

Empreinte Énergétique de YoloV8 pour la Détection de Piétons et de Véhicules

Yoann Dupas^{*,**}, Olivier Hotel^{*},
Grégoire Lefebvre^{*}, Christophe Cérin^{***}

^{*}Orange, ^{**}Université Grenoble Alpes,
^{***}Université Sorbonne Paris Nord

1 Introduction

Dans le cadre des véhicules autonomes, la perception de l'environnement joue un rôle majeur. La caméra visible est le capteur principal pour permettre une perception visuelle mais elle peut être complétée par d'autres capteurs tels que les caméras infrarouges ou LiDaR¹. Néanmoins, l'ajout de capteurs induit une augmentation de la consommation énergétique, de par l'acquisition des nouvelles données, mais aussi par le traitement des images par les modèles de réseaux de neurones. Des études proposent ainsi une évaluation de la consommation énergétique de différents modèles de réseaux de neurones [Desislavov et al. (2023)]. Cependant, il n'existe pas, à ce jour, d'étude sur l'empreinte énergétique du modèle YoloV8 [Jocher et al. (2023)]. Dans cette étude préliminaire, nous mesurons donc la précision et la consommation énergétique en entraînement et en inférence de deux configurations du modèle YoloV8 (i.e. nano et extra-large) avec des images visibles et infrarouges pour la détection de piétons et de véhicules.

2 Étude expérimentale et résultats

La détection de piétons et de véhicules peut être traitée soit sous différentes architectures de modèles (ex : sans ancrage, en une ou deux phases, basées Transformers [Amjoud et Amrouch (2023)]), soit par le traitement de différentes modalités (ex : visible, infrarouge, LiDaR). Parmi les modèles existants, nous avons retenu YoloV8 car ce modèle de bout-en-bout a l'avantage d'offrir un meilleur compromis entre rapidité d'exécution et précision de détection. Notre étude est construite sur la base de données Seeing Through The Fog [Bijelic et al. (2020)] présentant 13000 images visibles et infrarouges de scènes de conduite capturées sous différentes conditions lumineuses et météorologiques. L'empreinte énergétique du modèle YoloV8 est mesurée avec la librairie CodeCarbon [Jay et al. (2023)]. Après une recherche des hyperparamètres optimaux, nous avons entraîné les deux modèles YoloV8 (i.e. 3,2M de paramètres pour la version nano et 68,2M de paramètres pour la version extra-large) sur l'ensemble des données d'entraînement (i.e. 80% de la base de données).

1. En anglais : Light Detection and Ranging

Empreinte Énergétique de YoloV8 pour la Détection de Piétons et de Véhicules

	mAP@50 (<i>val</i> <i>test</i>)	mAP@50-95 (<i>val</i> <i>test</i>)	CE (kWh)	TE (heure)	MC ($10e-7$ kWh/img)	MTE (ms/img)
NV	63.4 65.1	35.4 35,5	0.843	4.21	14.05 ± 0.18	2.70 ± 0.72
NI	66.9 68.5	33.6 34.3	0.665	2.67	9.10 ± 0.14	2.71 ± 0.73
XV	74.5 74.5	46.5 46.6	4.980	16.35	27.04 ± 0.68	21.75 ± 0.91
XI	72.0 73.0	39.3 39.2	6.172	20.16	21.35 ± 1.04	21.97 ± 0.57

TAB. 1 – Performances des modèles YoloV8.

Les entraînements ont été effectués sur un serveur AMD² et les mesures énergétiques des inférences ont été produites sur la machine Dell Precision 5770³. Le tableau 1 présente nos résultats pour les modèles YoloV8 en version nano (N) et extra-large (X) pour les images visibles (V) et infrarouges (I). Les colonnes donnent les mesures suivantes : la précision mAP@50, la précision mAP@50-95, la consommation totale (CE), le temps d’exécution de l’entraînement (TE), la moyenne et l’écart type des consommations pour l’inférence d’une image (MC) et la moyenne et l’écart-type du temps d’inférence pour une image (MTE). Ces résultats montrent que le modèle YoloV8 entraîné sur des images visibles est plus précis que sur des images infrarouges (i.e. respectivement 74,5 contre 73,0 pour mAP@50 en test) avec des ordres de grandeurs similaires pour la consommation et le temps d’exécution pour l’entraînement et l’inférence. Les modèles extra-larges consomment 5,91 à 9,28 fois plus que les modèles nano lors de l’entraînement et consomment 1,92 à 2,35 fois plus que les modèles nano pour une inférence. Bien que la modalité infrarouge permette des détections dans des situations difficiles pour la modalité visible (ex : la nuit, sous la pluie), il existe des redondances de détections entre les modalités visible et infrarouge. Cette redondance est évaluée ici à 56,9% des détections. Diminuer ces redondances pourrait permettre donc une réduction de la consommation énergétique. Ces éléments constituent les pistes de réflexion pour nos futurs travaux.

Références

- Amjoud, A. B. et M. Amrouch (2023). Object detection using deep learning, cnns and vision transformers : A review. *IEEE Access* 11, 35479–35516.
- Bijelic, M., T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer, et F. Heide (2020). Seeing through fog without seeing fog : Deep multimodal sensor fusion in unseen adverse weather. In *2020 IEEE/CVPR*, pp. 11679–11689.
- Desislavov, R., F. Martínez-Plumed, et J. Hernández-Orallo (2023). Trends in ai inference energy consumption : Beyond the performance-vs-parameter laws of deep learning. *Sustainable Computing : Informatics and Systems* 38, 100857.
- Jay, M., V. Ostapenco, L. Lefèvre, D. Trystram, A.-C. Orgerie, et B. Fichel (2023). An experimental comparison of software-based power meters : focus on CPU and GPU. In *CCGrid*, pp. 1–13.
- Jocher, G., A. Chaurasia, et J. Qiu (2023). Ultralytics YOLOv8.

2. AMD Ryzen 7 1800X; NVIDIA GeForce RTX 2080 Ti; RAM : 64 Go DDR4

3. Intel® Core™ i7-12800H, vPro; NVIDIA RTX A3000 12 Go de mémoire GDDR6; RAM : 32 Go DDR5