*2017 Abnormal Event Detection in Videos using Generative Adversarial Nets*

Generator: U-Net, Discriminator: 2D CNN

In training steps, network is trained using normal frames and corresponding optical-flow images.

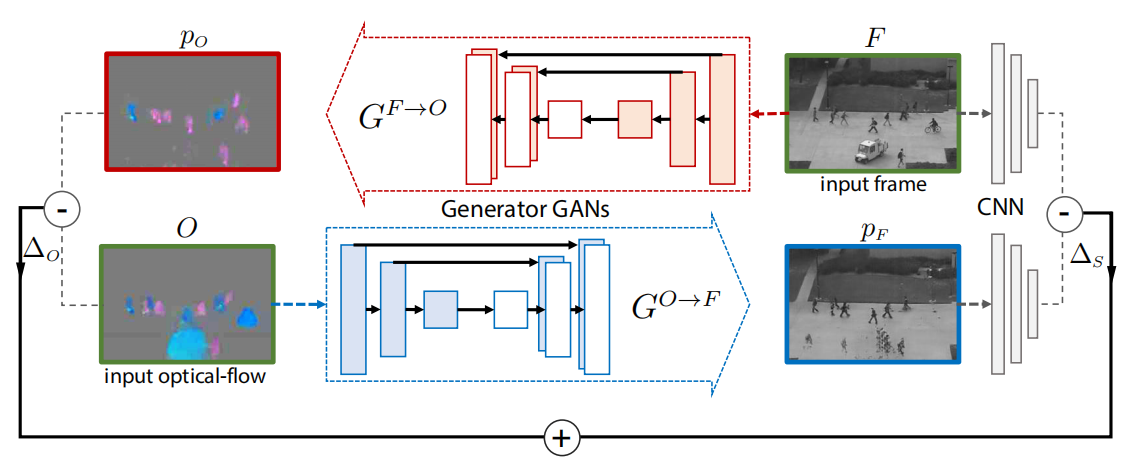
Top: A generator network takes as **input a frame (*F*)** and produces a **motion ()** image.

Bottom: A second generator network is fed with a **real optical-flow image (*O*)** and outputs an **appearance ()** reconstruction.

Since the networks have learned to generate normal reconstructions, they are not able to reconstruct appearance and motion information of the possible abnormal regions of the test frame.

Discriminator takes as input two images (the input of generator and generated image/the input of generator and label image) and output a scalar representing the probability that both its input images came from the real data.

At testing time, the real data are compared with both the appearance and motion representations, and abnormal areas are detected by computing local differences.



*F* is original image, and *O* is its corresponding optical-flow image.

Using *F*, an optical-flow reconstruction can be obtained: , which is compared with *O* using a simple pixel-by-pixel difference, obtaining:

Where highlights the local differences between the real optical flow and its reconstruction and these differences are higher in correspondence of those areas in which was not able to generate the abnormal behavior.

Similarly, using *O*, the appearance reconstruction can be obtained: . A simple pixel-by-pixel difference between *F* and is less informative than the difference computed in optical-flow channel.

Let be the representation of *F* in AlexNet and the corresponding representation of the appearance reconstruction.

A semantics-based difference between *F* and :

Finally, and are fused to obtain a unique abnormality map. Specifically, they first upsample to obtain with same resolution as . Then both and are normalized with respect to their corresponding channel-value range.

For each test video *V*, they compute the maximum value of all the elements of over all the input frames of *V*. The normalized optical-flow difference map is given by:

Similarly, the normalized semantic difference map is obtained using computed over all the elements of in all the frames of *V*:

The final abnormality heatmap *A* is obtained by summing and :

Frame-level abnormality detection:

Abnormality detection criterion is based on checking if the frame contains at least one predicted abnormal pixel. The procedure is applied over a range of thresholds to build ROC curve.

Pixel-level abnormality localization:

A true positive prediction should cover at least 40% the ground truth abnormal pixels. Otherwise the frame is counted as a false positive.