

Fusion de détecteurs d'objets par coordination multi-agent et a priori causaux

Alaa Daoud¹,
alaa.daoud@uphf.fr

Hiba Alqasir¹,
chloéhiba.alqasir@uphf.fr

Maxime Guèriau²,
maxime.gueriau@insa-rouen.fr

Corentin Bunel³
corentin.bunel@cerema.fr

¹ Université Polytechnique Hauts-de-France, CNRS, UMR 8201-LAMIH, F-59313, Valenciennes, France

² INSA Rouen Normandie, LITIS UR 4108, F-76000 Rouen, France

³ CEREMA, ENDSUM, Le Grand-Quevilly, France

Résumé

Les ensembles de détecteurs d'objets nécessitent une étape de fusion pour agréger des boîtes englobantes redondantes ou conflictuelles. Les approches classiques, telles que *Weighted Boxes Fusion (WBF)*, reposent sur des seuils fixes d'*Intersection over Union (IoU)* et sur des scores de confiance homogènes, ce qui limite leur adaptabilité, leur décentralisation et leur interprétabilité. Nous proposons deux contributions complémentaires : (i) une reformulation de la fusion comme un problème de coordination multi-agent distribuée, où chaque boîte englobante devient un agent autonome capable de coopération, de compétition et de négociation ; (ii) une pondération par fiabilité comme connaissance causale a priori, appliquée après agrégation spatiale et associée à un seuil d'*IoU* adaptatif interprété comme intervention contextuelle, séparant consensus géométrique et confiance épistémique. La fusion est ainsi interprétée comme un mécanisme de coordination régulé par des a priori épistémiques et des interventions contextuelles, reliant systèmes multi-agents et raisonnement causal pour une perception adaptative et interprétable.

Mots-clés

Détection d'objets, fusion adaptative, boîtes englobantes, systèmes multi-agents, coordination, a priori causaux.

Abstract

Object detection ensembles require fusion to resolve redundant or conflicting bounding boxes. Classical methods such as *Weighted Boxes Fusion (WBF)* rely on fixed *IoU* thresholds and uniform detector trust, limiting adaptability, decentralization, and interpretability. Two complementary contributions are proposed : (i) a reformulation of fusion as a distributed multi-agent coordination problem (AWBF), where each bounding box becomes an autonomous agent capable of cooperation, competition, and negotiation ; (ii) a reliability-aware extension (AR-WBF) modeling detector reliability as a causal prior applied after spatial aggregation, with an adaptive *IoU* threshold acting as a contextual intervention that separates geometric consensus from epistemic trust. Ensemble fusion is thus interpreted as a layered coordination mechanism regulated by epistemic priors and contextual interventions, bridging multi-agent systems and causal reasoning for adaptive and interpretable perception.

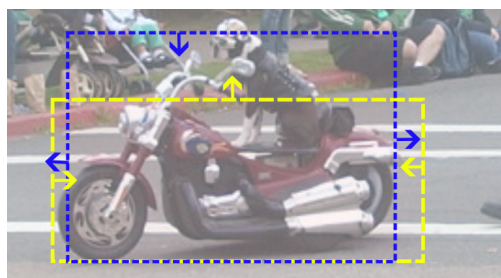


FIGURE 1 – Fusion adaptative de 2 boîtes englobantes sur une moto, image #495357 du jeu de données COCO

Keywords

Object detection, bounding-boxes, adaptive fusion, multi-agent systems, coordination, causal priors.

1 Problématique

La détection d'objets constitue un composant central des systèmes de perception embarqués, notamment en conduite autonome et en robotique. L'usage d'ensembles de détecteurs améliore la robustesse mais génère des prédictions multiples pour un même objet. Une étape de fusion devient alors nécessaire.

Les méthodes classiques telles que *Non-Maximum Suppression (NMS)* [5] ou *Weighted Boxes Fusion (WBF)* [6] réalisent une agrégation statique basée sur des seuils fixes d'*IoU* et des scores de confiance. NMS repose sur la suppression, tandis que WBF relève de la fusion. Sa supériorité sur NMS étant déjà établie dans [6], nous le prenons comme référence pour étudier des mécanismes de fusion. Cette approche soulève néanmoins plusieurs limites : 1. absence d'adaptation au contexte (taille, densité, variabilité des objets) ; 2. absence de modélisation explicite de la fiabilité relative des détecteurs ; 3. centralisation du processus décisionnel ; 4. faible interprétabilité des décisions de fusion. Ces constats nous conduisent aux questions suivantes :

RQ1 : Peut-on formuler la fusion de boîtes englobantes comme une coordination multi-agents distribuée permettant une décision adaptative et décentralisée ?

RQ2 : Comment intégrer explicitement la fiabilité des détecteurs en séparant consensus spatial et confiance épistémique, avec une interprétation causale de la fusion ?

2 Agentification de la fusion

Pour répondre à RQ1, nous proposons *Agentified WBF* (AWBF) [2], qui reformule la fusion comme un problème de coordination multi-agents adaptative (figure 1). Chaque boîte est incarnée par un agent autonome caractérisé par sa localisation, son score de confiance et sa capacité d'interaction locale.

2.1 Formalisation des interactions

Chaque agent propose une boîte englobante et peut interagir avec ses voisins en adoptant des comportements de compétition ou de coopération. Contrairement aux approches centralisées basées sur une métrique symétrique (IoU), nous utilisons une métrique asymétrique d'Intersection over Box (IoB), permettant de modéliser des relations directionnelles d'influence. L'IoB est définie par :

$$\text{IoB}(b_i, b_j) = \frac{|b_i \cap b_j|}{|b_i|},$$

ce qui permet de distinguer la force d'un agent i sur j de celle de j sur i .

Nous définissons alors une interaction compétitive entre deux agents A et B via leurs forces d'attaque et de défense :

$$S_{\text{attack}}(A, B) = s_A \cdot \text{IoB}(b_B, b_A),$$

$$S_{\text{defense}}(B, A) = s_B \cdot \text{IoB}(b_A, b_B),$$

où s_A et s_B désignent les scores de confiance des agents. Le résultat de l'interaction est donné par :

$$R(A, B) = S_{\text{attack}}(A, B) - S_{\text{defense}}(B, A).$$

Le comportement collectif est régulé par un seuil T :

- si $R(A, B) > T$, A domine et B est supprimé ;
- si $R(A, B) < -T$, B domine et A est supprimé ;
- sinon, les deux agents coopèrent et fusionnent leurs boîtes.

Le seuil T , fixé empiriquement, contrôle le niveau de coopération : des valeurs élevées favorisent la fusion, tandis que des valeurs faibles induisent des interactions plus compétitives. Les interactions sont réalisées localement au sein de clusters définis par recouvrement spatial.

2.2 Cadre général de coordination.

Plus généralement, la compétition entre agents peut être formulée comme un mécanisme d'enchères. Chaque agent i soumet une offre fondée sur une fonction d'utilité U_i combinant confiance, recouvrement spatial et fiabilité, et propose une enchère pondérée par recouvrement $\text{Bid}_{i \rightarrow j}$:

$$\text{Bid}_{i \rightarrow j} = U_i \cdot \text{IoB}(b_j, b_i).$$

$$U_i = \lambda_1 s_i + \lambda_2 \text{IoU}(b_i, b_j) + \lambda_3 r_i,$$

où $\lambda_1, \lambda_2, \lambda_3$ sont des coefficients pondérant respectivement la confiance, le recouvrement spatial et la fiabilité du détecteur.

Le gagnant est déterminé de manière décentralisée par comparaison des enchères, tandis que des valeurs proches (contrôlées par T) conduisent à une fusion coopérative. Cette formulation unifie les mécanismes d'attaque-défense et permet d'intégrer explicitement des facteurs de fiabilité dans la décision.

Dans ce cadre, trois mécanismes de coordination ont été instanciés et évalués : (a) une reproduction décentralisée de WBF par moyenne pondérée locale ; (b) un mécanisme compétitif d'attaque-défense fondé sur l'IoB ; (c) une négociation multi-tour permettant un ajustement itératif jusqu'à un consensus local.

Cette reformulation transforme la fusion en un processus décisionnel décentralisé et adaptatif, où la dynamique de consensus émerge des interactions locales. La fiabilité des détecteurs demeure toutefois implicite, car elle n'est pas distinguée formellement de la confiance locale.

3 Fiabilité comme prior causal

Pour répondre à RQ2, nous proposons une extension fiabiliste de WBF, appelée *Adaptive Reliability WBF* (AR-WBF) [1], reposant sur deux mécanismes.

3.1 Pondération différée par fiabilité

Nous définissons un prior de fiabilité $r_k \in [0, 1]$ pour chaque détecteur M_k , dérivé de sa performance validée (mAP normalisée). La fusion produit une boîte agrégée b^* et un score de confiance associé s^* , la fiabilité intervenant uniquement dans la calibration finale du score :

$$b^* = \frac{\sum s_i b_i}{\sum s_i}$$

$$s^* = \left(\frac{1}{n} \sum s_i \right) \times (\alpha + \beta r_k)$$

où b_i et s_i désignent respectivement la boîte englobante et le score de confiance de la détection i , et n le nombre de détections dans le cluster.

Les paramètres α et β contrôlent respectivement le poids du consensus moyen et l'influence du prior de fiabilité ($\alpha + \beta = 1$). Cette paramétrisation rend explicite le compromis entre confiance collective et crédibilité épistémique.

En pratique, r_k est estimé à partir du mAP du détecteur M_k sur un jeu de validation, normalisé dans $[0, 1]$.

$$r_k = \frac{\text{mAP}_k - \min_j(\text{mAP}_j)}{\max_j(\text{mAP}_j) - \min_j(\text{mAP}_j) + \epsilon}.$$

Dans une lecture causale, r_k agit comme un prior modulant la crédibilité des détections issues de M_k , indépendamment du consensus spatial, ce qui permet de dissocier perception géométrique et confiance épistémique. Cette pondération différée évite qu'un détecteur peu fiable n'influence la géométrie du consensus et limite la suppression excessive, notamment dans des ensembles hétérogènes.

3.2 Intervention contextuelle via IoU

En notant σ_A la variance normalisée des tailles d’objets dans le cluster, nous définissons le seuil adaptatif :

$$\tau(c, A) = \tau_0 + \alpha_c \sigma_A.$$

Ce mécanisme agit comme une intervention sur la frontière décisionnelle : les petits objets ou contextes incertains déclenchent une fusion plus permissive, tandis que les objets stables imposent une contrainte plus stricte. La fusion devient ainsi sensible au contexte.

4 Experimentation

Nous évaluons AWBF sur le jeu de données COCO [3] en reprenant le benchmark complet proposé dans [6], comprenant 21 détecteurs et leurs prédictions associées. Ce protocole permet de comparer AWBF à WBF dans un cadre décentralisé, tout en conservant les mêmes entrées et conditions expérimentales. Les différentes variantes de AWBF — fusion coopérative (AWBF-f), compétition (AWBF-c) et négociation multi-tour (AWBF-n) — sont ensuite évaluées afin d’analyser l’impact des mécanismes de coordination sur le processus de fusion. Les performances sont mesurées à l’aide des métriques standards du protocole COCO, notamment la précision moyenne (AP) et le rappel moyen (AR), évalués à différents seuils IoU et pour plusieurs tailles d’objets (petits, moyens et grands).

TABLE 1 – Performances de WBF et des variantes AWBF sur le benchmark COCO, évaluées en AP et AR selon la taille des objets.

	Précision moyenne (AP)			Rappel moyen (AR)		
	Petits	Moyens	Grands	Petits	Moyens	Grands
WBF	0.605	0.731	0.846	0.800	0.850	0.867
AWBF-f	0.610	0.766	0.675	0.664	0.706	0.819
AWBF-c	0.322	0.636	0.632	0.653	0.743	0.840
AWBF-n	0.622	0.780	0.701	0.684	0.741	0.859

Les résultats montrent que les variantes AWBF peuvent améliorer la précision moyenne pour les petits et moyens objets, en particulier avec AWBF-n, qui obtient les meilleurs scores AP sur ces deux catégories. En revanche, WBF conserve un avantage net pour les grands objets ainsi que pour l’ensemble des métriques de rappel. Ces résultats suggèrent que la coordination adaptative peut favoriser la précision de certaines détections, mais qu’elle tend également à réduire la couverture globale des objets détectés par rapport à l’agrégation déterministe de WBF.

La figure 2 illustre l’évolution des boîtes englobantes au cours du processus de fusion adaptative, à trois étapes successives : état initial, intermédiaire et final, sur des images issues du jeu de données COCO. Ces exemples mettent en évidence les ajustements successifs produits par les mécanismes de coordination lors de la construction du consensus spatial.

Afin d’analyser finement l’impact de la fiabilité sur le processus de **AR-WBF**, nous adoptons un protocole contrôlé



FIGURE 2 – Exemples d’évolution des boîtes englobantes sous AWBF aux différentes étapes du processus de fusion adaptative (initiale, intermédiaire et finale) sur des images de COCO.

TABLE 2 – Performances de WBF et AR-WBF sur le jeu de validation COCO avec YOLOv5x et EfficientDet-B2, évaluées en AP et AR selon la taille des objets.

	Précision moyenne (AP)			Rappel moyen (AR)		
	Petits	Moyens	Grands	Petits	Moyens	Grands
WBF	0.281	0.453	0.547	0.481	0.676	0.785
AR-WBF	0.281	0.458	0.553	0.503	0.701	0.801

en sélectionnant deux détecteurs issus du benchmark, plutôt que l’ensemble complet des modèles. Les expériences sont menées sur un sous-ensemble de 500 images du jeu de validation COCO, sélectionnées aléatoirement.

Nous retenons **YOLOv5x** [4] et **EfficientDet-B2** [7], représentant des architectures complémentaires : YOLOv5x favorisant le rappel, et EfficientDet-B2 privilégiant la précision. Ce choix permet d’isoler les effets de la pondération par fiabilité dans un cadre hétérogène mais interprétable.

Les prédictions de ces modèles sont évaluées individuellement puis fusionnées via WBF et AR-WBF. Le prior de fiabilité r_k est estimé à partir du mAP sur un sous-ensemble de validation COCO, puis normalisés avant la fusion. Pour assurer une comparaison équitable, tous les modèles sont pondérés uniformément et un seuil IoU de base de 0.5 est utilisé. Les performances sont évaluées selon le protocole COCO (AP et AR).

La table 2 présente les résultats obtenus avec WBF et AR-WBF sur les prédictions de YOLOv5x et EfficientDet-B2. AR-WBF obtient des performances comparables à WBF pour les petits objets, tout en améliorant systématiquement les métriques AP et AR pour les objets moyens et grands. Les gains observés en rappel suggèrent que l’introduction d’une pondération par fiabilité permet de mieux préserver les détections pertinentes lors du processus de fusion, tout en maintenant une cohérence spatiale entre les boîtes agrégées.

Figure 3 compare les résultats de fusion entre WBF et AR-WBF sur des échantillons sélectionnés à partir des prédic-

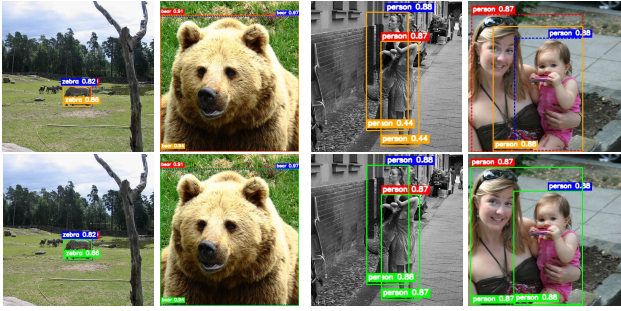


FIGURE 3 – Similarité (à gauche) et divergence (à droite) entre WBF (en haut) et AR-WBF (en bas) sur des exemples issus de YOLOv5x et EfficientDet-B2.

tions de YOLOv5x (bleu) et EfficientDet-B2 (rouge). Dans de nombreux cas, les deux méthodes produisent des résultats similaires. Cependant, des divergences apparaissent lorsque l’IoU se situe à proximité du seuil de fusion, conduisant à des décisions d’agrégation différentes entre WBF et AR-WBF.

Lorsque le recouvrement entre boîtes se situe à proximité du seuil de fusion, WBF peut produire des décisions d’agrégation instables, soit en fusionnant des détections spatialement discordantes, soit en conservant des duplications. AR-WBF réduit ces effets grâce à un seuil IoU adaptatif et à une pondération par fiabilité appliquée après regroupement, ce qui permet d’obtenir une fusion plus cohérente spatialement.

5 Synthèse et perspectives

Les résultats expérimentaux présentés dans la section précédente mettent en évidence deux dynamiques complémentaires dans le processus de fusion.

D’une part, AWBF introduit une coordination distribuée entre détecteurs afin de reproduire le comportement de WBF dans un cadre décentralisé. Les résultats montrent que certaines variantes, en particulier AWBF-n, améliorent la précision moyenne pour les petits et moyens objets (cf. table 1). Ces gains suggèrent que les mécanismes de coordination peuvent faciliter l’agrégation de détections partiellement redondantes. En revanche, WBF conserve de meilleures performances sur les objets de grande taille ainsi que sur plusieurs métriques de rappel, indiquant que les interactions entre agents peuvent également introduire une variabilité dans le consensus spatial produit lors de la fusion.

D’autre part, AR-WBF introduit une pondération des détecteurs fondée sur leur fiabilité estimée. Les résultats présentés dans la table 2 montrent des améliorations cohérentes des métriques AP et AR pour les objets moyens et grands. Cette tendance est également visible dans l’analyse des gains relatifs présentée dans la table 3. Contrairement à AWBF, dont les performances varient fortement selon les métriques considérées, AR-WBF présente des gains plus homogènes sur les indicateurs liés à la précision de localisation et au rappel.

TABLE 3 – Gains relatifs (%) obtenus par AWBF et AR-WBF par rapport à WBF sur différentes métriques AP et AR du protocole COCO.

	AP ₇₅ (%)	AP _M (%)	AR ₅₀ (%)	AR ₇₅ (%)
AWBF	−8.4	3.5	4.9	−10.1
AR-WBF	1.2	0.5	1.0	2.3

Ces deux approches répondent ainsi à des objectifs complémentaires. AWBF cherche à modéliser des mécanismes de coordination distribuée entre détecteurs, tandis que AR-WBF introduit une calibration globale fondée sur la fiabilité des modèles. Le processus de fusion peut ainsi être vu comme une combinaison entre interactions locales et pondération globale des prédictions.

Les mécanismes présentés s’appuient sur des résultats expérimentaux détaillés dans [2, 1], dont nous synthétisons ici les éléments essentiels afin de mettre en évidence la complémentarité entre coordination multi-agents et calibration par fiabilité.

AWBF introduit une coordination distribuée entre détecteurs, capable d’améliorer certaines métriques de précision au prix d’une plus grande variabilité du consensus spatial. À l’inverse, AR-WBF stabilise le processus de fusion grâce à une pondération globale fondée sur la fiabilité des modèles. Ces observations suggèrent qu’une combinaison entre coordination locale et calibration par fiabilité constitue une direction prometteuse pour la fusion multi-sources.

Références

- [1] Alaa Daoud and Hiba Alqasir. From consensus to causality : Adaptive reliability fusion for object detection ensembles. In *CALM-26*, 04 2026.
- [2] Alaa Daoud, Corentin Bunel, and Maxime Guériau. Introducing multiagent systems to av visual perception sub-tasks : A proof-of-concept implementation for bounding-box improvement. In *ATT-24 @ ECAI*, 2024.
- [3] Tsung-Yi Lin et al. Microsoft coco : Common objects in context. In *ECCV*, pages 740–755. Springer, 2014.
- [4] Glenn Jocher, Ayush Chaurasia, Alex Stoken, Jirka Borovec, et al. Yolov5 by ultralytics, 2023. YOLOv5x model variant.
- [5] Alexander Neubeck and Luc Van Gool. Efficient non-maximum suppression. In *ICPR’06*, volume 3, pages 850–855. IEEE, 2006.
- [6] Roman Solovyev, Weimin Wang, and Tatiana Gabruseva. Weighted boxes fusion : Ensembling boxes from different object detectors. *Image and Vision Computing*, 2021.
- [7] Mingxing Tan, Ruoming Pang, and Quoc V Le. Efficientdet : Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10781–10790, 2020.