Apprentissage actif pour l'annotation d'images aériennes appliqué aux suivis environnementaux

Mathieu Laroze[2 3] Romain Dambreville[1 3] Chloé Friguet[1] Ewa Kijak[2] Sébastien Lefèvre[1]

[1] Univ. Bretagne Sud, UMR 6074, IRISA, 56000 Vannes, France
[2] Univ. Rennes 1, UMR 6074, IRISA, 35000 Rennes, France
[3] WIPSEA, 35690 Acigné, France.

1 Introduction

Les nouvelles résolutions apportées par les capteurs aéroportés (drones, avions) ont facilité les suivis environnementaux pour la biodiversité à une nouvelle échelle d'objets auparavant trop réduite pour une interprétation humaine. Cependant les images ainsi acquises représentent des données de grande dimension souvent encore analysées visuellement par des experts du domaine. Cette étude manuelle des images est donc coûteuse en temps et ne permet pas de suivi régulier à grande échelle. Ce type d'image a récemment suscité l'intérêt de la communauté en vision par ordinateur où la détection automatique d'objets est un thème de recherche actuel apportant une solution à ces nouvelles applications [1]. Si certaines de ces applications environnementales ont déjà bénéficié de ces développements, la disponibilité de bases de données annotées apparaît de nos jours comme l'un des verrous pour de nouveaux projets, l'annotation étant une tâche longue et coûteuse à mettre en place.

L'objectif de la méthode présentée ici est d'assister l'annotation d'images aériennes dans le contexte de l'étude d'objets d'intérêt pour un suivi environnemental en introduisant un procédé d'apprentissage actif. L'apprentissage actif est un procédé itératif qui demande à un utilisateur externe d'étiqueter des instances inconnues, via des requêtes. La sélection de ces instances est basée sur leur capacité d'apport d'information et leur incertitude vis-à-vis du modèle de classification [3, 5].

A notre connaissance, il n'existe pas de requête sur image entière dans un cas de détection multi-objets. Dans notre méthode, la requête est contrainte par le besoin que ces instances appartiennent à un même groupe, dans notre cas l'image. En proposant l'image entière à annoter, nous cherchons à faciliter la tâche de l'annotateur en gardant le contexte de l'instance proposée pour l'annotation. Ainsi, l'objectif est de réduire le nombre d'interactions réalisées par l'annotateur sur l'ensemble des images lors de son annotation.

Le système par apprentissage actif est présenté dans la section 2, puis une évaluation de ses performances est menée sur un cas d'images réelles de suivi environnemental dans la section 3. Enfin, les perspectives envisagées sont présentées dans la section 4.

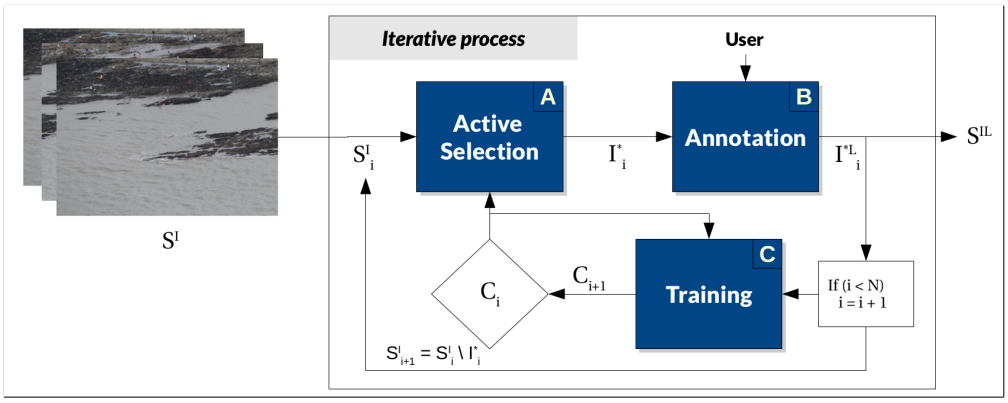

FIGURE 1 – Schéma global

2 Méthode d'apprentissage actif pour l'annotation d'images aériennes

Un ensemble de $K+1$ images forme l'ensemble des images à annoter. A l'aide d'une fenêtre glissante, n régions d'intérêt sont extraites de chaque images. Les instances formant cette image sont ensuite calculées à l'aide d'un Histogramme de Gradients Orientés (HOG). Nous considérons le problème de détection d'objet dans notre cas comme binaire : pour une instance donnée, extraite d'une image, le classifieur doit déterminer si elle contient un objet d'intérêt (instance positive) ou non (instance négative). Le classifieur utilisé est un SVM avec un noyau RBF. Le classifieur SVM a notamment été étudié dans le cas de l'apprentissage actif [4] et présente un intérêt quant à la sélection d'instances incertaines pour la réduction de l'espace des hypothèses afin de définir sa fonction de décision. Le système global est présenté sur la Figure 1 et détaillé ci-dessous.

Sélection Active (Bloc A – Figure 1). La sélection des instances est contrainte par la sélection de l'image. Nous définissons donc un score par image basé sur l'intérêt de cette image pour le ré-entraînement. Pour une itération i donnée, la prédiction du classifieur $\mathcal{C}_i$ sur l'ensemble non étiqueté nous permet de créer deux groupes : (1) les instances "certaines", prédites comme positives ou négatives avec une probabilité supérieure à $0,8$ et (2) les instances "incertaines", prédites avec une probabilité entre $[0,4; 0,6]$.

Le score d'une image à l'itération i est calculé à partir des proportions de chacune de ces catégories relativement à l'ensemble des instances appartenant à l'image. Soit P_i (resp. U_i) le ratio d'instances certaines positives (resp. incertaines). Le score de l'image k, noté s_k, est alors défini par la moyenne harmonique de la densité des instances certaines et incertaines de l'image :

$$s_k = \frac{P_i \times U_i}{P_i + U_i} \tag{1}$$

Pour l'itération i, l'image soumise à annotation I_i^* vérifie alors : $I_i^* = \text{argmax}_k\{s_k\}$. Cette image est présentée à l'annotateur pour obtenir une étiquette sur chacune de ses instances puis est retirée de l'ensemble S_{i+1}^I.

Interaction avec l'annotateur (Bloc B – Figure 1). L'annotateur est supposé parfait dans notre simulation. L'annotateur ne peut donc pas introduire d'erreur dans le système en donnant une mauvaise étiquette. Cela correspond à notre cas d'usage où l'annotateur est un expert (biologiste, écologue). A chaque itération, l'annotateur peut (1) corriger une fausse détection (=identifier un faux positif) ou (2) ajouter une détection manquée (=identifier un faux négatif). Les vrais positifs et vrais négatifs sont validés implicitement par les actions précédentes.

Entraînement du classifieur (Bloc C – Figure 1). Une fois l'annotation réalisée, l'ensemble d'entraînement étiqueté $\mathcal{L}_i$ est créé avec les nouvelles instances étiquetées. Le classifieur $\mathcal{C}_i$ est entraîné sur un sous-ensemble de $\mathcal{L}_i$ afin d'éviter un déséquilibre des classes trop important. Nous avons élaboré et comparé trois stratégies différentes pour définir ce sous-ensemble. Il est composé de : (UC) instances incertaines ; (UC+C) instances avec un ratio équilibré entre les instances certaines et incertaines ; ou bien (UC+C+EK) instances avec un ratio équilibré entre les instances certaines et incertaines ainsi que toutes les instances positives contenues dans l'image après correction de l'utilisateur (*Extra-Knowledge*).

Initialisation et arrêt du système. Lors de l'initialisation, $\mathcal{I}_0$ est étiquetée : l'ensemble des instances positives de l'image sont identifiées, les autres sont considérées comme négatives. Ainsi, l'étiquette y est connue pour chaque instance de l'image $\mathcal{I}_0$. Le classifieur $\mathcal{C}$ est entraîné sur l'ensemble $\{x_\ell^0; y_\ell^0\}_{\ell=1\cdots n}$. Les sélections d'image sont ensuite réalisées itérativement jusqu'à épuisement des images non annotées dans S^I.

3 Expériences sur données réelles

Les données. Les expériences sont réalisées sur un ensemble de 7 images de pêcheurs à pied sur la côte atlantique dans la zone du Parc Naturel Régional du Golfe du Morbihan (Bretagne, France) [1]. Les données ont été obtenues dans le cadre d'une étude de suivi de pression de l'activité de pêche à pied sur son environnement [2]. Une vérité terrain sur ces images a été réalisée manuellement et représente 651 pêcheurs dans l'ensemble des images. Afin de réduire la taille des images, les 7 images sont divisées en 28 sous-images de taille 1750×1167 pixels. Les instances contenues dans une même image constituent donc 28 ensembles initialement non étiquetés. Ces 28 images contiennent ainsi en moyenne 25 objets (pêcheurs). Pour les expériences suivantes, les 28 images sont réparties en deux ensembles d'entraînement (23 images) et de test (5 images) afin d'évaluer les performances du classifieur tout au long du processus. Quatre combinaisons entraînement/test différentes sont créées tout en respectant la distribution globale des objets par image. Dans chacune de ces combinaisons, en moyenne 114 objets sont présents dans le jeu d'apprentissage, soit $17,5\%$ des objets au total. L'image utilisée pour initialiser le modèle est impactant pour la suite des itérations. Pour chacune des combinaisons entraînement/test, les résultats présentés ci-dessous sont la moyenne des résultats obtenus sur l'ensemble des initialisations possibles.

Évaluation du nombre d'interactions réalisées par l'annotateur. Pour l'image donnée $\mathcal{I}_i^*$ soumise à l'annotateur lors de l'itération i, le nombre d'interactions est défini par la somme des faux positifs et des faux négatifs détectés par l'annotateur

1. Nous remercions le Parc Naturel Regional du Golfe du Morbihan et l'Agence Française pour la Biodiversité pour le jeu de données.

	Nb d'itérations où NInter>NObj		Gain d'interaction	
	moy. (e-t)	min/max	moy (e-t)	min/max
UC	$15,9\ (6,1)$	$4/21$	$20,5\ (26,5)$	$0/87$
UC+C	$13,9\ (5,5)$	$2/21$	$41,8\ (38,9)$	$0/\mathbf{135}$
UC+C+EK	$\mathbf{4}\ (0,9)$	$\mathbf{2/6}$	$\mathbf{77,5}\ (16,6)$	$\mathbf{43}/102$

TABLE 1 – Évaluation des différentes stratégies de ré-entraînement en gain d'interactions

dans $\mathcal{I}_i^*$. Le gain en interactions est donc défini par la différence entre le nombre d'objets dans $\mathcal{I}_i^*$ et le nombre d'interactions. Ce gain quantifie l'apport du système d'un point de vue utilisateur.
Nous introduisons également dans nos expériences une règle simulant l'utilisateur dans le cas où pour une image à annoter, le nombre d'interactions nécessaires est supérieur au nombre d'objets présents dans l'image (cf. NInter>NObj dans la Table 1). Dans ce cas, le nombre d'interactions est considéré comme égal au nombre d'objets dans l'image. Cela simule ainsi la préférence de l'utilisateur d'annoter depuis zéro l'image plutôt que de réaliser un nombre d'interactions non optimales.
La Table 1 reporte le nombre moyen d'itérations pour lesquelles cette règle est appliquée. Ces étapes interviennent lors des premières itérations alors que le système n'est pas robuste et que de nombreuses fausses détections sont affichées pour l'annotateur. Le nombre d'itérations avec application de cette règle indique le nombre d'itérations nécessaires au modèle pour proposer des détections cohérentes à l'annotateur. La stratégie UC+C+EK nécessite en moyenne 4 itérations sur 23 pour arriver à une prédiction stable, là où les autres stratégies ont besoin de 16 (UC) et 14 (UC+C) itérations en moyenne.
Enfin, pour évaluer les performances globales de la méthode proposée, le gain d'interactions en moyenne (moy.) ainsi que leur écart-type (e-t) sont présentés dans la Table 1. Le gain maximal est obtenu avec la stratégie UC+C avec un gain de 135, cependant l'écart-type et les gains minimaux obtenus lors des plus mauvaises initialisations montrent que la stratégie UC+C+EK est plus robuste avec 43 interactions gagnées au minimum. En moyenne, 537 objets sont présents dans le jeu d'entraînement, ainsi un gain de $77,5$ interactions correspond à $14,3\%$ d'interactions gagnées sur le procédé d'annotation.

4 Conclusion

Nous avons introduit un système actif pour l'annotation d'un ensemble d'instances sur des images, réduisant le coût d'annotation lors de la création d'une vérité terrain. Habituellement, les procédés d'apprentissage actif font une sélection sur l'ensemble des données. Dans notre approche, nous avons appliqué une contrainte de sélection des instances à l'appartenance à une même image afin de créer une requête par image. Cette requête permet à l'annotateur de garder le contexte global d'annotation. La sélection de l'image à annoter est basée sur la maximisation d'un score calculé à partir de la répartition des instances en deux catégories, certaines et incertaines, à l'aide d'un classifieur. Cette stratégie a été évaluée en termes de nombre d'interactions gagnées lors de l'annotation active. En comparant différentes stratégies de ré-entraînement à l'aide de cette mesure, nous avons pu évaluer l'impact de l'*Extra Knowledge* apporté par l'utilisateur dans le procédé actif. Nous avons également identifié que l'initialisation était une étape importante pour le procédé. Bien qu'en dehors du sujet de cet article, l'étude de cette étape permettrait un gain en efficacité et en robustesse de notre système.

Références

[1] Luis F Gonzalez, Glen A Montes, Eduard Puig, Sandra Johnson, Kerrie Mengersen, and Kevin J Gaston. Unmanned aerial vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation. *Sensors*, 16(1) :97, 2016.

[2] Mathieu Laroze, Luc Courtrai, and Sébastien Lefèvre. Human detection from aerial imagery for automatic counting of shellfish gatherers. In *International Conference on Computer Vision Theory and Applications (VISAPP)*, Rome, Italy, 2016.

[3] Burr Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.

[4] Simon Tong and Daphne Koller. Support vector machine active learning with applications to text classification. *J. Mach. Learn. Res.*, 2 :45–66, March 2002.

[5] Devis Tuia, Michele Volpi, Loris Copa, Mikhail Kanevski, and Jordi Munoz-Mari. A survey of active learning algorithms for supervised remote sensing image classification. *IEEE Journal of Selected Topics in Signal Processing*, 5(3), 2011.