Zusammenfassung der Literatur

**Paper 1**

1. **General infos**
2. Main idea:
   1. unsupervised learning/ semi-supervised
   2. pattern recognition
   3. key problem: to design or learn a **feature representation** of input sample space to d-dimensional vectors, that captures “normal” motion and spatial appearance patterns.
   4. goal: to identify abnormal patterns or motions in data that are infrequent or rare events
3. metric
   1. reconstruction error: measuring the approximation error geometrically in a vector space
   2. prediction error: the posterior probability of a given model
4. obstacles
   1. anomalies are rarely annotated
   2. labeled data rarely available to train a network
5. VAD dataset
   1. UCSD Ped1/Ped2 [2]
      1. pedestrian videos from campus surveillance
      2. anomalies:
         * **prohibited objects**: appearance of a cyclist, a wheelchair, and a car in the scene that is usually populated with pedestrians walking along the road
         * **abnormal movements**: People walking in unusual locations
   2. CUHK Avenue Dataset
      1. from campus surveillance
      2. Anomalies: strange actions such as a person throwing papers or bag, moving in unusual directions, and appearance of unusual objects like bags and bicycle
   3. ShanghaiTech [2]
      1. from campus surveillance
   4. UCFCrime [2]
      1. Anomalies: accidents, robbery, and theft.
   5. the Subway entry and exit datasets
      1. anomalies: people moving in the wrong direction, loitering and so on
   6. The Train dataset
      1. moving people in a train
      2. anomalies: unusual movements of people in the train
   7. the Queen Mary University of London U-turn dataset
      1. normal traffic
      2. anomalies: such as jaywalking and movement of a fire engine
   8. LV dataset
      1. online video anomaly detection.
6. **VAD Models**
7. reconstruction Models:
   1. Concept:
      1. minimize the reconstruction error of training samples from the normal distribution
      2. Anomalies represent poorly reconstructed deviations
   2. Principal component analysis (PCA),
      1. linear approximations
      2. reduction in dimensionality is used to capture the anomalous behavior as that not well reconstructed samples
   3. Autoencoders (AEs)
      1. non-linear point-wise transform
      2. neural network trained by back-propagation and provides dimensionality reduction by reducing the reconstruction error on the training set
      3. maps input x to latent space z
      4. variations
         * convolutional AutoEncoder (ConvAE) [2](14)
           + encodes temporally stacked images with 2D convolutional encoders

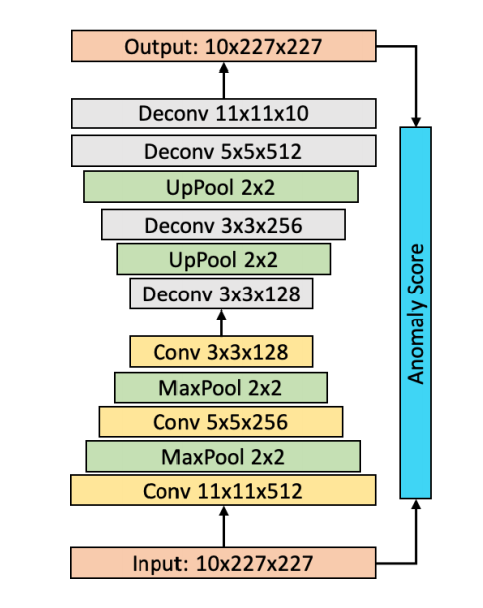


Figure. ConvAE

* + - * + anomaly score map: per-pixel reconstruction error and mean squared error (MSE) is computed as a frame-level anomaly score
      * ConvLSTMAE [2](6)
        + models spatial and temporal features separately
        + 2D CNN encoder first captures spatial information
        + a multi-layer ConvLSTM recurrently encodes temporal features

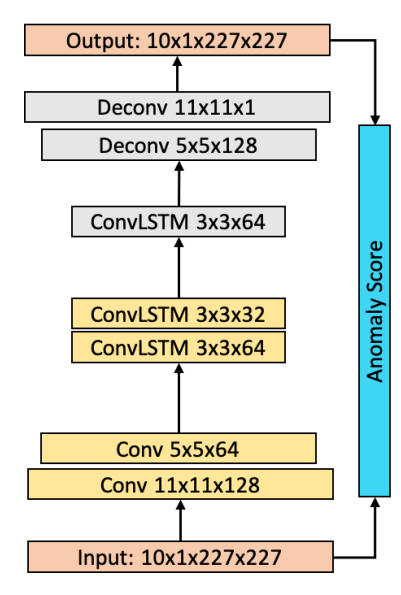


Fig. ConvLSTMAE

* + - * MemAE [2](13)
      * TAD: RNN encoder-decoders [2](44)
        + models normal bounding box trajectories in traffic scenes
        + encode past trajectories and ego motion and to predict future object bounding boxes.
* standard deviation of predictions serves as the anomaly score

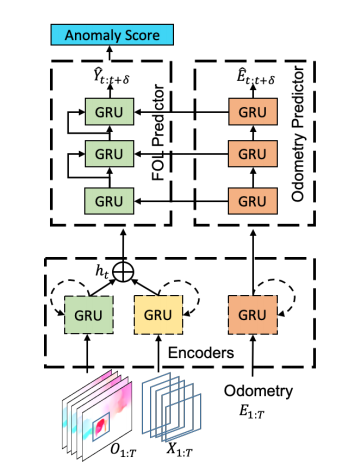


Fig. TAD

1. spatio-temporal predictive models:
   1. Concept:
      1. minimize the prediction error on spatio-temporal sequences from the training series
      2. learn a generative model to predict the current frame or its encoded representation using the past frames
   2. autoregressive models
      1. variations
         * stacked RNN [2](29)
   3. convolutional Long-Short-Term-Memory (LSTM) models

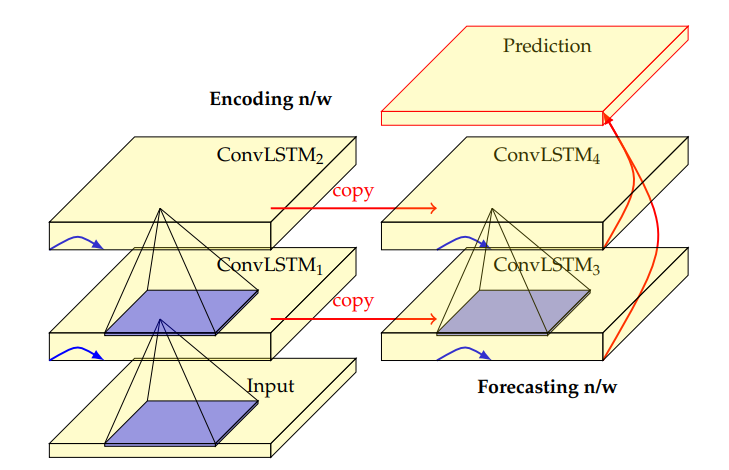


Figure 1. A convolutional LSTM architecture for spatio-temporal prediction

* + 1. variations:
       - Convolutional LSTM Auto-Encoder [2](30,6,28)

1. deep generative models
   1. Concept:
      1. minimizing the reconstruction error as well as distance between generated and training distribution
      2. modeling the likelihood of normal video samples
   2. Variational Autoencoders (VAE)
   3. Generative Adversarial Networks (GAN)
      1. Variations
         * AnoPred [2](26)
           + Input: four continuous previous RGB frames
           + Output: applying UNet to predict a future RGB frame

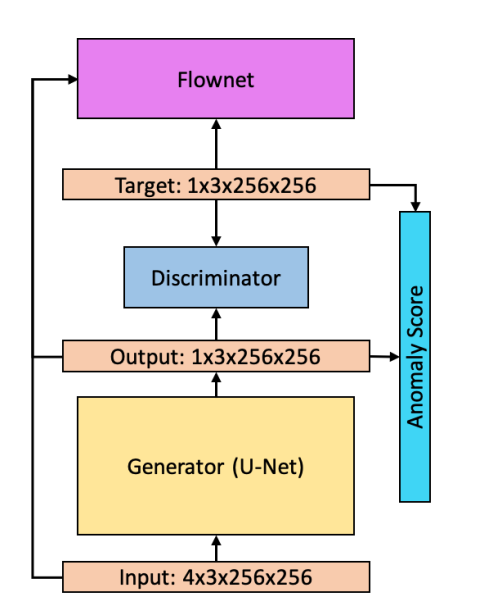
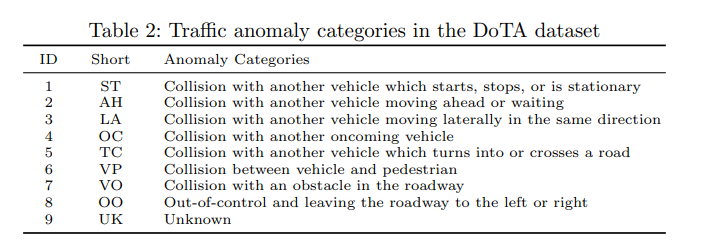


Fig. AnoPred

* 1. Adversarially trained AutoEncoders (AAE)

**Paper2**

1. Motivation
   1. Accurate perception of accident scenarios demands detection, localization, and classification of on-road anomalies for proper reaction and event data recording
   2. challenging problem due to dynamic foreground and background, perspective projection, and complicated scenes
   3. future frame prediction is difficult
2. Main idea
   1. traffic anomaly detection
   2. egocentric videos
   3. a when-wherewhat pipeline to detect, localize, and recognize anomalous events
3. Pipeline
   1. When-Where-What
      1. When the anomalous event starts and ends
         * three temporal partitions:
         * precursor: normal video preceding the anomaly,
         * the anomaly window,
         * post-anomaly:
           + normal activity following the anomaly
           + end: the time when all anomalous objects are out of the field of view or are stationary.
      2. Where the anomalous regions are in each video frame
      3. What the anomaly type is.
   2. Two tasks
      1. video anomaly detection (VAD):
         * per-frame anomaly scores for answer when
         * per-pixel or per-object anomaly scores answer where
      2. video action recognition (VAR)
         * classifies video type to answer what
         * methods: R(2+1)D and SlowFast
4. Existing Action Recognition methods
   1. Two-stream networks (33)
   2. temporal segment networks (TSN) (39)
   3. 3D convolutional networks (C3D) (36):
      1. spatiotemporal modeling,
   4. R(2+1)D (37)
   5. SlowFast (10)
   6. reinforce encoder-decoder(RED) (11)
      1. action prediction and online action recognition
   7. temporal recurrent network (TRN) (41)
      1. online action detection
5. existing VAD dataset in egocentric traffic videos
   1. StreetAccident dataset
      1. on-road accidents with 620 video clips collected from dash cameras.
      2. Anomalies: last ten frames of each clip
   2. A3D dataset
      1. 1,500 anomalous videos
      2. abnormal events are annotated with the start and end times
   3. DADA dataset
      1. for driver attention prediction
6. new dataset
   1. Detection of Traffic Anomaly (DoTA)
   2. 4,677 videos with 1280 ⇥ 720 resolution
   3. With richer temporal, spatial, and categorical annotations using Scalabel
   4. seven common trac participant categories:
      1. person, car, truck, bus, motorcycle, bicycle, and rider
      2. following the BDD100K style
   5. 18 anomaly categories
      1. 9 categories in list
      2. each category is split to ego-involved and non-ego



* 1. largest traffic anomaly dataset to-date and the first supporting traffic anomaly studies across when-where-what perspectives.
  2. Link: https://github.com/MoonBlvd/ Detection-of-Trac-Anomaly

1. New effective metric
   1. spatial-temporal area under curve (STAUC)