

Sujet de stage M2 ou M1 en Informatique 2025

« Réidentification de chiens par clustering à partir de vidéos issues de pièges photographiques. »

Encadré par Marc Chaumont, Gérard Subsol, et Eugênio Dias Ribeiro Neto.

Vannes : Equipe OBELIX, IRISA, Rue Yves Mainguy, 56000 Vannes. <https://www-obelix.irisa.fr/>

Montpellier : Équipe ICAR, LIRMM, Bât. 5, 860 rue de St Priest 34095 Montpellier.

<https://www.lirmm.fr/icar/>

marc.chaumont@irisa.fr, gerard.subsol@lirmm.fr, ediasribeiro@lirmm.fr

Note : Le lieu du stage peut avoir lieu à Vannes ou à Montpellier en fonction de la préférence de l'étudiant.

Dans le cadre de la thèse d'Eugênio Dias Ribeiro Neto "Détection et suivi des interactions individuelles entre animaux basés sur l'IA à partir de pièges photographiques" et du projet SEAdogSEA [2], le LIRMM (Gérard Subsol), l'IRISA (Marc Chaumont) et le CIRAD (Michel De Garine-Witchatitsky, Hélène Guis) travaillent ensemble sur la réidentification et le suivi de chiens à partir des vidéos capturés par des pièges photographiques en environnement non-contrôlé, à l'aide d'un pipeline automatique basé sur l'IA [1]. L'objectif est de proposer un pipeline automatique multi-étapes pour identifier les individus-chiens à partir de l'ensemble des vidéos issues d'une campagne de capture.

À partir de milliers de vidéos, le pipeline permet :

1. D'extraire les imagerie de chiens,
2. De représenter chaque imagerie par un vecteur caractéristique (issu d'un réseau de deep learning entraîné pour la réidentification de chiens),
3. De regrouper les vecteurs par un algorithme de clustering.

Le pipeline permet de regrouper les imagerie d'un même individu-chien. La figure 1 illustre un chien identifié par le pipeline où chaque imagerie provient d'une vidéo différente.

Cluster 5

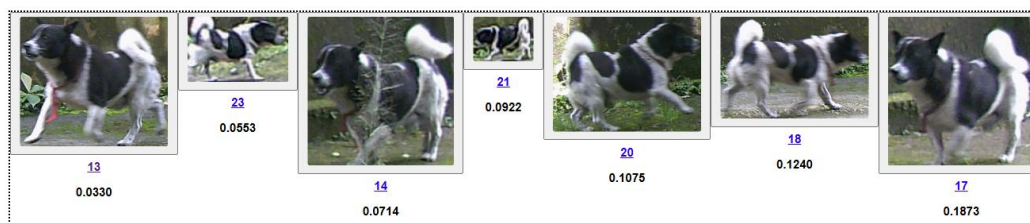


Fig. 1 : Exemple d'un même chien identifié (et ré-identifié) par notre pipeline.

Le résultat obtenu par le pipeline permettra de calculer des paramètres démographiques, en particulier avec la méthode de capture-marquage-recapture (CMR) [6], utilisée en écologie pour estimer la taille d'une population animale.

Le stage portera principalement sur l'étape 3 du pipeline, et plus exactement sur l'étude et la comparaison des algorithmes de clustering afin d'identifier une méthode efficace pour distinguer les individus-chiens dans notre contexte. Nous disposons de deux bases annotées (B1 et B2), correspondant à un même site (avec des individus en commun). Ces deux bases sont des sous-ensembles d'une base plus large. Les objectifs du stage sont les suivants :

- 1- Étudier les métriques d'évaluations et déterminer les plus pertinentes pour afin d'évaluer le résultat de clustering dans notre contexte, à la fois sans vérité terrain (nombre de clusters inconnu) et avec vérité terrain (nombre de clusters connu),
- 2- Répertorier les différentes formules pour le calcul d'une population dans le cadre de la méthode CMR,
- 3- Étudier les différents algorithmes de clustering,
- 4- Proposer un benchmark des différentes méthodes pour nos bases annotées. Utiliser les métriques (objectifs 1 et 2) pour évaluer les solutions et les paramètres optimaux de chaque méthode.

Ce stage vise à répondre de nombreuses questions dont celles-ci :

- 1- Comment choisir le nombre de clusters K , ou fixer des bornes inférieure et supérieure sur ce nombre ?
- 2- Entre les métriques qui ne nécessitent pas la vérité terrain (sans K défini), lesquelles sont les plus pertinentes (corrélées) à celles utilisant une vérité terrain ?
- 3- Comment le choix de K influence-t-il l'estimation de la taille de population (CMR) ? Existe-il une relation linéaire ? Comment une erreur sur K se propage sur l'estimation de la taille de la population ?
- 4- Peut-on exploiter les annotations d'un sous-ensemble de la base pour améliorer l'estimation sur la base complète ?
- 5- Existe-il une différence de performance lorsque nous analysons les bases séparément (B1, B2) ou leur union (B1 $\cup$ B2) ?
- 6- Peut-on intégrer des métriques dans un algorithme de deep learning et obtenir une prédiction de population plus précise ?

Compétences / Qualités :

- M2 ou M1 en informatique ;
- Bonnes notions de machine learning ;
- Bonnes notions des outils tels que Python, Numpy, Pandas, Scikit-Learn, Pytorch ;
- Anglais scientifique ;
- Autonome ;
- Organisé ;
- Intérêt pour la recherche scientifique et pour les applications d'IA appliqués à l'écologie.

Conditions de stage :

Durée : 4 à 6 mois avec un début en février ou mars 2025.

Le stage se déroulera à l'IRISA à Vannes ou au LIRMM à Montpellier.

Rémunération : Gratification de l'ordre de 650 euros par mois.

RÉFÉRENCES :

1. Ré-identification de chiens à partir de vidéos en environnement non-contrôlé, Cyril Barrelet, Eugenio Dias Ribeiro Neto, Marc Chaumont, Gérard Subsol, Etienne Loire, Michel De Garine-Witchatitsky, XXIXème Colloque Francophone de Traitement du Signal et des Images, GRETSI'2023, Grenoble, France, 28 août – 01 septembre 2023, 4 pages. URL: https://www.lirmm.fr/~chaumont/publications/GRETSI2023_BARRELET_DIAS-RIBERO-NETO_SUBSOL_LOIRE_DE-GARINE-WITTCCHATITSKY_ReId_Chien.pdf
2. de Garine-Witchatitsky et al, 2024. One health in practice: Seadogsea! a socio-ecological approach for the study and management of zoonotic diseases associated with free-roaming domestic dogs in south-east asia, in: 3rd Joint AITVM–STVM International Conference, Montpellier, France. URL: <https://www.lirmm.fr/~chaumont/publications/AITVM%E2%80%93STVM-2024-SEAdogSEA-DogZoonoticDiseasesSoutheastAsia-POSTER.pdf>
3. Willians et al, 2025. Animal re-identification in video through track clustering. Pattern Analysis and Applications. URL: <https://link.springer.com/article/10.1007/s10044-025-01497-8>
4. Miele et al, 2021. Revisiting animal photo-identification using deep metric learning and network analysis. Methods in Ecology and Evolution. URL: <https://www.lirmm.fr/~chaumont/publications/AITVM%E2%80%93STVM-2024-SEAdogSEA-DogZoonoticDiseasesSoutheastAsia-POSTER.pdf>
5. Sundaresan et al, 2025. Adapting the Re-ID Challenge for Static Sensors. IET Computer Vision. URL: <https://arxiv.org/abs/2412.00290>
6. La méthode de capture-marquage-recapture. URL: <https://eduscol.education.fr/document/25501/download>