Novelty Journals

ISSN 2394-7349

International Journal of Novel Research in Engineering and Science
Vol. 13, Issue 2, pp: (8-23), Month: September 2026 - February 2027, Available at: www.noveltyjournals.com

Smart-Trafic: Approche intelligente de perception visuelle, prévision à court terme et évaluation du contrôle adaptatif de la circulation à Kinshasa

Christophe Ebaka Bongelo[1]*, Mukala Patrick[1], Nsenge Mpia Héritier[2], Telesphorus Nsumbu Lukamba[1], Blaise Kapalala Kapenda[3], Yves Nsota Olang[4], Emmanuel Moyiba Mayayi[5]

[1]Université Pédagogique Nationale (UPN), Mention de Mathématiques, Informatique et Statistique, Kinshasa, République démocratique du Congo.

[2]Université de l'Assomption au Congo (UAC), Butembo, République Démocratique du Congo.

[3]Institut supérieur pédagogique (ISP) Gombe, Kinshasa, République Démocratique du Congo.

[4]Institut supérieur de Commerce (ISC) d'Idiofa, Idiofa, République Démocratique du Congo.

[5]Institut supérieur pédagogique (ISP) Yumbi, Yumbi, République démocratique du Congo.

DOI: https://doi.org/10.5281/zenodo.23801929

Published Date: 30-October-2026

Résumé : À Kinshasa, la circulation routière réunit des véhicules et des usagers très différents dans des espaces où les occlusions, les déplacements désordonnés et la diversité des points de vue compliquent l'observation automatique. Nous étudions Smart-Trafic, une chaîne expérimentale associant détection visuelle, suivi multi-objet, construction d'indicateurs temporels, prévision à court terme et comparaison de stratégies de signalisation. Un détecteur YOLOv8n, adapté à sept catégories locales, est évalué sur un ensemble de 358 images et 482 annotations, dont 56 images et 67 objets réservés au test. Il obtient une précision de 0,5500, un rappel de 0,3502 et une mAP@0,50:0,95 de 0,1803. L'ablation met en évidence, dans les conditions expérimentales retenues, l'intérêt du pré-entraînement et des augmentations de données. Pour le score de pression calculé à partir des vidéos, la persistance présente la plus faible MAE à une minute ; le LSTM obtient les erreurs absolues les plus faibles à trois et cinq minutes, mais son R^2 à cinq minutes demeure négatif (−0,2184). Dans le modèle de files agrégées, l'heuristique prédictive est la plus performante ; en microsimulation SUMO/TraCI, PPO réduit par rapport au feu fixe la file moyenne de 44,79 %, l'attente de 79,65 % et l'écart d'équité de 41,94 %, au prix d'une diminution du débit de 11,11 %. Ces résultats éclairent les compromis entre qualité de perception, prévision et décisions de contrôle. Ils établissent une faisabilité expérimentale, et non une efficacité démontrée en situation routière réelle.

Mots-clés : systèmes de transport intelligents ; Kinshasa ; YOLOv8n ; ByteTrack ; prévision de trafic ; apprentissage par renforcement ; PPO ; SUMO ; robustesse expérimentale.

Novelty Journals

ISSN 2394-7349

International Journal of Novel Research in Engineering and Science
Vol. 13, Issue 2, pp: (8-23), Month: September 2026 - February 2027, Available at: www.noveltyjournals.com

Abstract: **Urban traffic in Kinshasa involves heterogeneous road users, frequent occlusions, and limited ground-truth data, which complicate the use of computer vision for traffic monitoring. This study examines Smart-Trafic, an experimental pipeline linking object detection, multi-object tracking, short-term forecasting, and simulated traffic-signal control. A YOLOv8n detector fine-tuned for seven locally relevant classes was evaluated on 358 images containing 482 annotations, with 56 images and 67 instances reserved for testing. Test precision, recall, and mAP@0.50:0.95 were 0.5500, 0.3502, and 0.1803, respectively. Ablation results indicate descriptive benefits from transfer learning, data augmentation, and a 640-pixel input resolution. For a video-derived congestion-pressure proxy, persistence achieved the lowest one-minute mean absolute error; LSTM had the lowest errors at three and five minutes, although its five-minute R^2 remained negative (−0.2184). A predictive heuristic ranked first in the aggregate-queue analytical environment. In a separate SUMO/TraCI evaluation, PPO reduced mean queues by 44.79%, waiting time by 79.65%, and a fairness-gap indicator by 41.94% relative to fixed-time control, while throughput fell by 11.11%. The main contribution is an integrated, resource-aware experimental framework that makes trade-offs between perception reliability, forecasting, and control explicit. No effectiveness in real-world intersection operations is claimed.**

Keywords: **heterogeneous traffic; Kinshasa; computer vision; YOLOv8n; ByteTrack; traffic forecasting; PPO; SUMO; resource-constrained machine learning.**

1. INTRODUCTION

1.1. Contexte et problème de recherche

Les difficultés de circulation à Kinshasa ne tiennent pas uniquement au nombre de véhicules. Sur un même axe, voitures particulières, minibus, motos, tricycles, camions, chariots et piétons partagent souvent un espace dont les usages évoluent au fil de la journée. Pour un dispositif fondé sur des caméras, cette diversité introduit des occlusions fréquentes, des objets de petite taille et des ambiguïtés entre catégories. La croissance urbaine et les contraintes d'infrastructure renforcent l'intérêt d'outils capables de produire des observations localisées et vérifiables (JICA, 2019, 2024). Pourtant, observer un trafic aussi hétérogène ne se réduit pas à transférer sans précaution un modèle entraîné sur des scènes routières standardisées.

Une chaîne de gestion intelligente repose sur plusieurs inférences successives. Un objet manqué à la détection peut disparaître du comptage ; un comptage biaisé peut altérer les séries temporelles ; une prévision imparfaite peut, à son tour, modifier l'évaluation d'une politique de feux. Les travaux sur la détection, le suivi et l'apprentissage par renforcement apportent des outils utiles, mais leurs résultats isolés ne démontrent pas la fiabilité du système complet. Nous examinons donc chaque module avec ses propres critères, puis les compromis produits lorsque ces modules sont réunis dans une expérimentation cohérente.

1.2. Question, objectif et périmètre

La question étudiée est la suivante : dans quelle mesure des vidéos du réseau routier de Kinshasa, exploitées sur un ordinateur grand public, permettent-elles de construire une chaîne de perception, de prévision et de contrôle simulé dont les sorties restent interprétables et traçables ? L'objectif est de concevoir et d'évaluer une preuve de concept adaptée à des ressources de calcul limitées. Nous distinguons ainsi les événements produits par les algorithmes, les quantités directement mesurées et les effets estimés par simulation ; seule cette distinction permet de juger correctement de la portée des résultats.

1.3. Contributions et portée de la démonstration

Cette étude apporte quatre éléments complémentaires. Elle définit d'abord une taxonomie de sept classes, adaptée à la présence locale des motos, des tricycles et des chariots. Elle associe ensuite YOLOv8n à ByteTrack et à une règle de franchissement, tout en évaluant séparément la détection et la production des événements de comptage. Elle compare également une référence par persistance à trois prédicteurs appris, pour des horizons de une, trois et cinq minutes. Enfin, elle confronte quatre politiques de signalisation dans un modèle analytique puis dans SUMO, afin de rendre visibles les changements de classement induits par la dynamique du simulateur. L'ensemble est exécuté dans une configuration matérielle accessible, autour d'une carte GTX 1650 de 4 Go.

Le périmètre est délibérément expérimental. Le test de détection reste de petite taille ; les franchissements calculés n'ont pas été confrontés à un inventaire manuel complet ; les prévisions concernent des indicateurs dérivés des vidéos ; et la validation microscopique porte sur deux sessions distinctes d'un même axe routier. Ces éléments limitent les conclusions générales, mais permettent d'étudier avec précision où une architecture intégrée fonctionne et où l'incertitude demeure.

2. TRAVAUX CONNEXES ET POSITIONNEMENT

2.1. Détection d'objets dans des scènes routières hétérogènes

Les architectures YOLO ont contribué à rendre la détection d'objets compatible avec des contraintes de latence fortes (Redmon et al., 2016). Leur évolution a amélioré le traitement multi-échelle et l'apprentissage de représentations visuelles, sans supprimer la dépendance aux données d'entraînement (Terven et al., 2023). Des applications de YOLOv8 au trafic confirment l'intérêt de ces architectures pour la surveillance visuelle (Bakirci, 2024a ; You et al., 2024). Leur précision ne se transfère toutefois pas automatiquement à des scènes présentant d'autres classes, d'autres angles de prise de vue ou des conditions d'occultation plus marquées.

La définition des catégories mérite donc autant d'attention que le choix du détecteur. Une moto à deux roues et un moto-tricycle peuvent avoir des caractéristiques visuelles proches, alors qu'ils correspondent à des usages et à des dimensions différentes dans la circulation. Akujobi et al. (2026) montrent l'intérêt de catégories spécifiques pour le comptage de tricycles au Nigeria. Les analyses menées sur des images de transports collectifs à Nairobi et Kigali soulignent également l'effet de la densité et des variations d'image sur les algorithmes (Kyatha et Taneja, 2025). Ces observations étayent l'emploi d'une ontologie locale, sans dispenser d'une vérification des confusions entre catégories.

2.2. Du suivi visuel à la mesure du trafic

Le suivi multi-objet relie les détections d'une image à l'autre afin de construire des trajectoires. ByteTrack réutilise notamment des détections de confiance plus faible pour limiter la fragmentation des pistes pendant les occultations (Zhang et al., 2022). Le franchissement d'une ligne virtuelle transforme ensuite certaines trajectoires en événements directionnels. Il importe toutefois de ne pas confondre événement algorithmique et passage réellement observé : une classe mal attribuée, un changement d'identifiant ou une détection manquante peut affecter le résultat final.

2.3. Prévision et contrôle de la signalisation

Pour prévoir une série temporelle courte, la persistance constitue un point de comparaison indispensable. Les forêts aléatoires (Breiman, 2001), les méthodes de boosting et les réseaux LSTM (Hochreiter et Schmidhuber, 1997) peuvent capter des relations non linéaires, à condition de disposer de données appropriées et d'un protocole temporel sans fuite d'information. S'agissant des feux, PPO optimise une politique à partir d'interactions avec un environnement (Schulman et al., 2017) ; sa performance dépend fortement de la récompense et des contraintes de contrôle. SUMO modélise les véhicules individuellement, offrant une description plus riche qu'un système de files agrégées, mais sa proximité avec la réalité dépend de sa calibration (Alvarez López et al., 2018).

2.4. Lacune scientifique spécifique

Le défi principal ne réside pas nécessairement dans la conception d'un nouveau réseau neuronal, mais dans l'articulation vérifiable de modules déjà éprouvés. Les progrès obtenus en apprentissage léger pour d'autres tâches, par exemple l'architecture embarquée de Mpia et al. (2025), illustrent l'intérêt d'une conception attentive aux ressources de calcul ; leurs performances ne sauraient toutefois être extrapolées au trafic routier. Notre positionnement privilégie ainsi l'adéquation aux classes locales, la traçabilité des sorties, la comparaison à des méthodes simples et l'analyse critique du passage d'un modèle agrégé à une microsimulation.

3. MATERIELS ET METHODES

3.1. Dessin de l'étude et séparation des expériences

L'expérimentation est structurée en quatre volets : adaptation d'un détecteur sur des images annotées, extraction de trajectoires et de franchissements à partir de vidéos, prévision d'indicateurs temporels et évaluation de politiques de signalisation. Les composants partagent une architecture logicielle commune sans être confondus avec une unique

validation de terrain. La figure 1 présente l'organisation fonctionnelle, tandis que la figure 2 suit le cheminement des données depuis les vidéos jusqu'au contrôle simulé. Cette lecture en deux niveaux aide à séparer les responsabilités des modules et leurs interfaces.

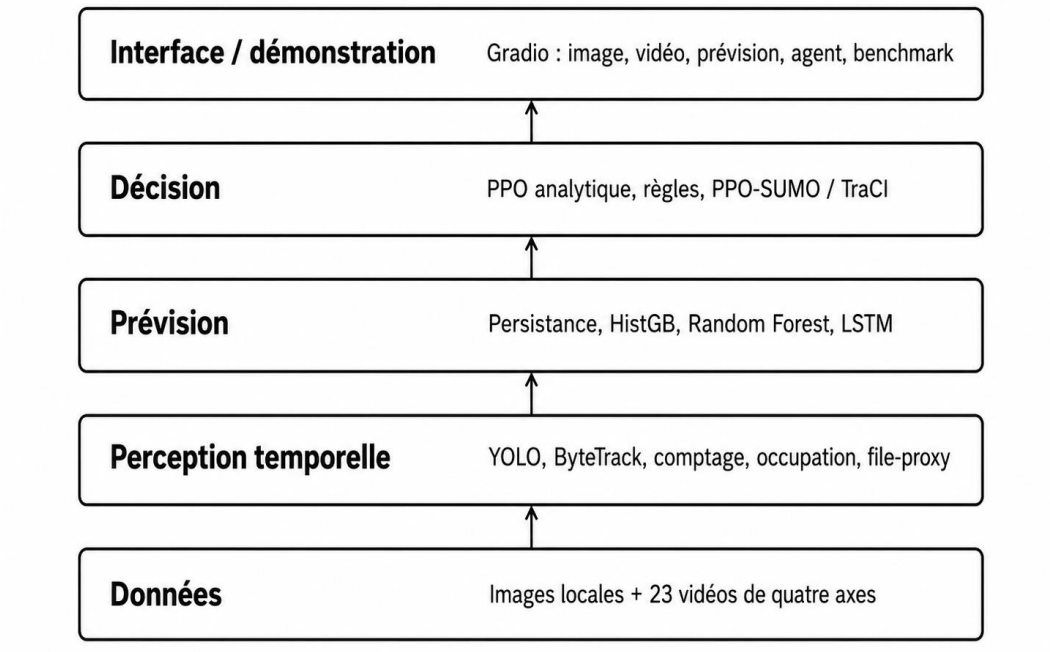


Figure 1. Architecture fonctionnelle de Smart-Trafic, organisée en cinq couches.

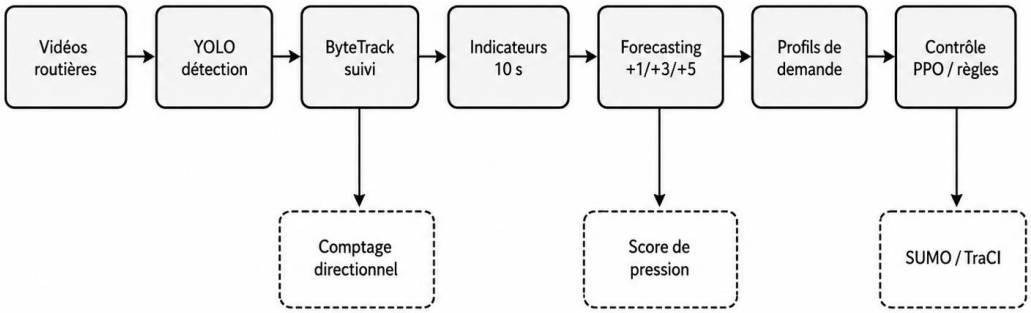


Figure 2. Enchaînement expérimental de la vidéo aux décisions simulées.

3.2. Sources vidéo, corpus annoté et ontologie

Les séquences vidéo étudiées couvrent quatre grands axes de Kinshasa : le boulevard du 30 Juin, l'avenue Bypass, l'avenue de la Libération et la route des Poids Lourds. Le corpus comprend 23 vidéos, représentant environ trois heures et quarante-neuf minutes d'enregistrement. Il s'agit d'un échantillon de situations de circulation, non d'un tirage représentatif de l'ensemble du réseau. Les différences d'emplacement, de caméra et de période d'observation font partie des sources possibles de variation que nous prenons en compte dans l'interprétation des performances.

L'ensemble dédié à la détection réunit 358 images et 482 boîtes annotées selon sept catégories : chariot_pousse_pousse, voiture, pieton, moto_2_roues, moto_tricycle, minibus_bus et camion. Le contrôle des fichiers n'a révélé ni paire image/annotation orpheline ni annotation syntaxiquement invalide. Cette vérification technique ne permet cependant pas de garantir que tous les objets visibles ont été annotés, notamment dans les scènes très chargées. Le tableau 1 indique la répartition entre entraînement, validation et test ; la rareté de certaines classes, en particulier le camion, reste une limite de l'échantillon.

Novelty Journals

ISSN 2394-7349

International Journal of Novel Research in Engineering and Science
Vol. 13, Issue 2, pp: (8-23), Month: September 2026 - February 2027, Available at: www.noveltyjournals.com

Tableau 1. Répartition du corpus de détection entre les sous-ensembles

Sous-ensemble	Images	Objets annotés	Fonction
Entraînement	252	348	Apprentissage des poids
Validation	50	67	Sélection et réglage
Test	56	67	Évaluation finale

Les 56 images du test ont été réservées à l'évaluation finale, tandis que l'ensemble de validation a servi au choix du checkpoint. Cette séparation évite l'utilisation directe des images de test pour l'ajustement des poids. En revanche, faute d'une documentation complète des liens entre images, vidéos et sessions, le risque de quasi-doublons entre partitions ne peut être définitivement écarté. Les résultats correspondent donc à ce découpage précis et ne constituent pas une estimation exhaustive de la généralisation à de nouveaux sites.

3.3. Détecteur, réglages et ablation

Le modèle principal est YOLOv8n, initialisé avec des poids issus de COCO puis adapté aux sept catégories. Les images sont redimensionnées à 640 × 640 pixels ; la taille de lot est de 8, la graine aléatoire de 42, le nombre de workers de 4 et le calcul GPU exécuté sur CUDA 0. Un maximum de 100 époques et une patience de 20 sont fixés pour l'arrêt précoce. L'entraînement s'interrompt après 72 époques, avec un meilleur checkpoint à l'époque 52. Les calculs ont été réalisés sous Windows 11 sur une GeForce GTX 1650 disposant de 4 Go de mémoire vidéo. L'étude d'ablation constitue un second ensemble d'essais, limité à 50 époques par configuration.

L'ablation fait varier séparément l'initialisation, les augmentations et la résolution. Les quatre conditions sont : poids pré-entraînés à 640 pixels avec augmentations ; entraînement depuis zéro avec les mêmes paramètres ; poids pré-entraînés à 640 pixels sans augmentation ; et poids pré-entraînés à 512 pixels avec augmentation. Les scores sont comparés sur le même découpage de test. Comme une seule série de résultats est disponible pour chaque condition, les différences décrivent des effets observés et non des effets assortis d'une incertitude statistique estimée.

3.4. Critères d'évaluation de la perception et du comptage

Nous évaluons la détection à l'aide de la précision, du rappel et du score F1, où TP désigne les vrais positifs, FP les faux positifs et FN les faux négatifs. Le recouvrement entre une boîte prédite et sa référence est mesuré par l'intersection sur union (IoU). Les relations utilisées sont les suivantes :

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

$$F_1 = \frac{2PR}{P + R}$$

$$IoU = \frac{aire(Bp \cap Br)}{aire(Bp \cup Br)}$$

La mAP@0,50 est calculée pour un seuil d'IoU de 0,50 ; la mAP@0,50:0,95 agrège les seuils compris entre 0,50 et 0,95, par pas de 0,05. Ces scores caractérisent la détection, non la continuité des trajectoires ni la justesse des franchissements.

Pour le comptage, une ligne verticale est placée au milieu de l'image normalisée (x = 0,50), avec une bande de tolérance relative de 0,035 afin de limiter les franchissements répétés dus aux oscillations des boîtes. Une trajectoire doit comporter au moins trois images pour être prise en compte ; le sens dépend de la direction du franchissement. La figure 3 illustre cette géométrie. Nous considérons les nombres retournés comme des événements détectés, sans les assimiler à une vérité terrain de comptage. La comparaison à des passages annotés manuellement permettrait ensuite de calculer des erreurs par classe et par direction.

Novelty Journals

ISSN 2394-7349

International Journal of Novel Research in Engineering and Science
Vol. 13, Issue 2, pp: (8-23), Month: September 2026 - February 2027, Available at: www.noveltyjournals.com

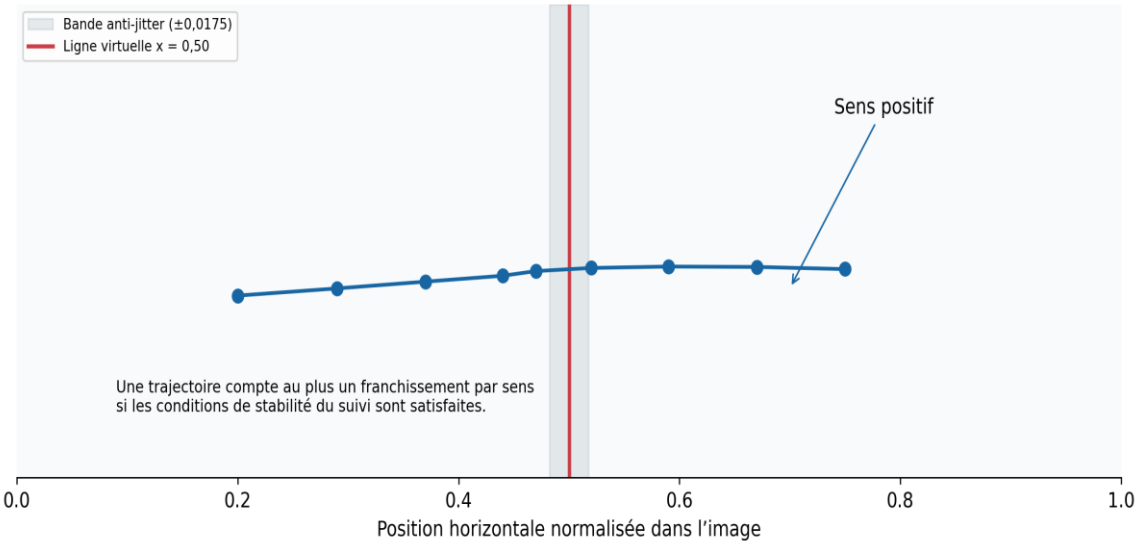


Figure 3. Ligne virtuelle de franchissement et bande de tolérance anti-oscillation.

3.5. Construction des séries et protocole de prévision

Les trajectoires produites par la chaîne visuelle sont agrégées par fenêtres de dix secondes. Nous construisons des cibles à une, trois et cinq minutes, dont new_tracks, mean_queue_proxy, mean_occupancy_ratio et un score synthétique de pression. Ces variables sont des indicateurs indirects calculés à partir de la vidéo : elles ne représentent pas, sans étalonnage supplémentaire, un débit de référence, une longueur de file exprimée en mètres ou une pression physique mesurée sur la chaussée.

La persistance, HistGradientBoosting, Random Forest et LSTM sont comparés selon les horizons. La construction de la base produit respectivement 1 002, 738 et 484 exemples exploitables, dont 235, 175 et 115 cas test. Les observations sont construites dans l'ordre temporel, chaque cible étant située après sa fenêtre d'entrée. Le protocole conserve une séparation chronologique des observations ; la durée précise des fenêtres d'entrée et les restrictions entre sessions ne sont toutefois pas toutes renseignées. Les résultats doivent, par conséquent, être interprétés dans les limites de ce découpage, sans présumer une généralisation à d'autres journées de circulation.

Pour comparer des erreurs exprimées dans les mêmes unités que les cibles, nous retenons l'erreur absolue moyenne (MAE) et la racine de l'erreur quadratique moyenne (RMSE). Le coefficient R^2 mesure l'écart à une prédiction constante fondée sur la moyenne du test ; le WAPE rapporte la somme des erreurs absolues à la somme des valeurs absolues observées. Pour n observations, ces métriques s'écrivent:

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i|$$

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2}$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

$$WAPE = \frac{\sum |y_i - \hat{y}_i|}{\sum |y_i|}$$

Dans ces expressions, y_i est la valeur observée, $\hat{y}_i$ la prévision, $\bar{y}$ la moyenne des observations test et n leur nombre. Le WAPE est indéfini lorsque la somme des valeurs absolues observées est nulle ; un R^2 négatif indique qu'un modèle est moins performant que la référence constante fondée sur cette moyenne au sens de l'erreur quadratique.

Novelty Journals

ISSN 2394-7349

International Journal of Novel Research in Engineering and Science
Vol. 13, Issue 2, pp: (8-23), Month: September 2026 - February 2027, Available at: www.noveltyjournals.com

3.6. Politiques de contrôle et validité du simulateur

Quatre stratégies sont étudiées : un plan de feux fixe, une politique réactive, une heuristique fondée sur la demande prévue et un agent Proximal Policy Optimization (PPO). Elles sont d'abord comparées dans un système de files agrégées sur 15 épisodes, puis dans SUMO/TraCI sur 10 épisodes. Nous examinons la file moyenne, le pic de file, l'attente, le débit, une mesure d'écart d'équité et la récompense du contrôleur. Réduire l'attente ou les files est favorable, mais une baisse simultanée du débit demeure un coût opérationnel ; les objectifs doivent donc être lus conjointement. Certains paramètres internes de la fonction de récompense et des indicateurs agrégés ne sont pas disponibles dans l'évaluation présentée ; les résultats sont donc comparés en priorité sur des métriques descriptives communes.

Pour la microsimulation, PPO est entraîné sur la session VID_Bypass_20260624_125718_train.csv, puis évalué, après gel des paramètres, sur VID_Bypass_20260727_163318_test.csv. Les deux fichiers proviennent de sessions distinctes du même axe Bypass. Cette organisation permet un test hors session d'entraînement, sans constituer une évaluation intersites. Le réseau SUMO reproduit une dynamique de véhicules individuels ; il n'a cependant pas été calibré sur des mesures indépendantes de mouvements tournants d'un carrefour réel, et l'ensemble des réglages de phases, comportements et transitions ne peut être reconstitué à partir des résultats rapportés.

3.7. Environnement d'exécution et traçabilité

Le dispositif logiciel associe Python, PyTorch, Ultralytics, OpenCV, scikit-learn, Gymnasium, Stable-Baselines3, SUMO/TraCI et Gradio. L'architecture a été choisie pour pouvoir exécuter perception, suivi et expérimentation sur un ordinateur portable plutôt que sur une station de calcul spécialisée. Les valeurs de latence présentées plus loin concernent l'inférence rapportée par l'outil d'évaluation du détecteur ; elles ne doivent pas être confondues avec le temps complet d'acquisition, d'agrégation, de suivi et d'affichage. La reproductibilité exacte dépend également de l'étude des bibliothèques, de la configuration matérielle et de l'état des données.

4. RESULTATS

4.1. Composition du corpus et courbe d'apprentissage

La composition du corpus est fortement déséquilibrée : 135 motos à deux roues, 96 moto-tricycles, 94 voitures, 69 minibus ou bus, 45 piétons, 35 chariots et 8 camions. La figure 4 en donne la distribution. Cette concentration d'exemples dans quelques classes impose de regarder les performances par catégorie, et non le seul score global. La figure 5 retrace par ailleurs les fluctuations du mAP sur l'ensemble de validation, qui ont conduit à retenir les poids de l'époque 52.

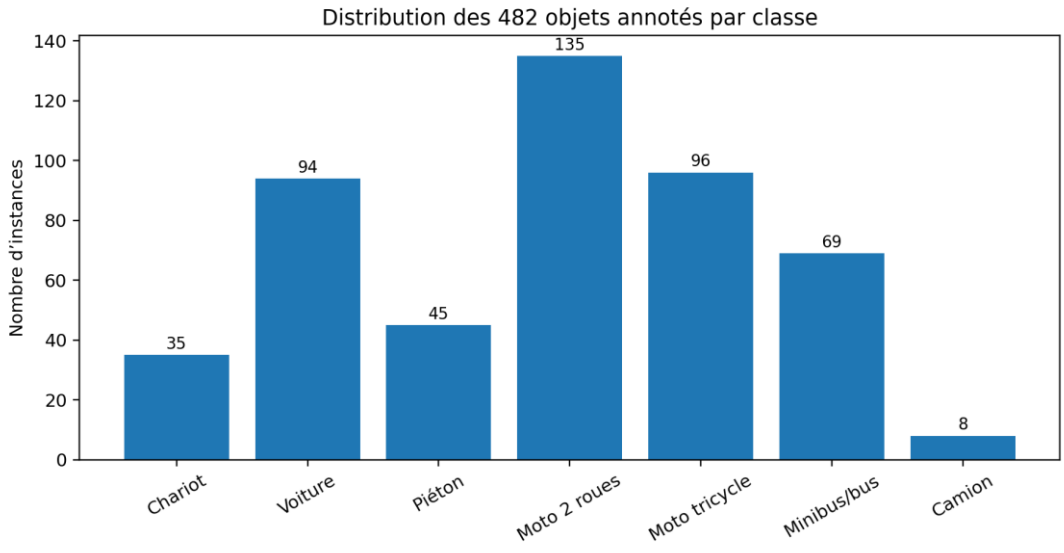


Figure 4. Répartition des 482 annotations entre les sept catégories.

Novelty Journals

ISSN 2394-7349

International Journal of Novel Research in Engineering and Science
Vol. 13, Issue 2, pp: (8-23), Month: September 2026 - February 2027, Available at: www.noveltyjournals.com

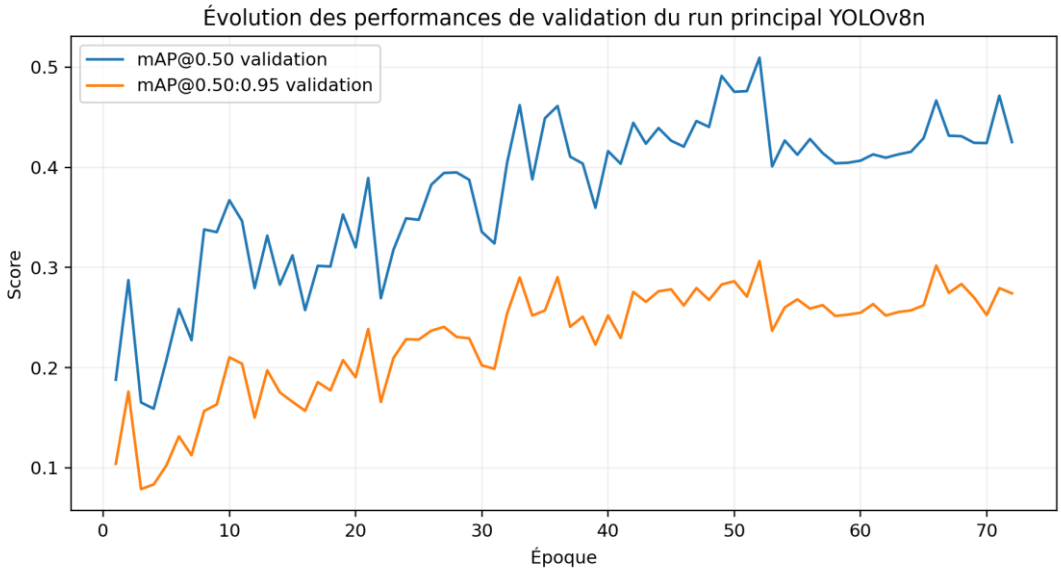


Figure 5. Évolution des métriques de validation au cours de l'entraînement de YOLOv8n.

Le journal d'entraînement totalise 72 époques et 621,239 secondes de calcul, soit environ 10,35 minutes. Au checkpoint de l'époque 52, la précision de validation atteint 0,6955, le rappel 0,4141, la mAP@0,50 0,5095 et la mAP@0,50:0,95 0,3065. La fluctuation des courbes souligne la sensibilité de cette phase au faible nombre d'images de validation. Ces résultats ont servi à sélectionner le modèle ; ils sont distincts de l'évaluation finale sur le test réservé.

4.2. Performance finale du détecteur

Les performances sur les 56 images et les 67 objets du test sont inférieures aux valeurs de validation. Le tableau 2 rassemble les principales mesures, dont la mAP@0,50:0,95 de 0,1803 et le temps d'inférence d'environ 10,16 ms par image. La différence entre validation et test suggère une variabilité importante selon les scènes, sans permettre d'en attribuer la cause à un seul facteur.

Tableau 2. Performances globales du détecteur sur le test réservé

Mesure	Valeur	Lecture
Précision	0,5500	Exactitude conditionnelle des prédictions positives
Rappel	0,3502	Sensibilité limitée sur le test
mAP@0,50	0,3258	Détection à IoU 0,50
mAP@0,75	0,2078	Localisation plus exigeante
mAP@0,50:0,95	0,1803	Robustesse de localisation limitée
Temps d'inférence annoncé	10,16 ms/image	Latence du rapport ; pas un FPS de pipeline

Avec un rappel global de 0,3502, le détecteur manque une fraction importante des objets annotés. Cette limite est particulièrement sensible pour le comptage : une trajectoire ne peut être reconstruite de manière fiable si les observations sont trop intermittentes. La différence entre la mAP à 0,50 et celle moyennée de 0,50 à 0,95 montre également que la localisation des boîtes se dégrade à mesure que le critère de recouvrement devient plus strict. Ces scores proviennent du protocole d'évaluation du modèle et ne correspondent pas directement à un pourcentage de passages routiers correctement comptés.

4.3. Hétérogénéité des classes

Le tableau 3 détaille les performances obtenues pour chacune des sept catégories, afin de mettre en évidence les erreurs masquées par les moyennes globales.

Tableau 3. Résultats de détection par classe sur le test réservé

Classe	n	Précision	Rappel	F1	mAP50	mAP50–95
Chariot	7	1,0000	0,0000	0,0000	0,1445	0,0461
Voiture	13	0,1206	0,1538	0,1352	0,1529	0,1019
Piéton	8	0,2612	0,3750	0,3079	0,2005	0,1075
Moto 2 roues	9	0,3768	0,4444	0,4078	0,3822	0,1197
Moto-tricycle	18	0,7005	0,7778	0,7371	0,8136	0,5014
Minibus/bus	10	0,3908	0,7000	0,5016	0,5825	0,3835
Camion	2	1,0000	0,0000	0,0000	0,0043	0,0017

Le tableau 3 et la figure 6 révèlent des performances très hétérogènes. Le moto-tricycle atteint la meilleure mAP@0,50 (0,8136) sur 18 objets de test, alors que les motos à deux roues atteignent 0,3822 et les voitures 0,1529. L'analyse visuelle des vidéos met pourtant en évidence des attributions erronées de certaines motos à la classe des tricycles. La performance obtenue sur des images fixes ne suffit donc pas à garantir une classification fiable dans les séquences suivies. Les deux camions du test présentent un rappel nul ; la valeur de précision de 1,0000 affichée par le logiciel, associée à ce rappel nul, ne doit pas être lue comme une détection parfaite.

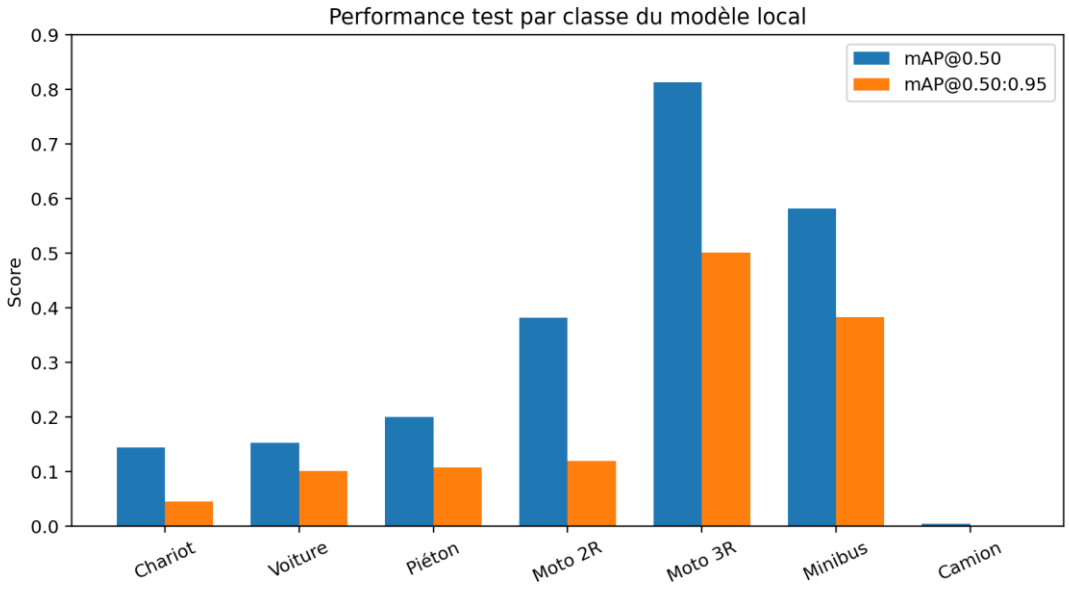


Figure 6. Performances de détection par catégorie sur le test réservé.

4.4. Effet des choix d'entraînement : étude d'ablation

Le tableau 4 regroupe les quatre configurations de l'ablation et leurs trois métriques de mAP.

Tableau 4. Étude d'ablation de YOLOv8n : résultats sur le test

Configuration	mAP@0,50	mAP@0,50:0,95	mAP@0,75
Pré-entraîné · 640 · augmentations	0,3738	0,2220	0,2310
Depuis zéro · 640 · augmentations	0,2656	0,1379	0,1289
Pré-entraîné · 640 · sans augmentation	0,2980	0,1868	0,2299
Pré-entraîné · 512 · augmentations	0,3266	0,1967	0,2130

Novelty Journals

ISSN 2394-7349

International Journal of Novel Research in Engineering and Science
Vol. 13, Issue 2, pp: (8-23), Month: September 2026 - February 2027, Available at: www.noveltyjournals.com

Les résultats de l'ablation sont réunis dans le tableau 4 et représentés dans la figure 7. Le pré-entraînement améliore la mAP@0,50:0,95 de 0,1379 à 0,2220, soit 0,0841 point absolu. Dans le même protocole, les augmentations correspondent à un écart de 0,0352 point, et le passage de 512 à 640 pixels à 0,0253 point. Ces observations appuient l'intérêt du transfert de représentations lorsqu'un petit corpus est disponible. Elles ne permettent pas d'établir la stabilité des gains entre graines aléatoires, et le score de la configuration de référence à 50 époques n'est pas interchangeable avec celui du modèle principal soumis à l'arrêt précoce.

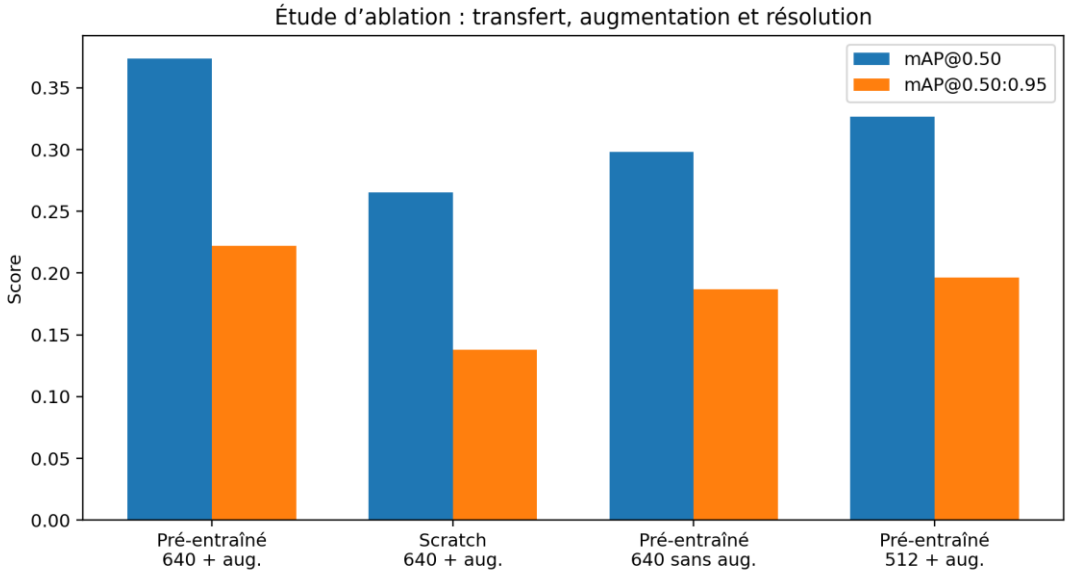


Figure 7. Effet descriptif du pré-entraînement, des augmentations et de la résolution.

4.5. Détection en interface et franchissements

La figure 8 donne un exemple d'utilisation de l'interface Gradio. Dans cette scène, le modèle identifie un minibus ou bus avec une confiance d'environ 0,88 et une moto à deux roues avec une confiance proche de 0,64. D'autres usagers présents dans l'image restent non détectés. Cet exemple illustre le fonctionnement de l'interface et les difficultés de perception, mais ne constitue pas une mesure de précision indépendante.

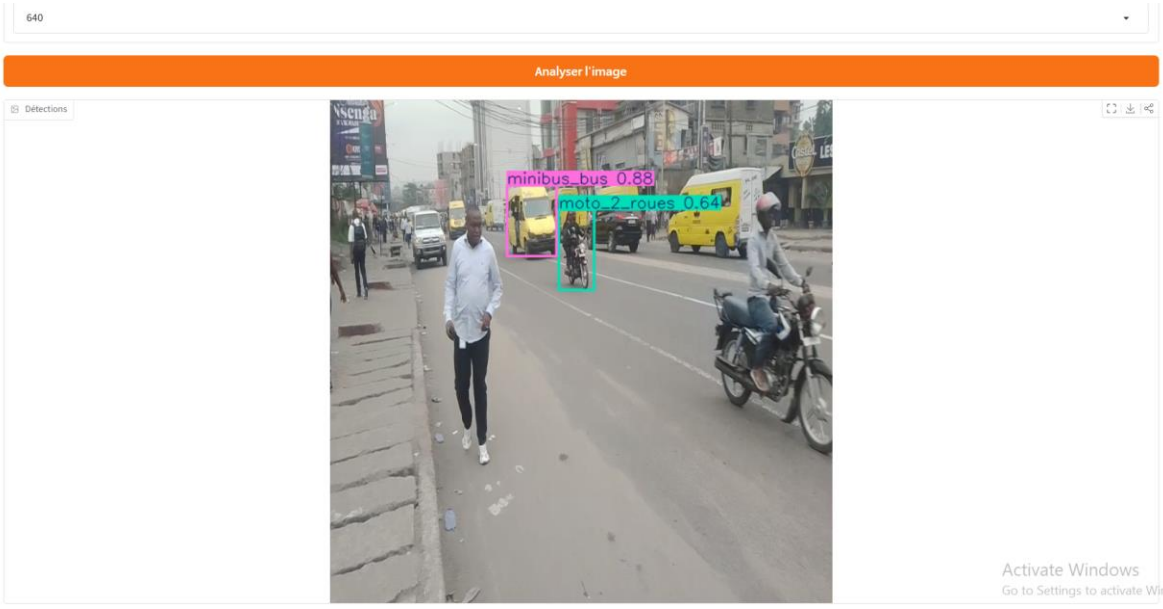


Figure 8. Exemple de détection d'usagers dans l'interface Gradio.

Le contrôle préliminaire associant YOLOv8n pré-entraîné sur COCO à ByteTrack a produit 2 656 détections cumulées sur 210 images, avec deux franchissements de voitures vérifiés sur un court extrait. Un essai de franchissement plus long produit 130 événements : 97 dans le sens positif et 33 dans le sens négatif. La distribution calculée est de 2 chariots, 11 voitures, 2 piétons, 33 motos à deux roues, 72 tricycles, 7 minibus ou bus et 3 camions. La figure 9 présente la sortie de l'interface. En l'absence d'annotations manuelles de référence pour cette séquence, ces effectifs sont des comptages logiciels, non des estimations validées du flux réel.



Figure 9. Franchissements calculés, ventilés par catégorie et par sens.

La part élevée de tricycles dans ces événements attire l'attention, car les essais visuels révèlent justement des confusions avec les motos à deux roues. Le biais peut provenir de la reconnaissance des classes, de l'association des trajectoires ou de leur interaction. Une comparaison événement par événement à un comptage humain permettrait de distinguer ces hypothèses. En l'état, les 130 franchissements attestent que le mécanisme de suivi et de comptage est opérationnel sur le plan logiciel, non qu'il mesure exactement 130 passages.

4.6. Prévision du score de pression

Pour comparer les méthodes de prévision sur une base commune, le tableau 5 présente leurs erreurs et, lorsqu'ils sont disponibles, leurs coefficients R^2 et WAPE.

Tableau 5. Prévision du score de pression aux trois horizons.

Horizon	Modèle	MAE	RMSE	R^2	WAPE
+1 min	Persistance	0,0401	0,0550	0,2041	0,2328
+1 min	HistGradientBoosting	0,0498	—	—	—
+1 min	Random Forest	0,0479	—	—	—
+1 min	LSTM	0,0493	—	—	—
+3 min	Persistance	0,0532	—	—	—
+3 min	HistGradientBoosting	0,0415	—	0,0475	—
+3 min	Random Forest	0,0426	—	0,0287	—
+3 min	LSTM	0,0412	—	0,0150	—
+5 min	Persistance	0,0593	—	−1,8219	—
+5 min	HistGradientBoosting	0,0457	—	−0,4490	—
+5 min	Random Forest	0,0451	—	−0,4034	—
+5 min	LSTM	0,0412	—	−0,2184	—

Novelty Journals

ISSN 2394-7349

International Journal of Novel Research in Engineering and Science
Vol. 13, Issue 2, pp: (8-23), Month: September 2026 - February 2027, Available at: www.noveltyjournals.com

Le tableau 5 rassemble les résultats pour le score de pression. À une minute, la persistance présente la plus petite MAE (0,0401), ce qui en fait une référence difficile à dépasser à cet horizon. À trois minutes, le LSTM obtient la plus faible MAE (0,0412), mais son R^2 demeure proche de zéro (0,0150) ; le boosting présente, sur ce critère, un R^2 légèrement supérieur (0,0475). À cinq minutes, le LSTM conserve la plus faible MAE (0,0412), avec un R^2 négatif (−0,2184). Ainsi, une erreur absolue favorable ne signifie pas nécessairement que les variations de la série sont bien expliquées. La figure 10 facilite la comparaison des MAE d'un horizon à l'autre.

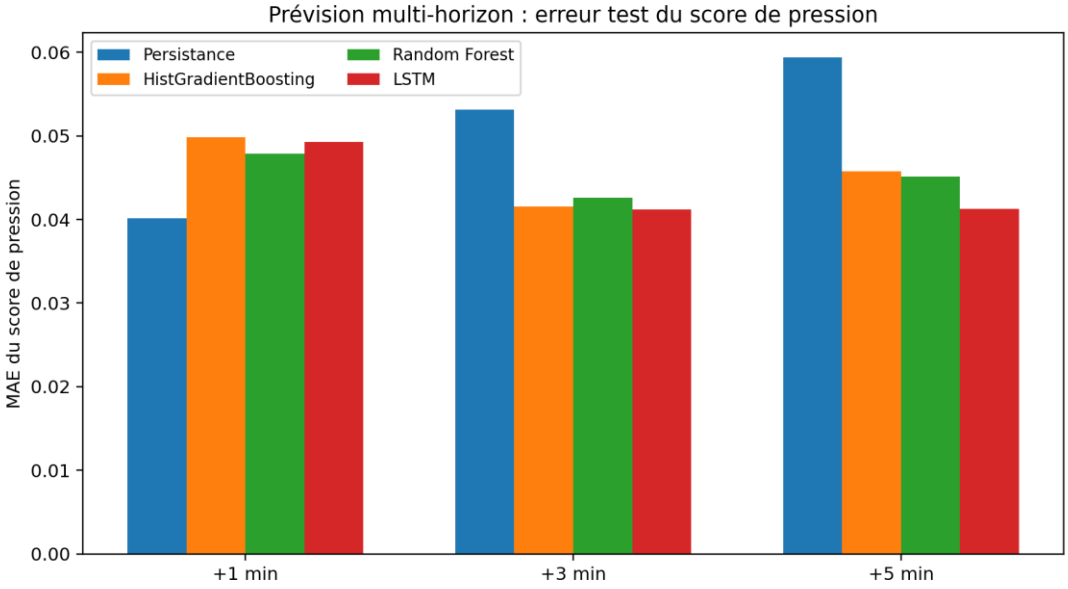


Figure 10. Comparaison des erreurs absolues moyennes du score de pression selon l'horizon.

Les autres indicateurs présentent des comportements différents. Pour new_tracks, Random Forest atteint des MAE de 14,64, 15,13 et 19,11 à une, trois et cinq minutes. Pour mean_queue_proxy, les meilleurs résultats descriptifs sont respectivement obtenus par HistGradientBoosting (0,2423), Random Forest (0,2647) et HistGradientBoosting (0,3094). Pour mean_occupancy_ratio, les plus petites MAE sont celles de la persistance (0,1170), de HistGradientBoosting (0,1316) et de Random Forest (0,1333). Ces indicateurs restent tributaires de la caméra, du détecteur et de la définition des proxys ; ils ne remplacent pas des mesures de terrain calibrées.

4.7. Comparaison dans l'environnement analytique

Les résultats des quatre contrôleurs dans le modèle à files agrégées sont consignés dans le tableau 6.

Tableau 6. Comparaison des politiques dans l'environnement analytique (15 épisodes).

Politique	File moy.	File max.	Attente cumulée	Débit	Arrêts	Récompense
Fixe	52,49	104,04	33 368,13	772,90	650,90	−18,45
PPO	12,48	39,66	7 973,90	838,29	568,62	7,53
Heuristique prédictive	3,02	13,85	1 919,13	872,17	501,16	14,16
Réactive	8,26	23,02	5 264,44	868,93	618,88	10,65

Le tableau 6 compare les quatre stratégies dans l'environnement analytique. L'heuristique prédictive produit la file moyenne la plus faible (3,02), l'attente cumulée la plus courte (1 919,13) et le débit le plus élevé (872,17) parmi les politiques étudiées. PPO améliore largement le plan fixe, mais reste derrière la règle réactive et l'heuristique pour plusieurs critères essentiels. Le résultat rappelle qu'une stratégie explicite peut offrir un compromis particulièrement favorable dans un modèle dont les hypothèses sont connues ; un agent appris n'apporte pas automatiquement une amélioration supplémentaire.

Novelty Journals

ISSN 2394-7349

International Journal of Novel Research in Engineering and Science
Vol. 13, Issue 2, pp: (8-23), Month: September 2026 - February 2027, Available at: www.noveltyjournals.com

4.8. Test indépendant de session dans SUMO/TraCI

L'agent PPO entraîné dans SUMO atteint 50 176 pas d'apprentissage, pour une longueur moyenne d'épisode de 104 et une récompense moyenne d'entraînement de −18,8. Le calcul dure environ 1 431 secondes. L'explained_variance de 0,956 et l'approx_kl d'environ 0,00144 renseignent sur la dynamique de l'optimisation de la politique, sans démontrer son efficacité sur la circulation. La comparaison finale est conduite sur dix épisodes associés à la session Bypass maintenue hors entraînement ; le tableau 7 en présente les moyennes.

Le tableau 7 présente l'évaluation hors session d'entraînement des quatre politiques dans SUMO.

Tableau 7. Comparaison des politiques dans SUMO sur une session Bypass distincte (10 épisodes)

Politique	File moy.	Pic de file	Attente moy.	Écart équité	Débit moy.	Récompense
Fixe	140,02	203,7	6,875	24,78	985,1	−30,29
PPO	77,31	102,9	1,399	14,39	875,7	−13,83
Heuristique prédictive	116,32	187,1	4,365	68,53	953,3	−28,76
Réactive	86,69	137,7	1,856	16,66	896,7	−16,31

Dans SUMO, PPO réduit la file moyenne de 140,02 à 77,31, l'attente moyenne de 6,875 à 1,399 et l'écart d'équité de 24,78 à 14,39. Rapportées au feu fixe, ces différences représentent des diminutions respectives de 44,79 %, 79,65 % et 41,94 %. Le débit recule cependant de 985,1 à 875,7, soit 11,11 %. La stratégie réactive conserve un débit légèrement supérieur à celui de PPO (896,7), avec des files et attentes un peu plus importantes. L'heuristique prédictive maintient un débit relativement élevé (953,3), mais son écart d'équité atteint 68,53. La figure 11 met en regard ces critères, rappelant qu'aucune politique ne domine toutes les autres sur chaque dimension.

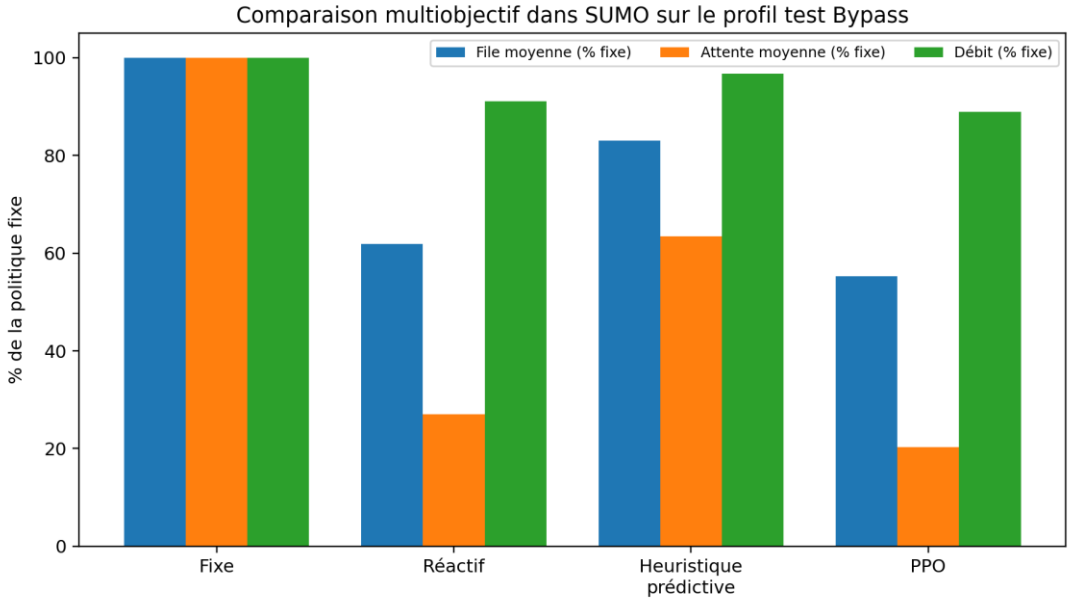


Figure 11. Comparaison multiobjectif des politiques de contrôle dans SUMO.

5. DISCUSSION

5.1. Contribution principale : révéler les dépendances et les compromis

Ces résultats montrent pourquoi la performance d'un système de mobilité intelligent ne peut se résumer à un unique indicateur. D'un côté, les catégories locales améliorent la pertinence de la description du trafic, mais le faible rappel et les confusions visuelles limitent la confiance accordée aux comptes. De l'autre, la prévision la plus sophistiquée n'est pas systématiquement la plus juste : la persistance conserve un avantage à une minute, alors qu'un LSTM obtient une erreur

absolue plus faible à cinq minutes sans expliquer correctement les fluctuations. Enfin, l'ordre des politiques varie nettement entre les deux simulateurs. La contribution centrale de Smart-Trafic tient à cette mise en relation des performances, des erreurs et des objectifs de contrôle.

5.2. Pourquoi l'heuristique et PPO changent-ils de rang ?

La divergence entre l'heuristique prédictive et PPO invite à examiner la dynamique des environnements. Dans le modèle analytique, l'évolution des files est gouvernée par des règles agrégées qui répondent directement aux profils de demande ; une heuristique construite pour ces relations peut y être particulièrement efficace. SUMO introduit des véhicules individuels, des transitions de signalisation et des interactions qui modifient les conséquences d'une décision. PPO peut alors privilégier la diminution des files et des attentes au détriment d'une partie du débit. Cette explication est compatible avec les observations, sans établir un lien causal certain : l'absence de calibration et de description exhaustive de la récompense empêche de départager toutes les causes possibles du changement de classement.

5.3. Menaces sur la validité

Quatre limites déterminent la portée de l'étude. La première concerne les mesures : le score de pression, l'occupation visuelle et la file estimée sont des variables indirectes, sans équivalence automatique avec des grandeurs physiques relevées sur la route. La deuxième touche au découpage des données : la proximité possible entre certaines images et le manque de détail sur les frontières temporelles limitent la capacité à exclure tout risque de fuite. La troisième est statistique : le test comporte seulement 67 objets, dont deux camions, et la comparaison SUMO repose sur dix épisodes sans intervalles de confiance. Enfin, la généralisation externe reste ouverte puisque les politiques n'ont pas été évaluées sur plusieurs intersections calibrées et indépendantes. Ces réserves ne retirent pas leur intérêt aux expériences ; elles précisent ce que les chiffres permettent réellement de conclure.

5.4. Ce qu'exigerait une preuve opérationnelle

Une étape suivante devrait privilégier la qualité des observations plutôt que l'accumulation de modèles. Des caméras fixes, installées sur plusieurs intersections et calibrées géométriquement, permettraient de constituer des séries couvrant heures de pointe, périodes creuses, week-ends et conditions météorologiques variées. Un sous-ensemble annoté par plusieurs personnes offrirait des références indépendantes pour la détection, les identités de suivi et les franchissements par direction. Pour l'apprentissage des politiques, des réseaux SUMO reconstruits à partir de la géométrie des carrefours et des comptages tournants rendraient possibles des comparaisons multi-scénarios, à plusieurs graines et avec intervalles de confiance. L'introduction de contraintes explicites sur le service minimal et les pertes de débit serait particulièrement utile.

5.5. Données, gouvernance et acceptabilité

L'observation vidéo de l'espace public pose également des questions de gouvernance. Les images peuvent contenir des visages, des plaques et des habitudes de déplacement. Une perspective de déploiement implique donc de limiter les données collectées, de contrôler les accès, de définir des durées de conservation et d'évaluer l'anonymisation avant partage. La confiance dans un système de transport intelligent dépend autant de ces garanties que de la précision de ses prédictions.

6. CONCLUSION

Smart-Trafic montre qu'une chaîne légère associant perception visuelle, suivi, séries temporelles et contrôle simulé peut être mise en œuvre sur des ressources informatiques limitées, tout en tenant compte de catégories caractéristiques du trafic à Kinshasa. Le détecteur YOLOv8n atteint une mAP@0,50:0,95 de 0,1803 et un rappel de 0,3502 sur le test réservé : ces résultats traduisent une faisabilité technique accompagnée de limites importantes pour le comptage. Les prévisions mettent en évidence la valeur de références simples, en particulier la persistance à une minute. Dans le modèle analytique, l'heuristique prédictive domine plusieurs indicateurs ; dans SUMO, PPO diminue fortement les files et les attentes, mais au prix d'une baisse de débit de 11,11 %. Au-delà de ces scores, l'étude établit l'intérêt d'une lecture multiobjectif et d'une évaluation séparée des maillons de la chaîne. La suite du travail repose sur la validation manuelle des mesures, la séparation contrôlée des sessions et la calibration des intersections simulées avant toute décision de régulation en conditions réelles.

REFERENCES

[1] Abid, A., Abdalla, A., Abid, A., Khan, D., Alfozan, A., & Zou, J. (2019). Gradio: Hassle-free sharing and testing of ML models in the wild. arXiv.

[2] Akujobi, M. U., Abubakar, A. U., Mailabari, R. J., Thuku, I. T., Musa, S. Y., Visa, I. M., & Abioye, A. O. (2026). Keke-aware vehicle counting for traffic measurement using YOLO: Dataset and field evaluation. Applied Sciences, 16(9), 4316. https://doi.org/10.3390/app16094316

[3] Alvarez López, P., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., & Wießner, E. (2018). Microscopic traffic simulation using SUMO. In 2018 21st International Conference on Intelligent Transportation Systems (ITSC) (pp. 2575-2582). IEEE.

[4] Bakirci, M. (2024a). Utilizing YOLOv8 for enhanced traffic monitoring in intelligent transportation systems (ITS) applications. Digital Signal Processing, 152, 104594.

[5] Bakirci, M. (2024b). Enhancing vehicle detection in intelligent transportation systems via autonomous UAV platform and YOLOv8 integration. Applied Soft Computing, 164, 112015.

[6] Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. In 2016 IEEE International Conference on Image Processing (ICIP) (pp. 3464-3468). IEEE.

[7] Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32.

[8] Eclipse SUMO. (2026). Eclipse SUMO 1.27.1 [Computer software]. Zenodo.

[9] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585, 357-362.

[10] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.

[11] Huang, L., & Qu, X. (2023). Improving traffic signal control operations using proximal policy optimization. IET Intelligent Transport Systems.

[12] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90-95.

[13] JICA (2019). Project for urban transport master plan in Kinshasa city (PDTK): Final report (Vol. 1).

[14] JICA. (2024). Étude de collecte de données sur le secteur routier de la ville de Kinshasa en République Démocratique du Congo – Rapport final.

[15] Khrisne, D. C., Sudarma, M., Giriantari, I. A. D., Wiharta, D. M., & Kurniawan, Y. (2026). Traffic signal control via proximal policy optimization with reward shaping to minimize waiting time and violations. JOIV: International Journal on Informatics Visualization, 10(4), 1578–1587. https://doi.org/10.62527/joiv.10.4.4518

[16] Kyatha, N., & Taneja, J. (2025). Recovering origin–destination flows from bus CCTV: Early results from Nairobi and Kigali. arXiv:2512.00424. https://arxiv.org/abs/2512.00424

[17] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In Computer Vision - ECCV 2014 (pp. 740-755). Springer.

[18] Mpia, H. N., Kavalami, M. M., Lusenge, G. K., Ushindi, K. P., Kyalengekania, D.-D. K., & Ciswaka, O. M. (2025). Lightweight deep learning system for infant-cry recognition with real-time notification in resource-constrained environments. Machine Learning with Applications, 22, 100791. https://doi.org/10.1016/j.mlwa.2025.100791

[19] Padilla, R., Passos, W. L., Dias, T. L. B., Netto, S. L., & da Silva, E. A. B. (2021). A comparative analysis of object detection metrics with a companion open-source toolkit. Electronics, 10(3), 279.

[20] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. In Advances in Neural Information Processing Systems, 32.

[21] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

[22] Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., & Dormann, N. (2021). Stable-Baselines3: Reliable reinforcement learning implementations. Journal of Machine Learning Research, 22(268), 1-8.

[23] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 779-788). IEEE.

[24] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv.

[25] Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. (2023). A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Machine Learning and Knowledge Extraction, 5(4), 1680-1716.

[26] Ugboko, E., & Jo, S. (2025). Adapting AI-based speed violation detection systems for Africa: A case study with Nigeria. African Transport Studies, 3, 100035.

[27] Ultralytics. (2026). Ultralytics YOLO documentation: Train, validation, prediction and tracking.

[28] United Nations, Department of Economic and Social Affairs, Population Division. (2025). World Urbanization Prospects 2025: Summary of results (UN DESA/POP/2025/TR/No. 12).

[29] Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. In 2017 IEEE International Conference on Image Processing (ICIP) (pp. 3645-3649). IEEE.

[30] Xiao, F., Lu, J., Li, L., Tu, W., & Li, C. (2025). Advances in reinforcement learning for traffic signal control: A review of recent progress. Intelligent Transportation Infrastructure, 4, liaf009.

[31] You, L., Chen, Y., Xiao, C., Sun, C., & Li, R. (2024). Multi-object vehicle detection and tracking algorithm based on improved YOLOv8 and ByteTrack. Electronics, 13(15), 3033.

[32] Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-object tracking by associating every detection box. In Computer Vision - ECCV 2022 (pp. 1-21). Springer