# 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]

[2]COSE, 5 bis Route de Saint-Leu, 95360 Montmagny, France

[1]Laboratoire de Traitement et Transport de l'Information, Université Sorbonne Paris Nord
99 avenue Jean-Baptiste Clément, 93430 Villetaneuse, France

**Résumé –** L'*Intersection over Union* (IoU) est un critère mesurant la proximité entre deux boîtes englobantes. C'est un critère fondamental pour la détection d'objets car il permet à la fois d'entraîner les modèles de détection (en tant que fonction de coût) et de les évaluer. C'est la valeur d'IoU entre les boîtes vraies et prédites qui détermine si une détection est correcte ou non, selon un seuil défini. La détection d'objets de petite taille est un problème persistent en vision par ordinateur, notamment lorsque l'on dispose de peu de données annotées (*few-shot learning*). La faible supervision disponible empêche l'apprentissage d'une localisation robuste, ce qui est particulièrement néfaste pour les petits objets. En effet, lorsque un objet est petit, quelques pixels d'écart entre la boîte englobante prédite et la boîte annotée suffisent pour que la prédiction soit considérée comme une fausse détection. Nous proposons dans cet article *Scale-adaptive Intersection over Union* (SIoU), un nouveau critère de similarité contrôlable et adaptatif en fonction de la taille des objets. Premièrement, SIoU permet de trouver un meilleur équilibre entre petits et grands objets pendant l'entraînement de méthodes de détection few-shot, pour lesquelles les petits objets sont extrêmement problématiques. Des expériences sur quatre jeux de données distincts montrent des performances de détection supérieures en utilisant SIoU comme fonction de coût. Deuxièmement, en étant plus laxiste envers les petits objets, SIoU est mieux aligné avec la perception humaine que l'IoU, ce qui en fait un critère d'évaluation plus adéquat.

**Abstract –** The Intersection over Union (IoU) measures the similarity between bounding boxes. It is a fundamental criterion both for training and evaluating object detection models. The IoU between a predicted and ground truth box determines whether a detection is correct. Detecting small objects is a well-known issue in computer vision, especially when only a few labeled images are available (few-shot learning). The weak supervision is insufficient to learn robust localization, which is particularly harmful for small targets. A predicted box shifted from only a few pixels from the ground truth is considered a false detection when objects are small. Therefore, we propose Scale-adaptive Intersection over Union, a new box similarity criterion which adapts to object size. First, as a loss function, SIoU finds a better balance between small and large objects. This is particularly desirable for the training of few-shot methods for which small objects are extremely challenging. Experiments on four datasets showcase the superiority of SIoU as a loss function. Secondly, according to the user study that we conducted, SIoU is better aligned than IoU with human perception, making it a more suitable criterion for model evaluation.

## 1 Introduction

Malgré les avancées récentes en détection d'objets, il est toujours difficile de localiser les objets de petite taille. Ce problème est largement exacerbé lorsque les annotations sont rares (régime d'apprentissage *few-shot*) comme démontré par [1]. La faible supervision disponible dans ce cas ne permet pas l'apprentissage d'une localisation robuste. Les erreurs de localisation sont d'autant plus problématiques pour les petits objets. En effet, l'*Intersection over Union* (IoU) qui mesure la similarité entre les boîtes englobantes, est invariante par rapport à la taille des objets. Cela signifie que si deux boîtes $b_1$ et $b_2$ sont décalées de $\rho$ pixels, multiplier $\rho$ et les coordonnées de $b_1$ et $b_2$ par la même quantité ne change pas leur IoU. Cela semble être une propriété souhaitable, mais lorsque l'on s'intéresse aux petits objets, cela est contestable. L'IoU est souvent utilisé avec des seuils, par exemple on sélectionne comme détections correctes uniquement les boîtes dont l'IoU avec une vérité terrain est supérieur à 0.5. Ce seuil est fixe pour tous les objets, mais du à l'invariance de l'IoU, cela signifie qu'une bonne détection d'un petit objet doit avoir une erreur de localisation absolue $\varepsilon_{loc}$ (c.a.d le décalage absolu en pixel) beaucoup plus petite qu'un objet de grande taille. Or, le ratio entre l'erreur de localisation et la taille des objets, pour un détecteur entraîné, augmente largement pour les petits objets (voir Figure 1 droite). Bien qu'ils soient entraînés avec l'IoU, les réseaux existants ne sont pas indépendants de la taille des objets et sont beaucoup moins précis pour détecter les petites cibles. Pour résoudre ce problème, nous proposons *Scale-adaptive Intersection over Union* (SIoU), une extension de l'IoU qui s'adapte en fonction de la taille des objets.

SIoU est paramétrable et permet de favoriser plus ou moins les petits objets, tout en conservant un comportement proche de l'IoU pour les grands objets. La notion de taille d'objet est définie concrètement dans [4] : les petits objets sont ceux dont l'aire $\omega^2$ ne dépasse pas $32^2$ pixels, les moyens vérifient $32 < \omega \leq 96$ et les grands $\omega > 96$. SIoU peut être utilisé comme fonction de coût afin de favoriser les petits objets pendant l'entraînement. Nos expériences montrent que cela apporte des gains significatifs pour la détection en régime few-shot (FSOD) sur des images naturelles (sur les jeux de données Pascal VOC [5] et MS COCO [4]) ainsi que sur des images aériennes (sur les jeux de données DOTA [6] et DIOR [7]). Nous concentrons notre analyse sur le régime few-shot car celui-ci

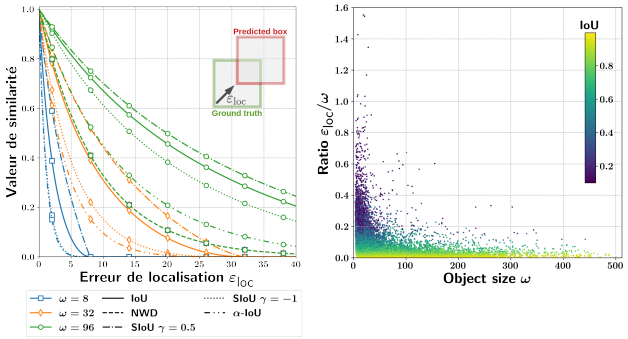


Figure 1. **(Gauche)** Évolution des critères IoU, NWD [2], SIoU avec $\gamma = 0.5$ et $\gamma = -1$ ainsi qu'$\alpha$-IoU [3] en fonction de l'erreur de localisation $\varepsilon_{\text{loc}}$ pixels entre une prédiction et un label pour différentes tailles d'objets $\omega \in \{8, 32, 96\}$. **(Droite)** Ratio entre l'erreur de localisation $\varepsilon_{\text{loc}}$ et la taille de l'objet $\omega$ pour un modèle de détection entraîné sur le dataset DOTA. Chaque point représente l'erreur de localisation du modèle pour un objet dans DOTA.

est plus impacté par la présence de petits objets et reflète mieux les cas d'application réels. Cependant des résultats cohérents pour la détection classique sont également présentés. SIoU peut également être utilisé comme critère de similarité lors de l'évaluation des modèles. Une étude subjective que nous avons réalisée auprès de 74 personnes montre qu'en moyenne les humains sont plus laxistes que l'IoU pour les petits objets. Ainsi SIoU est un critère plus pertinent que l'IoU pour évaluer les modèles de détection et construire des applications plus adaptées à des utilisateurs humains.

# 2 SIoU : nouveau critère proposé

## 2.1 IoU et ses variantes

Avant de présenter le nouveau critère, rappelons tout d'abord la définition de l'IoU et certaines de ses variantes. On calcule l'IoU entre deux boîtes englobantes $b_1 = [x_1, y_1, w_1, h_1]^T$ et $b_2 = [x_2, y_2, w_2, h_2]^T$ comme le rapport de l'aire de leur intersection sur l'aire de leur union. Ici, $x_i$ et $y_i$ représentent les coordonnées du centre de la boîte $b_i$, tandis que $w_i$ et $h_i$ représentent sa largeur et sa hauteur. De nombreuses variantes de l'IoU ont été proposées dans la littérature. L'une des plus connues est *Generalized IoU* (GIoU) [8] qui généralise l'IoU lorsque l'intersection entre $b_1$ et $b_2$ est vide. Pour cela, GIoU soustrait à l'IoU un terme mesurant l'écart entre les deux boîtes. Ainsi, $\text{GIoU}(b_1, b_2) \in [-1, 1]$ alors que $\text{IoU}(b_1, b_2) \in [0, 1]$. GIoU est particulièrement efficace en tant que fonction de coût : $\mathcal{L}_{\text{GIoU}}(b_1, b_2) = 1 - \text{GIoU}(b_1, b_2)$ car cela réduit les problèmes d'optimisation de $\mathcal{L}_{\text{IoU}}$ lorsque les boîtes ne se recouvrent pas. $\alpha$-IoU [3] est une autre variante de IoU, elle porte IoU à la puissance $\alpha$. $\alpha$-IoU pénalise d'avantage les prédictions ayant un grand IoU avec une boîte vraie, cela a pour but d'améliorer la précision des détections de manière générale. $\alpha$ permet de régler la précision désirée. Plus récemment, [2] a proposé une alternative afin de mieux détecter les petits objets. Son principe repose sur le calcul d'une distance de Wasserstein normalisée (NWD) entre deux distributions Gaussiennes ajustées aux boîtes $b_1$ et $b_2$. Ce critère est alors plus laxiste pour des boîtes de petite taille. Cependant, lorsque $b_1$ et $b_2$ ont les mêmes dimensions, NWD devient équivalent à une distance euclidienne entre les centres des boîtes et perd ainsi son comportement variable selon la taille des objets.

## 2.2 Scaled-adaptative Intersection over Union

Afin de résoudre les difficultés de détection des petits objets en régime Few-Shot (FS), nous proposons *Scale-Adaptive Intersection over Union* (SIoU) défini comme suit :

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

où $p$ est une fonction de la taille des objets, ainsi les valeurs d'IoU sont plus ou moins amplifiées selon la taille des objets. Les paramètres $\gamma \in ]-\infty, 1]$ et $\kappa > 0$ permettent de régler la force de l'amplification et la vitesse avec laquelle le comportement de l'IoU est rétabli pour les grands objets : $\lim_{\tau \to +\infty} \text{SIoU}(b_1, b_2) = \text{IoU}(b_1, b_2)$ ($\tau = \frac{1}{2}(w_1 h_1 + w_2 h_2)$). Ainsi, SIoU permet d'obtenir un comportement variable mais contrôlé en fonction de la taille des objets, tout en conservant des propriétés proches de l'IoU. Lorsque $\gamma > 0$, SIoU attribue des valeurs de similarité plus grandes aux petits objets, cela permet de mieux s'accorder à la perception humaine (voir Section 3). A l'inverse, avec $\gamma < 0$, SIoU génère des valeurs plus faibles pour les petits objets, nous verrons dans la Section 4 que cela a un intérêt pour l'entraînement des modèles de détection. La Figure 1 (gauche) permet d'observer la différence de comportement entre les critères évoqués et proposés dans les sections précédentes (IoU, SIoU, NWD et $\alpha$-IoU), pour différentes tailles d'objets.

SIoU peut ensuite être généralisé comme l'IoU en GIoU. Il suffit simplement de remplacer IoU par GIoU (noté $g(b_1, b_2)$ ci-dessous) dans l'équation 1 :

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

Ainsi, nous définissons les fonctions des coûts $\mathcal{L}_{\text{SIoU}}(b_1, b_2) = 1 - \text{SIoU}(b_1, b_2)$ et $\mathcal{L}_{\text{GSIoU}} = 1 - \text{GSIoU}(b_1, b_2)$. Ces fonctions de coût peuvent directement remplacer $\mathcal{L}_{\text{IoU}}$ et $\mathcal{L}_{\text{GIoU}}$. Il est important de noter la similitude de formalisme entre $\alpha$-IoU et SIoU. Cependant, il existe une différence cruciale entre les deux. Le comportement d'$\alpha$-IoU ne change pas avec la taille des objets contrairement à SIoU (voir Figure 1, gauche) et ne contribue ni à l'amélioration des performances des petits objets, ni à un meilleur alignement avec la perception humaine.

# 3 Accord avec la perception humaine

Outre son utilité pour l'entraînement de modèles de détection, IoU est aussi nécessaire à leur évaluation. Ces modèles sont généralement évalués avec la *mean Average Precision* (mAP) à un seuil d'IoU (par exemple 0.5). Cela signifie que l'on considère une prédiction de boîte englobante comme correcte uniquement si elle a un IoU supérieur au seuil défini avec une boîte vraie. Pour les petits objets, cela est problématique car un décalage de seulement quelques pixels suffit pour passer d'une prédiction correcte à un faux positif. Ainsi, les performances de détection pour les petits objets sont souvent basses bien que

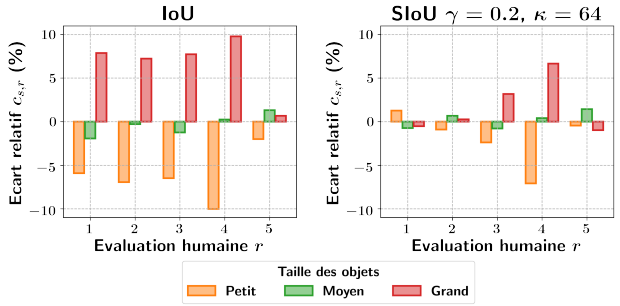


Figure 2. Valeur moyenne d'IoU **(gauche)** et de SIoU **(droite)** pour différentes tailles d'objet $s$ et notation $r \in \{1, 2, 3, 4, 5\}$. Les scores sont donnés comme l'écart relatif avec le score moyen pour chaque $r$.

Tableau 1. Corrélation (Kendall's $\tau$) entre les valeurs des différents critères et la notation humaine. Pour SIoU, $\gamma = 0.2$ and $\kappa = 64$, pour $\alpha$-IoU, $\alpha = 3$.

| | **IoU** | **SIoU** | **$\alpha$-IoU** | **NWD** |
|---|---|---|---|---|
| $r$ | 0.674 | **0.701** | 0.674 | 0.550 |

les prédictions associées semblent tout à fait satisfaisantes pour un observateur humain. En effet, un humain aura tendance à être plus laxiste que l'IoU envers les petits objets. Nous avons réalisé une étude subjective dans laquelle un observateur doit noter une prédiction par rapport à une boîte vraie sur une échelle allant de 1 (très mauvaise localisation) à 5 (très bonne localisation). 74 personnes (spécialistes et non-spécialistes du domaine) ont participé à cette étude pour un total de plus de 3000 notations. Il en ressort qu'un humain attribue en moyenne des notes plus hautes à des petits objets qu'à des grands lorsque le ratio $\varepsilon_{\text{loc}}/\omega$ est fixe. Autrement dit, l'oeil humain est plus laxiste que l'IoU avec les petits objets. Pour rappel, l'IoU est constante lorsque $\varepsilon_{\text{loc}}/\omega$ est fixe. SIoU, au contraire relaxe cette invariance et s'aligne mieux avec la perception humaine. Cela est visible dans la Figure 2, qui représente l'écart relatif d'IoU (gauche) et de SIoU (droite) groupé par notation $r$ et taille d'objet $s$, par rapport à la moyenne pour chaque notation $r$. Plus précisément, on définit :

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

où $\mathcal{C}_{s,r}$ représente l'IoU ou SIoU moyen pour une taille d'objet $s$ et une notation $r$. Cette figure montre d'abord que pour attribuer une notation $r$, un humain a un seuil d'IoU plus bas pour les petits objets que pour les grands. Ensuite, elle montre que SIoU est beaucoup mieux adaptée à la perception humaine, car les seuils de SIoU pour une notation $r$ sont presque identiques peu importe la taille des objets. Cela démontre un meilleur alignement de SIoU avec la perception humaine. Pour renforcer cela, le Tableau 1 montre les valeurs de corrélation entre les différents critères étudiés ici (IoU, SIoU, $\alpha$-IoU et NWD) et la perception humaine. Là encore, SIoU est supérieur aux autres alternatives. Ici, nous avons choisit $\gamma = 0.2$ et $\kappa = 64$ afin de maximiser l'accord avec la perception humaine. Avoir un critère de similarité plus proche de la perception humaine est crucial pour l'élaboration de modèles mieux adaptés au jugement humain, ce qui est particulièrement utile pour des applications de détection (photo-interprétation, radiologie, etc.).

Tableau 2. Comparaison des performances few-shot en utilisant différents critères, IoU, $\alpha$-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 : petits (**S**), moyens (**M**), grands (**L**) et toutes tailles confondues (**All**), avec $\gamma = -3$, $\kappa = 16$ et $\alpha = 3$.

| | **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 |
| **$\alpha$-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** |

# 4 Résultats expérimentaux

Pour prouver l'efficacité de SIoU en tant que fonction de coût, une série d'expériences est présentée dans cette section. Nous nous concentrons ici sur le régime few-shot car celui-ci est plus exigeant envers les petits objets et plus proche des applications réelles. Les expériences sont principalement menées sur des images aériennes (DOTA et DIOR) mais des résultats sur des images naturelles sont également rapportés.

**Détails d'implémentation.** En régime d'apprentissage few-shot, on se concentre principalement sur les performances pour les classes nouvelles, c'est-à-dire les classes pour lesquelles très peu d'exemples annotés sont disponibles. Pour toutes nos expériences nous avons utilisé 10 *shots* (c.a.d 10 images annotées par classe). Les résultats sur les classes de base sont également inclus dans les tableaux à titre indicatif. Nous nous sommes basés sur XQSA [9], une récente méthode de détection few-shot fonctionnant à plusieurs échelles pour améliorer la détection des petits objets. Cette méthode est basée sur le détecteur FCOS [10] et est entraînée de manière épisodique d'abord sur les classes de base puis sur les classes nouvelles.

Tout d'abord, nous comparons l'influence des différents critères évoqués et proposés sur les performances de détection few-shot. Ces résultats sont rapportés dans le Tableau 2. On observe une claire domination de SIoU et de GSIoU (les critères sont séparés en deux groupes selon si ils attribuent des valeurs identiques ou non lorsque les boîtes ont une intersection vide). SIoU et GSIoU améliorent largement les performances sur les classes nouvelles, notamment pour les objets de petite taille.

Il est important de noter que les valeurs de $\gamma$ et $\kappa$ doivent être choisies avec précaution lorsque SIoU (ou GSIoU) est employé dans le calcul de la fonction de coût. En effet, on a vu dans la Section 3 que l'IoU était trop stricte pour les petits objets et qu'en utilisant $\gamma = 0.2$ on obtient un comportement plus proche de la perception humaine (c'est le cas dès que $\gamma > 0$). Cependant lorsque l'on s'intéresse à l'entraînement, $\gamma > 0$ revient à réduire le coût des petits objets ($\mathcal{L}_{\text{GSIoU}} = 1 - \text{GSIoU}(b_1, b_2)$). Cela modifie alors l'équilibre de l'optimisation entre les petits et grands objets. L'entraînement se concentre alors à réduire le coût généré par les grands objets et on obtient des performances moins bonnes sur les petits objets. Ainsi, lorsque SIoU et GSIoU sont utilisés en tant que fonctions de coût, il convient de choisir des valeurs négatives de $\gamma$. Cela augmente le coût généré par les petits objets et l'entraînement se focalise sur la détection de petites cibles. On peut voir cela de manière assez explicite dans le Tableau 3 qui regroupe les performances sur DOTA pour dif-

Tableau 3. Évolution des performances de détection en régime few-shot sur DOTA pour différentes valeurs de $\gamma$, avec $\kappa = 16$ fixé.

| | Classes de base | | | | Nouvelles Classes | | | |
|---|---|---|---|---|---|---|---|---|
| $\gamma$ | **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 |

Tableau 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.

| | 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** |

férentes valeurs de $\gamma$. $\kappa$ a beaucoup moins d'influence sur les performances et est fixé à 16 dans nos expériences.

Afin de démontrer la robustesse de GSIoU, nous avons réalisé des expériences sur quatre jeux de données : DOTA [6] et DIOR [7] (contenant des images aériennes), ainsi que Pascal VOC [5] et MS COCO [4] (images naturelles). Utiliser GSIoU plutôt que GIoU comme fonction de coût améliore les performances de détection des classes nouvelles, en particulier pour les petits objets (voir Tableau 5). On remarque néanmoins que pour les images naturelles (Pascal VOC et MS COCO), les gains sont légèrement plus faibles que sur les images aériennes. Cela s'explique par le fait que les objets sont largement plus petits dans les images aériennes que dans les images naturelles.

Pour finir, nous avons également mené des expériences pour la détection en régime classique (c.a.d avec beaucoup d'annotations) sur DOTA et DIOR (voir Tableau 4). Là encore, l'utilisation de GSIoU comme fonction de coût améliore les performances de détection. En revanche, les gains sont moins conséquents car les performances sont déjà bien meilleures qu'en régime few-shot.

# 5 Conclusion

Dans cet article nous avons mis en avant la sous-optimalité de l'IoU à la fois pour l'entraînement et l'évaluation de modèles de détection d'objets dans des images, en particulier lorsqu'il s'agit de détecter de petits objets. Afin de répondre à ce problème, nous avons proposé *Scaled-adaptive Intersection over Union* (SIoU), une extension de l'IoU permettant de mieux contrôler le comportement de ce critère en fonction de la taille des objets. SIoU peut être facilement paramétré pour mieux s'accorder avec la perception humaine ou pour améliorer l'entraînement des méthodes de détection. Dans ce dernier cas, l'utilisation de SIoU comme fonction de coût permet des gains significatifs de performance sur des images naturelles et aériennes en régime d'apprentissage few-shot.

Tableau 5. 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 mais $\gamma = -1$ pour Pascal VOC et MS COCO.

| | | 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** |


## Remerciements

Nous remercions l'entreprise COSE et le LabCom IRISER (ANR-21-LCV3-0004) pour le financement de ces travaux.


# Références


[1] Pierre Le Jeune and 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.

[2] Chang Xu, Jinwang Wang, , Wen Yang, Huai Yu, Lei Yu, and Gui-Song Xia. Detecting tiny objects in aerial images : A normalized wasserstein distance and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing (ISPRS J P & RS)*, 2022.

[3] Jiabo He, Sarah Erfani, Xingjun Ma, James Bailey, Ying Chi, and Xian-Sheng Hua. Alpha-iou : A family of power intersection over union losses for bounding box regression. *NEURIPS*, 34 :20230–20242, 2021.

[4] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco : Common objects in context. In *ECCV*, pages 740–755. Springer, 2014.

[5] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2) :303–338, 2010.

[6] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. Dota : A large-scale dataset for object detection in aerial images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3974–3983, 2018.

[7] Ke Li, Gang Wan, Gong Cheng, Liqiu Meng, and Junwei Han. Object detection in optical remote sensing images : A survey and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159 :296–307, 2020.

[8] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. 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*, pages 658–666, 2019.

[9] Pierre Le Jeune and Anissa Mokraoui. A comparative attention framework for better few-shot object detection on aerial images. *arXiv*, 2022.

[10] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. Fcos : Fully convolutional one-stage object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9627–9636, 2019.