



REST Journal on Emerging Trends in Modelling and Manufacturing

Vol: 9(4), December 2023

REST Publisher; ISSN: 2455-4537 (Online)

Website: <https://restpublisher.com/journals/jemm/>

DOI: <https://doi.org/10.46632/jemm/9/4/6>



Improving Robot Image Perception Design Using Artificial Intelligence

*R. Suresh, S. Ponlatha, K. Giri, R. Chandralekha

Mahendra Engineering College, Namakkal, Tamil Nadu, India.

*Corresponding Author Email: sureshr@mahendra.info

Abstract: In this current decade, there are numerous growth has been identified in science and technology. Traditional robots functions and its performance are frequently failed in fulfilling the people's production and life needs, whereas the robot development has evolved into increasingly intelligent as well as networked. For these, the key technology for mobile robot location as well as its navigation, environment perception technology remains a priority and interesting area of study in the robotic fields. Methods for image processing have had an important effect on the vast majority of robotics as well as automation in industry. Real-time approaches are commonly utilized in complicated automation tasks, assist in making decision or effective guidance of machinery and robots, whereas post-processing is usually applied for retrospective technology and its process evaluation. In addition, a human pose detection system is built using Deep Learning (DL) as well as robotics whereas each part is recommended. Ultimately, the overall system efficiency is verified. The outcomes indicate that once the precision approaches High-Resolution Net (HRNet-32). The model's inferences performance is boosted by around 2.4 times, while its parameters are made compressed over approximately 67%. The results evaluation with existing Convolutional Neural Network (CNN) has illustrate the method's effectiveness and the posture detection method in robot's human has ability in resolving human posture detection effectively in fitness activities.

Keywords: Deep learning (DL), Robotics, image processing, food industry, High-Resolution Net (HRNet).

1. INTRODUCTION

A present, robots are steadily expanded their presence in retail malls, people's lives, fire safety, and traffic fields where they now serve an essential role. Robots have offered people convenience in a variety of manners, including inspection robots and sweeping robots. However, there are certain technological limitations in the traditional method. The main limitation identified is robots contain only simple functions which are more expensive as well as poor accuracy in positioning. They are able to perform basic assignments and exhibit clear inadequacies when it comes to complicated tasks assisting clients in pouring water, retrieving items, and various parts of complicated processes. As a result, humans have proposed increased standards for robots with respect to practicality with the goal of obtaining additional ecological data as well as improving their comprehension and environment analysis. Robots currently utilize the optical system, Lidar, as well as infrared induction for collecting data about the environment. Since Lidar and infrared induction have high signal transmission speeds, they require signal return, which limits their utility. However, the visual sensor doesn't need signal return as well as offers a diverse variety of applications which is frequently utilized [1]. Even the panorama camera is going to be employed as the robot's sensor. Despite the development of different cameras, the robot's visual sensing technology has progressively developed, yet there remain certain shortcomings. For instance, the camera's depth is utilized most frequently today is capable of gathering data on the surrounding environment, but the expense is considerable with environmental data can be obtained seldom as

well as the computation amount is significant that leading to comparatively poor real-time performance. Robots' ability to interpret and utilize data is rather limited [2]. Individuals get data regarding their surroundings mostly through their visual system. As a result, humans as well as higher-beings rely on their vision as their main visual source. Mobile robots' vision systems for sensing their surroundings primarily consist of a monocular camera, a binocular camera, a multiocular camera, as well as a folding and reflecting panorama camera [3]. Compared with different vision systems, a refraction panoramic camera has infrequently employed in mobile robots. As a result, the next section will employ a panoramic camera as a vision system for mobile robots to complete their environment comprehension. The primary ways for collecting panoramic images involve image mosaic as well as fusion systems, fisheye lenses, and refraction panoramic imaging techniques. In the 1970s, international countries began researching the omni-directional system of vision. The "Wanderer," which depends upon the hyperboloid fold-reflection panoramic vision framework, was developed in the 1990s at Carnegie Mellon University's Robotics Institute in the United States. The robot's hyperboloid fold-reflection panoramic technology had been installed on its top [4]. Lehigh University's laboratory in US used an omni-directional retinal system for investigating the recognition and location of objects in motion [5]. The system of vision may collect data in the $360^{\circ} \times 210^{\circ}$ range due to its ability to capture an extensive variety of field angles. Furthermore, experimental simulations have reveals that an omni-directional vision system may achieve target identification as well as monitoring accuracy rates of exceeding 90%, demonstrating the system's viability. HRNet begins with high-resolution image data as input as well as builds networks of sub-net from high resolution to low resolution images layer by layer over network have producing higher-level parallel phases. Parallel integrative connectivity to multi-resolution sub-networks established from high-resolution images are established gradually as the network complexity grows [6]. This technique allows for the merging of visual data across several scales. HRNet has a distinctive parallel multi-resolution sub-network topology, which has a positive impact on estimating the human poses [7]. Despite HRNet can substantially enhance the accuracy of predictions and obtain optimal outcomes using numerous public datasets, the parameters of its model are large and need a lot of computational power [8, 9]. As a result, to prevent losing excessive computational accuracy, HRNet must be upgraded to lessen its dependency on hardware assets. Data extraction compresses the human posture detection system and minimizes its parameter count. In addition, a novel data distillation strategy is used during training for models. This method outperforms previous data extraction methods with respect to method training effectiveness. The primary research involvement to the understanding subjected is for detecting the image processing method for fitness movement using soft robotic technology, as well as investigate the overall efficiency of the entire system using HRNet as the skeleton model, by improving the detecting of human posture method as well as data extraction. In this study, a robot perception system utilizing HRNet for development that addresses issues like the typical robot sensor's restricted perspective range as well as enhances the robots autonomous. The technology may additionally increase robots' effectiveness in providing services to humans and assist individuals lessen their reliance on robots, yet it also allows for autonomous operations in tough environments. They are going to address various training techniques that either enhance classic supervised learning methods to dynamic perception or use reinforcement learning techniques for delivering dynamic perception data.

2. LITERATURE REVIEW

As discussed in Section 1, training active perception models requires a shift from traditional static perception methods, presenting a distinctive challenge. This distinction arises from the necessity for active perception models to not only grasp the static aspects of object recognition but also to encompass the dynamics inherent in the perception process, allowing them to generate control feedback. Tosidis et al., have discussed initial category of methods utilize realistic simulation environments for simulating the effect of various movements and allow the agent to learn how perception accuracy varies when performing different actions [10]. This approach provides great flexibility since any action can be simulated and the effect of the movement of a robot can be easily obtained. However, such approaches are computationally demanding, since they rely on realistic simulation environments and graphics engines, such as Webots slowing down the training process. Zheng et al., have introduced training active perception models also departs from the typical supervised learning approach that is followed in many perception applications, such as face recognition, object detection and pose estimation [11]. Active perception models should not only analyze and understand their input but also provide some kind of control feedback that can be then subsequently used for improving perception accuracy. Indeed, Passalis and Tefas have discussed early DL-based approaches extended embedding-based active perception methods into active ones by including an additional head that predicts the subsequent best movement that a robot should perform in improving confidence of face recognition confidence [12]. This approach assumes that the robot moves on a predefined trajectory around the target in order to be compatible with the multi-view dataset employed. Then, the model is trained to both maximize face recognition confidence, following a

contrastive learning objective, as well as to regress the direction of movement leading to the best face recognition accuracy. Ginargyros et al., have proposed for object detection which consist of rich dataset for potential movements was built using a simulation environment. This approach enabled the models to learn the object detection confidence manifold for different types of objects, e.g., cars and humans, while taking into account possible occlusions, allowing them to perform control tailored to the unique characteristics of different cases. To this end, a separate navigation proposal network was trained according to the confidence manifold of each object, enabling the model to learn to propose trajectories that will maximize object detection confidence. At the same time, this paper also revealed limitations that are often intrinsic to the current state-of-the-art object detection models, since it provided a structured approach for revealing the confidence manifold of object detectors [13]. T.C. Chen and S.Y. Yu have illustrated the vegetable and fresh fruits appearance is a single aspect that influences customer impressions. In reply to customer demands, robot vision is a novel instrument for monitoring and enhancing the appearance of products. It allows you to determine product features like as ripeness, and lack of flexibility are the important factors for a consumer's initial perception of a product. This technology can be used for a variety of purposes include packaging of goods form assessment, flaw detection, and quality inspection. Several investigators examined the possibility of utilizing fluorescence imaging with hyperspectral linear scans for identifying pollution on the surface of apples by developing an intuitive multidimensional algorithm [14]. Licciardo et al. developed a bespoke hardware design method for a Fully Convolutional Neural Network (CNN) to create an embedded human gesture recognition system (HGR). The system has good recognition accuracy for resting as well as sitting postures. The development of optimized hardware uses a minimum pressure sensors amount that enables for staying near to the data source through edge computing as well as promotes the embedded HGR system design [6]. This study incorporates multi-scale features along with interest distribution, as well as feature information, to investigate network structure design with respect to the remaining network structure for accomplishing better human pose detection accuracy in three dimensions. The design system's primary component is the human pose detecting module. Once the target detection modules acquires the target location data of an individual, the human posture detection module needs to recognize the important spots in the image[16]. The HRNet model is used as the fundamental skeleton framework for this module for identifying human poses. The feature pyramid module is introduced following the implementation of HRNet NN, as well as the feature map of single-resolution has been replaced with the multi-resolution attributes of Feature Pyramid Networks (FPN) for improved detection accuracy. The experimental results on the MPII Human Pose dataset show that the technique proposed in this study greatly improves the human pose estimation accuracy. Once the FPN get included towards the HRNet method, the performance of detection method gets improved with 0.03 detection to each features parts as the results finding. Once accumulating 3 x 3 convolution module before upsampling, the small target detection purpose get modified and experimental results have improved high while compared with the feature pyramid usage with the complete implementation as 0.17. There are certain parts gets improved with 0.5 that determines suggested method produced high accuracy in detecting human pose estimation as well as detection accuracy result is better with multiple experiments.

3. RESEARCH METHODOLOGY

This paper investigates the impact of various lighting conditions on image quality. The change in the quantity of near-infrared radiation absorbed from the tissue was determined. In the acquisition framework, the light resource has the greatest impact on the image impact. The good-quality image is required to appropriate venous perception as well as localization. As a result, in this study, the experimental design of light source was performed to investigate the imaging impact of the infrared near cameras in various light sources, as well as the imaging influence when modifying the light intensity of the light source intensity using different input energy for determining the best imaging circumstances. An optical flow approach is frequently employed in a variety of applications that involve or benefit from image processing. The most exciting applications are in surgical robots, wherein the digital video stream from the endoscopy remains continually available. In the food and beverage business, the meat processing plants with automation has the ability for earnings from optical flow-based systems. The most major issues in meat sector automation, like in the overwhelming majority of the food as well as agriculture industries, are dealing with high variability caused by natural ecosystems, as well as detecting and managing disease and deformed animals. Every equipment and robot framework that designed in automation meat industry has ability to collaborate with the extremely wide range of mechanical properties such as elasticity, size, surface properties, weight, etc. of the "targeted objects". A possible approach is for making the equipment by design resistant of large variance. This indicates the fact that the device's mechanical construct may complete its responsibilities with no the assistance of sensors or "intelligence" despite the targeted object's unknown and/or unidentified attributes as well as coordinates which is as long as it remain within normal boundaries. Figure 1 depicts a hypothetical meat industry gripper which is capable of grabbing target objects

among 2 and 6 cm with a displacement of 4 cm due to the circumferential action of the gripping fingers. The most modern option for increasing robustness is to use sensors as well as intelligence embedded within it. In the present scenario, the most significant aspect of the devices is not the hardware design, rather the intelligent parts as well as software that operate it. In terms of decision-making location, the food sector automation offers two fundamental possibilities. The RoBUTCHER project, outlined in Section I-A, combines each of these methodologies. The robot system has been functionally centralized, with one machine collecting the majority of the sensor data, running the Virtual Reality as well as AI modules, and controlling the robots. In contrast, efficient end-of-arm tools with gripper and knife function as embedded systems with individual specialized communication lines and/or incorporated compute modules. Among of the robotic grippers, the gripper used in the RoBUTCHER project comprises a smart gripper with an embedded microcontroller, sensors, and a Raspberry Pi Zero. Thus, the mechanical design is shown in Figure 1. However, the mechanism provides encompassing gripping motion, fail-safe grasping, and a significant clamping force by the design. Therefore, its integrated sensors as well as built-in intelligence are the features which distinguish the gripper as well as enable it to perform the complicated tasks necessary for completely automated meat processing.

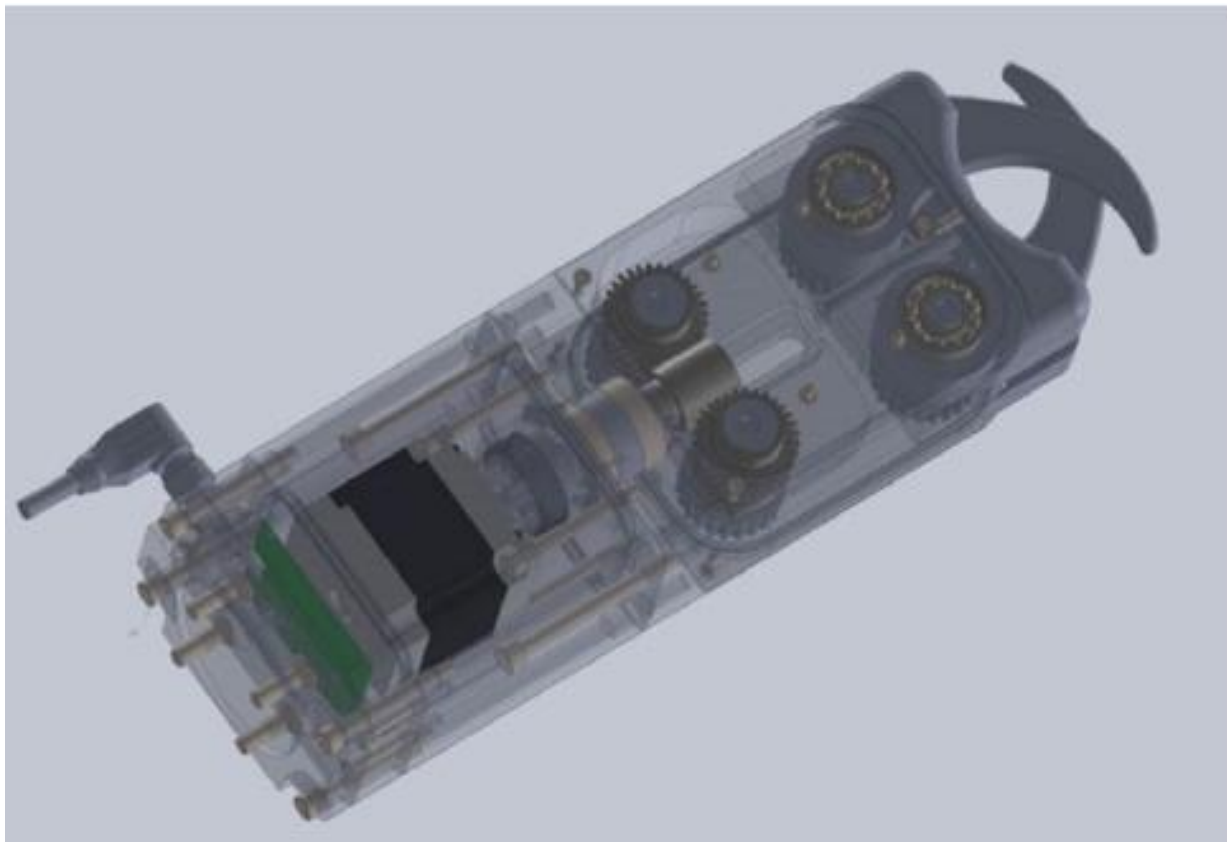


Figure 1. smart gripper used in the RoBUTCHER project

Figure 1 shows the structure of the HRNet method whereas the method is broken into four stages. Each stage is further separated into two components namely the multi-parallel and multi-resolution convolution module as well as the multi-scale feature fusion module. Each level introduces a branch to the preceding one. The recently inserted branches are the result of a convolutional merging of every single feature maps from the previous phase. The measure under consideration is just half of the preceding segment. The channel amount gets double that of the last branch. The approach offers significant application value for optimizing and upgrading the embedded HGR system due to its extensive improvement and recognition processes.

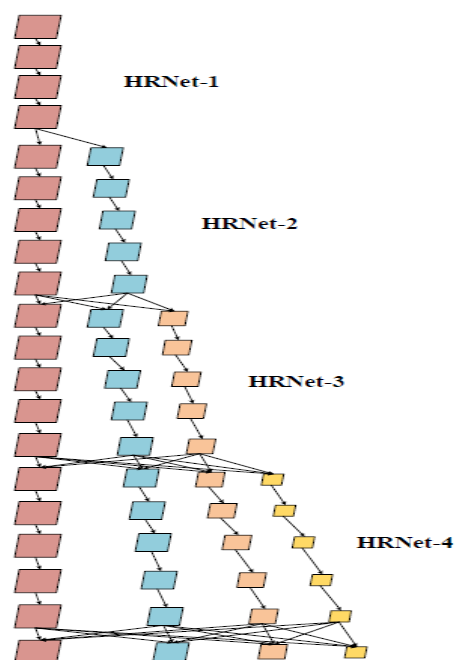


Figure 2. Architecture of HRNet

The parameters of the HRNet architecture are rather large. To minimize the model size, each of the final three phases' widths have been decreased to 18. In addition, this work employs HRNetV2 to enhance the model. Low-resolution feature maps are upsampled using feature maps of all resolutions and then merged with high-resolution feature maps. The revised feature graphing output may significantly enhance the model's effect while reducing accuracy loss resulting from method parameters. The parameters of the method are compressed on the basis of equivalent method accuracy for increasing the system's inference time. To begin, information distillation uses training data to generate a framework with numerous parameters as well as great accuracy. Subsequently on the training data, labeling are produced to allow the tiny model to acquire information from the large, effectively passing on the larger model's data to the smaller.

4. RESULT AND DISCUSSION

Despite AI-based image processing has grown in prominence in the current decade whereas hard computing approaches remain feasible as well as for certain cases it is highly efficient. Optical flow is an excellent instance for these techniques with numerous different algorithms available in among the most popular image processing packages. Optical flow methods are typically used for the food business in real-time because it's an excellent device for detecting slipping, which has become frequently encountered with automated gripping. HRNet is used as the skeleton model, whereas the HRNet-48 model is the instructional concept. The HRNet-18 model is a model for learning, with the effect of HRNet-32 serving as an example for the research. The widths of the high-resolution subnetworks for the last three stages are represented by 18, 32, and 48, respectively. The key experimental component is the HRNet-18 model that implements the method using the Pytorch framework. Once images have been gathered certain evaluation measure indices the frequently required to assess the image quality. This experimental research discusses the evaluation of two major QoS parameters are PSNR and Q-Factor. According to this experiment, there are three different natural light source considered as sensor lights are 750nm, 850nm and 950nm. However, natural light comprises many different kinds of distinct light wavelengths and can be distinguished by homogeneous light, parallel light, and the same lighting conditions at all angles. Therefore, to improve image, imaging outcomes with multiple-point light sources must be combined to form multi-point light sources.

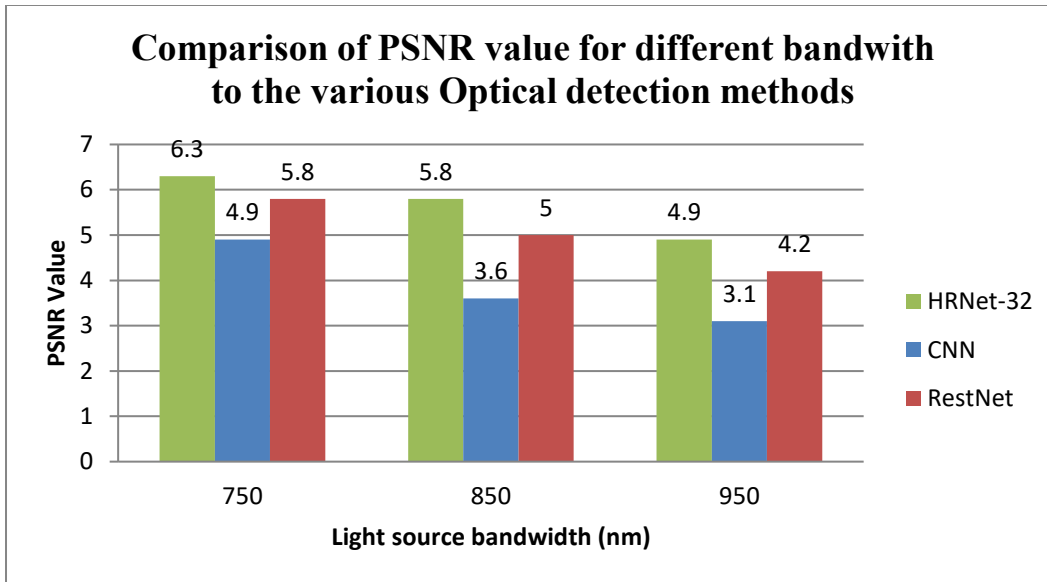


Figure 3. Comparison of PSNR values for various optical detection methods

Figure 3 illustrates the PSNR values for three different models of DL method for optical detection from three different light sources namely 750nm, 850nm and 950nm. In this research, the result assist in identifying the PSNR value get decreased along with increase of light source bandwidth but the accuracy in detection of optical light source is better in noise detection in HRNet-32 which consist of PSNR value as 6.3, 5.8 and 4.9 for 750nm, 850nm and 950nm respectively. HRNet-32 has better intense of light source while compared to RestNet and CNN are 5.8, 5.0 and 4.2 and 4.9, 3.6 and 3.1 respectively.

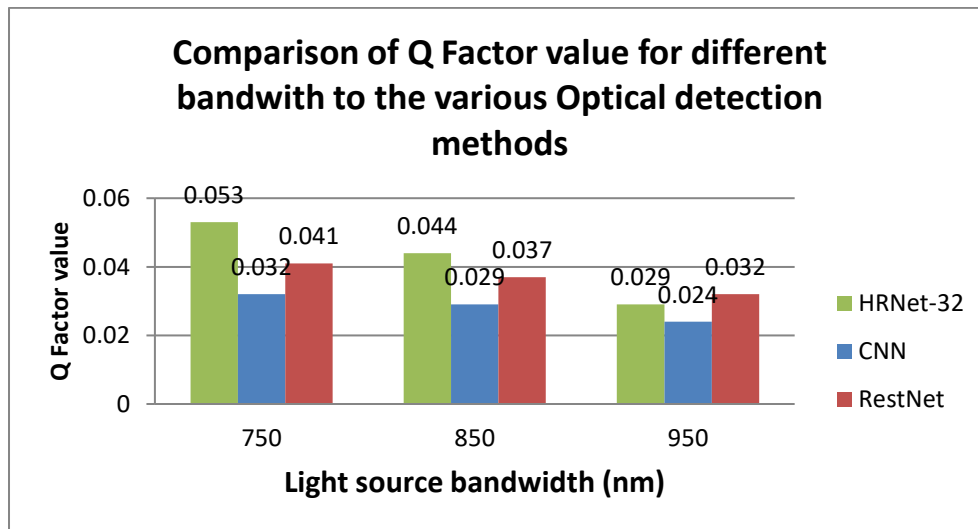


Figure 4. Comparison of Q Factor values for various optical detection methods

Figure 4 illustrates the Q Factor value for three different models of DL method for optical detection from three different light sources namely 750nm, 850nm and 950nm. In this research, the result assist in identifying the Q Factor value get decreased along with increase of light source bandwidth but the accuracy in detection of optical light source is better in noise detection in HRNet-32 which consist of Q Factor value as 0.053, 0.044 and 0.029 for 750nm, 850nm and 950nm respectively. HRNet-32 has better intense of light source while compared to RestNet and CNN are 0.041, 0.037 and 0.032 as well as 0.032, 0.029 and 0.024 respectively.

5. CONCLUSION

As a fundamental study area, image processing is critical to the growth of AI. Stereo images not only contribute to the interpretation of 2-D images, but they additionally perform a vital role in the determination of 3-D visual dimensions. As a result, identifying the offset pixels which correspond to the left and right angles in stereoscopic images is critical for accomplishing super-resolution as well as parallax estimates. An upgraded technique that depends upon the HRNet-32 model is proposed for enhancing the accuracy of human posture estimation. The layer of the feature pyramid is introduced after the HRNet NN, as well as the single-resolution feature map has been replaced with the multi-resolution feature of FPN for enhancing the detection accuracy. Once the accumulation of 3x3 convolution module before upsampling is for detecting small target is motivated. The experimental outcome determines that HRNet-32 has better detection of human posture estimation precisely than other traditional RestNet and CNN methods for various light source bandwidth. The evaluation of detecting human posture estimation is done through PSNR value and Q Factor that has high intense of light source detection in 750nm while compared with 850nm and 950nm. Thus, the detection of small target human posture estimation is done using 750nm light source is high using HRNet-32 while compared with 850nm and 950nm with HRNet-32.

REFERENCES

- [1]. Rosenberger P, Cosgun A, Newbury R, et al. Object-independent human-to-robot handovers using real time robotic vision. *IEEE Robotics and Automation Letters*, 2020, 6(1): 17-23.
- [2]. Coşar S, Fernandez-Carmona M, Agrigoroaie R, et al. ENRICHME: Perception and Interaction of an Assistive Robot for the Elderly at Home. *International Journal of Social Robotics*, 2020, 12(3): 779-805.
- [3]. Ibarz J, Tan J, Finn C, et al. How to train your robot with deep reinforcement learning: lessons we have learned. *The International Journal of Robotics Research*, 2021, 40(4-5): 698-721.
- [4]. Martín-Martín R, Brock O. Coupled recursive estimation for the online interactive perception of articulated objects. *The International Journal of Robotics Research*, 2022, 41(8): 741-777.
- [5]. Chen A I, Balter M L, Maguire T J, et al. Deep learning robotic guidance for autonomous vascular access. *Nature Machine Intelligence*, 2020, 2(2): 104-115.
- [6]. S. Seong and J. Choi, "Semantic segmentation of urban buildings using a high-resolution network (HRNet) with channel and spatial attention gates." *Remote Sensing*, vol.13, no.16, pp. 3087,2021.
- [7]. K. Sengupta and P. R. Srivastava, "HRNET: AI-on-Edge for Mask Detection and Social Distancing Calculation." *SN computer science*, vol.3, no.2, pp. 1-15,2022.
- [8]. M. Hidaka, D. Matsuoka, D. Sugiyama, et al. "Pixel-level image classification for detecting beach litter using a deep learning approach." *Marine Pollution Bulletin*, vol.175, Article ID 113371, 2022.
- [9]. R. Tao, Y. Zhang, L. Wang, et al. "U-High resolution network (U-HRNet): cloud detection with high-resolution representations for geostationary satellite imagery." *International Journal of Remote Sensing*, vol.42,no.9, pp.3511-3533,2021.
- [10]. Tosidis, P., Passalis, N., and Tefas, A. (2022). Active vision control policies for face recognition using deep reinforcement learning. In *Proceedings of the 30th European Signal Processing Conference*, pages 1087– 1091.
- [11]. Zheng, C., Wu, W., Chen, C., Yang, T., Zhu, S., Shen, J., Kehtarnavaz, N., and Shah, M. (2023). Deep learningbased human pose estimation: A survey. *ACM Computing Surveys*, 56(1):1–37.
- [12]. Passalis, N. and Tefas, A. (2020). Leveraging active perception for improving embedding-based deep face recognition. In *Proceedings of the IEEE International Workshop on Multimedia Signal Processing*, pages 1– 6.
- [13]. Ginargyros, S., Passalis, N., and Tefas, A. (2023). Deep active perception for object detection using navigation proposals. In *Proceedings of the IEEE Symposium Series on Computational Intelligence (accepted)*, page 1.
- [14]. Chen, T.C., and Yu. S.Y. (2022) "The review of food safety inspection system based on artificial intelligence, image processing, and robotic." *Food Science and Technology* 42: e35421.
- [15]. G. D. Licciardo, A. Russo, A. Naddeo, et al. "A Resource Constrained Neural Network for the Design of Embedded Human Posture Recognition Systems." *Applied Sciences*, vol.11, no.11, pp. 4752, 2021
- [16]. J. Stenum, C. Rossi and R. T. Roemmich, "New generation of smart highway: framework and insights." *Journal of Advanced Transportation*, 2021.
- [17]. Dong, Fangjie & Yang, Miaowei & Yang, Pengshen. (2023). Human Pose Estimation based on HRNet and Feature Pyramids. *Highlights in Science, Engineering and Technology*. 39. 1239-1244. 10.54097/hset.v39i.6734.