Fish4label: Accomplishing an Expert Task without Expert Knowledge *

Jiyin He Jacco van Ossenbruggen Arjen P. de Vries

{j.he, jacco.van.ossenbruggen, arjen.de.vries}@cwi.nl

Centrum Wiskunde en Informatica, Science Park 123
1098XG, Amsterdam, the Netherlands

ABSTRACT

Obtaining large quantities of labeled data of sufficient quality is non-trivial, especially when expert knowledge is required. Experts are scarce and expensive, while laymen lack the necessary knowledge to perform the task. In this demo paper, we present an image labeling tool *Fish4label*. By carefully converting an object recognition task to a visual similarity comparison task, our tool enables laymen to identify fish species in images extracted from video footage taken by underwater cameras, a task that typically requires profound domain knowledge in marine biology.

Keywords

Image labeling, Domain specific knowledge

1. INTRODUCTION

Crowd-sourcing is shown to be an effective strategy [2, 7–9] in creating large scale image annotation datasets. We introduce Fish4label, an image labeling tool that aims to collect ground truth data in order to train classification models that identify fish species on video footage of Taiwanese coral reefs. Different from most of the previous work in crowd-sourcing image labels, where no or little expert knowledge is required for the labeling task, our labeling task requires highly specialized domain knowledge. Similar problems include Foldit [3] and Galaxy zoo [5], where scientific problems are turned into games or less complicated sub-problems, for which "citizens' wisdom" contributes to the scientific solutions.

2. SYSTEM DESCRIPTION

The Fish4label system consists of two labeling interfaces: (i) an expert labeling interface, and (ii) a non-expert gaming interface.

Expert labeling interface

The expert labeling interface is used to assist experts (e.g., marine biologists, coral reef specialists, frequent divers) to create ground truth labels. These labels, while often small in quantity, are assumed to be of high quality. They are considered a gold standard, e.g., for assessing the quality of non-expert labels, and serve as feedback for non-experts so that they may learn and improve.

As pre-processing, the set of images that needs to be labeled is clustered (manually or with automatic clustering algorithms, e.g., [1]). At each screen, images within a cluster are presented to the expert user. See Fig. 1. The labeling process consists of two steps. **Step 1**: The user is asked to enter a species name that would apply to the majority of the images within the cluster. Once the name is entered, all images within the clusters are assigned the same name. **Step 2**: When there exist images that do not belong to the same cluster, i.e., images assigned a wrong species names after step 1, the user can select that "wrong" image, and correct its species name.

In the worst case, the expert users would manually enter the species name for each image when each single image within a cluster belongs to a different species. In the best case, when the cluster is pure, they only need to enter the label once for a cluster. Users can further indicate 1) their confidence in the labels they have entered, and 2) "bad images", i.e., images that contain no fish, half fish or multiple fish. As our images are extracted from video footage using automatic fish detection algorithms, the above "bad images" may occur due to detection errors.

Non-expert gaming interface

The gaming interface is used to collect labels from laymen players. These labels are often large in quantity but noisy, therefore each im-

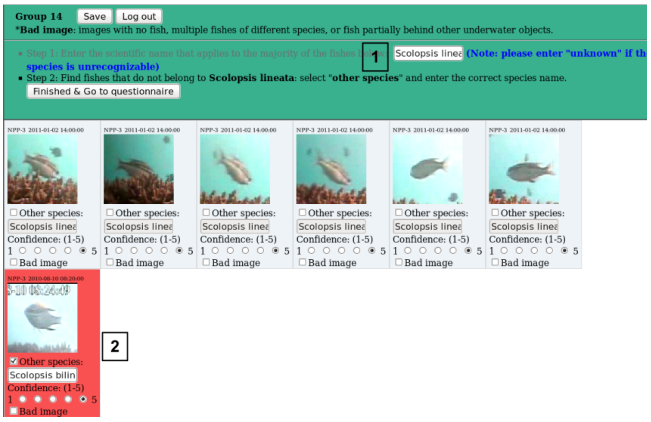
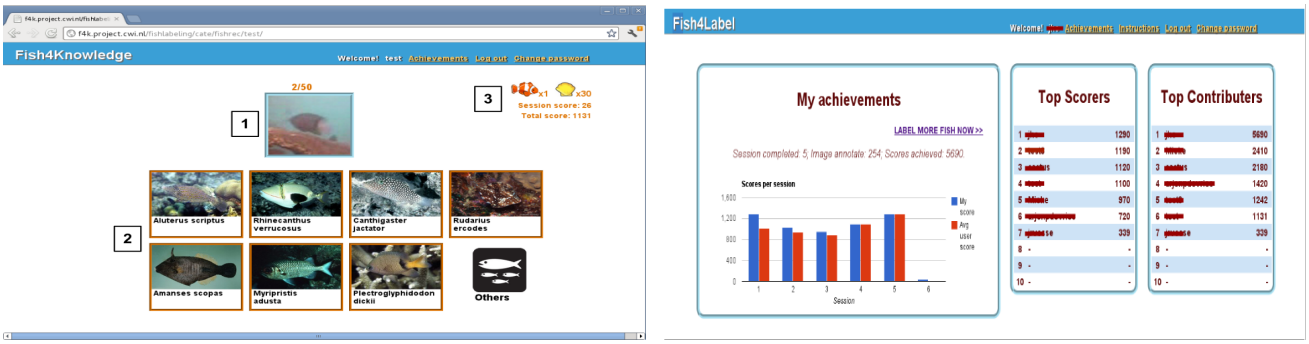


Figure 1: Expert labeling interface. (1): entering species name for the cluster; (2): correct individual species names.

*This paper describes the application demo of the study reported in [6].

OAIR'13, May 22-24, 2013, Lisbon, Portugal.



(a) Labeling interface for non-expert players

(b) Achievment summary

Figure 2: Non-expert labeling game interface. (1): query image; (2): candidate images; (3): feedback scores.

age is labeled multiple times by multiple players. Post-processing is needed to aggregate the obtained labels into a final label.

On the labeling screen, a *query image* (image to be labeled) and a set of *candidate images* (potential labels) are presented to the players. See Fig. 2(a). The players are asked to compare the query image to the candidate images. They select one of the candidates if they believe this candidate image should belong to the same species as the fish in the query image. If none of the candidates is similar enough to the query image, the "others" icon should be selected.

Each image to be labeled is assigned a priority score. In a default setting, if an image has been assigned many labels, then it has less priority than images that have received no or very few labels. For instance, in the beginning all images are assigned a score of 1, and are updated as the labeling process goes on, computed as 1 divided by the number of labels it has received.

Candidate images are prototype images of different species obtained from Fishbase [4]. To avoid overloading players with too many candidates, only 7 candidate images are shown for a query image. Candidate images are selected based on their similarity to the query image. Different similarity measures can be applied; similarity scores are pre-computed and stored in the back-end database when deploying the system.

For each click, a feedback score is shown to the user. We consider two types of feedback scores: (1) expert feedback, computed as the percentage of expert labels that agree with the chosen label, if available; and (2) peer-agreement, computed as the percentage of players' labels that agree with the chosen label. Notice that with peer-agreement, the feedback score of a same {image, label} pair may change as more people play the game. In particular, when there are very few labels the scores are sensitive to erroneous decisions made by individual players, and these scores are likely to introduce bias to players' decisions. To handle this situation, for initial runs, labels generated by automatic methods or manual runs without feedback are used.

Within the system, different methods of computing image priority, candidate similarity, and feedback scores can be easily extended and deployed.

Players can view their "achievements" in the achievement screen (Fig. 2(b)). We show three types of statistics: (i) achievements of the current user, including the number of sessions they have played, the number of images they have labeled, as well as their per-session scores compared to that of the average scores achieved by other players; (ii) top scorers: the top 10 players in terms of the highest single session scores they have achieved; and (iii) top contributors: the top 10 players in terms of the cumulative scores they have achieved. These scores are displayed in order to encourage users to aim for higher scores and play more sessions.

Table 1: A comparison of resources used by experts and non-expert players during fish labeling.

Type	Candidates source	Verification source
Experts	From their knowledge	Textbook
Non-experts	Given by the system	System feedback

Table 1 shows a comparison between the resources used by experts and non-experts in the two labeling settings. In [6] we report studies on user behavior with respect to our labeling system.

Acknowledgements

This research was funded by European Commission FP7 grant 257024, in the Fish4Knowledge project (www.fish4knowledge.eu).

3. REFERENCES

[1] B. J. Boom, P. X. Huang, J. He, and R. B. Fisher. Supporting ground-truth annotation of image datasets using clustering. In *ICPR*, pages 1542–1545, 2012.

[2] Y.-Y. Chen, W. H. Hsu, and H.-Y. M. Liao. Learning facial attributes by crowdsourcing in social media. In *WWW'11*, pages 25–26, 2011.

[3] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beene, A. Leaver-Fay, D. Baker, and Z. P. & Foldit players. Predicting protein structures with a multiplayer online game. *Nature*, pages 756 – 760, 2010.

[4] R. Froese and D. Pauly. Fishbase, version (02/2013), 2013. URL `http://www.fishbase.org`.

[5] Galaxy Zoo. URL `http://www.galaxyzoo.org/`.

[6] J. He, J. van Ossenbruggen, and A. P. de Vries. Do you need experts in the crowd? A case study in image annotation for marine biology. In *OAIR'13*, 2013.

[7] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. Labelme: A database and web-based tool for image annotation. *Int J Comput Vision*, 77(1-3):157–173, 2008.

[8] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *CHI'04*, pages 319–326, 2004.

[9] J. Yuen, B. C. Russell, C. Liu, and A. Torralba. Labelme video: Building a video database with human annotations. In *ICCV*, pages 1451–1458, 2009.