The Cognitive Philosophical Problems in Visual Attention and Its Influence on Artificial Intelligence Modeling

Jing-Jing Zhao

The Cognitive Philosophical Problems in Visual Attention and Its Influence on Artificial Intelligence Modeling

Jing-jing Zhao

College of Liberal Arts and Science, National University of Defense Technology, Changsha, Hunan, China

jjzhao1983@126.com

ABSTRACT. Human perception of visual scenes has distinct initiative and purpose characteristics. Human beings can quickly extract information of their interest from massive visual input, giving priority to processing. This selective perception is visual attention. With the rise of artificial intelligence, studying the mechanism of human visual attention and establishing the computational model of visual attention has become a new research hotspot. What is the essence of visual attention? What are the basic units of visual attention? What are the factors that affect visual attention? The in-depth analysis of cognitive philosophy in visual attention helps to understand the mechanism of visual attention and to establish an effective artificial intelligence model. The fact shows that the performance of the computer simulation model can be improved by taking full account of the influence of high-level factors such as task, expectation, memory, knowledge and experience in modeling.

KEYWORDS: visual attention, visual cognition, AI Philosophy, computational model of visual attention

1 Introduction

Most of human's perception of the world comes from vision. Compared with the massive visual information received by the retina, the neural resources of our visual system are very limited. So there is a contradiction, the total amount of information provided by the perceptual system is far beyond the maximum capacity of the cortex to store information[1]. However, this so-called bottleneck effect does not bring any discomfort to our visual perception. That's because the cognitive process of human vision is an active process of interacting with the outside world. Through eye scanning, it can make the region of interest to the retinal foveal feature to recognize the target priority, using high resolution foveal. The selective perception is visual attention, which can filter visual information, give priority to important information, and control behaviour based on the information.

As far as physiological characteristics are concerned, visual attention is the basic function developed by primates in the long term evolution process for survival needs.

Visual attention can help to find the enemy and from the primate class animal food real-time massive visual information, so the visual information is significant and the environment and memory are closely related. As far as the psychological characteristics are concerned, visual attention reflects the importance of human psychological activity initiative and consciousness, so most of the mathematical modeling of visual attention simulate the filtering operation in the human brain in the information process. As far as the characteristics of machine vision are concerned, the construction of machine vision system requires quick response to complex and changeable external environment through attention mechanism, so as to respond in real time. Therefore, visual attention is related to specific tasks. At present, the research on visual attention mechanism is mainly focused on the cognitive study of visual attention and the computer simulation modeling of visual attention. Scholars in the field of cognition mainly study the mechanism of visual attention mechanism and the construction of cognitive models. Scholars in the fields of computer vision, pattern recognition and artificial intelligence are mainly concerned with how to establish a computational model of visual attention by combining the research results of visual physiology and cognition. The effectiveness of the computer simulation model depends largely on the idea of modeling. The idea of modeling often comes from the understanding of the cognitive mechanism of visual attention. Therefore, the in-depth analysis of the cognitive mechanism of visual attention and the analysis of the cognitive philosophical problems have a great guiding role in improving the thinking of modeling and improving the performance of artificial intelligence model.

2 Cognitive philosophical problems of visual attention

Although we all know visual attention, so far, the cognitive mechanism of visual attention has not been conclusive. In terms of cognitive philosophy, the main philosophical issues discussed include: what is the essence of visual attention? What are the basic units of visual attention? What are the factors that affect the visual attention? Is there a conscious participation in the process of attention? How to understand the relationship between attention and recognition and so on.

2.1 What is the essence of visual attention?

At first, some theories suggest that visual attention function is the front end of the visual system, so it is necessary to find the area of interest first, and then carry on the more elaborate processing of the back end such as the target recognition. They have studied from the point of view of biological evolution[1]. The representative scholar is BroadBent, who emphasizes that the function of visual attention is actually a filter. When the external visual stimulation enters the filter, it passes through several channels to receive preliminary processing first, and then carries out fine processing

of object identification and semantic analysis through subsequent processing[2]. Einhauser et al. pointed out that the correct relationship between visual attention and perception has not been solved to a large extent. Visual attention and perceptual models should be integrated, rather than simply taking visual attention as a preprocessing step for target recognition[3]. Grossberg thinks that the basic unit of visual attention is the surface and boundary. He thinks that there is a link between attention and learning, expectation, competition and consciousness[4]. Fazl et al. proposed the ARTSCAN neural model, and proposed the concept of attention coverage. The prediction results of the model have high consistency with the psychological experimental data[5]. It provides a unified explanation for target attention and spatial attention on how to work together and how to learn content in the scene together. Chikkerur et al. believes that attention is part of the reasoning process that solves the visual task of "where there is"[6]. They emphasize that the main role of the visual system is to infer the identity and position of the target in the visual scene。 They introduced Bayesian inference theory for visual attention, and integrated the target identity information and target location information of the dorsal pathway.

2.2 What are the basic units of visual attention?

The basic unit of visual attention has always been a controversial issue. At present, there are several viewpoints on this issue. Based on the discernibility view, the attention system is restricted by the number of discernibility, so the unit of visual attention is related to the system resolved properties[7]. According to the time view, visual attention selects the attributes that occur simultaneously in time[8,9]. From the viewpoint of feature, visual salience is based on various characteristics of objects, which is the combination of these characteristics in guiding visual attention[10,11]. The space based view believes that attention is like a "Spotlight" moving in the visual area. Attention needs to be focused on the specific space area in the visual scene, only the visual information falling into the area can be followed by subsequent analysis, and the information that is not in that area is ignored[12]. From the viewpoint of objects, visual attention plays a role in perceiving units or objects that have been well organized in the pre attention stage. Visual attention can directly select discrete objects in the visual area, instead of concentrating attention on a certain spatial area in the visual scene[13-15]. Therefore, when the attention is paid to a certain object, the various components of the object can be parallel processing in time, while other objects can only be processed in time sequence. At present, more and more psychological experimental results provide support for object-based visual attention theory. The process of attention is consciously involved. Attention is related to perception and recognition.

2.3 What are the factors that affect the visual attention?

At present, a hot topic in the field of visual cognition is to study the main factors that affect visual attention, that is, what affects our visual attention. According to the physiological research results, ventral pathways and dorsal pathways are involved in this process in different ways. Visual attention can be bottom-up and driven by image data. Visual attention can also be top-down and conceptually driven [16]. With the advent of the eye motion recorder, many experimental data sets related to visual attention have been set up. It is possible to obtain more information about attention from the human visual experience.

This paper takes the data set [17,18] released by Massachusetts Institute of Technology as an example to analyze the key factors that affect visual attention. The data set contains 1003 images. From the content point of view, the image of the data set covers natural landscape images, human scene images, and images related to portraits, characters, animals, buildings, and so on. The researchers invited 15 people to look at the images and record the focus distribution of each person's gaze when they watched the image through an eye motion recorder. The researchers recorded each of the 15 people watching the first 5 focuses of each image. Then, these focal points in the same image are labeled, and the overall distribution of the focus of each image is gotten, as shown in Figure 1.

Fig. 1. Eye focus distribution collected from the eye movement data set [17,18]

By analyzing the focus distribution map of the dataset, we can find [17]: People tend to pay attention to faces, people and texts, followed by body parts, animals and cars. When there are portraits in the image scene, almost all observers observe the faces instinctively, regardless of whether the parts of the face are clear. When text information exists in the scene images, the observers tend to pay attention to these textual features. These data indicate that visual attention is related to the knowledge structure, personal experience and memory in our brain. Figure. 2 is a schematic diagram of the relationship between human visual focus and image content provided

by data set publishers. It can be seen from the Figure.2 that objects with high-level semantic features are more likely to arouse the attention of subjects, and this kind of attention is generally unified.

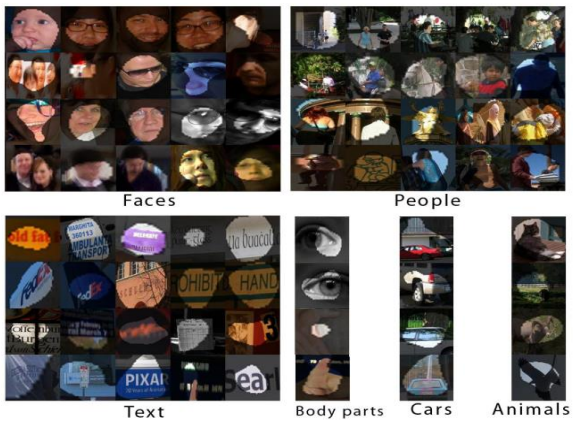

Fig. 2. The image content of the attention focus of people [17,18]

3 The computational model of visual attention

Human perception of visual scene has a distinct initiative and purpose, and the perceived content is closely related to the specific visual tasks and memory mechanisms. Therefore, visual perception can be regarded as an active process under certain conditions. As the new research direction and hot spot of computer vision, the core of bionic vision is to simulate the active perception ability of the biological vision system and accomplish the corresponding task or purpose from the mechanism. The establishment of visual attention computer simulation model is one of the important research contents, which can help us create a bionic visual system which is closer to the human visual cognitive process, in order to accomplish more computer vision tasks.

In the field of computer vision modeling, from the international conferences and periodicals published in the field of computer vision, it can be seen that the computational model of visual attention is one of the current research focus. At first, the research team of California Institute of Technology and the research team of University of Southern California presented their respective representative models respectively. In recent years, with the in-depth research, Massachusetts Institute of Technology and other institutions have carried out related research. The first theoretical framework for visual attention was put forward in 1985 [19]. After several years of exploration and efforts, Itti and Koch put forward the first computer simulation model of visual attention in 1998 [20]. The model has been universally

recognized by the academic community, and has become the benchmark of performance evaluation for current visual attention computation models. Walther and others extended the above model, and extended the selected unit to the prototype by hierarchical feedback connection, and realized the continuous multi target recognition in the scene[21]. Navalpakkam and others integrated the bottom-up attention model and top-down statistical knowledge to further optimize the speed of target detection[22]. With the deepening of research, scholars have actively explored the mechanism of attention. Instead of sticking to the original model framework, scholars put forward a series of innovative models from other perspectives, such as: computational models based on spectrum analysis[23-25], computational models based on entropy[26], computational models based on area[27], computational models based on graph theory[28], computational models based on machine learning[29,30]. Machine learning based modeling is the current research hotspot.

4 Core issues and strategies for constructing visual attention computational models

Although the human visual system has evolved to a nearly perfect stage, our knowledge of our own visual system is very limited. Thankfully, science always advances and progresses. On the one hand, scientists have developed eye tracking system to obtain the focus distribution of attention when humans are free to watch, and to get "ground truth" through the analysis of the human visual focus. It reflects the response of human visual system to visual scene content, and describes the distribution of saliency in visual scenes. With the appearance of "ground truth", we have real experimental data. On the other hand, with the development of cognitive philosophy, it provides a good theoretical guidance for computer modeling. We can try to think about the core issues of the visual attention computational model from the perspective of cognitive philosophy, and give corresponding coping strategies.

4.1 The basic unit selection problem of visual attention computational model

The biggest difference in the basic unit selection of the visual attention computational model is that the visual attention calculation should be based on the basic unit of the object or the space position as the basic unit. In addition, the definition of "object" is also controversial. Does it refer to objects with strict semantic characteristics or some similar target areas? From a psychological point of view, several viewpoints have been supported by experiments. However, from the perspective of cognition, object based visual attention computing is more conducive to subsequent visual analysis. Theoretically, features with more semantic information are more conducive to subsequent identification and analysis. There is no semantic feature in the pure area, and it has no significant effect on the identification and analysis of subsequent dorsal pathways. Therefore, on the basic unit selection

problem, the idea of object based modeling is more in line with human visual characteristics.

4.2 Feature selection of visual attention computational model

In view of the problem of feature selection in visual attention modeling, different visual attention models are divided mainly in the need to add semantic information on the basis of primary visual features. From a physiological point of view, the computation of visual attention is actually a very complex neural activity, which is not only influenced by the current input signal, but also influenced by the prior knowledge of the memory. Brain science has shown that in the connections between the cerebral cortex and the thalamus, the information transmitted backward is more than an order of magnitude higher than the information transmitted forward. The prior knowledge in memory is a summary of the cerebral cortex for the historical input signal and its response. It has a high level of generality and is a high-level representation of the scene. In addition, cognitive studies show that in the long term memory, the coding of scene images is mainly semantic code. Therefore, from the point of view of physiological structure and cognitive experience, the establishment of visual attention computational model needs to take into account the effects of the primary visual features and the visual features of the semantic information. It is necessary to integrate the integrated effect of the current visual input signal and the historical input signal in memory.

4.3 Selection of multiple visual features fusion strategies for visual attention computational models

In view of the problem of feature selection in visual attention modeling, different visual attention models are divided mainly in the need to add semantic information on the basis of primary visual features. From a physiological point of view, the computation of visual attention is actually a very complex neural activity, which is not only influenced by the current input signal, but also influenced by the prior knowledge of the memory. Brain science has shown that in the connections between the cerebral cortex and the thalamus, the information transmitted backward is more than an order of magnitude higher than the information transmitted forward. The prior knowledge in memory is a summary of the cerebral cortex for the historical input signal and its response. It has a high level of generality and is a high-level representation of the scene. In addition, cognitive studies show that in the long term memory, the coding of scene images is mainly semantic code. Therefore, from the point of view of physiological structure and cognitive experience, the establishment of visual attention computational model needs to take into account the effects of the primary visual features and the visual features of the semantic information. It is

necessary to integrate the integrated effect of the current visual input signal and the historical input signal in memory [31].

5 Conclusion

Computer simulation modeling of visual attention has always been a fascinating research topic in the field of artificial intelligence. Therefore, we need to dig into and analyse the cognitive philosophical problems in visual attention, to recognize the basic elements that affect visual attention, to understand the relationship between attention and consciousness, to identify the basic units of visual attention and to understand the mechanism of visual attention. In this way, we can better extract effective visual features in artificial intelligence modeling, build a simulation model based on task and high-level semantic information, and calculate visual attention based on image data to solve more artificial intelligence problems.

6 Acknowledgement

This work was supported by the National Social Science Fund of China（Visual attention research based on the field of artificial intelligence）, Hunan Provincial NSF of China under Grant 2015JJ3018, and Pre research project of National University of Defense Technology under Grant JS17-03-19.

References

1. Zhu Baorong. Applied Psychology [M]. Beijing: Tsinghua University Press, 2009.
2. Broadbent D. Perception and Communication [M]. London: Pergamon Press,1958.
3. Einhauser W, Spain M, Perona P. Objects predict fixations better than early saliency [J]. Journal of Vision,2008,8(14):1-26.
4. Grossberg S. Linking Attention to Learning, Expectation, Competition, and Consciousness [R], 2003.
5. Fazl A, Grossberg S, Mingolla E. View-invariant object category learning: How spatial and object attention are coordinated using surface-based attentional shrouds [J]. Cognitive Psychology,2009,58:1-48.
6. Chikkerur S, Serre T, Tan C, et al. What and where: A Bayesian inference theory of attention [J]. Vision Research,2010.
7. Isabelle M, Robert M S. Effects of contrast and size on orientation discrimination [J]. Vision Research, 2004, 44:57-67.
8. Egeth H E, Yantis S.Visual attention: control, representation, and time course [J]. Annual Review of Psychology, 1997, 48:269-297.
9. Deco G, Pollatos O, Zihl J. The time course of selective visual attention: Theroy and experiments [J]. Vision Research, 2002, 42:2925-2945.

10. Treue S, Trujillo J C M. Feature-based attention influences motion processing gain in macaque visual cortex [J]. Nature, 1999, 399(6736):575-579.
11. Saenz M, Buracas G T, Boynton G M. Global effects of feature-based attention in human visual cortex [J]. Nature Neuroscience, 2002, 5(7):631-632.
12. Treisman A M, Gelade G. A feature integration theory of attention [J]. Cognitive Psychology, 1980, 12(1):97-136.
13. Treisman A. Feature binding, attention and object perception [J]. Philosophical Transactions of the Royal Society, Series B, 1998, 353:1295-1306.
14. Scholl B J. Objects and attention: the state of the art [J]. Cognition, 2001, 80:1-46.
15. Scholl B J, Pylyshynb Z W, Feldman J. What is a visual object-evidence from target merging in multiple object tracking [J]. Cognition, 2001, 80:159-177.
16. Christof Koch. The Quest for consciousness [M]. Shanghai science & technology press,2012.
17. Judd T, Ehinger K, Durand F, and Torralba A. Learning to predict where humans look [C]. In IEEE International Conference on Computer Vision, 2009.
18. http://people.csail.mit.edu/tjudd/WherePeopleLook/index.html.
19. Koch C, Ullman S. Shifts in Selective Visual Attention:Towards the Underlying Neural Circuitry [J]. Human Neurobiology. 1985, 4 (4): 219–227.
20. Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Trans Pattern Anal Mach Intell, 1998;20(11):1254–9.
21. Walther D, Koch C. Modeling attention to salient proto-objects [J]. Neural Networks,2006,19:1395-1407.
22. Navalpakkam V, Itti L. An integrated model of top-down and bottom-up attention for optimizing detection speed [C]. 2006.
23. Hou X, Zhang L. Saliency detection: A spectral residual approach [C]. In IEEE Conf. Computer Vision and Pattern Recognition. 2007.
24. Li J, Levine D, An X, et al. Visual saliency based on scale space analysis in the frequency domain [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence. 2012.
25. Guo C, Zhang L. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression [J]. IEEE Trans. Image Processing. 2010, 19 (1): 185–198.
26. Bruce N, Tsotsos J. Saliency Based on Information Maximization [C]. In Advances in Neural Information Processing Systems. 2006.
27. Gao D, Han S, Vasconcelos N. Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition [J]. IEEE Trans. Pattern Analysis and Machine Intelligence. 2009: 989–1005.
28. Harel J, Koch C, Perona P. Graph-Based Visual Saliency [J]. In Advances in Neural Information Processing Systems. 2007.
29. Li J, Tian Y, Huang T, Gao W. Probabilistic Multi-Task Learning for Visual Saliency Estimation in Video [J]. International journal of computer vision, 2010, 90(2): 150-165.
30. Judd T, Ehinger K, Durand F, and Torralba A. Learning to predict where humans look [C]. In IEEE International Conference on Computer Vision, 2009.