0.521	0.506	0.515	0.611	0.651	0.103	0.367	0.418	0.482	0.532
$\mathcal{L}_{\alpha-CIoU}$	0.319	0.361	0.450	0.541	0.574	0.521	0.585	0.522	0.638	0.651	0.080	**0.378**	0.449	0.506	0.540
$\mathcal{L}_{\alpha-DIoU}$	0.320	0.379	0.440	0.499	0.555	0.495	**0.590**	0.510	0.632	0.652	0.054	0.355	0.440	0.498	0.538
$\mathcal{L}_{M2IoU}$	0.327	0.379	**0.464**	**0.539**	**0.581**	0.358	0.550	0.573	0.627	0.637	**0.132**	0.318	**0.451**	**0.567**	**0.625**

Table 5: Dice Coefficient Scores on test dataset

Dataset	Dental-j1vge	BCCD	Kvasir
$\mathcal{L}_{IoU}$	0.840	0.883	0.771
$\mathcal{L}_{GIoU}$	0.835	0.888	0.783
$\mathcal{L}_{CIoU}$	0.833	0.881	0.765
$\mathcal{L}_{DIoU}$	0.796	0.883	0.800
$\mathcal{L}_{\alpha-IoU}$	0.850	0.831	0.765
$\mathcal{L}_{\alpha-GIoU}$	0.827	0.886	0.683
$\mathcal{L}_{\alpha-CIoU}$	0.845	0.821	0.766
$\mathcal{L}_{\alpha-DIoU}$	0.803	**0.893**	**0.803**
$\mathcal{L}_{M2IoU}$	**0.891**	0.889	0.784

lead to diagnostic errors and compromise the model's intended purpose.

5.4 FLOPS comparison with CIoU

Minimizing FLOPS reduces energy consumption, which is an active area of concern especially in ML Development. Assuming $atan(x)$ takes 30 FLOPS for a floating decimal x, and each of the arithmetic operations addition, subtraction, division, and multiplication take 1 FLOP each, the CIoU loss function requires 64 more FLOPS on an average than M2IoU loss per ground-truth prediction pair. Let B be the number of candidate bounding boxes generated by the model after Non-Maximal Suppression (NMS), D be the number of data points used for training and E be the total number of training epochs. Thus, the overall difference in the number of FLOPS, ΔF is $O(B * D * E)$. This number is typically of the order of 10^6 FLOPS for training on a few thousand images. Utilizing the M2IoU loss function in place of CIoU can potentially reduce the computational burden by three orders of magnitude.

6 Conclusion

In this paper, we introduced a new loss function and metric named M2IoU. We demonstrated that this new metric learns faster than the other loss functions and hence can be very useful in medical object detection where high accuracy is desirable. It is a better choice in most of the performance metrics on downstream tasks like object detection. We also showed an analytical FLOPS comparison with CIoU and demonstrated that our loss function performs better than CIoU with fewer computations.

As for future work, we would like to introduce a mechanism/ algorithm to readily fine-tune the value of the hyper-parameter α depending on the dataset in hand. We would also like to extend M2IoU loss to 3-D medical object detection (like 3D MRI).

7 ACKNOWLEDGMENTS

This research study was conducted using human subject data made available in open access. No funding was received to conduct this study. The authors have no additional financial or non-financial interests to disclose.

References

[1] bitcamp. 2023. dental Dataset. https://universe.roboflow.com/bitcamp/dental-j1vge. https://universe.roboflow.com/bitcamp/dental-j1vge [Accessed on 2023-11-08].

[2] Lorenzo Ciampiconi, Adam Elwood, Marco Leonardi, Ashraf Mohamed, and Alessandro Rozza. 2023. A survey and taxonomy of loss functions in machine learning. *arXiv preprint arXiv:2301.05579* (2023).

[3] cosmicad and akshaylamba. [n. d.]. BCCD Dataset. https://github.com/Shenggan/BCCD_Dataset.. [Accessed 09-11-2023].

[4] Nilay Ganatra. 2021. A comprehensive study of applying object detection methods for medical image analysis. In *2021 8th international conference on computing for sustainable global development (INDIACom)*. IEEE, 821–826.

[5] Jiabo He, Sarah Erfani, Xingjun Ma, James Bailey, Ying Chi, and Xian-Sheng Hua. 2021. Alpha-IoU: A family of power intersection over union losses for bounding box regression. *Advances in Neural Information Processing Systems* 34 (2021), 20230–20242.

[6] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. 2023. *Ultralytics YOLOv8*. https://github.com/ultralytics/ultralytics

[7] Mengfang Li, Yuanyuan Jiang, Yanzhou Zhang, and Haisheng Zhu. 2023. Medical image analysis using deep learning algorithms. *Frontiers in Public Health* 11 (2023), 1273253.

[8] Andreas S Panayides, Amir Amini, Nenad D Filipovic, Ashish Sharma, Sotirios A Tsaftaris, Alistair Young, David Foran, Nhan Do, Spyretta Golemati, Tahsin Kurc, et al. 2020. AI in medical imaging informatics: current challenges and future directions. *IEEE journal of biomedical and health informatics* 24, 7 (2020), 1837–1857.

[9] Konstantin Pogorelov, Kristin Ranheim Randel, Carsten Griwodz, Sigrun Losada Eskeland, Thomas de Lange, Dag Johansen, Concetto Spampinato, Duc-Tien Dang-Nguyen, Mathias Lux, Peter Thelin Schmidt, Michael Riegler, and Pål Halvorsen. 2017. KVASIR: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection. In *Proceedings of the 8th ACM on Multimedia Systems Conference* (Taipei, Taiwan) *(MMSys'17)*. ACM, New York, NY, USA, 164–169. https://doi.org/10.1145/3083187.3083212

[10] Xiaoliang Qian, Niannian Zhang, and Wei Wang. 2023. Smooth giou loss for oriented object detection in remote sensing images. *Remote Sensing* 15, 5 (2023), 1259.

[11] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. 2019. Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 658–666.

[12] Ma Siliang and Xu Yong. 2023. MPDIoU: A Loss for Efficient and Accurate Bounding Box Regression. *arXiv preprint arXiv:2307.07662* (2023).

[13] Zhaohui Zheng, Ping Wang, Wei Liu, Jinze Li, Rongguang Ye, and Dongwei Ren. 2020. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. In *The AAAI Conference on Artificial Intelligence (AAAI)*.

[14] Zhaohui Zheng, Ping Wang, Dongwei Ren, Wei Liu, Rongguang Ye, Qinghua Hu, and Wangmeng Zuo. 2021. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. *IEEE transactions on cybernetics* 52, 8 (2021), 8574–8586.

[15] Dingfu Zhou, Jin Fang, Xibin Song, Chenye Guan, Junbo Yin, Yuchao Dai, and Ruigang Yang. 2019. Iou loss for 2d/3d object detection. In *2019 international conference on 3D vision (3DV)*. IEEE, 85–94.

[16] Zongwei Zhou. 2021. *Towards annotation-efficient deep learning for computer-aided diagnosis*. Ph. D. Dissertation. Arizona State University.