Using Gamification to Create and Label Photos That Are Challenging for Computer Vision and People*

Piotr Kotlinski†
Electrical Engineering and Computer Science
National Chiao-Tung University Hsinchu Hsinchu City Taiwan
piokot83@gmail.com

Xi-Jing Chang
Institute of Applied Art National Chiao-Tung University Hsinchu City Taiwan
siliconcrystal.aa08g@nctu.edu.tw

Yang Chih-Yun
Institute of Applied Art
National Chiao-Tung University Hsinchu City Taiwan
jean.aa08g@nctu.edu

Wei-Chen Chiu
Computer Science
National Chiao-Tung University Hsinchu City Taiwan
walon@cs.nctu.edu

Yung-Ju Chang
Computer Science
National Chiao-Tung University
Hsinchu City Taiwan
armuro@cs.nctu.edu.tw

ABSTRACT

It would be hard to overstate the importance of Computer Vision (CV), applications of which can be found from self-driving cars, through facial recognition to augmented reality and the healthcare industry. Recent years have witnessed dramatic progress in visual-object recognition, partially ascribable to the availability of labeled data. Unfortunately, recognition of obscure, unclear and ambiguous photos that are taken from unusual angles or distances remains a major challenge, as recently shown by the creation of the ObjectNet [1]. This paper complements that work via a game in which obscure, unclear and ambiguous photos are collaboratively created and labeled by the players, who adopt the role of detectives collecting evidence against in-game criminals. The game rules enforce the creation of images that are challenging to identify for CV and people alike, as a means of ensuring the high quality of players' input.

CCS CONCEPTS

• Human-centered computing • Ubiquitous and mobile computing • Ubiquitous and mobile devices • Smartphones

KEYWORDS

Human-Computer Interactions, Mobile Crowdsourcing, Gamification, Computer Vision

UbiComp/ISWC '20 Adjunct, September 12–16, 2020, Virtual Event, Mexico

ACM ISBN 978-1-4503-8076-8/20/09.
https://doi.org/10.1145/3410530.3414420

ACM Reference Format:

Piotr Kotlinski, Xi-Jing Chang, Yang Chih-Yun, Yung-Ju Chang, Wei-Chen Chiu 2020. Using Gamification to Create and Label Photos That Are Challenging For Computer Vision and People. In P*roceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2020 International Symposium on Wearable Computers (UbiComp/ISWC 20), September 12-16, 2020, Cancun, México*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3410530.3414420

1 Introduction

In recent years, hardware progress has made it affordable for people to acquire a range of wearable devices including fitness trackers, personal cameras, and even virtual-reality equipment. In addition to their main functions, such devices produce a vast amount of contextual data that is hard to analyze, yet can possess attributes crucial for domains like personalized fitness training, home automation and even healthcare. Unfortunately, if the user moves away from the object of focus, looks at it from an unusual angle, or holds it in an unusual way, object-recognition performance can drop significantly. This, in turn, can reduce the usefulness of these devices, sometimes to nil, or even place the user in danger (e.g., in the case of wearable object-detection devices used by blind people to navigate in complex urban environments). Thus, more efficient object recognition is a critical area for the future development of this kind of equipment.

As part of that effort, this paper focuses on analyzing photos that are obscure, unclear, ambiguous, or taken from unusual angles, and/or contain surprising context [3]. Without a proper training dataset, CV accuracy can drop drastically, as was shown during the creation of the ObjectNet [1]. Yet, gathering such images is expensive and time-consuming, as they are not a common byproduct of daily life. Following the success of research by Luis von Ahn [2], we decided to create a Game With A Purpose (GWAP) to address this issue and that could complement ObjectNet in its current state; i.e., devise game mechanics that

Figure 1: Creating cases panels

will motivate people to provide such photos and create multiple labels for them. To achieve this, we chose a detective theme, as making obscure photos of unusual objects makes intuitive sense in the context of collecting evidence, and guessing the correct label shows similarities to solving criminal cases, e.g., by identifying rare stolen objects. We decided to target mobile devices as the most convenient for this type of a game, what should lead to creation of diverse data. As a check on whether players input could contribute to ObjectNet, we analyzed its recognition factor using Inception, ResNet, MobileNet and VGG16, pre-trained on weights from ImageNet. We provide results related to players' performances in the game. Finally, players attitudes' towards the game tasks were measured using a post-game questionnaire.

2 About Detective Pig

The goal of the game we designed, Detective Pig (Fig. 1), is to achieve a rank higher than that of the other players, making it highly competitive. To win, the player must correctly identify the subjects of photos taken by other players, while her/his own photos remain difficult or impossible for them to identify. As well as making the game more fun, this structure motivates people to provide obscure and unusual photos as a means of preventing others from achieving high scores. To give the players a sense of conflict, game villains were created. Players are shown pop-up reminders that one of these villains committed a crime, and that players should gather evidence against him. The overarching story is provided in the form of an animated introduction.

2.1 Game Rules and Gameplay

All the game's mechanics are focused on making photos and guessing the appropriate labels for the photos made by others. Success at the latter activity is rewarded with different types of in-game incentives. Its most important progress indicators are players' ranks (achieved for guessing and uploading photos) and the quantity of in-game currency that they own, known as Oinks. The graphics were designed to evoke Film Noir detective stories.

2.2 Creating Cases

To create a new case, players have to either make a new photo using the Evidence panel, or can use a photo previously stored on

Figure 2: An example of solving case

their device, which gives them full control over data creation, like suggested in previous studies [7] . Players also have to choose a photo's category (e.g., "Animals"), label the main object in it, and estimate how difficult it will be for other players to guess its label correctly. Players are rewarded with 100 Oinks for each new case they create. To check their input, we created a second app called Detective Pig's PhotoChecker, which either accepts the new case photo – meaning that the other players can see it, and gain additional Oinks by correctly guessing its label – or rejects it.

2.3 Solving Cases

In the second main game phase—solving cases—for each photo, additional labels from the players other than the player who contribute that photo are collected. It is stylized to make players feel like they are solving a criminal case in Detective Pig's office (Fig 2). This must be done within a randomly generated time limit, indicated by a timer visible at the top of the screen and by the animations that the player sees. A "guess" by a player consisted of typing the name of an object that s/he thinks is depicted in a case photo, and pressing the "Check" button. Every photo could be drawn by multiple users at random to increase the number of potential labels. Player could draw the same photo multiple times until successful guessing. Every wrong guess is penalized by the loss of 10 seconds. However, the guessing player need not run out the clock, but can give up at any time. If the player is able to guess the label, s/he is rewarded with a varying amount of Oinks, depending on the photo's difficulty. In the end, a pop-up is shown asking the player to rate the quality of the photo s/he just saw by giving it like or dislike, and allowing him/her to either try again or go back to the "office."

2.4 In-Game Rewards

A higher game rank can be attained by correctly guessing the labels of photos and uploading them, while medals are awarded for solving or creating certain numbers of cases. Additionally, a leaderboard is visible on the Detective Pig website, and the top

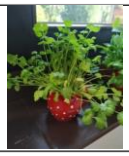
Hard label	Ambiguity	Wrong impression	Specificity
Label: chandalier	Label: coriander	Label: jellyfish	Label: wardrobe
Guesses: lamp(4), light(7), ring(1), filter(1)	**Guesses:** strewberry(10), plant(12), flowerpot(4), window(1)	**Guesses:** mushroom(2), medusa(4), satellite(1), start(1)	**Guesses:** wardrobe(6), dresser(5), cabnet(5)
			

Table 1: Examples of the most common mistakes

three players' nicknames are shown in the in-game News.

2.5 News

As well as being an important aspect of the game's Film Noir stylization, the News provides additional possibilities for player engagement, by giving feedback about game progress, providing information about top players, and assigning special missions.

2.6 Special Mission

Three *special missions* were given to the players, all involving photo-taking. Specifically, we asked them to make photos of 1) glasses with black frames, to expand ObjectNet's ~~"G~~lasses" category; 2) big jars, to expand ObjectNet's ~~"J~~ars" category; and 3) big trees, to expand a potential ~~"T~~rees" category. This resulted in players providing 26 photos of trees (20 of which met the requirement ~~"b~~ig"), 12 photos of glasses (all correct) and 17 photos of jars (15 of which could be classified as ~~"b~~ig").

3 Preliminary Game-Test Results

Participants were recruited through advertisements in Facebook groups for students in Taiwan, as that population is likely both to know English and to be interested in computer games. Each volunteer was required to have an Android cellphone or tablet running Android version 6.0 or higher. In all, 87 people volunteered, of whom seven were rejected because they did not meet one or more of these requirements. The game test took place from May 24 to June 1, 2020. The remaining 80 players were sent emails with links to the game on Google Play at the same time. On June 10, the post-game questionnaire was sent to all of them. During gameplay, we tried various additional methods of engaging players, including the creation of a website with a gallery of photos from the game, as well as the special missions alluded to above, which were also used to check if we were able to expand ObjectNet categories.

In all, the players tried to solve a case 5,315 times and guessed 10,540 times. They were able to solve 48% of the cases (2,544). Just over a third of their first guesses were correct (n=1,796).

Label: table	Label: candy	Label: headphones
likes: 18 dislikes: 1 user guessed: 9/19 times	likes: 15 dislikes: 4 user guessed: 0/19 times	likes: 20 dislikes: 4 user guessed: 7/24 times
		

Table 2. Examples of most liked photos

Crucially, the user-provided labels in almost 70% of total guesses could be used effectively as additional labels for the photos. However, when their first guesses were incorrect, players exhibited problems recovering, and were able to go on to solve the case in question only 748 times (i.e., in around 14% of all case-solving attempts and 7% of all guesses). We divided main reasons for players' failure into the following four distinct categories.

1. **Hard label:** the correct label consisted of a difficult or rarely used word
2. **Ambiguity:** the photo included many objects
3. **Wrong impression:** the object in the photo reminded the player of something else
4. **Specificity:** the user-provided label was either too specific or too general

The players uploaded 1,324 images. We had to reject 90 of them: eight as being impossible to guess, eight as having labels incompatible with game rules, four as having been downloaded from the Internet, five as concerning politics or religion, and the remaining 65 as duplicative of photos uploaded to the game already.

The game included ~~"l~~ike" and ~~"d~~islike" buttons for each photo. Players gave 4,406 likes and 1,095 dislikes, equating to 80% and 20% of their solving attempts, respectively. Our analysis of most-liked and most-disliked photos indicated that the established game rules were compatible with players' preferences. Provided that the photo was focused on the main object, the guesser tended to give it a like, whereas ambiguous photos tended to be given dislikes. Photos with tricky labels were especially disliked.

Players very rarely used labels that would make guessing impossible. Their most common mistake was providing a label that was quantitatively inconsistent with the number of the objects of the same type that were visible in the photo (44 cases).

4 Computer Vision Performance Test

The main purpose of this test was to check if the labeled photos created during gameplay could complement ObjectNet. Therefore, our main interest was in checking the general recognition factor. Overall, this was lower than in ObjectNet: with Inception recognizing 37% of all our users' photos; MobileNet, 30%;

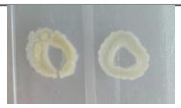
Label: kettle	Label: penne	Label: tablet
likes: 8 dislikes: 11 user guessed: 0/18 times	likes: 3 dislikes: 8 user guessed: 0/191 times	likes: 3 dislikes: 15 user guessed: 0/18 times
		

Table 3. Most-disliked photo examples

Resnet, 31%; and VGG16, 30%. None of the big trees and jars input during special missions were recognized. However, most of the special-mission photos of glasses were identified (Inception: 60%, MobileNet: 42%, ResNet: 50%, VGG16: 60%).

Additionally, we conducted a second test, during which all four of the CNNs mentioned above were treated as if they were players. This meant that the predicator had to be identical with the label given by the photo-uploading user, and only the top three predicators were checked (players gave three guesses per photo on average). Under these conditions, Inception was able to guess correctly in only 7% of cases, MobileNet in 6%, ResNet in 6%, and VGG16 in 5% in comparison to 24% of correct guesses of human players. After comparison between the results of both experiments, our conclusions were similar to those of - Are we done with ImageNet [4]. While many objects in the photos were recognized, the ImageNet labels were not always the best predictors, and consistently obtaining valid results required us to go through predictions manually. In addition, our tests showed that the players were choosing correct difficulty levels for the photos they were uploading in an unbiased way – the higher the difficulty level chosen by the uploading player was, the more problems with recognizing object on the photo the CNNs had (Table 4).

5 Questionnaire

Just like in various prior studies [5, 6], players main motivation for playing the game was wanting to help science (Mode: 7; Median: 6; SD: 1.5). On the other hand, many of our players decided to play because they found the game tasks to be entertaining: a very important factor, especially when it comes to titles as niche as the game presented in this work. Their third-highest motivation was curiosity about what would happen next, implying that future iterations of Detective Pig should include more surprises. Among all the in-game incentives, the two most motivating were the leaderboard (Mode: 7; Median: 5; SD: 1.8) and achieving higher rank (Mode: 6; Median: 5; SD: 1.8). Rewards like medals and titles appeared to be less interesting. Players found the game mechanics easy to learn and intuitive (Mode: 7; Median: 7; SD: 1.01). Most importantly, the majority of the players said they would like to play Detective Pig in the future (Mode: 6; Median: 6; SD: 1.23), and that they were willing to recommend it to friends (Mode: 6; Median: 6; SD: 1.19). Additionally, they reported that creating cases that would be hard for others to solve gave them satisfaction (Mode: 6; Median: 6; SD: 1.14).

Photo difficulty level	Inception	ResNet	MobileNet	VGG16
1	228/492	193/492	178/492	185/492
2	164/433	133/433	145/433	134/433
3	48/224	45/224	47/224	42/224
4	20/86	20/86	17/86	19/86
5	4/43	3/43	2/43	5/43

Table 4. Number of photos recognized per difficulty level

6 Conclusions and Future Work

Although it seems that the game achieved its goals, some changes to it will also clearly be required. First, this iteration only counted one answer as correct, which was unnerving for players in many cases, e.g., when they guessed "screen" and the only acceptable answer was "monitor". The case-creating player should therefore be instructed to provide multiple synonymous labels. Second, the current process of checking and approving photos is very time-consuming and tiresome, and should be automated. Lastly, the rules for accepting photos should be made stricter.

ACKNOWLEDGMENTS

We thank Michael Angkawidjaja and Jin-An Lin for the help in testing "Detective Pigs" usability. This research was supported in part by the Ministry of Science and Technology, R.O.C (MOST 108-2218-E-009 -050).

REFERENCES

[1] Andrei Barbu, David Mayo, Julian Alverio, William Luo, C. Wang, Dan Gutfreund, Josh Tenenbaum and Boris Katz (2019). ObjectNet: A large-scale bias-controlled dataset for pushing the limits of object recognition models. In 33rd Annual Conference on Neural Information Processing Systems (pp. 9449-9458). NIPS, Vancouver, Canada.

[2] Luis von Ahn and Laura Dabbish (2004). Labeling images with a computer game. In ACM Conference on Human Factors in Computing Systems (pp. 319-326) ACM, New York, United States.

[3] James Vincent (2019).The mind-bending confusion of "hammer on a bed" shows computer vision is far from solved. These images are designed to fool. Retreived from: https://www.theverge.com/2019/12/12/21012410/machine-vision-ai-adversarial-images-dataset-objectnet-mit-algorithms

[4] Lucas Beyer, Olivier Hénaff, Alexande Kolesnikov, Xiaohua Zhai, and Aäron Oord (2020). Are we done with ImageNet? Retrieved from: https://arxiv.org/abs/2006.07159

[5] Pei-Yu Chi, Matthew Long, Akshay Gaur, Abhimanyu Deora, Anurag Batra, and Daphne Luong (2019). Crowdsourcing images for global diversity. In Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices (pp. 1-10). ACM, New York, United States.

[6] Pei-Yu Chi, Anurag Batra, and Maxwell Hsu (2018). Mobile crowdsourcing in the wild: Challenges from a global community. In Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices (pp. 410-415). ACM, Barcelona, Spain.

[7] Yung-Ju Chang, Gaurav Paruthi, Hsin-Ying Wu, Hsin-Yu Lin, Mark W. Newman (2017). An investigation of using mobile and situated crowdsourcing to collect annotated travel activity data in real-word settings. In International Journal of Human-Computer Studies, Vol 102 (pp. 81-102)