Extension de l'Intersection over Union pour améliorer la détection d'objets de petite taille en régime d'apprentissage few-shot

Pierre Le Jeune[1,2], Anissa Mokraoui[1]
[1]L2TI, Université Sorbonne Paris Nord, [2]COSE

I – Inconvénients de l'*Intersection over Union*

L'*Intersection over Union* (IoU) n'est pas adaptée pour la détection de petits objets. Un décalage d'un pixel entre deux boites englobantes entraine une baisse beaucoup importante de l'IoU pour des petits objets (voir Fig. 1). On considère les objets comme petits lorsque leur surface a est inférieures à 32^2 pixels, moyens lorsque $32^2 < a \leq 96^2$ et grands sinon. L'origine de cette baisse est la propriété d'invariance de l'IoU par rapport la taille des objets pour des décalages proportionnels à la taille. Les détecteurs d'objets usuels ne possèdent pas cette propriété : ils sont moins précis avec les petits objets.

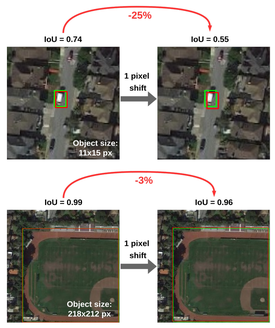

Fig. 1 : Comparaison des diminutions d'IoU pour des objets de tailles différentes.

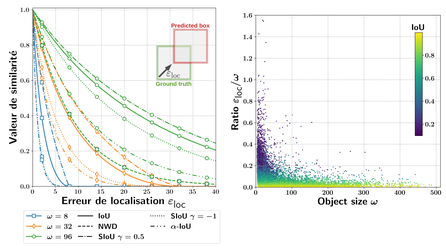

Fig. 2 : **(Gauche)** Évolution des critères IoU, NWD [1], SIoU avec $\gamma = 0.5$ et $\gamma = -1$ ainsi qu'α-IoU [2] en fonction de l'erreur de localisation ε_{loc} pixels entre une prédiction et un label pour différentes tailles d'objets $\omega \in \{8, 32, 96\}$. **(Droite)** Ratio entre l'erreur de localisation ε_{loc} et la taille de l'objet ω pour un modèle de détection entraîné sur le dataset DOTA. Chaque point représente l'erreur de localisation du modèle pour un objet dans DOTA.

Cette baisse est problématique car un tout petit décalage de localisation peut faire passer une boite englobante en dessous des seuils fixés dans les algorithmes de détection. L'IoU est utilisée à de multiples endroits : fonction de coût, sélection des exemples, post-processing et évaluation.

II – *Scaled-adaptative Intersection over Union*

Pour résoudre les problèmes liés à l'IoU, nous proposons un critère adaptatif et contrôlable en fonction de la taille des objets :

$$\text{SIoU}(b_1, b_2) = \text{IoU}(b_1, b_2)^p$$
$$\text{avec} \qquad p = 1 - \gamma \exp\left(-\frac{\sqrt{w_1 h_1 + w_2 h_2}}{\sqrt{2}\kappa}\right), \tag{1}$$

γ et κ contrôlent la direction et la force de SIoU. Lorsque $\gamma < 0$, les petits objets sont défavorisés, lorsque $\gamma \geq 0$, ils sont favorisés. κ détermine la vitesse à laquelle on récupère le comportement de l'IoU lorsque la taille des objets augmente.

Generalized IoU (GIoU) [3] peut aussi être étendue de cette manière :

$$\text{GSIoU}(b_1, b_2) = \begin{cases} \text{g}(b_1, b_2)^p & \text{if g}(b_1, b_2) \geq 0 \\ -|\text{g}(b_1, b_2)|^p & \text{if g}(b_1, b_2) < 0 \end{cases}. \tag{2}$$

C'est plus souvent GIoU qui est utilisée comme fonction de coût pour l'entrainement de modèles de détection. Dans ce cas, $\mathcal{L}_{\text{GIoU}}(b_1, b_2) = 1 - \text{GIoU}(b_1, b_2)$, avec $b_1 = [x_1, y_1, w_1, h_1]^T$ et $b_2 = [x_2, y_2, w_2, h_2]^T$ deux boites englobantes. On peut de manière similaire définir $\mathcal{L}_{\text{GSIoU}}(b_1, b_2) = 1 - \text{GSIoU}(b_1, b_2)$.

Avec $\gamma \geq 0$, SIoU évite une baisse trop rapide du critère de similarité des boites englobantes tout en conservant le comportement classique de l'IoU pour les objets de plus grande taille.

III – Accord avec la perception humaine

La propriété d'invariance en la taille des objets de l'IoU n'est problématique uniquement pour les détecteurs d'objets qui sont moins précis pour les petits objets. La perception humaine est également moins précise pour les petits objets. Une étude subjective a été menée pour le confirmer :

- 75 participants et environ 3000 comparaisons.
- Présentation aux participants de deux rectangles avec une IoU aléatoire.
- Les participants devaient noter sur une échelle de 1 à 5 la qualité du second rectangle (prédiction) par rapport au premier (annotation vraie).

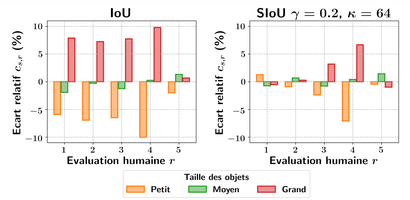

Fig. 3 : Perception humaine versus IoU et SIoU.

$$c_{s,r} = \frac{\mathcal{C}_{s,r} - \frac{1}{|\mathcal{S}|}\sum_{s \in \mathcal{S}} \mathcal{C}_{s,r}}{\frac{1}{|\mathcal{S}|}\sum_{s \in \mathcal{S}} \mathcal{C}_{s,r}}, \tag{3}$$

où $\mathcal{C}_{s,r}$ représente l'IoU ou SIoU moyen pour une taille d'objet s et une notation r.

La perception humaine est plus laxiste envers les petits objets : on se satisfait de manière identique d'une détection avec une IoU plus faible lorsque l'objet est petit.

⇒ **SIoU est donc plus adaptée que l'IoU pour la conception de systèmes destinés à faciliter les tâches d'opérateurs humains**, e.g. repérage de points d'intérêt, diagnostic médical.

IV – Résultats expérimentaux

La détection de petits objets en régime few-shot est extrêmement difficile [4], pour cette raison, les expériences ci-dessous se concentrent sur la détection few-shot dans des images aériennes.

Influence de γ sur les performances :

- $\gamma < 0$ donne de bien meilleurs résultats sur les petits objets.
- Les modèles sont optimisés pour minimiser $\mathcal{L}_{\text{GSIoU}}(b_1, b_2) = 1 - \text{GSIoU}(b_1, b_2)$. Ainsi, favoriser les petits objets ($\gamma > 0$) revient à réduire la contribution des petits objets à la loss. L'entrainement se focalise alors sur les objets plus grands.
- Un critère unique mieux aligné que IoU avec la perception humaine, et qui favorise les petits objets pendant l'entrainement n'est pas possible.

	Classes de base				Nouvelles Classes			
γ	All	S	M	L	All	S	M	L
0.5	47.09	21.29	54.67	65.48	30.50	8.83	44.97	65.89
0.25	45.94	21.60	54.39	63.40	30.96	12.53	42.37	64.14
0	52.41	26.94	61.17	63.00	41.03	24.01	52.13	69.78
-0.5	52.80	**27.16**	61.19	64.61	41.06	25.20	50.18	**72.04**
-1	53.03	23.20	61.53	66.68	42.77	27.55	**52.01**	70.76
-2	**54.06**	23.68	**62.69**	66.62	43.67	30.04	51.69	69.66
-3	52.91	22.14	61.19	66.02	**45.88**	**34.83**	51.26	70.78
-4	53.59	22.50	62.48	66.18	42.43	27.56	51.79	68.70
-9	53.11	20.98	62.13	**67.00**	42.63	30.53	48.89	68.62

Tab. 1 : Évolution des performances de détection en régime few-shot sur DOTA pour différentes valeurs de γ, avec $\kappa = 16$ fixé.

Résultats expérimentaux en détection *few-shot* :

Pour démontrer la supériorité de SIoU sur les critères existants, des comparaisons sont menées sur des datasets aériens (DOTA [5] et DIOR [6]) et naturels (Pascal VOC [7] et MS COCO [8]).

	Classes de base				Nouvelles Classes			
Loss	All	S	M	L	All	S	M	L
IoU	50.67	**25.83**	57.49	68.24	32.41	10.06	47.87	67.09
α-IoU	46.72	13.24	55.21	**69.94**	33.95	12.58	46.58	**74.50**
SIoU	**53.62**	24.07	**61.91**	67.34	**39.05**	**16.59**	**54.42**	**74.49**
NWD	50.79	19.19	58.90	67.90	41.65	28.26	50.16	65.06
GIoU	52.41	**26.94**	61.17	63.00	41.03	24.01	**52.13**	69.78
GSIoU	**52.91**	22.14	**61.19**	**66.02**	**45.88**	**34.83**	51.26	**70.78**

Tab. 2 : Comparaison des performances few-shot en utilisant différents critères, IoU, α-IoU, SIoU, NWD, GIoU, et GSIoU, comme fonction de coût. La mAP est rapportée avec un seuil d'IoU à 0.5 et selon la taille des objets avec $\gamma = -3$, $\kappa = 16$ et $\alpha = 3$.

		Classes de base				Nouvelles Classes			
		All	S	M	L	All	S	M	L
DOTA	GIoU	52.41	**26.94**	61.17	63.00	41.03	24.01	**52.13**	69.78
	GSIoU	**52.91**	22.14	**61.19**	**66.02**	**45.88**	**34.83**	51.26	**70.78**
DIOR	GIoU	58.90	10.38	40.76	80.44	47.93	9.85	47.61	68.40
	GSIoU	**60.29**	**11.28**	**43.24**	**81.63**	**52.85**	**13.78**	**53.73**	**71.22**
Pascal	GIoU	51.09	**13.93**	**40.26**	62.01	48.42	18.44	36.06	59.99
	GSIoU	**54.47**	13.88	40.13	**66.82**	**55.16**	**22.94**	**36.24**	**67.40**
COCO	GIoU	19.15	**8.72**	22.50	30.59	26.25	11.96	23.95	38.60
	GSIoU	**19.57**	8.41	**23.02**	**31.07**	**27.11**	**12.81**	**26.02**	**39.20**

Tab. 3 : Comparaison des performances de détection entre GIoU et GSIoU sur 4 datasets : DOTA , DIOR, Pascal VOC et MS COCO, en régime few-shot. $\gamma = -3$ et $\kappa = 16$ pour DOTA et DIOR et $\gamma = -1$ pour Pascal VOC et MS COCO.

SIoU apporte un gain de performance significatif sur les petits objets. Cela génère une amélioration substantielle sur les images aériennes car celles-ci contiennent plus de petits objets. Sur des images aériennes, les gains sur les petits objets sont également observés.

Résultats expérimentaux en détection classique :

Finalement, SIoU est également utilisée pour la détection d'objet classique, les gains obtenus sont alors réduits. L'abondance d'annotations dans le cas classique est suffisante pour apprendre à détecter correctement les petits objets.

	DOTA				DIOR			
FCOS	All	S	M	L	All	S	M	L
GIoU	34.9	17.4	36.6	43.3	48.1	10.1	40.3	63.2
GSIoU	**36.8**	**17.5**	**40.4**	**45.2**	**49.2**	**11.0**	**41.2**	**66.1**

Tab. 4 : Performance de détection classique sur DOTA et DIOR avec GIoU et GSIoU ($\gamma = -3$ et $\kappa = 16$). Ici la mAP est calculée comme une moyenne avec plusieurs seuils (de 0.5 à 0.95) comme c'est le cas en détection classique.

V – Conclusions et Perspectives

- SIoU ($\gamma < 0$) permet un gain considérable de performance sur les petits objets.
- Ce gain est réduit lorsque beaucoup d'annotations sont disponibles ou lorsque les petits objets sont rares.
- SIoU ($\gamma > 0$) est un critère plus adapté pour l'évaluation de modèle de diffusion car mieux aligné avec la perception humaine.
- Un critère unique basé sur l'IoU ne peut pas à la fois être optimal pour l'entrainement et être aligné avec la perception humaine.
- Des expériences sont nécessaires pour évaluer l'influence de SIoU dans les autres composants des modèles de détection, notamment la sélection d'exemples et la suppression non maximale.

Remerciements

Les auteurs remercient l'entreprise COSE ainsi que le LabCom IRISER (ANR-21-LCV3-0004) pour leur collaboration étroite et le financement de ce projet.

Références

[1] Chang XU et al. « Detecting Tiny Objects in Aerial Images : A Normalized Wasserstein Distance and A New Benchmark ». In : *ISPRS Journal of Photogrammetry and Remote Sensing (ISPRS J P & RS)* (2022).

[2] Jiabo HE et al. « Alpha-IoU : A Family of Power Intersection over Union Losses for Bounding Box Regression ». In : *NEURIPS* 34 (2021), p. 20230-20242.

[3] Hamid REZATOFIGHI et al. « 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*. 2019, p. 658-666.

[4] Pierre LE JEUNE et Anissa MOKRAOUI. « Improving Few-Shot Object Detection through a Performance Analysis on Aerial and Natural Images ». In : *Proceedings of the 30th European Signal Processing Conference (EUSIPCO)*. 2022.

[5] Gui-Song XIA et al. « DOTA : A large-scale dataset for object detection in aerial images ». In : *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, p. 3974-3983.

[6] Ke LI et al. « Object detection in optical remote sensing images : A survey and a new benchmark ». In : *ISPRS Journal of Photogrammetry and Remote Sensing* 159 (2020), p. 296-307.

[7] Mark EVERINGHAM et al. « The pascal visual object classes (voc) challenge ». In : *International journal of computer vision* 88.2 (2010), p. 303-338.

[8] Tsung-Yi LIN et al. « Microsoft coco : Common objects in context ». In : *ECCV*. Springer. 2014, p. 740-755.