Out-of-Distribution Approach

Pierre Minier, Alexandre Valade

Enseirb-Matmeca, Bordeaux-INP

17 Janvier 2023

Introduction

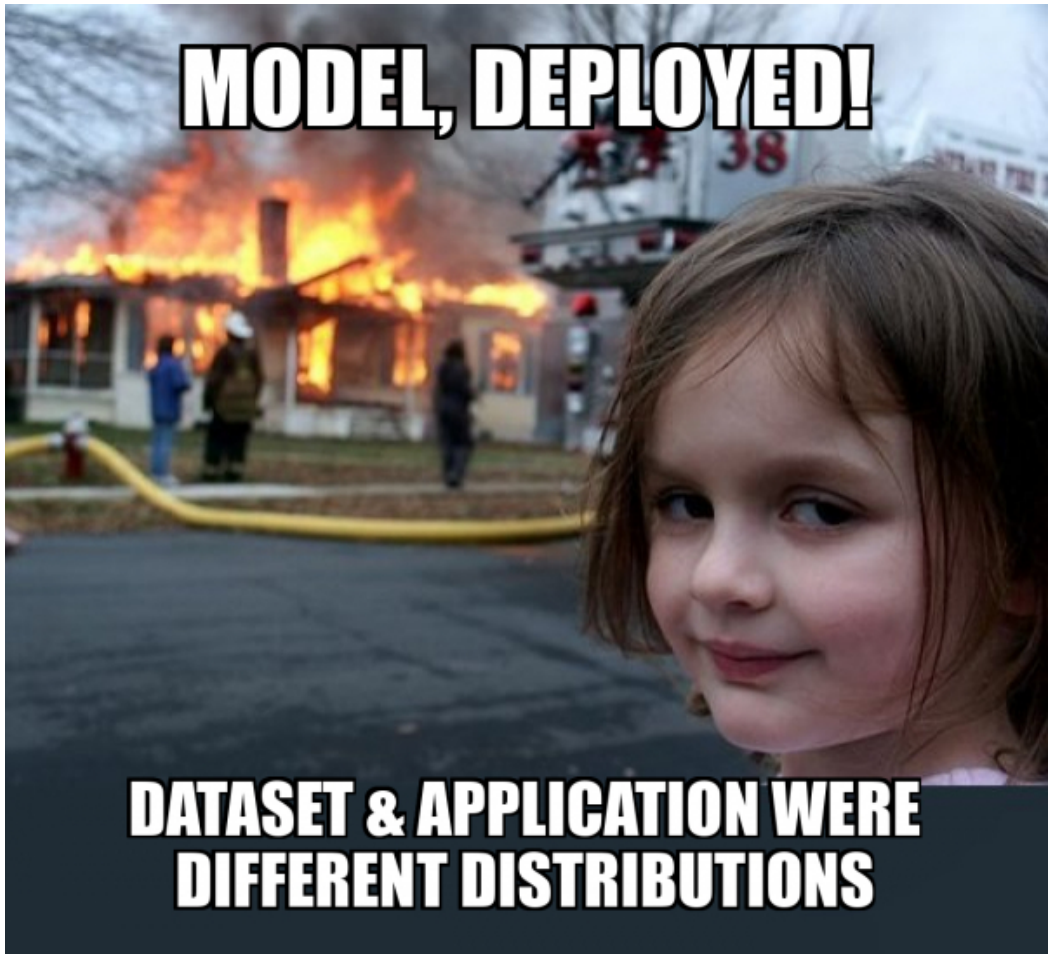

Figure – Disaster AI

Contenu de la présentation

Principe (1)

Problème :

Hypothèse du monde clos

Tout ce qui n'est pas dans mes données d'entraînement est faux.

Conséquence : Impossible de traiter correctement une donnée qui sort du cadre de l'entraînement. Cependant, la donnée est quand même traitée.

Risques : Détecter une maladie dans des données erronées (biomédical), erreurs de prédictions, etc.

Principe (2)

Figure – Chiffre 1

Figure – Lettre m

Lien pour tester directement

Principe (3)

Données dans la distribution : In-D

Données dont les propriétés statistiques sont similaires à une classe de données d'entraînement.

Données hors de la distribution : OOD

Données dont les propriétés statistiques ne correspondent à aucune classe de données d'entraînement.

Principe (4)

Causes d'OOD :

- Malfonctionnement d'un capteur
- Changement de distribution des données

Objectif : Détecter les données qui sont hors de la distribution, de manière fiable.

Méthodes communes :

- Détection par génération
- Détection par classification

Principe (5)

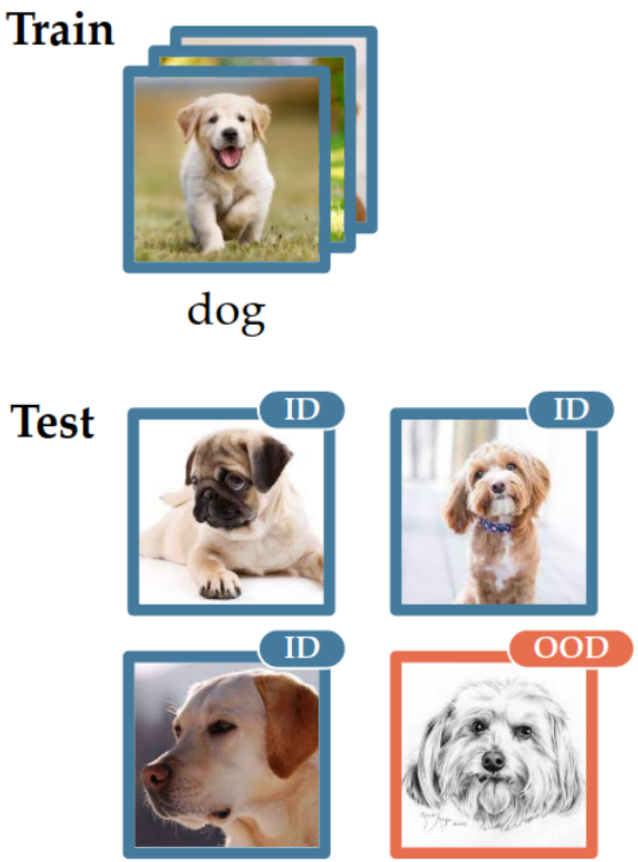

Figure – Anomalie de style

Principe (5)

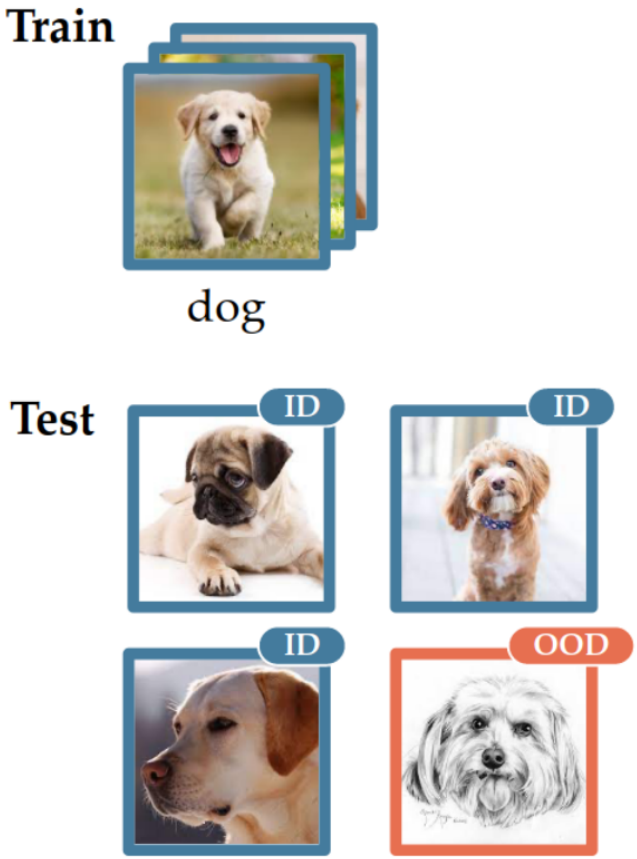

Figure – Anomalie de style

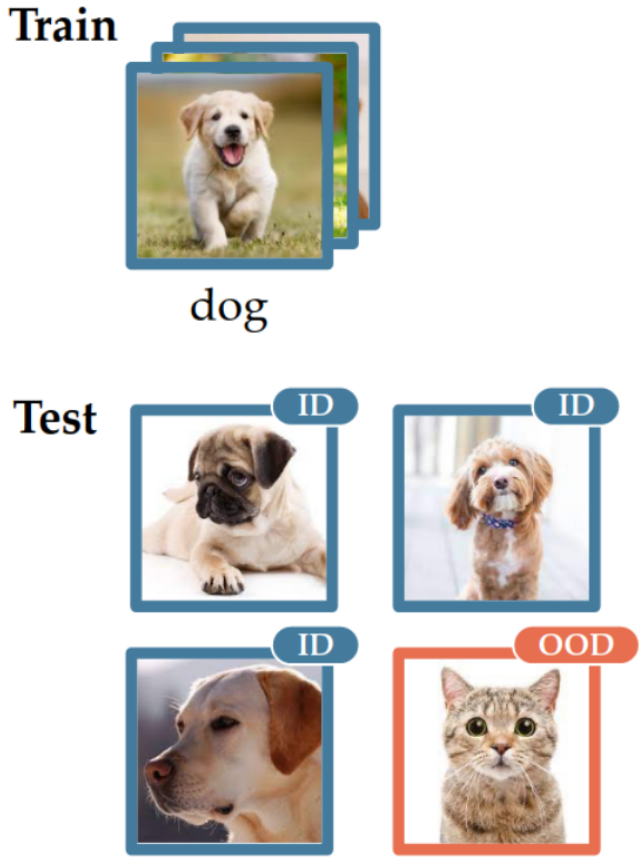

Figure – Anomalie sémantique

Principe (6)

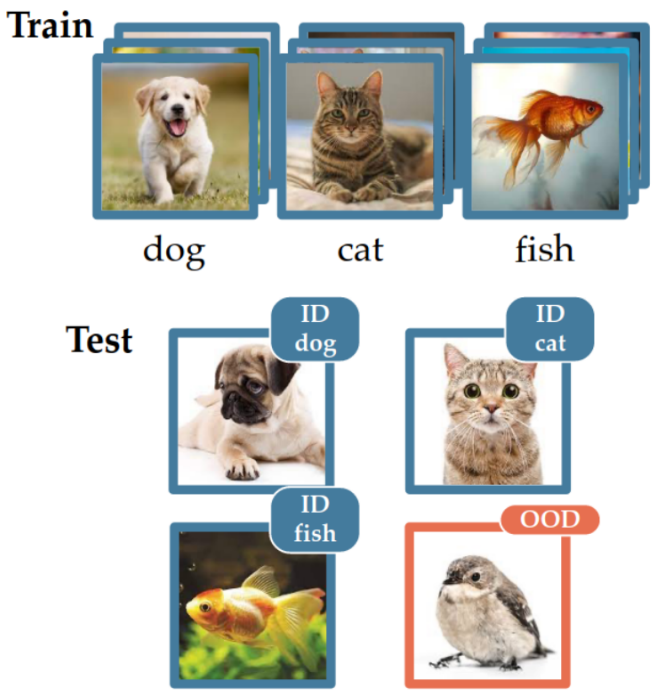

Figure – Classification avec réjection

Principe (6)

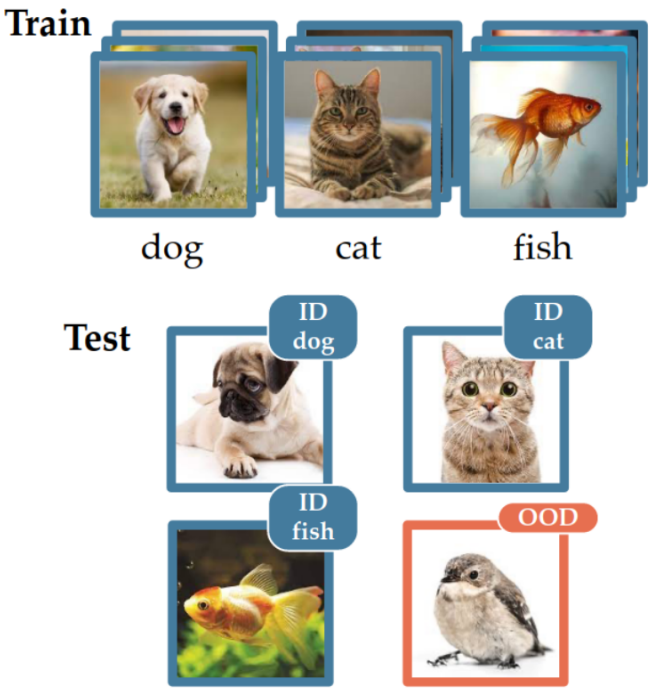

Figure – Classification avec réjection

Figure – Détection d'intrus

Principe (7)

Opportunités :

- Arrêter le traitement d'une donnée hors de la distribution
- Inclure une nouvelle classe au modèle

Cadres d'application :

- Conduite autonome
- Surveillance
- Biomédical

MoodCat : motivations

MoodCat : Masked OOD Catcher

Objectifs :

- Détection de OOD par un classificateur formé sur des In-D.
- Détection efficace sans affecter la précision du classificateur.
- Plug-and-play : peu de manipulations pour l'intégrer

Présentation de la méthode

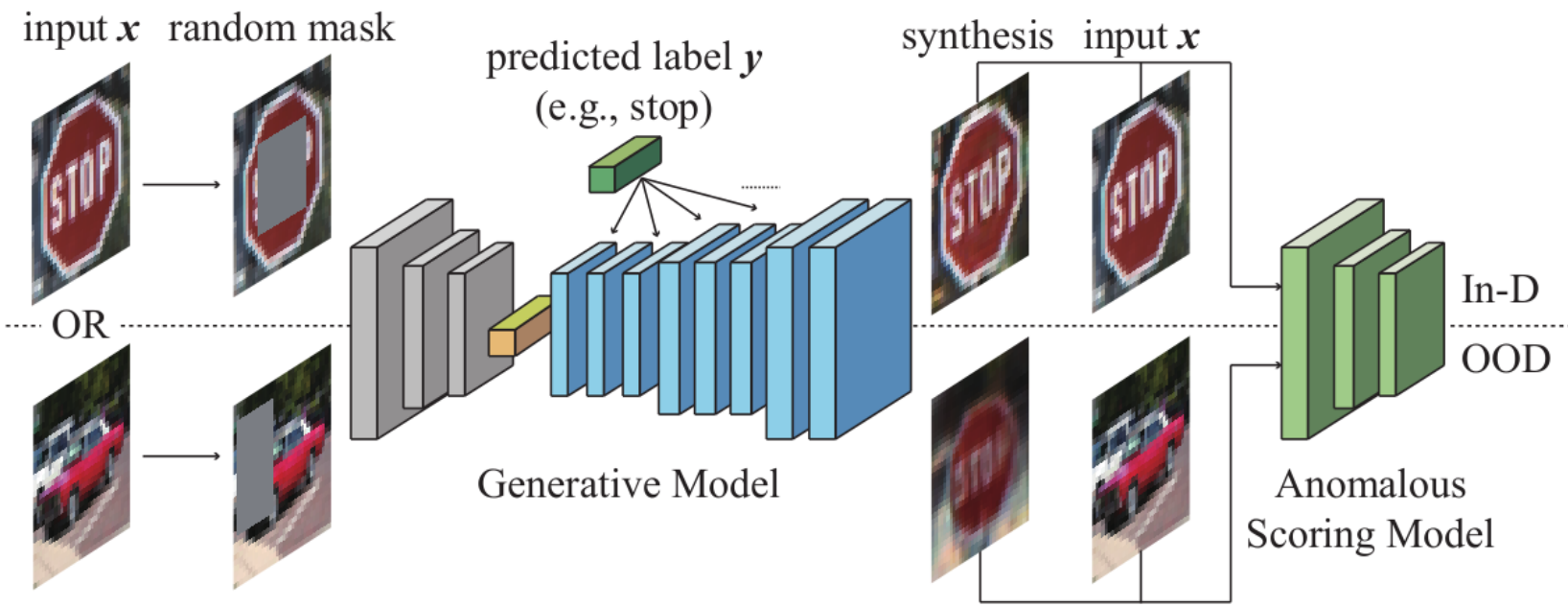

Figure – Pipeline de MoodCat

Masquage aléatoire

Génération de contenu

Les images possèdent des redondances : il faut en retirer suffisamment pour ne pas reconstruire l'image d'origine.

Deux avantages :

- Meilleur compréhension de la sémantique (OOD)
- Résumer les informations depuis l'ensemble de l'image (In-D)

Modèle génératif

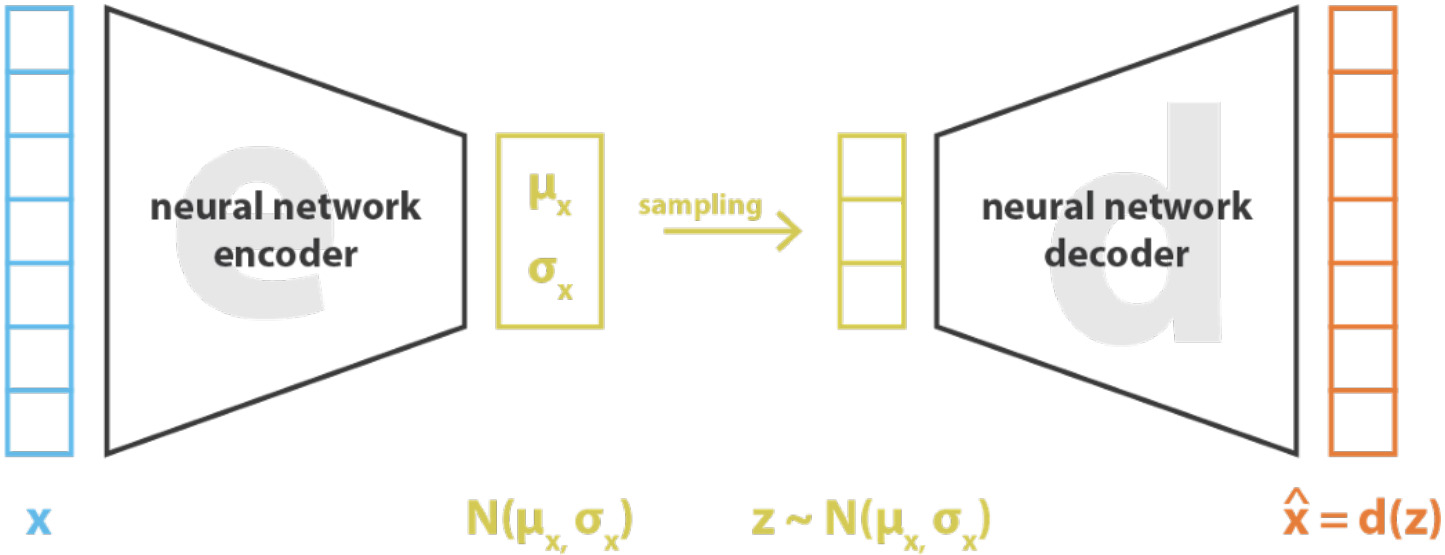

Figure – VAE entrainé In-D ; DKL pour régulariser la distribution latente

Utilisation du décodeur

Génération de contenu en utilisant un remplissage aléatoire de la partie masquée de l'image d'origine en entrée du décodeur.

Score

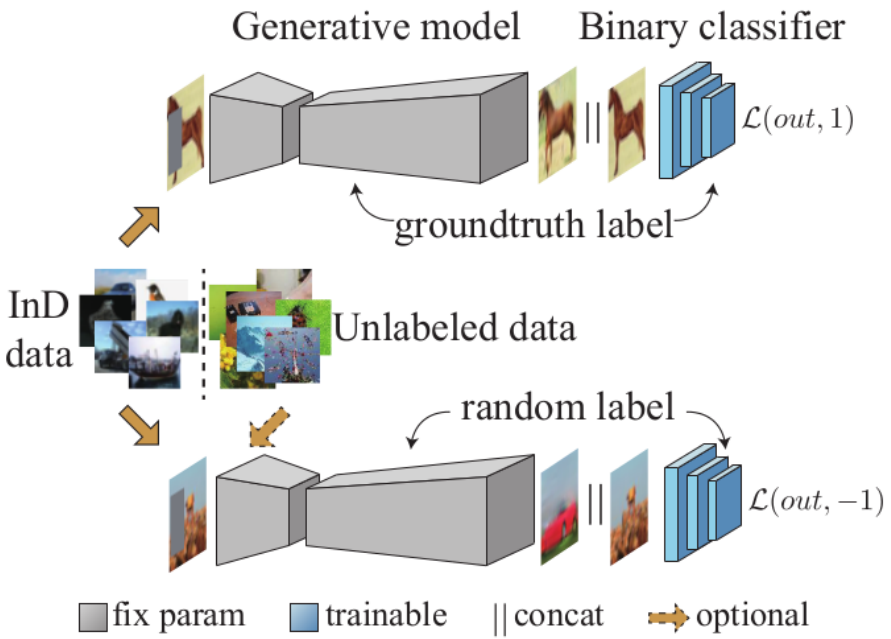

Figure – Entraînement du classifieur binaire

Coût de Hinge

$$\mathcal{L}_{\mathcal{C}_b} = ReLU(1 - \mathcal{C}_b((x, x'_y), y)) + ReLU(1 + \mathcal{C}_b((x, x'_{y'}), y'))$$

Méthodes classiques prédictives de détection d'OOD

Méthode d'OOD classique : Ajouter de l'OOD dans les données d'entrainement.

Méthodes classiques prédictives de détection d'OOD

Méthode d'OOD classique : Ajouter de l'OOD dans les données d'entrainement.

Problèmes :

- Biaise les détections
- Ne peut pas inclure toutes les sources d'OOD

Méthodes classiques prédictives de détection d'OOD

Méthode d'OOD classique : Ajouter de l'OOD dans les données d'entrainement.

Problèmes :

- Biaise les détections
- Ne peut pas inclure toutes les sources d'OOD

Idée : Détecter les données OOD avec un réseau Transformer

OODFormer

Architectures utilisées :

- ViT (**Vi**sion **T**ransformer)
- DEIT (**D**ata **E**fficient **I**mage **T**ransformer)

OODFormer

Architectures utilisées :

- ViT (**Vi**sion **T**ransformer)
- DEIT (**D**ata **E**fficient **I**mage **T**ransformer)

Ensembles d'entrainement :

- CIFAR-10
- CIFAR-100
- ImageNet30

Architecture ViT

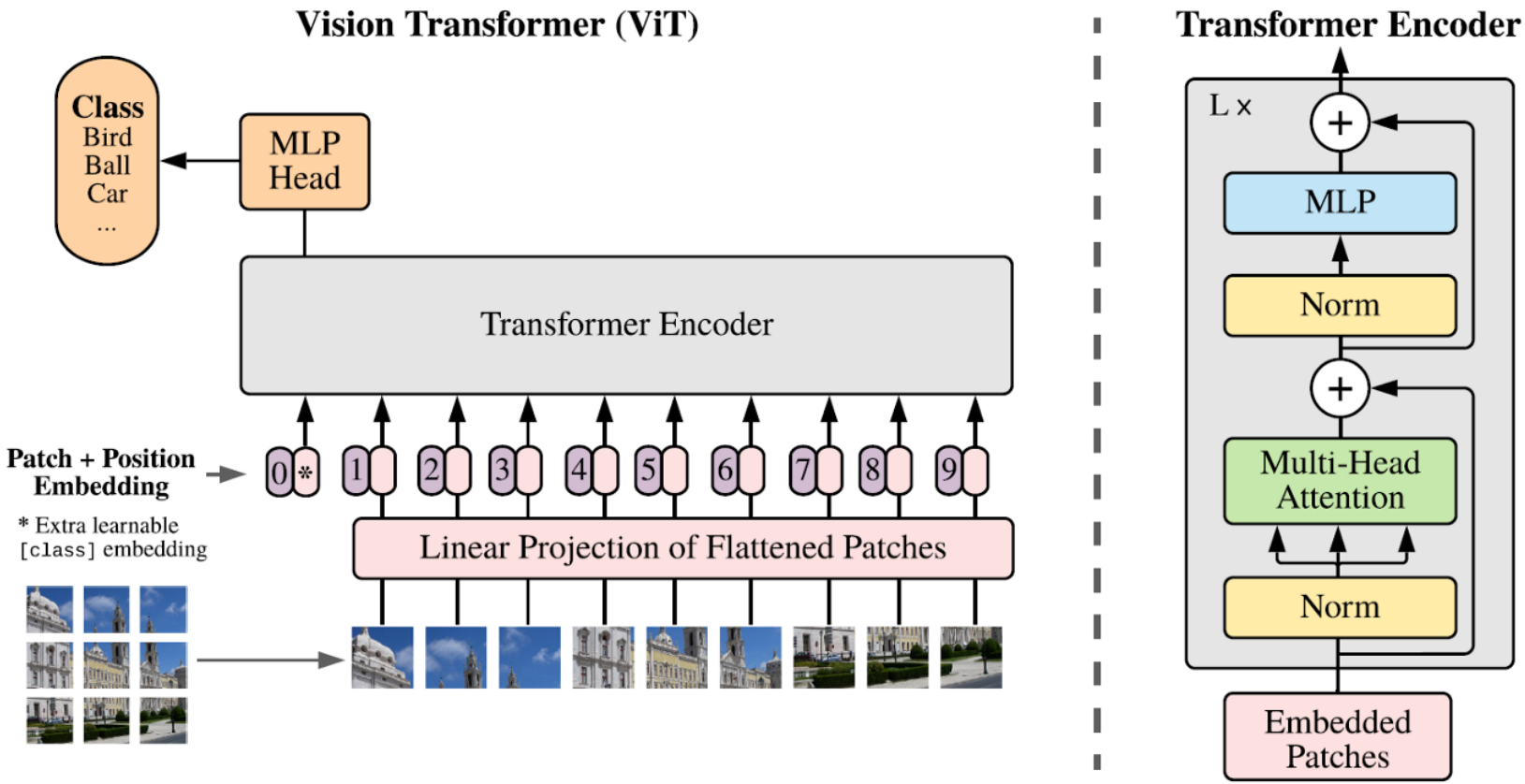

Figure – Architecture VIT

Principe d'OODFormer

Critères de discrimination

- Confiance de la prédiction Softmax
- Distance à la classe la plus similaire

Principe d'OODFormer

Critères de discrimination

- Confiance de la prédiction Softmax
- Distance à la classe la plus similaire

Hypothèse : Une donnée qui n'est pas dans la distribution d'entraînement entraînera soit une confiance très faible dans la prédiction, soit une distance importante dans l'espace des features.

Pourquoi un transformer ?

Attention : Mesure la similarité entre patchs d'une même image, en fonction de sa classe

Pourquoi un transformer ?

Attention : Mesure la similarité entre patchs d'une même image, en fonction de sa classe

Echantillon ID : Attention focalisée sur l'objet classifié

Pourquoi un transformer ?

Attention : Mesure la similarité entre patchs d'une même image, en fonction de sa classe

Echantillon ID : Attention focalisée sur l'objet classifié

Echantillon OOD : Attention moins focalisée (décors), prédiction non fiable

Pourquoi un transformer ?

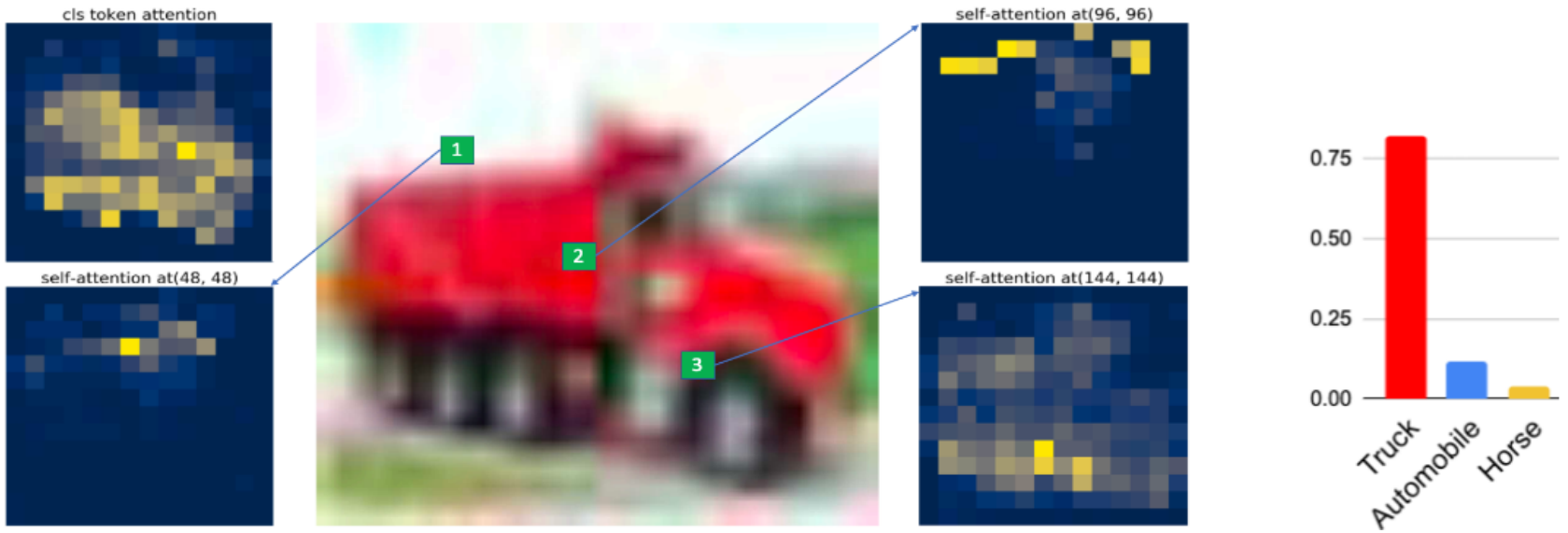

Figure – Attention pour un échantillon dans la distribution d'entraînement.

Pourquoi un transformer ?

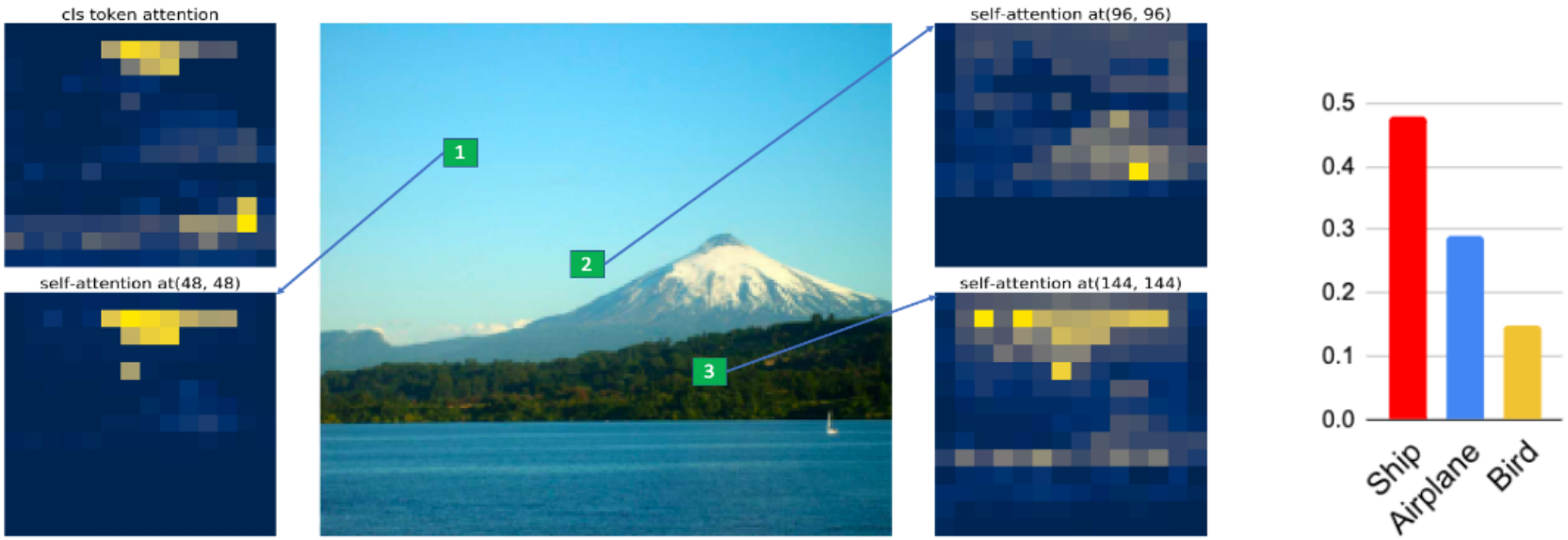

Figure – Attention pour un échantillon hors de la distribution d'entraînement.

Conclusion

	MoodCat	**OODFormer**
Méthode	génératif	prédictif
Cadre d'application	binaire	multi-classe
Technologie	VAE	Transformer
Nécessite une base OOD	Non	Non

Bibliographie

[1] Rajat Koner et al. *OODformer : Out-Of-Distribution Detection Transformer*. 2021. DOI : 10.48550/ARXIV.2107.08976. URL : https://arxiv.org/abs/2107.08976.

[2] Jingkang Yang et al. *Generalized Out-of-Distribution Detection : A Survey*. 2021. DOI : 10.48550/ARXIV.2110.11334. URL : https://arxiv.org/abs/2110.11334.

[3] Yijun Yang, Ruiyuan Gao et Qiang Xu. *Out-of-Distribution Detection with Semantic Mismatch under Masking*. 2022. DOI : 10.48550/ARXIV.2208.00446. URL : https://arxiv.org/abs/2208.00446.