Image and Vision Computing 69 (2018) 1–8

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Negative results in computer vision: A perspective☆

Ali Borji

Center for Research in Computer Vision, University of Central Florida, Orlando, FL., USA

ARTICLE INFO

Article history:
Received 4 October 2017
Accepted 16 October 2017
Available online 26 October 2017

Keywords:
Negative results
All results
Computer vision
Biological vision
Statistical testing

ABSTRACT

A negative result is when the outcome of an experiment or a model is not what is expected or when a hypothesis does not hold. Despite being often overlooked in the scientific community, negative results are results and they carry value. While this topic has been extensively discussed in other fields such as social sciences and biosciences, less attention has been paid to it in the computer vision community. The unique characteristics of computer vision, particularly its experimental aspect, call for a special treatment of this matter. In this manuscript, I will address what makes negative results important, how they should be disseminated and incentivized, and what lessons can be learned from cognitive vision research in this regard. Further, I will discuss matters such as experimental design, statistical hypothesis testing, explanatory versus predictive modeling, performance evaluation, model comparison, reproducibility of findings, the confluence of computer vision and human vision, as well as computer vision research culture.

1. Introduction

What is a negative result? One may characterize a negative result as "when a hypothesis does not hold" or "when the outcome of an experiment or a model is not what is expected". Such a definition, however, could be one out of many possible definitions. One may argue that an unexpected result is actually a good useful positive result to share. Another possible definition is that a negative result is when the performance is not better given metrics such as accuracy. Regardless of how negative results are defined, such challenging and sometimes inconclusive findings are often discouraged and buried in the drawers and computers. Therefore, the publication record reflects only a tiny slice of the conducted research. In some sense they fabricate the "dark matter" of science. Such findings, however, still hold value. At the very least they can save resources by preventing researchers from repeating the same experiments. Perhaps the main reason for an overwhelmingly high number of negative results not put forward for dissemination is the lack of incentives. Interestingly, some researchers have even argued that most published findings are false [1]. Some also claim that hiding negative results is unethical. Nevertheless, negative results have been and continue to be constructive in the advancement of the science (e.g., Michelson-Morley experiment [2]).

To answer whether negative results are important in computer vision, should be published, or even if it makes sense to talk about them, first we need to investigate how computer vision research is conducted relative to scientific practices and methodologies conducted in other fields such as social or biological sciences. Computer vision research consists of a mixture of theoretical and experimental research. A small fraction of publications introduce principled theories for vision tasks (e.g., optical flow [3]). A large number of publications report models and algorithms (e.g., for solving the object detection problem) that are more powerful than contending models. Thus, compared to other fields, computer vision is relatively less hypothesis-driven and more practical. Some negative results offer invaluable insights regarding strength and shortcomings of existing models and theories, while others provide smart baselines. The emphasis has traditionally been placed on improving existing models in terms of performance over benchmark datasets. While some papers conduct statistical tests, it is not the common practice. As in some other fields, there is a high tendency among computer vision researchers to submit positive results as such results are often considered to be more novel by the reviewers.

Computer vision has its own unique characteristics making it distinct from other fields, thereby demanding a specific treatment of negative results. Firstly, vision is an extremely hard problem which has baffled many smart people throughout the history. The complexity of the problem makes it difficult to run controlled experiments and come up with a universal theory of vision. Secondly, often a lot of variables are involved in building vision algorithms and in analyzing large amounts of data. Further, fair comparison of several

☆ This paper has been recommended for acceptance by Sinisa Todorovic.

E-mail address: aborji@crcv.ucf.edu.

https://doi.org/10.1016/j.imavis.2017.10.001
0262-8856/

competing models using multiple evaluation scores exacerbates the problem. To address these, it would be very helpful to borrow from other fields (e.g., natural sciences) where experimental design and statistical testing are integral parts of the scientific research.

The common practice in experimental hypothesis-driven fields (e.g., cognitive science) includes carefully formulating a hypothesis, identifying and controlling confounding factors, designing the right stimulus set, collecting high quality data, and performing appropriate statistical tests. These are complicated to perform in computer vision research as often many factors are involved. In particular, statistical analysis becomes very challenging in the presence of many parameters and models. This makes it complicated to decide which statistical test is needed or when statistical analysis is critical to conduct. Principled and systematic gauging of the progress (rather than relying on trials and error and luck) helps judge what truly works and what does not and, hence steer the research in the right direction. For instance, we might have not given up on neural networks easily if we did more careful rigorous analyses in the past.

Notice that dealing with negative results is a very controversial topic and still unsettled in many fields. So, do not expect this writing to touch on all of the aspects. Rather, here, I try to shed light on some less explored matters and put computer vision in a broader perspective with respect to science in general, and its related fields such as Neuroscience and Cognitive Science, in particular. Indeed, further discussion is needed in the vision community to converge to a consensus regarding treatment of negative results.

In what follows, first I elaborate on science versus engineering and where computer vision fits. I will continue with a comparison of computer and human vision research and how they relate to each other in terms of goals, research methodologies and practices. This is followed by discussions of negative results and statistical analysis in the context of computer vision. Section 6 considers the dissemination of negative results. Finally, a wrapup is presented in the epilogue.

2. Computer vision: engineering or science?

Let's start with the question of whether computer vision is a scientific or an engineering discipline, or both. Science is concerned with understanding fundamental laws of nature, whereas engineering involves the application of science to create technology, products and services useful for society. Science asks questions about nature while engineers design solutions to problems.

As a scientific discipline, computer vision is concerned with gaining high-level understanding from digital images, video sequences, views from multiple cameras, or multi-dimensional data. It seeks to automate tasks that the human visual system can do and involves the development of a theoretical and algorithmic basis to achieve automatic visual understanding. Further, it deals with constructing a physical model of the scene (i.e., how the scene is created), how light interacts with the scene, as well as low-, intermediate-, and high-level descriptions of the scene content [4]. In other words, the ultimate goal of computer vision is image understanding, the ability not only to recover image structure but also to know what it represents. As a technological and engineering discipline, computer vision seeks to apply its theories and models for the construction of computer vision systems and applications.

Science and engineering are complementary and are beautifully and happily married in computer vision. We have a very solid in-depth scientific understanding of phenomena such as image formation, depth perception, stereoscopic vision, color perception and optical flow. Some engineering applications, among many, include biometrics (robust face and fingerprint recognition), optical character recognition, gesture recognition, motion capture, game playing, structure from motion, image stitching, machine inspection, retail, 3D model building, medical imaging, automotive safety, autonomous cars, assistive systems, and surveillance (in traffic and security). In this respect, computer vision is both theoretical (e.g., optical flow formulation) and experimental (e.g., model replication, parameters tuning, hacks, and tricks).

3. Computer vision and biological vision

Vision is a broad interdisciplinary area. Both computer and human vision systems share the same objective which is converting light into useful signals from which accurate models of the physical world can be constructed. This information helps an agent (e.g., be it a robot or a human) live, act, and survive in its environment.

For a long time, human vision research has been concentrated on understanding the principles and mechanisms by which biological visual systems (with higher emphasis on primate vision) operate. This is in essence a reverse engineering (or inverse graphics) task. Likewise, computer vision research seeks a theory and engineering implementation. Despite sharing the same goal, they own unique characteristics. Early human visual sensory mechanisms, including the retina and the Lateral Geniculate Nucleus (LGN), are much more elaborate than current digital cameras (CCD sensors). Neural networks in higher visual areas (e.g., visual ventral stream) accommodate a sophisticated hierarchical processing through cascades of filtering (modeled as convolution), pooling, lateral inhibition, and normalization mechanisms. The result is a selective and invariant representation of the objects and scenes. This is somewhat akin to what Convolutional Neural Networks (CNNs) [5] do. Almost half of the human brain (considered to be the most complex known physical systems and thus a major scientific challenge) is devoted directly or indirectly to vision. The entire brain needs about 20 W to operate (enough to run a dim light bulb). A processor as smart as the brain requires at least 10 to 20 MW of electricity to operate [6]. As to processing speed, the brain is still faster than the fastest supercomputers [7]. A remarkable capability of human vision is attention (a.k.a active vision) which allows selecting the most relevant and informative part of the massive incoming visual stimulus (at a rate of 10^8–10^9 bits/s) [8]. Both human and computer vision systems have their own biases. Human vision is extremely sensitive to faces and optical illusions. Similarly, computer vision systems get easily fooled by adversarial examples [9]. One thing that we know, almost for sure, is that vision should be solved by frameworks that start with extracting simple features and build increasingly more complex ones. This is mainly because the world we live in is compositional.

There has indeed been a cross-pollination in the two fields (e.g., [10–21]). On the one hand, experimental paradigms and psychophysics tools in cognitive vision have been exploited to study the behavior of computer vision algorithms or to interpret how they work. For example, Parikh and Zitnick [22] employed the image jumbling paradigm, introduced in [23], to inspect whether some computer vision algorithms capture local or global scene information. Deng et al. [24] used the bubbling paradigm, proposed by Gosselin and Schyns [25], to model fine grained object recognition. The rapid (or ultra rapid) serial visual presentation [26,27], has been utilized to investigate the quality of images generated by Generative Adversarial Networks [28]. Vondrick et al. [29] and Fong et al., [30] leveraged human recognition biases to improve machine classifiers. On the other hand, computational tools have been exploited heavily to understand how human vision works. For example, deep convolutional networks have recently been used to study the representational space in the visual ventral stream (e.g., [31]). Moreover, a plethora of computer vision, image processing, and machine learning tools have been utilized in biological vision research for the purposes such as stimulus design, discovering cues humans might rely on in solving a task, and modeling single neurons and neural populations.

In terms of performance, while computer vision has made large strides, it is still nowhere near human vision. In general, it seems that