Margin Learning Embedded Prediction for Video Anomaly Detection (Supplementary Material)

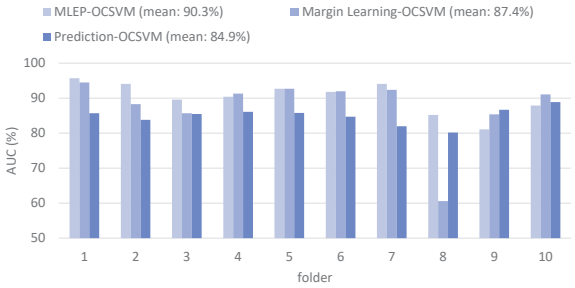
(a) GT on a normal frame (b) Pred on a normal frame (c) GT on a normal frame (d) Pred on an abnormal frame

Figure 1: Failure cases for normal and abnormal frames when imposing $-\|\hat{I}^p_{t+T} - I^p_{t+T}\|_1$ in the prediction loss. GT denotes Ground-Truth and Pred denotes Prediction.

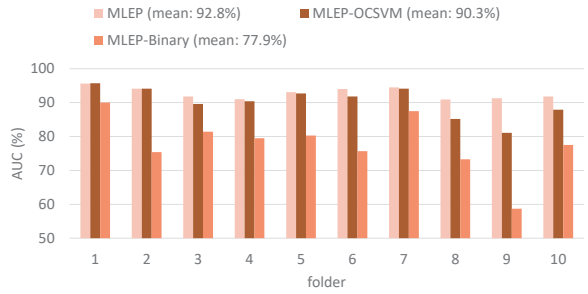
1 Evaluation of different components in MLEP

Network architectures in the prediction module. In this section, we compare our proposed network with some state-of-the-art prediction networks [Villegas *et al.*, 2017] [Ronneberger *et al.*, 2015] [Isola *et al.*, 2017]. For fair comparison, we only replace our proposed network in the MLEP with other networks and keep other modules unchanged. The reason why our network outperforms UNet is that we do not adopt skip connection which may ignore the margin learning and favor the prediction of abnormal frames. In addition, our network has a larger gap between normal and abnormal scores than other advanced prediction networks [Villegas *et al.*, 2017] [Isola *et al.*, 2017], where larger gap helps the classification between normal and abnormal events. Thus, our network is more suitable for anomaly detection with a few anomalies than other ones.

The losses in the future prediction module. As we mention before, it is straightforward to deploy $-\|\hat{I}^p_{t+T} - I^p_{t+T}\|_1$ in the prediction loss to enlarge the gap between normal and abnormal data. As shown in Figure 1, however, directly enforcing abnormal frames being unpredictable leads to a failure for predicting the whole image even for normal data and the AUC drops from 92.8% to 88.5%. As shown in Figure 1, normal and abnormal frames share the same background in the scene, so that enforcing the abnormal frames to be with a large prediction error will distort the prediction of normal frames. Therefore, we do not regularize this constraint in the prediction module, while the margin learning module can handle this.



(a) Comparison of different features



(b) Comparison of different classifiers

Figure 2: We conduct comparison with different features and classifiers on the Avenue dataset with 10-fold cross validation. MLEP achieves the best result in term of feature and classifier. Best view in color.

Comparison with features learned with different regularizers. To evaluate the features learned by the prediction module, margin learning module and MLEP, we train different one-class SVMs based on different features, whose results are shown in Figure 2a. On the one hand, comparing MLEP-OCSVM with Prediction-OCSVM in Figure 2a, we can see that using same classifier, the learned features by MLEP could improve by 5.4% over prediction based method in terms of mean AUC under 10-folds case on Avenue dataset. On the other hand, by comparing MELP-OCSVM with Margin Learning-OCSVM in Figure 2a, the learned features by MELP could improve by 2.7% than margin learning based method in terms of mean AUC under 10-folds case on the Avenue dataset. It can be seen that prediction and margin learning modules are both important for anomaly detection. Therefore, combining these two modules is reasonable.

Table 1: Evaluation of different network architectures on the Avenue dataset. Larger gap means it is easier to tell anomalies from normal events.

	MCNet	UNet	Cycle GAN Generator	**Ours**
Mean AUC	88.1%	87.2%	88.0%	**92.8%**
Score on normal frames	0.726	0.771	0.810	0.799
Score on abnormal frames	0.402	0.447	0.485	0.250
Gap between normal and abnormal scores	0.324	0.324	0.325	**0.548**

Comparison of MLEPs with different classifiers. To further validate the advantages of MLEP over other classification methods, we use the features extracted with our framework to train a one-class classifier and a binary classifier for normal and abnormal frame classification. The results are shown in Figure 2b. This figure shows that the prediction module in MLEP is about 2% better than OCSVM and about 10% better than the binary classifier on the Avenue dataset. The poor performance of binary classification is probably caused by the unbalanced data distribution. The observed types of anomalies are limited in the training set, while the binary classifier cannot handle the unseen types of anomalies in testing phase.

References

[Isola *et al.*, 2017] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *CVPR*, 2017.

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015.

[Villegas *et al.*, 2017] R. Villegas, J. Yang, S. Hong, X. Lin, and H. Lee. Decomposing motion and content for natural video sequence prediction. *ICLR*, 2017.