

Séminaire IA

18 et 19 septembre 2024

Réseaux de neurones profonds pour l'analyse des flux d'audio et de vidéo surveillance

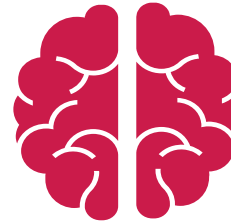
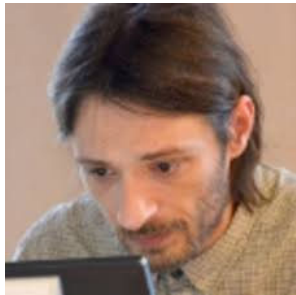
Sébastien AMBELLOUIS - COSYS/LEOST



**Université
Gustave Eiffel**

Préambule

- Les acteurs de cette recherche au COSYS-LEOST



David SODOYER

Sébastien
AMBELLOUIS

Cyril MEURIE

Et nos doctorants et post-doctorants : Christophe CONIGLIO, Tuan-Hung VU, Rémi DUFOUR, Mohamed-Amine HADDED, Tony MARTEAU, Olivier LAURENDIN



Contexte applicatif

... depuis les années 2000

- Automatisation de certaines fonctions de surveillance
 - Exploitation / interprétation du contenu des flux de vidéo et audio surveillance

SÛRETÉ

SÉCURITÉ

Enceinte d'un véhicule



Bus
Projet TEsS, EVAS
(ANR)



Métro



Train
Projet BOSS (EU-Celtic),
SURTRAIN (ANR), DÉGIV (FUI)

Infrastructure

Projet STATUE (ANR)



Contexte applicatif

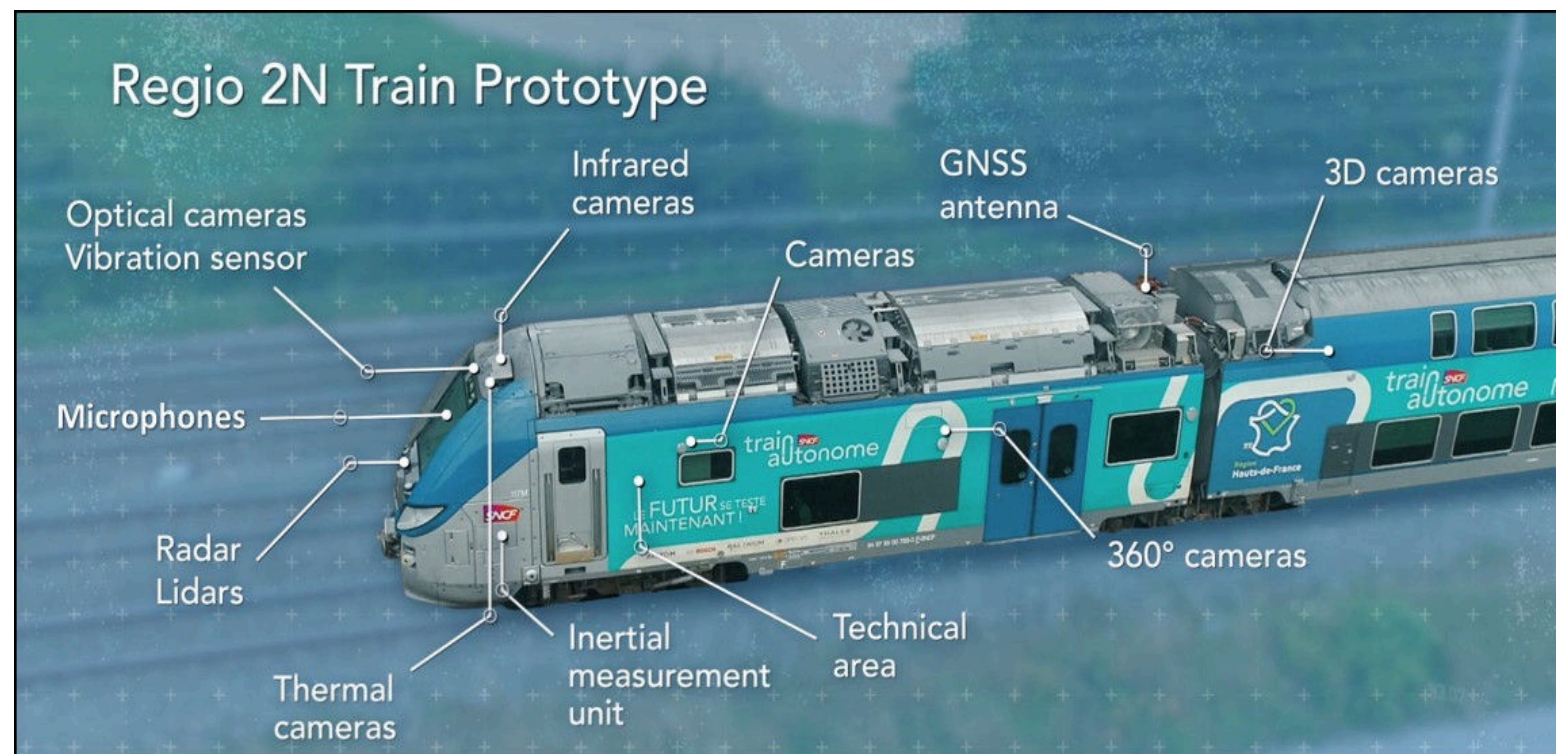
... aujourd'hui et depuis 2019



GOA3
Conduite automatisée
avec personnel à bord



GOA4
Train
totalement autonome



Contexte applicatif

... les événements d'intérêt

- Quels sont les événements à détecter ou identifier (les « classes ») ?
 - Dégradation du matériel
 - Chute de passagers
 - Actes d'incivilité ou de violence
 - Bagage abandonné
 - Passagers bloqués entre les portes
- **Enceinte d'une voiture**
 - Détection d'obstacles en voies
 - Détection d'une voie endommagée
 - Détection en bord de voie (agents / piétons / feu / animal)
 - État passage à niveau

Contexte applicatif

... les contraintes

- Les contraintes liées à l'environnement
 - La variation d'échelle
 - Les occultations
 - Point de vue variable des caméras
 - Le champ de vision parfois limité
 - Bruit dans les signaux
 - Diversité des environnements et des situations
- Les contraintes liées aux données
 - Le manque de données (bases de données ouvertes vs privées)
 - Déséquilibre du nombre d'échantillons par « classe »
 - Présence d'événements rares



© SNCF – CIM Le Mans



© Railenium



© Consortium Projet BOSS

Contexte applicatif

... l'objectif

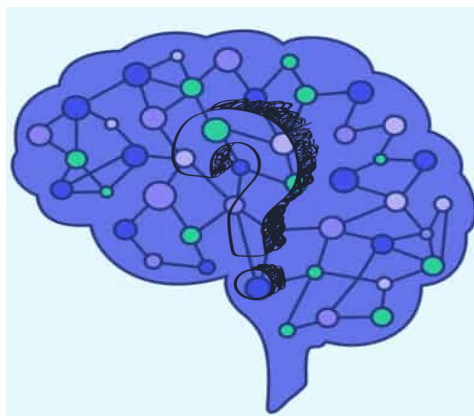


Séquence d'images



Signal sonore

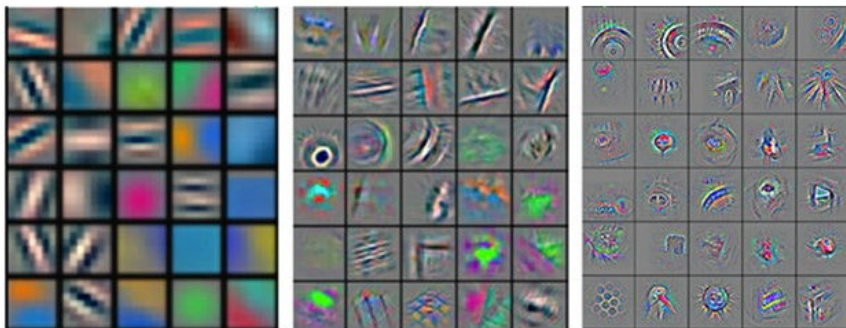
Réseau de neurones profond



Détection d'un événement anormal

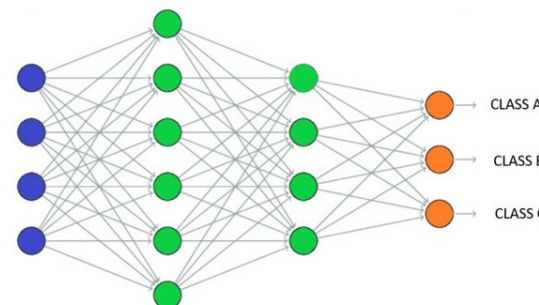
Identification d'un des événements d'intérêt

Espace de représentation



© Krizhevsky et al.

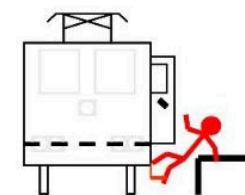
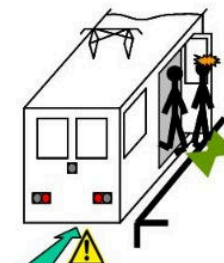
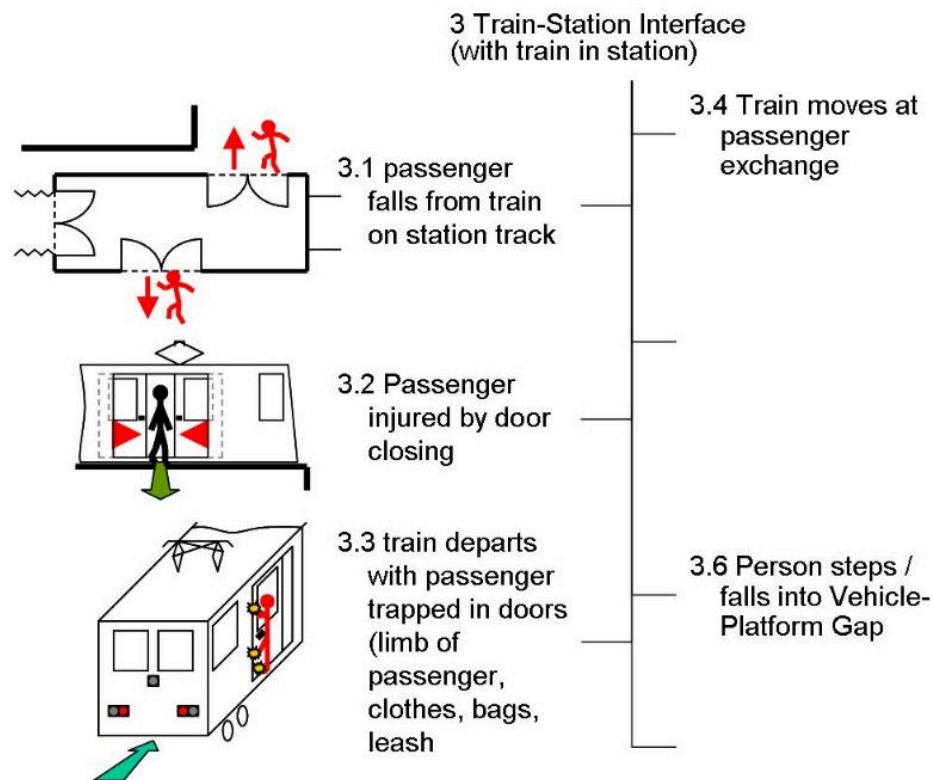
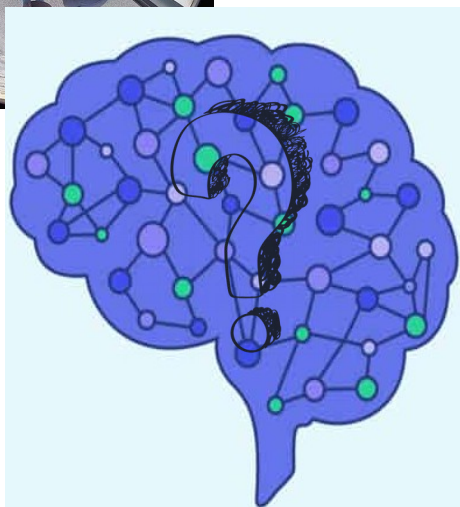
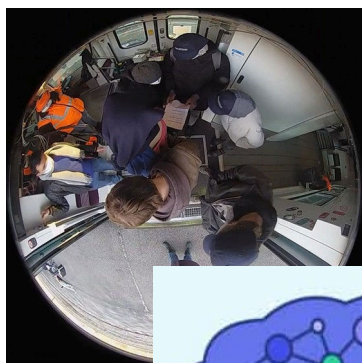
Classifieur



Contexte applicatif

... le cas d'usage

Détection d'anomalies aux portes d'un train autonome



© Railenium

Méthodologie

... les données au cœur de l'apprentissage

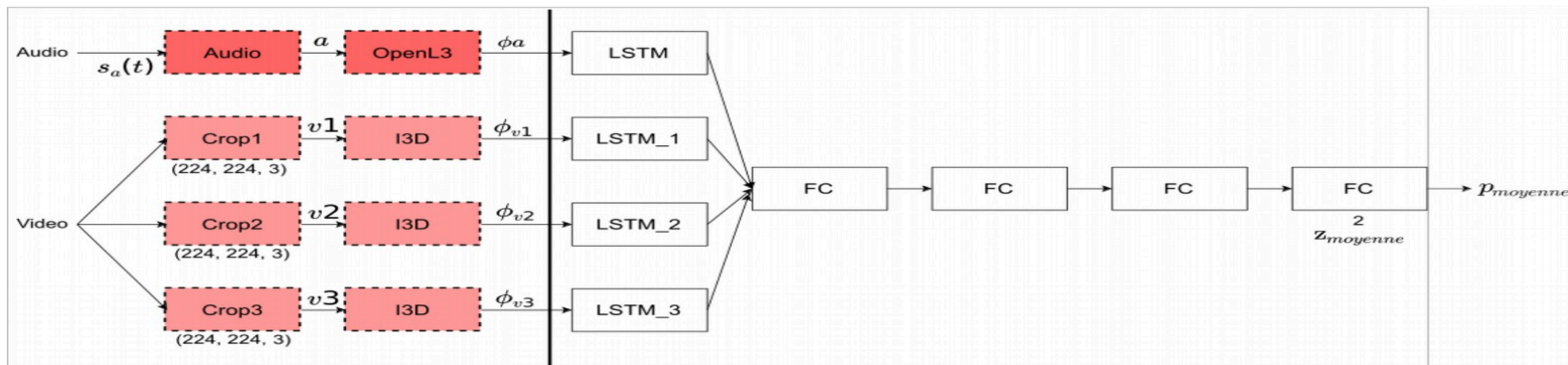
- Elles doivent être **en quantité suffisante** pour s'assurer d'avoir une base d'apprentissage, une base de validation et une base de test
- **S'assurer de la diversité et la représentativité des événements anormaux.** Les événements anormaux étant souvent **rare**s, il est difficile de collecter suffisamment d'exemples. Il faut gérer ce déséquilibre pour :
 - entraîner des modèles **robustes**
 - Obtenir des modèles **non biaisés** vers la détection des événements normaux uniquement ou qui montreraient des performances variables d'un événement à l'autre
- La labellisation manuelle des événements est une tâche fastidieuse et sujette à des erreurs : utilisation de méthodes automatisées ou semi-automatisées d'annotation
- Ethique : Vie privée et protection des données. La capture et l'analyse des vidéos de transport en commun impliquent des questions de vie privée. Il faut **garantir que l'utilisation de ces bases de données respecte les lois sur la protection des données personnelles, telles que le RGPD en Europe**



Méthodologie

... l'apprentissage supervisé : approche classique

- Pour identifier chaque événement : apprentissage « supervisé »
 - La base de données d'apprentissage doit comporter des échantillons pour chaque classe d'événements à identifier i.e. chaque donnée est associée à un label (son appartenance à une classe)
 - Les classes doivent être équilibrées
 - Le réseau appris doit permettre d'associer une classe « événement » lorsqu'on lui présente une ou des images
- Thèse CIFRE SNCF de Tony Marteau, *Approche neuronale profonde pour la reconnaissance conjointe audio-vidéo de violences dans un environnement ferroviaire embarqué*, encadrée par David Sodoyer et al., 2023

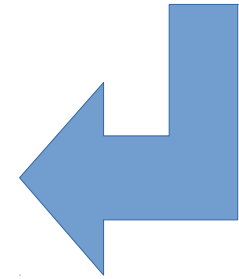


Méthodologie

... *apprentissage « non » et « auto » supervisé*

- Chaque événement est considéré comme une anomalie : apprentissage « non » et « auto » supervisé
 - La base de données d'apprentissage ne comporte aucune anomalie → seule la base de test en comporte
 - Définition de tâches « **prétextes** » (tâches proxy) → **apprentissage de l'espace de représentation des données**
 - **Le modèle apprend à reconstruire des données normales → anomalie détectée en fonction des erreurs de reconstruction**

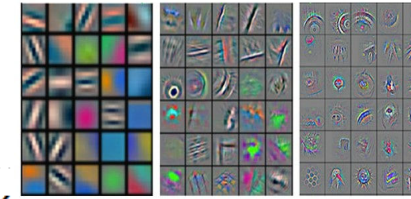
Très peu d'instances
des événements
anormaux dans les
bases de données



Méthodologie

... utilisation de modèles génératifs

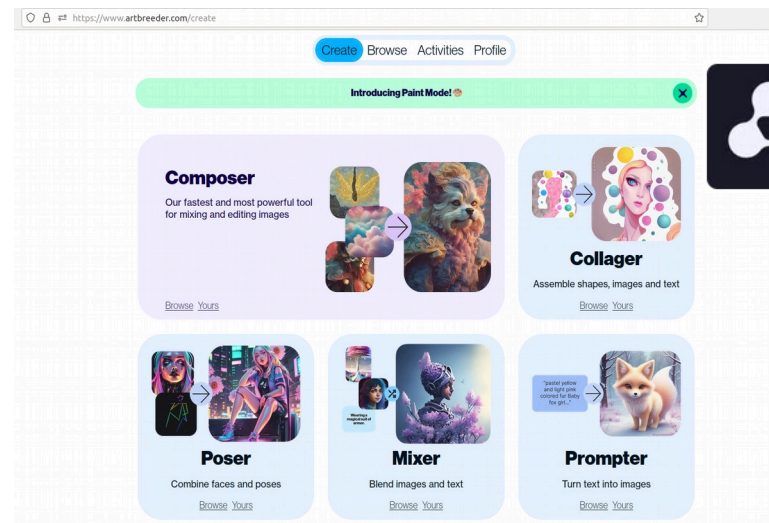
Espace de représentation



- Définition de tâches « prétextes » (tâches proxy) → apprentissage de l'espace de représentation des données
- Le modèle apprend à reconstruire des données normales → anomalie détectée en fonction des erreurs de reconstruction

Un modèle génératif est appris pour capturer la distribution des données d'entraînement → générer de nouvelles données similaires.

Generative Adversarial Networks (GAN), Conditional GAN, AutoEncoder (AE), Variational AE



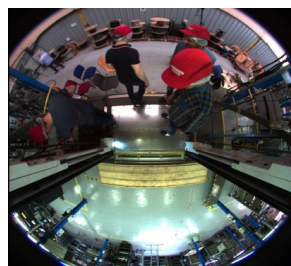
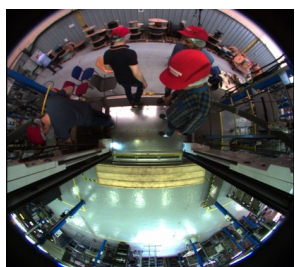
Artbreeder



Méthodologie

... utilisation de modèles génératifs

Tâche « proxy » : Reconstruction en apparence

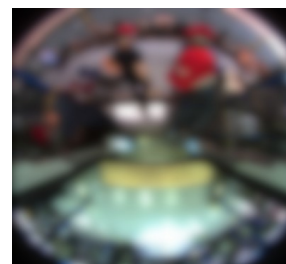


$$\theta_* = \underset{\theta}{\operatorname{argmin}} L(X, \theta)$$

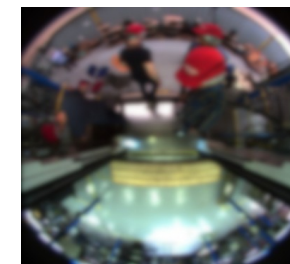


Sortie au cours de l'apprentissage

Itération 1

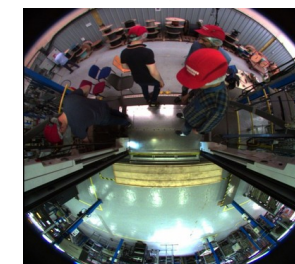


Itération 2



.....

Itération n

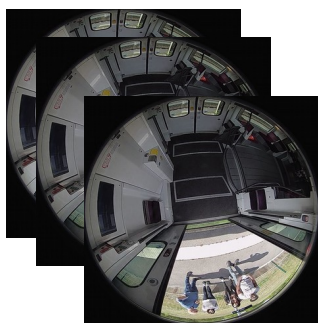


Images d'apprentissage = aucune anomalie

Méthodologie

... utilisation de modèles génératifs

Tâche « proxy » : Prédiction de l'image à l'instant t



Images successives
entre $t-n$ et $t-1$



$$\theta_* = \underset{\theta}{\operatorname{argmin}} L(X, \theta)$$

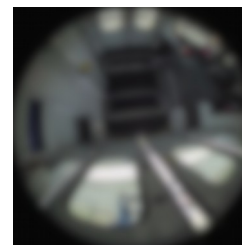


Images à l'instant t

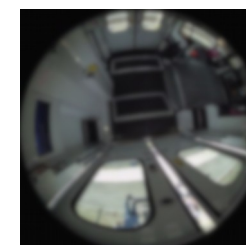


Sortie au cours de l'apprentissage

Itération 1



Itération 2



.....

Itération n

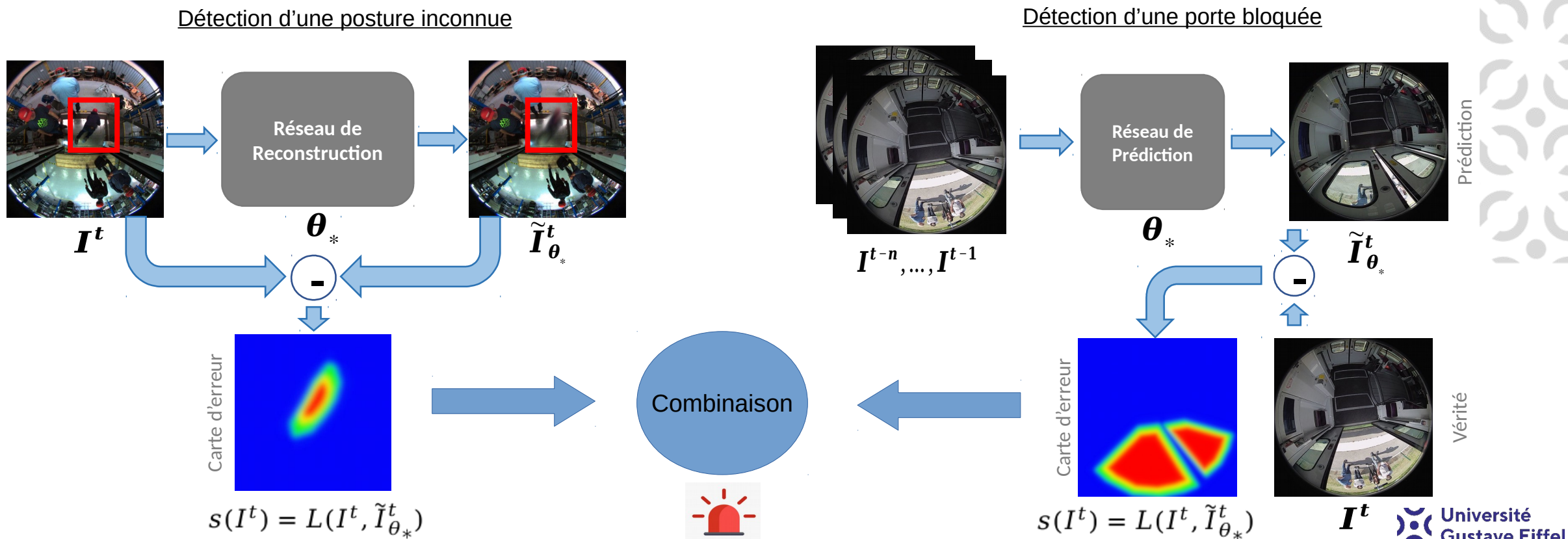


Images d'apprentissage = aucune anomalie

Méthodologie

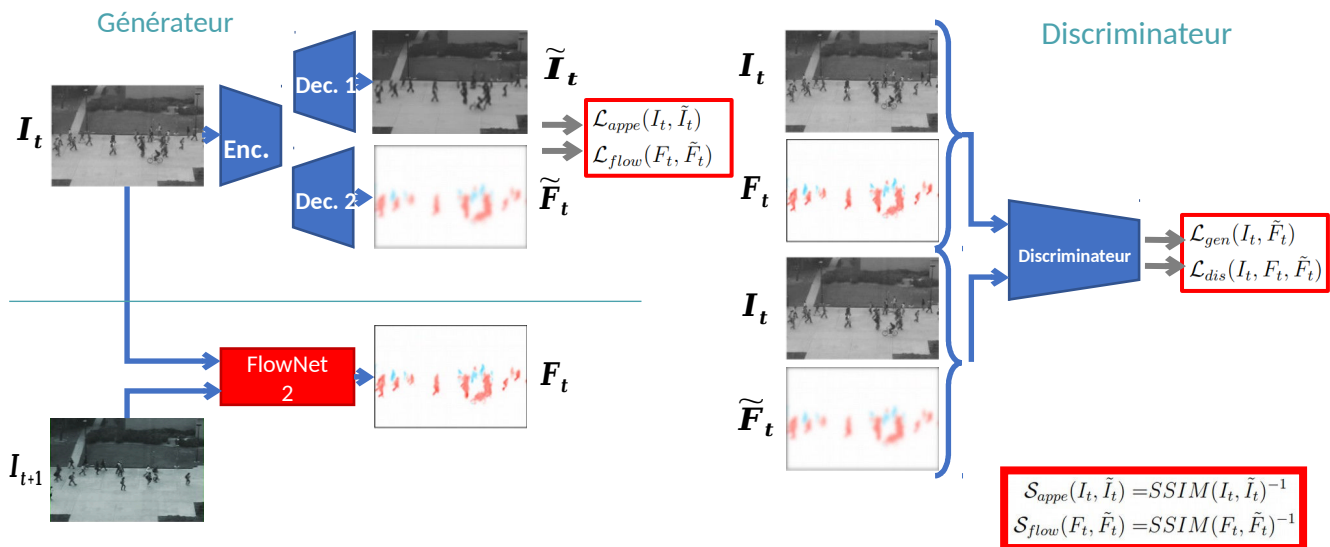
... détection d'une anomalie

- Inférer chaque réseau entraîné sur les images d'entrée d'une séquence et calculer l'erreur de reconstruction ou de prédiction
- Combiner les erreurs calculées pour déterminer si oui ou non une anomalie est en cours



Résultats

... architecture GAN adaptée



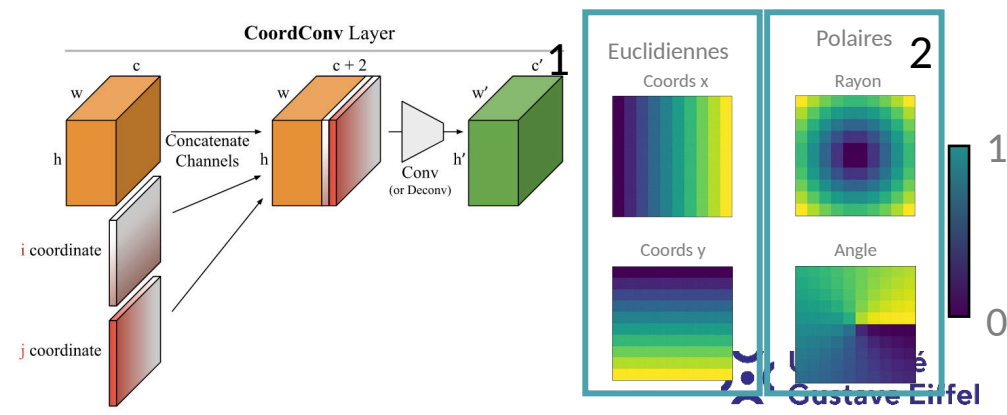
Masque de Ruzicka

Trong Nguyen Nguyen and Jean Meunier. Anomaly Detection in Video Sequence With Appearance-Motion Correspondence. In 2019 IEEE/CVF International Conference on Computer Vision (ICCV)

Apparence dépendante de la position dans l'image

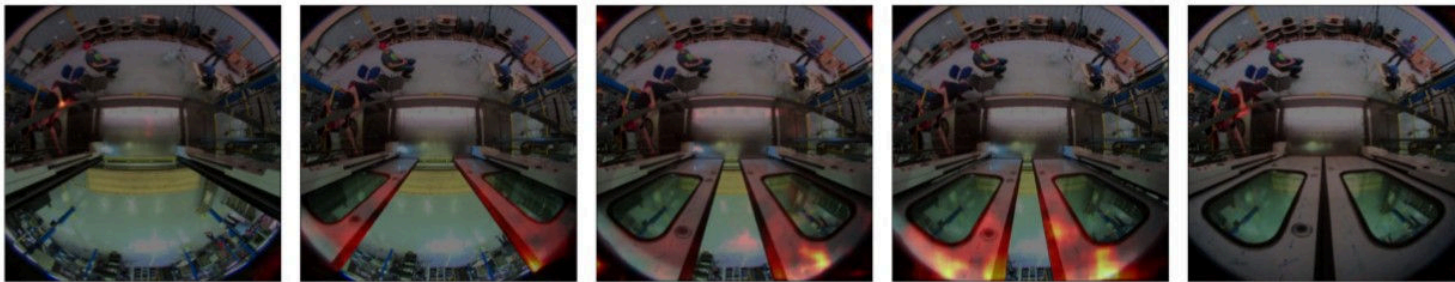
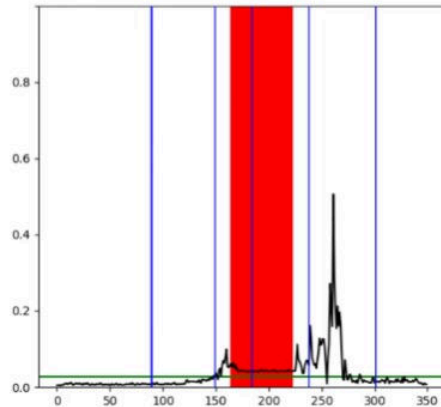


Couche CoordConv



Méthodologie

... architecture GAN adaptée



	gm ROC		
	Appa	Flot	Aléa.
Portes en Interruption de Déplacement	0.783	0.874	0.5
Instance Prise dans les Portes	0.819	0.888	
Usager Pris dans les Portes	0.819	0.888	
Bagage Pris dans les Portes, usager à l'Intérieur du train	0.819	0.796	
Bagage Pris dans les Portes, usager à l'Extérieur du train	0.819	0.888	
Instance au seuil de la Porte Pendant leur Fermeture	0.774	0.84	
Échange passager entravé	0.46	0.481	
Chute d'un usager	0	0.099	

Conclusions et perspectives

- Problématique IA et analyse audio/vidéo au COSYS-LEOST
 - Apprentissage machine : supervisé, non et auto supervisé → événements rares
 - Résultats qui sont ceux rencontrés dans l'état de l'art mais dans des contextes d'application beaucoup plus contraints et en présence d'événements complexes (interactions)

- Faire en sorte que le modèle s'adapte « facilement » à d'autres véhicules ferroviaires (mobilier et agencement différents), qu'il soit robuste à la variabilité des situations etc.
 - Création de base de données → synthèse ou système génératif
- Complexité calculatoire (apprentissage et inférence) → énergivore
- Utilisation d'une autre technologie de caméras : caméra neuromorphique → moins consommatrice, données plus « sparses », réseau à impulsions, apprentissage plus difficile, implantation sur du matériel spécifique moins consommateur

Sébastien AMBELLOUIS

COSYS-LEOST

sebastien.ambellouis@univ-eiffel.fr

Publications

Mohamed Amine Hadded, Ankur Mahtani, Sébastien Ambellouis, Jacques Boonaert, Hazem Wannous. *Application of Rail Segmentation in the Monitoring of Autonomous Train's Frontal Environment*. ICPRAI 2022, International Conference on Pattern Recognition and Artificial Intelligence, Jun 2022

Olivier Laurendin, Sebastien Ambellouis, Ankur Mahtani, Anthony Fleury. Pedestrian and Automatic Doors Abnormal Interactions Detection using Multi-Task Self-Supervised Learning. Workshop in Video Understanding and its Applications, BMVC, 2023

Tony Marteau, David Sodoyer, Sebastien Ambellouis, Sitou Afanou. Level fusion analysis of recurrent audio and video neural network for violence detection in railway. EUSIPCO 2022, European Signal Processing Conference, Aug 2022

Olivier Laurendin, Sébastien Ambellouis, Anthony Fleury, Ankur Mahtani, Sanaa Chafik, et al.. *Hazardous Events Detection in Automatic Train Doors Vicinity Using Deep Neural Networks*. AVSS 2021, IEEE International Conference on Advanced Video and Signal Based Surveillance, Nov 2021

Tuan-Hung Vu, Jacques Boonaert, Sébastien Ambellouis, Abdelmalik Taleb-Ahmed. Multi-Channel Generative Framework and Supervised Learning for Anomaly Detection in Surveillance Videos. Sensors, 2021