

Aurélien COSTE, Esteban THÉVENON, Lony RIFFARD, Florian MACHENAUD

Rapport Final

Rappel du sujet/besoin et cahier des charges

Notre projet est porté par le Parc National des Écrins et par l'OFB (Office Français de la Biodiversité). L'objectif est de connaître la fréquentation des parcs par les visiteurs et usagers afin d'assurer au mieux leurs missions de protection, de connaissance, ainsi que d'accueil du public.

Les données de flux physiques (eco-compteurs, enquêtes...) collectés avec des méthodes traditionnelles demandent à être consolidées en terme d'analyse et confrontés avec des solutions innovantes en devenir (big data, informations géographiques, IA...).

En 2022, dans le contexte de hausse de la fréquentation post covid le Parc national des Ecrins a redéfini ses objectifs en matière de suivi des flux de visiteurs.

Cela a amené le Parc national des Ecrins à collaborer avec l'OFB à l'été 2023 pour expérimenter et consolider une solution de suivi innovante par l'utilisation de l'Intelligence artificielle associée à des capteurs photos. Solution mise au point par Mathieu Garel, chercheur à l'OFB, et qui a été testée (robustesse, sur 4 sites de terrain), comparée à d'autres solutions (Deepfaune...) et qui a fait l'objet de développements pour la consolider et l'intégrer dans une chaîne de travail propre au PNE. La chaîne de travail testée en 2023 va de la collecte de terrain, en passant par le floutage des visages, la détection, la classification, et la visualisation des données dans l'application Ist SOS.

A noter que le cœur du projet repose sur un script de détection automatique d'objets (ici les "humains") dans des images, basé sur le modèle YOLOv4 entraîné sur le jeu de données COCO. Ce script utilise le modèle de réseau de neurones yolov4 pour détecter des objets dans des images provenant d'un serveur FTP ou d'un répertoire local. Il permet ainsi de compter automatiquement, sans visualisation par l'utilisateur, le nombre de personnes présentes sur des images, notamment dans le cadre de suivis de la fréquentation réalisés avec des pièges photos à déclenchement automatique. Le script compte le nombre maximum d'humains au sein de chaque séquence, retenu comme taille de groupe. Le travail d'expérimentation et de consolidation a été mené dans le cadre du stage d'Aurélien Coste, 4ème année à Polytech Grenoble de mai à août 2023.

Les usages des données de fréquentation sont diverses :

- alimenter les indicateurs de données socio-économiques pour évaluer le poids de la randonnée dans les espaces naturels,
- suivi des flux à des fins d'observation,
- travaux de recherche appliquée pour étudier les interactions faune / pratiques récréatives (cas de l'OFB dans les massifs de Belledonne et des Bauges par ex.).

La solution de suivi de la fréquentation par capteurs photos associé à de l'IA permet de répondre à des besoins des gestionnaires d'espaces protégés et à des projets de recherche appliquée en espaces naturels. Le projet présente un fort potentiel d'utilisation par d'autres structures que le PNE et l'OFB. A l'issue de cette nouvelle phase de R&D, l'essaimage de l'usage de cette solution pourrait se faire à court-moyen

terme dans le cadre de réseaux constitués (Parcs nationaux, conservatoires d'espaces naturels, réserves naturelles, organismes de recherche...).

L'objectif du projet est d'approfondir l'expérimentation menée en 2023 et poursuivre la consolidation de l'outil mis au point par Mathieu Garel en intégrant notamment les derniers développements (floutage des cadres de détection plutôt que les visages).

Il s'agira de mener un travail exploratoire afin de :

- renforcer et tester la fiabilité de la solution,
- de développer l'outil (sens de passage, upgradage de briques, etc.),
- de tester l'utilisation de l'outil dans un contexte d'étude précis à partir de jeu de données fournies,
- et de proposer des scénarios d'évolution tel que l'utilisation de l'IA embarquée et l'IoT.

Le Parc national a souhaité que le projet se déroule en méthode Agile. Dans le détail, voici les tâches que l'on nous a confiées:

- Réaliser une analyse de l'outil développé en mode prototype à l'été 2023 (résultats, analyse "critique" du processus et recommandations d'améliorations) + ré investiguer les outils équivalents, y compris commerciaux
- Analyser, critiquer et redéfinir au besoin l'architecture globale du système et de la chaîne de travail (objets, script d'IA, floutage des images, serveurs, data visualisation...),
- Identifier les obligations en matière de respect de la vie privée du service (RGPD) et du droits à l'image (risques d'atteinte au respect de la vie privée ou à d'autres aspects réglementaires...)
- Étudier et définir la sécurité du projet
- Approfondir la faisabilité de développer la reconnaissance des sens de passage des piétons par l'IA-
Réaliser les développements pour la détection des sens de passage : modélisation de l'évolution du modèle IA, réalisation de codes, tests de classification, tests de fiabilité, développement des API en V2, développement de l'outil Ist SOS pour la visualisation double sens...
- Analyser la pertinence et le cas échéant upgrader vers yolov8 pour monter en capacité (rapidité)
- Réaliser les tests de la solution développée à partir de sets de données fournies
- Continuer la calibration des modèles utilisés sur d'autres éléments que le simple comptage des humains avec notamment la caractérisation de la pratique (randonnée, ski de randonnée, VTT, présence ou non de chien)
- Documenter les développements sur le dépôt Github du projet.
- Analyser la faisabilité et élaboration d'un modèle mathématique/cartographique pour essayer d'analyser les flux d'un site avec une logique de maillage d'itinéraires. Cas d'étude du site du Lauvitel en Oisans.
- Exploiter les données connexes (niveaux d'équipements des piétons, chiens, faune...) et les rendre visualisables dans l'outil IST SOS.

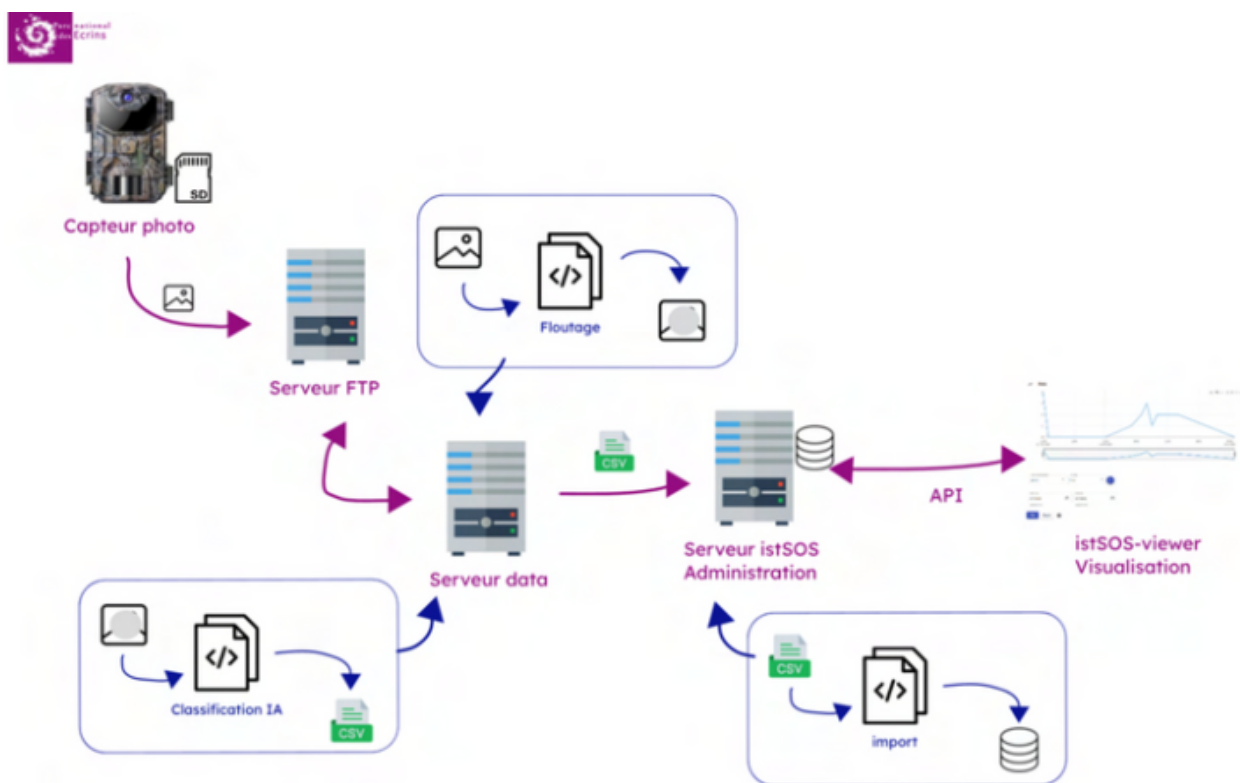
- Prospective et recommandations sur l'évolution de la chaîne de travail à moyen terme : remplacement des capteurs photos par IA embarquée, serveurs, ...
- Volet prospectif :
 - Prospector et estimer le coût d'objectif d'un scénario à moyen terme de remplacement des capteurs photos par le développement d'une solution d'IA embarquée avec transmission LoRaWAN pour 50 objets (développement du capteur intégrant l'IA et déploiement d'un mode de transmission à distance) -> Changement à 5 objets sur une zone précise suite à un entretien.
 - Prospector pour la réalisation d'une solution de dénombrements de tentes de bivouac sur photos (set de donnée disponible au besoin).

Technologies employées

- YOLOv8 : il s'agit d'un ensemble de modèles de classification, de détection, de segmentation, de suivi (tracking) et de détermination de squelette humain (pose) développé par d'Ultralytics. Ces modèles sont plus performants que le modèle actuellement utilisé dans yolov4-attendance. De plus, la diversité des modèles nous permet d'apporter des informations supplémentaires telles que le sens de passage grâce au modèle YOLOv8-pose qui sait générer un squelette sur un corps humain, en y identifiant des points (articulations, et points du visage).
- CLIP : CLIP est un réseau de neurones qui apprend efficacement les concepts visuels de la supervision du langage naturel. CLIP peut être appliqué à n'importe quel repère de classification visuelle en fournissant simplement les noms des catégories visuelles à reconnaître, de la même façon que les capacités « zero-shot » du GPT-2 et du GPT-3. C'est un outil développé par OpenAI.
- istSOS : Logiciel de visualisation de données, utile pour voir les résultats des analyses. Outil déjà utilisé par le PNE, d'où son utilisation dans le projet.

Architecture techniques

En ce qui concerne l'architecture technique, nous avons la contrainte de devoir réaliser un outil utilisable sur la chaîne de travail actuelle.



Tout d'abord, cette chaîne de travail commence avec les pièges photos. Ces pièges sont utilisés pour prendre des photos des visiteurs lorsqu'ils passent devant l'appareil. Ces photos sont stockées sur une carte SD propre à chaque appareil. Ces données sont ensuite récoltées à la main et déposées sur le serveur FTP.

Cette chaîne de travail est composée de 3 serveurs. Le premier est un serveur FTP. Les photos récoltées des pièges photos sont stockées sur ce serveur. Ensuite, il y a un serveur data, qui permet de faire tourner le modèle, de récolter les données des photos sous forme de CSV, et de flouter les photos traitées afin de respecter la vie privée des visiteurs. Ces données de sorties CSV sont ensuite envoyées vers un dernier serveur d'administration et de visualisation de données. Ce dernier serveur permet d'utiliser les données CSV pour les visualiser avec l'outil istSOS. À préciser qu'il ne s'agit pas physiquement de serveurs différents, ce sont en réalité trois machines virtuelles sur un puissant serveur physique.

Réalisations techniques

Tout d'abord, nous avons pris en main les technologies étudiées. Lors de nos recherches, nous avons constaté que les technologies évoluaient rapidement, et que Ultralytics avaient sorti une version bien plus optimisée et performante que YOLOv4. Nous avons donc décidé de les comparer, en faisant une étude.

Pour cela, nous avons besoin d'un jeu de donnée, si possible du parc ou de l'OFB. Ces images doivent être annotées pour que l'on puisse extraire de ces images, des valeurs fiables pour comparer avec ce que le modèle trouve. Nous pourrions également annoter des images si nous n'en avons pas assez, mais cela prend énormément de temps et 6 semaines, c'est court. Nous nous sommes donc rendu dans les locaux de Mathieu Garel, chercheur à l'OFB. Nous y avons récupéré les jeux de données annotés. Par la suite, nous avons créé un code qui nous permettait d'extraire les métadonnées des photos, et de les répertorier dans un fichier JSON.

Après, avoir récupéré les données, il faut choisir les modèles à comparer. Il faut savoir que YOLOv8 possède plusieurs modèles de différentes tailles. Certains modèles sont très légers, mais moins précis, et d'autres

plus lourds, mais plus précis. Il existe 5 versions, de la plus légère à la plus lourde : N (nano), S (small), M (medium), L (large), X (extra-large).

Ensuite, nous avons fait tourner les différents modèles, et comparé les sorties des modèles aux annotations des photos, effectuées par des humains.

Nous avons donc trouvé les résultats suivants :

20200709_20200802 (12 885 images)

Test avec 16Go de RAM, SSD, AMD Ryzen 5600H 2.28GHz, GPU RTX 3060 Mobile (PC Gamer)

Nom du modèle	Temps de calcul (min)	Classifications correctes et comptages corrects (%) 100_BTCF	Classifications correctes et comptages corrects (%) 101_BTCF	Classifications correctes et comptages corrects en moyenne(%)
yolov4	28'22" (1682.22sec)	95.2%	93.0%	94.67%
yolov8n	5'43" (342.83sec)	95.3%	92.6%	94.66%
yolov8s	5'46" (346.3sec)	94.8%	93.1%	94.39%
yolov8m	7'04" (424.7sec)	97.9%	95.9%	97.42%
yolov8l	7'55" (474.7sec)	98.2%	96.5%	97.81%
yolov8x	9'39" (579sec)	98.8%	97.5%	98.50%

Test avec 8Go de RAM, SSD, Intel I5 11eme Gen 2.4GHz (PC Bureautique)

Nom du modèle	Temps de calcul (s)	Classifications correctes et comptages corrects (%) 100_BTCF	Classifications correctes et comptages corrects (%) 101_BTCF
yolov8n	29'01" (1741sec)	95.3%	92.6%

On s'est donc rendu compte que les dernières versions de YOLO étaient plus précises, mais surtout bien plus efficaces, avec un temps de traitement bien plus petit. Après une réunion avec les porteurs du projet, il a été décidé que l'on utiliserait en version de base, la version "M" de YOLOv8. Cette version permet d'atteindre une précision très bonne, tout en restant assez léger. Bien sûr, les utilisateurs du code peuvent choisir le modèle qui leur convient le mieux, cela via un fichier de configuration et une documentation.

Une fois le modèle choisi, une partie de l'équipe s'est penchée sur la détection de la direction des passants. Pour cela, on utilise une variation du modèle de YOLOv8, qui s'appelle YOLOv8-pose, et qui permet de positionner des points du squelette sur un corps humain détecté. On retrouve ici un exemple de squelette généré par le modèle :



En posant des règles sur la position des jambes, des épaules ou encore des yeux, nous arrivons donc à déterminer la direction de chaque personne sur une photo (ce qui était jusqu'ici impossible).

Pendant ce temps, une autre équipe réalisait l'analyse de CLIP. Il s'agit d'un réseau de neurones capable d'apprendre efficacement les concepts visuels de la supervision du langage naturel. CLIP nous permettrait de déterminer le sexe des passants, leurs catégories d'âge, ou encore leurs types d'activité. Des tests ont donc été effectués pour déterminer la précision de l'outil. Nous avons pu voir que pour certaines choses, comme les activités, la précision était satisfaisante, mais pour d'autres critères, tels que l'âge, la précision n'est pas très haute (~50%).

Nous avons pu trouver une solution alternative à CLIP. Il existe des modèles de YOLOv8 entraînés sur différents jeux de données (datasets), avec différents objets pouvant être détectés. Jusqu'ici, nous utilisons un modèle entraîné sur le jeu de données COCO, qui permet de détecter 80 objets. Il existe d'autres modèles, qui, cette fois-ci, sont entraînés sur le jeu de données d'OpenImagesV7 (OIV7), un dataset de Google, qui permet de détecter près de 600 objets.

L'avantage potentiel d'utiliser un des modèles entraînés avec OIV7 est que parmi les objets détectés, on peut en conclure un certain type d'activité. Par exemple, si le modèle reconnaît des équipements de randonnée, avec un sac à dos, on pourra en conclure que le passant fait de la randonnée. Dans le cas où le modèle détecterait des roues de vélo, un vélo ou un casque de vélo, alors on pourra en conclure que le passant est un cycliste. Ainsi, en posant un certain nombre de règles, il serait possible de détecter différentes activités (randonnée, VTT, ski de randonnée).

Une fois ces règles posées et vérifiées, nous avons comparé la précision de ces règles à celle de CLIP. Après avoir analysé les résultats, nous en avons conclu que le modèle de CLIP était plus précis pour le sexe et aussi précis que le modèle d'OIV7 avec les règles que nous avons posé pour les activités. Il serait donc plus judicieux d'utiliser CLIP lors de notre analyse. Néanmoins, notre chaîne de travail étant déjà complexe, nous avons préféré garder OIV7, ce qui nous permet d'utiliser un seul modèle (pour le qualitatif), et donc d'avoir un meilleur temps de calcul.

Lorsque tous les différents modèles ont été choisis, il a fallu les fusionner dans un seul code pour en faire un seul outil. Après une réunion technique avec Théo Lechémi, il nous a conseillé de faire une version différente en cas d'utilisation d'un FTP (File Transfert Protocol). Jusqu'ici, nous téléchargeons toutes les images à étudier, puis nous appliquons le traitement dessus. Il s'est avéré que pour une utilisation impliquant un serveur FTP, il est plus judicieux de télécharger, de classer et de supprimer les images une à une. Ainsi, il n'est pas nécessaire d'avoir beaucoup d'espace de stockage sur le serveur. Après avoir effectué cette version DCD (Download, Classify, Delete), le client était pleinement satisfait de l'outil.

Pendant qu'une partie de l'équipe travaillait sur le développement de la solution, une autre partie de l'équipe prospectait sur une installation LoRa dans le parc. Cette solution permettrait de se passer de l'intervention humaine entre le piège photo et le serveur FTP, lors de la récupération des photos. Cette équipe a donc fait un document détaillant différentes possibilités d'installation pour le site du Lauvitel. Évidemment, cette solution a un coût, qu'il fallait mettre en évidence. Pour cela, nous avons fait appel à des spécialistes d'équipements LoRa. Ces derniers nous ont conseillé des appareils correspondants à nos besoins, ainsi que le coût de ces derniers. En ce qui concerne les tests, nous avons prévu de tester le modèle sur une carte Raspberry Pi, afin de savoir si ce genre de carte pouvait répondre à nos besoins. Malheureusement, nous n'avons pas eu le temps de tester cette piste.

Gestion de projet (méthode, planning prévisionnel et effectif, gestion des risques, rôles des membres ...)

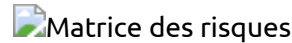
Ce projet étant très court et en considérant que nous avons des compétences techniques, très similaires, nous avons décidé d'organiser notre groupe horizontalement, en d'autres termes, chaque personne a touché à tout dans ce projet sans attribuer de rôles spécifiques à quiconque. Cependant, nous avons tout de même nommé Aurélien Coste chef du projet puisqu'il s'agit de la suite directe de son stage de l'année dernière, il a donc permis de faciliter l'interface entre l'équipe et le client ainsi que d'orienter nos solutions dans la bonne direction grâce à sa connaissance de l'environnement de travail du parc des Écrins tout en organisant le travail de l'équipe.

D'un point de vue méthodologique, nous avons opté pour l'agile avec un point hebdomadaire avec nos clients. Ce qui nous a permis de réaliser notre solution de la manière la plus proche possible aux besoins du client. Afin de nous accorder temporellement, nous avons utilisé l'outil de gestion de projet intégré à GitHub. Ce dernier nous a permis de créer un Gantt dans lequel nous ajoutions nos tâches ainsi que les personnes assignées à ces dernières, dans un second temps, notre client ainsi que tous les collaborateurs pouvaient accéder à notre Gantt en lecture, ce qui permettait des échanges encore plus rapides et avec une transparence parfaite sur le projet.

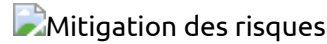


Une des forces de notre organisation a été notre état des lieux initial, ainsi que notre analyse des technologies, des outils et des risques liés à ce projet. Grâce à cette étude, nous avons pu adapter notre travail afin de faciliter le développement de la solution.

Matrice des risques



Mitigation des risques



Outils (collaboration, CD/CI ...)

En ce qui concerne les outils d'organisation, nous avons centralisé au maximum les informations. Pour cela, nous avons utilisé la fonctionnalité *Projet* de GitHub. Cela permet d'avoir une liste de toutes les tâches à faire, celles qui sont en cours de développement et celles terminées (un backlog). Nous avons aussi la possibilité de réaliser une *Roadmap*, ce qui permet d'ajouter une vision temporelle de nos tâches.

En ce qui concerne notre organisation, nous avons plusieurs dépôts. Un dépôt avec la documentation, sur lequel n'importe quel membre du groupe peut commit sans faire de PL. En revanche, sur les dépôts de développement, nous avons mis en place un système de pull request, sur lequel au moins 2 membres du groupe doivent approuver la pull request avant que le commit soit effectué.

Métriques logiciels : lignes de code, langages, performance, temps ingénieur (d'après vos journaux), la répartition des lignes de code et des commits en pourcentage entre les membres du projet ...)

À propos des lignes de code, nous avons environ 700 lignes de code sur le programme principal. À côté de ça, nous avons aussi d'autres programmes qui permettent par exemple l'extraction des métadonnées des images (qui contiennent les annotations de l'OFB). Ce programme fait environ 150 lignes, et le programme qui permet de détecter la direction des passants environ 200 lignes. Il y a également quelques lignes de codes sur la comparaison des modèles et sur l'utilisation de CLIP, environ 100 lignes pour le tout.

Lors du projet, nous avons aussi dû faire de la documentation, notamment sur la comparaison des différents modèles, parmi eux, nous avons réalisé une comparaison sur tous les différents modèles de YOLOv8 et sur le modèle utilisé précédemment sur YOLOv4.

En ce qui concerne la détection du sexe, des activités et de l'âge, nous avons aussi mené une enquête sur la précision de l'outil CLIP et du modèle entraîné sur le dataset de Google OpenImages V7.

Notre projet comprenait aussi une partie prospective sur une installation LoRaWAN au sein du parc, afin de se passer de l'intervention humaine pour récupérer les photos et les uploader sur un serveur. Cette analyse aura pris beaucoup de temps lors de notre projet, car nous avons essayé d'être le plus complet, à la fois sur l'explication du fonctionnement, que sur le choix technique des gateways (nous avons fait appel à EBDS, une entreprise spécialisée dans l'équipement LoRa). Ce document leur a permis une bonne compréhension du sujet, d'avoir un aperçu des différentes problématiques liées à ce genre d'installation, mais aussi tous les avantages que cela pourrait avoir (pas de stockage de photos, pas d'intervention humaine sur place).

En ce qui concerne la répartition du projet, nous estimons que chacun a fourni une quantité de travail égale. Certains ont fait plus de code pendant que d'autres ont plus travaillé sur la partie documentation et rédactionnelle de la prospection LoRaWAN. Nous avons très peu fait de travail personnel en dehors des heures attribuées au projet (~5% du projet a dû être fait en travail supplémentaire).

Pour les commits, au total (seulement liés au code) nous avons la répartition suivante :

- Lony : 5
- Aurélien : 29
- Esteban : 46
- Florian : 17

Ces commits offrent un aperçu de la participation de chacun pour le code, mais ne représentent en rien la participation de chacun pour le projet. Étant donné qu'il y a eu une grande partie rédactionnelle dans ce projet, certains d'entre nous ont dû passer plus de temps sur de la documentation et la prospection LoRaWAN.

Conclusion (Retour d'expérience)

Concernant nos retours, nous avons été très contents de poursuivre le projet qu'avait entamé Aurélien. Le fait de l'avoir dans l'équipe a permis de prendre en main plus facilement l'outil pour ceux qui ne le connaissait pas. Le fait d'avoir des porteurs de projet disponibles chaque semaine pour des meetings a aussi grandement aidé à avancer de manière efficace dans la bonne direction.

Ce projet était concret, et nous estimons que c'est ce qui fait sa force. De plus, il est possible qu'un article scientifique soit rédigé sur ce sujet et sur cet outil, ce qui est très gratifiant pour le groupe et notre travail.

Si nous avons un seul regret sur ce projet, c'est d'avoir été aussi limité sur le temps, car 6 semaines ne nous ont pas permis de pousser les analyses au maximum et d'intégrer des modèles comme CLIP à notre outil, ce qui aurait pu améliorer davantage la précision des informations qualitatives. Cependant, nous restons très contents du livrable que nous avons pu réaliser au cours des 6 semaines.

En conclusion, nous sommes très heureux d'avoir pu répondre à la demande des clients, qui nous ont d'ailleurs félicité pour notre travail réalisé de manière très professionnel.

Glossaire

- PNE : Parc National des Écrins
- OFB : Office Français de la Biodiversité
- IA : Intelligence Artificielle
- YOLO : You Only Look Once
- COCO : Common Objects in Context
- FTP : File Transfert Protocol
- RGPD : Règlement Général sur la Protection des Données
- CLIP : Contrastive Language-Image Pretraining
- API : Application Programming Interface
- CSV : Comma-Separated Values
- JSON : JavaScript Object Notation
- OIV7 : OpenImageV7

Références

- Ultralytics - YOLOv8 - <https://docs.ultralytics.com/fr/models/yolov8/>
- EBDS - Gateways LoRaWAN - <https://www.ebds.eu/produits/equipements/gateways-lorawan>
- PNE/OFB - YOLOv4 Attendance - <https://github.com/Attendance-PNE-OFB/yolov4-attendance>
- Ultralytics - Pose Estimation with YOLOv8 - <https://www.youtube.com/watch?v=Y28xXQmju64>
- OpenAI - CLIP - <https://openai.com/research/clip>
- COCO Dataset - <https://cocodataset.org/#home>
- Google - OpenImagesV7 - <https://storage.googleapis.com/openimages/web/index.html>
- Aurélien COSTE - rapport de stage - <https://data.ecrins-parcnational.fr/documents/stages/2023-09-rapport-stage-Aurelien-Coste-photos-IA-frequentation.pdf>
- Aurélien COSTE - présentation de stage - <https://data.ecrins-parcnational.fr/documents/stages/2023-09-restitution-stage-Aurelien-Coste-photos-IA-frequentation.pdf>