

## REFERENCES

- [1] J. Heaton, “Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The MIT Press, 2016, 800 pp, ISBN: 0262035618,” *Genet. Program. Evolvable Mach.*, vol. 19, no. 1–2, pp. 305–307, Jun. 2018, doi: 10.1007/s10710-017-9314-z.
- [2] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
- [3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
- [4] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” Dec. 10, 2015, *arXiv*: arXiv:1512.03385. doi: 10.48550/arXiv.1512.03385.
- [5] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, “Squeeze-and-Excitation Networks,” May 16, 2019, *arXiv*: arXiv:1709.01507. doi: 10.48550/arXiv.1709.01507.
- [6] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional Block Attention Module,” in *Computer Vision – ECCV 2018*, vol. 11211, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., in Lecture Notes in Computer Science, vol. 11211. , Cham: Springer International Publishing, 2018, pp. 3–19. doi: 10.1007/978-3-030-01234-2\_1.
- [7] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, “ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA: IEEE, Jun. 2020, pp. 11531–11539. doi: 10.1109/CVPR42600.2020.01155.
- [8] L. Yang, R.-Y. Zhang, L. Li, and X. Xie, “SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks”.
- [9] D. Ouyang *et al.*, “Efficient Multi-Scale Attention Module with Cross-Spatial Learning,” in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics*,

- Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece: IEEE, Jun. 2023, pp. 1–5. doi: 10.1109/ICASSP49357.2023.10096516.
- [10] W. Lu, Y. Yang, and L. Yang, “Fine-grained image classification method based on hybrid attention module,” *Front. Neurorobotics*, vol. 18, p. 1391791, May 2024, doi: 10.3389/fnbot.2024.1391791.
- [11] J. Ye *et al.*, “YOLOv8 Architectural Scene Section Recognition Method Based on SimAM-EMA Hybrid Attention Mechanism,” *Sensors*, vol. 25, no. 10, p. 3060, May 2025, doi: 10.3390/s25103060.
- [12] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” Jan. 30, 2015, *arXiv*: arXiv:1409.0575. doi: 10.48550/arXiv.1409.0575.
- [13] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” Sep. 11, 2020, *arXiv*: arXiv:1905.11946. doi: 10.48550/arXiv.1905.11946.
- [14] Y. Le and X. Yang, “Tiny ImageNet Visual Recognition Challenge”.
- [15] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?”.
- [16] J. Howard and S. Ruder, “Universal Language Model Fine-tuning for Text Classification”.
- [17] W. Rawat and Z. Wang, “Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review,” *Neural Comput.*, vol. 29, no. 9, pp. 2352–2449, Sep. 2017, doi: 10.1162/neco\_a\_00990.
- [18] V. Dumoulin and F. Visin, “A guide to convolution arithmetic for deep learning,” Jan. 11, 2018, *arXiv*: arXiv:1603.07285. doi: 10.48550/arXiv.1603.07285.
- [19] S. H. S. Basha, S. R. Dubey, V. Pulabaigari, and S. Mukherjee, “Impact of Fully Connected Layers on Performance of Convolutional Neural Networks for Image Classification,” *Neurocomputing*, vol. 378, pp. 112–119, Feb. 2020, doi: 10.1016/j.neucom.2019.10.008.
- [20] M. D. Zeiler and R. Fergus, “Visualizing and Understanding Convolutional Networks,” in *Computer Vision – ECCV 2014*, vol. 8689, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., in Lecture Notes in Computer Science, vol. 8689. , Cham: Springer International Publishing, 2014, pp. 818–833. doi: 10.1007/978-3-319-10590-1\_53.

- [21] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Apr. 10, 2015, *arXiv*: arXiv:1409.1556. doi: 10.48550/arXiv.1409.1556.
- [22] N. A. Al-Humaidan and M. Prince, “A Classification of Arab Ethnicity Based on Face Image using Deep Learning Approach,” *IEEE Access*, 2021.
- [23] R. K. Srivastava, K. Greff, and J. Schmidhuber, “Highway Networks,” Nov. 03, 2015, *arXiv*: arXiv:1505.00387. doi: 10.48550/arXiv.1505.00387.
- [24] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” Mar. 02, 2015, *arXiv*: arXiv:1502.03167. doi: 10.48550/arXiv.1502.03167.
- [25] H. Touvron, A. Vedaldi, M. Douze, and H. Jégou, “Fixing the train-test resolution discrepancy,” Jan. 20, 2022, *arXiv*: arXiv:1906.06423. doi: 10.48550/arXiv.1906.06423.
- [26] S. Kornblith, J. Shlens, and Q. V. Le, “Do Better ImageNet Models Transfer Better?,” Jun. 17, 2019, *arXiv*: arXiv:1805.08974. doi: 10.48550/arXiv.1805.08974.
- [27] B. Hariharan, J. Malik, and D. Ramanan, “Discriminative Decorrelation for Clustering and Classification,” in *Computer Vision – ECCV 2012*, vol. 7575, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds., in Lecture Notes in Computer Science, vol. 7575. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 459–472. doi: 10.1007/978-3-642-33765-9\_33.
- [28] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated Residual Transformations for Deep Neural Networks,” Apr. 11, 2017, *arXiv*: arXiv:1611.05431. doi: 10.48550/arXiv.1611.05431.
- [29] J. Fu *et al.*, “Dual Attention Network for Scene Segmentation,” Apr. 21, 2019, *arXiv*: arXiv:1809.02983. doi: 10.48550/arXiv.1809.02983.
- [30] L. Chen *et al.*, “SCA-CNN: Spatial and Channel-wise Attention in Convolutional Networks for Image Captioning,” Apr. 12, 2017, *arXiv*: arXiv:1611.05594. doi: 10.48550/arXiv.1611.05594.
- [31] W. Zeng, “Image data augmentation techniques based on deep learning: A survey,” *Math. Biosci. Eng.*, vol. 21, no. 6, pp. 6190–6224, 2024, doi: 10.3934/mbe.2024272.

- [32] K. Alomar, H. I. Aysel, and X. Cai, "Data Augmentation in Classification and Segmentation: A Survey and New Strategies," *J. Imaging*, vol. 9, no. 2, p. 46, Feb. 2023, doi: 10.3390/jimaging9020046.
- [33] I. Corley, C. Robinson, R. Dodhia, J. M. L. Ferres, and P. Najafirad, "Revisiting pre-trained remote sensing model benchmarks: resizing and normalization matters," May 22, 2023, *arXiv*: arXiv:2305.13456. doi: 10.48550/arXiv.2305.13456.
- [34] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," Jan. 04, 2019, *arXiv*: arXiv:1711.05101. doi: 10.48550/arXiv.1711.05101.
- [35] M. Zolfaghari, M. Saniee Abadeh, and H. Sajedi, "Channel-spatial attention modules in convolutional neural networks for image classification," *Sci. Rep.*, vol. 16, no. 1, p. 2348, Dec. 2025, doi: 10.1038/s41598-025-32055-1.
- [36] V. Shankar, R. Roelofs, H. Mania, A. Fang, B. Recht, and L. Schmidt, "Evaluating Machine Accuracy on ImageNet".
- [37] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Apr. 17, 2017, *arXiv*: arXiv:1704.04861. doi: 10.48550/arXiv.1704.04861.
- [38] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design," Jul. 30, 2018, *arXiv*: arXiv:1807.11164. doi: 10.48550/arXiv.1807.11164.
- [39] P. Goyal *et al.*, "Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour," Apr. 30, 2018, *arXiv*: arXiv:1706.02677. doi: 10.48550/arXiv.1706.02677.