

TriCCOT

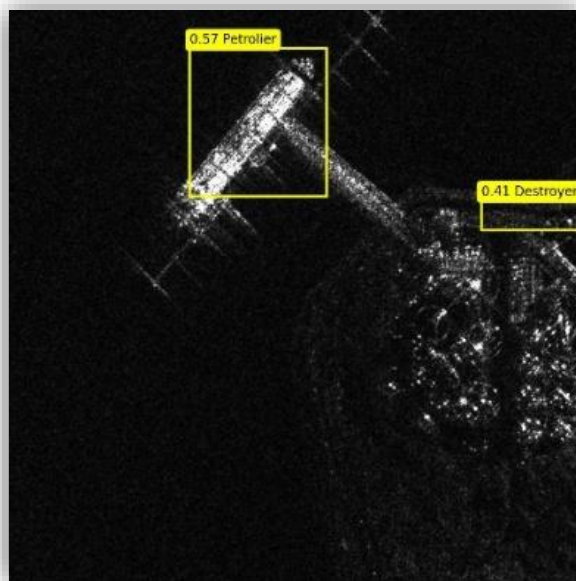
A new paradigm for AI in space

Adrien DORISE – CNES / IRT
PFIA 2026

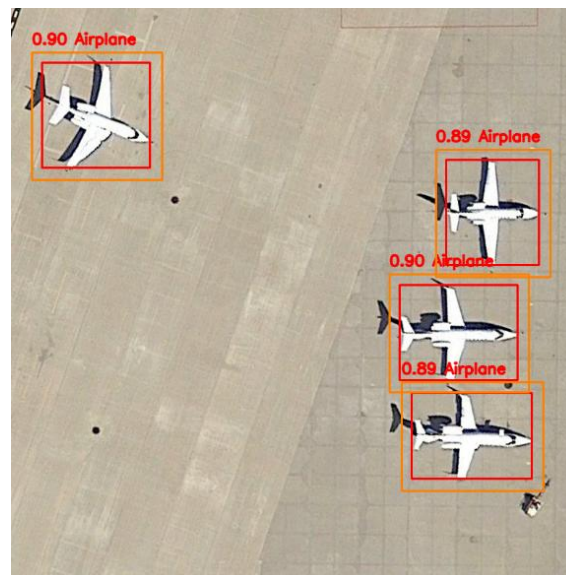
Intelligence artificielle et remote sensing



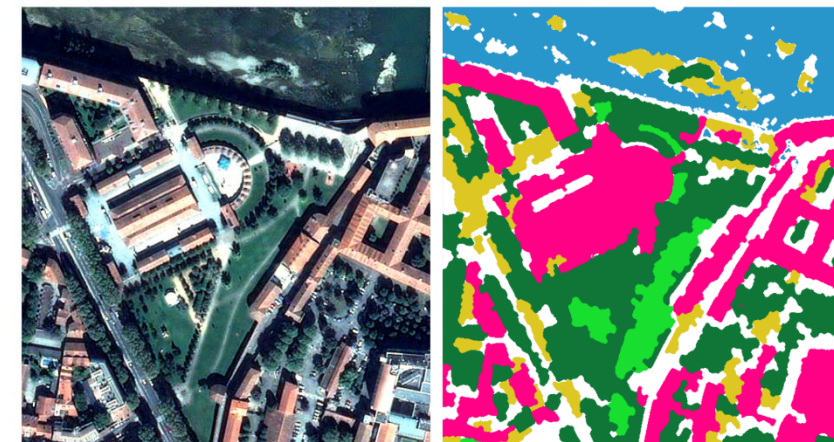
- L'IA est omniprésente lorsqu'il s'agit **d'extraire des informations** à partir des images d'observations de la Terre.
 - Land cover, détection de navire, détection de changement
- Ces traitements sont réalisés au sol.



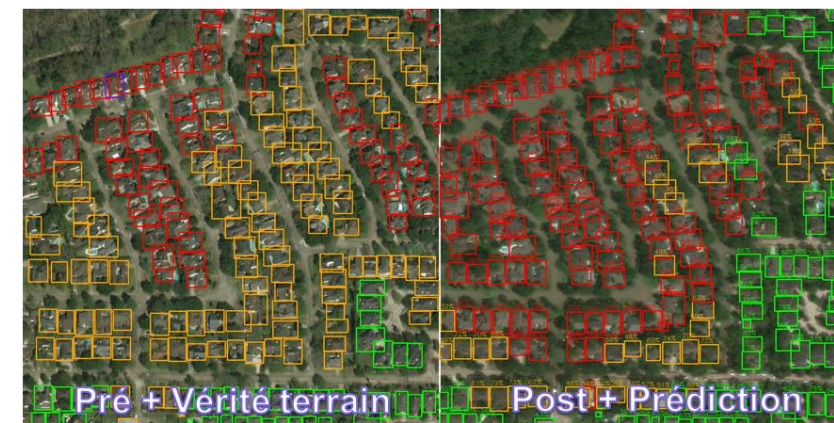
Surveillance maritime



Détection d'objet



Land Use



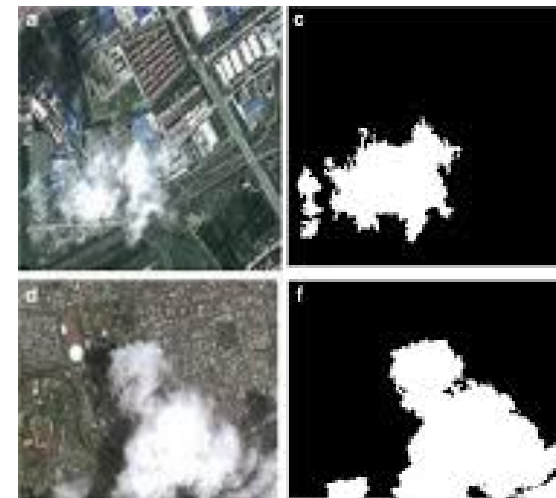
Evaluation de dégâts

Intérêt de l'IA bord

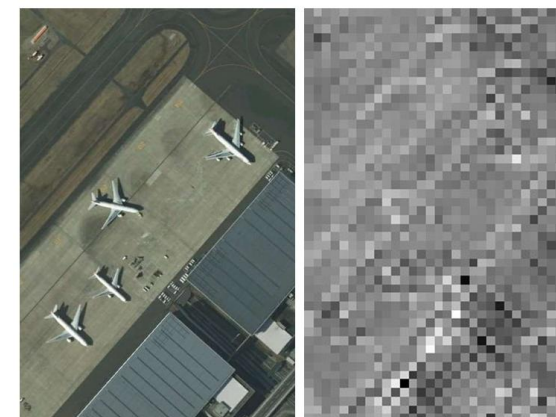


Sélectivité

- Réduire le volume de données transmis.
- Eviter le traitement d'images ne contenant pas d'information pertinente.



Cloud detection



Compression



Sélection de patches pour l'imagerie haute résolution

Intérêt de l'IA bord

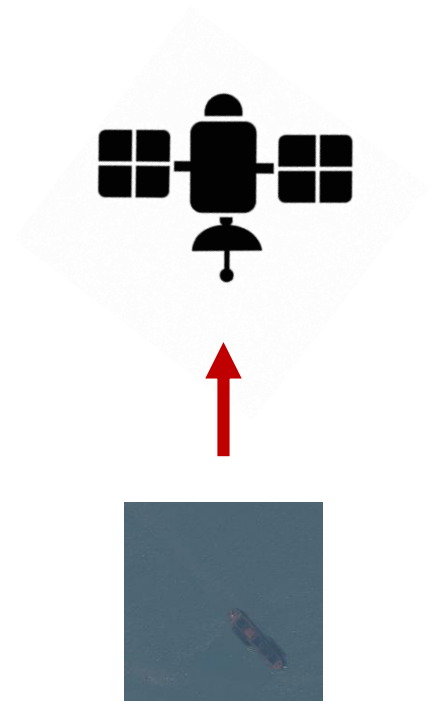


Sélectivité

- Réduire le volume de données transmis.
- Eviter le traitement d'images ne contenant pas d'information pertinente.

Réactivité

- Diminuer la latence entre le moment où l'information est perçue par le satellite, et le moment où l'information est exploitée.



Intérêt de l'IA bord

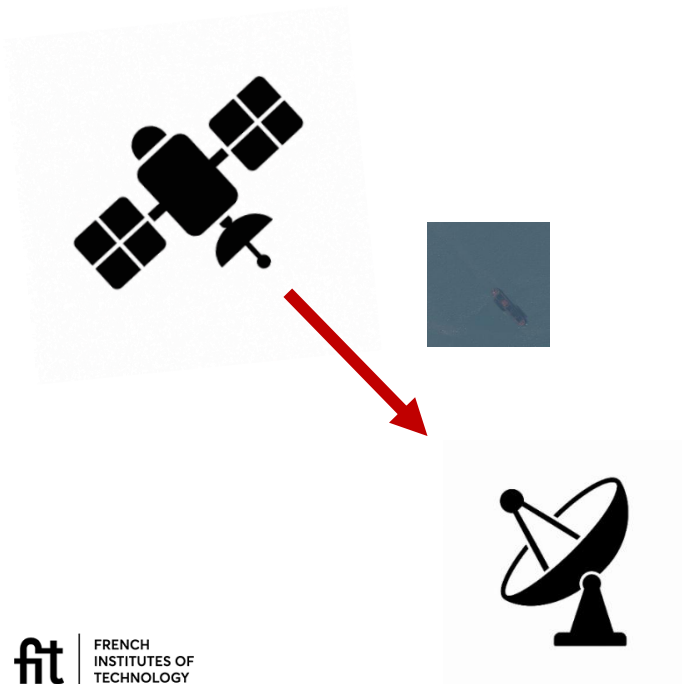


Sélectivité

- Réduire le volume de données transmis.
- Eviter le traitement d'images ne contenant pas d'information pertinente.

Réactivité

- Diminuer la latence entre le moment où l'information est perçue par le satellite, et le moment où l'information est exploitée.



Interêt de l'IA bord

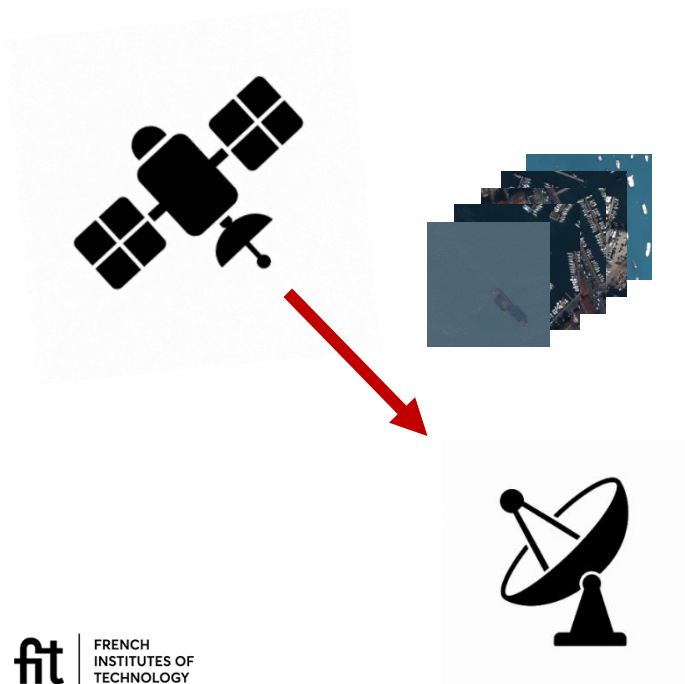


Sélectivité

- Réduire le volume de données transmis.
- Eviter le traitement d'images ne contenant pas d'information pertinente.

Réactivité

- Diminuer la latence entre le moment où l'information est perçue par le satellite, et le moment où l'information est exploitée.



Intérêt de l'IA bord

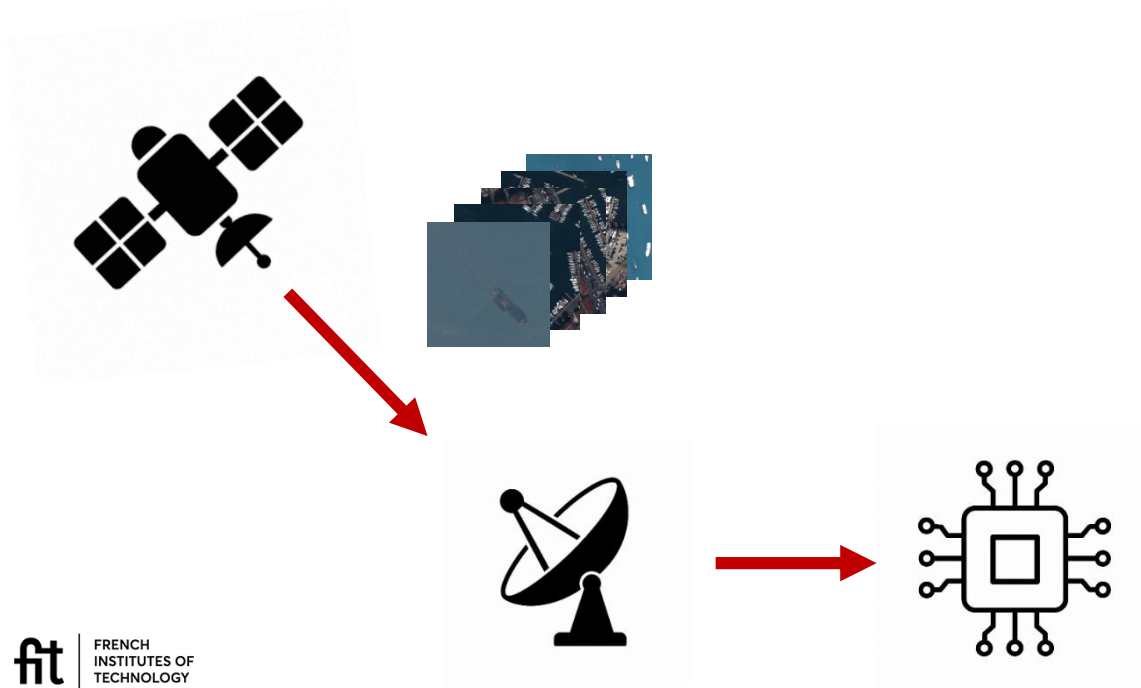


Sélectivité

- Réduire le volume de données transmis.
- Eviter le traitement d'images ne contenant pas d'information pertinente.

Réactivité

- Diminuer la latence entre le moment où l'information est perçue par le satellite, et le moment où l'information est exploitée.



Intérêt de l'IA bord

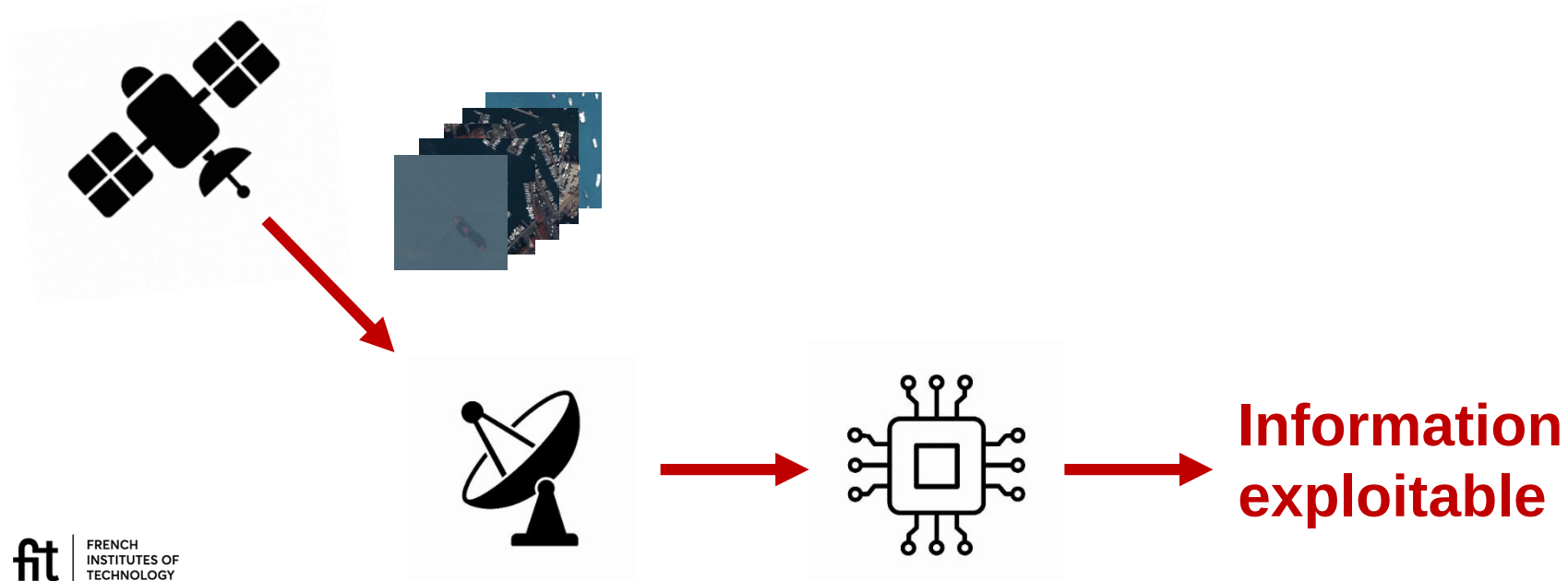


Sélectivité

- Réduire le volume de données transmis.
- Eviter le traitement d'images ne contenant pas d'information pertinente.

Réactivité

- Diminuer la latence entre le moment où l'information est perçue par le satellite, et le moment où l'information est exploitée.



Intérêt de l'IA bord



Sélectivité

- Réduire le volume de données transmis.
- Eviter le traitement d'images ne contenant pas d'information pertinente.

Réactivité

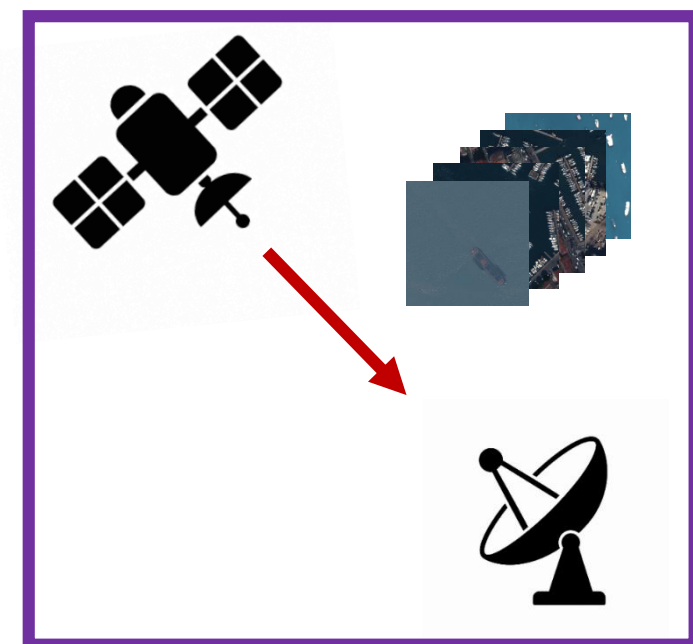
- Diminuer la latence entre le moment où l'information est perçue par le satellite, et le moment où l'information est exploitée.
- Permettre une **priorisation** des informations.



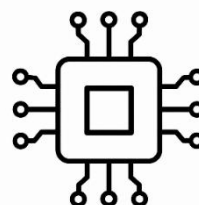
Wildfire detection



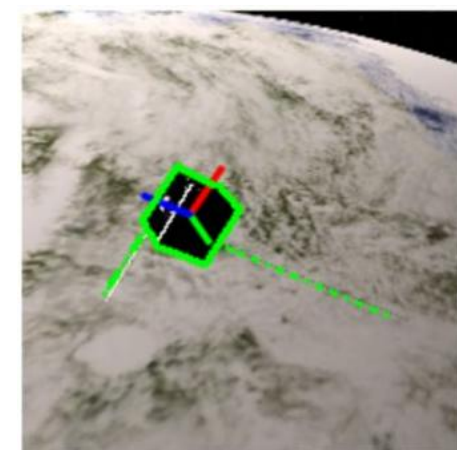
Tracking



IA bord pour un traitement accéléré



Information exploitable



Pose estimation for space RDV

La problématique de l'image brute



Une vision dégradée

- Les traitements IA des images satellites sont réalisés sur des images ayant été restaurées (produits L1C, L2A,...).
- Reproduire fidèlement ces chaînes de traitement nécessite d'importantes ressources de calcul, incompatibles avec le traitement bord.



Université d'Artois vu sur Google Maps

La problématique de l'image brute



Une vision dégradée

- Les traitements IA des images satellites sont réalisés sur des images ayant été restaurées (produits L1C, L2A,...).
- Reproduire fidèlement ces chaînes de traitement nécessite d'importantes ressources de calcul, incompatibles avec le traitement bord.



Université d'Artois vu sur Google Maps



Université d'Artois vu par un satellite

La problématique de l'image brute



Devrais-je m'en soucier ?

La problématique de l'image brute



Devrais-je m'en soucier ?

OUI

La problématique de l'image brute

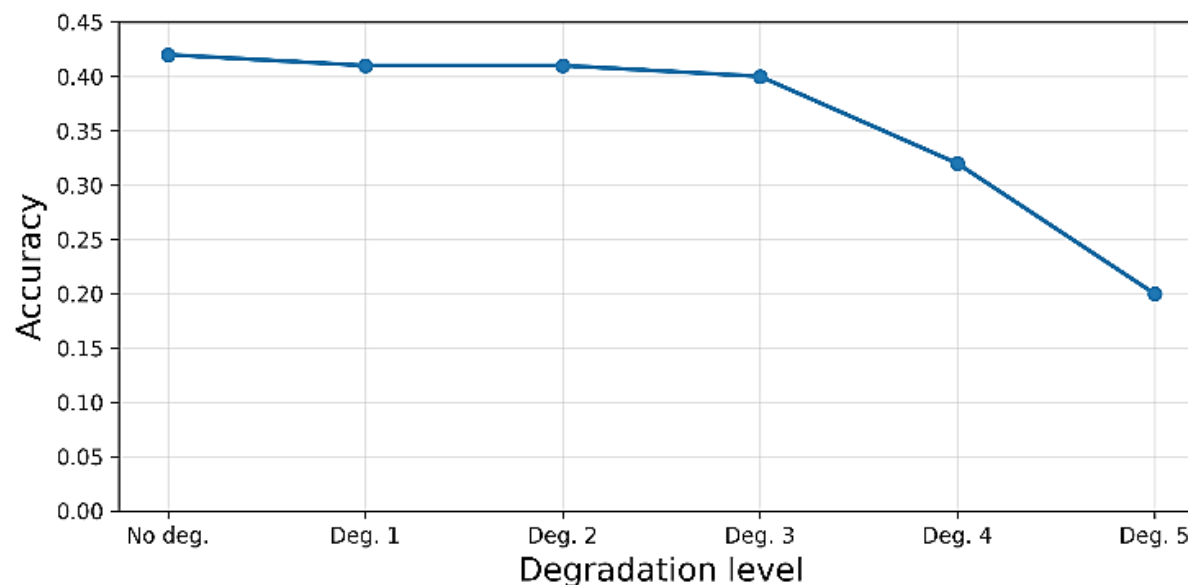


Devrais-je m'en soucier ?

- Nos travaux ont montré que les images brutes dégradent les performances des modèles IA (détection de navire).

Nos solutions, compatibles CPU et FPGA :

- Une méthode de restauration compatible avec le hardware satellite.
- Un modèle capable de traiter directement la donnée brute.



Performance des modèles de détection en fonction du niveau de dégradation

Restauration d'images embarquée : ConvBEERS



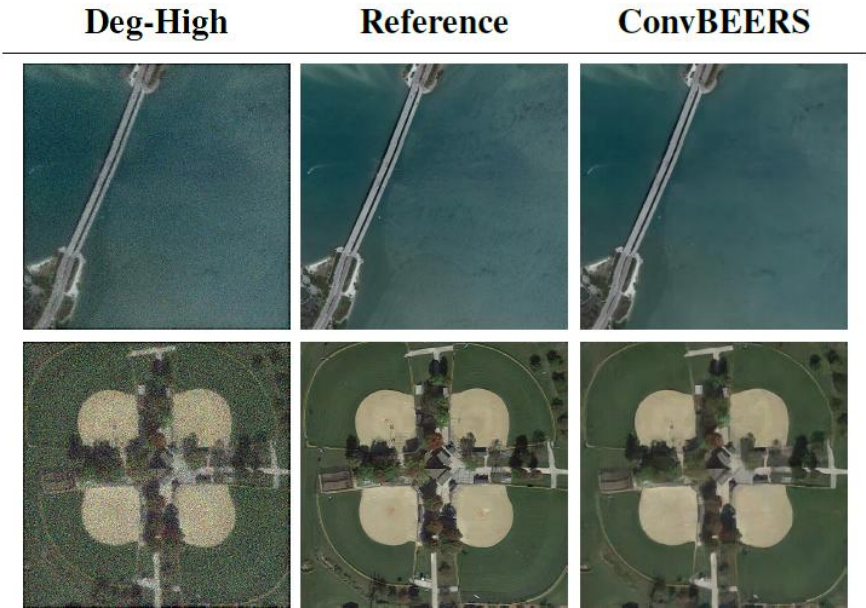
ConvBEERS est notre alternative aux techniques de restauration traditionnelles

ConvBEERS est testé sur un pipeline de détection d'objets avec YOLO-X.

- Résultat de détection : ConvBEERS **améliore significativement la détection.**
- Résultat embarqué : ConvBEERS **s'exécute en temps réel** sur car FPGA.

Conclusion

ConvBEERS permet l'intégration de tâches IA à bord en temps réel.



Exemples de restauration

Dataset	Set	mAP50	mAP90	Prec.	Rec.
Reference	N/A	75.90%	54.10%	86.60%	69.30%
Deg-High	Deg.	67.30%	46.29%	63.80%	60.70%
	Rest.	72.37%	50.57%	85.03%	66.36%

Résultats de détection

Method	Latency (s)↓	FPS↑	Power (W)↓
Baseline (CPU)	1724	0.2	22.5
ConvBEERS (DPU)	42	7.2	29.6

Performances hardware sur Xilinx Versal

La détection d'objet sur images brutes



Des modèles qui manquent de contexte

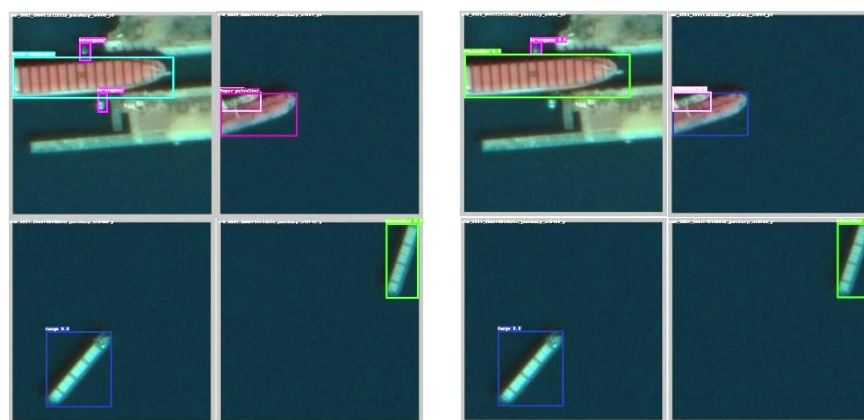
- Les analyses montrent que les modèles de détection sont aptes à localiser les objets d'intérêts dans les images brutes.
- C'est la **classification** qui est le plus impactée par la donnée brute.

Hypothèses

- La recherche **locale** des réseaux convolutionnels est perturbé sur des données dégradées
- Une recherche de contexte **global** devrait aider les modèles de détection.

Solution proposée

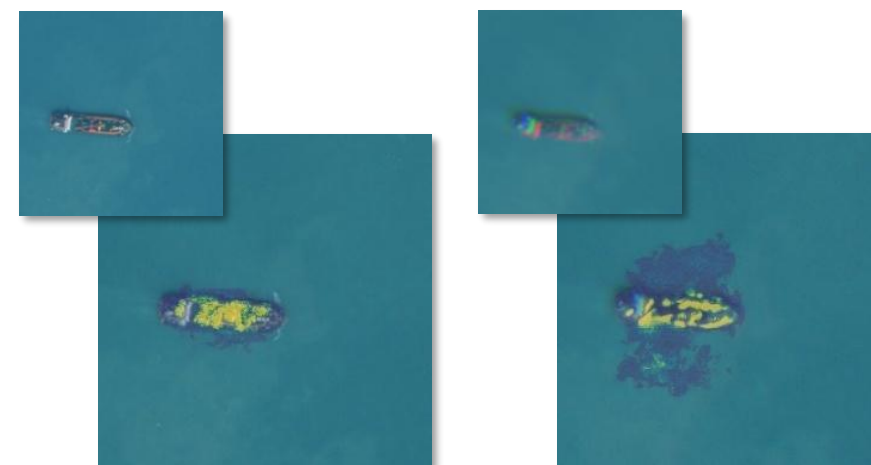
- Utiliser le **mécanisme d'attention** pour améliorer les performances des modèles.



Vérité terrain

Prédictions Yolo11n

Analyse des prédictions sur images brutes MAXAR



Saliency maps sur
image L1

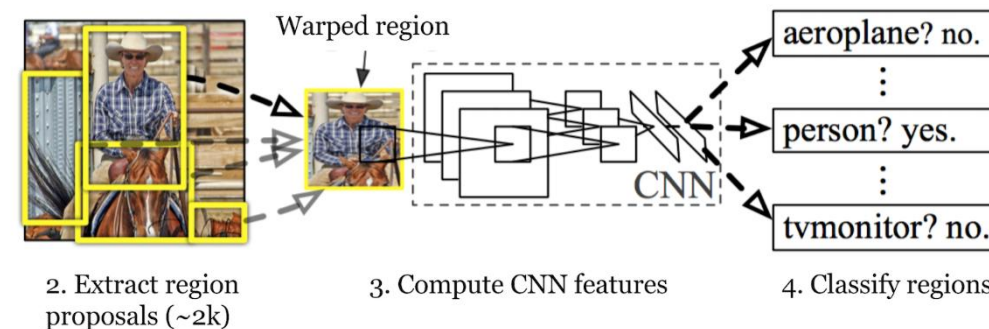
Saliency maps sur
image brute

TriCCOT : la détection d'objet sur images brutes

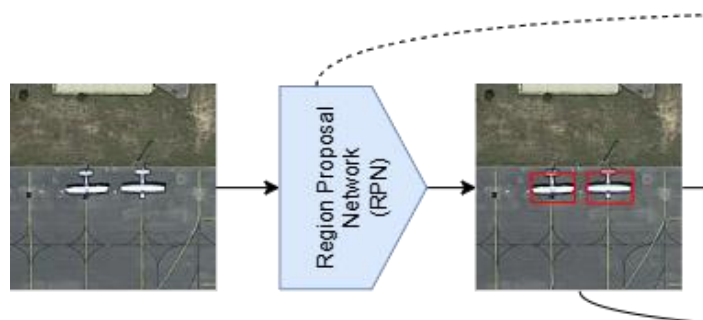


TriCCOT : Tri-part CNN Conformal Transformer

- **Region Proposal Network (RPN) :**
 - Détecte les objets et **propose des boîtes englobantes.**
 - Pas de classification.



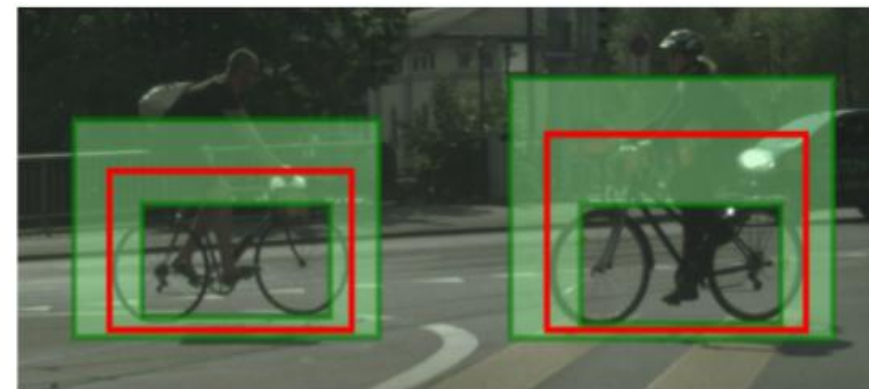
RPN similaire au modèle RCNN



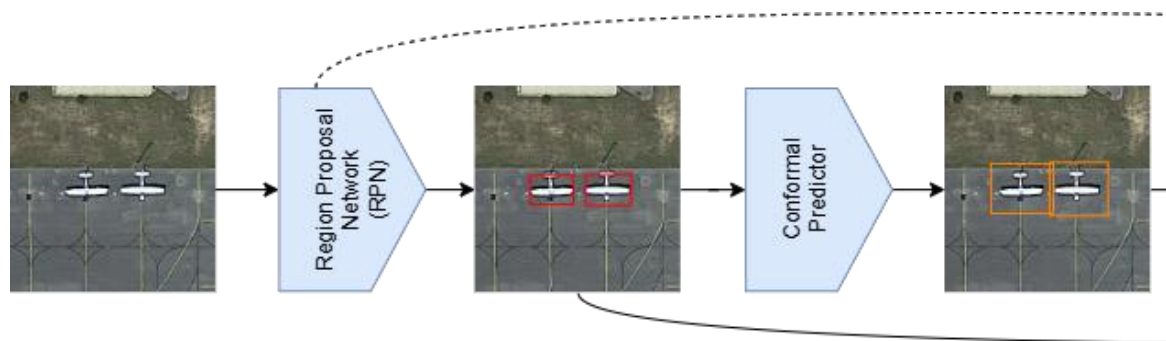
TriCCOT : la détection d'objet sur images brutes

TriCCOT : Tri-part CNN Conformal Transformer

- **Region Proposal Network (RPN) :**
 - Détecte les objets et **propose des boîtes englobantes.**
 - Pas de classification.
- **Prédicteur conforme :**
 - $\hat{Y}_{\text{new}}^j - d_{\alpha/4}^j, \quad j \in \{x_{\min}, y_{\min}, x_{\max}, y_{\max}\}$
 - **Compense les défauts** du RPN.
 - Agrandi les boîtes, permettant d'ajouter du contexte et d'éviter les problématiques de « snapped box ».



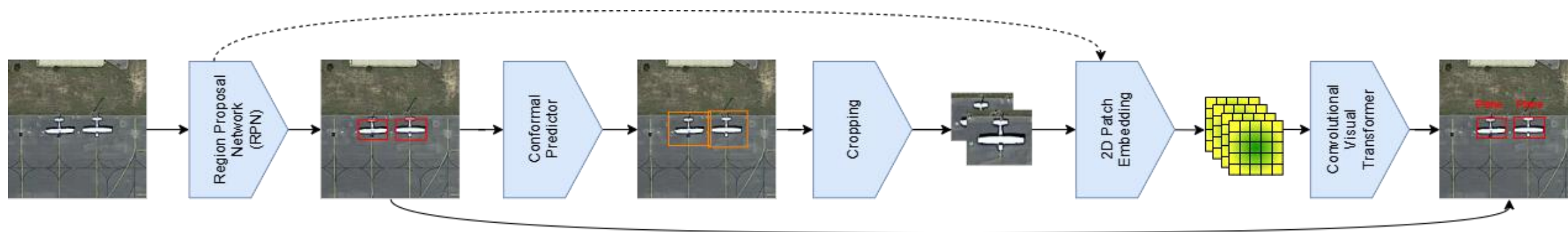
Exemple de prédiction conforme



TriCCOT : la détection d'objet sur images brutes

TriCCOT : Tri-part CNN Conformal Transformer

- **Region Proposal Network (RPN) :**
 - Détecte les objets et **propose des boîtes englobantes**.
 - Pas de classification.
- **Prédicteur conforme :**
 - $\hat{Y}_{\text{new}}^j - d_{\alpha/4}^j, \quad j \in \{x_{\min}, y_{\min}, x_{\max}, y_{\max}\}$
 - **Compense les défauts** du RPN.
 - Agrandi les boîtes, permettant d'ajouter du contexte et d'éviter les problématiques de « snapped box ».
- **Classification par un Visual Transformer (ViT) :**
 - Prédiction sur les objets croppés pour limiter la taille et maximiser l'information.
 - Mécanisme d'attention **pour combler les lacunes des réseaux convolutionnels**.



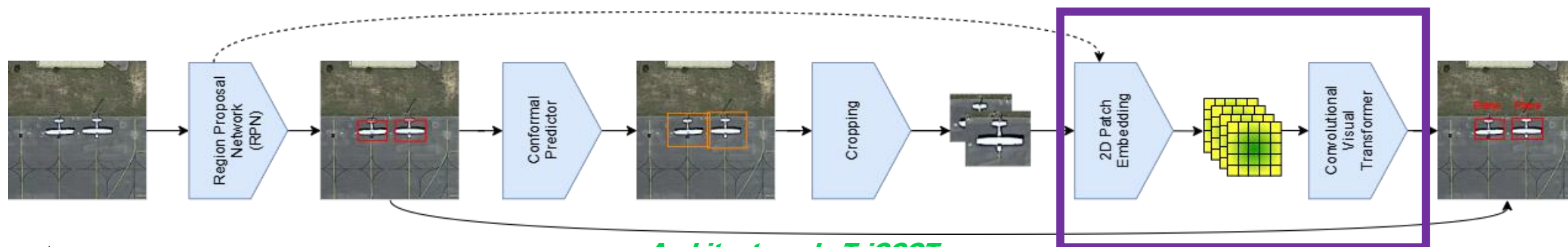
Architecture de TriCCOT

TriCCOT : la détection d'objet sur images brutes

TriCCOT : Tri-part CNN Conformal Transformer

- **Region Proposal Network (RPN) :**
 - Détecte les objets et **propose des boîtes englobantes**.
 - Pas de classification.
- **Prédicteur conforme :**
 - $\hat{Y}_{new}^j - d_{\alpha/4}^j, \quad j \in \{x_{min}, y_{min}, x_{max}, y_{max}\}$
 - **Compense les défauts** du RPN.
 - Agrandi les boîtes, permettant d'ajouter du contexte et d'éviter les problématiques de « snapped box ».
- **Classification par un Visual Transformer (ViT) :**
 - Prédiction sur les objets croppés pour limiter la taille et maximiser l'information.
 - Mécanisme d'attention **pour combler les lacunes des réseaux convolutionnels**.

Aper-GATES : un transformer sur FPGA



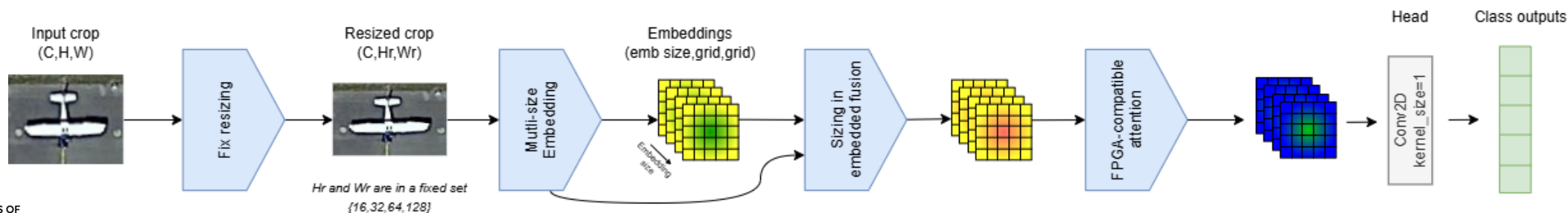
Architecture de TriCCOT

Aper-GATES : un transformer sur FPGA



Un transformer en plusieurs parties

- L'image est redimensionnée selon un set de dimensions prédéfini.
- L'image passe dans un **multi-size embedding** qui vient encoder l'information.
- L'embedding sert d'input à notre **mécanisme d'attention**.
- La **matrice d'attention** est ensuite utilisée pour **définir la classe de l'objet**.



Aper-GATES : un transformer sur FPGA

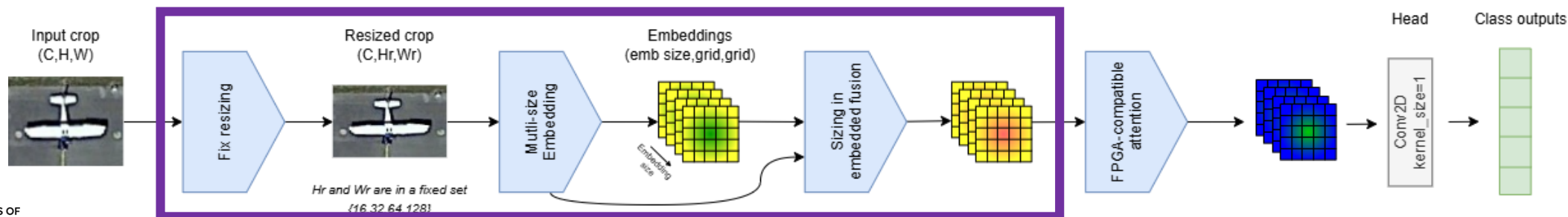


Multi-size embedding

- Les images en entrée du transformer sont des crops réalisées à partir des boîtes englobantes du prédicteur conforme.
 - Les tailles des crops ne sont pas fixes**, et varient en fonction de la taille de l'objet à classifier.
- Cependant, l'information liée à la taille et au ratio d'un objet est utile pour déterminer sa classe.



Exemple de taille d'objets variable

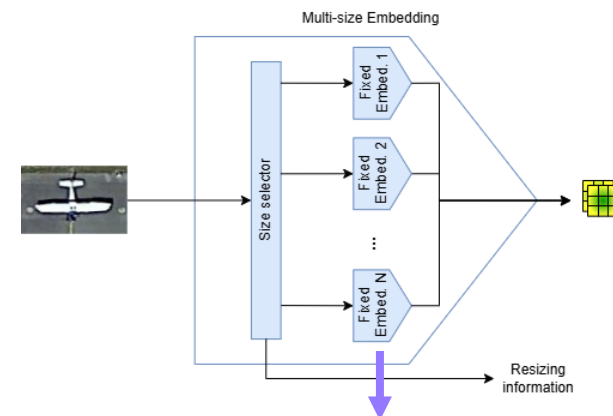


Aper-GATES : Aperture-Gated Attention Transformer for Embedded Systems

Aper-GATES : un transformer sur FPGA

Multi-size embedding

- Les images en entrée du transformer sont des crops réalisées à partir des boîtes englobantes du prédicteur conforme.
 - Les tailles des crops ne sont pas fixes**, et varient en fonction de la taille de l'objet à classifier.
- Cependant, l'information liée à la taille et au ratio d'un objet est utile pour déterminer sa classe.
- L'image est redimensionnée selon **des tailles prédéfinies** (16,32,64,128), et traitée dans un encodeur spécifique.
- L'encoder s'assure que l'espace latent **en sortie possède une taille fixe**.

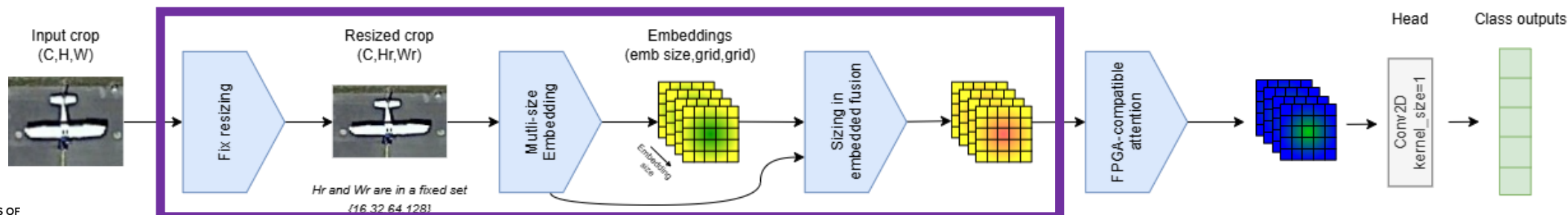


$$E_i(x) = \phi_i^{(2)}(\phi_i^{(1)}(x))$$

$$\phi_i^{(1)}(x) = \text{ReLU}\left(\text{BN}\left(\text{Conv}(x; s_{H,1}^{(i)}, s_{W,1}^{(i)})\right)\right)$$

$$\phi_i^{(2)}(z) = \text{ReLU}\left(\text{BN}\left(\text{Conv}(z; s_{H,2}^{(i)}, s_{W,2}^{(i)})\right)\right)$$

$$s_{H,1}^{(i)} s_{H,2}^{(i)} = \frac{H_i}{G}, \quad s_{W,1}^{(i)} s_{W,2}^{(i)} = \frac{W_i}{G}$$



Aper-GATES : Aperture-Gated Attention Transformer for Embedded Systems

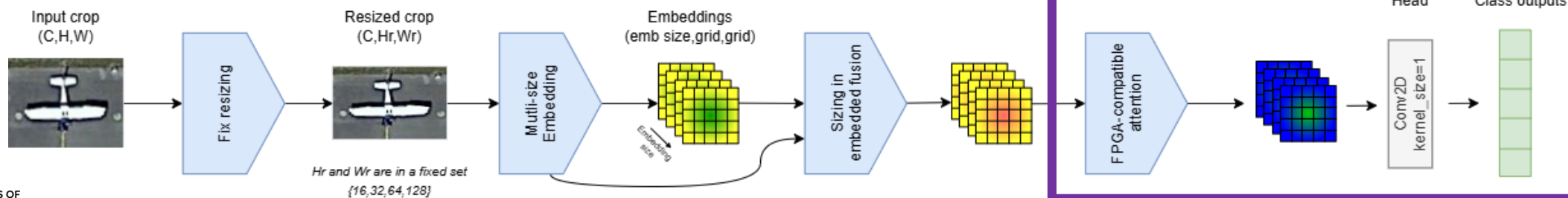
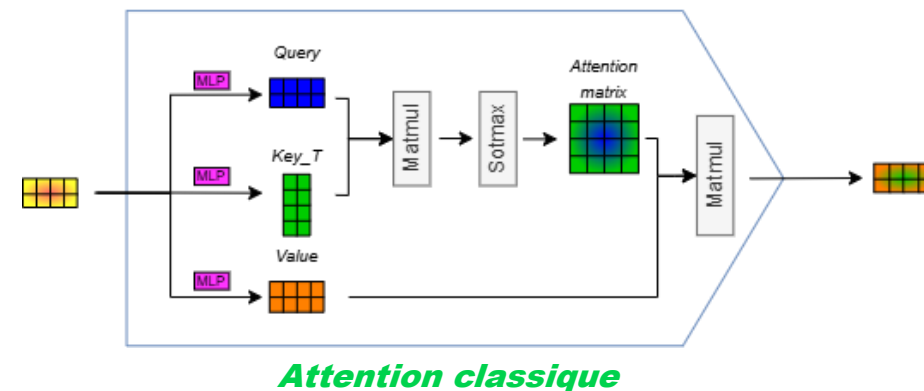
Aper-GATES : un transformer sur FPGA



Le mécanisme d'attention n'est pas compatible avec un déploiement FPGA

$$Attention = Softmax\left(\frac{Q \cdot K^T}{scale}\right) \cdot V$$

- Produit de matrices à hautes dimensions.
- Complexité spatiale quadratique $O(n^2)$.



Aper-GATES : Aperture-Gated Attention Transformer for Embedded Systems

Aper-GATES : un transformer sur FPGA



Le mécanisme d'attention n'est pas compatible avec un déploiement FPGA

$$Attention = Softmax\left(\frac{Q \cdot K^T}{scale}\right) \cdot V$$

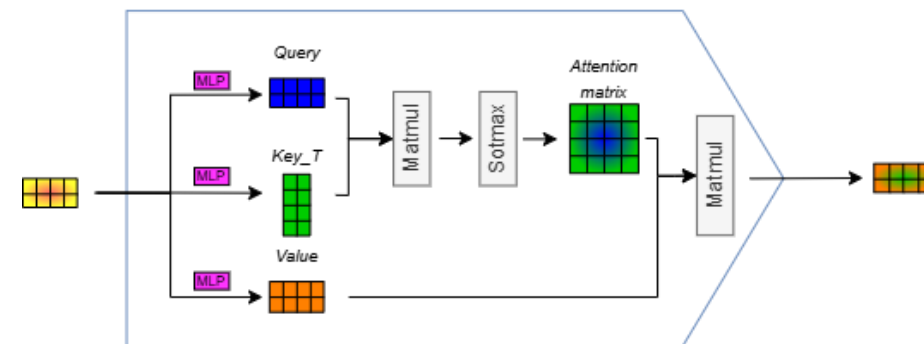
- Produit de matrices à hautes dimensions.
- Complexité spatiale quadratique $O(n^2)$.

Notre solution : Aper-GATES

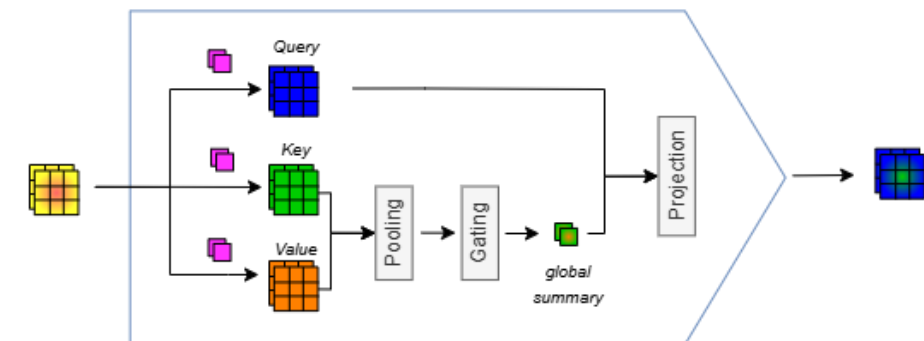
$$AperGATES = \phi(\sigma(\text{Pool}(K \odot V) \odot Q)$$

- Aper-GATES reprend le principe du « **gated mechanism** », utilisant une fonction σ pour contrôler l'importance d'une information globale.
- Utilisation d'**element-wise multiplication** $K \odot V$.
- Ici, on évalue l'importance des informations **contenues dans l'image**.

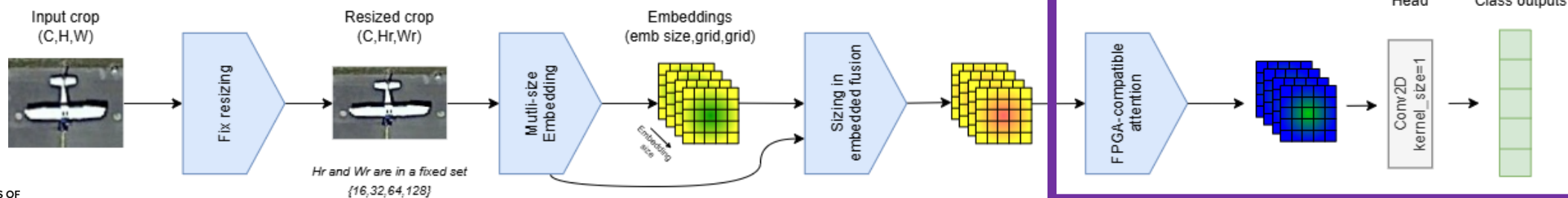
AperGATES : l'un des premiers transformers sur DPU



Attention classique



Attention Aper-GATES



Aper-GATES : Aperture-Gated Attention Transformer for Embedded Systems

TriCCOT : Un modèle robuste à la donnée brute



Résultats

TriCCOT est testé sur les datasets DIOR original & DIOR dégradé.

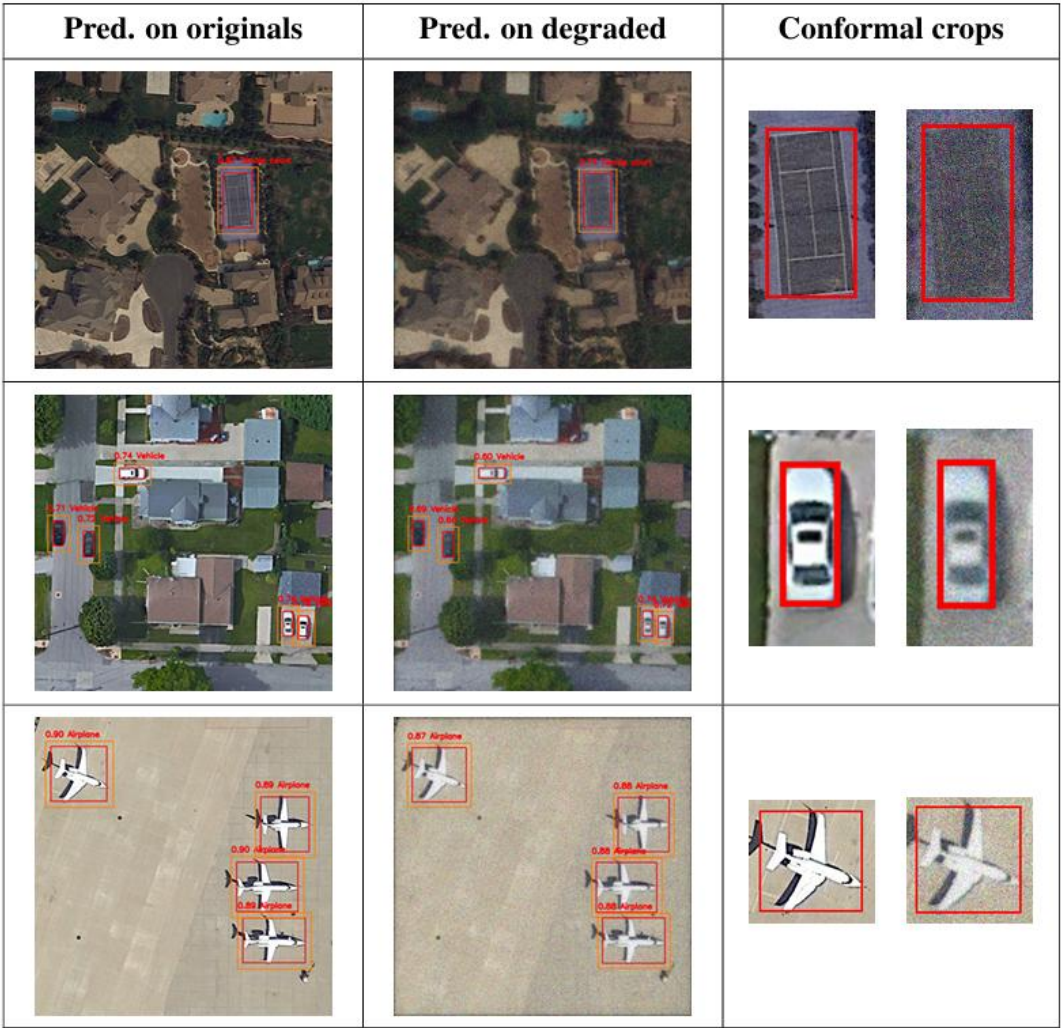
- TriCCOT est proche de l'état de l'art sur les images restaurées.
- TriCCOT **surpasse l'état de l'art** sur les images dégradées.
- TriCCOT permet un **traitement temps réel** avec ~30 FPS sur Xilinx Versal.

Method	Params	GFlops	Original		Raw	
			mAP@50	mAP@50:95	mAP@50	mAP@50:95
Faster R-CNN [46]	40.0 M	89.0	74.79	47.37	69.24	40.41
YoloX-S [19]	8.9 M	26.8	78.73	46.07	72.55	40.07
DeTR [8]	41.5 M	38.4	49.85	28.70	49.10	27.68
TriCCOT (ours)	3.5 M	8.2	76.49	45.48	74.77	42.16

Résultat sur le dataset DIOR

Method	Hardware	Prec.	Size (Mb)	Latency (ms) ↓	FPS ↑	Power (W) ↓	Pxls/s/W ↑
DETR [8]	CPU	float32	162.7	14 266.5	0.07	22.5	1148
YOLOX-S [19]	DPU	int8	9.2	74.1	13.5	25	251 257
TriCCOT (ours)	DPU	int8	6.4	35.4	28.6	25	462 824

Résultat hardware sur Xilinx Versal



Prédictions sur le dataset DIOR

Conclusion



Les verrous ralentissant l'utilisation de l'IA à bord des satellites se réduisent !

- Les coûts pour porter les algorithmes diminuent.
- ConvBEERS et TriCCOT permettent de gérer la donnée brute.
- L'IA bord permet de nouveaux cas d'usage inaccessibles avant.

N'hésitez pas à me contacter si le sujet vous intéresse !



Merci !