[1]   S. Wu, S. Zhong, and Y. Liu, "Deep residual learning for image steganalysis," *Multimed. Tools Appl.*, 2017, doi: 10.1007/s11042-017-4440-4.

[2]   M. Sornam, K. Muthusubash, and V. Vanitha, "A Survey on Image Classification and Activity Recognition using Deep Convolutional Neural Network Architecture," *2017 9th Int. Conf. Adv. Comput. ICoAC 2017*, pp. 121–126, 2018, doi: 10.1109/ICoAC.2017.8441512.

[3]   S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," pp. 1–23, 2020, [Online]. Available: http://arxiv.org/abs/2001.05566.

[4]   A. Kumar, S. Verma, and H. Mangla, "A Survey of Deep Learning Techniques in Speech Recognition," *Proc. - IEEE 2018 Int. Conf. Adv. Comput. Commun. Control Networking, ICACCCN 2018*, pp. 179–185, 2018, doi: 10.1109/ICACCCN.2018.8748399.

[5]   Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object Detection with Deep Learning: A Review," *IEEE Transactions on Neural Networks and Learning Systems*. 2019, doi: 10.1109/TNNLS.2018.2876865.

[6]   D. W. Otter, J. R. Medina, and J. K. Kalita, "A Survey of the Usages of Deep Learning in Natural Language Processing," vol. XX, no. X, pp. 1–22, 2018, [Online]. Available: http://arxiv.org/abs/1807.10854.

[7]   S. Zerdoumi *et al.*, "Image pattern recognition in big data: taxonomy and open challenges: survey," *Multimed. Tools Appl.*, 2018, doi: 10.1007/s11042-017-5045-7.

[8]   A. Rehman, A. Paul, A. Ahmad, and G. Jeon, "A novel class based searching algorithm in small world internet of drone network," *Comput. Commun.*, 2020, doi: 10.1016/j.comcom.2020.03.040.

[9]   M. M. Rathore, A. Paul, A. Ahmad, M. Anisetti, and G. Jeon, "Hadoop-Based Intelligent Care System (HICS)," *ACM Trans. Internet Technol.*, 2017, doi: 10.1145/3108936.

[10]   A. Krizhevsky, I. Sutskever, and G. E. Hinton, "2012 AlexNet," *Adv. Neural Inf. Process. Syst.*, 2012, doi: http://dx.doi.org/10.1016/j.protcy.2014.09.007.

[11]   K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.

[12]   K. He, X. Zhang, S. Ren, and J. Sun, "ResNet," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2016, doi: 10.1109/CVPR.2016.90.

[13]   R. Pascanu, T. Mikolov, and Y. Bengio, "Understanding the exploding gradient problem," *Proc. 30th Int. Conf. Mach. Learn.*, 2012, doi: 10.1109/72.279181.

[14]   Y. Bengio, P. Simard, and P. Frasconi, "Learning Long-Term Dependencies with Gradient Descent is Difficult," *IEEE Trans. Neural Networks*, 1994, doi: 10.1109/72.279181.

[15]   J. Gu *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognit.*, 2018, doi: 10.1016/j.patcog.2017.10.013.

[16]   D. E. Rumelhart, "David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams," *Nature*, 1986.

[17]   J. Zhang, Q. Lei, and I. S. Dhillon, "Stabilizing gradients for deep neural networks via efficient SVD parameterization," in *35th International Conference on Machine Learning, ICML 2018*, 2018.

[18]   J. Bergstra and Y. Bengio, "Slow, decorrelated features for pretraining complex cell-like networks," in *Advances in Neural Information Processing Systems 22 - Proceedings of the 2009 Conference*, 2009.

[19]   K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, doi: 10.1109/ICCV.2015.123.

[20]   X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Journal of Machine Learning Research*, 2010.

[21]   D. Mishkin and J. Matas, "All you need is a good init," in *4th International Conference on Learning*

Representations, ICLR 2016 - Conference Track Proceedings, 2016.

[22] D. Sussillo, "Random Walks: Training Very Deep Nonlinear Feed-Forward Networks with Smart Initialization," *ICLR*, 2015.

[23] LeCun Yann, Cortes Corinna, and Burges Christopher, "THE MNIST DATABASE of handwritten digits," *Courant Inst. Math. Sci.*, 1998.

[24] A. Ahmad *et al.*, "Toward modeling and optimization of features selection in Big Data based social Internet of Things," *Futur. Gener. Comput. Syst.*, 2018, doi: 10.1016/j.future.2017.09.028.

[25] H. Ullaha *et al.*, "Optic disc segmentation and classification in color fundus images: a resource-aware healthcare service in smart cities," *J. Ambient Intell. Humaniz. Comput.*, vol. 0, no. 0, p. 0, 2018, doi: 10.1007/s12652-018-0988-8.

[26] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected crfs with Gaussian edge potentials," in *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011*, 2011.

[27] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, "FitNets: Hints for thin deep nets," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.

[28] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," pp. 1–9, 2015, [Online]. Available: http://arxiv.org/abs/1503.02531.

[29] Z. Chen, V. Badrinarayanan, C. Y. Lee, and A. Rabinovich, "GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks," in *35th International Conference on Machine Learning, ICML 2018*, 2018.

[30] W. Lin, M. E. Khan, and M. Schmidt, "Fast and simple natural-gradient variational inference with mixture of exponential-family approximations," in *36th International Conference on Machine Learning, ICML 2019*, 2019.

[31] S. Shankar, V. Piratla, S. Chakrabarti, S. Chaudhuri, P. Jyothi, and S. Sarawagi, "Generalizing across domains via cross-gradient training," in *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 2018.

[32] C. Wu *et al.*, "Variance reduction for policy gradient with action-dependent factorized baselines," in *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 2018.

[33] A. M. Saxe, J. L. McClelland, and S. Ganguli, "Exact solutions to the nonlinear dynamics of learning in deep linear neural networks," in *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, 2014.

[34] D. Xie, J. Xiong, and S. Pu, "All you need is beyond a good init: Exploring better solution for training extremely deep convolutional neural networks with orthonormality and modulation," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 5075–5084, 2017, doi: 10.1109/CVPR.2017.539.

[35] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms," pp. 1–6, 2017, [Online]. Available: http://arxiv.org/abs/1708.07747.

[36] A. Krizhevsky, "Learning Multiple Layers of Features from Tiny Images," *... Sci. Dep. Univ. Toronto, Tech. ...*, 2009, doi: 10.1.1.222.9220.

[37] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.