

DAFTAR PUSTAKA

- [1] M. H. Arshad, M. Bilal, and A. Gani, "Human Activity Recognition: Review, Taxonomy and Open Challenges." *Sensors (Basel, Switzerland)*, vol. 22, no. 17, p. 6463, aug 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/22/17/6463><http://www.ncbi.nlm.nih.gov/pubmed/36080922><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC9460866>
- [2] O. Mazhar, S. Ramdani, and A. Cherubini, "A Deep Learning Framework for Recognizing Both Static and Dynamic Gestures," *Sensors*, vol. 21, no. 6, p. 2227, mar 2021. [Online]. Available: <https://doi.org/10.3390/s21062227><https://www.mdpi.com/1424-8220/21/6/2227>
- [3] P. Kumar, S. Chauhan, and L. K. Awasthi, "Human Activity Recognition (HAR) Using Deep Learning: Review, Methodologies, Progress and Future Research Directions," *Archives of Computational Methods in Engineering*, vol. 31, no. 1, pp. 179–219, jan 2024. [Online]. Available: <https://link.springer.com/10.1007/s11831-023-09986-x>
- [4] J. Yang, S. Tang, Y. Xu, Y. Zhou, and L. Xie, "MaskFi: Unsupervised Learning of WiFi and Vision Representations for Multimodal Human Activity Recognition," feb 2024. [Online]. Available: <http://arxiv.org/abs/2402.19258>
- [5] B. Chen, W. Wongso, Z. Li, Y. Khaokaew, H. Xue, and F. Salim, "COMODO: Cross-Modal Video-to-IMU Distillation for Efficient Egocentric Human Activity Recognition," pp. 1–11, mar 2025. [Online]. Available: <http://arxiv.org/abs/2503.07259>
- [6] M. B. Shaikh and D. Chai, "RGB-D Data-Based Action Recognition: A Review," *Sensors*, vol. 21, no. 12, p. 4246, jun 2021. [Online]. Available: <https://www.mdpi.com/1424-8220/21/12/4246>
- [7] Z. Leng, A. Iyer, and T. Plötz, "Scaling Human Activity Recognition: A Comparative Evaluation of Synthetic Data Generation and Augmentation Techniques," *Proceedings of ACM Conference (Conference'17)*, vol. 1, no. 1, jun 2025. [Online]. Available: <http://arxiv.org/abs/2506.07612>
- [8] S. H. Khan and R. Iqbal, "A Comprehensive Survey on Architectural Advances in Deep CNNs: Challenges, Applications, and Emerging Research Directions," mar 2025. [Online]. Available: <https://arxiv.org/abs/2503.16546v1>
- [9] L. Shen, Y. Sun, Z. Yu, L. Ding, X. Tian, and D. Tao, "On Efficient Training of Large-Scale Deep Learning Models," *ACM Computing Surveys*, vol. 57, no. 3, nov 2024. [Online]. Available: [/doi/pdf/10.1145/3700439?download=true](https://doi.org/10.1145/3700439?download=true)
- [10] G. Menghani, "Efficient Deep Learning: A Survey on Making Deep Learning Models Smaller, Faster, and Better," *ACM Computing Surveys*, vol. 55, no. 12, dec 2023. [Online]. Available: [/doi/pdf/10.1145/3578938?download=true](https://doi.org/10.1145/3578938?download=true)
- [11] G. Jocher and J. Qiu, "Ultralytics YOLO11," 2024. [Online]. Available: <https://github.com/ultralytics/ultralytics>

- [12] R. Khanam and M. Hussain, “YOLOv11: An Overview of the Key Architectural Enhancements,” oct 2024. [Online]. Available: <https://arxiv.org/abs/2410.17725v1><http://arxiv.org/abs/2410.17725>
- [13] S. Alghyaline, “A Real-Time Street Actions Detection,” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 2, pp. 322–329, 2019. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85063584229&doi=10.14569%2Fijacsa.2019.0100243&partnerID=40&md5=b0ced8ead827f54a6ed654267026cba8><http://thesai.org/Publications/ViewPaper?Volume=10&Issue=2&Code=ijacsa&SerialNo=43>
- [14] F. Serpush and M. Rezaei, “Complex Human Action Recognition Using a Hierarchical Feature Reduction and Deep Learning-Based Method,” *SN Computer Science*, vol. 2, no. 2, pp. 1–15, 2021. [Online]. Available: <https://doi.org/10.1007/s42979-021-00484-0>
- [15] Z. Jiang, J. Yang, N. Jiang, S. Liu, T. Xie, L. Zhao, and R. Li, “YOWOV2: A Stronger yet Efficient Multi-level Detection Framework for Real-Time Spatio-Temporal Action Detection,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 15202 LNAI, no. 8, pp. 33–48, 2025.
- [16] T. Liu, X. He, L. He, and F. Yuan, “A video drowning detection device based on underwater computer vision,” *IET Image Processing*, vol. 17, no. 6, pp. 1905–1918, 2023.
- [17] R. Luo, A. Anand, F. Zulkernine, and F. Rivest, “YOLO-I3D: Optimizing Inflated 3D Models for Real-Time Human Activity Recognition,” *Journal of Imaging 2024, Vol. 10, Page 269*, vol. 10, no. 11, p. 269, oct 2024. [Online]. Available: <https://www.mdpi.com/2313-433X/10/11/269/htm><https://www.mdpi.com/2313-433X/10/11/269>
- [18] D. M. N. Dang, V. H. Duong, J. C. Wang, and N. B. Duc, “YOWOV3: An Efficient and Generalized Framework for Human Action Detection and Recognition,” aug 2024. [Online]. Available: <https://arxiv.org/abs/2408.02623v2>
- [19] N. Alzahrani, O. Bchir, and M. M. B. Ismail, “YOLO-Act: Unified Spatiotemporal Detection of Human Actions Across Multi-Frame Sequences,” *Sensors*, vol. 25, no. 10, p. 3013, may 2025. [Online]. Available: <https://www.mdpi.com/1424-8220/25/10/3013>
- [20] L. Zhou, X. Liu, X. Guan, and Y. Cheng, “CSSA-YOLO: Cross-Scale Spatiotemporal Attention Network for Fine-Grained Behavior Recognition in Classroom Environments,” *Sensors*, vol. 25, no. 10, p. 3132, may 2025. [Online]. Available: <https://www.mdpi.com/1424-8220/25/10/3132>
- [21] O. Köpüklü, X. Wei, and G. Rigoll, “You Only Watch Once: A Unified CNN Architecture for Real-Time Spatiotemporal Action Localization,” nov 2019. [Online]. Available: <https://arxiv.org/abs/1911.06644v5><http://arxiv.org/abs/1911.06644>

- [22] R. M. B. Suryo Wicaksono, I. Wayan Mustika, and S. Sulistyo, "Contextualizing YOLOv11 with Spatiotemporal Approach Using YOWOV3 for Multiple Human Activity Recognition," in *2025 International Conference on Data Science and Its Applications (ICoDSA)*. IEEE, jul 2025, pp. 643–648. [Online]. Available: <https://ieeexplore.ieee.org/document/11157317/>
- [23] M. M. Islam, S. Nooruddin, F. Karray, and G. Muhammad, "Human activity recognition using tools of convolutional neural networks: A state of the art review, data sets, challenges, and future prospects," *Computers in Biology and Medicine*, vol. 149, 2022.
- [24] R. Sapkota, R. Qureshi, M. Flores-Calero, C. Badgujar, U. Nepal, A. Poulouse, P. Z. Zenorobotics, L. Billings, U. Bhanu, P. Vaddevolu, S. Khan, M. Shoman, H. Yan, and M. Karkee, "YOLOv12 to Its Genesis: A Decadal and Comprehensive Review of The You Only Look Once (YOLO) Series," jun 2024. [Online]. Available: <https://arxiv.org/abs/2406.19407v6>
- [25] S. J. Kim and M. J. Kim, "Human Action Recognition Using YOLOv11Ultralytics: A Comprehensive Study for Real-Time Applications," *Journal of Advanced Artificial Intelligence, Engineering and Technology*, vol. 01, no. 02, 2025. [Online]. Available: <https://artificial-intelligence.engineering-technology.wren-research-journals.com/1/article/view/14/10>
- [26] N. Jegham, C. Y. Koh, M. Abdelatti, and A. Hendawi, "YOLO Evolution: A Comprehensive Benchmark and Architectural Review of YOLOv12, YOLO11, and Their Previous Versions," pp. 1–20, mar 2025. [Online]. Available: <https://arxiv.org/abs/2411.00201>
- [27] I. Ofeidis, D. Kiedanski, and L. Tassiulas, "An Overview of the Data-Loader Landscape: Comparative Performance Analysis," *Proceedings - 2024 IEEE International Conference on Big Data, BigData 2024*, pp. 360–367, sep 2022. [Online]. Available: <https://arxiv.org/pdf/2209.13705>
- [28] V. Iglovikov, "Need for Speed: A Comprehensive Benchmark of JPEG Decoders in Python," pp. 1–7, jan 2025. [Online]. Available: <http://arxiv.org/abs/2501.13131>
- [29] G. Leclerc, A. Ilyas, L. Engstrom, S. M. Park, H. Salman, and A. Mądry, "FFCV: Accelerating Training by Removing Data Bottlenecks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2023-June, pp. 12 011–12 020, jun 2023. [Online]. Available: <https://arxiv.org/pdf/2306.12517>
- [30] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electronic Markets*, vol. 31, no. 3, pp. 685–695, sep 2021. [Online]. Available: <https://link.springer.com/article/10.1007/s12525-021-00475-2>
- [31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," *Advances in Neural Information Processing Systems*, vol. 2017-Decem, pp. 5999–6009, jun 2017. [Online]. Available: <https://arxiv.org/abs/1706.03762v5><http://arxiv.org/abs/1706.03762>

- [32] Y. Gong, G. Liu, Y. Xue, R. Li, and L. Meng, "A survey on dataset quality in machine learning," *Information and Software Technology*, vol. 162, p. 107268, oct 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0950584923001222>
- [33] V. Kalokyri, N. S. Tachos, C. N. Kalantzopoulos, S. Sfakianakis, H. Kondylakis, D. I. Zaridis, S. Colantonio, D. Regge, N. Papanikolaou, T. P.-I. Consortium, K. Marias, D. I. Fotiadis, and M. Tsiknakis, "AI Model Passport: Data and System Traceability Framework for Transparent AI in Health," jun 2025. [Online]. Available: <https://arxiv.org/pdf/2506.22358>
- [34] M. Steidl, M. Felderer, and R. Ramler, "The pipeline for the continuous development of artificial intelligence models—Current state of research and practice," *Journal of Systems and Software*, vol. 199, p. 111615, may 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0164121223000109>
- [35] D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, M. Young, J.-F. Crespo, and D. Dennison, "Hidden Technical Debt in Machine Learning Systems," *Advances in Neural Information Processing Systems*, vol. 28, 2015.
- [36] N. Polyzotis, S. Roy, S. E. Whang, and M. Zinkevich, "Data management challenges in production machine learning," *Proceedings of the ACM SIGMOD International Conference on Management of Data*, vol. Part F1277, pp. 1723–1726, may 2017. [Online]. Available: <https://dl.acm.org/doi/pdf/10.1145/3035918.3054782>
- [37] J. Lanchantin, T. Wang, V. Ordonez, and Y. Qi, "General Multi-Label Image Classification With Transformers," pp. 16 478–16 488, 2021.
- [38] E. Zimmermann, J. Szeto, J. Pasquero, and F. Ratle, "Benchmarking a Benchmark: How Reliable is MS-COCO?" nov 2023. [Online]. Available: <https://arxiv.org/abs/2311.02709v1><https://arxiv.org/abs/2311.02709>
- [39] Z. Huang, "Research on Segmentation Technology of Process Image Based on Segmentation Algorithm," *Proceedings - 2023 International Conference on Computer Simulation and Modeling, Information Security, CSMIS 2023*, pp. 392–398, 2023.
- [40] Q. Huang, Y. Xiong, and D. Lin, "Unifying Identification and Context Learning for Person Recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2217–2225, jun 2018. [Online]. Available: <https://arxiv.org/pdf/1806.03084>
- [41] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8693 LNCS, no. PART 5. Springer Verlag, may 2014, pp. 740–755.
- [42] C. Lyu, W. Zhang, H. Huang, Y. Zhou, Y. Wang, Y. Liu, S. Zhang, and K. Chen, "RTMDet: An Empirical Study of Designing Real-Time Object Detectors," dec 2022. [Online]. Available: <https://arxiv.org/pdf/2212.07784>

- [43] C. Gu, C. Sun, D. A. Ross, C. Vondrick, C. Pantofaru, Y. Li, S. Vijayanarasimhan, G. Toderici, S. Ricco, R. Sukthankar, C. Schmid, and J. Malik, “AVA: A Video Dataset of Spatio-temporally Localized Atomic Visual Actions,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 6047–6056, may 2017. [Online]. Available: <https://arxiv.org/abs/1705.08421v4>
- [44] J. Zhao, Y. Zhang, X. Li, H. Chen, B. Shuai, M. Xu, C. Liu, K. Kundu, Y. Xiong, D. Modolo, I. Marsic, C. G. Snoek, and J. Tighe, “TubeR: Tubelet Transformer for Video Action Detection,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2022-June, pp. 13 588–13 597, apr 2021. [Online]. Available: <https://arxiv.org/pdf/2104.00969>
- [45] D. R. Beddiar, B. Nini, M. Sabokrou, and A. Hadid, “Vision-based human activity recognition: a survey,” *Multimedia Tools and Applications*, vol. 79, no. 41-42, pp. 30 509–30 555, nov 2020. [Online]. Available: <https://link.springer.com/article/10.1007/s11042-020-09004-3>
- [46] N. Gupta, S. K. Gupta, R. K. Pathak, V. Jain, P. Rashidi, and J. S. Suri, “Human activity recognition in artificial intelligence framework: a narrative review,” *Artificial Intelligence Review*, vol. 55, no. 6, pp. 4755–4808, jan 2022. [Online]. Available: <https://link.springer.com/article/10.1007/s10462-021-10116-x>
- [47] F. Shafizadegan, A. R. Naghsh-Nilchi, and E. Shabaninia, “Multimodal vision-based human action recognition using deep learning: a review,” *Artificial Intelligence Review*, vol. 57, no. 7, p. 178, jun 2024. [Online]. Available: <https://link.springer.com/10.1007/s10462-024-10730-5>
- [48] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem. IEEE Computer Society, jun 2016, pp. 779–788. [Online]. Available: <https://arxiv.org/abs/1506.02640v5>
- [49] J. Redmon and A. Farhadi, “YOLO9000: Better, Faster, Stronger,” *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 6517–6525, dec 2016. [Online]. Available: <https://arxiv.org/abs/1612.08242v1>
- [50] —, “YOLOv3: An Incremental Improvement,” Tech. Rep. [Online]. Available: <https://pjreddie.com/yolo/>.<https://pjreddie.com/media/files/papers/YOLOv3.pdf>
- [51] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” *arXiv*, apr 2020. [Online]. Available: <http://arxiv.org/abs/2004.10934>
- [52] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “Scaled-YOLOv4: Scaling Cross Stage Partial Network,” nov 2020. [Online]. Available: <http://arxiv.org/abs/2011.08036>
- [53] T. Li, J. Liu, W. Zhang, Y. Ni, W. Wang, and Z. Li, “UAV-Human: A large benchmark for human behavior understanding with unmanned aerial vehicles,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 16 261–16 270, 2021.

- [54] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” jul 2022. [Online]. Available: <https://arxiv.org/abs/2207.02696v1>
- [55] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, “YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information,” feb 2024. [Online]. Available: <https://arxiv.org/abs/2402.13616v2>
- [56] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, and G. Ding, “YOLOv10: Real-Time End-to-End Object Detection,” no. NeurIPS, pp. 1–28, may 2024. [Online]. Available: <http://arxiv.org/abs/2405.14458>
- [57] “Brief summary of YOLOv8 model structure. Issue ultralytics/ultralytics.” [Online]. Available: <https://github.com/ultralytics/ultralytics/issues/189>
- [58] P. Hidayatullah, N. Syakrani, M. R. Sholahuddin, T. Gelar, and R. Tubagus, “YOLOv8 to YOLO11: A Comprehensive Architecture In-depth Comparative Review,” apr 2025. [Online]. Available: <http://arxiv.org/abs/2501.13400>
- [59] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, “YOLOX: Exceeding YOLO Series in 2021,” pp. 1–7, aug 2021. [Online]. Available: <http://arxiv.org/abs/2107.08430>
- [60] J. Carreira and A. Zisserman, “Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset,” *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2017-Janua, pp. 4724–4733, feb 2018. [Online]. Available: <https://arxiv.org/abs/1705.07750v3><http://arxiv.org/abs/1705.07750><http://ieeexplore.ieee.org/document/8099985/>
- [61] C. Feng, Y. Zhong, Y. Gao, M. R. Scott, and W. Huang, “TOOD: Task-aligned One-stage Object Detection,” *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3490–3499, aug 2021. [Online]. Available: <http://arxiv.org/abs/2108.07755>
- [62] Albumentations, “Image Benchmarks.” [Online]. Available: <https://albumentations.ai/docs/benchmarks/image-benchmarks/>
- [63] “Penonton Kerumunan Stadion - Foto gratis di Pixabay.” [Online]. Available: <https://pixabay.com/id/photos/penonton-kerumunan-stadion-1536341/>
- [64] Heart News West Country, “Fans at Ashton Gate Stadium, Bristol, celebrate England winner against Wales at Euro 2016,” Jun. 2016. [Online]. Available: <https://www.youtube.com/watch?v=Jvmmkx1dtmA>