**Proposal for Thesis: A safety camera solution pipeline to alert anomaly behaviors.**

**Motivation:**

* Video anomaly detection represents a significant area of research and development, with applications spanning critical domains. The most common applications can be:
  + Security and Surveillance: Video anomaly detection can help identify unusual activities, such as unauthorized access, theft, or vandalism. It can alert security personnel to take immediate action.
  + Traffic Monitoring: It can be used to detect accidents, traffic jams, or any unusual behavior on the roads. This helps in managing traffic flow and ensuring road safety.
  + Healthcare: In hospitals, video anomaly detection can monitor patients for unusual movements or behaviors, which might indicate a fall or other medical emergencies.
  + Retail: In stores, it can help identify suspicious activities like shoplifting or unusual customer behavior, improving loss prevention strategies.
* Despite the advancements in algorithmic approaches, the practical deployment of video anomaly detection systems with specialized data remains challenging due to the complexities associated with model training, hyperparameter optimization, etc.
* These challenges often require specialized expertise, limiting the accessibility of this technology to a restricted group of researchers and practitioners.
* Meanwhile, data from each user might be private and not able to be public for practitioners, thus, user shall need to self-train the models.
* Finally, in current domain, researchers often focus on improving models’ accuracy and precision, but rarely put infer time into consideration. Therefore, despite the existence of many advanced domain adaptation methods, it is difficult to deploy these methods on resource-constrained devices due to their model complexity.
* This work addresses these limitations by presenting an automated pipeline for video anomaly detection. The proposed system facilitates the training of anomaly detection models from user-provided video data, thereby eliminating the need for extensive manual configuration. Also, sample models in pipeline shall be tested to see possible FPS support.

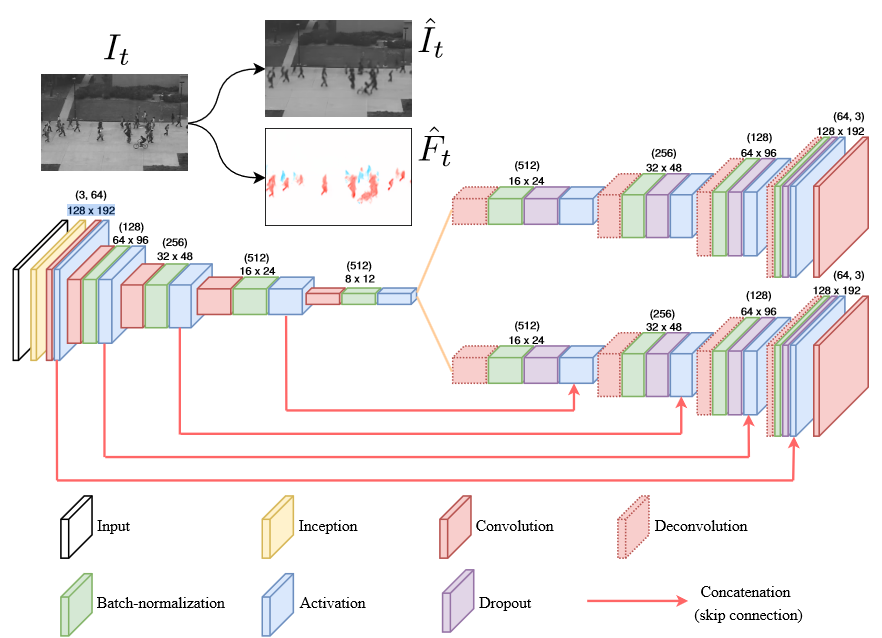
**Problem description:**

* In this alert system, most important part is flexibility for user to use their own data to train for specialized use cases to detect anomalies.
* The main components of this system shall be:
  + A simple training pipeline application to get customized input data from user
  + Set of selectable lightweight models for most common types of anomaly detection (Reconstruction, Prediction or Hybrid)
  + The data as images from security cameras
  + The anomaly detection component (output from customized training model)
  + The alert component to inform user on anomalies.

**Related works:**

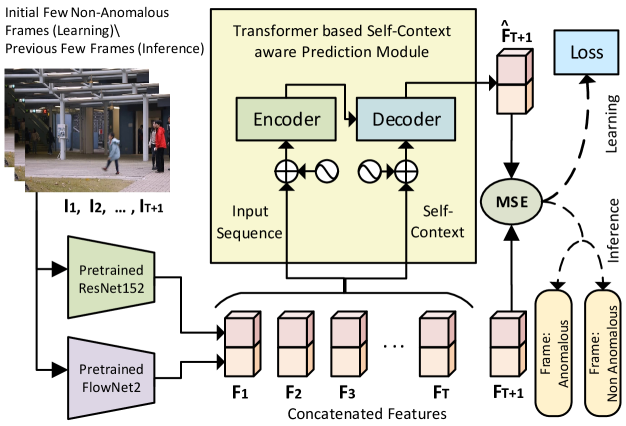
**Anomaly Detection in Video Sequence with Appearance-Motion Correspondence** [https://arxiv.org/pdf/1908.06351]

* The model includes two processing streams. The first one is performed via a Conv-AE to learn common appearance spatial structures in normal events. The second stream is to determine an association between each input pattern and its corresponding motion represented by an optical flow of 3 channels (xy displacements and magnitude)



* Analysis: The model applies unsupervised learning to both reconstruction and prediction to score the anomaly point. However, due to multilayers with multi-head architect, the model is too heavy for real-time application deployment.

**Transformer based self-context aware prediction for few-shot anomaly detection in videos** [<https://arxiv.org/html/2503.00670v1>]

* The model based on one-class few-shot learning of the transformer network only using the initial few frames in that video. The transformer network’s encoder gets the features of a few consecutive video frames as the input and its decoder predicts the feature vector of the subsequent frame in the video as the output. The input features into the encoder are employed as the input to the decoder as well, which allows them to act as a self-context attending over all the frames in the input sequence.
* **Analysis**: The model leverages self-context via Transformer to reduce need of data, which make it effective for customize data on new anomaly types. However, the Self-Attention Mechanism scales quadratically with the number of tokens, which is computationally intensive for temporal data. This leads to even author reduce number of Encoder/ Decoder layers to adapt for specific use cases. This makes the solution not suitable for end user, who might not familiar with model tuning.

**Objectives of the thesis:**

* Gain comprehensive understanding of vision Anomaly Detection model
* Implement solution based on camera, using customized input data from user, suggest standard models with required fps to detect anomalies.
* Identify suitable benchmark for the solution

**Scope of the thesis:**

* The thesis primarily concentrates on techniques that enable fast anomalies detection with acceptable false positive.
* Available datasets: UCSD, CUHK Avenue, ShanghaiTech Campus, UCF-Crime, XD-Violence
* Custom dataset can also be added as demonstration.

**Draft solution:**

* To leverage custom dataset with least labeling requires for user, we can focus on Weakly-Supervised, Unsupervised models.
* The pipeline shall include:
  + Data preprocessing: Video/ Image sets shall be loaded into correct format before feeding to models. Also, augmentation could also be applied to generalize data:
    - Spatial Transformations
    - Temporal Transformations
    - Occlusion
    - Etc
  + Feature extraction backbones:
    - There shall be backbones which can capture spatio, temporal or spatio-temporal features
    - Adaptive head shall be provided to feed feature map to anomaly detection head
  + Anomaly detection head which include:
    - Reconstruction/ Prediction or both techniques
    - Relevant Loss functions
    - Anomaly score calculation
    - Decision base on anomaly score
  + Evaluation section for user:
    - User shall be able to feed test video to model to check whether the anomaly detection is match the expectation or not
  + Limitation:
    - Due to unsupervised training, it is difficult to understand in case False Positive, False negative is detected.
    - A simple reconstruction distance map can give rough information on this
* Contribution of the thesis:
  + Re-test on available models for possibility to deploy on real-time usage
  + Comparison for impact of different backbones, techniques to overall performance and infer time.
  + A simple pipeline application for end user