DAFTAR PUSTAKA

Altuncu, E., Franqueira, V. N. L., & Li, S. (2024). Deepfake: definitions, performance metrics and standards, datasets, and a meta-review. *Frontiers in Big Data*, *7*(Dl), 1–31. https://doi.org/10.3389/fdata.2024.1400024

Bommasani, R., Hudson, D. A., & others. (2021). On the opportunities and risks of foundation models. *ArXiv Preprint ArXiv:2108.07258*. https://arxiv.org/abs/2108.07258

Borji, A. (2023). Qualitative failures of image generation models and their application in detecting AI-generated images. *ArXiv Preprint ArXiv:2202.03945*.

Brock, A., Donahue, J., & Simonyan, K. (2019). Large scale GaN training for high fidelity natural image synthesis. *7th International Conference on Learning Representations, ICLR 2019*, 1–35.

Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, *106*, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011

Chen, Y., Yuille, A., & Zhou, Z. (2023). Which layer is learning faster? {A} systematic exploration of layer-wise convergence rate for deep neural networks. *International Conference on Learning Representations*. https://openreview.net/forum?id=wlMDF1jQF86

Corvi, R., Cozzolino, D., Zingarini, G., Poggi, G., Nagano, K., & Verdoliva, L. (2023a). On the detection of synthetic images generated by diffusion models. *ICASSP 2023--IEEE International Conference on Acoustics, Speech and Signal Processing*, 1–5. https://doi.org/10.1109/ICASSP49357.2023.10095167

Corvi, R., Cozzolino, D., Zingarini, G., Poggi, G., Nagano, K., & Verdoliva, L. (2023b). On The Detection of Synthetic Images Generated by Diffusion Models. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *2023-June*. https://doi.org/10.1109/ICASSP49357.2023.10095167

d'Ascoli, S., Touvron, H., Leavitt, M. L., Morcos, A. S., Biroli, G., & Sagun, L. (2022). ConViT: improving vision transformers with soft convolutional inductive biases. *Journal of Statistical Mechanics: Theory and Experiment*, *2022*(11). https://doi.org/10.1088/1742-5468/ac9830

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., & others. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations*. https://openreview.net/forum?id=YicbFdNTTy

Forum, iNaturalist C. (2026). *Article on AI generated image(s) on iNat*. https://forum.inaturalist.org/t/article-on-ai-generated-image-s-on-inat/76984

Frank, J., Herbert, F., Ricker, J., Schönherr, L., Eisenhofer, T., Fischer, A., Dürmuth, M., & Holz, T. (2024). A Representative Study on Human Detection of Artificially Generated Media Across Countries. *Proceedings - IEEE Symposium on Security and Privacy*, 55–73. https://doi.org/10.1109/SP54263.2024.00159

Gonzalez, R. C., & Woods, R. E. (2018). *Digital Image Processing* (4th ed.). Pearson.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. https://www.deeplearningbook.org

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, *27*.

Goyal, P., Dollár, P., Girshick, R., Noordhuis, P., Wesolowski, L., Kyrola, A., Tulloch, A.,

Jia, Y., & He, K. (2018). *Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour*. http://arxiv.org/abs/1706.02677

Grommelt, P., Webers, J., & Paulus, J. (2024). Fake or {JPEG}? Revealing common biases in generated image detection datasets. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10603–10613. https://doi.org/10.1109/CVPR52733.2024.01007

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked autoencoders are scalable vision learners. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15979–15988. https://doi.org/10.1109/CVPR52688.2022.01553

He, K., Girshick, R., & Dollar, P. (2019). Rethinking imageNet pre-training. *Proceedings of the IEEE International Conference on Computer Vision*, *i*, 4917–4926. https://doi.org/10.1109/ICCV.2019.00502

Högemann, M., Betke, J., & Thomas, O. (2025). What you see is not what you get anymore: a mixed-methods approach on human perception of AI-generated images. *Frontiers in Artificial Intelligence*, *8*(November), 1–14. https://doi.org/10.3389/frai.2025.1707336

Hughes, R. T., Zhu, L., & Bednarz, T. (2021). Generative Adversarial Networks–Enabled Human–Artificial Intelligence Collaborative Applications for Creative and Design Industries: A Systematic Review of Current Approaches and Trends. *Frontiers in Artificial Intelligence*, *4*(April), 1–17. https://doi.org/10.3389/frai.2021.604234

Ji, C., Xie, Z., Li, R., Yang, Z., & Hou, Z. (2025). Vision transformer based classification of sewer defects weighted loss model. *Tunnelling and Underground Space Technology*, *156*, 106271. https://doi.org/10.1016/j.tust.2024.106271

Jia, Z., Yuan, Z., Duan, X., Zhang, J., Zhou, J., & Jain, A. K. (2026). CoDA: Color Distribution Probing for Efficient and Generalizable AI-Generated Image Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–17. https://doi.org/10.1109/TPAMI.2026.3696399

Kamali, N., Nakamura, K., Chatzimparmpas, A., Hullman, J., & Groh, M. (2024). *How to Distinguish AI-Generated Images from Authentic Photographs*. *June*. http://arxiv.org/abs/2406.08651

Kamali, N., Nakamura, K., Kumar, A., Chatzimparmpas, A., Hullman, J., & Groh, M. (2025). Characterizing Photorealism and Artifacts in Diffusion Model-Generated Images. In *Conference on Human Factors in Computing Systems - Proceedings* (Vol. 1, Issue 1). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713962

Krubha, Y. S., Hou, A., Vester, B., Walker, W., Wang, X., Lin, L., & Hu, S. (2025). Robust AI-Generated Face Detection with Imbalanced Data. *Proceedings of the International Conference on Multimedia Information Processing and Retrieval, MIPR*, 470–476. https://doi.org/10.1109/MIPR67560.2025.00080

Kumar, A., Raghunathan, A., Jones, R., Ma, T., & Liang, P. (2022a). Fine-tuning can distort pretrained features and underperform out-of-distribution. *International Conference on Learning Representations*. https://openreview.net/forum?id=UYneFzXSJWh

Kumar, A., Raghunathan, A., Jones, R., Ma, T., & Liang, P. (2022b). Fine-Tuning Can Distort Pretrained Features and Underperform Out-of-Distribution. *ICLR 2022 - 10th International Conference on Learning Representations*, *10*(Id), 1–54.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. *Proceedings of the IEEE International Conference on Computer Vision*, 2980–2988. https://doi.org/10.1109/ICCV.2017.324

Liu, Z., Ojha, U., & Wu, G. (2023). {LSDA}: Large scale diffusion-based augmentation for

robust object detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. https://arxiv.org/abs/2311.11919

Loshchilov, I., & Hutter, F. (2017). {SGDR}: Stochastic gradient descent with warm restarts. *International Conference on Learning Representations*. https://openreview.net/forum?id=Skq89Scxx

Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. *International Conference on Learning Representations*. https://openreview.net/forum?id=Bkg6RiCqY7

Murabayashi, A. (2023). What Does Generative AI Mean for Bird and Nature Photography? *Audubon Magazine*. https://www.audubon.org/magazine/what-does-generative-ai-mean-bird-and-nature-photography

Narang, S., Diamos, G., Elsen, E., Micikevicius, P., Alben, J., Garcia, D., Ginsburg, B., Houston, M., Kuchaiev, O., Venkatesh, G., & Wu, H. (2018). Mixed precision training. *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 1–12.

Nguyen, H. H., Yamagishi, J., & Echizen, I. (2024). Exploring Self-Supervised Vision Transformers for Deepfake Detection: A Comparative Analysis. *Proceedings - 2024 IEEE International Joint Conference on Biometrics, IJCB 2024*. https://doi.org/10.1109/IJCB62174.2024.10744497

Nichol, A., & Dhariwal, P. (2021). Improved Denoising Diffusion Probabilistic Models. *Proceedings of Machine Learning Research*, *139*, 8162–8171.

Ojha, U., Li, Y., & Lee, Y. J. (2023a). Towards universal fake image detectors that generalize across generative models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 24480–24489. https://doi.org/10.1109/CVPR52729.2023.02345

Ojha, U., Li, Y., & Lee, Y. J. (2023b). Towards Universal Fake Image Detectors that Generalize Across Generative Models. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2023-June*, 24480–24489. https://doi.org/10.1109/CVPR52729.2023.02345

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Reddy, C. K. K., Reddy, P. A., Janapati, H., Assiri, B., Shuaib, M., Alam, S., & Sheneamer, A. (2024). A fine-tuned vision transformer based enhanced multi-class brain tumor classification using MRI scan imagery. *Frontiers in Oncology*, *14*(July), 1–23. https://doi.org/10.3389/fonc.2024.1400341

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10684–10695. https://doi.org/10.1109/CVPR52688.2022.01042

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

Version, T., & Graphics, C. (n.d.). *Seeing Is No Longer Believing : A Survey on the State of Deepfakes*.

Wang, S.-Y., Wang, O., Zhang, R., Owens, A., & Efros, A. A. (2020). {CNN}-generated images are surprisingly easy to spot... for now. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8695–8704. https://doi.org/10.1109/CVPR42600.2020.00872

Wang, S. Y., Wang, O., Zhang, R., Owens, A., & Efros, A. A. (2020). CNN-Generated Images Are Surprisingly Easy to Spot.. For Now. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 8692–8701. https://doi.org/10.1109/CVPR42600.2020.00872

Ye, P., Huang, Y., Tu, C., Li, M., Chen, T., He, T., & Ouyang, W. (2023). *Partial Fine-Tuning: A Successor to Full Fine-Tuning for Vision Transformers*. http://arxiv.org/abs/2312.15681

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems*, *27*.

Zheng, K., Huang, W., & Sun, L. (2023). *TransMed: Large Language Models Enhance Vision Transformer for Biomedical Image Classification*. http://arxiv.org/abs/2312.07125

Zhu, M., Chen, H., Yan, Q., Huang, X., Lin, G., Li, W., & others. (2023). {GenImage}: A million-scale benchmark for detecting {AI}-generated image. *Advances in Neural Information Processing Systems*, *36*.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, *109*(1), 43–76. https://doi.org/10.1109/JPROC.2020.3004555