R11/Ishank Juneja/16D070012

March 3, 2020

Crowd-Sourcing is the method of employing a large number of people to a task, typically by connecting them to the project through the internet. The paper by Ahn and Dabbish is one of the early works to employ crowd sourcing effectively. In the paper, the authors describe the methodologies employed to implement their image-labelling project.

Despite the best efforts of the Computer Vision community, there isn't any software available that labels images accurately across a wide range of contexts. This limitation of CV was especially true in 2004, when the paper was written, however, even today the importance of high quality pre-labelled image data cannot be over stated. This is the case since despite the accuracy of modern deep learning methods, expanding their scope and capabilities is dependent largely on the availability of such labelled images. The problem that the paper solves through crowd-sourcing is precisely this - labelling images accurately **at scale**. The authors achieved this by deploying a **game they call ESP**.

The labelling game - ESP - is played between two players who both try to guess a good label for the same image shown to both of them. The players get points and proceed to a new image once they can agree upon a single descriptor (with fewer than 13 characters) for the picture. Players viewing the same image are randomly assigned to each other and cannot communicate by any means. By itself this makes the game fun and challenging and the use of *taboo* words also makes the game even more enjoyable. Taboo words are certain words which cannot be chosen as guesses by the players. These taboo words are derived from labels that an earlier pair of players would have agreed upon. By forcing participants to look for words other than the most obvious first choice, the game also generates several high quality labels for a single image.

The key to making a crowd-sourcing task successful is the presence of the right incentives for participants. Monetary compensation could act as an incentive when the budget to do so is available. However, due to financial constraints, this strategy is likely to fail at scale. In the absence of monetary incentives, the activity being fun for the participants is crucial. By designing a fun game environment, the team was able to achieve engagement and good quality results.

In their paper, the authors also thoroughly validate the accuracy of the crowd sourced labels. First they verify that the results obtained on searching for a few popular keywords such as *car* are accurate. Then they invite 15 volunteers

who are asked to assign a few labels to some randomly selected subset of their dataset and lastly a different set of volunteers were asked to rate the quality of the sourced labels. The results from all these tests suggested that the sourced labels were very high quality.

The paper presents their methodology as a superior alternative to other methods at the time including Computer Vision (CV) and text scraping from the web. Results cited by the authors suggest that the state of the art CV image labelling method at the time was quite narrow and limited. The other option, which was the prevailing method, was text scraping from the web-page hosting the image. Although sound in principle this approach presents a challenge since the text adjacent to images is scarce and hard to condense into a single label. The key takeaway from the work is that crowd-sourcing labelled data at scale is effective given the right incentives. The impact of this work is quite visible in the form of image based *Captchas* sourcing object recognition labels from unsuspecting participants surfing the web!

A question I have after the reading the paper is

- Do the authors claims of Computer Vision not being able to label images in a useful way still hold up today? It bugs me since the paper was written in 2004 before the wave of Deep Learning based Computer Vision arrived.