Sujet proposé : Calibration croisée spatio-catégorielle de la confiance pour la détection d'objets longue distance en milieu maritime

Nom de l'auteur : Adrien Golebiewski [1,2]

[1] STIM – Mines Paris – PSL — [2] LERITY

1. Identification des cas d'usages

La surveillance optronique longue distance en milieu maritime est une fonction stratégique pour les forces armées et la sécurité maritime. Déployés sur les systèmes optroniques de Lerity [1], des réseaux de neurones profonds automatisent la détection et le pistage d'objets d'intérêt maritimes pour réduire la charge cognitive des opérateurs face aux menaces offshore/onshore. Ce pistage automatisé s'avère essentiel pour maintenir une veille panoramique continue, y compris en présence de bruit maritime intense ou de météo dégradée.

Problématique & verrous scientifiques

Les modèles de détection (type YOLO) souffrent d'un biais critique : une sur-confiance systématique, amplifiée sur les cibles éloignées de petite taille (≤ 32² px). Ce manque d'information, combiné au bruit maritime, entraîne des impacts opérationnels majeurs :

- **Saturation de l'opérateur (Alarm fatigue) & Risque de perte de menaces :** Fausses alarmes prioritaires dues à des scores artificiellement élevés attribués au bruit de fond. A l'inverse des scores relativement faibles mais représentant des détections de réelles menaces.
- **Perte de séparabilité TP/FP :** Compression des vrais positifs (TP) et fausses alarmes (FP) dans la zone haute (0,7 – 0,9), rendant le filtrage par seuil inefficace.
- **Instabilité du pistage (Tracking) :** Ruptures de pistes (ex: algorithme SORT) dues aux oscillations du score de confiance et mauvaise priorisation des zooms optiques des systèmes.

Valeur ajoutée & apports opérationnels

Notre approche restaure le sens probabiliste des sorties du réseau sans altérer ses performances brutes de détection. L'intégration d'une brique de recalibration des confiances en fonction de la taille et tenant compte des classes d'objet répond à un enjeu majeur de fiabilisation et d'appui/validation de la décision :

- **Fiabilisation mathématique :** Alignement strict entre le score de confiance affiché et la précision réelle du modèle. Réduction de l'erreur d'estimation sur tout le spectre de résolution. Permet une interprétation plus juste de la détection par l'opérateur.
- **Décompression des scores & Pertinence du score de confiance :** Poussée des FP vers les basses confiances (< 0,2) et maintien des TP en zone haute (> 0,8). L'opérateur peut ainsi fixer un seuil de confiance (Th_{conf}) efficace pour éliminer le bruit sans rejeter les cibles éloignées.
- **Fusion multi-capteurs & zero-latency :** Probabilités fiables pour les systèmes de fusion (RADAR/AIS) via une transformation directe sur les logits, garantissant un surcoût nul (0 ms).

Passage à l'Échelle

- **Phase Académique :** Évaluation et correction du biais sur le Singapore Maritime Dataset (SMD) [3].
- **Phase Opérationnelle :** Qualification et intégration sur les données réelles de Lerity.

2. Approche technique proposée

L'approche repose sur un post-traitement paramétrique non destructif mesurant et corrigeant le biais selon la classe et l'échelle spatiale :

- **Pipeline YOLOv8 & appariement IoU :** Association spatio-temporelle (IoU ≥ 0,5) sur le dataset SMD pour isoler les TP et FP avant le calcul de la calibration.
- **Granularité spatio-catégorielle :** Subdivision croisée selon la classe c et la catégorie de taille t. Le découpage standard COCO [4] est étendu par les quantiles Q33 et Q66 pour isoler les très petites cibles

(small_1 à small_3).

- **Ajustement de température croisé (CS-CTS discrète) & stratégie de repli :** Les logits z_i sont ajustés par une matrice de températures T_{c,t} optimisée selon la méthode du Temperature Scaling de Guo et al. [5]. Une stratégie de fallback hiérarchique à 3 niveaux sécurise l'estimation pour les cellules à faible effectif (Level 1 si n ≥ 20–30, Lvl 2 vers T*c globale, Lvl 3 vers T = 1.0 neutre si sous-représentation absolue).
- **Limites opérationnelles du discret :** Le découpage rigide par taille crée des discontinuités en paliers. Par exemple, lors du suivi d'un navire s'approchant de la côte, sa surface croît continuellement : au franchissement du seuil de taille, la température peut basculer brutalement, provoquant un saut artificiel du score qui déstabilise les seuils d'alarme IHM (risque de fausses alarmes ou manques d'alerte).
- **Extension continue (Continuous CS-CTS) :** Pour éliminer tout saut temporel, la température est modélisée par une fonction log-linéaire continue de la surface de boîte a par classe :

$$T(a, c) = w_c \cdot \log(a) + b_c$$

Cette fonction est optimisée globalement par NLL (avec fallback neutre T = 1.0 si N < 100). Elle supprime tous les effets de seuil IHM tout en maintenant de hauts gains sur les petites cibles.

Éléments de validation & Résultats expérimentaux

L'évaluation s'appuie sur la métrique ECE (Expected Calibration Error) [6] et l'analyse de distribution des confiances par taille et classe d'objet :

- **Gain global & séparabilité (SMD) :** Réduction relative de 43 % de l'erreur globale (ECE de 0,0467 à 0,0266 en bins fixes et 0,0270 en bins adaptatifs). Les FP sont recompressés vers < 0,2 et les TP maintenus > 0,8, sans dégrader la détection brute.
- **Analyse par taille et par classe :** Erreur divisée par 2 sur les très petits objets (small_1, gain ΔECE de +0,0430 en discret et +0,0404 en continu). Correction de la sur-confiance sur "others" (T = 2,24, gain +0,0764) et de la sous-confiance sur "Sailboat" (T = 0,65, gain +0,0508).
- **Impact sur la détection (mAP) :** La méthode maintient la capacité discriminante (Resolution stable) et améliore le mAP50 sur les très petites cibles (+0,0286 sur small_1, passant de 0,7093 à 0,7380).

Tableau – Impact de la calibration croisée sur l'ECE et régimes de température T* ajustés (fixed bins/adaptive bins)

Métrique \ Échelle & Classe	GLOBAL	SIZE small_1	Class « others »	Class « Sailboat »
ECE Avant	0,0467 / 0,0414	0,0854 / 0,0707	0,1323 / 0,1039	0,1100 / 0,1105
ECE Après	0,0266 / 0,0270	0,0424 / 0,0475	0,0559 / 0,0535	0,0592 / 0,0592
Gain (ΔECE)	**+0,0201 / +0,0144**	**+0,0430 / +0,0232**	**+0,0764 / +0,0504**	**+0,0508 / +0,0513**
Reliability Avant	0,0031 / 0,0027	0,0114 / 0,0065	0,0248 / 0,0219	0,0176 / 0,0166
Reliability Après	0,0013 / 0,0012	0,0077 / 0,0045	0,0092 / 0,0055	0,0059 / 0,0058
Gain (ΔReliability)	**+0,0018 / +0,0015**	**+0,0037 / +0,0020**	**+0,0156 / +0,0164**	**+0,0117 / +0,0108**
Régime T* & Effet	Global scaling	T > 1 (Sur-conf.)	T = 2,24 (Sur-conf.)	T = 0,65 (Sous-conf.)

Note : Les valeurs de Gain (ΔECE) correspondent à la réduction absolue de l'erreur. T > 1 indique une correction de sur-confiance, T* < 1 une correction de sous-confiance. La classe « others » représente des objets de petite taille à la surface de l'eau.*

Niveau de maturité, perspectives & conclusion

Ce prototype pré-opérationnel applique une transformation ultra-légère sur les logits embarqués, sans aucun impact sur la latence des calculateurs Lerity. Il démontre des résultats très prometteurs sur données open-source. L'approche continue constitue la solution optimale retenue pour les systèmes opérationnels Lerity : elle résout l'instabilité des seuils d'alarme liée aux sauts de taille tout en divisant par 2 l'erreur globale et en améliorant le mAP des cibles éloignées. Les travaux futurs porteront sur sa qualification en conditions maritimes réelles.

Références :

[1] Lerity. https://www.lerity-alcen.com

[2] STIM – Mines Paris PSL https://www.stim.minesparis.psl.eu/

[3] D. Kumar et al., *Singapore Maritime Dataset*, 2015.

[4] T.-Y. Lin et al., *Microsoft COCO: Common Objects in Context*, ECCV 2014.

[5][6] C. Guo et al., *On Calibration of Modern Neural Networks*, ICML 2017.