

DE PLÉIADES À PLÉIADES NEO : ENJEUX ET OPPORTUNITÉS. L'EXPÉRIENCE D'UN PRODUCTEUR DE DONNÉES.

Sophie Delpierre¹, Maxime Chauvin², Nahossio Gongga-Saholiariliva¹, Camille Lhenry¹, Yann Desmarais¹, Jérôme Leroux², Jules Roger², Baptiste Benard², Philippe Albert¹, Ivan Pristhepa¹

1: Geofit Montpellier, Le Syracuse, 2 av. Monteroni d'Arbia, 34920 Le Crès

2: Geofit Nantes, 7 rue Alfred Kastler, 44307 Nantes

Résumé

Avec 50 années d'expérience dans le traitement de l'information géospatiale, le Groupe GEOFIT intervient entre autres dans les domaines de la photogrammétrie et de la télédétection. Une de ses entités emploie plus de 50 personnes en France pour produire des données cartographiques, en deux ou trois dimensions, à partir d'acquisitions aériennes ou de données satellitaires. GEOFIT Montpellier est notamment, en tant que sous-traitant, attributaire depuis 2016 d'un marché de la Direction Générale de l'Armement destiné à la production de cartes au 50 000^{ème} et de plans de villes au 5000^{ème} : le programme GEOMAPS, qui fait suite au programme TOPOBASE des années 2005-2015. Depuis, les équipes de GEOFIT ont visualisé, traité, ajusté, étalonné, échantillonné, des millions de km² d'images Pléiades pan-sharpées à 50 cm de résolution. Ces mêmes images ont été utilisées pour l'entraînement de modèles d'intelligence artificielle dans un objectif de produire des couches d'appui internes à la production vecteur. Plusieurs dizaines de photo-interprètes ont "pané/balayé" et interprété de multiples paysages, urbains, péri-urbains, ruraux, plus ou moins anthropisés et ont eu à produire des bases de données vectorielles conformes aux exigences militaires. Depuis l'année 2023, ces mêmes équipes de data-scientists, ingénieurs, analystes images, photo-interprètes font évoluer leur métier en appliquant leur expérience aux nouvelles images issues de Pléiades Neo. Quels sont les implications et impacts ? Est-ce une simple évolution ou un "game changer" ? Il y a-t-il une opportunité à utiliser de nouveaux canaux ? Qu'en est-il des standards vecteurs ? L'IA sait-elle et peut-elle faire des prédictions Pléiades Neo depuis un apprentissage Pléiades ? GEOFIT a évoqué ces questions lors du séminaire Pléiades Neo du 5 octobre 2023 (Toulouse, France).

Mots clés : Production vecteur, Optique très haute résolution, Pléiades, Pléiades Neo, Photo-interprétation, Résolution, Résolution mixte, Apprentissage profond, Segmentation sémantique, Apprentissage par transfert

Abstract

GEOFIT has 50 years' experience in geospatial information processing and is involved in fields such as photogrammetry and remote sensing. GEOFIT Montpellier employs over 50 people in France to produce two- and three-dimensional cartographic data from aerial and satellite acquisitions. As a subcontractor, GEOFIT Montpellier has been awarded a contract since 2016 by the French Defence Procurement Agency (DGA) to produce 50,000-scale maps and 5,000-scale city maps : the Geomaps program, following on the Topobase program of 2005-2015. Meanwhile, GEOFIT teams have visualized, processed, adjusted, calibrated and sampled millions of km² of Pleiades pan-sharpened images at 50 cm resolution. These images were used to train artificial intelligence models, with the aim of producing internal support layers for vector production. Several dozen photo-interpreters "panning" and interpreting multiple landscapes, urban, peri-urban, rural, more or less anthropized, had to produce vector databases in line with military requirements. Since 2023, these same teams of data scientists, engineers, image analysts and photo-interpreters have been evolving their profession by applying their experience to the new images produced by Pleiades Neo. What are the advantages and disadvantages ? Is this a simple evolution or a 'game changer' ? Are the new channels useful ? What about vector standards ? Is AI capable of making Pleiades Neo inferences from Pleiades training ? GEOFIT addressed these questions at the Pleiades Neo Seminar held in Toulouse (France) on October 5, 2023.

Keywords : Vector Production, VHR, PNEO, PHR, Digitation mapping, Resolution, Combined Resolution, Deep Learning, Semantic Segmentation, Transfer learning

1. Introduction

La photo-interprétation est un processus d'analyse et d'extraction d'informations à partir d'images géoréférencées (photographies aériennes ou images satellite) permettant de comprendre la surface terrestre et ses caractéristiques. En tant que producteur, GEOFIT a développé des chaînes de production vecteur par photo-interprétation s'adaptant régulièrement aux avancées technologiques, que l'on se place au niveau de l'imagerie et des capteurs d'acquisition ou au niveau algorithmique et software avec le développement d'outils d'aide à la photo-interprétation.

Dans le cadre de marchés Défense, GEOFIT intervient comme producteur de données vectorielles. Jusqu'alors, les équipes de production ont pu bénéficier d'imagerie Très Haute Résolution (THR) Pléiades Haute Résolution (PHR) de résolution 50 cm. Elles exploitent depuis quelques mois des images satellite de résolution encore plus fine, *i.e.* des Pléiades Neo (PNEO) dont la résolution de 30 cm est quasi équivalente à des orthophotographies de résolution similaire. Cette gamme de résolution est exploitée dans certains projets (hors Défense) à des échelles de rendu cartographique inférieures au 1/3000^{ème} en zone rurale et au 1/1000^{ème} en zone urbaine et péri-urbaine (Figure 1).



FIGURE 1 – Compositions Colorées vraies couleurs, CC-VC, en zone urbaine (1/500^{ème}). **Haut** : PHR, **Bas** : PNEO.

Une évaluation de l'impact de l'intégration de ces nouvelles données dans notre outil de production nous a permis de saisir cette nouveauté comme une opportunité d'évaluer nos processus et la donnée d'entrée. Cela nous a permis également de faire évoluer en interne la mise à disposition de "couche d'appui" THR provenant

d'indicateurs innovants issus de modèles utilisant l'apprentissage profond. Ces couches d'appui sont souvent des masques vectorisés issus de prédictions des modèles avec ou sans post-traitement (vectoriel ou image). Elles servent de support et d'aide à la photo-interprétation.

Dans le cadre de notre présentation et de cet article, nous abordons les avantages et inconvénients compilés lors des premières exploitations des images PNEO sur certains chantiers cartographiques.

Dans un premier temps, le passage des PHR aux PNEO est évalué empiriquement et qualitativement, nous parlons du ressenti et de l'expertise des photo-interprètes.

Dans un second temps, nous exposons quels sont les apports de ces données dans nos processus de production de couches d'appui internes pour aider nos photo-interprètes à l'extraction vecteur.

2. Photo-interprétation

Les sections qui suivent ont pour objectif d'illustrer les éléments essentiels relevés par les photo-interprètes lors de l'utilisation de données multi-sources (images) ou du passage d'un type d'image à un autre avec une ou des mises à jour à l'échelle. Les exemples les plus significatifs sont présentés afin d'éclairer tant le processus de production des données vecteur que le travail du photo-interprète, et ce, du point de vue du producteur de données. Les images utilisées sont pan-sharpées, la résolution est de 50 cm (70 cm ré-échantillonnées à 50 cm) pour les PHR, la résolution est de 30 cm pour les PNEO.

GEOFIT réalise la photo-interprétation pour les productions GEOMAPS en exploitant la bande proche-infrarouge¹ et en utilisant des compositions colorées infrarouges (*i.e.* Proche-Infrarouge / Rouge / Vert). Cette approche permet de mieux différencier la végétation des sols nus ainsi que d'autres éléments anthropiques (*cf.* Figure 2). Ces compositions colorées (CC) sont nommées CC-PIR dans la suite du document. Dans les sections suivantes, les images pourront être proposées en Composition colorée « vraies couleurs » à titre d'illustration, elles sont alors nommées CC-VC dans le reste du document. Les bandes Red Edge ou Deep Blue des PNEO ne sont pas exploitées sur nos différents chantiers.

2.1. Spécifications MUVD

Dans le cadre de GEOMAPS, pour la filière de production des plans de villes (F5), les règles d'extraction sont définies par le modèle de référence MUVD : MGCP Urban Vector Data (**Defence Geospatial Information Working Group (DGIWG 262) (2021)**) pour une cible cartographique 1/5000^{ème}). Cette cible cartographique, les règles d'extraction, restent identiques quel que soit le capteur utilisé. La Figure 4 illustre le fait que le rendu

1. Note : les vecteurs de production GEOMAPS ne sont pas présentés dans cette étude. Les illustrations vecteurs visibles sur certaines figures sont issus d'une saisie GEOFIT produite à des fins de communication

vectorel reste identique (les règles d'extraction étant similaires) quel que soit le capteur utilisé, bien que l'opérateur / photo-interprète appréhende différemment l'espace en fonction de la résolution et de la netteté de l'image.

Cependant, la précision de l'écart de pointé nécessaire pour remplir les spécifications GEOMAPS doit être explicitement inférieure ou égale à un pixel, quelle que soit la résolution de l'image. De ce fait, les objets dont les limites pouvaient être floues en PHR sont dorénavant concernés par les spécifications en PNEO et, du fait de la taille réduite des pixels sur PNEO le positionnement du pointé est moins permissif que sur les images PHR.



FIGURE 2 – Compositions Colorées PIR (1/4000^{ème}).
Haut : PHR, Bas : PNEO.

2.2. Simple comparaison visuelle

En photo-interprétation, le type d'image, les résolutions spatiales et spectrales ainsi que la saisonnalité (temporalité) viennent contraindre ou faciliter la perception des objets de surface, leur extraction ainsi que leur restitution. Viennent s'ajouter aux caractéristiques intrinsèques des capteurs et images, la capacité d'analyse de la morphologie des objets, de leurs dimensions, de leurs couleurs et teintes, de leur texture ou encore de l'ombrage et du relief qui apportent des informations indispensables à l'interprétation et l'extraction des objets (Figure 2). En dernier point, les éléments cités précédemment ont des impacts selon la(les) production(s) cible(s).

L'une des premières conséquences liée à l'usage d'images de résolution plus fine que des PHR est de permettre au photo-interprète de scanner sa zone d'étude de manière « optimisée ». En effet, le photo-interprète analyse et appréhende plus rapidement le paysage et ses éléments. Il se dirige donc plus rapidement vers sa cible cartographique.

En milieu urbain, la délimitation des bâtiments et l'identification du type de toit majoritaire sur la zone d'étude

s'avèrent être facilitées par la finesse de résolution et des détails présents à l'image.

L'interprétation est de fait plus intuitive sur la PNEO. La Figure 3 est un exemple de comparaison visuelle entre les images PHR et PNEO en amont du processus d'extraction vectorielle. Il est ainsi possible de distinguer des niveaux de détail et de netteté différents.

Il est également notable que la couleur de certains éléments anthropiques tend moins vers le rouge dans les images PNEO. Cela s'explique probablement par des méthodes de correction atmosphérique différentes. Les images PHR GEOMAPS utilisent une correction *Top-Of-Atmosphere* tandis que les images PNEO utilisent une correction *Bottom-Of-Atmosphere* (Poilvé et al. (2022)). Cela apporte un rendu radiométrique plus proche de la réalité physique.



FIGURE 3 – CC-PIR (1/1250^{ème}).
Haut : PHR, Bas : PNEO.



FIGURE 4 – CC-PIR (1/500^{ème}). Vecteurs relatifs aux bâtiments en rouge (saisie interne GEOFIT) Gauche : PHR, Droite : PNEO.

2.3. Objets particuliers et structuration du paysage

Le fait d'avoir une résolution plus fine permet un repérage plus rapide des objets particuliers ou des objets structurants du paysage, notamment en milieu urbain. Les Figures 5 et 6, montrent qu'il est plus facile d'identifier les superstructures de toit, les détails visibles tels que les indentations ou encore les différences de niveau et de matériaux des toits.



FIGURE 5 – Détails sur des CC-VC, distinction des indentations (1/500^{ème}). **Haut** : PHR **Bas** : PNEO.



FIGURE 6 – Détails sur des CC-VC — distinction des niveaux et indentations (1/600^{ème}). **Haut** : PHR **Bas** : PNEO.

Les images PNEO apportent également un « confort d'interprétation », notamment pour l'identification des petits objets qui, bien que visibles sur les images PHR, sont plus facilement repérables sur les PNEO. L'extraction est plus facile, l'œil cherche moins la délimitation des contours.

Les Figures 7 à 10 illustrent (i) un exemple d'objets du type antenne ou pylône pour lesquels la structure, la base et l'ombre sont plus visibles ; (ii) des exemples d'objets du type « petits réservoirs d'eau » ou (iii) l'exemple d'objets fins du type canaux urbains ou passage d'accès à des habitations. Ces exemples apparaissent plus nettement et distinctement sur les images PNEO.



FIGURE 7 – Détection d'antenne/Pylône (1/500^{ème}). **Haut** : PHR, **Bas** : PNEO.



FIGURE 8 – Exemples de deux sites industriels (1/500^{ème}). **Gauche** : PHR **Droite** : PNEO.



FIGURE 9 – Détection de petits réservoirs d'eau(1/500^{ème}). **Haut** : PHR, **Bas** : PNEO.



FIGURE 10 – Détection des canaux (1/500^{ème}). **Haut** : PHR, **Bas** : PNEO.

Il est également plus aisé pour le photo-interprète de déterminer les limites et fonctions des sites industriels grâce à la meilleure interprétation des petits objets, les composants (*e.g.* bassins, citernes, cheminées, aérateurs...) et des objets fins (*e.g.* clôtures, convoyeurs, pipelines...), Figure 8.

2.4. Zones urbaines denses, zones rurales

Dans des zones urbaines extrêmement denses, là où il était parfois complexe de faire des délimitations des bâtiments et structures de toit en s'appuyant sur des PHR, la résolution des PNEO 30 cm permet une individualisation qui n'aurait peut-être pas pu être réalisée sur des PHR 50 cm (Figure 11). On comprend que les images PNEO permettent de faciliter l'interprétation grâce à la finesse des détails, mais le mode d'extraction doit être plus minutieux afin de respecter les critères d'extraction de la cible GEOMAPS F5 (qui s'adaptent en partie à la résolution de l'image).



FIGURE 11 – Zoom au 1/500^{ème} sur une PHR et zoom au 1/300^{ème} sur une PNEO. **Haut** : PHR, **Bas** : PNEO.

Le fait d'utiliser des images PNEO permet une meilleure visibilité des détails, mais cela implique aussi une précision plus importante dans le pointé pour s'adapter aux spécifications, car la résolution est plus importante (voir sous-section 2.1). De ce fait, l'opérateur doit adapter son seuil de saisie pour un rendu similaire. Il est en moyenne au 500^{ème} sur des PHR 50 cm et au 300^{ème} sur des PNEO (Figure 12).



FIGURE 12 – Zoom au 1/500^{ème} sur des PHR et zoom au 1/300^{ème} sur des PNEO. **Haut** : PHR, **Bas** : PNEO.

Il en va de même quelle que soit la thématique extraite. Ici pour l'occupation du sol, pour un paysage avec des délimitations précises, l'opérateur est en moyenne au 2000^{ème} sur des Pléiades 50 cm et au 1500^{ème} sur des PNEO, Figure 13.



FIGURE 13 – Zoom au 1/2000^{ème} sur une PHR et zoom au 1/1500^{ème} sur une PNEO. **Gauche** : PHR, **Droite** : PNEO.

2.5. Impact de la parallaxe

La hauteur des bâtiments et l'angle d'acquisition des images influencent fortement l'extraction des vecteurs en milieu urbain. Les Figures 14 et 15 illustrent bien l'effet de la parallaxe (dévers) sur des bâtiments de moyenne et de grande taille. L'identification et le positionnement des bases des bâtiments sont plus précis avec une image PNEO de 30 cm qu'avec une image PHR de 50 cm. Toutefois, malgré cet avantage, il reste indispensable d'appliquer un décalage systématique des vecteurs pour corriger l'effet de parallaxe, car les contours sont extraits à partir des toits.



FIGURE 14 – Parallaxe sur de grands ensembles(1/500^{ème}). **Haut** : PHR, **Bas** : PNEO.



FIGURE 15 – Parallaxe sur des immeubles de taille moyenne (1/500^{ème}). **Haut** : PHR, **Bas** : PNEO.

2.6. Bilan

À cible équivalente, la production de données cartographiques est impactée par le passage à la PNEO 30 cm.

D'un côté, l'interprétation est plus aisée :

- avec une meilleure différenciation et identification des objets ;
- avec moins de doutes ou d'hésitations sur les délimitations et les codages attributaires (par exemple les types de toits) ;
- avec un travail d'interprétation plus reposant pour les yeux des photo-interprètes.

De l'autre côté, l'extraction est plus minutieuse :

- il est nécessaire de respecter les parallaxes et d'ajuster continuellement les emprises à la base,
- un phénomène de mesure et « re-mesure » des indentations induit une augmentation relative du temps de saisie sur certains objets.
- La subdivision des bâtiments (en réponse aux règles d'extraction) est plus importante dans la mesure où la résolution des images permet de mieux interpréter l'individualisation des objets, ainsi que des différents niveaux qui les composent.
- les niveaux de zoom sont plus fins, ce qui apporte un pointé plus précis, mais un temps plus long à la saisie.

Ainsi, il est nécessaire d'établir un équilibre entre gain à l'interprétation et perte de temps à l'extraction (dont l'importance est variable et est fonction des thématiques et des paysages). *Qu'en est-il pour l'évolution des modèles d'apprentissage et des prédictions IA ? Les sections suivantes abordent cette question du changement de résolution dans la production de couches d'appui à la photo-interprétation.*

3. L'IA au service de la production cartographique

3.1. L'application de l'IA à l'imagerie satellite

L'IA, et plus particulièrement l'apprentissage profond, a connu un essor significatif dans le domaine de l'analyse d'images depuis les années 2010. Entraînés sur de larges bases de données annotées, les réseaux de neurones sont capables de détecter, classer et segmenter des zones spécifiques : leurs applications dans le domaine de l'imagerie satellite sont donc nombreuses.

3.2. L'IA comme pré-production cartographique

L'utilisation de l'IA pour une pré-production cartographique a pour but d'identifier et de détourner les objets d'intérêt sur des images satellite. Cette phase préliminaire permet la production de couches d'appui. L'intégration dans une production des objets issus de l'IA peut être directe ou indirecte, ainsi les objets sont des objets de référence pouvant être totalement ou partiellement intégrés dans la production en fonction du contexte. La qualité de la pré-production ainsi que les exigences spécifiques au produit final conditionnent l'orientation d'utilisation des couches d'appui et donc de ce type de pré-production.

Pour réaliser cette pré-production, l'IA est utilisée pour deux tâches spécifiques : la **segmentation sémantique** et la **segmentation d'instance**. La segmentation sémantique a pour but d'associer à chaque pixel d'une image un même label correspondant à une classe. Par exemple, le label "1" est associé à tous les objets de la classe "bâtiment". La segmentation d'instance attribue un label unique à chaque occurrence d'une même classe. En reprenant l'exemple précédent, chaque bâtiment se verrait attribuer un label différent, permettant leur distinction mutuelle ainsi que leur différenciation par rapport aux autres classes (utile dans un environnement urbain dense).

3.3. Framework IA

Le framework IA développé pour l'aide à la production cartographique s'articule autour de trois modules : la création de bases de données d'apprentissage, l'entraînement des modèles IA et l'inférence sur les données à traiter avec le modèle retenu (Figure 16).

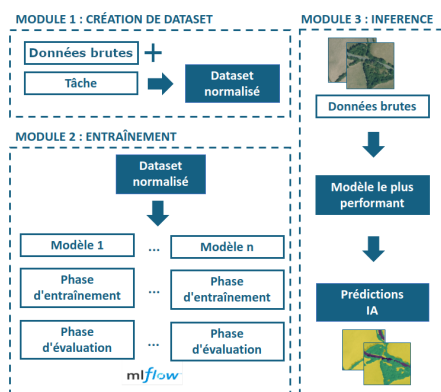


FIGURE 16 – Schéma du framework IA développé - GEOFIT.

Le premier module permet de créer un jeu de données adapté à l'entraînement supervisé d'un modèle d'apprentissage profond, indépendamment de son architecture et de la tâche requise (segmentation sémantique ou segmentation d'instance). Il est également indépendant du format des données d'entrée, pour les images ou leur vérité terrain.

Le second module est un module de suivi et de comparaison des entraînements permettant de sélectionner le meilleur modèle et ses hyper-paramètres optimaux pour la tâche à effectuer. Ce module s'appuie sur **MLflow (Zaharia et al. (2018))**, une plateforme de MLOps mettant en place une base de données des modèles entraînés et de l'ensemble des métriques utilisées pendant les entraînements.

Différentes métriques sont utilisées lors de l'apprentissage d'un modèle. Durant l'entraînement, une **métrique de coût (loss)**, calculée à chaque époque sur les données d'apprentissage, quantifie l'écart entre les prédictions du modèle et la vérité terrain, permettant d'adapter continuellement les poids du modèle (Figure 17).

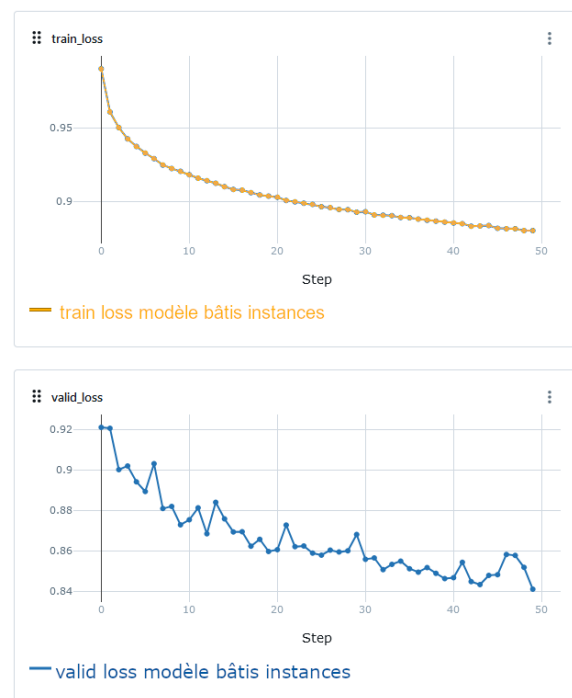


FIGURE 17 – Fonction de coût (loss) sur les données d'entraînement et de validation enregistrée par MLflow.

En phase de test, la métrique *Intersection over Union (IoU)* est calculée pour évaluer les performances du modèle sur de nouvelles données. Elle met en correspondance la zone prédite par le modèle et la zone réelle de l'objet, en calculant le rapport entre la zone d'intersection et la zone d'union de la prédiction et la vérité terrain (Figure 18).

Le dernier module permet d'exploiter le modèle choisi pour réaliser des détections IA de manière simple et robuste. Pour traiter efficacement les images satellites sur la carte graphique des serveurs de calcul, celles-ci sont découpées en tuiles régulières (*i*). Un recouvrement est

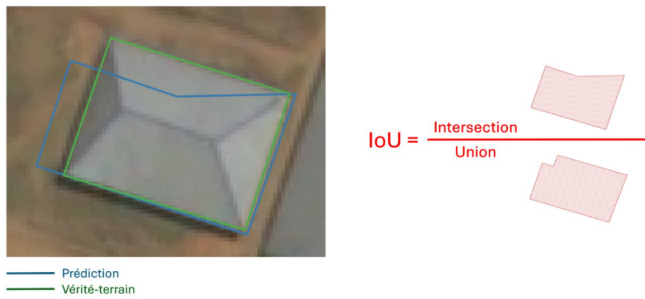


FIGURE 18 – Schéma de la métrique IoU.

utilisé entre les tuiles de manière à maximiser la probabilité qu'un objet se trouve au centre de celle-ci. En effet, un objet centré a plus de contexte et est donc plus facile à prédire pour le modèle.

Pour agréger les différentes prédictions sur un même objet, un système de vote majoritaire, privilégiant les prédictions des tuiles où l'objet est centré, est mis en place (ii). Le résultat est une image de même taille que l'image traitée en entrée avec une valeur de prédiction par pixel (iii) (Figure 19).

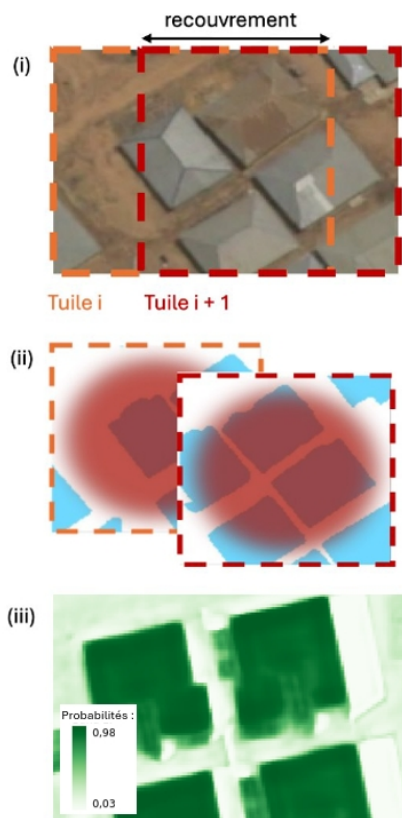


FIGURE 19 – Schéma du système de vote majoritaire.

3.4. Modèles utilisés

La production des couches d'appui fait référence à deux architectures de l'état de l'art : **U-Net** (segmentation sémantique) et **MaskRCNN** (segmentation d'instance). Ces architectures de réseaux de neurones ont toutes les deux fait leurs preuves pour leurs tâches respectives dans divers domaines d'applications (par exemple : Yu et al. (2019), Vuola et al. (2019), Cao et al. (2023), Hou et al. (2021)).

U-Net est un réseau de neurones convolutifs développé par **Ronneberger et al. (2015)** pour la classification et la segmentation sémantique. Son architecture en forme de "U" présente les composantes principales suivantes (Figure 20) :

- L'encodeur, qui a pour but d'extraire les informations utiles de l'image de manière condensée à plusieurs échelles spatiales.
- Le décodeur, qui extrait des données condensées par l'encodeur. Il s'agit des informations les plus pertinentes pour construire une couche de "caractéristiques" (*feature maps*) de résolution équivalente à l'image à traiter. Cette couche sera utilisée par la "tête de segmentation" (*segmentation head*) du réseau pour effectuer la segmentation sémantique.
- Les connexions directes, qui transfèrent les informations de l'encodeur au décodeur. Elles comprennent les détails spatiaux souvent perdus lors des opérations de *pooling* (couche permettant de sous-échantillonner l'image pour réduire sa dimension).

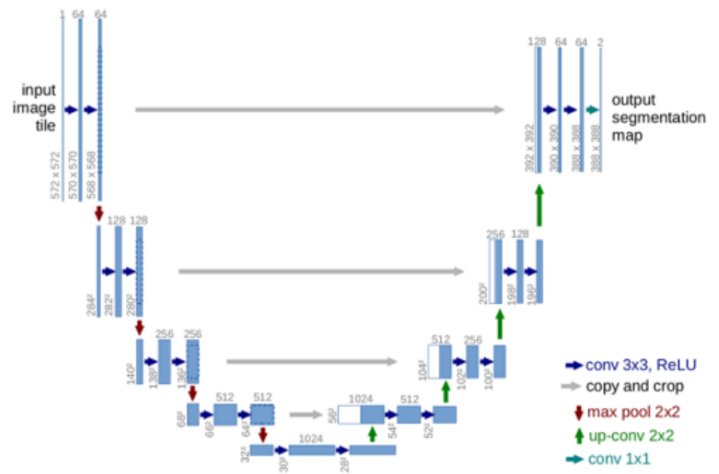


FIGURE 20 – Schéma de l'architecture du réseau U-Net.

MaskRCNN est un modèle de segmentation d'instance proposé par **He et al. (2017)**. Ces composantes principales sont les suivantes (Figure 21) :

- Une *backbone* permettant d'extraire les caractéristiques de l'image pour construire des *feature maps*, jouant un rôle similaire à l'encodeur de U-Net ;
- Les composantes RPN (*Region Proposal Network*) et RoI Align (*Region of Interest Alignment*), dont le but est de proposer plusieurs régions, formes et tailles pouvant contenir des objets d'intérêt à partir des *feature maps* ;
- La composante de détection d'objet, dont le rôle est de prédire, pour chaque RoI, la catégorie de l'objet contenu ainsi que les coordonnées de sa boîte englobante ;
- La composante de détection d'instance, qui ajoute une branche de segmentation générant un masque binaire pour chaque région d'intérêt. Ce masque indique quels pixels, à l'intérieur de la région d'in-

térêt, appartient à l'objet à détecter.

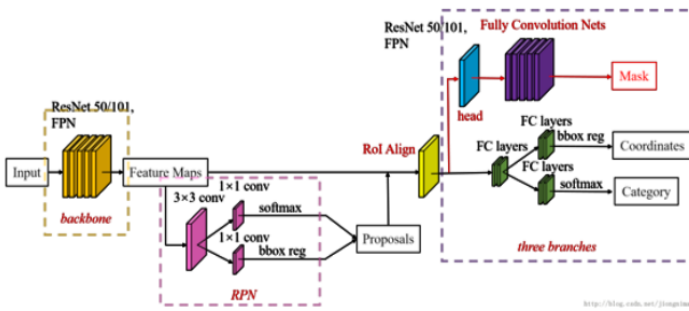


FIGURE 21 – Schéma de l'architecture du réseau MaskRCNN.

3.5. L'apport des Pléiades NEO pour l'IA

L'objectif est d'étudier les améliorations potentielles des pré-productions cartographiques IA liées à l'utilisation d'images PNEO par rapport aux images PHR.

En premier lieu, les résultats d'étude interne montrent qu'il existe des améliorations directes lors de prédictions réalisées avec des modèles historiquement entraînés sur des images PHR sur de nouvelles images PNEO. Des analyses numériques et "à dire d'experts" ont été conduites pour mettre en évidence ce gain de performance sur les classes d'objets suivantes : bâtiments (cf. Section §3.5.2) et réseaux routiers (cf. Section §3.5.3).

Pour aller plus loin, nous présentons quelques résultats d'études internes provenant de l'utilisation de nouveaux modèles utilisant différentes stratégies d'entraînement sur des données hybrides PHR - PNEO (cf. Section §3.6). Ces études se concentrent sur la tâche de segmentation sémantique pour les thématiques du "bâti" et des "routes" sur des zones urbaines.

3.5.1. Données utilisées

Les mêmes modèles IA entraînés sur des données PHR ont été appliqués sur des couples d'images PHR / PNEO. Les données utilisées sont décrites dans le tableau 1.

Paysage	Date PHR	Date PNEO
Résidentiel et industriel	06/02/2024	04/10/2023
Résidentiel	06/10/2023	04/10/2023
Résidentiel et industriel	06/10/2023	04/10/2023
Désertique	10/01/2024	19/01/2022

TABLEAU 1 – Dates des images PHR et NEO en fonction des thématiques d'étude.

Les trois premiers échantillons sont issus d'images d'une capitale dans la corne de l'Afrique avec différents niveaux de densité urbaine. Leurs caractéristiques sont les suivantes : **zone 1** : 6.8 km² pour 15500 bâtiments et 900 routes, **zone 2** : 0.16km² pour 2000 bâtiments et 34 routes, **zone 3** : 1.5 km² pour 8000 bâtiments et 310

routes. Le dernier échantillon (**zone 4**) est une zone désertique de 1 km² dans l'ouest de l'Afrique avec des objets épars et peu nombreux (moins de 10 routes et structures) (Figure 22). Afin d'évaluer la qualité des prédictions IA, une vérité terrain vecteur issue d'une production manuelle GEOFIT (réalisée pour cette communication) a été utilisée. L'étude porte sur les thématiques "bâti" et "réseau routier".



FIGURE 22 – Extrait des 4 zones Pléiades 50 cm (PHR).

3.5.2. Thématique bâti

Pour la thématique "bâti", une analyse automatique des prédictions IA sur images PHR / PNEO a été mise en place. Il s'agit d'un calcul d'IoU entre les objets bâtis issus de la vérité terrain et ceux prédits par les modèles IA. Le résultat de ce calcul est une couche vecteur contenant le score IoU pour chaque bâti de la vérité terrain. Par exemple, si un bâti a été omis par le modèle, le score IoU sera nul. Comme la figure 23 l'illustre sur les prédictions IA sur les petits bâtiments, le nombre d'omissions sur l'image PNEO est inférieur à ce qui peut être relevé sur l'image PHR. Cette analyse automatique montre que les images PNEO permettent d'obtenir systématiquement de meilleurs résultats pour la thématique "bâti".

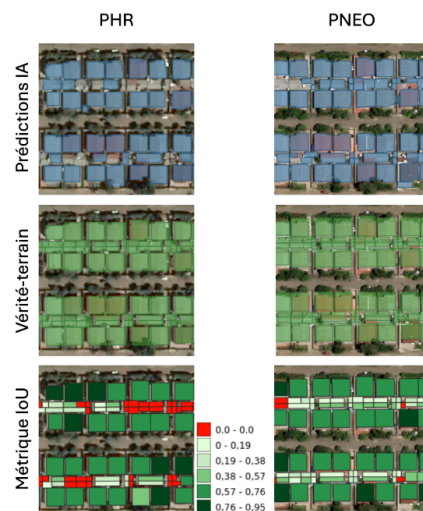


FIGURE 23 – Exemple de comparaison automatique avec la métrique IoU sur un échantillon de la zone 3

Le tableau 2 reporte les gains sur le score d'IoU moyen par zone et sur le taux d'omissions des prédictions PNEO par rapport aux prédictions PHR. La zone 4 n'a pas été traitée automatiquement, car les vecteurs de vérité terrain sont des ponctuels et non des polygones.

Zone	Mean IoU (%)	Omission (%)
Zone 1	+13	-20
Zone 2	+12	-28
Zone 3	+8	-34

TABEAU 2 – Gains sur l'IoU moyenne et sur le taux d'omissions des prédictions sur images PNEO par rapport à PHR.

Pour compléter cette évaluation automatique, une analyse "à dire d'experts" a été réalisée pour mettre en évidence les différences entre les prédictions issues d'images PHR et PNEO. La figure 24 montre que l'utilisation d'images PNEO permet une meilleure individualisation des bâtiments dans un contexte urbain dense (cf. Figure 24, colonne (1)) ainsi que de limiter les omissions cf. Figure 24, colonne (2)).

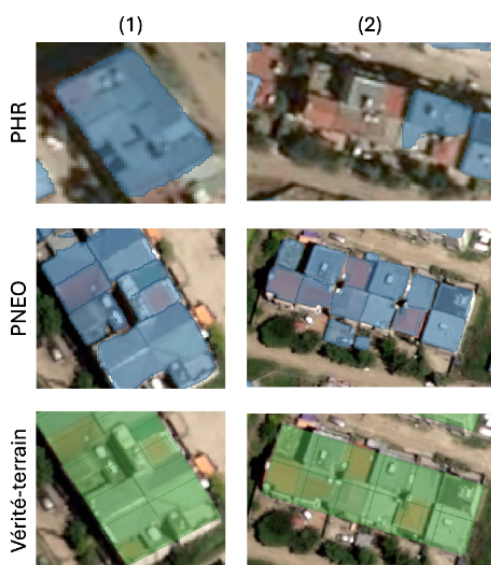


FIGURE 24 – Exemples de prédictions de bâtis sur images PHR et PNEO.

Cependant, de fausses détections sont relevées sur les camions et les rochers (Figure 25).



FIGURE 25 – Exemples de fausses détections bâtis sur PNEO.

Une de nos hypothèses, expliquant cette confusion, est relative au fait que les modèles utilisés pour faire les prédictions sur les images PNEO ont été entraînés uniquement sur des images PHR.

Ainsi, des camions ou des rochers, à une résolution de 30 cm, apparaissent similaires à des bâtiments à une résolution de 50 cm pour le modèle.

3.5.3. Thématique réseau routier

Pour la thématique réseau routier, seule une analyse "à dire d'experts" a été menée pour qualifier les prédictions IA issues des images PHR et PNEO. La Figure 26 montre que dans un contexte urbain, avec un quadrillage régulier, les prédictions du réseau routier sur les deux types d'images sont similaires.



FIGURE 26 – Exemple de prédictions du réseau routier (bleu) sur PHR / PNEO avec vérité terrain (rouge).

Cependant, les images PNEO permettent d'identifier des chemins carrossables non-visibles sur des images PHR, notamment dans des zones moins densément peuplées (Figure 27). C'est tout particulièrement le cas pour la zone 4 de l'expérience concernant une zone désertique où les chemins se confondent dans la végétation et le paysage.

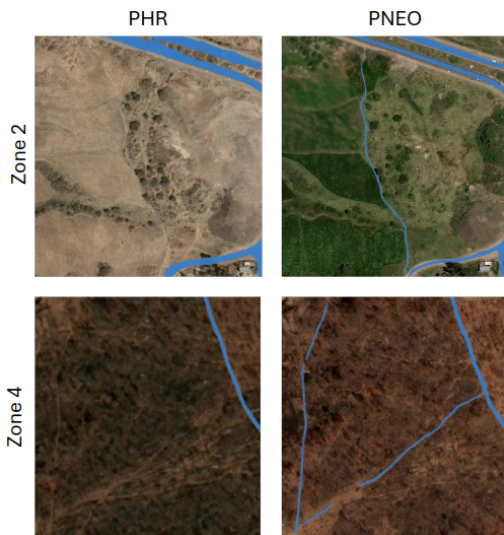


FIGURE 27 – Identification des chemins carrossables sur les images PNEO.

3.6. Entraînements sur des données PNEO

L'objectif est d'explorer les modes d'adaptation des modèles IA entraînés sur des images PHR afin qu'ils puissent être performants sur des images PNEO. Les défis identifiés proviennent des différences entre les images PHR et PNEO qui sont listés ci-après :

- Différence de **résolution spatiale** : Les images PHR ont une résolution de 50 cm, contre 30 cm pour les images PNEO. La résolution plus élevée des images PNEO révèle des détails supplémentaires qui ne sont pas visibles sur les images PHR. De plus, la différence de résolution provoque une différence de taille apparente des objets. Par exemple, un camion sur une image PNEO aura une taille apparente similaire à celle d'un bâtiment sur une image PHR.
- Différence de **colorimétrie** : les bandes spectrales et les processus de pré-traitement peuvent différer entre les deux systèmes.
- Différence de **qualité** : les données PNEO présentent une meilleure qualité en termes de rapport signal/bruit et donc moins d'artefacts que les données PHR.

Ces différences entre images PHR et PNEO peuvent affecter les performances d'un modèle IA entraîné exclusivement sur des données PHR. Quatre stratégies d'adaptation sont explorées :

- La première stratégie (**stratégie 1**) consiste en une **utilisation directe** (*zero-shot learning*) des modèles PHR sur les images PNEO. L'idée est d'utiliser la capacité de généralisation de ces modèles : si les caractéristiques apprises durant leurs entraînements sont suffisamment générales et robustes, ils seront capables de généraliser leur apprentissage à des images PNEO. Il s'agit de la stratégie utilisée pour les comparaisons exposées dans

la section 3.5.2.

- La deuxième stratégie envisagée (**stratégie 2**) est aussi une **utilisation directe** des modèles PHR, mais en effectuant au préalable un **pré-traitement** sur les images PNEO pour réduire leur résolution spatiale à celle des images PHR. Ainsi, les objets présents sur les images PNEO sous-échantillonnées auront des caractéristiques et tailles similaires aux objets présents sur les images PHR utilisées lors de l'entraînement des modèles.

- La troisième stratégie (**stratégie 3**) utilise un **ré-entraînement** des modèles PHR sur des images PNEO en utilisant une approche d'**apprentissage par transfert** (Yosinski et al. (2014)). Cette technique consiste à utiliser un modèle pré-entraîné (ici, les modèles PHR) comme base pour l'entraînement d'un nouveau modèle sur de nouvelles données (ici, les images PNEO). L'avantage de cette méthode est de faire bénéficier au nouveau modèle des connaissances acquises par le modèle précédent. Cela permet de réduire le temps et les coûts d'entraînement, car moins d'images PNEO annotées seront nécessaires pour que le nouveau modèle converge.

- Enfin, la dernière stratégie (**stratégie 4**) envisagée consiste en un **ré-entraînement complet** de nouveaux modèles sur une base d'apprentissage hybride composée d'images PHR (38 000 images) et PNEO (20 000 images).

3.7. Résultats obtenus pour les différentes stratégies

Afin de comparer les performances des modèles issues des différentes stratégies sur les quatre zones de l'étude, les métriques **Intersection-over-Union (IoU)**, **rapport (recall)** et **précision (precision)** ont été calculées sur chaque prédiction.

- La métrique IoU est définie dans la section §3.3.
- La précision mesure la proportion des prédictions positives correctes par rapport à l'ensemble des prédictions du modèle.
- Le rappel mesure la proportion de vrais positifs qui ont été correctement identifiés par le modèle par rapport au nombre d'éléments positifs dans la vérité terrain.

L'ensemble de ces métriques est calculé sur des images PNEO issues des trois zones urbaines décrites précédemment au regard d'une vérité terrain issue de productions cartographiques par photo-interprétation.

Le tableau 3 présente ces métriques moyennées sur l'ensemble des zones pour chacune des stratégies pour la segmentation des bâtiments. Le tableau 4 contient les mêmes métriques calculées sur les routes. La Figure 28 illustre les résultats pour chacune des stratégies pour les PHR et les PNEO.



FIGURE 28 – Comparaison des prédictions des modèles en fonction de la stratégie 1 à 4. À **gauche**, prédictions sur les bâtiments de la zone 1. À **droite**, sur les routes de la zone 2.

Stratégie	mIoU	mRec.	mPrec.
Stratégie 1	0.730	0.785	0.912
Stratégie 2	0.736	0.797	0.907
Stratégie 3	0.843	0.917	0.912
Stratégie 4	0.856	0.926	0.919

TABLEAU 3 – Bâtiments : Métriques moyennées pour toutes les zones par stratégies

Stratégie	mIoU	mRec.	mPrec.
Stratégie 1	0.491	0.516	0.912
Stratégie 2	0.493	0.521	0.912
Stratégie 3	0.316	0.333	0.854
Stratégie 4	0.514	0.537	0.921

TABLEAU 4 – Réseau Routier : Métriques moyennées pour toutes les zones par stratégies

- La **stratégie 1** constitue notre modèle "baseline" à comparer aux nouveaux entraînements utilisant les données PNEO. Il correspond à un modèle similaire à celui employé dans les comparaisons précédentes ayant appris uniquement sur PHR (pour plus de détails, se référer aux analyses numériques et "à dire d'experts" en 3.5.2 et 3.5.3). Cette stratégie semble directement fournir des prédictions "acceptables" pour les bâtiments et "moyennes" pour les routes.

- Le sous-échantillonnage (**stratégie 2**) semble produire des résultats avec des bâtiments plus complets (cf. Figure 28) ce qui améliore le rappel. Cependant, il existe encore des confusions importantes avec des objets plus petits, comme décrit dans la section précédente. Il semble même que la perte de résolution diminue légèrement la précision globale du modèle pour les bâtiments.

- L'apprentissage par transfert (**stratégie 3**) améliore nettement les prédictions pour les modèles spécialisés sur le bâti, mais dégrade les performances pour les routes.

De nombreuses routes de petite et moyenne taille semblent manquer, ce qui diminue fortement le rappel. Une hypothèse pourrait être que les caractéristiques des routes dans les deux images sont trop éloignées du fait d'une prise de vue à une période différente (voir tableau 1).

- L'apprentissage avec une base de données contenant des données des deux types (**stratégie 4**) amène des résultats nettement meilleurs sur l'ensemble des trois métriques évaluées. C'est particulièrement le cas pour les routes qui sont bien mieux détectées et où de plus petits chemins sont moins omis. Pour le bâti, les segmentations sont un peu mieux délimitées si on les compare à celles obtenues avec les modèles par transfert, ce qui améliore légèrement la précision.

Pour conclure, **la stratégie 4 semble être la plus performante**, elle produit les résultats les plus précis surtout pour la segmentation des bâtiments. **L'apprentissage avec un jeu de données hybrides** exploite au mieux les données issues du capteur PNEO. Les images des deux sources sont suffisamment proches pour permettre l'apprentissage de modèles de segmentation efficaces sur des images de plus haute résolution. La quantité de données PHR associées à des images de meilleure qualité et résolution permet d'obtenir le meilleur modèle de cette étude.

3.8. Limites et perspectives

Les résultats montrent un très bon transfert des modèles entraînés sur des données PHR lors des prédictions sur des données PNEO. Les caractéristiques des images provenant des deux sources semblent également suffisamment proches pour être utilisées lors d'un apprentissage avec des données mixtes. Cependant, la quantité de données PNEO étant moins élevée, il est très probable que le modèle ainsi conçu favorise légèrement les prédictions sur les images PHR.

Une analyse complémentaire pourrait mettre en évidence ces disparités, il s'agirait de comparer, de nouveau, les deux types d'images avec un nouveau modèle exclusivement entraîné sur des images PNEO (avec une taille du jeu de données similaire).

De plus, la résolution spatiale de PNEO, plus fine, permet d'envisager des modèles de plus en plus performants en exploitant au mieux les caractéristiques multi-échelles dans les images. Ces modèles permettront de mieux répondre aux enjeux de la production cartographique, tout particulièrement par rapport à la grande diversité de domaines et de spécifications.

Plusieurs pistes peuvent être envisagées comme l'apprentissage par transfert ou la création d'un "modèle fondation" (ex : **Oquab et al. (2024)**). Voir la définition dans le rapport "AI Foundation Models" : "Initial Report" de **CMA (2022)** dédié à la cartographie entraînée sur une grande quantité de données diverses avec des méthodes d'auto-supervision (**He et al. (2022)**, **Caron et al. (2021)**). Il pourrait ensuite être adapté à des tâches spécifiques en fonction des besoins de production. Une autre possibilité serait des modèles multi-modaux utilisant différentes sources de données ou permettant une plus grande adaptabilité à de nouveaux domaines ou nouvelles tâches (**Kirillov et al. (2023)**).

4. Conclusion

Le passage des images Pléiades à 50 cm de résolution vers des images PNEO à 30 cm de résolution dans notre processus de production présente des avantages et des inconvénients que nous avons pu aborder à travers un focus sur la photo-interprétation et le regard du photo-interprète.

L'interprétation des images est plus simple, avec une meilleure identification des objets et une réduction des incertitudes. Cependant, l'extraction des données devient plus complexe et demande plus de temps, en raison de la nécessité de respecter les détails et de réajuster les mesures.

L'enjeu est de trouver un équilibre entre ces aspects, tout en abordant l'impact de cette évolution sur les modèles d'apprentissage et les prédictions en IA qui viennent alimenter nos productions. Les études et évaluations menées nous ont permis de mieux cerner les gains de performance et l'impact de l'utilisation de résolutions plus fines sur notre framework IA.

- **Dans un premier temps**, la comparaison directe de paires d'images PHR/PNEO a été effectuée en utilisant un modèle préexistant. En analysant la précision de ce modèle au regard d'une production cartographique manuelle, un gain de performance significatif a été observé sur les images PNEO (+11% pour l'IoU moyenne et -27% d'omissions).

- **Dans un second temps**, une évaluation de nouveaux modèles entraînés à l'aide des images PNEO a été réalisée. Son objectif était de déterminer si un apprentissage conjoint avec des données historiques PHR

permettrait un gain de performance. Trois stratégies ont été comparées au modèle "baseline" historique : sous-échantillonnage, apprentissage par transfert et ré-entraînement hybride. C'est cette dernière approche qui a produit le modèle le plus précis sur des données PNEO.

En tant que producteur de données vecteur, GEOFIT a adapté ses chaînes de production pour répondre à l'opportunité d'avoir des images satellite de résolution plus fine sur ses chantiers.

Une adaptation et de nouveaux procédés de saisie ont permis aux photo-interprètes de gagner en expertise tout en étant toujours conforme à la cible du standard vecteur produit.

Il s'agit d'une évolution majeure qui pourrait encore apporter de nouvelles opportunités tant pour la photo-interprétation que pour le(s) Framework(s) IA de GEOFIT notamment (i) avec les nouveaux canaux (Deep blue et Red edge) qui ne sont pas encore exploités, mais qui pourraient avoir de sérieux avantages notamment en gain de production dans certaines zones géographiques, mais également (ii) avec l'implémentation et la collecte itérative d'images PNEO pour les bases d'apprentissage qui permettront de rendre des prédictions de plus en plus précises et exploitables dans nos productions.

Remerciements

L'équipe GEOFIT remercie Airbus Defence and Space pour la fourniture des images qui ont permis de réaliser cet article. Crédits :

- Images Pléiades (PHR) : Copyright © CNES, 2023-2024, Distribution Airbus DS.
- Images Pléiades Neo (PNEO) : Copyright © Airbus DS, 2022-2024, Distribution Airbus DS.

Références

- Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., and Wang, M. (2023). Swin-unet : Unet-like pure transformer for medical image segmentation. In Karlinsky, L., Michaeli, T., and Nishino, K., editors, *Computer Vision – ECCV 2022 Workshops*, pages 205–218, Cham. Springer Nature Switzerland.
- Caron, M., Touvron, H., Misra, I., Jegou, H., Mairal, J., Bojanowski, P., and Joulin, A. (2021). Emerging properties in self-supervised vision transformers. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9630–9640.
- CMA (2022). Ai foundation models : Initial report. Competition and Markets Authority (UK). URL https://assets.publishing.service.gov.uk/media/65081d3aa41cc300145612c0/Full_report_.pdf.
- Defence Geospatial Information Working Group (DGIWG 262) (2021). Multinational geospatial co-production program urban vector data (muvd) metadata specification. Technical report. URL <https://portal.dgiwg.org/files/72193>.
- He, K., Chen, X., Xie, S., Li, Y., Dollár, P., and Girshick, R. (2022). Masked autoencoders are scalable vision learners. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15979–15988.

- He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In *IEEE International Conference on Computer Vision (ICCV)*.
- Hou, Y., Liu, Z., Zhang, T., and Li, Y. (2021). C-unet : Complement unet for remote sensing road extraction. *Sensors*, 21(6).
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., and Girshick, R. (2023). Segment anything. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3992–4003.
- Oquab, M., Darcet, T., Moutakanni, T., Vo, H. V., Szafraniec, M., Khalidov, V., Fernandez, P., HAZIZA, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.-Y., Li, S.-W., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jegou, H., Mairal, J., Labatut, P., Joulin, A., and Bojanowski, P. (2024). DINOv2 : Learning robust visual features without supervision. *Transactions on Machine Learning Research*.
- Poivlé, H., Bonfort, T., Varoquaux, V., and Elin, C. (2022). Model-based dehazing for optimised image display and analysis ready data. In *VHR-RODA 2022, ESRIN*.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net : Convolutional networks for biomedical image segmentation. In Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham. Springer International Publishing.
- Vuola, A. O., Akram, S. U., and Kannala, J. (2019). Mask-rcnn and u-net ensembled for nuclei segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 208–212.
- Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*.
- Yu, Y., Zhang, K., Yang, L., and Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on mask-rcnn. *Computers and Electronics in Agriculture*, 163 :104846.
- Zaharia, M., Chen, A., Davidson, A., Ghodsi, A., Hong, S. A., Konwinski, A., Murching, S., Nykodym, T., Ogilvie, P., Parkhe, M., et al. (2018). Accelerating the machine learning lifecycle with mlflow. *IEEE Data Eng. Bull.*, 41(4) :39–45.