Deep Learning for Computer Vision: A Brief Review

Ksheera R Shetty[1], Vaibhav S Soorinje[2], Prinson Dsouza[3], Swasthik[4]
Department of Computer Science and Engineering[1,2,3,4]
Alva's Institute Of Engineering and Technology, Mangalore, India

Abstract: *Over the last years deep learning methods have been shown to outperform previous state-of-the-art machine learning techniques in several fields, with computer vision being one of the most prominent cases. This review paper provides a brief overview of some of the most significant deep learning schemes used in computer vision problems, that is, Convolutional Neural Networks, Deep Boltzmann Machines and Deep Belief Networks, and Stacked Denoising Autoencoders. A brief account of their history, structure, advantages, and limitations is given, followed by a description of their applications in various computer vision tasks, such as object detection, face recognition, action and activity recognition, and human pose estimation. Finally, a brief overview is given of future directions in designing deep learning schemes for computer vision problems and the challenges involved therein.*

Keywords: Convolutional Neural Networks, Stacked Denoising Autoencoders, Deep Belief Networks

REFERENCES

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

[2]. Y. LeCun, B. Boser, J. Denker et al., "Handwritten digit recognition with a back-propagation network," in Advances in Neural Information Processing Systems 2 (NIPS*89), D. Touretzky, Ed., Denver, CO, USA, 1990.

[3]. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[4]. G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," Neural Computation, vol. 18, no. 7, pp. 1527–1554, 2006.

[5]. TensorFlow, Available online: https://www.tensorflow.org.

[6]. B. Frederic, P. Lamblin, R. Pascanu et al., "Theano: new features and speed improvements," in Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop, 2012, http://deeplearning.net/software/theano/.

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

[8]. W. Ouyang, X. Zeng, X. Wang et al., "DeepID-Net: Object Detection with Deformable Part Based Convolutional Neural Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 7, pp. 1320–1334, 2017.

[9]. A. Diba, V. Sharma, A. Pazandeh, H. Pirsiavash, and L. V. Gool, "Weakly 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.

[10]. N. Doulamis and A. Voulodimos, "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.

[11]. N. Doulamis, "Adaptable deep learning structures for object labeling/tracking under dynamic visual environments," Multimedia Tools and Applications, pp. 1–39, 2017.

[12]. L. Lin, K. Wang, W. Zuo, M. Wang, J. Luo, and L. Zhang, "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.

[13]. S. Cao and R. Nevatia, "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.

[14]. A. Toshev and C. Szegedy, "DeepPose: 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.

[15]. X. Chen and A. L. Yuille, "Articulated pose estimation by a graphical model with image dependent pairwise relations," in Proceedings of the NIPS, 2014.

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

[17]. J. Long, E. Shelhamer, and T. Darrell, "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.

[18]. D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex," The Journal of Physiology, vol. 160, pp. 106–154, 1962.

[19]. K. Fukushima, "Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," Biological Cybernetics, vol. 36, no. 4, pp. 193–202, 1980.

[20]. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2323, 1998.

[21]. Y. LeCun, B. Boser, J. S. Denker et al., "Backpropagation applied to handwritten zip code recognition," Neural Computation, vol. 1, no. 4, pp. 541–551, 1989.

[22]. M. Tygert, J. Bruna, S. Chintala, Y. LeCun, S. Piantino, and A. Szlam, "A mathematical motivation for complex-valued convolutional networks," Neural Computation, vol. 28, no. 5, pp. 815– 825, 2016.

[23]. M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Is object localization for free? - Weakly-supervised learning with convolutional neural networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, pp. 685– 694, June 2015.

[24]. C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '15), pp. 1–9, Boston, Mass, USA, June 2015.

[25]. Y. L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in Proceedings of the ICML, 2010.

[26]. D. Scherer, A. Muller, and S. Behnke, "Evaluation of pooling¨ operations in convolutional architectures for object recognition," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface, vol. 6354, no. 3, pp. 92–101, 2010.

[27]. H. Wu and X. Gu, "Max-Pooling Dropout for Regularization of Convolutional Neural Networks," in Neural Information Processing, vol. 9489 of Lecture Notes in Computer Science, pp. 46– 54, Springer International Publishing, Cham, 2015

[28]. K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," in Computer Vision – ECCV 2014,vol.8691of Lecture Notes in Computer Science, pp. 346–361, Springer International Publishing, Cham, 2014.