Étude sur les outils pour améliorer la production de données géospatiales

Fabien Amarger[1], Anaïs Cipel[1], Nathalie Noblecourt[1], Jehanne Portefaix[1], Pierre-Marie Brunet[2]

[1] Digitanie
Cour Guillaut 09700 Saverdun
prénom.nom@digitanie.org
[2] CNES
Centre spatial de Toulouse
18 avenue Edouard Belin 31401 Toulouse
prénom.nom@cnes.fr

Résumé

L'observation de la Terre par imagerie satellite s'impose comme un levier majeur pour répondre aux enjeux climatiques, socio-économiques et d'aménagement du territoire. La chaîne de valeur associée repose, en grande partie, sur la disponibilité de données d'entraînement fiables pour des modèles d'IA (détection, classification, segmentation), lesquelles sont issues d'une photo-interprétation exigeante : vectorisation des objets, puis labellisation des entités. Depuis 2019, des collaborations successives entre le CNES et Digitanie ont permis d'industrialiser la production de ces données géospatiales et de mettre à l'épreuve des modèles sur des projets emblématiques du CNES[1]. Ces projets ont démontré la pertinence de l'annotation manuelle pour produire la base d'apprentissage, mais l'évolution des technologies doit permettre une optimisation de ces processus de production. La consolidation d'un outillage collaboratif d'annotation, couplé à des pré-annotations issues de modèles semi-automatisés, constitue un enjeu décisif : accélérer la production tout en préservant la qualité et l'homogénéité des données produites. Nous proposons, dans cet article, les résultats d'une étude sur les outils collaboratif d'annotations spécifiques aux données géospatiales pour :

- *accroître la productivité sans dégrader la qualité ;*
- *assurer un suivi de production transparent et auditable ;*
- *intégrer des modèles d'annotation automatisée lorsque cela est pertinent ;*
- *s'insérer dans l'écosystème SIG et documentaire existant (projections, formats, données contextuelles ainsi que conformité géométrique et topographique).*

Mots-clés

Observation de la terre, base d'apprentissage, annotations, images satellites, données géospatiales.

1 Introduction

L'observation de la Terre par imagerie satellite s'impose comme un levier majeur pour répondre aux enjeux climatiques, socio-économiques et d'aménagement du territoire. La chaîne de valeur associée repose, en grande partie, sur la disponibilité de données géospatiales d'entraînement fiables pour des modèles d'IA (détection, classification, segmentation), lesquelles sont issues d'une photo-interprétation exigeante : vectorisation des objets, puis labellisation des entités. Depuis 2019, des collaborations successives entre le CNES et Digitanie ont permis d'industrialiser la production de ces données géospatiales et de mettre à l'épreuve des modèles sur des projets emblématiques du CNES. Ces projets ont démontré la pertinence de l'annotation manuelle pour produire la base d'apprentissage, tout en mettant en lumière des axes d'amélioration : optimisation des coûts de production, harmonisation entre opérateurs, renforcement du pilotage et de la traçabilité des annotations, ainsi que fluidification des échanges avec les clients.

Dans ce contexte, la consolidation d'un outillage collaboratif d'annotation, couplé à des pré-annotations issues de modèles de vision, constitue un enjeu décisif : accélérer la production tout en préservant (ou améliorant) la qualité et l'homogénéité des données produites.

Au regard de la multitude des outils existants, nous avons réalisé une étude exhaustive des outils pouvant répondre à ce besoin. Nous présentons, dans cet article, les résultats de cette étude. Pour cela, nous aborderons tout d'abord la méthode de sélection des outils de l'étude, ensuite, nous détaillerons le protocole expérimental mis en place. Nous analyserons les résultats obtenus avant de conclure et d'apporter des perspectives.

2 Sélection des outils

L'état de l'art a été construit selon une démarche visant à recenser, analyser et comparer les outils d'annotation d'images géospatiales haute résolution (HR) ou très haute

résolution (THR), c'est-à-dire une résolution de 30 à 50 cm. La recherche s'est appuyée sur plusieurs sources complémentaires : les documentations officielles des outils (sites éditeurs, dépôts GitHub), les articles scientifiques et publications techniques décrivant des approches ou modèles d'annotation et de segmentation, ainsi que des rapports institutionnels (CNES, ESA, SpaceNet, etc.) et de la littérature grise (blogs spécialisés, retours d'expérience industriels). Les solutions identifiées ont ensuite été classées en différentes catégories (plateformes commerciales, logiciels open source, modèles IA spécialisés, outils internes), afin de couvrir à la fois les solutions clés en main et les modèles d'automatisation réutilisables dans un processus de production. Cette approche garantit que l'étude repose sur une base documentée, exhaustive et actualisée, permettant de comparer chaque outil selon des critères objectifs et reproductibles.
Dans le cadre de l'état de l'art, un panel de 26 outils (c.f. tableau 1 d'annotation a été recensé afin d'obtenir une vision complète du paysage technologique actuel. Si certains d'entre eux ont été sélectionnés pour l'expérimentation comparative, d'autres n'ont pas été retenus. Leur examen reste néanmoins important, car il illustre la diversité des approches disponibles et justifie les choix opérés.

- **C1 :** Solution adaptée aux données géospatiales THR
- **C2 :** Solution qui offre une gestion des projections (au minimum les références mondiales)
- **C3 :** Solution qui propose une interopérabilité avec des environnement SIG
- **C4 :** Solution qui permet l'intégration de données contextuelles
- **C5 :** Solution qui permet une gestion collaboration de la production
- **C6 :** Solution permettant une traçabilité
- **C7 :** Contraintes de déploiement et de licence

Chaque critère a été évalué selon la grille suivante :

- **C :** conforme
- **P :** partiel (possible avec adaptation / connecteur / prétraitement / en option)
- **NC :** non conforme
- **ND :** non déterminé

Une première liste de critères de sélection des outils a été définie. Ces critères sont opérationnalisés par des indicateurs mesurables grâce à des échelles de notations dédiées. Nous définissons tout d'abord le critère **Qualité** pour mesurer la robustesse et la fiabilité globale de l'outil : précision des annotations, stabilité et maturité logicielle. Le deuxième critère **Interopérabilité** permet d'évaluer la capacité d'intégration dans les chaînes de production existantes : compatibilité SIG (QGIS, ArcGIS), formats supportés (GeoTIFF, Shapefile, GeoJSON), disponibilité d'API. Ensuite le critère **Annotation assistée par IA** mesure la présence de fonctionnalités de pré-annotation ou d'annotation assistée par IA (outils de numérisation avancée, baguette magique, interdiction des chevauchements...). Nous

Outil	C1	C2	C3	C4	C5	C6	C7	Retenu ?
BRIMA	NC	ND	ND	ND	ND	ND	ND	Non
CVAT + SAM	P	P	P	NC	P	ND	C	Oui
Dataloop	ND	ND	ND	ND	ND	ND	P	Non
FlyPix AI	NC	NC	ND	ND	ND	ND	ND	Non
Geo4i	C	C	C	ND	ND	C	ND	Oui
GeoAI (Esri)	C	C	C	C	P	P	P	Oui
IMERIT Annotator	ND	ND	ND	ND	P	ND	NC	Non
KalDA	ND	ND	ND	ND	ND	ND	ND	Non
Kili Technology	P	ND	P	ND	C	P	P	Oui
Label Studio	P	NC	P	NC	P	ND	C	Oui
Labelbox	P	ND	P	ND	C	ND	P	Oui
LabelImg	NC	NC	NC	NC	NC	NC	C	Non
LabelMe	NC	NC	NC	NC	NC	NC	C	Non
Label-Studio	NC	NC	P	NC	C	P	C	Non
Makesense	NC	NC	NC	NC	NC	NC	P	Non
Mapflow.ai	P	P	P	NC	NC	NC	P	Oui
Pixano	NC	NC	NC	NC	ND	ND	P	Non
RectLabel	NC	NC	NC	NC	NC	NC	P	Non
Roboflow	NC	NC	ND	ND	ND	ND	ND	Non
Snail	P	NC	NC	NC	NC	NC	C	Non
SuperAnnotate	P	NC	P	ND	ND	ND	P	Non
Supervisely	P	ND	P	ND	C	C	P	Oui
TagLab	NC	NC	ND	ND	ND	ND	ND	Non
VIA	NC	NC	NC	NC	NC	NC	P	Non
VoTT	NC	NC	NC	NC	NC	NC	P	Non

TABLE 1 – Grille d'évaluation des outils et modèles selon les critères définis

analysons ensuite le critère **Ergonomie** qui évalue la facilité d'utilisation, la fluidité de l'interface, l'expérience utilisateur, et la gestion collaborative. Et enfin le critère **Coût** qui permet d'analyser le coût d'acquisition, de déploiement et de maintenance, ainsi que le modèle économique (SaaS, licence, open source).
Sur les 26 outils étudiés sur cette base, nous sommes arrivés à une sélection de 7 outils éligibles pour une étude plus approfondie. Le comparatif des critères pour ces outils éligibles est présenté avec le tableau 2.

3 Protocole expérimental

L'expérimentation est une phase centrale de l'étude : elle vise à confronter, dans des conditions contrôlées et reproductibles, les outils identifiés lors de l'état de l'art. L'objectif est d'aller au-delà de l'analyse théorique pour mesurer de manière scientifique et objective leurs performances réelles sur des jeux de données représentatifs. Le protocole a été conçu de façon à répondre à trois enjeux principaux : la productivité, la qualité ainsi que l'ergonomie et la gestion de projet. Afin de garantir la validité scientifique de l'étude, la méthodologie repose sur une sélection d'échantillons représentatifs (différents types de territoires et classes d'objets), puis une comparaison systématique des résultats obtenus selon un mode de production entièrement manuel (baseline) et ceux obtenus selon un mode de production hybride (semi-automatisation/automatisation +

Outil	Qualité	Interopérabilité	Annotation assistée par IA	Ergonomie	Coût
Kili Technology	Excellent	Très bonne	Oui	Très bonne	Coûteux
CVAT+SAM	Très bonne	Bonne	Oui	Moyenne	Variable selon le type de licence
Geo4i	Très bonne	Excellente	Oui	Bonne	Coûteux
Supervisely	Bonne	Bonne	Partielle	Bonne	Faible
Labelbox	Excellente	Bonne	Partielle	Bonne	Coûteuse
Mapflow.ai	Bonne	Bonne	Oui	Bonne	Faible
GeoAI (Esri)	Excellente	Excellente	Oui	Très bonne	Très coûteux

TABLE 2 – Comparatif des 7 outils sélectionnés

post-édition). Les étapes du protocole (standardisation, annotation, contrôle qualité, mesures de robustesse) assurent la traçabilité et la reproductibilité des résultats. En somme, cette approche expérimentale va permettre d'objectiver les bénéfices et les limites des différentes solutions, et d'identifier celles qui présentent le meilleur potentiel de déploiement dans un flux de production.

L'étude s'appuie sur les jeux de données fournis par le CNES et produits lors des projets antérieurs (AI4GEO)[2]. Ces jeux incluent des images satellites de très haute résolution (THR – Pleiades NEO) accompagnées des couches vectorielles issues de la photo-interprétation entièrement manuelle réalisée par Digitanie avec QGIS (baseline), ainsi que des données exogènes contextuelles pour aider à la photo-interprétation (OSM, Google Street View, ...). Pour s'assurer d'une évaluation représentative, l'échantillon d'expérimentation doit couvrir plusieurs contextes territoriaux : zones urbaines denses, péri-urbaines, rurales. Les unités expérimentales sont des tuiles d'images géoréférencées d'une surface nominale de 50 000 m², sélectionnées de manière à être homogènes au regard du contexte territorial. Chaque scène sera accompagnée d'un référentiel de vérité terrain issu des vecteurs et métriques produites lors de la photo-interprétation manuelle.

Afin d'évaluer de manière fiable et comparable les performances des différents outils testés, il est essentiel de définir un protocole d'annotation clair et standardisé. Celui-ci repose sur deux approches complémentaires : une baseline manuelle, qui sert de référence, et une procédure hybride intégrant l'annotation assistée via l'outil, suivie d'une post-édition humaine. Cette organisation permet à la fois de mesurer le gain potentiel en productivité et en qualité lié à l'automatisation, tout en conservant un cadre de comparaison robuste et reproductible. Le protocole inclut également une étape préalable de standardisation afin d'assurer la cohérence des annotations entre outils et annotateurs. Enfin, un contrôle qualité systématique, basé sur une double relecture par un expert du domaine, garantit la fiabilité des résultats et permet de corriger les divergences lorsque nécessaire. Les éléments suivants sont figés avant expérimentation :

- schéma de classes (nomenclature) ;
- règles de vectorisation et tolérances (unité minimale de collecte, gestion des angles, simplification, règles topologiques) ;
- formats d'échange et conventions de nommage (exports, versions, identifiants de tâche) ;

3.1 Les procédures de production

3.1.1 Procédure baseline (annotation manuelle)

La baseline est l'annotation complète réalisée manuellement, à partir de l'image et des couches contextuelles, selon les règles de standardisation. Pour chaque scénario (territoire × classes), un ensemble de tuiles est annoté manuellement. Le temps est mesuré par phase (annotation, correction des erreurs topologiques et géométriques) et un relecteur effectue une revue afin de limiter la variabilité inter-annotateur. La production baseline correspond au flux actuellement utilisé dans les projets CNES/Digitanie, reposant sur la vectorisation manuelle réalisée dans QGIS à partir des images satellites et des couches contextuelles.

3.1.2 Procédure hybride (pré-annotation + post-édition)

La procédure hybride ajoute une étape de pré-annotation automatique (modèle intégré à l'outil ou pipeline externe), suivie d'une post-édition humaine visant à atteindre un niveau de qualité satisfaisant. Les paramètres de pré-annotation (seuils, classes, post-traitements) sont documentés et versionnés. Le temps de post-édition est mesuré séparément afin d'estimer le gain net.

3.2 Indicateurs et métriques

Comme défini dans l'introduction de cette section, nous souhaitons comparer les outils suivant 3 axes : la qualité, la productivité et l'ergonomie/gestion de projet. Pour permettre la comparaison de ces trois axes, nous définissons, pour chacun, la meilleure manière de les évaluer.

Qualité : Pour déterminer la qualité de la production, nous comparons chaque production avec une baseline. Cette baseline, produite manuellement, permettra de servir de référentiel. La méthodologie de production manuelle appliquée à Digitanie a permis de démontrer une très grande qualité dans ses réalisations. Un des objectifs est d'essayer de ne pas dégrader cette qualité, c'est pour cela que nous comparons avec cette base. La qualité est analysée suivant le calcul de la précision, du rappel, mais aussi de la F-mesure pour chaque cas étudié. Nous détaillerons la méthode de calcul de ces métriques dans une prochaine section.

Productivité : Afin de déterminer s'il y a un gain de productivité, nous comptabiliserons le temps nécessaire à la réalisation de la même production. Nous mettrons en perspective ces temps avec les temps nécessaires concernant la baseline. Nous considérons le temps total comme la somme

du temps de production en utilisant l'outil (avec une aide plus ou moins efficace de l'outil) avec le temps de correction des erreurs de l'outil ainsi que le temps de contrôle. Nous pourrons alors comparer le temps total entre la production de la baseline et le temps total nécessaire pour toute la production.

Ergonomie et gestion de projet : Enfin, pour évaluer l'ergonomie et la gestion de projet, nous définissons des questionnaires à remplir pour chaque outil. Ces questionnaires sont remplis par chaque opérateur réalisant la production. Une partie de ces critères ont déjà été évalués lors de la phase de l'état de l'art, mais ils sont reconsidérés ici au regard de l'expérimentation. Les critères considérés sont :

- Interopérabilité
  - **Formats des données :** la possibilité d'importer et exporter dans des formats standards
  - **Projections :** la prise en compte de la projection des images
  - **Compatibilité :** compatibilité avec d'autres outils SIG
  - **Intégration de modèles :** possibilité d'intégration de modèles de vision
- Ergonomie
  - **Prise en main :** la facilité à s'approprier l'outil
  - **Interface :** la clarté de compréhension de l'interface
  - **Facilité d'utilisation :** présence d'outils pratiques, de raccourcis clavier, paramétrages, etc.
  - **Numérisation avancées :** gestion des accrochages, éviter les chevauchements, remodeler une entité, etc.
- Gestion de projet
  - **Suivi de l'avancement :** facilité de suivi de l'avancement
  - **Collaboration :** outils collaboratifs (commentaires, etc.)
  - **Indicateurs :** présence d'indicateurs et de métriques

Ces critères sont évalués sur une échelle de 0 à 5, ce qui nous permet ensuite de comparer ces résultats entre les outils.

4 Annotation de la baseline

Pour confectionner la baseline, nous avons sélectionné des images représentatives de 3 types de zones : urbain dense, péri-urbain et rural. De cette manière, nous pourrons comparer les résultats en fonction de ce type de zone, pour identifier si des outils sont plus performants sur une zone plutôt qu'une autre.

De la même manière, la nomenclature des classes à annoter couvre une grande partie des éléments que nous pouvons observer sur ces images.

FIGURE 1 – Classes utilisées pour l'annotation

Ces 3 zones d'intérêt et la nomenclature ont été déterminées suite à l'expérience significative de la collaboration entre Digitanie et le CNES.

Dans le cadre de la zone urbain dense, une image provenant de la ville de Paris a été sélectionnée. Nous obtenons le résultat présenté avec l'image 2.

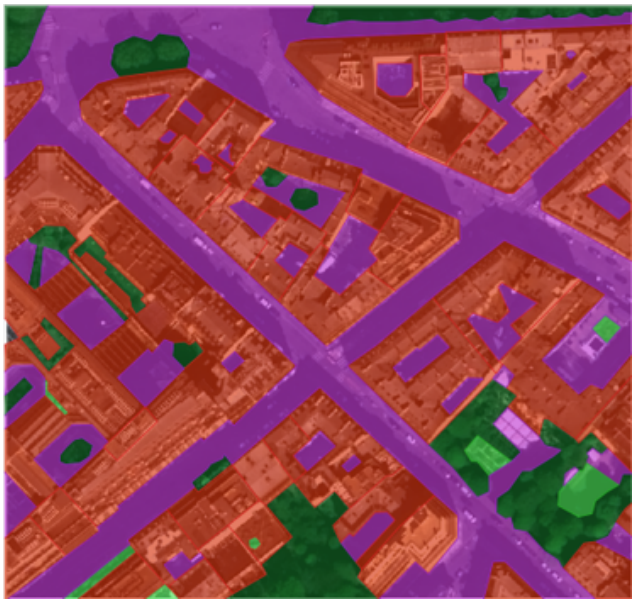

FIGURE 2 – Résultat de l'annotation manuelle pour la zone urbaine dense (Paris)

Dans le cadre de la zone péri-urbaine, une image provenant des alentours de la ville de Nice a été sélectionnée. Nous obtenons le résultat présenté avec l'image 3.

Dans le cadre de la zone rurale, une image provenant des alentours de la ville de Nouméa a été sélectionnée. Nous obtenons le résultat présenté avec l'image 4.

Ces trois baselines seront utilisées comme référentiels pour comparer les résultats obtenus par les outils sélectionnés.

5 Analyse des résultats

Pour les expérimentations, nous avons exclu de l'étude 3 outils sur les 7 sélectionnés :

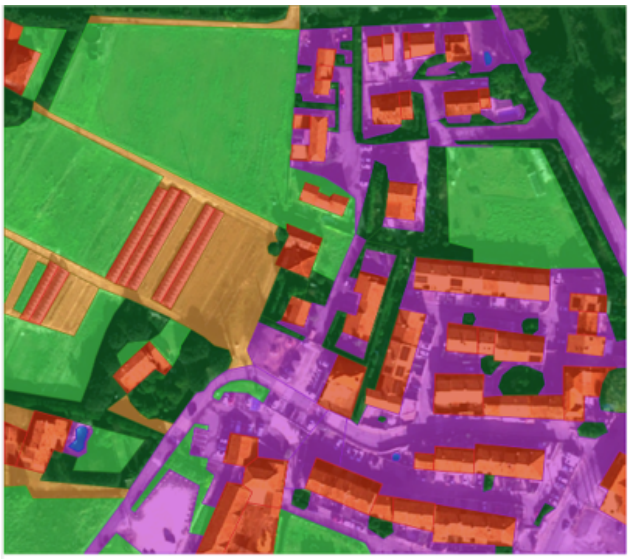

FIGURE 3 – Résultat de l'annotation manuelle pour la zone péri-urbaine (Nice)

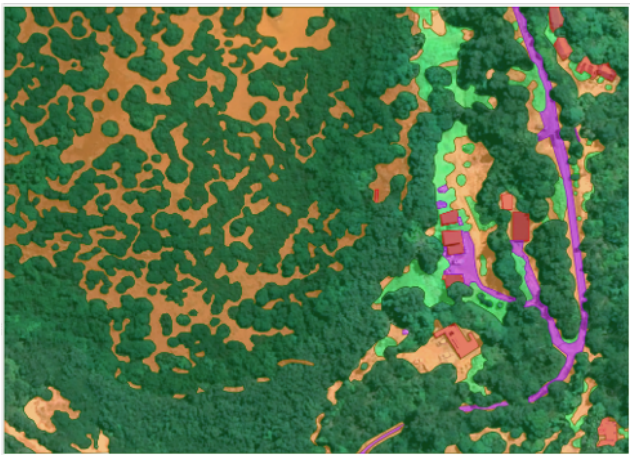

FIGURE 4 – Résultat de l'annotation manuelle pour la zone rurale (Nouméa)

Labelbox présente de bons résultats en termes de qualité, d'interopérabilité et d'ergonomie. En revanche, des contraintes techniques importantes ont été rencontrées lors de l'import et de l'exploitation des données géospatiales. L'outil nécessite notamment l'hébergement préalable des images sur une infrastructure cloud accessible par URL, ce qui complique l'utilisation de jeux de données internes. Par ailleurs, l'export des annotations s'effectue sous forme de masques raster associés à des métadonnées (.ndjson), nécessitant un traitement intermédiaire pour reconstruire des couches vectorielles exploitables dans un environnement SIG. Cet outil a par conséquent été exclu de la phase expérimentale.

Geo4i constitue une solution robuste et collaborative, adaptée à des environnements industriels exigeants. Cependant, nous n'avons obtenu qu'une version de démonstration de cet outil, ne permettant pas l'import et le traitement de jeux de données externes. Il n'a donc pas été possible d'appliquer le protocole expérimental (import des tuiles, annotation selon les scénarios définis, mesure des temps de production). Cet outil a par conséquent été exclu de la phase expérimentale.

L'outil **GeoAI (Esri)** est très bien référencé, grâce à sa forte qualité, son interopérabilité native et son intégration poussée dans l'écosystème ArcGIS, mais son coût et son verrouillage propriétaire doivent être pris en compte. GeoAI dispose de nombreux modèles à disposition, Mais il nous a été impossible de réussir à avoir une sortie exploitable.

Nous avons donc considéré les **4 outils restants pour analyser les résultats : Kili, CVAT+SAM, Supervisely et Mapflow.ai.** Pour chacun, nous avons tout d'abord utilisé l'outil d'aide à l'annotation pour pouvoir réaliser les premières annotations. Néanmoins, nous nous sommes rendu compte que la qualité des résultats obtenus par ces outils était loin des attentes en comparaison avec la baseline. C'est pour cela que nous avons ensuite comptabilisé le temps nécessaire, en retravaillant avec QGIS (l'outil utilisé pour produire la baseline) pour corriger les erreurs, combler les manques et affiner les résultats jusqu'à obtenir une annotation complète de la tuile.

L'image 5 est le résultat obtenu pour la tuile de la zone "Urbain dense" (Paris) en utilisant l'outil Kili. L'image de gauche présente le résultat obtenu pour la baseline, l'image du milieu est le résultat obtenu en utilisant l'outil d'aide à l'annotation de Kili. On remarque que cet outil n'a pas permis de remplir l'intégralité de la tuile et comporte des erreurs et des creux. Le niveau de qualité n'est donc pas satisfaisant comme présenté précédemment. Pour tendre vers un résultat se rapprochant le plus de la baseline possible, nous avons fait un post traitement (en utilisant QGIS, comme pour la baseline) pour atteindre le résultat obtenu sur l'image de droite. Le temps de traitement est plus court en comptabilisant toute la chaîne de traitement en utilisant l'outil Kili. Néanmoins, le résultat est encore un peu éloigné de la baseline. Pour mesurer cette distance, nous appliquons un calcul de précision et de rappel sur les deux étapes de traitement en fonction de la baseline.

La précision et le rappel sont calculés en superposant les polygones de la baseline et des résultats obtenus par l'outil. De cette manière, nous avons calculé la superficie correspondant à la même classe sur les deux annotations. Ce calcul est effectué pour le RAW et pour la post-édition. La précision est calculée avec la formule suivante :

$$precision = S_{inter}/S_{resultat} \quad (1)$$

Avec S_{inter} la surface de l'intersection pour la même classe entre la baseline et le résultat obtenu par l'outil et $S_{resultat}$ la surface pour cette classe dans le résultat obtenu par l'outil.

Le rappel est calculé avec la formule suivante :

$$rappel = S_{inter}/S_{baseline} \quad (2)$$

Avec S_{inter} l'intersection comme pour la précision et $S_{baseline}$ la surface de cette classe dans la baseline. Le F1-score est calculé avec une moyenne harmonique.

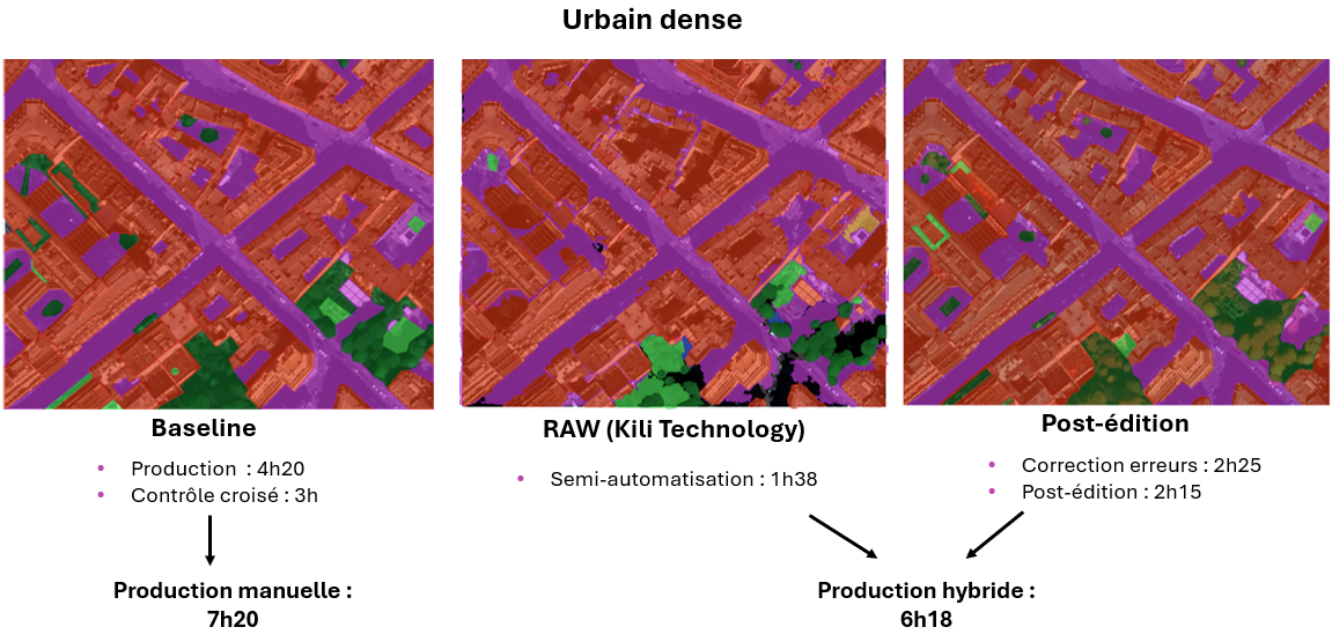

FIGURE 5 – Exemple de post traitement sur la tuile de Paris avec l'utilisation de l'outil Kili Technology

5.1 Résultats bruts (RAW)

Le tableau 3 présente les résultats obtenus pour le calcul de la précision, du rappel et du F1-score pour les 3 zones, pour les 4 outils sélectionnés.

Nous observons que les scores sont, globalement, supérieurs pour l'outil Supervisely, sur les 3 zones, pour les 6 classes. Les 4 outils ont des résultats sensiblement supérieurs pour la zone "péri urbain". Les outils produisent généralement de bons résultats pour les classes de grande taille et bien contrastées, telles que "Building", tandis que les classes plus hétérogènes comme "Hardscape" ou "Bare ground" présentent davantage d'erreurs de segmentation. Néanmoins, les scores globaux n'atteignent pas la qualité d'une annotation manuelle. Nous pouvons en conclure que l'utilisation des outils sans post-édition n'est pas réellement satisfaisante, quel que soit l'outil.

5.2 Résultats de la post-édition

Le tableau 4 présente les résultats après une post-édition. Cette édition est le traitement dans le logiciel QGIS des résultats obtenus par les outils. En utilisant la même méthodologie que pour la baseline, nous tentons, ici, de se rapprocher au plus de la qualité de celle-ci. Ce post-traitement n'est pas une reprise de l'intégralité de la tuile (qui reviendrait à tout refaire comme pour la baseline) mais une édition pour corriger les erreurs topologiques et des ajustements de labellisation et de segmentation.

Nous observons que, même avec du post traitement, les résultats n'atteignent pas le niveau d'une production manuelle.

Le tableau 5 présente les temps qui ont été nécessaires pour le traitement des différentes zones pour chaque outil. Ces temps sont confrontés aux temps nécessaires pour la réalisation de la baseline. De cette manière, nous pouvons observer quels outils ont nécessité plus ou moins de temps pour réaliser tout le traitement. Nous remarquons que l'outil Supervisely est celui qui génère le moins d'erreurs. Néanmoins, ce n'est pas toujours l'outil qui demande le moins de temps en post-édition. En fonction des géométries générées, il est plus ou moins facile de pouvoir les corriger à posteriori. Si les géométries générées comportent beaucoup de détails, parfois trop justement, il peut être nécessaire de reprendre toute la géométrie. Nous remarquons que l'outil Kili technology nécessite beaucoup plus de temps pour les zones peri urbain et rural que pour la baseline. Ce qui s'explique par une performance certaine de cet outil pour la détection des classes "Building" et "Hardscape" très présentes dans la classe urbain dense (c.f. tableau 4). L'outil CVAT+SAM a nécessité moins de temps que la baseline pour la zone peri urbain, là où tous les autres outils ont demandé bien plus de temps. La pertinence de cet outil pour la détection de l'eau (c.f. tableau 3) avec une précision certaine permet de simplifier le traitement de cette tuile. Néanmoins, la post-édition a permis à d'autres outils, notamment Mapflow.ai d'atteindre un score bien plus élevé sur la détection de l'eau sur cette tuile, mais avec des temps plus élevés.

De manière générale, les outils sont plutôt adaptés pour de la détection de zone urbaine dense comme Paris, car nous observons des temps de traitements, pour tous les outils, inférieurs à la baseline avec des scores entre 0.79 et 0.87 en moyenne sur ces outils. Ces résultats sont très intéressants dans la perspective d'améliorer la productivité sur ces zones, bien que la qualité soit diminuée.

Le tableau 6 présente les résultats des questionnaires concernant l'interopérabilité, l'ergonomie et la gestion de projet. Ces questionnaires ont été remplis par les opérateurs durant la réalisation de la production. Ils permettent de mesurer le confort d'utilisation de l'outil grâce à des critères

URBAIN DENSE (Paris)

	CLASSES																		Moyenne
	Bare ground			Building			Hardscape			Tall Vegetation			Low Vegetation			Water			
	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	
Kili Technology	0,00	0,00	0,00	**0,90**	**0,85**	**0,87**	0,76	0,75	0,75	0,35	**0,96**	0,51	0,03	0,06	0,04	**1,00**	**1,00**	**1,00**	0,53
CVAT+SAM	0,00	0,00	0,00	**0,87**	**0,91**	**0,89**	0,76	0,80	0,78	0,66	0,73	0,69	0,00	0,00	0,00	**1,00**	**1,00**	**1,00**	0,56
Supervisely	**1,00**	**1,00**	**1,00**	**0,91**	**0,91**	**0,91**	0,76	**0,85**	**0,80**	0,70	**0,85**	0,77	0,16	0,70	0,26	**1,00**	**1,00**	**1,00**	0,79
Mapflow.ai	**1,00**	**1,00**	**1,00**	**0,89**	**0,93**	**0,91**	0,09	**0,96**	0,16	0,69	0,78	0,73	0,00	0,00	0,00	**1,00**	**1,00**	**1,00**	0,63

PERI URBAIN (Nice)

	CLASSES																		Moyenne
	Bare ground			Building			Hardscape			Tall Vegetation			Low Vegetation			Water			
	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	
Kili Technology	0,79	0,45	0,57	0,78	**0,85**	**0,81**	0,62	**0,81**	0,70	0,41	**0,84**	0,55	0,46	**0,82**	0,59	0,57	0,64	0,60	0,64
CVAT+SAM	0,41	0,21	0,28	**0,83**	**0,92**	**0,87**	0,52	0,75	0,61	0,55	**0,85**	0,67	0,54	**0,94**	0,69	**0,83**	0,39	0,53	0,61
Supervisely	0,75	0,43	0,55	**0,85**	**0,85**	**0,85**	0,71	0,80	0,75	0,71	0,75	0,73	**0,84**	**0,88**	**0,86**	0,68	0,80	0,74	0,75
Mapflow.ai	0,00	0,00	0,00	0,80	**0,93**	**0,86**	0,11	0,63	0,19	0,59	0,66	0,62	0,00	0,00	0,00	0,00	0,00	0,00	0,28

RURAL (Nouméa)

	CLASSES																		Moyenne
	Bare ground			Building			Hardscape			Tall Vegetation			Low Vegetation			Water			
	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	
Kili Technology	0,11	0,60	0,19	**0,82**	**0,81**	**0,81**	0,39	0,75	0,51	**0,86**	**0,87**	**0,86**	0,19	0,19	0,19	**1,00**	**1,00**	**1,00**	0,59
CVAT+SAM	0,10	0,78	0,18	0,71	**0,92**	**0,80**	0,45	0,39	0,42	0,65	**0,86**	0,74	0,06	0,01	0,02	**1,00**	**1,00**	**1,00**	0,53
Supervisely	0,66	0,71	0,68	0,80	**0,89**	**0,84**	0,42	0,67	0,52	**0,84**	**0,91**	**0,87**	0,35	0,67	0,46	**1,00**	**1,00**	**1,00**	0,73
Mapflow.ai	0,00	0,00	0,00	0,73	**0,92**	**0,81**	0,25	0,54	0,34	**0,95**	0,78	**0,86**	0,00	0,00	0,00	**1,00**	**1,00**	**1,00**	0,50

TABLE 3 – Résultats RAW

URBAIN DENSE (Paris)

	CLASSES																		Moyenne
	Bare ground			Building			Hardscape			Tall Vegetation			Low Vegetation			Water			
	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	
Kili Technology	**1,00**	**1,00**	**1,00**	**0,94**	**0,94**	**0,94**	**0,92**	**0,83**	**0,87**	**0,85**	**0,99**	**0,91**	0,23	**0,92**	0,37	**1,00**	**1,00**	**1,00**	**0,85**
CVAT+SAM	**1,00**	**1,00**	**1,00**	**0,92**	**0,94**	**0,93**	**0,90**	0,80	**0,85**	0,80	**0,82**	**0,81**	0,09	0,64	0,16	**1,00**	**1,00**	**1,00**	0,79
Supervisely	**1,00**	**1,00**	**1,00**	**0,93**	**0,95**	**0,94**	**0,92**	**0,87**	**0,89**	**0,91**	**0,85**	**0,88**	0,34	**0,90**	0,49	**1,00**	**1,00**	**1,00**	**0,87**
Mapflow.ai	**1,00**	**1,00**	**1,00**	**0,97**	**0,94**	**0,95**	**0,94**	**0,96**	**0,95**	**0,92**	0,80	**0,86**	0,00	0,00	0,00	**1,00**	**1,00**	**1,00**	0,79

PERI URBAIN (Nice)

	CLASSES																		Moyenne
	Bare ground			Building			Hardscape			Tall Vegetation			Low Vegetation			Water			
	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	
Kili Technology	**0,97**	0,78	**0,86**	**0,87**	**0,88**	**0,87**	**0,87**	**0,82**	**0,84**	0,61	**0,90**	0,73	**0,92**	**0,86**	**0,89**	**0,91**	0,66	0,77	**0,83**
CVAT+SAM	**0,85**	0,73	0,79	**0,95**	**0,93**	**0,94**	**0,92**	**0,84**	**0,88**	**0,83**	**0,87**	**0,85**	**0,85**	**0,97**	**0,91**	**0,83**	0,44	0,58	**0,82**
Supervisely	0,75	0,45	0,56	**0,85**	**0,86**	**0,85**	0,71	**0,81**	0,76	0,74	0,79	0,76	**0,90**	**0,90**	**0,90**	0,68	0,80	0,74	0,76
Mapflow.ai	**0,85**	**0,82**	**0,83**	**0,81**	**0,96**	**0,88**	**0,95**	0,76	**0,84**	**0,92**	0,78	**0,84**	0,00	0,00	0,00	**0,99**	0,69	**0,81**	0,70

RURAL (Nouméa)

	CLASSES																		Moyenne
	Bare ground			Building			Hardscape			Tall Vegetation			Low Vegetation			Water			
	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	R	P	F1	
Kili Technology	0,64	0,70	0,67	**0,86**	**0,85**	**0,85**	0,43	0,78	0,55	**0,92**	**0,89**	**0,90**	0,40	0,38	0,39	**1,00**	**1,00**	**1,00**	0,73
CVAT+SAM	0,50	**0,86**	0,63	**0,96**	**0,94**	**0,95**	0,72	0,45	0,55	**0,97**	**0,90**	**0,93**	0,60	0,61	0,60	**1,00**	**1,00**	**1,00**	0,78
Supervisely	0,72	0,77	0,74	**0,87**	**0,90**	**0,88**	0,73	0,78	0,75	**0,94**	**0,91**	**0,92**	0,60	**0,89**	0,72	**1,00**	**1,00**	**1,00**	**0,84**
Mapflow.ai	0,74	0,80	0,77	**0,96**	**1,00**	**0,98**	0,50	0,65	0,57	**0,99**	**0,81**	**0,89**	0,00	0,00	0,00	**1,00**	**1,00**	**1,00**	0,70

TABLE 4 – Résultats post édition

URBAIN DENSE (Paris)

	nb erreurs geom	nb erreurs topo	Temps traitement	Temps post-édition		Temps total
				corrections erreurs	corrections seg/lab	
BASELINE	0	0	07:20	00:00	00:00	**07:20**
Kili Technology	0	130	01:38	02:25	02:15	**06:18**
CVAT+SAM	0	130	00:12	02:15	02:10	**04:37**
Supervisely	0	0	01:40	00:00	02:35	**04:15**
Mapflow.ai	41	123	00:03	02:00	01:50	**03:53**

PERI URBAIN (Nice)

	nb erreurs geom	nb erreurs topo	Temps traitement	Temps post-édition		Temps total
				corrections erreurs	corrections seg/lab	
BASELINE	0	0	06:00	00:00	00:00	**06:00**
Kili Technology	205	727	02:12	10:00	03:50	16:02
CVAT+SAM	0	110	00:26	02:55	02:10	**05:31**
Supervisely	0	1	02:00	00:01	08:15	10:16
Mapflow.ai	47	119	01:45	02:30	06:30	10:45

RURAL (Nouméa)

	nb erreurs geom	nb erreurs topo	Temps traitement	Temps post-édition		Temps total
				corrections erreurs	corrections seg/lab	
BASELINE	0	0	04:40	00:00	00:00	**04:40**
Kili Technology	5	373	03:24	06:50	02:20	12:34
CVAT+SAM	0	68	00:13	01:10	03:35	04:58
Supervisely	0	0	04:00	00:00	01:20	05:20
Mapflow.ai	5	235	00:01	02:05	02:45	04:51

TABLE 5 – Comparaison des temps de production manuelles et hybrides

	Interopérabilté					Ergonomie					Gestion de projet				Note globale
	Formats des données	Projections	Compatibilité	Intégration de modèles	Note moyenne	Prise en main	Interface	Facilité d'utilisation	Numérisation avancés	Note moyenne	Suivi de l'avancement	Collaboration	Indicateurs	Note moyenne	
Kili Technology	4	3	2	1	**2,5**	4	2	3	3	**3,0**	4	5	3	**4,0**	**3,2**
CVAT+SAM	0	0	0	3	0,8	3	4	3	0	2,5	5	4	0	**3,0**	2,1
Supervisely	2,5	2	3	4	**2,9**	4	5	4	2	**3,8**	4	4	2	**3,3**	**3,3**
Mapflow.ai	3	2	4	0	**2,3**	3	3	4	5	**3,8**	5	0	0	1,7	2,6

TABLE 6 – Résultats de l'analyse de l'ergonomie et de la gestion de projet

plus qualitatifs. Nous remarquons que l'outil Supervisely se démarque des autres concernant l'interopérabilité, notamment avec l'intégration des modèles, alors que Kili Technology est particulièrement pertinent sur les formats de données. CVAT+SAM est assez bas concernant l'interopérabilité, car il a obtenu les notes de 0 en formats des données, en projection et en compatibilité. Concernant l'ergonomie, là encore, Supervisely est au-dessus du lot, notamment avec sa très bonne note concernant l'interface, bien qu'il manque de fonctionnalités de numérisation avancées. Enfin concernant la gestion de projet, l'outil Kili Technology se démarque particulièrement surtout concernant la collaboration.

6 Conclusion et perspectives

Notre étude porte sur le comparatif des outils d'annotation pour de la photo interprétation d'images satellites dans un objectif de formalisation de l'occupation du sol. Nous avons, pour cela, comparé 26 outils avec des critères de sélection en étudiant les documentations et publications en lien avec ces outils. Cette étude nous a permis de sélectionner 7 outils. Le protocole a pu être appliqué pour 4 d'entre eux, permettant de les comparer plus en détail. Ce comparatif étendu se concentre sur les temps de production pour chacun en fonction d'une baseline réalisée manuellement. Nous observons que la qualité des résultats obtenus par ces outils ne permettent pas d'atteindre le résultat de la baseline. Afin de tendre vers un résultat satisfaisant, nous appliquons une post-édition pour corriger certaines erreurs, topologiques par exemple. La comparaison de ces résultats et au regard du temps passé à utiliser l'outil et à corriger ces erreurs en post édition ne permettent, non seulement pas d'atteindre la qualité de la baseline mais en plus ce résultat prend parfois plus de temps à obtenir. Au regard de cette analyse, un traitement purement manuel pour ce type de tâches reste le plus adapté et le plus optimisé pour atteindre une qualité satisfaisante.

Afin d'améliorer cette étude, nous pouvons considérer plusieurs biais. Tout d'abord, nous avons réalisé ce comparatif sur un ensemble restreint d'images et d'outils. Il pourrait être intéressant de multiplier le nombre d'images et d'outils pour vérifier si nous observons les mêmes résultats. Ensuite, chaque traitement a été réalisé par une personne, il y a donc un biais de subjectivité dans l'analyse et l'interprétation de l'image et des critères qualitatifs. Il pourrait être intéressant de multiplier les personnes qui réalisent ce travail pour permettre d'éviter ce biais.

La perspective majeure est de considérer qu'il n'existe, aujourd'hui, pas d'outil réellement dédié à la photo interprétation qui permettent aussi de faciliter la gestion de projet et le contrôle qualité tout en maximisant la qualité au niveau de ce qu'il est possible d'atteindre avec QGIS. L'outil, QGIS, reste une référence dans ce domaine et permet de réaliser des traitements très poussés. Il serait compliqué d'envisager le développement d'un nouvel outil pouvant le concurrencer. Néanmoins, créer une plateforme qui permette de synchroniser des projets QGIS grâce à une extension tout en garantissant des fonctionnalités de gestion de projet (suivi d'avancement, contrôle qualité, livraison client, etc.) serait une combinaison idéale pour permettre d'optimiser ce que nous souhaitons réaliser ici. De plus, avec l'avènement des modèles de visions qui se développent de plus en plus, leur intégration dans QGIS est très souvent privilégiée. Les résultats obtenus dans cette étude ne permettent pas de conclure que l'utilisation d'un modèle de vision améliore la productivité, mais nous sommes convaincus qu'à l'avenir, ce sera le cas. Aujourd'hui, les outils sont attachés à un ou plusieurs modèles sans pouvoir intégrer d'autres modèles. C'est un frein pour certains bons outils associés à de mauvais modèles. Créer une interface standardisée permettrait de s'abstraire de ces contraintes techniques et de pouvoir utiliser n'importe quel modèle.

Références

[1] Fabien Amarger, Nathalie Noblecourt, Jehanne Portefaix, and Pierre-Marie Brunet. L'insertion au service de l'intelligence artificielle : Modèle d'apprentissage pour l'annotation d'images satellite pour le consortium AI4GEO. In *Conférence Nationale sur les Applications de l'Intelligence Artificielle*, APIA-2025, Dijon, France, June 2025.

[2] Pierre-Marie Brunet, Pierre Lassalle, Simon Baillarin, Bruno Vallet, Arnaud Le Bris, Gaëlle Romeyer, Guy Le Besnerais, Flora Weissgerber, Gilles Foulon, Vincent Gaudissart, et al. Ai4geo : A data intelligence platform for 3d geospatial mapping. In *24th ISPRS Congress Commission II : Imaging Today, Foreseeing Tomorrow*, volume 43, pages 817–823, 2021.