(IJAER) 2021, Vol. No. 22, Issue No. IV, October **e-ISSN: 2231-5152, p-ISSN: 2454-1796**

A COMPREHENSIVE REVIEW OF THE APPLICATION OF COMPUTER VISION BY LEVERAGING MACHINE LEARNING TECHNIQUES

Shreya Bhardwaj

Panjab University, Chandigarh, India

ABSTRACT

This paper conducts a scientific investigation into the significance of machine learning and its applications to computer vision. Tasks like computer vision and natural language processing have become extremely fast and accurate due to recent advancements in artificial intelligence, deep learning, computing resources, and the availability of large training datasets. As a result, artificial intelligence is a hot topic in computing. In artificial intelligence, deep learning is a subfield of machine learning. Can effectively use machine learning techniques to complete tasks related to image processing. As a result, complex images will be better-understood thanks to machine learning. Computer vision is related to object recognition, tracking, and detection. Computer vision advances significantly in this field thanks to convolutional neural network-based algorithms like YOLO and R-CNN. Pattern recognition algorithms based on machine learning models are very good, but they typically need a lot of computing power and huge data sets.

In computer vision, algorithms for real-time object detection will be very important. In most cases, a graphics processing unit is required for the neural network to accelerate the execution of machine learning models. This review paper briefly discusses the real-time object detection and machine learning algorithms developed by various researchers worldwide. In addition, the various techniques used to identify a specific image object are examined in this paper.

INTRODUCTION

The visual system of the human body has always been crucial to human life. The most significant discoveries humans have made can be attributed to human vision. The most important sense that humans have is vision. The human visual system is extremely complex, quick, and precise. It enables us to carry out strenuous activities like driving, learning and instructing. As a result, progress in the field of computer vision began to accelerate. Digital image processing employs computer algorithms to process images using digital computers. Computer vision is a part of image processing. Different computer algorithms are used to carry out the image processing task in digital image processing. Beginning with image acquisition, intricate digital image processing algorithms perform feature extraction, classification, object recognition, and pattern recognition. The accuracy of systems and image processing tasks will benefit from machine learning. The machine can derive a high-level understanding from digital images or videos, similar to how humans perceive the world, thanks to advances in computer vision. The goal of this development is to automate human-assisted tasks. When we look at an image, we immediately understand what it is, where the objects are, and how we can interact with them. Understanding the content of videos and digital images is the ultimate objective of computer vision systems. This includes methods for getting, analyzing, processing, and removing relevant data from the real world. In the past few decades, numerous academics have studied and developed computer vision, which has improved rapidly. Machine learning's fastest-growing subfield is deep learning. It is made up of multiple artificial neural network layers that are hidden. Artificial intelligence has been significantly influenced by developments in deep learning in various fields. How deep learning algorithms have been used in image processing and other computing fields is presented chronologically in this review paper. This review paper provides a general overview of machine learning, its application in computer vision, and its development over the past few decades.

LITERATURE REVIEW

Review the text by Konrad Ahlin, Benjamin Joffe, and others. A paper titled "Autonomous Leaf Picking Robots Using Deep Learning and Visual-Servoing" was published in 2016 [18]. They have developed an autonomous leaf-picking robot using the deep learning model. Convolutional Neural Networks and Monoscopic Depth Analysis are used in image processing and visual servoing, respectively, by this robot's camera sensor. This robotic agent could precisely locate and grasp a plant leaf in an unstructured environment. The robot only used one camera for identification in this experiment. The object detection algorithm picked up at least one leaf for each frame in which the plant was present. The camera's images had a resolution of 1280 by 720. On an NVIDIA GTX 880M GPU, tested the algorithm's computational performance. Each frame took an average of 0.7 seconds to process.

Po-Yu, Liu, et al., A paper on deep learning-based image reconstruction was published in 2018. When the noise is very high, this paper suggests a deep learning model that can statistically improve Poisson image de-noising significantly over conventional algorithms. The demand for an advanced Poisson image de-noising algorithm has increased due to the expanding camera market. The authors' proposed architecture includes a hybrid of convolutional and deconvolutional neural network layers with symmetric connections. In tests [19], the authors' de-noising system outperformed standard traditional algorithms by 0.38 dB, 0.68 dB, and 1.04 dB on average in PSNR. In addition, this de-noising network can outperform the benchmark algorithm while consuming less computational power. This network only has six layers because the computer can do calculations. Even though it can learn the parameters from the data alone, this network can still perform Poisson de-noising without being explicitly taught the noise characteristics.

Shahein Tajmir, Hyunkwang Lee, et al. A convolutional neural network-based, fully automated deep learning system for bone age assessment was the subject of a 2017 paper. The authors developed a deep learning system to automatically segment radiographs of the hand and wrist and carry out automated bone age assessments. The developed bone age assessment model uses a refined convolutional neural network to generate radiology reports with a deep learning system. The trained model's features to assess bone age are shown in the attention maps generated using the input occlusion method. For both genders, the proposed design achieves 57.32 and 61.40% accuracies using a fine-tuned CNN and an Image-Net dataset that has already been trained [20]. The trained algorithm looks at similar parts of the hand and wrist to determine bone age. As a decision-supporting system, a bone age assessment model that is completely automated can be used in the clinical setting. The BAA model, developed by authors using machine learning, is more accurate and effective than the standard approach. Can use machine learning in various computer vision applications, including robotics, medical diagnosis, education, electronic trading platforms, transportation, business decisions, and remote sensing. The applications of artificial neural networks in computing and data science are numerous.

Object detection and recognition (e.g., [7], [8], object motion tracking (e.g., [9], [10]) action recognition (e.g., [11], [12]) and image segmentation (e.g., [13], [14]), in which deep learning excels, are examples of machine learning applications in computer vision. The term "deep learning" refers to abstract layer evaluation and hierarchical approaches to problem-solving.

One way to get better output and speed up the processing process is through deep learning. In the last ten years, natural language processing (NLP) and deep learning techniques for pattern recognition have been utilized in machine vision. It would be extremely difficult to discuss every neural network application, given the scope of this paper. However, can also apply it to several real-world issues. As a result, real-time object detection systems and convolutional neural networks are our primary areas of interest. The development of computer vision applications has accelerated due to the emergence of new machine learning algorithms.

A. Autonomous vehicle development is greatly facilitated by computer vision. By 2030, McKinsey projects that 15% of all cars sold will be autonomous. Traffic accidents cause 2.2% of global deaths each year. Autonomous vehicles of this kind can significantly reduce traffic accidents. An autonomous vehicle is used to improve traffic flow in transportation. Computer vision enables us to make traffic management procedures smarter and more effective. We can see a paradigm shift in the automotive industry, making travelling extremely safe with advanced computer vision technology.

B. Medicine

In the last ten years, deep learning has been used in medical applications. For instance, the utilization of deep learning algorithms in neural networks results in superior performance for detecting and predicting diseases. Processed an MRI image of human organs for disease detection. Even though these methods could benefit from the success of this deep learning technique, unfortunately, the current approach to supervised machine learning is constrained by the need for training datasets, high computational resources, and high-quality training data like MRI images. Can overcome this by providing more powerful GPUs and better training data sets for specific applications. The deep learning algorithms in optical coherence tomography (OCT) are achieving remarkable results. Images can typically be thought of as the manual creation of convolutional matrices [15]. A framework can predict and detect retinal pathologies using deep learning, lowering OCT costs. In addition, diabetic retinopathy can be detected using CNN on retinal images. The machine-learning method detects and predicts more accurately than most well-known ophthalmologists. It can also utilize machine learning models in drug research by predicting molecular structure and simulating properties like toxicity and the capacity to bind to other molecules. Additionally, it can simulate isolated chemical or biological processes without the need for costly software, and the system as a whole is significantly faster than conventional methods.

CONCLUSIONS

The present work provides a comprehensive overview of previous research on machine learning and its application to computer vision. In addition, this paper discusses recent developments in real-time object detection, object recognition, and deep learning by several researchers over the past ten years. Object detection, pattern recognition, semantic segmentation, action recognition, and face recognition are just a few visual tasks that can accomplish with the help of computer vision methods. In particular, general-purpose object recognition and detection should be quick, accurate, and capable of recognizing many objects. YOLO is a deep convolutional neural network-based real-time object detection algorithm. YOLO's real-time object detection algorithm is comparable in speed and accuracy.

REFERENCES

[1] Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi (2016), You Only Look Once: Unified, Real-Time Object Detection University of Washington, Allen Institute for AI, *Facebook AI Research*, arXiv: 1506.02640v5 [cs.CV] 9 May 2016.

[2] Sumit Das and Aritra Dey (2015), Applications of Artificial Intelligence in Machine Learning: Review and Prospect, *International Journal of Computer Applications* (0975 – 8887) Volume 115 – No. 9, April 2015.

[3] https://towardsdatascience.com/

[4] W. S. McCulloch and W. Pitts, (1943), A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biology*, vol. 5, no. 4, pp. 115–133, 1943.

[5] Subhashini Venugopalan, Marcus Rohrbach and others (2015), Sequence to Sequence – Video to Text, arXiv:1505.00487v3 [cs.CV] 19 Oct 2015.

[6] R. Girshick, Jeff Donahue, Trevor Darrell and Jitendra Malik (2015), Fast R-CNN, arXiv preprint arXiv: 1504.08083, 2015.

[7] Shaoqing Ren and R. Girshick, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, arXiv: 1506.01497 [cs.CV].

[8] M. Everingham, L. Van Gool, C. K. I.Williams, J.Winn, and A. Zisserman. (2007), The PASCAL Visual Object Classes Challenge 2007 (VOC2007) *International Journal of Computer Vision*, 2007.

[9] M. Everingham, L. Van Gool, C. K. I.Williams, J.Winn, and A. Zisserman. (2010), The PASCAL Visual Object Classes Challenge 2010 (VOC2010) *International Journal of Computer Vision*, 2010.

[10] P. Dollár, C. Wojek, B. Schiele and P. Perona (2012), Pedestrian Detection: An Evaluation of the State of the Art PAMI, *IEEE Transactions on Pattern Analysis and Machine Intelligence.*

[11] Paul Viola and Michael J Jones. (2004), Robust real-time face detection, *IJCV*, 57(2):137–154, 2004.

[12] C. Chen, A. Seff, A. Kornhauser and J. Xiao , (2015), Deep Driving: Learning Affordance for Direct Perception in Autonomous Driving, *15th IEEE International Conference on Computer Vision ICCV*, 2015.

[13] J. E. Hoo, and K. C. Lim. (2016), Accuracy and Error Study of Horizontal and Vertical Measurements with Single View Metrology for Road Surveying, *ARPN Journal of Engineering and Applied Sciences*. 11(12):7872-6, 2016.

[14] Yonglong Tian, Ping Luo, Xiaogang Wang, and Xiaoou Tang, (2015), Pedestrian detection aided by deep learning semantic tasks, In *CVPR*, 2015.

[15] Lee CS, Tyring AJ, Deruyter NP, Wu Y, Rokem A, Lee AY. (2017), Deep-learning based automated segmentation of macular edema in optical coherence tomography, *Biomed Opt Express*. 2017;8(7):3440-3448. Published 2017 Jun 23. doi:10.1364/BOE.8.003440.

[16] https://www.forbes.com/sites/bernardmarr/2019/04/08/7

[17] Marra, Francesco & Poggi, Giovanni & Sansone, Carlo & Verdoliva, Luisa (2017), A Deep Learning Approach for Iris Sensor Model Identification. Pattern Recognition, *Letters*. 113. 10.1016/j.patrec.2017.04.010.

[18] Ahlin, Konrad et al. (2016), Autonomous Leaf Picking Using Deep Learning and Visual-Servoing. *IFAC-Papers On Line* 49 (2016): 177-183.

[19] Liu, Po-Yu, and Edmund Y. Lam. (2018), Image reconstruction using deep learning, arXiv preprint arXiv: 1809.10410 (2018).

[20] Lee, H., Tajmir, S., Lee, J. et al. (2017), Fully Automated Deep Learning System for Bone Age Assessment. *J Digit Imaging* 30, 427–441(2017).

[21] Paul Viola and Michael Jones (2001), Rapid Object Detection using a Boosted Cascade of Simple Features, *Accepted Conference on Computer Vision and Pattern Recognition 2001*

[22] Dalal and B. Triggs. (2005), Histograms of oriented gradients for human detection, *In CVPR.*

[23] Alex Krizhevsky, Ilya Sutskever ad Geoffrey E. Hinton (2012), ImageNet Classification with Deep Convolutional Neural Networks, *Advances in Neural Information Processing Systems 25 (NIPS 2012)* January 2012.

[24] R. Girshick, J. Donahue, T. Darrell, and J. Malik. (2014), Rich feature hierarchies for accurate object detection and semantic segmentation, *In CVPR*.

[25] Uçar, Ayşegül, et al. (2017), Object Recognition and Detection with Deep Learning for Autonomous Driving Applications. *Simulation,* vol. 93, no. 9, Sept. 2017, pp. 759–769, doi: 10.1177/0037549717709932.

[26] J. Long, E. Shelhamer, and T. Darrell (2015), Fully convolutional networks for semantic segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15),* pp. 3431–3440, IEEE, Boston, Mass, USA, June 2015.

[27] S. Ren, K. He, R. Girshick, and J. Sun. (2015), Faster R-CNN: Towards real-time object detection with region proposal networks, arXiv: 1506.01497, 2015.

[28] W. S. McCulloch and W. Pitts, (1943), A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biology,* vol. 5, no. 4, pp. 115–133, 1943.

[29] Mxnet, Available online: http://mxnet.io.

[30] W. Ouyang, X. Zeng, X. Wang et al., Deep ID-Net: Object Detection with Deformable Part Based -amazing-examples-of-computer-and-machine-vision-inpractice/# 4da402eb1018.

[31] A. Diba, V. Sharma, A. Pazandeh, H. Pirsiavash, and L. V. Gool, (2017), Theyakly Supervised Cascaded Convolutional Networks, in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5131–5139, Honolulu, HI, July 2017.

[32] N. Doulamis and A. Voulodimos, (2016), FAST-MDL: Fast Adaptive, Supervised Training of multi-layered deep learning models for consistent object tracking and classification, in *Proceedings of the 2016 IEEE International Conference on Imaging Systems and Techniques, IST 2016*, pp. 318–323, October 2016.

[33] N. Doulamis, (2017), Adaptable deep learning structures for object labelling/tracking under dynamic visual environments, *Multimedia Tools and Applications*, pp. 1–39, 2017.

[34] L. Lin, K.Wang,W. Zuo, M.Wang, J. Luo, and L. Zhang, (2016), A deep structured model with radius-margin bound for 3D human activity recognition, *International Journal of Computer Vision*, vol. 118, no. 2, pp. 256–273, 2016.

[35] S. Cao and R. Nevatia, (2016), Exploring deep learning based solutions in fine grained activity recognition in the wild, in *Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 384–389, Cancun, December 2016.

[36] A. Toshev and C. Szegedy, (2014), Deep Pose: Human pose estimation via deep neural networks, in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014*, pp. 1653–1660, USA, June 2014.

[37] H.Noh, S.Hong, and B.Han, (2015), Learning de convolution network for semantic segmentation, *in Proceedings of the 15th IEEE International Conference on Computer Vision, ICCV 2015,* pp. 1520–1528, Santiago, Chile, December 2015.

[38] J. Long, E. Shelhamer, and T. Darrell, (2015), Fully convolutional networks for semantic segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15),* pp. 3431–3440, IEEE, Boston, Mass, USA, June 2015.

[39] K. Fukushima, (1980), Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, *Biological Cybernetics*, 1980.

[40] P. Arbel´aez, B. Hariharan, C. Gu, S. Gupta, L. Bourdev, and J. Malik., (2012), Semantic segmentation using regions and parts In *CVPR*, 2012.

[41] H. A. Rowley, S. Baluja and T. Kanade. (1998), Neural network based face detection, *TPAMI.*

[42] P. Sermanet, K. Kavukcuoglu, S. Chintala, and Y. LeCun. (2013), Pedestrian detection with unsupervised multi-stage feature learning". In *CVPR*

[43] K. Simonyan and A. Zisserman.(2014), Very Deep Convolutional Networks for Large-Scale Image Recognition". arXiv preprint, arXiv:1409.1556.

[44] R. Girshick, P. Felzenszwalb, and D. McAllester. "Discriminatively trained deformable part models".

[45] Tai Sing Lee and David Mumford (2003), Hierarchical Bayesian inference in the visual cortex 1434 J. Opt. Soc. Am. A/Vol. 20, No. 7/July 2003.

[46] https://en.wikipedia.org/wiki/Machine_learning

[47] Karpathy, Andrej, and Li Fei-Fei. (2015), Deep visual-semantic alignments for generating image descriptions. *Proceedings of the IEEE conference on computer vision and pattern recognition.*