Problem : People Fall detection

People's Fall is a serious concern especially since it is life threatening sometimes. Hence we need a solution that generally identifies fall of people and specifically at staircases, escalators, steps etc.

Library/Algorithms to use

YOLO CV Library, Pose Estimation algorithm, open source models OWL-ViT, OWL-V2

Input

Offline Videos in mp4 format

Evaluation Criteria

High level of Accuracy of predictions

Extremely low False Positives

Output &amp:

(a) Fall Detection Approach document – A writeup explaining various possible approaches and provide a recommendation for an approach to be implemented. Recommend using Zero shot object detection models for training Phase

(b) Provide skeletal solution of training and prediction phases

Approaches:

YOLO CV Library:

YOLO (You Only Look Once) is a fast multi-object detection algorithm which uses CNN capable of real-time object detection in images and videos. It divides the input image into grids and predicts bounding boxes and class probabilities within each grid cell. In our case, we can train our model on a labeled dataset containing images or video frames of people falling. Now the model can be used to detect falls in real-time by analyzing each frame and identifying the presence of falling individuals and to specify fall-related classes, like "falling," " lying on the ground," etc.We can detect people in video using pre-trained model without custom training it.Then we Fine-tune the YOLO model to optimize performance for fall detection scenarios.

Pose Estimation Algorithm:

Pose estimation algorithms analyze human body poses and can be utilized for fall detection. By estimating the positions of key body joints, such as the shoulders, hips, and knees, we can infer the pose of an individual. to analyze body poses and detect abnormalities indicative of falls,Train the pose estimation model on the annotated dataset to recognize key body keypoints and infer body poses accurately. Develop custom post-processing techniques to interpret pose estimation outputs and identify fall-related poses, such as "lying down" or "unbalanced posture."Falls often exhibit distinct body orientations and motions, which can be detected by monitoring changes in joint positions over time. Here we should train a model on fall-specific poses and use it to track and analyze poses in real-time video streams.

OWL-ViT and OWL-V2 Models:
These are open-source deep learning models that leverage transformers for image analysis tasks. These models have shown promising performance in image classification and object detection. The goal is to learn the text embedding that aligns with the image embedding. A multi-head attention pooling (MAP) is used for aggregating image representations.To use these models for fall detection, we can employ transfer learning techniques. First, we train the models on a large dataset of labeled video frames containing falls. After training, we can fine-tune it for fall detection . By analyzing the output predictions of the models, falls can be identified.

Combining algorithms mentioned above:
This combination helps us achieve better results.
We preprocess the data by resizing videos, extracting frames. For object detection, we use the YOLO CV Library, training a model to detect individuals and fall-related classes. We use pose estimation algorithms to analyze body poses and detect abnormalities that indicate falls. Using pre-trained OWL-ViT and OWL-V2 models, we further extract features and classify fall events.In this approach, we integrate these techniques, and emphasize data collection, model training, fine-tuning, and post-processing. By combining the outputs of object detection and pose estimation algorithms and evaluating performance metrics like precision and recall, we achieve optimal accuracy while minimizing false positives.

Zero Shot object detection:
we can implement zero-shot detection,here models detect falls without explicitly training.
Here we should be able to create semantic embeddings or attributes that describe falls . These embeddings capture essential characteristics of falls, such as body pose, motion dynamics. This can be done from domain knowledge, expert input, or learned directly from data using techniques like attribute learning.Then we should choose a model which is capable of detecting falls based on input data and these embeddings. We can use convolutional neural networks (CNNs), recurrent neural networks (RNNs), or OWL-ViT.
The zero-shot detection involves the chosen model to perform fall detection by leveraging semantic embeddings. Our input data is processed, and feature representations are extracted. These features are then matched with semantic embeddings or attributes associated with falls using similarity measures or other techniques.
Using similarity scores that are obtained,we apply thresholding techniques to determine whether a fall has occurred based on these scores . If the similarity score exceeds threshold, the input data is classified as a fall . Additional factors, like presence of obstacles or environmental factors, can be considered to improve detection accuracy.
Performance is evaluated based on validation set containing fall-related data, and the model is fine-tuned and hyperparameters adjusted to optimize detection accuracy and minimize false positives.

References:

https://www.machinelearningnuggets.com/object-detection-with-vision-transformer-for-open-world-localization-owl-vit/
https://opencv-tutorial.readthedocs.io/en/latest/yolo/yolo.html
https://viso.ai/deep-learning/pose-estimation-ultimate-overview/