# Video Anomaly Detection by Estimating Likelihood of Representations

Yuqi Ouyang  Victor Sanchez
Signal and Information Processing Lab, Department of Computer Science,
The University of Warwick
yuqi.ouyang@warwick.ac.uk  v.f.sanchez-silva@warwick.ac.uk

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

## Intro

Video anomaly detection refers to detecting abnormal activities or events in a scene. Since it is very challenging to collect and label examples of all possible types of abnormal events in a scene, video anomaly detection is usually solved as an unsupervised learning task, namely, training model with normal data for the objective of detecting outliers.

We propose a deep probabilistic model (named GMM-DAE) to transfer this task into a density estimation problem where latent manifolds are generated by a deep denoising autoencoder and clustered by expectation maximization. Evaluations on several benchmark datasets show the strengths of our model, achieving outstanding performance on challenging datasets.

## Model: Overall

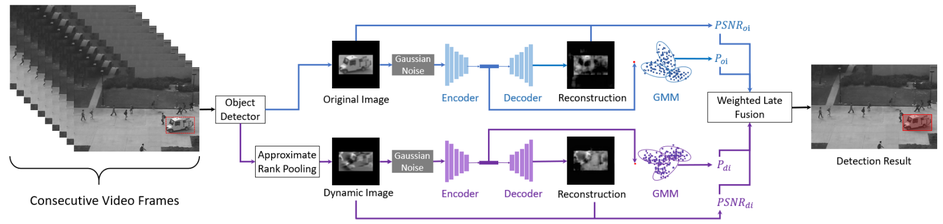


Figure 1: The proposed GMM-DAE model. An object detector is applied to generate patches. A dynamic image is computed by approximate rank pooling for motion info collection. Two denoising autoencoders (DAE) are trained to reconstruct data. Low-dimensional manifolds are clustered with two Gaussian Mixture Models (GMM). The final anomaly score is computed by fusing the reconstruction errors and latent likelihoods.

## Model: Denoise to Reconstruct

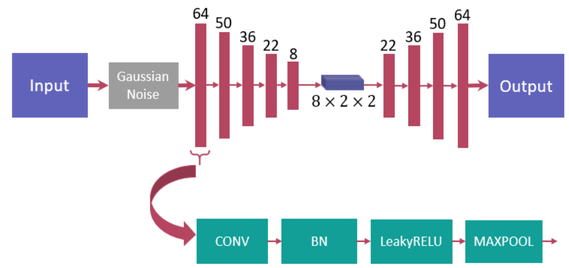


Figure 2: Architecture of the DAE in the GMM-DAE model. CONV: convolutional layer. BN: batch normalization.

## Training to Reconstruct

$$L\left(\boldsymbol{X},\hat{\boldsymbol{X}}\right)=\frac{1}{n}\sum_{i=1}^{n}\|x_i-\hat{x}_i\|_2^2+\beta\|\boldsymbol{W}\|_2^2$$

$\boldsymbol{X}$: $n$ uncorrupted images.
$\hat{\boldsymbol{X}}$: $n$ output images.
$\|\boldsymbol{W}\|_2^2$: $\ell_2$-regularization on the weights by a factor $\beta$.

## Training to Estimate Density

$$\phi_j=\frac{D(\boldsymbol{z_j})}{\sum_{i=1}^{k}D(\boldsymbol{z_i})}\qquad \mu_j=\frac{1}{D(\boldsymbol{z_j})}\sum_{z\in\boldsymbol{z_j}}z$$

$$\Sigma_j=\frac{1}{D(\boldsymbol{z_j})}\sum_{z\in\boldsymbol{z_j}}(z-\mu_j)(z-\mu_j)^T$$

$$\gamma_{ij}=\frac{\omega_{ij}\cdot\phi_j\cdot\exp-\frac{1}{2}(z_i-\mu_j)^T\Sigma_j^{-1}(z_i-\mu_j)}{\sqrt{|2\pi\Sigma_j|}}$$

$$\hat{\phi}_j=\frac{\sum_{i=1}^{n}\gamma_{ij}}{n}\qquad \hat{\mu}_j=\frac{\sum_{i=1}^{n}\gamma_{ij}z_i}{\sum_{i=1}^{n}\gamma_{ij}}$$

$$\hat{\Sigma}_j=\frac{\sum_{i=1}^{n}\gamma_{ij}(z_i-\hat{\mu}_j)(z_i-\hat{\mu}_j)^T}{\sum_{i=1}^{n}\gamma_{ij}}$$

KMeans++ to Initialize Gaussian Mixture $\left\{N(\phi_1,\mu_1,\Sigma_1),N(\phi_2,\mu_2,\Sigma_2),\cdots,N(\phi_k,\mu_k,\Sigma_k)\right\}$

E-Step to Compute Posterior Likelihood $\gamma_{ij}$

M-Step to Update Gaussian Mixture into $\left\{N(\hat{\phi}_1,\hat{\mu}_1,\hat{\Sigma}_1),N(\hat{\phi}_2,\hat{\mu}_2,\hat{\Sigma}_2),\cdots,N(\hat{\phi}_k,\hat{\mu}_k,\hat{\Sigma}_k)\right\}$

## Anomaly Inference

$$PSNR(x,\hat{x})=10\log_{10}\frac{max(x)}{MSE(x,\hat{x})}$$

$$P(z)=\log\sum_{j=1}^{k}\frac{\hat{\phi}_j\cdot\exp-\frac{1}{2}(z-\hat{\mu}_j)^T\hat{\Sigma}_j^{-1}(z-\hat{\mu}_j)}{\sqrt{|2\pi\hat{\Sigma}_j|}}$$

$$A(x^t)=-\left[\lambda_1P(z_{x^t})+\lambda_2\cdot PSNR(x^t,\hat{x}^t)+\lambda_3P(z_{d^t})+\lambda_4\cdot PSNR(d^t,\hat{d}^t)\right]$$

$$A(I^t)=max\{A(x_1^t),A(x_2^t),\ldots,A(x_n^t)\}$$

$PSNR(x,\hat{x})$: Reconstruction accuracy.
$P(z)$: Likelihood of latent representation $z$.
$A(x^t)$: Anomaly score of frame patch $x^t$.
$A(I^t)$: Anomaly score of frame $I^t$.

## Experiments

| Model | UCSD Ped2 | CUHK Avenue | ShanghaiTech |
|---|---|---|---|
| MPPCA [14] | 69.3 | - | - |
| MPPCA+SFA [25] | 61.3 | - | - |
| MDT [25] | 82.9 | - | - |
| Unmasking [13] | 82.2 | 80.6 | - |
| AMDN [42] | 90.8 | - | - |
| FRCN action [10] | 92.2 | **89.8** | - |
| Conv-AE [9] | 90.0 | 70.2 | 60.9 |
| STAE [43] | 91.2 | 77.1 | - |
| GANs [31] | 93.5 | - | - |
| FFP+MC [18] | 95.4 | 85.1 | 72.8 |
| LSA [1] | 95.4 | - | 72.5 |
| MLAD [41] | **99.21** | 71.54 | - |
| sRNN-AE [22] | 92.21 | 83.48 | 69.63 |
| MemAE [8] | 94.1 | 83.3 | 71.2 |
| UnetGAN [28] | 96.2 | 86.9 | - |
| MLEP-FP [19] | - | 89.2 | 73.4 |
| MemAE2020 [30] | 97.0 | 88.5 | 70.5 |
| SDOR [29] | 83.2 | - | - |
| SAGC [26] | - | - | 76.1 |
| GMM-DAE | **96.5** | **89.3** | **81.2** |

Table 1: Frame-level AUROC curve (%) comparison with other baseline models, on three benchmark datasets (Higher is better). Current SOTA performance is indicated in bold text. (Please find all references in the manuscript)

| Method | AUROC |
|---|---|
| DAE+OP | 90.1 |
| DAE+OP+DP | 93.9 |
| DAE+OP+OL | 94.2 |
| DAE+OP+OL+DP+DL (GMM-DAE) | **96.5** |
| AE+OP+OL+DP+DL | 95.8 |

Table 2: Frame-level AUROC curve (%) on the UCSD Ped2 dataset using different components of the proposed GMM-DAE model. O*: Frame patch; D*: Dynamic frame patch; *P: PSNR value; *L: Likelihood value.

## Performance Analysis

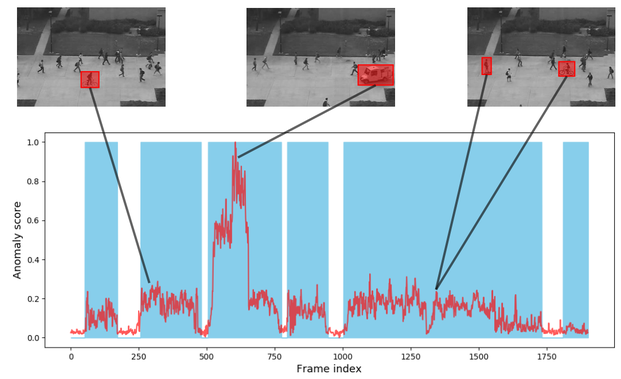


Figure 3: Performance analysis on the UCSD Ped2 dataset. Higher anomaly score indicates more abnormal events happening. Blue areas indicate ground truth abnormal frames. Red boxes indicate correct detections. Activities from left to right in the first row: cycling, vehicle moving, skateboard riding & cycling.

## Latent Space Visualization

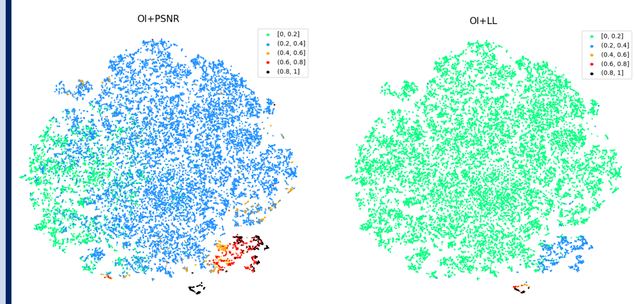


Figure 4: Distribution of the generated manifolds based on the normalized anomaly scores using PSNR values (left: OI+PSNR) and latent likelihood values (right: OI+LL) on the UCSD Ped2 test videos. Each color represents a range of anomaly scores.