**Youtube Link for Task 4 Q8**

<https://www.youtube.com/watch?v=tQt33GdmoQM&ab_channel=Lee>

**Task 2**

**AA2.8**

1. Image warping and perspective correction are aspects in computer vision which can be considered crucial since they help with enabling image rectification, distortion correction, and object alignment for applications such as document scanning, 3D reconstruction, and augmented reality. There are various algorithms which help to transform images, correcting skewness, distortion, or perspective errors.  
     
   Homography transformation is an algorithm used to rectify perspective distortion. This is done by mapping points between two planes using a 3x3 transformation matrix. The use case for this algorithm ranges from image stitching [1, 2, 3], multi-scale gigapixel photography [4, 5] multispectral image fusion [6, 7], and planar object tracking [8,9] among other use cases.  
     
   Neural Networks are also being used to automate perspective correction. [10] used neural networks for data augmentation to be used in detecting COVID-19 via X-Ray images. The key features of Neural Networks for perspective corrections are that they are trained on large datasets to learn different transformation patterns and that they can adapt to complex distortions without requiring manual control points.  
     
   [11] proposed a model for dynamically replacing clothes on a model using image transformation on a specific garment onto the model depending on their pose. This model uses Affine transformation to move the input in-shop clothes closer to the target body pose and Thin-Plate Spline Transformation to generate a warped version of the input garment by taking advantage of the previously used transformation result.

**Task 3**

**KU 3.1**

Pose estimation and human activity recognition (HAR) are increasingly important factors in computer vision and machine learning, with their application ranging from Augmented Reality to health monitoring.

[16] stated that pose estimation typically involves the recognition of 18 key body points and joints’ location. Two prominent algorithms in this field are OpenPose ([12]) and High-Resolution Network (HRNet) ([13]). OpenPose is a real-time multi-person pose estimation method that adopts a bottom-up approach by initially detecting all body parts in an image and grouping them per individual. OpenPose supports a maximum detection of 135 keypoints, including hands, face, and feet. On the other hand, HRNet Utilizes a top-down approach, starting with person detection followed by keypoint estimation for each detected individual. HRNet maintains high-resolution representations throughout the process, contributing to its superior performance.

[14] conducted a comparison analysis on HRNet and OpenPose and found that HRNet demonstrated higher accuracy compared to OpenPose. On their dataset, HRNet's best model achieved a mean Average Precision (mAP) of 77.0, surpassing OpenPose's 61.8 mAP with a 24.5% improvement. However, in terms of computational speed on GPU systems, OpenPose outperforms HRNet, achieving approximately 3.16 to 3.56 frames per second (FPS) compared to HRNet's 0.154 to 0.162 FPS. On CPU systems, both models exhibit slower performance, with OpenPose achieving 0.025 to 0.0325 FPS and HRNet 0.040 to 0.060 FPS.

Moving on to HAR which focuses on identifying physical activities from sensor data, commonly using accelerometers and gyroscopes. [15] detail different machine learning techniques which have been applied to this task, including Support Vector Machines (SVM) which is a supervised learning model that finds the optimal hyperplane separating different activity classes, Convolutional Neural Networks (CNNs) that are deep learning models that automatically learn spatial hierarchies of features from input data, effective in capturing local dependencies and Long Short-Term Memory Networks (LSTMs) which are a type of recurrent neural network capable of learning long-term dependencies, suitable for sequential data like time-series sensor readings.

In a study conducted by [15] which compare these algorithms on accelerometer and gyroscope data:

* **CNNs**: Achieved the highest accuracy at 91%.
* **LSTMs**: Closely followed with 90% accuracy.
* **Random Forests**: Obtained 88% accuracy.
* **SVM**: Recorded 85% accuracy.

In conclusion, in the case of pose estimation, HRNet offers superior accuracy, while OpenPose provides faster processing speeds, making the choice dependent on the task’s requirements. As for HAR, deep learning models like CNNs and LSTMs demonstrate high accuracy, but considerations regarding computational resources and real-time processing requirements are key variables in selecting the appropriate algorithm.

**Task 4 – Question 3**

Google's MediaPipe framework is a well known method for recognizing and following human hands in computer vision applications. Using machine learning models that can recognize hands and estimate poses, MediaPipe offers a complete hand tracking solution. It locates hand landmarks with great accuracy by combining CNNs with specific keypoint recognition. [15].

Using a trained deep learning model, the pipeline locates the hand's main features, including the fingertips, knuckles, and palm center, after first identifying a hand inside an image or video frame. Even under difficult real-world circumstances including shifting hand orientations, occlusions, and lighting, it performs well. The usage of pre-trained models that leverage substantial datasets of annotated hand photographs for training is responsible for the model's resilience [16].

**Task 4 – Question 6**

The model training overall concluded positively, this is shown by the reduction in loss values for box\_loss, cls\_loss, and dfl\_loss on both training and validation sets which indicates successful learning and convergence.

The model appears to also be successfully learning to detect objects, as evidenced by the improvement of key performance metrics such as precision, recall, mAP50, and mAP50-95 over 50 epochs which indicate a good model training process.

A collage of different types of bread

AI-generated content may be incorrect.  
Example identification for all types of bread

A blue squares with white text

AI-generated content may be incorrect.  
Confusion Matrix

A graph of a variety of colors

AI-generated content may be incorrect.  
Precision-Recall curve for all classes

A group of graphs showing different types of data

AI-generated content may be incorrect.  
Overall results of trainingA diagram of a diagram

AI-generated content may be incorrect.  
Training Pipeline

**Task 4 – Question 7**

  
Example identification of hand holding bread

**Bibliography**

1. M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 59–73, 2007.
2. J. Gao, S. J. Kim, and M. S. Brown, "Constructing image panoramas using dual-homography warping," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 49–56.
3. H. Guo, S. Liu, T. He, S. Zhu, B. Zeng, and M. Gabbouj, "Joint video stitching and stabilization from moving cameras," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5491–5503, 2016.
4. D. J. Brady, M. E. Gehm, R. A. Stack, D. L. Marks, D. S. Kittle, D. R. Golish, E. M. Vera, and S. D. Feller, "Multiscale gigapixel photography," *Nature*, vol. 486, no. 7403, pp. 386–389, 2012.
5. R. Shao, G. Wu, Y. Zhou, Y. Fu, L. Fang, and Y. Liu, "LocalTrans: A multiscale local transformer network for cross-resolution homography estimation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 14890–14899.
6. J. Ying, H.-L. Shen, and S.-Y. Cao, "Unaligned hyperspectral image fusion via registration and interpolation modeling," *IEEE Trans. Geosci. Remote Sens.*, 2021.
7. Y. Zhou, A. Rangarajan, and P. D. Gader, "An integrated approach to registration and fusion of hyperspectral and multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3020–3033, 2019.
8. X. Zhan, Y. Liu, J. Zhu, and Y. Li, "Homography decomposition networks for planar object tracking," in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, 2022, pp. 3234–3242.
9. H. Zhang and Y. Ling, "HVC-Net: Unifying homography, visibility, and confidence learning for planar object tracking," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 701–718.
10. M. Elgendi, M. U. Nasir, Q. Tang, D. Smith, J.-P. Grenier, C. Batte, B. Spieler, W. D. Leslie, C. Menon, R. R. Fletcher, N. Howard, R. Ward, W. Parker, and S. Nicolaou, "The effectiveness of image augmentation in deep learning networks for detecting COVID-19: A geometric transformation perspective," *Front. Med.*, vol. 8, Art. no. 629134, Mar. 2021, doi: 10.3389/fmed.2021.629134.
11. M. Fincato, F. Landi, M. Cornia, F. Cesari, and R. Cucchiara, "VITON-GT: An Image-based Virtual Try-On Model with Geometric Transformations," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2020, pp. 1-8. DOI: 10.1109/ICPR48806.2021.9412052. [Online]. Available: <https://www.researchgate.net/publication/350124727>.
12. CMU-Perceptual-Computing-Lab, "OpenPose," GitHub, 2025. [Online]. Available: https://github.com/CMU-Perceptual-Computing-Lab/openpose. [Accessed: Mar. 14, 2025].
13. HRNet, "HRNet: High-Resolution Representations for Vision Tasks," GitHub, 2025. [Online]. Available: https://github.com/HRNet. [Accessed: Mar. 14, 2025].
14. K. Wason, “Human pose estimation: OpenPose vs. HRNet,” Medium, 2020. [Online]. Available: https://kartikwason.medium.com/human-pose-estimation-openpose-vs-hrnet-e8fa37768929. [Accessed: Mar. 14, 2025].
15. S. Kleanthous, N. Tsapatsoulis, and C. Christodoulou, “A comparative analysis of hybrid deep learning models for human activity recognition,” *Sensors*, vol. 20, no. 19, p. 5688, 2020. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/33036479/>. [Accessed: Mar. 14, 2025].
16. P. N. Huu, N. N. Thi, and T. P. Ngoc, “Proposing posture recognition system combining MobileNetV2 and LSTM for medical surveillance,” *IEEE Access*, doi: 10.1109/ACCESS.2021.3138778. [Online]. Available: <https://www.researchgate.net/publication/357359945_Proposing_Posture_Recognition_System_Combining_MobilenetV2_and_LSTM_for_Medical_Surveillance>. [Accessed: Mar. 17, 2025].