

REFERENCES

- [1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, p. 84–90, May 2017. [Online]. Available: <https://doi.org/10.1145/3065386>
- [2] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in *Computer Vision – ECCV 2014*. Cham: Springer International Publishing, 2014, pp. 818–833.
- [3] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” April 2015. [Online]. Available: <https://arxiv.org/abs/1409.1556>
- [4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 1–9.
- [5] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 2261–2269.
- [6] M. Tan and Q. V. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” September 2020. [Online]. Available: <https://arxiv.org/abs/1905.11946>
- [7] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16x16 words: Transformers for image recognition at scale,” June 2021. [Online]. Available: <https://arxiv.org/abs/2010.11929>
- [8] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, “Swin transformer: Hierarchical vision transformer using shifted windows,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 9992–10 002.

- [9] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, “A convnet for the 2020s,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 11 966–11 976.
- [10] H. Zheng, J. Fu, T. Mei, and J. Luo, “Learning multi-attention convolutional neural network for fine-grained image recognition,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 5219–5227.
- [11] C. Liu, H. Xie, Z.-J. Zha, L. Ma, L. Yu, and Y. Zhang, “Filtration and distillation: Enhancing region attention for fine-grained visual categorization,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, pp. 11 555–11 562, Apr. 2020. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/6822>
- [12] Z. Yang, T. Luo, D. Wang, Z. Hu, J. Gao, and L. Wang, “Learning to navigate for fine-grained classification,” in *Computer Vision – ECCV 2018*. Cham: Springer International Publishing, October 2018, pp. 438–454.
- [13] P. Li, J. Xie, Q. Wang, and Z. Gao, “Towards faster training of global covariance pooling networks by iterative matrix square root normalization,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2018, pp. 947–955.
- [14] Y. Wang, V. I. Morariu, and L. S. Davis, “Learning a discriminative filter bank within a cnn for fine-grained recognition,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2018, pp. 4148–4157.
- [15] C. Liu, H. Xie, Z. Zha, L. Yu, Z. Chen, and Y. Zhang, “Bidirectional attention-recognition model for fine-grained object classification,” *IEEE Transactions on Multimedia*, vol. 22, no. 7, pp. 1785–1795, July 2020.
- [16] Y. Ding, Y. Zhou, Y. Zhu, Q. Ye, and J. Jiao, “Selective sparse sampling for fine-grained image recognition,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019, pp. 6598–6607.
- [17] Y. Chen, Y. Bai, W. Zhang, and T. Mei, “Destruction and construction learning for fine-grained image recognition,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019, pp. 5152–5161.
- [18] T. Hu, H. Qi, Q. Huang, and Y. Lu, “See better before looking closer: Weakly supervised data augmentation network for fine-grained visual classification,” March 2019. [Online]. Available: <https://arxiv.org/abs/1901.09891>

- [19] H. Li, X. Zhang, Q. Tian, and H. Xiong, “Attribute mix: Semantic data augmentation for fine grained recognition,” in *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, December 2020, pp. 243–246.
- [20] Z. Wang, S. Wang, H. Li, Z. Dou, and J. Li, “Graph-propagation based correlation learning for weakly supervised fine-grained image classification,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, pp. 12 289–12 296, Apr. 2020. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/6912>
- [21] R. Du, D. Chang, A. K. Bhunia, J. Xie, Z. Ma, Y.-Z. Song, and J. Guo, “Fine-grained visual classification via progressive multi-granularity training of jigsaw patches,” in *Computer Vision – ECCV 2020*. Cham: Springer International Publishing, November 2020, pp. 153–168.
- [22] P. Zhuang, Y. Wang, and Y. Qiao, “Learning attentive pairwise interaction for fine-grained classification,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, pp. 13 130–13 137, Apr. 2020. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/7016>
- [23] Y. Rao, G. Chen, J. Lu, and J. Zhou, “Counterfactual attention learning for fine-grained visual categorization and re-identification,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 1005–1014.
- [24] Y. Hu, X. Liu, B. Zhang, J. Han, and X. Cao, “Alignment enhancement network for fine-grained visual categorization,” *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 17, no. 1s, Mar. 2021. [Online]. Available: <https://doi.org/10.1145/3446208>
- [25] Y. Zhao, J. Li, X. Chen, and Y. Tian, “Part-guided relational transformers for fine-grained visual recognition,” *IEEE Transactions on Image Processing*, vol. 30, pp. 9470–9481, November 2021.
- [26] F. Zhang, M. Li, G. Zhai, and Y. Liu, “Multi-branch and multi-scale attention learning for fine-grained visual categorization,” in *MultiMedia Modeling*. Cham: Springer International Publishing, January 2021, pp. 136–147.
- [27] D. Liu, L. Zhao, Y. Wang, and J. Kato, “Learn from each other to classify better: Cross-layer mutual attention learning for fine-grained visual classifi-

- cation,” *Pattern Recognition*, vol. 140, p. 109550, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0031320323002509>
- [28] Z.-H. Zhou, “A brief introduction to weakly supervised learning,” *National Science Review*, vol. 5, no. 1, pp. 44–53, 08 2017. [Online]. Available: <https://doi.org/10.1093/nsr/nwx106>
- [29] A. Sauer and A. Geiger, “Counterfactual generative networks,” 2021. [Online]. Available: <https://arxiv.org/abs/2101.06046>
- [30] E. M. Kenny and M. T. Keane, “On generating plausible counterfactual and semi-factual explanations for deep learning,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 13, pp. 11 575–11 585, May 2021. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/17377>
- [31] P. Gupta, A. Biswas, H. Admoni, and D. Held, “Object importance estimation using counterfactual reasoning for intelligent driving,” *IEEE Robotics and Automation Letters*, vol. 9, no. 4, pp. 3648–3655, 2024.
- [32] J. Sun, M. Gao, H. Wang, and Q. Dong, “Recursive counterfactual deconfounding for image recognition,” *Knowledge-Based Systems*, vol. 315, p. 113245, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0950705125002928>
- [33] T. Mesnard, T. Weber, F. Viola, S. Thakoor, A. Saade, A. Harutyunyan, W. Dabney, T. S. Stepleton, N. Heess, A. Guez, E. Moulines, M. Hutter, L. Buesing, and R. Munos, “Counterfactual credit assignment in model-free reinforcement learning,” in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. Meila and T. Zhang, Eds., vol. 139. PMLR, 18–24 Jul 2021, pp. 7654–7664. [Online]. Available: <https://proceedings.mlr.press/v139/mesnard21a.html>
- [34] F. Feng, J. Zhang, X. He, H. Zhang, and T.-S. Chua, “Empowering language understanding with counterfactual reasoning,” 2021. [Online]. Available: <https://arxiv.org/abs/2106.03046>
- [35] L. Chen, Y. Zheng, Y. Niu, H. Zhang, and J. Xiao, “Counterfactual samples synthesizing and training for robust visual question answering,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 11, pp. 13 218–13 234, 2023.

- [36] E. Abbasnejad, D. Teney, A. Parvaneh, J. Shi, and A. van den Hengel, “Counterfactual vision and language learning,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 10 041–10 051.
- [37] A. Parvaneh, E. Abbasnejad, D. Teney, J. Q. Shi, and A. van den Hengel, “Counterfactual vision-and-language navigation: Unravelling the unseen,” in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 5296–5307. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/39016cfe079db1bfb359ca72fcb3fd8-Paper.pdf
- [38] H. Kim, A. Zala, and M. Bansal, “CoSIm: Commonsense reasoning for counterfactual scene imagination,” in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, M. Carpuat, M.-C. de Marneffe, and I. V. Meza Ruiz, Eds. Seattle, United States: Association for Computational Linguistics, Jul. 2022, pp. 911–923. [Online]. Available: <https://aclanthology.org/2022.naacl-main.66/>
- [39] J. Ji, Z. Li, S. Xu, M. Xiong, J. Tan, Y. Ge, H. Wang, and Y. Zhang, “Counterfactual collaborative reasoning,” in *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, ser. WSDM ’23. New York, NY, USA: Association for Computing Machinery, 2023, p. 249–257. [Online]. Available: <https://doi.org/10.1145/3539597.3570464>
- [40] D. Delahaye, S. Chaimatanan, and M. Mongeau, *Simulated Annealing: From Basics to Applications*. Cham: Springer International Publishing, 2019, pp. 1–35. [Online]. Available: https://doi.org/10.1007/978-3-319-91086-4_1
- [41] A. Gülcü and Z. Kuş, “Multi-objective simulated annealing for hyperparameter optimization in convolutional neural networks,” *PeerJ Computer Science*, vol. 7, p. e338, Jan. 2021. [Online]. Available: <https://doi.org/10.7717/peerj-cs.338>
- [42] A. M. Sarfi, Z. Karimpour, M. Chaudhary, N. M. Khalid, M. Ravanelli, S. Mudur, and E. Belilovsky, “Simulated annealing in early layers leads to better generalization,” in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 20 205–20 214.

- [43] A. H. Correia, D. E. Worrall, and R. Bondesan, “Neural simulated annealing,” in *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, F. Ruiz, J. Dy, and J.-W. van de Meent, Eds., vol. 206. PMLR, 25–27 Apr 2023, pp. 4946–4962. [Online]. Available: <https://proceedings.mlr.press/v206/correia23a.html>
- [44] M. S. Haque, M. Fahim, and M. Ibrahim, “An exploratory study on simulated annealing for feature selection in learning-to-rank,” *International Journal of Intelligent Systems and Applications*, vol. 16, no. 4, pp. 86–103, 2024.
- [45] S. Maji, E. Rahtu, J. Kannala, M. Blaschko, and A. Vedaldi, “Fine-grained visual classification of aircraft,” 2013. [Online]. Available: <https://arxiv.org/abs/1306.5151>
- [46] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, “The caltech-ucsd birds-200-2011 dataset,” Jul 2011. [Online]. Available: <https://authors.library.caltech.edu/records/cvm3y-5hh21>
- [47] T.-Y. Lin, A. RoyChowdhury, and S. Maji, “Bilinear cnn models for fine-grained visual recognition,” in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1449–1457.
- [48] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 770–778.