

Old is \mathcal{G}^{old} : Redefining the Adversarially Learned One-Class Classifier Training Paradigm

Abstract

This is supplementary material for the paper.

1. Psuedo-Anomaly Module

Summarily, pseudo-anomaly module comprises of the following steps:

- Input two arbitrary training images, X_i and X_j .
- Regenerate the images using a frozen low epoch generator \mathcal{G}^{old} , so that the output images \hat{X}_i^{low} and \hat{X}_j^{low} essentially become a distorted form of the input images.
- Take a pixel-wise mean of these two images which results in a psuedo-anomaly image, $\hat{\bar{X}}$.
- Finally, pass this image to the Generator \mathcal{G} , hence mimicking its reconstruction behavior in the case of an unusual input.

Example images obtained at each step of the pseudo-anomaly module during the training of our framework on Caltech-256, MNIST, UCSD Ped2, are shown in Figure 1.

2. Training of the Discriminator for Anomaly Detection

Figure 2 shows the four types of images that are used in our model to train discriminator \mathcal{D} for anomaly detection. Original training images X and reconstructed images \hat{X} represent good quality reconstructions. Whereas, images \hat{X}^{low} reconstructed using a frozen low epoch generator \mathcal{G}^{old} and the reconstructed psuedo-anomalies \hat{X}^{pseudo} are provided as examples of bad quality reconstructions.

3. Results

Reconstruction examples and the anomaly scores obtained during test time are shown in Figure 3. It can be seen that the network predicts higher anomaly scores whenever the \mathcal{G} fails to regenerate anomalies. However, it is interesting to observe that in the case of reasonably good reconstructions of the anomalous objects, the model still successfully predicts higher anomaly scores. This particular property, as also shown in the ablation study of our framework, can be attributed to the regenerated pseudo-anomaly images obtained using our proposed pseudo-anomaly module. These images can help the discriminator learn about the subtle distortions that often appear in reconstructions of the anomaly inputs. It also makes our model robust against the necessity of having a hard criteria to stop the training hence ensures superior results across various training epochs.

In Figure 4, we present an overview of the performance shown by our model on various test videos of the UCSD Ped2 dataset. Additionally, a video file depicting these results in a visually dynamic form is also provided as part of this supplementary material.

4. Miscellaneous

Additional discussion about Figure 6 of the paper is provided below:

The state of the scores immediately after phase one training (iteration 0) shows that the \mathcal{D} is trained relatively better than the \mathcal{G} at this particular training step. Which means that all the images (inliers or outliers) being generated by \mathcal{G} are recognized by \mathcal{D} as fake. About the higher inlier scores, it can happen in two cases: either \mathcal{D} is over-trained or \mathcal{G} is under-trained. In both cases, the \mathcal{D} might latch on to the particular details that are present in generated images and can successfully classify the reconstructed inliers as fake images, hence the higher scores. The best performance of the baseline model is achieved at the particular sweet spot where \mathcal{G} is good enough to generate high quality inlier images without enough clues for the \mathcal{D} to latch on to and at the same time \mathcal{D} is good enough to recognize the outliers as fake while inliers as real. However, this state is not stable and, due to adversary, one further training step can cause the \mathcal{D} to overfit or underfit on the data coming from \mathcal{G} . This particularly explains why the adversarially learnt one-class classifier can depict high fluctuations in the performance with each training epoch. Moreover, it also justifies our proposed phase two training which, by successfully overcoming this problem, depicts stable performance across various training epochs.



Figure 1: Example images at each step of the psuedo-anomaly module for all three datasets. X_i and X_j are two arbitrary input images taken from the training dataset. The heavily distorted, \hat{X}_i^{low} and \hat{X}_j^{low} , are obtained by regenerating these input images individually through \mathcal{G}^{old} . A pixel wise mean of the \hat{X}_i^{low} and the \hat{X}_j^{low} is then taken to obtain a pseudo-anomaly image \hat{X} . Finally, this \hat{X} is regenerated using the generator \mathcal{G} to obtain \hat{X}^{pseudo} , which is the final output of our pseudo-anomaly module.

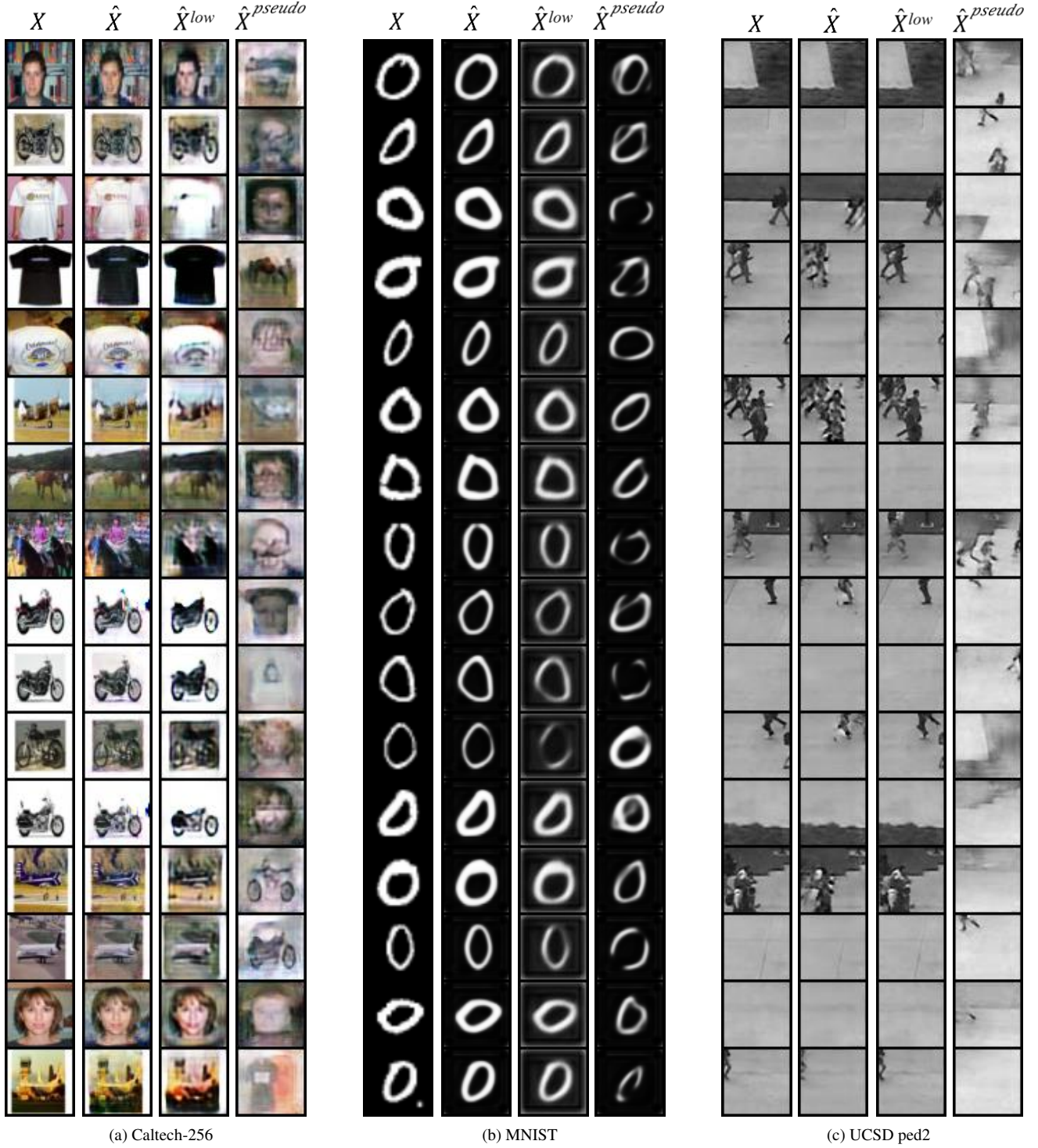


Figure 2: Examples of the images from each dataset, used for the phase-two training of OGNet framework. X are the input images, $\hat{X} = \mathcal{G}(X)$ are the images reconstructed using current generator, \hat{X}^{low} are the images reconstructed using a frozen low epoch generator \mathcal{G}^{old} , and \hat{X}^{pseudo} are the reconstructed anomalies obtained using our pseudo-anomaly module. (a) Inlier input images are taken from 5 classes (horse, airplanes, motorbikes, t-shirt, faces) of the Caltech-256 dataset. (b) Inlier input images are taken from class ‘0’ of the MNIST dataset. (c) Input patches are extracted from the training videos of the UCSD Ped2 dataset which only contain normal scenes of walking pedestrians.

X	$\mathcal{G}(X)$	$\mathcal{D}(\mathcal{G}(X))$	X	$\mathcal{G}(X)$	$\mathcal{D}(\mathcal{G}(X))$	X	$\mathcal{G}(X)$	$\mathcal{D}(\mathcal{G}(X))$	X	$\mathcal{G}(X)$	$\mathcal{D}(\mathcal{G}(X))$
		0.99			0.71			0.01			0.00
		0.06			0.17			0.85			0.12
		0.20			0.76			<u>0.29</u>			0.08
		0.00			<u>0.34</u>			0.99			0.00
		0.29			0.58			0.01			<u>0.49</u>
		0.00			0.33			<u>0.09</u>			0.02
		0.28			0.31			<u>0.20</u>			0.78
		0.27			0.70			1.00			0.00
		0.94			0.56			0.03			0.99
		0.99			0.23			0.99			0.99
		0.00			0.27			1.00			0.01
		0.35			0.68			0.99			<u>0.76</u>
		0.86			0.36			0.87			0.04
		0.52			0.25			0.99			0.12
		0.91			0.41			<u>0.29</u>			0.01
		0.02			0.57			0.79			0.24

(a) Caltech-256

(b) MNIST

(c) UCSD Ped2

Figure 3: Examples of the test images X from each dataset and their reconstructions $\mathcal{G}(X)$ as well as the respective anomaly scores $\mathcal{D}(\mathcal{G}(X))$ obtained through our network. Outliers / anomaly patches and inliers / normal patches are represented with red and green bars, respectively. Although reconstructions of various anomalous inputs are reasonably good in quality, our framework can still predict higher anomaly scores for such images. The output scores generating false alarms are marked as double underlined whereas the scores less than 0.5 for anomaly patches are highlighted as single underlined. Most of these low scores are due to the fact that the anomaly objects are extremely small in size compared to the actual size of the patches. Moreover, it is still fascinating to observe that the anomaly scores for these patches are relatively higher than the scores of most of the normal patches. (a) Training is performed on samples from horse, airplanes, motorbikes, t-shirt and faces classes whereas the outlier images are taken from the clutter class of the Caltech-256 dataset. (b) The model is trained on class '1' of the MNIST dataset and the outliers belong to rest of the classes. (c) Patches extracted from the test videos of the UCSD Ped2 dataset which contain normal scenes of walking pedestrians as well as anomalies such as bikers, skaters and carts.

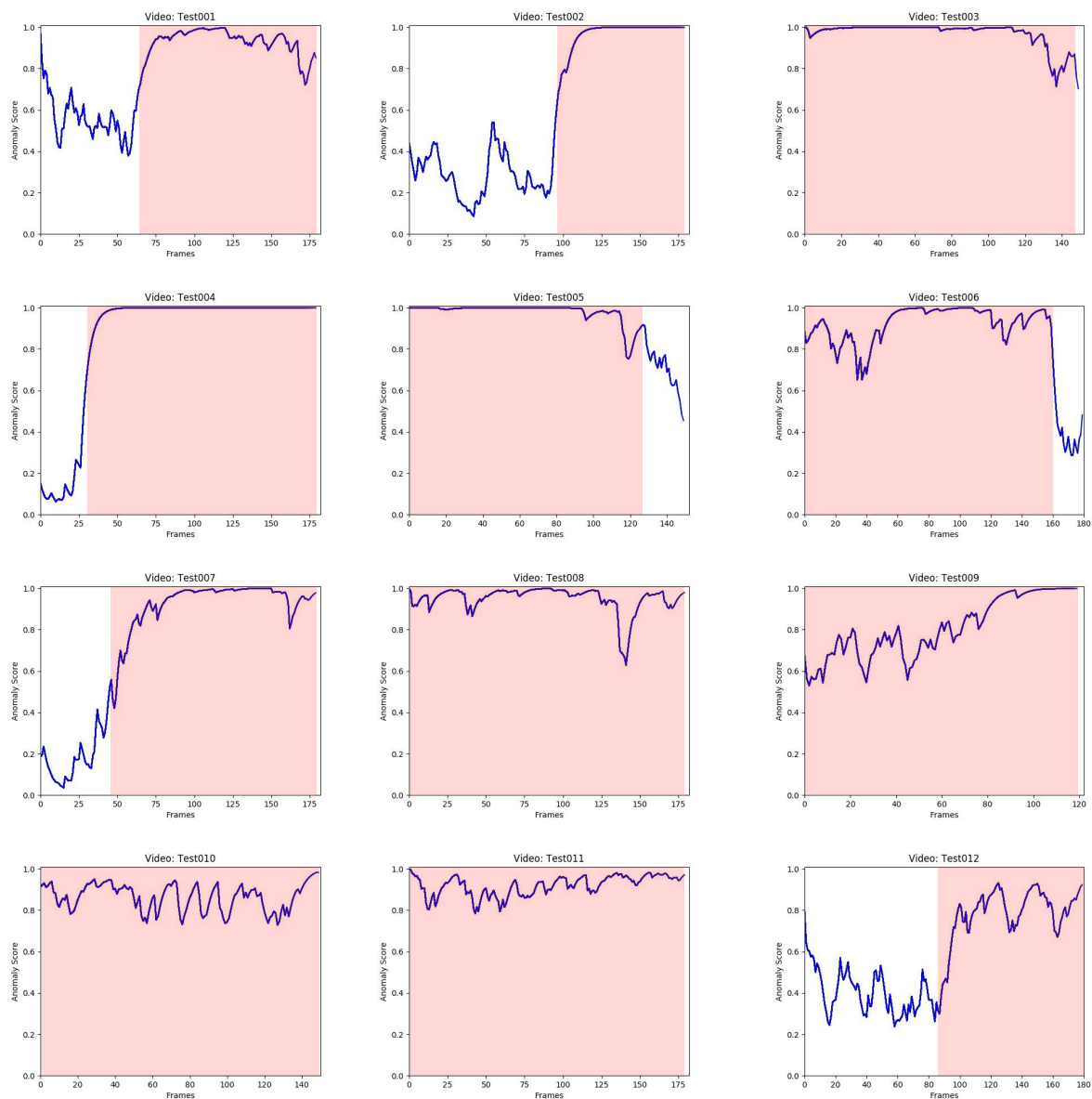


Figure 4: Frame-level anomaly scores ranging 0 to 1 (on the horizontal axis) for various test videos from the UCSD ped2 dataset, provided by our OGNNet framework. Ground truth are represented in pink and the network scores are plotted in blue.