

Fiabilisation des données d'infrastructure ferroviaire par intelligence artificielle

Housni Djellab¹, Edouard Donze¹, Yasmina Chiguer¹, Renan Hilbert², Gérald Petitjean³

¹ SNCF RESEAU

{housni.djellab, edouard.donze, yasmina.chiguer}@reseau.sncf.fr

² yPlus

rhilbert2@y-plus.fr

³ EURODECISION

gerald.petitjean@eurodecision.com

Mots-clés : *infrastructure ferroviaire, intelligence artificielle, référentiel de données, automatisation, machine learning, deep learning, computer vision, réseaux de neurones, transformers, CNN.*

1 Introduction et définition de la problématique

L'essor de IA et de la Computer Vision, constitue aujourd'hui un levier majeur d'innovation pour le secteur ferroviaire. Face à la complexité croissante des opérations, à l'étendue du réseau et aux exigences de sécurité, l'analyse automatisée d'images permet d'assister – voire de remplacer – des tâches manuelles historiquement longues, coûteuses et exposées aux erreurs humaines. Offrant ainsi aux gestionnaires d'infrastructures des capacités inédites pour moderniser les pratiques de maintenance, de surveillance et de pilotage.

Dans cette dynamique, SNCF Réseau a engagé la conception d'un service d'aide à la décision visant à améliorer la qualité des données d'infrastructure ferroviaire. Ce service constitue la première brique de la Tour Qualité de données et s'articule avec les travaux de rapprochement et de comparaison des données d'infrastructure. Il répond à un enjeu clé : identifier automatiquement les écarts entre la réalité terrain observée dans les images et les données enregistrées dans les systèmes opérationnels, afin de renforcer leur fiabilité.

Le projet s'est concentré sur un périmètre restreint d'objets critiques, avec une supervision métier tout au long du processus pour garantir la pertinence opérationnelle et adapter les modèles.

2 Description de la solution technique

La solution développée repose sur deux blocs fonctionnels complémentaires.

1) Détection automatique des objets ferroviaires par IA

Un modèle d'IA a été entraîné pour détecter automatiquement différents types d'objets ferroviaires à partir de sources d'images variées (ESV, SIM, drones). Pour cela, une démarche d'apprentissage supervisé a été déployée : préparation et annotation de jeux d'images, puis entraînement sur des modèles de deep learning.

Plusieurs architectures ont été testées :

- pour les images ESV et drones : un modèle pré-entraîné de type Transformer (DETR-ResNet101 via Hugging Face) ;
- pour les images SIM : des réseaux de neurones convolutifs (CNN), mieux adaptés à des images centrées et spécifiques.

2) Rapprochement, analyse des écarts et restitution

Le second bloc consiste à rapprocher les objets détectés avec les données d'infrastructure, en s'appuyant sur leur ferro-localisation (code ligne, code voie, PK). L'objectif est d'identifier automatiquement les incohérences entre la réalité observée dans les images et les données enregistrées dans les systèmes.

Les résultats sont ensuite traités et restitués sous forme d'indicateurs visuels et analytiques, destinés aux experts métiers, qui conservent un rôle essentiel dans la validation et les décisions de correction. Un post-traitement avancé permet la suppression des doublons, l'extraction d'informations textuelles et l'harmonisation des résultats, afin d'obtenir des données exploitables.

Enfin, un prétraitement systématique d'anonymisation garantit la conformité avec le RGPD et un usage responsable des images.

3 Résultats

L'outil a permis d'identifier automatiquement 6 types d'objets critiques à partir des images ESV, et 3 types d'objets critiques à partir des images SIM, avec un taux moyen de détection de 97,3 % et une précision de 0,95, démontrant la robustesse des modèles testés, sur données fictives et sur deux périmètres réels du réseau.

En termes de temps de calcul, l'entraînement sur les images ESV nécessite environ 9 heures, un délai qui peut être réduit en augmentant les ressources CPU allouées. Quant à l'inférence, elle prend environ 1 minute.

Certaines limites persistent (objets de petite taille, conditions visuelles difficiles, confusions ponctuelles), ce qui justifie un processus continu d'amélioration et d'enrichissement des données d'entraînement.

4 Conclusions et perspectives

Les travaux réalisés ont permis de développer et d'évaluer un système capable de détecter, rapprocher et analyser automatiquement des objets ferroviaires critiques à partir d'images terrain. La solution basée sur un modèle d'IA montre qu'une automatisation supervisée peut renforcer la fiabilité des données, réduire significativement les tâches manuelles, améliorer la productivité et homogénéiser les corrections apportées aux référentiels d'infrastructure.

À moyen et long termes, plusieurs pistes d'évolution sont ouvertes :

- élargissement du périmètre à l'ensemble du réseau ferré national,
- intégration de nouvelles typologies d'objets critiques,
- exploitation de sources d'images complémentaires,
- généralisation des mécanismes d'automatisation dans la Tour Qualité de données.

Cette étude pose ainsi les fondations d'un dispositif pérenne capable de soutenir la supervision et la qualité des données d'infrastructure, tout en apportant une méthodologie robuste et répliquable à d'autres domaines critiques.