References

[1] S. Singh, "Cousins of artificial intelligence," https://towardsdatascience.com/cousins-of-artificial-intelligence-dda4edc27b55, May 2018, accessed: (Sept, 2018).

[2] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, 1998, vol. 86, no. 11, pp. 2278–2324.

[3] T. Pfister, J. Charles, and A. Zisserman, "Flowing ConvNets for Human Pose Estimation in Videos," in *2015 IEEE International Conference on Computer Vision (ICCV)*. IEEE, Dec. 2015, pp. 1913–1921.

[4] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, Oct. 2017, pp. 2999–3007.

[5] M. R. Future, "Global User Interface Service Market, Type (Mobile Interface, Human Machine Interface, Query User Interface, Web Service Interface), Application (Education, Retail, Healthcare, Government, Market Intelligence, Consumer Electronics) - Forecast till 2030," https://www.marketresearchfuture.com/reports/user-interface-services-market-1086, 2022.

[6] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly, "Vision-based hand pose estimation: A review," *Computer Vision and Image Understanding*, Oct. 2007, vol. 108, no. 1-2, pp. 52–73.

[7] B. Doosti, "Hand pose estimation: A survey," *arXiv preprint arXiv:1903.01013*, Mar. 2019.

[8] M. A. Ahmed, B. B. Zaidan, A. A. Zaidan, M. M. Salih, and M. M. bin Lakulu, "A review on systems-based sensory gloves for sign language recognition state of the art between 2007 and 2017," *Sensors*, Jul. 2018, vol. 18, no. 7, p. 2208.

[9] G. Modanwal and K. Sarawadekar, "A New Dactylology and Interactive System Development for Blind–Computer Interaction," *IEEE Transactions on Human-Machine Systems*, Apr. 2018, vol. 48, no. 2, pp. 207–212.

[10] D. Gupta, B. Artacho, and A. Savakis, "Handypose: Multi-level framework for hand pose estimation," *Zum '98: The Z Formal Specification Notation*, 2022, vol. 128, p. 108674. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320322001558

[11] S. Jiang, P. Kang, X. Song, B. Lo, and P. Shull, "Emerging wearable interfaces and algorithms for hand gesture recognition: A survey," *IEEE Reviews in Biomedical Engineering*, 2022, vol. 15, pp. 85–102.

[12] W. Wu, C. Li, Z. Cheng, X. Zhang, and L. Jin, "YOLSE: Egocentric fingertip detection from single RGB images," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. IEEE, Oct. 2017, pp. 623–630.

[13] Y. Wang, C. Peng, and Y. Liu, "Mask-Pose Cascaded CNN for 2D Hand Pose Estimation From Single Color Image," *IEEE Transactions on Circuits and Systems for Video Technology*, Nov. 2019, vol. 29, no. 11, pp. 3258–3268.

[14] Y. Wang, B. Zhang, and C. Peng, "SRHandNet: Real-Time 2D Hand Pose Estimation With Simultaneous Region Localization," *IEEE Transactions on Image Processing*, 2020, vol. 29, pp. 2977–2986.

[15] W. Zhou, X. Jiang, X. Chen, S. Miao, C. Chen, S. Mei, and Y.-H. Liu, "HMTNet: 3d hand pose estimation from single depth image based on hand morphological topology," *IEEE Sensors Journal*, Jun. 2020, vol. 20, no. 11, pp. 6004–6011.

[16] Grzejszczak, Tomasz and Kawulok, Michal and Galuszka, Adam, "Hand landmarks detection and localization in color images," *Multimedia Tools and Applications*, Sep. 2016, vol. 75, no. 23, pp. 16 363–16 387.

[17] F. Gomez-Donoso, S. Orts-Escolano, and M. Cazorla, "Large-scale Multiview 3D Hand Pose Dataset," *Image and Vision Computing*, Jan. 2019, vol. 81, pp. 25–33.

[18] H. Joo, T. Simon, X. Li, H. Liu, L. Tan, L. Gui, S. Banerjee, T. Godisart, B. Nabbe, I. Matthews, T. Kanade, S. Nobuhara, and Y. Sheikh, "Panoptic studio: A massively multiview system for social interaction capture," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Jan. 2019, vol. 41, no. 1, pp. 190–204.

[19] C. Zimmermann, D. Ceylan, J. Yang, B. Russell, M. J. Argus, and T. Brox, "Freihand: A dataset for markerless capture of hand pose and shape from single rgb images," in *IEEE International Conference on Computer Vision (ICCV)*. IEEE, Oct. 2019. [Online]. Available: https://lmb.informatik.uni-freiburg.de/projects/freihand/

[20] C. Zimmermann and T. Brox, "Learning to Estimate 3D Hand Pose from Single RGB Images," in *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, Oct. 2017, pp. 4913–4921.

[21] J. Zhang, J. Jiao, M. Chen, L. Qu, X. Xu, and Q. Yang, "A hand pose tracking benchmark from stereo matching," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, Sep. 2017, pp. 982–986.

[22] G. Garcia-Hernando, S. Yuan, S. Baek, and T.-K. Kim, "First-Person Hand Action Benchmark with Rgb-D Videos and 3d Hand Pose Annotations," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Jun. 2018, pp. 409–419.

[23] G. Moon, S.-I. Yu, H. Wen, T. Shiratori, and K. M. Lee, "InterHand2.6m: A dataset and baseline for 3d interacting hand pose estimation from a single RGB image," in *Computer Vision – ECCV 2020*. Springer International Publishing, 2020, pp. 548–564.

[24] F. Lin, C. Wilhelm, and T. Martinez, "Two-Hand Global 3D Pose Estimation Using Monocular RGB," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, Jan. 2021.

[25] S. Brahmbhatt, C. Tang, C. D. Twigg, C. C. Kemp, and J. Hays, *ContactPose: A Dataset of Grasps with Object Contact and Hand Pose*. Springer International Publishing, Aug. 2020, pp. 361–378.

[26] X. Zhang, Q. Li, H. Mo, W. Zhang, and W. Zheng, "End-to-end hand mesh recovery from a monocular RGB image," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Oct. 2019, pp. 2354–2364.

[27] Z. Yu, L. Yang, S. Chen, and A. Yao, "Local and Global Point Cloud Reconstruction for 3d Hand Pose Estimation," *arXiv preprint arXiv:2112:06389*, Dec. 2021.

[28] M. Gardner and S. Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmospheric Environment*, Aug. 1998, vol. 32, no. 14-15, pp. 2627–2636.

[29] H. D. Block, "The Perceptron: A Model for Brain Functioning. I," *Reviews of Modern Physics*, Jan. 1962, vol. 34, no. 1, pp. 123–135.

[30] S. Gallant, "Perceptron-based learning algorithms," *IEEE Transactions on Neural Networks*, Jun. 1990, vol. 1, no. 2, pp. 179–191.

[31] M. Browne and S. S. Ghidary, "Convolutional neural networks for image processing: An application in robot vision," in *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2003, pp. 641–652.

[32] L. N. Smith, "Cyclical learning rates for training neural networks," in *Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on*, IEEE. IEEE, Mar. 2017, pp. 464–472.

[33] I. Loshchilov and F. Hutter, "SGDR: Stochastic Gradient Descent with Warm Restarts," *arXiv preprint arXiv:1608.03983*, Aug. 2016.

[34] W. Liu, A. Rabinovich, and A. C. Berg, "Parsenet: Looking Wider to See Better," *arXiv preprint arXiv:1506.04579*, Jun. 2015.

[35] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Apr. 2018, vol. 40, no. 4, pp. 834–848.

[36] A. Krizhevsky, "Learning Multiple Layers of Features from Tiny Images," Univ. Toronto, Technical Report, 2009.

[37] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Jun. 2009, pp. 248–255.

[38] S. Zagoruyko and N. Komodakis, "Wide Residual Networks," in *Procedings of the British Machine Vision Conference 2016*. British Machine Vision Association, 2016.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Identity Mappings in Deep Residual Networks," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 630–645.

[40] ——, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jun. 2016, pp. 770–778.

[41] G. Modanwal, "Development of a Novel Dactylology and Human Computer Interface Design for Visually Impaired," Ph.D. dissertation, IIT (BHU) varanasi, 2018.

[42] K. Oka, Y. Sato, and H. Koike, "Real-Time Tracking of Multiple Fingertips and Gesture Recognition for Augmented Desk Interface Systems," in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, ser. AFGR-02. IEEE, 2002, pp. 429–434.

[43] Z. Mo, J. P. Lewis, and U. Neumann, "Smartcanvas: A gesture-driven intelligent drawing desk system," in *Proceedings of the 10th international conference on Intelligent user interfaces*, ser. IUI05. ACM, Jan. 2005, pp. 239–243.

[44] J. Grudin, "Integrating paper and digital information on enhanceddesk: A method for realtime finger tracking on an augmented desk system," *ACM Trans. Comput.-Hum. Interact.*, Dec. 2001, vol. 8, no. 4, pp. 307–322.

[45] K. Li and X. Zhang, "A new fingertip detection and tracking algorithm and its application on writing-in-the-air system," in *2014 7th International Congress on Image and Signal Processing*, 2014, pp. 457–462.

[46] X. Zhang, Z. Ye, L. Jin, Z. Feng, and S. Xu, "A New Writing Experience: Finger Writing in the Air Using a Kinect Sensor," *IEEE MultiMedia*, 2013, vol. 20, no. 4, pp. 85–93.

[47] Y. Huang, X. Liu, X. Zhang, and L. Jin, "A Pointing Gesture Based Egocentric Interaction System: Dataset, Approach and Application," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2016, pp. 370–377.

[48] S. Mukherjee, S. A. Ahmed, D. P. Dogra, S. Kar, and P. P. Roy, "Fingertip detection and tracking for recognition of air-writing in videos," *Expert Systems with Applications*, 2019, vol. 136, pp. 217–229.

[49] P. Mistry and P. Maes, "SixthSense: A Wearable Gestural Interface," in *ACM SIGGRAPH ASIA 2009 Sketches*, ser. SIGGRAPH ASIA '09. New York, NY, USA: ACM, 2009, pp. 11:1–11:1.

[50] G. Modanwal, P. Mishra, and K. Sarawadekar, "A Robust Algorithm for Hand-Forearm Segmentation," in *Proceedings of the 2018 International Conference on Image and Graphics Processing*. ACM, 2018, pp. 102–105.

[51] G. Modanwal and K. Sarawadekar, "A Robust Wrist Point Detection Algorithm Using Geometric Features," *Pattern Recognition Letters*, Jul. 2018, vol. 110, pp. 72–78.

[52] V. Buchmann, S. Violich, M. Billinghurst, and A. Cockburn, "FingARtips: Gesture based direct manipulation in Augmented Reality," in *Proc. 2nd int. conf. CGIT*. ACM, 2004, pp. 212–221.

[53] K. Dorfmuller-Ulhaas and D. Schmalstieg, "Finger tracking for interaction in augmented environments," in *Proceedings IEEE and ACM International Symposium on Augmented Reality*, Oct. 2001, pp. 55–64.

[54] R. Zhang and X. Zhang, "Interaction Method Based on Data Glove in Virtual Environment," *Computer Engineering*, 2005, vol. 12.

[55] Y. Bai, Y. Zhang, M. Ding, and B. Ghanem, "Finding Tiny Faces in the Wild with Generative Adversarial Network," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Jun. 2018.

[56] B. H. Thomas and W. Piekarski, "Glove based user interaction techniques for augmented reality in an outdoor environment," *Virtual Reality*, Oct. 2002, vol. 6, no. 3, pp. 167–180.

[57] G. Gordon, M. Billinghurst, M. Bell, J. Woodfill, B. Kowalik, A. Erendi, and J. Tilander, "The Use of Dense Stereo Range Data in Augmented Reality," in *Proceedings. International Symposium on Mixed and Augmented Reality*, ser. ISMAR-02. IEEE Comput. Soc, 2002, pp. 14–23.

[58] T. Nakamura, S. Takahashi, and J. Tanaka, "Double-Crossing: A New Interaction Technique for Hand Gesture Interfaces," in *Asia-Pacific Conference on Computer Human Interaction*. Springer, 2008, pp. 292–300.

[59] Y. Wang, J. Min, J. Zhang, Y. Liu, F. Xu, Q. Dai, and J. Chai, "Video-Based Hand Manipulation Capture through Composite Motion Control," *ACM Trans. Graph.*, Jul. 2013, vol. 32, no. 4.

[60] Y. Xu, J. Gu, Z. Tao, and D. Wu, "Bare Hand Gesture Recognition with a Single Color Camera," in *2009 2nd International Congress on Image and Signal Processing*. IEEE, Oct. 2009.

[61] S. Phung, A. Bouzerdoum, and D. Chai, "Skin segmentation using color pixel classification: analysis and comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Jan. 2005, vol. 27, no. 1, pp. 148–154.

[62] G. Wu and W. Kang, "Robust Fingertip Detection in a Complex Environment," *IEEE Transactions on Multimedia*, Jun. 2016, vol. 18, no. 6, pp. 978–987.

[63] Lae-Kyoung Lee, Su-Yong An, and Se-Young Oh, "Robust fingertip extraction with improved skin color segmentation for finger gesture recognition in Human-robot interaction," in *2012 IEEE Congress on Evolutionary Computation*, Jun. 2012, pp. 1–7.

[64] C. Shan, T. Tan, and Y. Wei, "Real-time hand tracking using a mean shift embedded particle filter," *Pattern Recognition*, Jul. 2007, vol. 40, no. 7, pp. 1958–1970.

[65] R. Khan, A. Hanbury, J. Stöttinger, and A. Bais, "Color based skin classification," *Pattern Recognition Letters*, Jan. 2012, vol. 33, no. 2, pp. 157–163.

[66] K. Imagawa, S. Lu, and S. Igi, "Color-Based Hands Tracking System for Sign Language Recognition," in *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, ser. AFGR-98. IEEE Comput. Soc, 1998, pp. 462–467.

[67] D. J. Sawicki and W. Miziolek, "Human colour skin detection in CMYK colour space," *IET Image Processing*, Sep. 2015, vol. 9, no. 9, pp. 751–757.

[68] X. Yin and M. Xie, "Finger identification and hand posture recognition for human–robot interaction," *Image and Vision Computing*, Aug. 2007, vol. 25, no. 8, pp. 1291–1300.

[69] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Lecture Notes in Computer Science*. Springer International Publishing, 2015, pp. 234–241.

[70] Á. García-Casarrubios Muñoz, C. Sánchez Ávila, A. de Santos Sierra, and J. Guerra Casanova, *A Mobile-Oriented Hand Segmentation Algorithm Based on Fuzzy Multiscale Aggregation*. Springer Berlin Heidelberg, 2010, pp. 479–488.

[71] M. W. Krueger, *An Easy Entry Artificial Reality*. Elsevier, 1993, pp. 147–161.

[72] J. Segen and S. Kumar, "Gesture VR," in *Proceedings of the sixth ACM international conference on Multimedia - MULTIMEDIA '98*. ACM Press, 1998.

[73] G. Wu and W. Kang, "Vision-Based Fingertip Tracking Utilizing Curvature Points Clustering and Hash Model Representation," *IEEE Transactions on Multimedia*, Aug. 2017, vol. 19, no. 8, pp. 1730–1741.

[74] Malima, Ozgur, and Cetin, "A Fast Algorithm for Vision-Based Hand Gesture Recognition for Robot Control," in *2006 IEEE 14th Signal Processing and Communications Applications*, Apr. 2006, pp. 1–4.

[75] J. Segen and S. Kumar, "Human-computer interaction using gesture recognition and 3d hand tracking," in *Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No.98CB36269)*, Oct. 1998, pp. 188–192 vol.3.

[76] R. M. Gurav and P. K. Kadbe, "Real time finger tracking and contour detection for gesture recognition using OpenCV," in *2015 International Conference on Industrial Instrumentation and Control (ICIC)*, May 2015, pp. 974–977.

[77] C. Zhang, G. Wang, X. Chen, and H. Yang, "Bi-stream Region Ensemble Network: Promoting Accuracy in Fingertip Localization from Stereo Images," in *British Machine Vision Conference 2018, BMVC 2018, Newcastle, UK, September 3-6, 2018*. BMVA Press, 2018, p. 314.

[78] G. Wang, C. Zhang, X. Chen, X. Ji, J.-H. Xue, and H. Wang, "Bi-Stream Pose-Guided Region Ensemble Network for Fingertip Localization From Stereo Images," *IEEE Transactions on Neural Networks and Learning Systems*, Dec. 2020, vol. 31, no. 12, pp. 5153–5165.

[79] H. Liang, J. Yuan, and D. Thalmann, "Parsing the hand in depth images," *IEEE Transactions on Multimedia*, Aug. 2014, vol. 16, no. 5, pp. 1241–1253.

[80] C. Wang, Z. Liu, and S. Chan, "Superpixel-Based Hand Gesture Recognition With Kinect Depth Camera," *IEEE Trans. on Multimedia*, Jan. 2015, vol. 17, no. 1, pp. 29–39.

[81] O. Choi, Y.-J. Son, H. Lim, and S. C. Ahn, "Co-Recognition of Multiple Fingertips for Tabletop Human–Projector Interaction," *IEEE Transactions on Multimedia*, Jun. 2019, vol. 21, no. 6, pp. 1487–1498.

[82] Y. Huang, X. Liu, L. Jin, and X. Zhang, "DeepFinger: A Cascade Convolutional Neuron Network Approach to Finger Key Point Detection in Egocentric Vision with Mobile Camera," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, Oct. 2015, pp. 2944–2949.

[83] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional Pose Machines," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jun. 2016, pp. 4724–4732.

[84] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, "Hand Keypoint Detection in Single Images Using Multiview Bootstrapping," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jul. 2017, pp. 4645–4653.

[85] M. Li, J. Wang, and N. Sang, "Latent Distribution-Based 3D Hand Pose Estimation From Monocular RGB Images," *IEEE Transactions on Circuits and Systems for Video Technology*, Dec. 2021, vol. 31, no. 12, pp. 4883–4894.

[86] S. Guo, E. Rigall, Y. Ju, and J. Dong, "3D Hand Pose Estimation From Monocular RGB With Feature Interaction Module," *IEEE Transactions on Circuits and Systems for Video Technology*, Aug. 2022, vol. 32, no. 8, pp. 5293–5306.

[87] X. Sun, B. Xiao, F. Wei, S. Liang, and Y. Wei, *Integral Human Pose Regression*. Springer International Publishing, 2018, pp. 536–553.

[88] B. Yu and D. Tao, "Heatmap Regression Via Randomized Rounding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Nov. 2022, vol. 44, no. 11, pp. 8276–8289.

[89] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[90] S. Ong and S. Ranganath, "Automatic Sign Language Analysis: A Survey and the Future beyond Lexical Meaning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Jun. 2005, vol. 27, no. 6, pp. 873–891.

[91] K. Oka, Y. Sato, and H. Koike, "Real-Time Fingertip Tracking and Gesture Recognition," *IEEE Computer Graphics and Applications*, Nov. 2002, vol. 22, no. 6, pp. 64–71.

[92] M. Baldauf, S. Zambanini, P. Fröhlich, and P. Reichl, "Markerless Visual Fingertip Detection for Natural Mobile Device Interaction," in *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '11. New York, NY, USA: Association for Computing Machinery, 2011, pp. 539–544.

[93] Z. Ren, J. Meng, and J. Yuan, "Depth Camera Based Hand Gesture Recognition and Its Applications in Human-computer-interaction," in *2011 8th International Conference on Information, Communications & Signal Processing*. IEEE, 2011, pp. 1–5.

[94] S. Sridhar, A. Oulasvirta, and C. Theobalt, "Interactive Markerless Articulated Hand Motion Tracking Using RGB and Depth Data," in *2013 IEEE International Conference on Computer Vision*. IEEE, Dec. 2013, pp. 2456–2463.

[95] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jul. 2017, pp. 7263–7271.

[96] ——, "YOLOv3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[97] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv preprint arXiv:2004.10934*, 2020.

[98] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single Shot Multibox Detector," in *Proc. ECCV*. Springer, 2016, pp. 21–37.

[99] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, Oct. 2017, pp. 2980–2988.

[100] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in neural information processing systems*, vol. 39, no. 6. Institute of Electrical and Electronics Engineers (IEEE), Jun. 2015, pp. 91–99.

[101] P. Mishra and K. Sarawadekar, "Fingertips Detection in Egocentric Video Frames using Deep Neural Networks," in *International Conference on Image and Vision Computing New Zealand (IVCNZ)*, Dec. 2019, pp. 1–6.

[102] T. Amemiya, *Non-Linear Regression Models*. Elsevier, 1983, pp. 333–389.

[103] G. Tsoumakas and I. Katakis, "Multi-label classification," *International Journal of Data Warehousing and Mining*, Jul. 2007, vol. 3, no. 3, pp. 1–13.

[104] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Jun. 2018, pp. 4510–4520.

[105] I. J. Good, "Rational decisions," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1952, vol. 14, no. 1, pp. 107–114. [Online]. Available: http://www.jstor.org/stable/2984087

[106] G. Cybenko, D. P. O'Leary, and J. Rissanen, Eds., *The Mathematics of Information Coding, Extraction and Distribution*. Springer New York, 1999.

[107] W. James and C. Stein, "Estimation with quadratic loss," in *Springer Series in Statistics*. Springer New York, 1992, pp. 443–460.

[108] P. Mishra and K. Sarawadekar, "Polynomial Learning Rate Policy with Warm Restart for Deep Neural Network," in *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, Oct. 2019, pp. 2087–2092.

[109] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A New Backbone that can Enhance Learning Capability of CNN," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, Jun. 2020, pp. 1571–1580.

[110] P. J. Huber, "Robust Estimation of a Location Parameter," in *Breakthroughs in statistics*. Springer, 1992, pp. 492–518.

[111] Y. Yang and D. Ramanan, "Articulated Human Detection with Flexible Mixtures of Parts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Dec. 2013, vol. 35, no. 12, pp. 2878–2890.

[112] 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)*. IEEE, Jul. 2017, pp. 2261–2269.

[113] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on Imagenet Classification," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec. 2015, pp. 1026–1034.

[114] J. Liu, H. Ding, A. Shahroudy, L.-Y. Duan, X. Jiang, G. Wang, and A. C. Kot, "Feature Boosting Network For 3D Pose Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Feb. 2020, vol. 42, no. 2, pp. 494–501.

[115] S. Yang, J. Liu, S. Lu, M. H. Er, and A. C. Kot, "Collaborative learning of gesture recognition and 3d hand pose estimation with multi-order feature analysis," in *Computer Vision – ECCV 2020*. Springer International Publishing, 2020, pp. 769–786.

[116] Z. Fan, J. Liu, and Y. Wang, "Adaptive Computationally Efficient Network for Monocular 3D Hand Pose Estimation," in *Computer Vision – ECCV 2020*. Springer International Publishing, 2020, pp. 127–144.

[117] S. Gattupalli, A. R. Babu, J. R. Brady, F. Makedon, and V. Athitsos, "Towards Deep Learning Based Hand Keypoints Detection for Rapid Sequential Movements from RGB Images," in *Proc. PETRA*, ser. PETRA '18. Corfu, Greece: ACM, Jun. 2018, pp. 31–37.

[118] S. Guo, E. Rigall, L. Qi, X. Dong, H. Li, and J. Dong, "Graph-Based CNNs with Self-Supervised Module for 3D Hand Pose Estimation from Monocular RGB," *IEEE Transactions on Circuits and Systems for Video Technology*, Apr. 2021, vol. 31, no. 4, pp. 1514–1525.

[119] P. Panteleris, I. Oikonomidis, and A. Argyros, "Using a Single RGB Frame for Real Time 3d Hand Pose Estimation in the Wild," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, Mar. 2018, pp. 436–445.

[120] M. M. Alam, M. T. Islam, and S. M. Rahman, "Unified learning approach for egocentric hand gesture recognition and fingertip detection," *Zum '98: The Z Formal Specification Notation*, 2022, vol. 121, p. 108200.

[121] D. Avola, L. Cinque, A. Fagioli, G. L. Foresti, A. Fragomeni, and D. Pannone, "3D Hand Pose and Shape Estimation from RGB Images for Keypoint-Based Hand Gesture Recognition," *Zum '98: The Z Formal Specification Notation*, Sep. 2022, vol. 129, p. 108762.

[122] P. Mishra and K. P. Sarawadekar, "Fingertips Detection With Nearest-Neighbor Pose Particles From a Single RGB Image," *IEEE Transactions on Circuits and Systems for Video Technology*, May 2022, vol. 32, no. 5, pp. 3001–3011.

[123] J. Sanchez-Riera, K. Srinivasan, K. Hua, W. Cheng, M. A. Hossain, and M. F. Alhamid, "Robust RGB-D Hand Tracking Using Deep Learning Priors," *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, vol. 28, no. 9, pp. 2289–2301.

[124] X. Sun, Y. Wei, S. Liang, X. Tang, and J. Sun, "Cascaded Hand Pose Regression," in *Proc. CVPR*. IEEE, Jun. 2015, pp. 824–832.

[125] C. Wan, A. Yao, and L. Van Gool, "Hand Pose Estimation from Local Surface Normals," in *Proc. ECCV*, Springer. Springer International Publishing, 2016, pp. 554–569.

[126] F. Xiong, B. Zhang, Y. Xiao, Z. Cao, T. Yu, J. T. Zhou, and J. Yuan, "A2J: Anchor-to-Joint Regression Network for 3D Articulated Pose Estimation From a Single Depth Image," in *Proc. ICCV*. IEEE, Oct. 2019, pp. 793–802.

[127] V. Athitsos and S. Sclaroff, "Estimating 3D Hand Pose from a Cluttered Image," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern*

Recognition, 2003. Proceedings., ser. CVPR-03, vol. 2. IEEE Comput. Soc, 2003, pp. II–432.

[128] G. Rogez, M. Khademi, J. Supančič III, J. M. M. Montiel, and D. Ramanan, "3D hand pose detection in egocentric RGB-D images," in *Proc.ECCV*, Springer. Springer International Publishing, 2014, pp. 356–371.

[129] M. Baydoun, A. Betancourt, P. Morerio, L. Marcenaro, M. Rauterberg, and C. Regazzoni, "Hand Pose Recognition in First Person Vision through Graph Spectral Analysis," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2017, pp. 1872–1876.

[130] A. Newell, K. Yang, and J. Deng, "Stacked Hourglass Networks for Human Pose Estimation," in *Computer Vision – ECCV 2016*. Cham: Springer International Publishing, 2016, pp. 483–499.

[131] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep High-Resolution Representation Learning for Human Pose Estimation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jun. 2019, pp. 5693–5703.

[132] U. Iqbal, P. Molchanov, T. Breuel, J. Gall, and J. Kautz, "Hand pose estimation via latent 2.5d heatmap regression," in *Computer Vision – ECCV 2018*. Springer International Publishing, 2018, pp. 125–143.

[133] Y. Li, X. Wang, W. Liu, and B. Feng, "Pose Anchor: A Single-Stage Hand Keypoint Detection Network," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, vol. 30, no. 7, pp. 2104–2113.

[134] N. Santavas, I. Kansizoglou, L. Bampis, E. Karakasis, and A. Gasteratos, "Attention! A Lightweight 2d Hand Pose Estimation Approach," *IEEE Sensors Journal*, May 2021, vol. 21, no. 10, pp. 11 488–11 496.

[135] C. Wan, T. Probst, L. V. Gool, and A. Yao, "Dense 3D Regression for Hand Pose Estimation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Jun. 2018, pp. 5147–5156.

[136] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, vol. 39, no. 4, pp. 640–651.

[137] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015.

[138] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, 2015.

[139] *CMU Panoptic Studio Hand Dataset*, Accessed Sept. 14, 2019. [Online]. Available: http://domedb.perception.cs.cmu.edu/handdb.html

[140] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *Proc. ICML*, vol. 37, Lille, France, 2015, pp. 448–456.

[141] F. Chollet *et al.*, "Keras," https://keras.io, 2015.

[142] R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision (ICCV)*. IEEE, Dec. 2015, pp. 1440–1448.

[143] A. Virasova, D. Klimov, O. Khromov, I. Gubaidullin, and V. Oreshko, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," *Radioengineering*, 2021, pp. 115–126.

[144] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Sep. 2015, vol. 37, no. 9, pp. 1904–1916.

[145] Y. Zhu, R. Urtasun, R. Salakhutdinov, and S. Fidler, "segDeepM: Exploiting segmentation and context in deep neural networks for object detection," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 4703–4711.

[146] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective Search for Object Recognition," *International Journal of Computer Vision*, Apr. 2013, vol. 104, no. 2, pp. 154–171.

[147] D. Forsyth, "Object Detection with Discriminatively Trained Part-Based Models," *Computer*, 2014, vol. 47, no. 2, pp. 6–7.

[148] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *arXiv preprint arXiv:1312.6229*, 2013.

[149] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik, "Simultaneous Detection and Segmentation," in *Computer Vision – ECCV 2014*. Cham: Springer International Publishing, 2014, pp. 297–312.

[150] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik, "Hypercolumns for Object Segmentation and Fine-Grained Localization," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 447–456.

[151] J. Dai, K. He, Y. Li, S. Ren, and J. Sun, "Instance-Sensitive Fully Convolutional Networks," in *Proc. ECCV*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., 2016, pp. 534–549.

[152] A. Arnab and P. H. S. Torr, "Pixelwise Instance Segmentation with a Dynamically Instantiated Network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 879–888.

[153] M. Bai and R. Urtasun, "Deep Watershed Transform for Instance Segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2858–2866.

[154] A. Kirillov, E. Levinkov, B. Andres, B. Savchynskyy, and C. Rother, "Instancecut: From Edges to Instances with Multicut," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 7322–7331.

[155] K. Roy, A. Mohanty, and R. R. Sahay, "Deep Learning Based Hand Detection in Cluttered Environment Using Skin Segmentation," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017, pp. 640–649.

[156] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 936–944.

[157] A. Neubeck and L. Van Gool, "Efficient Non-Maximum Suppression," in *18^{th} International Conference on Pattern Recognition (ICPR'06)*, vol. 3, 2006, pp. 850–855.