



HAL
open science

Active Learning to Assist Annotation of Aerial Images in Environmental Surveys

Mathieu Laroze, Romain Dambreville, Chloé Friguet, Ewa Kijak, Sébastien
Lefèvre

► **To cite this version:**

Mathieu Laroze, Romain Dambreville, Chloé Friguet, Ewa Kijak, Sébastien Lefèvre. Active Learning to Assist Annotation of Aerial Images in Environmental Surveys. CBMI 2018 - 16th International Conference on Content-Based Multimedia Indexing, Sep 2018, La Rochelle, France. pp.1-6. hal-01845487

HAL Id: hal-01845487

<https://hal.science/hal-01845487v1>

Submitted on 20 Jul 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Active Learning to Assist Annotation of Aerial Images in Environmental Surveys

Mathieu Laroze¹, Romain Dambreville², Chloé Friguets³, Ewa Kijak⁴, and Sébastien Lefèvre⁵

IRISA, CNRS 6074, France

^{1,4} Univ. Rennes, ^{1,2} WIPSEA, ^{2,3,5} Univ. Bretagne-Sud, Vannes

{mathieu.laroze, romain.dambreville, chloe.friguet, ewa.kijak, sebastien.lefevre}@irisa.fr

Abstract—Nowadays, remote sensing technologies greatly ease environmental assessment using aerial images. Such data are most often analyzed by a manual operator, leading to costly and non scalable solutions. In the fields of both machine learning and image processing, many algorithms have been developed to fasten and automate this complex task. Their main common assumption is the need to have prior ground truth available. However, for field experts or engineers, manually labeling the objects requires a time-consuming and tedious process. Restating the labeling issue as a binary classification one, we propose a method to assist the costly annotation task by introducing an active learning process, considering a query-by-group strategy. Assuming that a comprehensive context may be required to assist the annotator with the labeling task of a single instance, the labels of all the instances of an image are indeed queried. A score based on instances distribution is defined to rank the images for annotation and an appropriate retraining step is derived to simultaneously reduce the interaction cost and improve the classifier performances at each iteration. A numerical study on real images is conducted to assess the algorithm performances. It highlights promising results regarding the classification rate along with the chosen re-training strategy and the number of interactions with the user.

Index Terms—Active learning, object detection, aerial images, data annotation, human-in-the-loop

I. INTRODUCTION

Nowadays, remote sensing technologies greatly ease environmental assessment over large study areas using aerial images, *e.g.* for monitoring and counting animals or ships. As drone prices are progressively decreasing, the acquisition of aerial high-resolution images targeting very different and specific objects of interest is now accessible to almost anyone. However, the resulting databases are most often analyzed by experts from each domain manually, leading to costly and non scalable processes. This challenging task recently found an echo in the computer vision community where automatic object detection algorithms appear as a providential solution [1]. If several environmental applications already successfully benefited from recent developments in computation capabilities [2]–[4], the required data annotation now appears as one of the main bottlenecks: it is a very tedious and time-consuming procedure. The aim of the method introduced hereafter is to assist the annotation process in aerial images with an active learning process considering a novel strategy.

We acknowledge the “Parc Naturel Régional du Golfe du Morbihan” and the “Agence Française pour la Biodiversité” for providing the dataset.

A. Motivation

The proposed strategy is motivated by an application to environmental survey where aerial images are used to count objects of interest. Several tens of images are available after an aerial campaign and are manually processed to identify and count these objects. In this field, an annotator can frequently need to inspect a wider area to determine whether an object is present or not. Our method is precisely derived from this observation and we adapt the usual active learning process to fit this assessment: a selected full image is proposed to the user for annotation, through bounding boxes correction or validation. It is assumed that the annotation task is therefore more efficient if the annotator has a good understanding of the context. We do not pretend that this query-by-group strategy should be used for any active annotation task but we believe it is pertinent to any application where a comprehensive context may be required to assist the annotator with the labeling task of a single instance. Besides, we focus on cases where there are only few and scattered positive instances per group. This imbalanced setting leads to the need for specific sampling for instance selection and active retraining strategy. Examples of the diversity of images and complexity of annotation of our specific application are shown in Fig. 2, where the high variability of several parameters from one image to another complicates the object detection task: acquisition parameters (acquisition height or angle), weather condition, but also the relatively small dimension of objects regarding the image size and the lack of prior hypothesis of localization on the image.

Hence in the following, the selection of the queried instances is constrained by requiring that they belong to a group. This strategy necessitates to (1) define a global score to rate the relevance of each group to be the next one to query, (2) define a performing training subset selection within the queried group and (3) define a metric to evaluate the interaction with the annotator. The noteworthy objective is to reduce the number of human interactions on the overall process, starting from a first annotated image to a fully annotated image dataset, rather than reaching the maximum final accuracy.

B. Related work

In active learning, the goal is to identify and select for query, some instances that are more useful than others for classification performances in an iterative process: well chosen, fewer representative instances are needed to achieve

similar performances. In classical active learning approaches, at each iteration, the labeling of one or several representative instances is queried one after another and the classifier is re-trained on the enlarged data including the new labeled instances. Queried instances are selected on criteria based on their informativeness or their uncertainty [5], [6]. Previous works combining active learning and object detection have focused on selecting queries at training time to reduce the number of instances needed to reach the performances of fully supervised learning. For example, [7]–[9] focus on designing the classifier during annotation time in an active learning setup, [10] considers a designed set based on Adaboost classifier, [11] an interactive way without active query. Nevertheless, these strategies aim to query the user with several patches to label and face the difficulty to annotate uncertain patch queries [10]. On the contrary, in our strategy, we aim to query the annotator with a full image (and not a set of patches coming from different images) to facilitate the annotator task. We thus should adapt our query criterion to query a whole image. Querying a full image to find a precise localization of an object has been studied in [7], [12]. In [7], the training images are crawled from the web and an active learning process selects the most uncertain images for annotation. [12] propose a method for training object detectors which only requires annotators to verify bounding-boxes produced automatically by the learning algorithm, in order to reduce the cost of annotation. However there is no active learning process involved. Both works consider images with objects of significant size compared to our case. Our main challenge is not so much the precise localization of objects than their detection.

The remainder of the paper is organized as follows. Our proposed active learning framework is introduced in Section II. Section III reports numerical experiments conducted on a real image dataset to assess the performances of the method. Finally, we conclude and provide future research directions in Section IV.

II. METHOD

A. Global workflow and notations

Our goal is to assist the annotation process for an object detection task in aerial images by introducing an iterative active process which global workflow is displayed in Fig. 1. The images collected through drone or plane flights over a specific area constitute a set of images S to annotate of size $K + 1$ (for convenience, as one image will be used for initialization). At each iteration i , the system queries an image to annotate. S_i denotes the image set with remaining non annotated images. The process ends when S_i is empty and a set of fully labeled images S^L has been created.

Each image, as in multiple instance learning, can be seen as a bag of instances extracted through a traditional object detection pipeline. An instance is a region of interest from an image, referred to as a patch, described by a feature vector $x_\ell \in \mathbb{R}^d$. An image I^k can be denoted in the feature space as $I^k = \{x_1^k, \dots, x_{p_k}^k\}$ with $k \in \{1, \dots, K + 1\}$ and p_k being

the number of instances extracted in image k . In our case, $\forall k, p_k = p$ is a constant determined by the patch extraction method used. The object detection problem on an image I^k can thus be restated as a binary classification for each instance x_i^k : the classifier C should determine its label $y_i^k \in \{0, 1\}$, depending on whether it contains an object of interest or not, 1 being the positive class. We note the number of total positive instances in S as N^+ . With this model, the object detection problem is strongly imbalanced between the positive class and the negative class because of the low distribution of objects of interest in comparison with the background. The main steps of the procedure are now described hereafter.

B. Initialization.

At the beginning ($i = 0$), one image I_0 is taken out of S to be fully labeled. The classifier C_1 is then trained on a balanced sample of labeled instances containing all the positives instances and negative instances chosen randomly. The active phase starts at $i = 1$ with the set S_1 containing the K remaining images.

C. Active phase: Score computation and selection of the image to be queried (Block A – Fig. 1).

At each iteration i , all instances of S_i are classified by C_i and grouped in two categories:

- 1) **Certain instances (positives and negatives)**: predicted respectively to their class with a probability greater than a fixed threshold t_c ,
- 2) **Uncertain instances**: predicted (either positive or negative) with a probability lying in an interval with fixed half-length $t_{unc} : [0.5 - t_{unc}, 0.5 + t_{unc}]$.

We propose to create a selective sampling query taking into account two main criteria: the number of certain positives P_i , and the number of uncertain instances U_i . As in our framework the instance selection is constrained by a group membership, a score taking into account a trade-off between these two criteria at the group level has to be defined to rate the relevance of each image to be the next one to query. The image score s_k of an image $I^k \in S_i$ is thus defined as a harmonic mean value combining the two criteria:

$$s_k = \frac{P_i \times U_i}{P_i + U_i} \quad (1)$$

This score allows to keep a trade-off between the densities of both certain data (that helps to improve a good prediction rate) and uncertain data (according to the well-known uncertainty sampling strategy). Its maximum is reached for images with a balanced setting between both. The image score s_k is computed for each image $I^k \in S_i$. The selected image I_i^* is the one with the largest score s_k and is submitted for annotation to the oracle.

D. Annotation: oracle/user interaction (Block B – Fig. 1).

The oracle is assumed to be a perfect oracle, the image being checked for annotation by a user who is expert in its field. The image presented to the user at each iteration contains

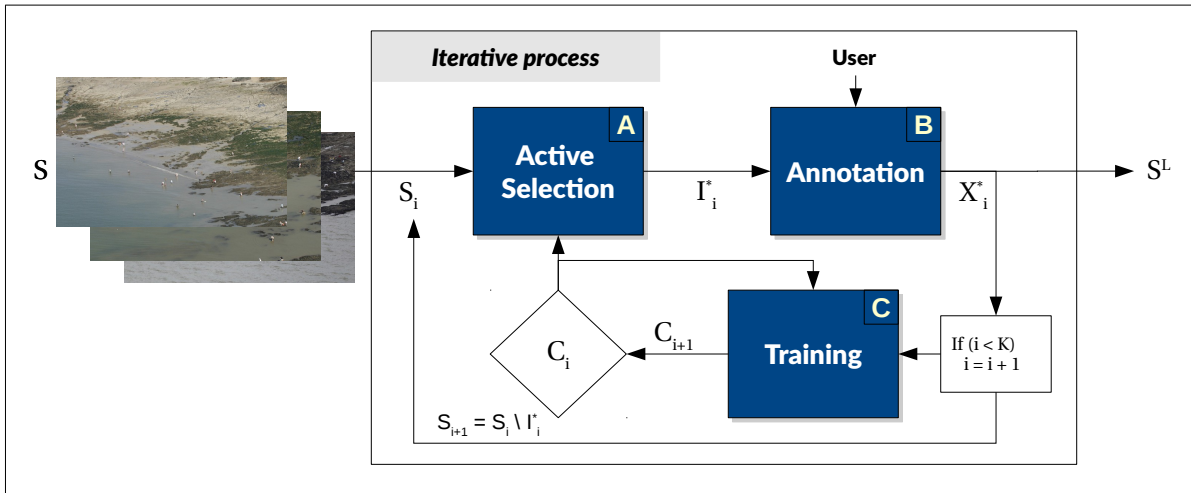


Fig. 1. Global workflow. Our system takes in input a set of images S . Once features have been extracted from the images, an iterative annotation process begins. At each iteration an image I^* is selected with the help of the classification made by C , and is given to an oracle for annotation. A subset $X^* \subset I^*$ is then used to retrain a classifier C . The iteration stops when the set S has been fully annotated by the oracle.

bounding boxes of certain positives instances predicted by the classifier. User interactions with I_i^* for the label annotation process can be either (1) to correct the false positives or (2) to add the missing positives (i.e. correct the false negatives). True positives and true negatives are validated implicitly.

At the i^{th} iteration, the number of interactions performed by the user is denoted \mathcal{A}_i . The evolution of the accumulated values of \mathcal{A}_i over the iterations of the algorithm is considered in the following as an evaluation criterion of the method: the method succeeds if, at the end of the algorithm, \mathcal{A}_K is less than the number of interactions needed to annotate the full dataset (i.e. N^+), the difference between the two values being the gain in interaction.

E. Classifier training (Block C – Fig. 1).

I^* being fully annotated, the label y_ℓ of each instance x_ℓ are known. As mentioned previously, the studied data are strongly imbalanced due to the distribution of the objects in a binary classification scheme and the instances extraction. Therefore retraining C on all the instances in I^* is not a viable solution. A subset of instances $X^* \subset I^*$ is selected and added to the previous training set. Then the classifier C is retrain from scratch on this newly concatenated training set. Selecting X^* is not trivial. We introduce three different strategies to select this subset:

- 1) **Uncertain (UC)** considers a balanced subset of positive and negative uncertain instances.
- 2) **Uncertain + Certain (UC+C)** is composed of a balanced subset of uncertain and certain instances;
- 3) **Uncertain + Certain + Extra-Knowledge (UC+C+EK)** is the previous subset (UC+C), enriched with Extra-Knowledge which refers to all corrected instances by the oracle, that would not have been selected with the previous strategies, i.e. false positives, false negatives and positives with medium confidence.

For each strategy, the balance of the set is done by having equal amount of positive and negative instances. All certain and uncertain positives instances are selected. Negative certain instances are sorted increasingly according to their probability and those with the highest probabilities are selected. Negative uncertain instances are sorted and selected regarding their closeness to the uncertain boundary probability (0.5). The number of selected negatives in each considered subset is set equally to the number of positives instances in this subset. The first two strategies mimic classical active learning selection strategies based on the classifier confidence. We consider these strategies in order to evaluate the gain of adding the different subsets to the training set, under the query-by-group constraint. Whereas the third one makes full use of the information provided by all the corrections on the selected image by the user.

F. End of the procedure.

The iterative process (computing image score, user interaction with the selected image), is performed *de novo* until all images are annotated.

III. EXPERIMENTAL RESULTS

A. Experiment settings

The experiments are carried out on a set of 7 images of humans gathering shellfish on the seashore in the Natural Park of Morbihan (South Brittany, France) during spring tides [4]. The aim is to evaluate the number of people on the seashore in this period of high attendance and deduce the pressure of their activity on the environment. The images were taken from a plane with different angles of acquisition inducing a high variability in the objects size and orientation (see Fig. 2).

651 humans have been manually annotated on this dataset to create the ground truth. This prior manual annotation is required to simulate the oracle and evaluate the classification



Fig. 2. Three images of our experimental dataset showing the diversity of the images in illumination, distribution of objects, angle of acquisition, which are main challenges of our dataset. These images respectively contain 15, 3, 26 shellfish gatherers.

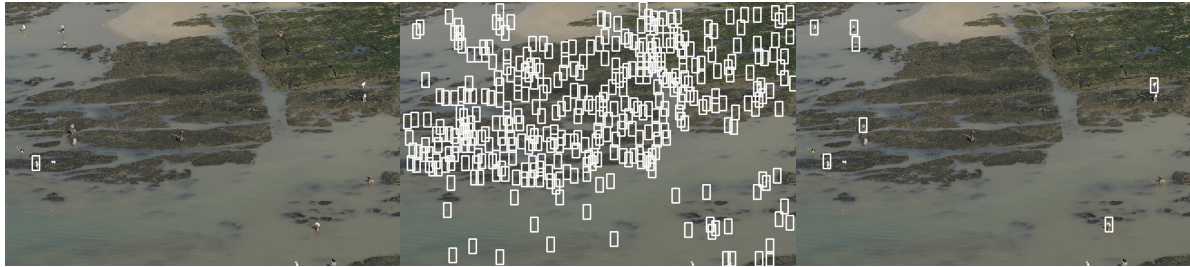


Fig. 3. We show evolution of the classifier on a test image through the iterative process at iteration 1, 2 and 6. The second image highlights the interest of a reset step for the annotator in case this image would have been selected for annotation.

performances. All labels are hidden and only disclosed when required at annotation step. According to [13], the average annotator error highly increases when counting more than 50 objects per image. We thus split each image in four parts, in order to reduce the number of humans per image for the annotator.

Our final set S , after splitting images, is composed of 28 sub-images of size 1750×1167 pixels. This set contains in average 25 objects per image. The set of 28 images is split into a training set ($K + 1 = 23$ images) and a test set (5 images) to assess the classification performances all along the iterative process. Four different splits in train/test are performed. In the four test sets, there are in average 114 positive instances. We take care that the distribution of the number of objects per image in each test set is consistent with the one of the corresponding training set. A summary of the data statistics are shown in Tab. I, highlighting the dataset imbalance between positive and negative examples.

Feature extraction. Humans size is varying from 20 to 60 pixels height, depending on the perspective. Instances (patches) are extracted with a sliding window of size 64×32 with a stride of 8. HOG features are computed for each instance with parameters set to a cell size of 8×8 and a block size of 2×2 to fit window sizes. The dimension of the output feature x_ℓ is $d = 1,152$ for each instance. This sliding window process gives $p = 29,670$ instances per image. Note that feature extraction is beyond the scope of the present paper. Even if the sliding-window algorithm is computationally expensive, it allows an exhaustive search for object positions.

TABLE I
DATASET STATISTICS (AVERAGE OVER THE 4 TRAIN/TEST SPLITS)

	Training set	Test set
# images	23	5
# total instances	682,410	148,350
# positive instances	537 (N^+)	114

Having no prior hypothesis on object positions, this exhaustive search ensures that all positive instances are present in the instance set for the image selection step, which is important for our oracle simulation.

Classifier. SVM has been studied in the active learning context and has shown good properties for active learning based on uncertainty sampling [14]. The classifier used here is an SVM with a RBF kernel ($C = 1, \gamma = 0.1$) and it is initialized with a first annotated image (seed).

Initialization. The procedure may depend upon the image chosen for initialization, hence averages computed over all possible initializations are reported, for each criterion. We empirically set thresholds for certain and uncertain data based on the Platt probability estimate with $t_c = 0.8$ and $t_{unc} = 0.1$. An instance is thus considered as certain if its probability to its respective label is in $[0.8; 1]$ or uncertain if it is in $[0.4; 0.6]$.

Oracle annotation. Before the image is proposed to the annotator, bounding boxes are drawn after a non maximum suppression algorithm on instances considered as positive by

TABLE II
EVALUATION OF CLASSIFIER AT THE END OF THE ANNOTATION PROCESS -
Average (std) over the 4 test sets.

	F-score	Recall	Precision
Random (weighted) guess	$7.7 \cdot 10^{-4}$	$7.7 \cdot 10^{-4}$	$7.7 \cdot 10^{-4}$
UC	0.22 (0.17)	0.58 (0.30)	0.25 (0.25)
UC+C	0.42 (0.10)	0.76 (0.07)	0.36 (0.13)
UC+C+EK	0.08 (0.04)	0.05 (0.03)	0.50 (0.20)

C with a confidence higher than 0.8. During the annotation phase, to simulate a perfect oracle, we assume a positive label is given to an instance if the overlap coverage with a positive localization of the ground truth is higher than 0.6.

B. Results

Evaluation of the classifier performances w.r.t. the re-training strategies. Each of the re-training strategies described in Section II-E provides a final classifier after the K iterations of the method. These classifiers are first evaluated regarding the F-score, recall and precision of the positive class computed on the test sets. Fig. 3 displays the evolution of the detection at different iterations on one of the image from the test set. Table II reports the detection scores and standard deviations of the three strategies, averaged over the 4 test sets and for the whole annotation process.

For purpose of comparison, performances for a random (weighted) guess are also reported. First, comparing UC and UC+C strategies, we can observe the improvement over F-score, precision and recall when adding certain instances in the retraining set. As we could expect, adding diversity to the uncertain training set, through certain examples, allows better detection performances. Secondly, taking into account the extra-knowledge (UC+C+EK) decreases the average F-score on the test sets with respect to the other strategies. More particularly, the recall sharply decreases as the precision increases. This effect is due to the false positives corrected by the annotator that are added to the training set. The correct labeling of these negative instances that are particularly hard to classify induces mechanically a reduction of the number of false positives. As a side effect, it also increases the number of false negatives, decreasing the recall.

This classifier score study shows that the procedure does not lead to a general classifier with suitable performance for an object detection task. Nevertheless, in the context of an annotation task, a crucial characteristic would rather be the number of interactions with the user during successive iterations and, ultimately, at the end of the process.

Therefore, in the next section, a novel annotation cost metric is introduced. It is based on the difference between the number of user interactions and the number of objects present in the image set, and gives a better interpretation of the retraining strategies for an annotation task.

Evaluation of the number of annotator interactions. For an image queried to the oracle for annotation at iteration i , the number of user interactions \mathcal{A}_i is defined as the sum of false

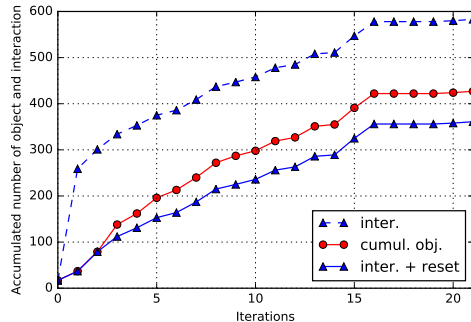


Fig. 4. Evolution of the accumulated number of interactions (*inter*) in comparison with the accumulated number of object (*cumul. obj.*) with a **disadvantageous** initialization for UC+C+EK strategy: two reset steps are needed at iterations 1 and 2. The curve *inter. + reset* illustrates that despite the two reset steps, the remaining iterations have provided a gain of annotation.

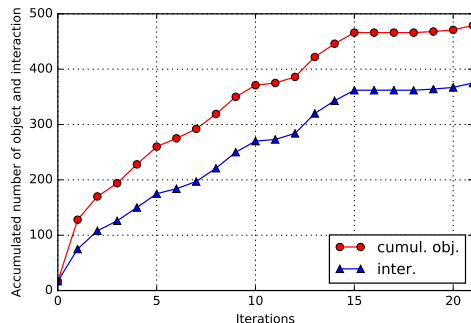


Fig. 5. Evolution of the accumulated number of interactions (*inter*) in comparison with the accumulated number of object (*cumul. obj.*) with an **advantageous** initialization for UC+C+EK strategy: no reset step has been needed by the user. The *cumul. obj.* curve is the upper bound of our method where no gain in annotation is made above this value.

positives and false negatives in I_i^* . The interaction gain is then the difference between \mathcal{A}_i and the actual number of objects (true positive instances) in I_i^* . It quantifies the enhancement of the annotation process from the user point of view.

As the initialization step can have an impact on this score, the proposed procedure is hence overall evaluated on results averaged over each initial image candidate. To better illustrate the wide variety of gain in interaction regarding the image chosen for initialization, two opposite situations (A) and (B) taken out from the set of possible initializations are more deeply investigated.

In the worst case scenario (A), the model predictions are still highly unsatisfying even after the first iterations, leading to poor predictions and many required corrections for the annotator. The number of interactions is larger than the number of objects, which does not fit the objective to ease the annotation process. In practice, in such a case, an annotator would annotate the image from scratch rather than correct many false bounding boxes (see the second image in Fig. 3). We therefore consider that the number of interactions with the annotator

TABLE III
EVALUATION OF RE-TRAINING STRATEGIES WITH RESPECT TO USER
INTERACTION GAIN

	Nb of iterations with reset steps		Interaction gain	
	mean (std)	min/max	mean (std)	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)	0/135
UC+C+EK	4 (0.9)	2/6	77.5 (16.6)	43/102

is the minimum between the number of interactions and the number of objects (inter+reset) in order to simulate a "reset" step by the annotator. Such a disadvantageous initialization, that requires few reset steps before the classifier contributes positively to the annotation process, is illustrated on Fig. 4.

On the other hand, in scenario (B), the initialization is relevant, and the number of interactions remains lower than the total number of objects. The system contributes to the annotation task directly after one image as shown in Fig. 5. This latter scenario is obviously the most preferred.

Table III reports the average number of iterations where this reset step is needed. We can see that UC+C+EK strategy only requires this action mainly during the first 4 iterations in average (out of 23 iterations). Besides, other retraining strategies still resort to it up to the 16th (UC) or 14th (UC+C) iteration in average. This is related to the differences in recall and precision for the three strategies. As explained previously, the proposed method gives prominence to a reduced false positive rate.

To assess the overall performance of the proposed method, the average and standard deviation of interactions gain for each retraining strategy is also reported in Tab. III. Even if the best result is achieved with UC+C strategy, with a maximum of 135 interactions gained, UC+C+EK performs best regarding the overall performance (average: 77.5 interactions gained). Indeed UC+C+EK strategy is more robust to the initialization step, while UC+C could even not perform any gain in worst cases. From a user point of view, the results reported here greatly and conversely differ from performances shown in Tab. II. User metric evaluations demonstrate that UC+C+EK is the strategy fitting our objectives, even though it does not lead to the best F-score in the end. Note also that the number of positive instances in the training sets is 537 in average. Proportionally, an interaction gain of 77.5 results in 14.3% less interactions.

IV. CONCLUSION AND DISCUSSION

We introduced an active learning annotation framework which reduces the annotation cost when creating a ground truth. Usual active learning algorithms perform instance selection from the whole set of input data. In the present work, our method query the best image to annotate to get all the objects in their comprehensive context for the annotator. We defined a score to rank the images and identify the one that should be annotated at each iteration, based on both uncertainty and true positives. The main objective is to reduce the number

of human interactions on the overall process, starting from a first annotated image, rather than reaching the maximum final accuracy. Therefore, the annotation cost is measured through the gain in interactions with respect to a labeling task from scratch. This user interaction metric represents the corrections of the classifier decisions by the annotator. At each iteration, the classifier is retrained according to a specific subset of data. Several strategies have been compared and their performances regarding the interaction gain have been discussed.

Several follow up investigations have been identified to strengthen the proposed architecture. We highlight that initialization is a crucial step to our design, and as such it requires further investigation to gain robustness in the process. Besides, scaling to larger datasets may require the introduction of an incremental classifier and a stopping criterion whose definition in the given context will not be straightforward.

REFERENCES

- [1] L. F. Gonzalez, G. A. Montes, E. Puig, S. Johnson, K. Mengersen, and K. J. Gaston, "Unmanned aerial vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation," *Sensors*, vol. 16, no. 1, pp. 97, 2016.
- [2] F. Maire, L. M. Alvarez, and A. Hodgson, "Automating marine mammal detection in aerial images captured during wildlife surveys: a deep learning approach," in *Australasian Joint Conf. on Artificial Intelligence*, 2015, pp. 379–385.
- [3] D. J. Klein, M. W. McKown, and B. R. Tershy, "Deep learning for large scale biodiversity monitoring," in *Bloomberg Data for Good Exchange Conference*, 2015.
- [4] M. Laroze, L. Courtrai, and S. Lefèvre, "Human detection from aerial imagery for automatic counting of shellfish gatherers," in *Inter. Conf. on Computer Vision Theory and Applications (VISAPP)*, 2016.
- [5] B. Settles, "Active learning literature survey," Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.
- [6] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Munoz-Mari, "A survey of active learning algorithms for supervised remote sensing image classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, 2011.
- [7] S. Vijayanarasimhan and K. Grauman, "Large-scale live active learning: Training object detectors with crawled data and crowds," *International Journal of Computer Vision*, vol. 108, no. 1, pp. 97–114, 2014.
- [8] S. Sivaraman and M. M. Trivedi, "A general active-learning framework for on-road vehicle recognition and tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 267–276, 2010.
- [9] T. Luo, K. Kramer, D. B. Goldgof, L. O. Hall, S. Samson, A. Remsen, and T. Hopkins, "Active learning to recognize multiple types of plankton," *Journal of Machine Learning Research*, vol. 6, pp. 589–613, 2005.
- [10] Y. Abramson and Y. Freund, *Active learning for visual object detection*, Department of Computer Science and Engineering, University of California, San Diego, 2006.
- [11] B. Le Saux, "Interactive Design of Object Classifiers in Remote Sensing," in *Inter. Conf. on Pattern Recognition (ICPR)*, 2014.
- [12] D. P. Papadopoulos, J. R. R. Uijlings, F. Keller, and V. Ferrari, "We don't need no bounding-boxes: Training object class detectors using only human verification," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 854–863.
- [13] A. D. Sarma, A. Jain, A. Nandi, A. Parameswaran, and J. Widom, "Surpassing humans and computers with jellybean: crowd-vision-hybrid counting algorithms," in *AAAI Conf. on Human Computation and Crowdsourcing*, 2015.
- [14] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *Journal of Machine Learning Research*, vol. 2, pp. 45–66, 2002.