Proceedings of the 2nd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/104/20241194

Integrating Vision: From Neuroscience to Artificial Intelligence

Haoshan Ye

College of Liberal Arts, Wenzhou Kean University, Wenzhou, China

1162284@wku.edu.cn

Abstract. This review article explores the complex interplay between neuroscience and artificial intelligence, focusing on vision processing and its applications. It starts by outlining the biological basis of vision, delving into how visual information is processed and encoded in the brain. The discussion then transitions to artificial intelligence, particularly machine vision, highlighting the advancements and technologies that mimic biological processes. A critical analysis compares how visual information is utilized in both biological organisms and artificial systems, with an emphasis on cognitive functions and neural encoding. The challenges of integrating vision across various sensory modalities are examined, underscoring the technological and cognitive limitations currently faced. The review culminates by identifying potential research paths aimed at closing the gap between neuroscience and AI. This involves enhancing the understanding and functionality of vision in multisensory contexts, striving to foster a more comprehensive approach to artificial intelligence that mirrors the complexity of human perception.

Keywords: Neuroscience, machine vision, cognitive functions, neural encoding, visual systems.

1. Introduction

The complexity and diversity of the material universe necessitate a multisensory approach to human perception, with vision playing a pivotal role. The human brain, akin to a sophisticated biological computer, integrates various sensory inputs to construct a subjective model of the real world, facilitating adaptation to dynamic environments. Vision is crucial, contributing to approximately 80% of the sensory information processed by the visual cortex [1]. Understanding how the brain encodes visual information is not only vital for comprehending human perception but also for advancing the theoretical foundations of consciousness.

Recent research has bridged the gap between neuroscience and artificial intelligence (AI), particularly through the development of technologies such as convolutional neural networks (CNNs). These advancements are inspired by the biological mechanisms of the visual cortex, notably the receptive fields that filter sensory input to aid in complex decision-making processes [2]. The integration of sensory information through neural systems underscores the importance of vision in both human cognition and the development of intelligent machines, marking a significant leap in both fields.

This review systematically explores the transformation from neural mechanisms of vision to AI algorithms that enhance machine perception and decision-making. It discusses the influence of visual processing on the formation of human consciousness and its simulation within AI systems, which could

DOI: 10.54254/2755-2721/104/20241194

revolutionize how machines perceive and interact with the world. Additionally, it examines the role of vision in multisensory and multimodal learning, highlighting how AI can benefit from mimicking human-like perception to develop more intuitive models that understand and interpret the complex nature of real-world environments.

2. Theoretical Foundations

2.1. Visual and neural

Human vision is a function with a complex nervous system and sensor underpinnings. First of all, the first step for humans to acquire visual signals is the eye. An eyeball basically consists of the cornea, pupil, iris, lens and the retina at the back of the eye.

The cornea, iris, pupil and lens are primarily used to collect and adjust the light entering the eye, but it is only after light has activated the photoreceptor cells in the retina where the light signals are converted into electrical signals and thus truly enter the entire neural network. The nervous system responsible for vision consists primarily of the retina, lateral geniculate nucleus (LGN), and visual corte. After the retina converts light signals into electrical, or neural, signals, these signals pass sequentially through the major visual pathways composed by the LGN and primary visual cortex (V1). In the order of neural signals transmitting, the secondary visual cortex (V2) comes next, where the visual signals will be further filtered and shunted to be passed on to more complex processing in the (parallel) visual cortex areas of V3 through V6.

2.2. Advancements in machine vision

Machine vision used to be a blueprint for when computers were invented, and that was to equip machines with visual capabilities close to human vision. However, this ambition has not been fully implemented until today. In recent years, with the development of artificial intelligence, machine vision has actually been interpreted more as the capability to “understand images”.

Much of the advancements acquired in machine vision are related to the development of artificial intelligence technology. The reduction of function is often based on the reduction of structure. The definition and development of machine vision derives from human vision, therefore, following this logic, it is inevitable to build human-like neural networks for machines. Artificial Neural Networks (ANN) are deep learning algorithmic models used in machine learning to simulate human neural structures. The ANN is structurally viewed as a structure similar to the human cortex. The input layer is where the data enters the network and then passes into one or more hidden layers, through which the data will flow through the entire algorithmic network to the output layer, which generates the network predictions [3]. Deep learning machine model is usually a supervised learning process, which requires labeling the data to explicitly tell the network what the data represents. Whereas the training of ANN model requires example images to be provided to the network. Trained on a large number of examples that have been labeled in advance with the type of output expected, the neural network will be able to give a prediction for a brand-new input image, in other words, which classification to output to.

DNN: Deep Neural Network (DNN) It is similar to ANN in that it also consists of three parts: an input layer, an intermediate network, and an output layer. But, unlike an ANN, the middle part of a DNN can contain multiple fully connected layers, or various techniques or mathematical formulas can be used to shape each layer of the intermediate network. A DNN can be considered as a type of ANN but not all DNNs are ANNs. if there are multiple hidden layers in an ANN model then it can also be recognized as a DNN. Hence a neural network structure can be called as a deep net as long as it has a sufficient number of hidden units in the middle layer in other words multiple layers of perceptions.

CNN: CNN is a special construction based on ANN and DNN, its structure consists of three layers: convolutional layer, pooling layer and fully connected layer. The architectural inspiration for CNNs is also largely related to the human optic nervous system; the structure of CNNs allows the entire network, like human vision, to perform rapid feature extraction and classification upon input. This allows CNNs

DOI: 10.54254/2755-2721/104/20241194

to navigate through large amounts of pixel information at once. Compared to DNNs, CNNs have a much better performance in classification tasks where multiple categories of inputs are detected at once.

3. System Analysis and Application Insights

3.1. Encoding of visual information

The visual system does an extraordinary level of coding of the visual signals that humans receive from the outside world, and as a result, this information-processing mechanism has brought humans a near-perfect image information-processing system [4]. Beginning with the major visual pathways composed of the retina, LGN, and V1, the visual system begins to categorize and process light and neural signals.

Sensory cells in the retina first have different preferences for different wavelengths of light, and at the same time the inner circle of the receptive field (RF) of ganglion cells in the retina has very different electrical feedbacks compared to the outer circle [5]. Simultaneously, research has shown that developing a preference for visual information in a particular direction occurs in the visual cortex of primates. A RF structure is also present in layer v1 similar to that in the retina. However, unlike the retina, the simple and complex cells in V1 present a preference for specific orientations. Also, the spatial connectivity of synapses in this structure may be related to the final orientation preference presented. This process does not occur in the retina, and there is no apparent directional preference in the transmission of neural signals from the LGN. This process does not occur in the primate retina, and although there are similar RF structures in the LGN, the neural signals that pass through the LGN during the frontal process of one-to-one transmission to V1 are also not pre-processed and shunted because of orientation preferences.

When the information comes to the V2 layer, it is further separated, and the cells in V2 are selective for corners. The research demonstrates that V2 is able to modulate V1's response to specific stimuli to some extent, and that this modulation comes from a top-down feedback channel [6, 7].

Throughout the primary visual pathway to V2, information is encoded in layers, and at the same time, the amount of information changes, and recent research points to this change as an inevitable degradation, i.e., a reduction in the amount of information, which is also in line with energy metabolism and transmission loss. During the encoding process from V1 to V2, shape, color, motion, and stereopsis are separated, and the information that passes through the V2 layer is sent to the subsequent different but parallel visual cortex for further processing and elaboration and eventually vision.

3.2. Vision's role in cognitive formation

Vision is essential in both memory and attention in human beings. In recent years some scientists have proposed a new view that the nervous system is task-oriented rather than sensory-based [8]. According to the idea of this perspective, the human cognitive system works primarily to achieve a particular task, while the role of the sensory system is secondary. Under this assumption, what task needs to be accomplished is an important prerequisite for determining what kind of conscious activity humans will engage in. The importance of attention in this section is undoubtedly significant, and the one thing that focuses most of a human's attention will generally be the most important task to be solved at the moment. Thus, vision, as the main information acquisition channel for humans, somewhat dominates the allocation of attention, which means determining the main task of the present.

As mentioned earlier, human vision is a perfect model for processing. Vision is capable of receiving information in a filtered manner and allocating attention to special important objects and ignoring irrelevant information. This ability is visual attention. This attention tends to come from two directions, bottom-up and top-down [9]. Bottom-up attention arises from the unconscious attraction of particular events occurring in the real environment for attention, whereas top-down attention arises from the direct allocation of higher brain regions. This bottom-up attention is then related to the screening process in the nervous system from the main pathway of vision to V2, where the visual system elicits a stronger neural response initially for specific colors, shapes, and states of motion. In contrast, attentional top-down allocation choices occur after vision formation, from decisions in prefrontal and parietal regions

DOI: 10.54254/2755-2721/104/20241194

[10, 11]. Bottom-up and top-down attention are essentially parallel at the same time, which means that working memory and attention allocation processes are difficult to separate in human cognitive processes and, meanwhile, interact very closely. Thus, human cognitive-consciousness processes are largely dependent on vision.

Noteworthy is the fact that if irreversible damage to the visual receptors results in a negative disturbance of the visual signal input, this perturbation will affect any part of the entire process involved in the visual system [12, 13]. The obstacles that come from prolonged exposure to such negative influences are about cognitive, motor and even a person's social development.

3.3. Applications of machine vision

3.3.1. Special object recognition. Machine vision is widely used in the task of recognizing a given special target [14]. Machine vision-enabled cameras can help humans monitor the survival of rare species in wild environments where it is inconvenient to have a permanent presence. Alternatively, machine vision can help humans quickly screen eligible individuals in animal husbandry, such as a Chinese research study that efficiently identified Chengdu Ma Goats in a herd through machine vision methods [15].

3.3.2. Automatic driving. Machine vision is also widely used in intelligent autonomous driving. Machine vision can be used in collision avoidance systems for recognizing other vehicles or objects and determining the distance from the other vehicle or object to its own vehicle body. Meanwhile, machine vision can also be used to recognize lanes, such as in recent research where a new lane detection scheme was proposed. This scheme is based on image data and uses Kirsch's algorithm to detect edges before using the VLF method to detect specific lane locations, which allows the system to obtain more accurate judgments [16].

3.3.3. Human movement recognition (HMR). The recognition of human motion is a very prospective topic. With the development of artificial intelligence and the maturity of machine vision, automatic interpretation of human motion is becoming increasingly necessary. Recognition of human movement is based in part on the recognition and tracking of human body parts, which can be categorized as gestures, actions, interactions, and group activities according to the parts and complexities involved in the movement [17]. This technology can also be applied to robots, and in the latest research, scientists attempted to enable two robots to interact with each other through gesture recognition [18]. Further, the robots can directly perform motion correction, state synchronization, etc. through HMR technology in order to achieve automated group cooperation of robots.

3.4. Integrating vision in multisensory and multimodal contexts

In the real world, visual information is rarely processed in isolation. The ability to integrate vision with other senses (e.g., auditory, tactile, and proprioceptive information) is critical for accurate perception and robust decision-making. According to the previously mentioned visual attention, as the sense that receives the largest amount of information, vision in many cases guides the process of receiving information by the other senses. Concurrently, vision also guides the brain to modify information from other receptors with preferences during multisensory information reception.

The impact of vision on multimodal integration in the field of artificial intelligence is something that needs to be discussed according to the application scenarios [19].

4. Challenges and Limitations

4.1. Neuroscientific boundaries

Despite significant progress in understanding the neural mechanisms behind vision, many aspects of visual perception remain largely unexplored. Since it is still not possible at this stage to fully explain the

DOI: 10.54254/2755-2721/104/20241194

functioning of the various brain regions and their specific functional counterparts, the support that neuroscience can directly provide for the development of machine vision systems is still limited.

4.2. Debating consciousness: scope and limitations

The role of perception, such as vision, in consciousness has been debated in psychology, neuroscience and philosophy. Although it has been argued that consciousness is task-directed rather than sensory-oriented, it is undeniable that the senses remain an integral part of conscious activity. While modern science is still unable to accomplish the deconstruction of the nervous system, there is also a lack of robustness in trying to completely explain human conscious activity. Thus, although in the course of research directed at the visual system, scientists have found some correspondence between neural structures and conscious activity, the relationship still suffers from verification incompleteness. At the same time, based on technical means and scientific ethical and moral limitations, some more in-depth research on the neural function of the human brain and the activity of consciousness can only be verified by carrying out on experimental animals.

4.3. Contrasting machine and human visual systems

Machine vision systems have made significant advances with in-depth research in neuroscience and the development of AI algorithms, but machine vision is still not as robust, flexible and versatile as human vision [20, 21]. One reason for this is that the human neural networks that serve as a reference for programming machine vision algorithms have not yet been fully parsed, and another is that it is not well understood how humans can use silicon-based ones to restore the function of their own carbon-based organs.

However, the established machine vision model can provide a method for quantitative analysis to reverse-validate human vision. For instance, CNN models are able to explain to some extent the degradation mechanism of neural signals as they are transmitted in layer V1 [22, 23]. Human vision and consciousness can guide the development of machine vision from the top down, and conversely models of machine vision can help scientists to further simulate, validate and explain human vision and consciousness [24].

5. Conclusion

This article has systematically explored the crucial role of vision in both neuroscience and artificial intelligence, illustrating its profound influence on human cognition and the development of intelligent systems. The integration of visual information processing from the biological perspective has significantly propelled advancements in AI, particularly in enhancing machine vision capabilities. By analyzing the neural mechanisms that underlie human visual perception, this work has illuminated the pathways through which sensory data is transformed into cognitive recognition and decision-making processes, offering a template for refining AI models.

The potential directions for future research are manifold. There is a pressing need to delve deeper into the uncharted territories of visual neuroscience to uncover more intricate details about sensory integration and processing. Such explorations could yield more sophisticated AI systems that mimic human perceptual and cognitive functions with greater fidelity. Furthermore, bridging the gap between the computational models of machine vision and the neurobiological processes of human vision promises to not only enhance the functionality and adaptability of AI but also to provide insights into the very nature of human consciousness and perception. These efforts will require interdisciplinary collaborations that merge insights from neuroscience, cognitive science, and computer engineering to pioneer innovations in both theoretical understanding and practical applications.

References

[1] Benson NC, Kupers ER, Barbot A, Carrasco M, Winawer J 2021 ELife 10 https://doi.org/10.7554/elife.67685

DOI: 10.54254/2755-2721/104/20241194

[2] Bowers JS, Malhotra G, Dujmović M, Montero ML, Tsvetkov C, Biscione V, Puebla G, Adolfi F, Hummel JE, Heaton RF, Evans BD, Mitchell J, Blything R 2022 Behav. Brain Sci. 1–74 https://doi.org/10.1017/s0140525x22002813
[3] Chokron S, Dutton GN 2022 J. Neural Transm. 130 409–424 https://doi.org/10.1007/s00702-022-02572-8
[4] Cox D, Dean T 2014 Curr. Biol. 24(18) R921–R929 https://doi.org/10.1016/j.cub.2014.08.026
[5] de Sousa AA, Todorov OS, Proulx MJ 2022 Neurosci. Biobehav. Rev. 134 104550 https://doi.org/10.1016/j.neubiorev.2022.104550
[6] Federer F, Ta'afua S, Merlin S, Hassanpour MS, Angelucci A 2021 Nat. Commun. 12(1) https://doi.org/10.1038/s41467-020-20505-5
[7] Geirhos R, Narayanappa K, Mitzkus B, Thieringer T, Bethge M, Wichmann FA, Brendel W 2021 NeurIPS Curran Associates https://proceedings.neurips.cc/paper/2021/hash/c8877cff22082a16395a57e97232bb6f-Abstract.html
[8] Jegham I, Ben Khalifa A, Alouani I, Mahjoub MA 2020 Forensic Sci. Int. Dig. Invest. 32 200901 https://doi.org/10.1016/j.fsidi.2019.200901
[9] Juang LH 2024 Multimedia Tools Appl. https://doi.org/10.1007/s11042-023-17989-w
[10] Li J, Ataman D, Sennrich R 2021 ArXiv Cornell Univ. https://doi.org/10.48550/arxiv.2109.03415
[11] Lindsay G 2020 J. Cogn. Neurosci. 33(10) 1–15 https://doi.org/10.1162/jocn_a_01544
[12] Lockhofen DEL, Mulert C 2021 Front. Neurosci. 15 https://doi.org/10.3389/fnins.2021.643597
[13] Pennartz CMA, Dora S, Muckli L, Lorteije JAM 2019 Trends Neurosci. 42(9) 589–603 https://doi.org/10.1016/j.tins.2019.07.005
[14] Peters B, Kriegeskorte N 2021 Nat. Hum. Behav. 5(9) 1127–1144 https://doi.org/10.1038/s41562-021-01194-6
[15] Pu J, Yu C, Chen X, Zhang Y, Yang X, Li J 2022 Animals 12(14) 1746 https://doi.org/10. 3390/ani12141746
[16] Rossi LF, Harris KD, Carandini M 2020 Nature 588(7839) 648–652 https://doi.org/10. 1038/s41586-020-2894-4
[17] Sagar V, Prabhakar CJ 2022 https://doi.org/10. 1109/discover55800.2022.9974938
[18] Smith ML, Smith LN, Hansen MF 2021 Comput. Ind. 130 103472 https://doi.org/10. 1016/j.compind.2021.103472
[19] Tuli S, Dasgupta I, Grant E, Griffiths TL 2021 ArXiv https://doi.org/10.48550/arXiv.2105.07197
[20] Kim YJ, Peterson BT, Crook JD, Joo HR, Wu J, Puller C, Robinson FR, Gamlin PD, Yau K, Viana F, Troy JB, Smith RG, Packer OS, Detwiler PB, Dacey DM 2022 Nat. Commun. 13(1) https://doi.org/10.1038/s41467-022-30405-5
[21] Yuille AL, Liu C 2020 Int. J. Comput. Vis. 129(3) 781–802 https://doi.org/10.1007/s11263-020-01405-z
[22] Zhong H, Wang R 2020 Cogn. Neurodyn. 15(2) 299–313 https://doi.org/10.1007/s11571-020-09599-1
[23] Zhong H, Wang R 2021a Adv. Cogn. Neurodyn. 251–251 https://doi.org/10.1007/978-981-16-0317-4_30
[24] Zhong H, Wang R 2021b Nonlinear Dyn. 105(4) 3551–3570 https://doi.org/10.1007/s11071-021-06648-0