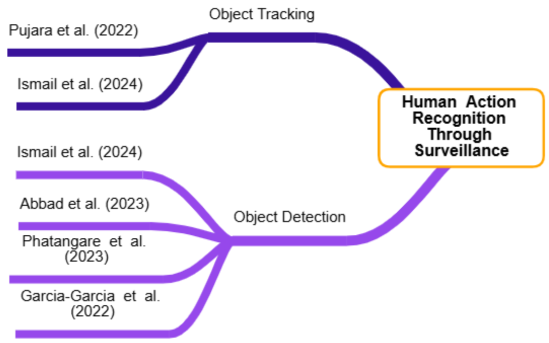
Literature Map



Comparison Table

Author(s), Year, reference	Topic	Techniques used	Datasets used	Training Images	Testing Images
Abbad et al. 2023 [1]	Unsafe Actions Detections for Humans using YOLOv7	YOLOv7	UT-Interaction and ISR-UoL 3D Social Activity	465	N/S
Ismail et al. 2024 [2]	Customer Activity Detection Using YOLOv8 and Status Order Algorithm	YOLOv8 and DeepSORT	Custom	N/S	N/S
Phatangare et al. 2023 [3]	Real Time Human Activity Detection using YOLOv7	YOLOv7	COCO and custom	N/S	N/S
Garcia-Garcia et al. 2022 [4]	Human Activity Recognition implanting the Yolo models	YOLOv5, YOLOv6 and YOLOv7	UR Fall detection and custom	70%	15%
Pujara et al. 2022 [5]	DeepSORT: Real Time & Multi-Object Detection and Tracking with YOLO and TensorFlow	YOLOv7 and DeepSORT	COCO and Custom	N/S	N/S

N/S – Representing 'Not Specified'.

Literature Review

Human Activity Recognition (HAR) has garnered significant attention due to its application in surveillance, healthcare, and human-computer interaction. Recent advancements have leveraged deep learning techniques, notably the You Only Look Once (YOLO) object detection algorithm and the DeepSORT tracking algorithm, to enhance detection and tracking accuracy.

Object Detection Using YOLO

In recent times the You Only Look Once (YOLO) family has become widely used in real-time object detection applications due to its speed and time-efficiency [1][2]. Unlike two-stage detectors such as R-CNN, Fast R-CNN, and Faster R-CNN [3], YOLO is a one-stage Convolutional Neural Network (CNN) detector, that simultaneously detects objects, their bounding boxes and classifies them by only requiring a single scan of an image [1] [4].

YOLO functions by splitting an image into a grid of S×S. Within each grid cell, multiple bounding boxes are predicted and each contains the coordinates, class probability score and confidence score for a potential detected object [3].

Following predictions, its common for multiple bounding boxes to overlap resulting in the same object being predicted. YOLO employs Non-Maximum Suppression (NMS) to resolve this problem. NMS suppresses bounding boxes that have a low final score and low Intersection over Union (IoU) score. The final score reflects the likelihood that an object exists, and that it is of a certain class. Final Score = Confidence Score × Class Probability Score. The IoU score reflects the likelihood that the object's predicted bounding box is closer to the ground-truth bounding box. IoU = Area of Intersection / Area of Union [3] [4].

A comparison of YOLO iterations and why other researchers have used it, is provided in the following section:

While conducting research on detecting five unsafe human actions, the researchers opted to use YOLOv7 since previous research had shown that compared to previous versions, the YOLOv7 version was the fastest and most accurate real-time object detection algorithm. Their model had obtained a result of precision 96%, recall of 95% and Mean Average Precision (mAP) of 99% at an IoU of 0.5 [1].

While conducting research on detecting three human actions, the researchers opted to work with YOLOv5, YOLOv6 and YOLOv7. Previous research made a comparison from YOLOv1 to YOLOv4 and concluded that YOLOv4 obtained the best results. After training the models, YOLOv7 obtained a better result in performance and activity classification, receiving a precision score of 96% in standing, 94% in sitting and 97% in falling [4].

While conducting research on a customer activity detection system in fast food courts, the researchers opted to use YOLOv8 for object detection due to its state-of-the-art (SOTA) performance and real-time processing capabilities. The system's obtained precision score for each activity was: 72% eating, 78% drinking, 75% sitting, 70% eating and drinking, and 70% empty table [2].

While conducting research on detecting three critical safety concerns in real-time, the researchers opted to use YOLOv7 as it was the most recent version, and it solved issues that previous YOLO iterations had. YOLOv7 obtained an average precision of 37.20% at an IoU of 0.5, which is higher than that of YOLOv5 when recognizing objects. YOLOv7 is faster than previous iterations since a Graphics Processing Unit (GPU) can be utilized. YOLOv7 makes use of

CSPDarknet53 which is a backbone network that is deeper and more potent than Darknet53 used in YOLOv5 [3].

Pre-processing techniques used with YOLO is provided in the following section:

Resizing the data to the same resolution size is important as the model would be inaccurate on different resolutions. In these studies, the images were resized to 640x640 [1] [4].

Pixel normalization is applied to the data by scaling the pixel values to the range [0,1] [1] [3].

Image smoothing and contrast enhancement is applied to data to increase the model's accuracy and reduce noise [3].

Data augmentation is used to enhance the training quality by modifying the training data using several techniques:

Abbad and Ibraheem had used several techniques to generate three modified versions of each image, resulting in a higher quality training dataset. 50% probability horizontal flipping, random cropping between 0% and 20%, random rotation between -10 and +10 degrees, random brightness of between -25% and +25%, and random exposure adjustment between -25% and +25% [1].

Phatangare et al. had used a few techniques, but there was never a mention of the values. Random rotation, horizontal flipping and random brightness modifications.

Object Tracking Using DeepSORT

Research was conducted on object tracking methodologies, analyzing and reviewing both object detection and tracking techniques to identify which algorithms are obtaining high accuracy in real-time detection. Object tracking involves a number of phases, including object detection, object classification, assigning unique identifiers for discovered objects and tracking the object.

Object tracking may be done in a variety of ways. Three popular techniques are the target tracking based on Mosse with Kernelized Correlation Filters (KCF), target tracking based on the Siam algorithm, and DeepSORT based on YOLO. Target tracking based on Mosse with KCF obtains high speeds and requires minimal hardware needs, however it is easy to lose the targeted object since the tracking frame scale is constant and cannot follow the target's changing scale.

DeepSORT is built upon the Simple Online and Realtime Tracking (SORT) algorithm, created to track multiple objects simultaneously. DeepSORT enhances SORT's object identification, since it would sometimes mistakenly assign a new or incorrect ID to an object.

DeepSORT based on YOLO uses the YOLO algorithm to detect objects, and after detection DeepSORT would be used for object tracking. DeepSORT starts by employing appearance feature extraction using the Residual Network (ResNet) which is a pre-trained CNN. Extracted features are used in future frame to re-identify objects by using the Hungarian algorithm. Motion prediction would be applied using Kalman filters to predict the future positions of the detected objects based on the objects' previous states.

Kalman filters function by predicting the object's movement by tracking its relative consistent speed and its previous predictable movement pattern. Certain situations might result in an object not being clearly visible, Kalman filters are used to predict the object's position, bridging the gap until the object becomes clearly detectable again [5].

Activity Recognition Methods

Two studies were analyzed from Abbad and Ibraheem, and Garcia-Garcia and Pinto-Elias. Both studies used YOLO for object detection and classification of activities [1] [4]. Although this solution is a success for both studies, in other environments a result with false positives can be obtained since multiple potential activities can be detected simultaneously around a person, and with no confident way of knowing whether a person is participating in the detected activity.

A customer activity detection system was developed by utilizing a restaurant's surveillance system. Four polygonal zones were drawn, each corresponding to a table's seating area. Activities at each table were determined based on the presence of objects in the table's spatial zone. A hierarchical rule set was created to prioritize and classify activities when multiple activities were detected simultaneously [2].

S. Phatangare, S. Kate, D. Khandelwal, A. Khandetod, and A. Kharade, used YOLO for object detection and classification, while activity recognition was handled by rule-based logic [3].

References

[1] M. F. Abbad and I. N. Ibraheem, "Unsafe Actions Detection for Humans using YOLOv7", 2023 International Conference on Artificial Intelligence Robotics, Signal and Image Processing (AIRoSIP) , 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10874011.

[2] M. Ismail et al, "Customer Activity Detection Using YOLOv8 and Status Order Algorithm", 2024 International Conference on Cyberworlds (CW), 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10917483.

[3] S. Phatangare, S. Kate, D. Khandelwal, A. Khandetod, and A. Kharade, "Real Time Human Activity Detection using YOLOv7", 2023 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10290168.

[4] S. Garcia-Garcia and R. Pinto-Elias, "Human Activity Recognition implenting the Yolo models", 2022 International Conference on Mechatronics, Electronics and Automotive Engineering (ICMEAE), 2022. [Online]. Available: https://ieeexplore.ieee.org/document/10414498.

[5] A. Pujara and M. Bhamare, "DeepSORT: Real Time & Multi-Object Detection and Tracking with YOLO and TensorFlow", 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), 2022. [Online]. Available: https://ieeexplore.ieee.org/document/10011018.