VISION-BASED HUMAN PRESENCE DETECTION PIPELINE BY MEANS OF TRANSFER LEARNING APPROACH

TANG JIN CHENG

MASTER OF SCIENCE

UNIVERSITI MALAYSIA PAHANG

SUPERVISOR'S DECLARATION

We hereby declare that We have checked this thesis, and, in our opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Master of Science.

(Supervisor's Signature)

Full Name : DR. AHMAD FAKHRI AB. NASIR

Position : SENIOR LECTURER

Date : 28 / 06 / 2023

(Co-supervisor's Signature)

Full Name : DR. ANWAR P.P. ABDUL MAJEED

Position : ASSOCIATE PROFESSOR

Date : 28 / 06 / 2023

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

(Student's Signature)

Full Name : TANG JIN CHENG

ID Number : MFS 20002

Date : 28 / 06 / 2023

VISION-BASED HUMAN PRESENCE
DETECTION PIPELINE BY MEANS OF
A TRANSFER LEARNING
APPROACH

TANG JIN CHENG

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Master of Science

Faculty of Manufacturing and Mechatronic Engineering Technology
UNIVERSITI MALAYSIA PAHANG

JUNE 2023

ACKNOWLEDGEMENTS

I must admit that this part is the most difficult section for me to write on because countless people have supported me throughout the research path of this master's degree as well as the production of the thesis. This work would not have been possible without the constant support, guidance, and assistance regardless in terms of academic, financial, and individual growth. My first and big appreciation goes to my supervisors, Dr. Ahmad Fakhri Ab. Nasir, and Dr. Anwar P.P. Abdul Majeed. I greatly appreciate their marvellous supervision, encouragement and advises. Sincere gratitude is extended to their generous participation in kind support, constructive feedback, guiding and advice throughout my research.

This study was carried under the financial assistance of TT Vision Holdings Berhad, I would also like to take this opportunity to express my gratitude towards TT Vision Holdings Berhad of the wonderful collaboration that have provided from their end, as well as funding for the study in collaboration with Universiti Malaysia Pahang via UIC200816 and RDU202405. Here, I must also thank my field supervisor, Mr. Lim Thai Li, that has suggested some valuable opinion on the image acquisition process. Without their cooperation, it would be a tough time for me to acquire a proper image dataset for the research. Meanwhile, my humble thanks go to the Innovative Manufacturing, Mechatronics and Sports Laboratory members for being supportive when I have doubt in the direction of my research, and I really feel fortunate to have been a part of the research group.

Further, yet importantly, my family deserves my deepest and endless gratitude for their unconditional, unequivocal, and unfailing supports. They are the one who always behind my back whenever I needed them. I could not express how much motivation they have gave in every important moment. This honour and accomplishment should always belong to my family who educates me since I am young. Thank you.

ABSTRACT

Over the last century, industrial robots have gained an immense amount of popularity in replacing the human workers due to their highly repetitive nature. It was a twist to the industries when the concept of cooperative robots, known as cobots, has been innovated. Sharing space between the cobots and human workers has considered as the most effective way of utilizing the cobots. Keeping in mind that the safety of the human workers is always the top priority of the cobot applications in the industries, many time and efforts have been invested to improve the safeness of the cobots deployments. Yet, the utilization of deep learning technologies is rarely found in accordance with human detection in the field of research, especially the transfer learning approach, providing that the visual perception has shown to be a unique sense that still cannot be replaced by other. Hence, this thesis aimed to leverage the transfer learning approach to fine-tune the deep learning-based object detection models in the human detection task. In relation to this main goal, the objectives of the study were as follows: establish an image dataset for cobot environment from the surveillance cameras in TT Vision Holdings Berhad, formulate deep learning-based object detection models by using the transfer learning approach, and evaluate the performance of various transfer learning models in detecting the presence of human workers with relevant evaluation metrics. Image dataset has acquired from the surveillance system of TT Vision Holdings Berhad and annotated accordingly. The variations of the dataset have been considered thoroughly to ensure the models can be well-trained on the distinct features of the human workers. TensorFlow Object Detection API was used in the study to perform the fine-tuning of the one-stage object detectors. Among all the transfer learning strategies, fine-tuning has chosen since it suits the study well after the interpretation on the size-similarity matrix. A total of four EfficientDet models, two SSD models, three RetinaNet models, and four CenterNet models were deployed in the present work. As a result, SSD-MobileNetV2-FPN model has achieved 81.1% AP with 32.82 FPS, which is proposed as the best well-balanced fine-tuned model between accuracy and speed. In other case where the consideration is taken solely on either accuracy or inference speed, SSD_MobileNetV1-FPN model has attained 87.2% AP with 28.28 FPS and CenterNet-ResNet50-V1-FPN has achieved 78.0% AP with 46.52 FPS, which is proposed to be the model with best accuracy and inference speed, respectively. As a whole, it could be deduced that the transfer learning models can handle the human detection task well via the fine-tuning on the COCO-pretrained weights.

ABSTRAK

Sepanjang abad yang lalu, robot industri telah mendapat populariti yang besar dalam menggantikan pekerja manusia kerana sifatnya yang sangat berulang. Ia merupakan titik perubahan bagi industri apabila konsep robot kolaboratif, atau kobot, telah diinovasikan. Kongsi ruang antara kobot dan pekerja manusia dianggap sebagai cara yang paling berkesan dalam penggunaan kobot. Dengan mengingati behawa keselamatan pekerja manusia hendaklah sentiasa diutamakan bagi aplikasi kobot dalam industri, banyak masa and usaha telah dilaburkan dalam meningkatkan keselamatan penggunaan kobot. Namun, dalam bidang penyelidikan, penggunaan teknik pembelajaran mendalam jarang intermui dengan pengesanan manusia, terutamanya kaedah pemindahan pembelajaran, memandangkan persepsi visual merupakan satu persepsi unik yang susah digantikan oleh persepsi-persepsi yang lain. Oleh itu, tesis ini bertujuan untuk menggunakan kaedah pembelajaran pemindahan untuk memperhalusi model pengesanan objek berasaskan pembelajaran mendalam dalam tugas pengesanan manusia. Sehubungan dengan matlamat utama ini, objektif kajian adalah seperti berikut: mewujudkan set data imej persekitaran robot daripada kamera pengawasan TT Vision Holdings Berhad, merumus model pengesanan objek berasaskan pembelajaran mendalam dengan menggunakan pendekatan pembelajaran pemindahan, dan menilai prestasi pelbagai model pembelajaran pemindahan dalam mengesan kewujudan pekerja manusia. Dataset imej telah diperoleh daripada sistem pengawasan TT Vision Holdings Berhad dan diannotasikan. Variasi set data telah dipertimbangkan dengan teliti untuk memastikan model boleh belajar ciri-ciri pekerja manusia dengan baik. API pengesanan objek TensorFlow telah digunakan dalam kajian untuk melakukan penalasan halus bagi model berperingkat satu. Di antara semua strategi pembelajaran pemindahan, penalaan halus telah dipilih kerana ia sesuai dengan kajian ini dengan baik selepas mentafsir daripada matriks keserupaan dan saiz. Sebanyak empat model EfficientDet, dua model SSD, tiga model RetinaNet dan empat model CenterNet telah digunakan dalam kajian ini. Sebagai akibatnya, model SSD-MobileNetV2-FPN telah mencapai 81.1% AP dengan 32.82 bingkai per saat, dicadangkan sebagai model penalaan halus yang mempunyai keseimbangan baik antara ketepatan dan kelajuan inferens. Dalam kes lain di mana hanya ketepatan atau kelajuan inferens dipertimbangkan, model SSD_MobileNetV2-FPN telah mencapai 87.2% AP dengan 28.28 bingkai per saat dan CenterNet-ResNet50-V1-FPN telah mencapai 78.0% dengan 46.52 bingkai per saat, masing-masing dicadangkan sebagai model yang mempunyai ketepatan dan kelajuan inferens yang terbaik. Secara keseluruhannya, model pembelajaran pemindahan boleh mengendalikan tugas pengesanan menusia dengan baik melalui penalaan halus pada pemberat pra-latih oleh dataset COCO.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

REFERENCES

Aaltonen, I., & Salmi, T. (2019). Experiences and expectations of collaborative robots in industry and academia: barriers and development needs. *Procedia Manufacturing*, *38*(2019), 1151–1158. https://doi.org/10.1016/j.promfg.2020.01.204

Aaltonen, I., Salmi, T., & Marstio, I. (2018). Refining levels of collaboration to support the design and evaluation of human-robot interaction in the manufacturing industry. *Procedia CIRP*, *72*, 93–98. https://doi.org/10.1016/j.procir.2018.03.214

Ahmadi, M., Ouarda, W., & Alimi, A. M. (2020). Efficient and Fast Objects Detection Technique for Intelligent Video Surveillance Using Transfer Learning and Fine-Tuning. *Arabian Journal for Science and Engineering*, *45*(3), 1421–1433. https://doi.org/10.1007/s13369-019-03969-6

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Van Esesn, B. C., Awwal, A. A. S., & Asari, V. K. (2018). *The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches*. http://arxiv.org/abs/1803.01164

Amin, F. M., Rezayati, M., van de Venn, H. W., & Karimpour, H. (2020). A Mixed-Perception Approach for Safe Human–Robot Collaboration in Industrial Automation. *Sensors*, *20*(21), 6347. https://doi.org/10.3390/s20216347

Amisse, C., Jijón-Palma, M. E., & Centeno, J. A. S. (2021). FINE-TUNING DEEP LEARNING MODELS FOR PEDESTRIAN DETECTION. *Boletim de Ciências Geodésicas*, *27*(2), 0–1. https://doi.org/10.1590/s1982-21702021000200013

Bamne, B., Shrivastava, N., Parashar, L., & Singh, U. (2020). Transfer learning-based Object Detection by using Convolutional Neural Networks. *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, *Icesc*, 328–332. https://doi.org/10.1109/ICESC48915.2020.9156060

Bdiwi, M., Pfeifer, M., & Sterzing, A. (2017). A new strategy for ensuring human safety during various levels of interaction with industrial robots. *CIRP Annals*, *66*(1), 453–456. https://doi.org/10.1016/j.cirp.2017.04.009

Bi, Z. M., Luo, C., Miao, Z., Zhang, B., Zhang, W. J., & Wang, L. (2021). Safety assurance mechanisms of collaborative robotic systems in manufacturing. *Robotics and Computer-Integrated Manufacturing*, *67*(January 2020), 102022. https://doi.org/10.1016/j.rcim.2020.102022

Borwarnginn, P., Thongkanchorn, K., Kanchanapreechakorn, S., & Kusakunniran, W. (2019). Breakthrough Conventional Based Approach for Dog Breed Classification Using CNN with Transfer Learning. *2019 11th International Conference on Information Technology and Electrical Engineering (ICITEE)*, *7*, 1–5.

https://doi.org/10.1109/ICITEED.2019.8929955

Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, *2017-Janua*, 1800–1807. https://doi.org/10.1109/CVPR.2017.195

Du, G., Long, S., Li, F., & Huang, X. (2018). Active Collision Avoidance for Human-Robot Interaction With UKF, Expert System, and Artificial Potential Field Method. *Frontiers in Robotics and AI*, *5*, 1–11. https://doi.org/10.3389/frobt.2018.00125

Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., & Tian, Q. (2019). CenterNet: Keypoint triplets for object detection. *Proceedings of the IEEE International Conference on Computer Vision*, *2019-Octob*, 6568–6577. https://doi.org/10.1109/ICCV.2019.00667

El Zaatari, S., Marei, M., Li, W., & Usman, Z. (2019). Cobot programming for collaborative industrial tasks: An overview. *Robotics and Autonomous Systems*, *116*, 162–180. https://doi.org/10.1016/j.robot.2019.03.003

Farahani, A., Pourshojae, B., Rasheed, K., & Arabnia, H. R. (2020). A Concise Review of Transfer Learning. *Proceedings - 2020 International Conference on Computational Science and Computational Intelligence, CSCI 2020*, 344–351. https://doi.org/10.1109/CSCI51800.2020.00065

Flacco, F., Kroger, T., De Luca, A., & Khatib, O. (2012). A depth space approach to human-robot collision avoidance. *2012 IEEE International Conference on Robotics and Automation*, 338–345. https://doi.org/10.1109/ICRA.2012.6225245

Gao, S. H., Cheng, M. M., Zhao, K., Zhang, X. Y., Yang, M. H., & Torr, P. (2021). Res2Net: A New Multi-Scale Backbone Architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(2), 652–662. https://doi.org/10.1109/TPAMI.2019.2938758

Grau, A., Indri, M., Bello, L. Lo, & Sauter, T. (2017). Industrial robotics in factory automation: From the early stage to the Internet of Things. *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*, *2017-Janua*, 6159–6164. https://doi.org/10.1109/IECON.2017.8217070

Gupta, J., Pathak, S., & Kumar, G. (2022). Deep Learning (CNN) and Transfer Learning: A Review. *Journal of Physics: Conference Series*, *2273*(1), 012029. https://doi.org/10.1088/1742-6596/2273/1/012029

Halme, R.-J., Lanz, M., Kämäräinen, J., Pieters, R., Latokartano, J., & Hietanen, A. (2018). Review of vision-based safety systems for human-robot collaboration. *Procedia CIRP*, *72*, 111–116. https://doi.org/10.1016/j.procir.2018.03.043

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition.

2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), *2016-Decem*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Heo, Y. J., Kim, D., Lee, W., Kim, H., Park, J., & Chung, W. K. (2019). Collision Detection for Industrial Collaborative Robots: A Deep Learning Approach. *IEEE Robotics and Automation Letters*, *4*(2), 740–746. https://doi.org/10.1109/LRA.2019.2893400

Hoeser, T., & Kuenzer, C. (2020). Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends. *Remote Sensing*, *12*(10). https://doi.org/10.3390/rs12101667

Hong, J., Cheng, H., Zhang, Y. D., & Liu, J. (2019). Detecting cerebral microbleeds with transfer learning. *Machine Vision and Applications*, *30*(7–8), 1123–1133. https://doi.org/10.1007/s00138-019-01029-5

Hoskins, G. O., Padayachee, J., & Bright, G. (2019). Human-Robot Interaction: The Safety Challenge (An inegrated frame work for human safety). *2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA)*, 74–79. https://doi.org/10.1109/RoboMech.2019.8704744

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *ArXiv*. http://arxiv.org/abs/1704.04861

Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2017). Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(8), 2011–2023. https://doi.org/10.1109/TPAMI.2019.2913372

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, *2017-Janua*, 2261–2269. https://doi.org/10.1109/CVPR.2017.243

Huang, J., Sun, C., Murphy, K., & Guadarrama, S. (2017). Speed / accuracy trade-offs for modern convolutional object detectors : Supplementary Materials. *Cvpr*, 7310–7319.

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. *ArXiv*, 1–13. http://arxiv.org/abs/1602.07360

Iman, M., Rasheed, K., & Arabnia, H. R. (2022). *A Review of Deep Transfer Learning and Recent Advancements*. http://arxiv.org/abs/2201.09679

Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *32nd International Conference on Machine Learning,* *ICML 2015*, *1*, 448–456. http://arxiv.org/abs/1502.03167

Iorga, C., & Neagoe, V.-E. (2019). A Deep CNN Approach with Transfer Learning for Image Recognition. *2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 1–6. https://doi.org/10.1109/ECAI46879.2019.9042173

Jiang, B. C., & Gainer, C. A. (1987). A cause-and-effect analysis of robot accidents. *Journal of Occupational Accidents*, *9*, 27–45. https://doi.org/10.1016/0376-6349(87)90023-X

Khalid, A., Kirisci, P., Ghrairi, Z., Thoben, K., & Pannek, J. (2017). Implementing Safety and Security Concepts for Human-Robot Collaboration in the context of Industry 4 . 0 Towards Implementing Safety and Security Concepts for Human-Robot- Collaboration in the context of Industry 4 . 0. *International MATADOR Conference on Advanced Manufacturing*, *July*, 1–7.

Kim, S., Noh, Y.-K., & Park, F. C. (2020). Efficient neural network compression via transfer learning for machine vision inspection. *Neurocomputing*, *413*, 294–304. https://doi.org/10.1016/j.neucom.2020.06.107

Kim, W. (2020). *The risk analysis for the introduction of collaborative robots in the Republic of Korea* [Iowa State University]. https://doi.org/10.31274/etd-20200624-157

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Kumar, S., Arora, S., & Sahin, F. (2019). Speed and Separation Monitoring using On-Robot Time-of-Flight Laser-ranging Sensor Arrays. *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*, 1684–1691. https://doi.org/10.1109/COASE.2019.8843326

Kumar, S., Savur, C., & Sahin, F. (2018). Dynamic Awareness of an Industrial Robotic Arm Using Time-of-Flight Laser-Ranging Sensors. *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2850–2857. https://doi.org/10.1109/SMC.2018.00485

Lasota, P. A., Fong, T., & Shah, J. A. (2017). A Survey of Methods for Safe Human-Robot Interaction. *Foundations and Trends in Robotics*, *5*, 261–349. https://doi.org/10.1561/2300000052

Law, H., & Deng, J. (2019). CornerNet: Detecting Objects as Paired Keypoints. *International Journal of Computer Vision*, *128*(3), 642–656. https://doi.org/10.1007/s11263-019-01204-1

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L.

(1990). *Handwritten digit recognition with a back-propagation network*. 396–404.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2323. https://doi.org/10.1109/5.726791

Li, X., Wang, W., Hu, X., & Yang, J. (2019). Selective kernel networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*, 510–519. https://doi.org/10.1109/CVPR.2019.00060

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2016). Feature Pyramid Networks for Object Detection. *Proceedings - 2019 IEEE Intl Conf on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking, ISPA/BDCloud/SustainCom/SocialCom 2019*, 1500–1504. https://doi.org/10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00217

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2018). Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(2), 318–327. https://doi.org/10.1109/TPAMI.2018.2858826

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016a). SSD: Single Shot MultiBox Detector. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Eccv* (Vol. 9905, pp. 21–37). Springer International Publishing. https://doi.org/10.1007/978-3-319-46448-0_2

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016b). SSD: Single shot multibox detector. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9905 LNCS*, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

Loshchilov, I., & Hutter, F. (2016). *SGDR: Stochastic Gradient Descent with Warm Restarts*. https://doi.org/10.48550/arxiv.1608.03983

Magrini, E., Ferraguti, F., Ronga, A. J., Pini, F., De Luca, A., & Leali, F. (2020). Human-robot coexistence and interaction in open industrial cells. *Robotics and Computer-Integrated Manufacturing*, *61*(June 2018), 101846. https://doi.org/10.1016/j.rcim.2019.101846

Mahardi, Wang, I.-H., Lee, K.-C., & Chang, S.-L. (2020). Images Classification of Dogs and Cats using Fine-Tuned VGG Models. *2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE)*, 230–233. https://doi.org/10.1109/ECICE50847.2020.9301918

Malm, T., Salmi, T., Marstio, I., & Montonen, J. (2019). Dynamic safety system for collaboration of operators and industrial robots. *Open Engineering*, *9*, 61–71. https://doi.org/10.1515/eng-2019-0011

Min, F., Wang, G., & Liu, N. (2019). Collision Detection and Identification on Robot Manipulators Based on Vibration Analysis. *Sensors*, *19*(5), 1080. https://doi.org/10.3390/s19051080

Mohammed, A., Schmidt, B., & Wang, L. (2017a). Active collision avoidance for human–robot collaboration driven by vision sensors. *International Journal of Computer Integrated Manufacturing*, *30*(9), 970–980. https://doi.org/10.1080/0951192X.2016.1268269

Mohammed, A., Schmidt, B., & Wang, L. (2017b). Active collision avoidance for human–robot collaboration driven by vision sensors. *International Journal of Computer Integrated Manufacturing*, *30*(9), 970–980. https://doi.org/10.1080/0951192X.2016.1268269

Ore, F. (2015). *Human - industrial robot collaboration: Simulation, Visualisation and Optimisation of Future Assembly Workstations* [Mälardalen University Sweden]. https://www.diva-portal.org/smash/get/diva2:816519/FULLTEXT02.pdf

Pasinetti, S., Nuzzi, C., Lancini, M., Sansoni, G., Docchio, F., & Fornaser, A. (2018). Development and Characterization of a Safety System for Robotic Cells Based on Multiple Time of Flight (TOF) Cameras and Point Cloud Analysis. *2018 Workshop on Metrology for Industry 4.0 and IoT*, 1–6. https://doi.org/10.1109/METROI4.2018.8439037

Pathak, Y., Shukla, P. K., Tiwari, A., Stalin, S., Singh, S., & Shukla, P. K. (2020). Deep Transfer Learning Based Classification Model for COVID-19 Disease. *IRBM*. https://doi.org/10.1016/j.irbm.2020.05.003

Realyvásquez-Vargas, A., Cecilia Arredondo-Soto, K., Luis García-Alcaraz, J., Yail Márquez-Lobato, B., & Cruz-García, J. (2019). Introduction and configuration of a collaborative robot in an assembly task as a means to decrease occupational risks and increase efficiency in a manufacturing company. *Robotics and Computer-Integrated Manufacturing*, *57*(December 2018), 315–328. https://doi.org/10.1016/j.rcim.2018.12.015

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Robla-Gomez, S., Becerra, V. M., Llata, J. R., Gonzalez-Sarabia, E., Torre-Ferrero, C., & Perez-Oria, J. (2017). Working Together: A Review on Safe Human-Robot Collaboration in Industrial Environments. *IEEE Access*, *5*, 26754–26773. https://doi.org/10.1109/ACCESS.2017.2773127

Shackleford, W., Cheok, G., Hong, T., Saidi, K., & Shneier, M. (2016). Performance Evaluation of Human Detection Systems for Robot Safety. *Journal of Intelligent & Robotic Systems*, *83*, 85–103. https://doi.org/10.1007/s10846-016-0334-3

Shuaib, M., Jaafar, J., Faizuddin, M., & Musa, S. (2022). Review on recent trends and challenges in deep learning and artificial intelligence. *AIP Conference Proceedings*, *2617*(1), 020010. https://doi.org/10.1063/5.0119678

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–14. http://arxiv.org/abs/1409.1556

Sultana, F., Sufian, A., & Dutta, P. (2020). A Review of Object Detection Models Based on Convolutional Neural Network. In *Advances in Intelligent Systems and Computing* (Vol. 1157, pp. 1–16). https://doi.org/10.1007/978-981-15-4288-6_1

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2016). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, 4278–4284. http://arxiv.org/abs/1602.07261

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2014). Going Deeper with Convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, *07-12-June*, 1–9. https://doi.org/10.1109/CVPR.2015.7298594

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). Rethinking the Inception Architecture for Computer Vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, *2016-Decem*, 2818–2826. https://doi.org/10.1109/CVPR.2016.308

Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? *IEEE Transactions on Medical Imaging*, *35*(5), 1299–1312. https://doi.org/10.1109/TMI.2016.2535302

Talukdar, J., Gupta, S., Rajpura, P. S., & Hegde, R. S. (2018). Transfer Learning for Object Detection using State-of-the-Art Deep Neural Networks. *2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)*, 78–83. https://doi.org/10.1109/SPIN.2018.8474198

Tan, J. T. C., Duan, F., Zhang, Y., Watanabe, K., Kato, R., & Arai, T. (2009). Human-robot collaboration in cellular manufacturing: Design and development. *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, *24*(5–6), 29–34. https://doi.org/10.1109/IROS.2009.5354155

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *36th International Conference on Machine Learning, ICML 2019*, *2019-June*, 10691–10700. http://arxiv.org/abs/1905.11946

Tan, M., Pang, R., & Le, Q. V. (2020). EfficientDet: Scalable and Efficient Object Detection. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10778–10787. https://doi.org/10.1109/CVPR42600.2020.01079

Tzu Ta, L. (2015). *LabelImg. Git code*. https://github.com/tzutalin/labelImg

Villani, V., Pini, F., Leali, F., & Secchi, C. (2018). Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. *Mechatronics*, *55*, 248–266. https://doi.org/10.1016/j.mechatronics.2018.02.009

Wu, X., Sahoo, D., & Hoi, S. C. H. (2020). Recent advances in deep learning for object detection. *Neurocomputing*, *396*(d), 39–64. https://doi.org/10.1016/j.neucom.2020.01.085

Xiao, Y., Tian, Z., Yu, J., Zhang, Y., Liu, S., Du, S., & Lan, X. (2020). A review of object detection based on deep learning. *Multimedia Tools and Applications*, *79*(33–34), 23729–23791. https://doi.org/10.1007/s11042-020-08976-6

Xie, S., Girshick, R., Dollar, P., Tu, Z., & He, K. (2017). Aggregated Residual Transformations for Deep Neural Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, *2017-Janua*, 5987–5995. https://doi.org/10.1109/CVPR.2017.634

Zagoruyko, S., & Komodakis, N. (2016). Wide Residual Networks. *Procedings of the British Machine Vision Conference 2016*, *2016-Septe*, 87.1-87.12. https://doi.org/10.5244/C.30.87

Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Lin, H., Zhang, Z., Sun, Y., He, T., Mueller, J., Manmatha, R., Li, M., & Smola, A. (2020). ResNeSt: Split-Attention Networks. *ArXiv*. http://arxiv.org/abs/2004.08955

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

Zou, Z., Shi, Z., Guo, Y., & Ye, J. (2019). Object Detection in 20 Years: A Survey. *ArXiv*, 1–39. http://arxiv.org/abs/1905.05055