*2020 3D-Convolutional Neural Network with Generative Adversarial Network and Autoencoder for Robust Anomaly Detection in Video Surveillance*

Generator: U-Net with 3D CNN. Discriminator: LRCN (CNN and LSTM).

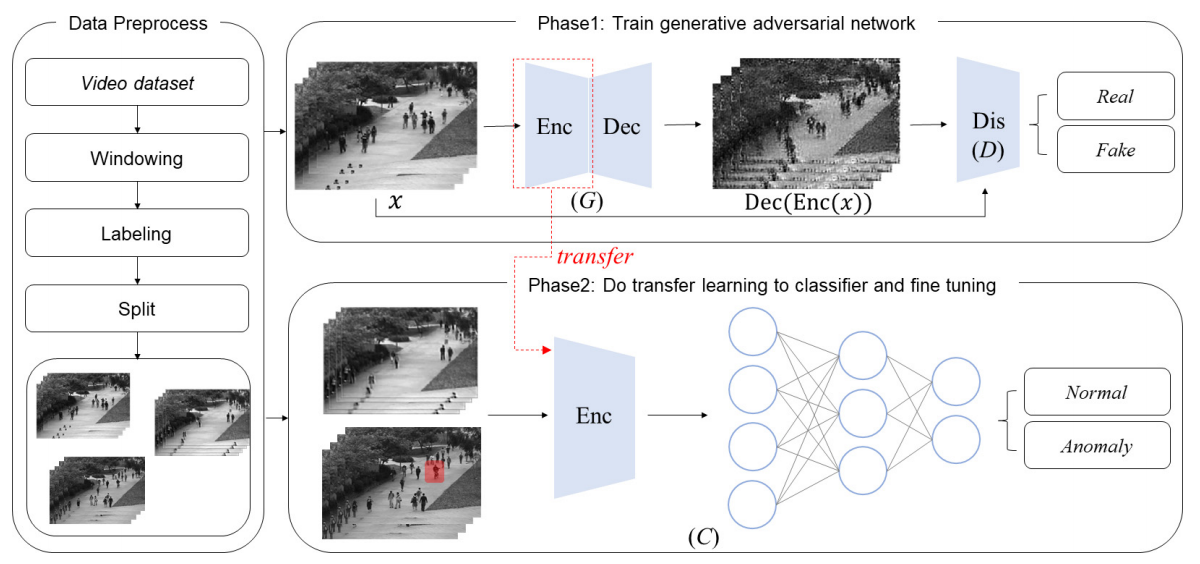
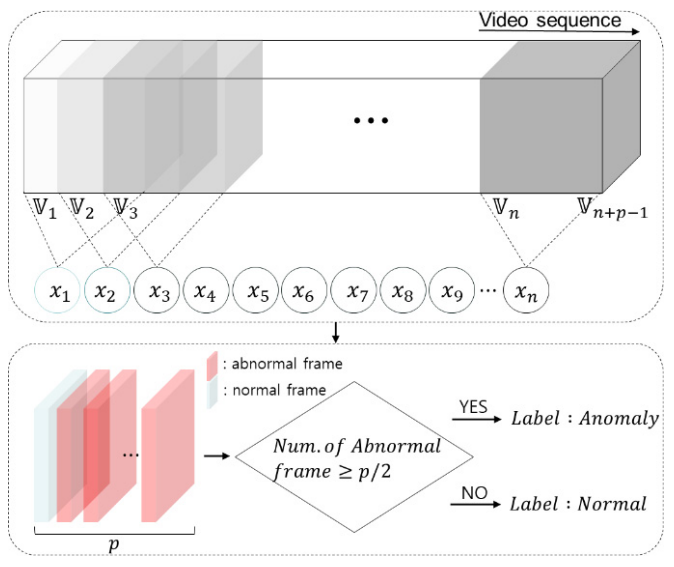
The network contains the preprocessing step and two learning phases.

The **first learning phase** trains GAN that receives a video clip as input and outputs a fake video clip of the same size. In this process, the decoder of generator learns more informative feature extraction.

In the **second phase**, VAD classifier consists of the encoder of generator and general classifier.

*p* is the size of window, which should be larger than 1, and *n* is the number of video clips *V*.

They do not consider point anomaly, and an anomaly occurs over several frames. They label anomaly if the video clip contains more than half anomaly frames.

The **generator** of the proposed method consists of 2 3D convolutional networks and two deconvolutional networks with 5 x 5 and 3 x 3 filter sizes respectively, and the number of channels is 16 and 32 in order. Skip connections and batch normalization is used to ensure learning stability.

The **discriminator** consists of 3 2D time-distributed convolutional networks with 7 x 7, 5 x 5 and 3 x 3 filter sizes. The channel of all convolutional networks is 16. LSTM layer receives the output of the last convolutional network. The output size of LSTM is fixed to 1024. The number of nodes in fully connected layer is set to 256 and 2. The activation function of the output layer is softmax and the function of all the other layers is RELU.

The last fully connected layer of the VAD classifier is designed to have sigmoid activation. The final output value *A* is in range [0, 1].

The *A* exceeds the threshold *T*, the given input is treated as an anomaly. Equal Error Rate (EER) is used as the threshold.