Vrije
Universiteit
Brussel

Methods for scientific research

Assignement 4 - Paper review

Bruno Rocha Pereira - 0529512

Paper n°9 : Donald Geman, Stuart Geman,Neil Hallonquista and Laurent Younes(2015) Visual Turing test for computer vision systems PNAS 112(12) 3618-3623

Wordcount : 708

1 Closed questions using Likert Scale

'Very Good', 'Good', 'Average', 'Bad', 'Very Bad'

- How confident are you ? Good

- How relevant is the paper ? Very Good

- Is the writing clear ? Good

- Is the method clearly explained ? Good

- Is it technically sound ? Good

- Are the findings relevant ? Good

- Is related work cited correctly ? Very Good

2 Summary

In this paper, the authors present a Virtual Turing test(VTT) adapted for computer vision systems. The VTT consists on

series of questions with only three types of answer possible: "yes", "no" or "ambiguous question". The questions are aimed to be asked to a computer vision system assisted by an operator on a particular known set of images. The operator has two roles : he either indicates that the question is ambiguous or provides the correct answer. The main goal of this query generator is to allow machines to associate semantics to images.

During the whole list of questions, every single question has to have an unpredictable answer and can be of 4 different types: existence questions, uniqueness questions, attribute questions and relationship questions. These questions have to be asked in a particular order, since for example a relationship question cannot be asked before instantiating at least 2 elements of an image.

3 Brief opinion

This paper is quite accessible and addresses what could be considered a futuristic problem. The authors managed to propose a valid VTT with the theory they had. The images

clear subdivisions and the queries example allow the reader to understand more quickly the whys and wherefores of this paper.

4 Wider explanation

The article was written with relatively easily understandable terms such that, in my honest opinion, a non-scientist person could understand it. However, in some parts of it, it suddenly becomes a little bit more technical and mathematical to describe simple concepts. For example, the formula for the relative frequency is, in my opinion, not needed and not very clear. Nonetheless, as stated previously, the images provided, as well as their caption are easily understandable and completely relevant to the topic.

The state of the art is clear, sufficient and completely matches the topic. The references are clearly stated and most of them are pretty recent since the field of computer vision is quite new to explore. Although, in order to include a wider choice of questions, the articles could have also come from fields other than computer science since the topic relates to a mat-

ter of human visual perception. This is however a completely different and higher level that is left to be achieved in the future that could have still been discussed at the end of the article.

The methodology used corresponds to what would be expected considering the topic approached. The query generator idea is pretty clear and how it works is explained multiple times as well. Every idea behind it is explicitly specified and explained with fairly understandable sentences and math formulas.

The supplemental information document[1] brings a lot more information to this paper. It allows the reader to get more specific data on how the query generator was implemented. It also explains the semantic of the vocabulary related to it. This document is definitely less accessible to a wider public than the paper in itself and is much more specific. It contains a lot of mathematical equations which even explained may be complex to understand.

Another thing that could be improved is that all along the

---

[1]http://www.pnas.org/content/suppl/2015/03/06/1422953112.DCSupplemental/pnas.1422953112.sapp.pdf

paper, the authors assume the objectivity of how ambiguous a question is, which can be discussed. For example, how much of a person does the frame have to contain to consider the question "Is there a person in this region" unambiguous? A broader public to collect the data ans annotate the images would have made it a little bit less objective. The last thing I see that could be ameliorated is that the criteria to match for the images to be considered are quite specific. The author did not consider any other image. Stretching out the constraints would have rendered a larger set of data and as a result, more reliable test results.

5 References

Donald Geman, Stuart Geman,Neil Hallonquista and Laurent Younes(2015) Visual Turing test for computer vision systems PNAS 112(12) 3618-3623