Open-set Supervised Video Anomaly Detection with Margin Learning Embedded Prediction (Supplementary Material)

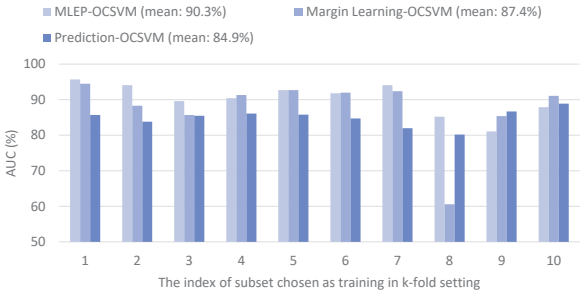
(a) GT on a normal frame (b) Pred on a normal frame (c) GT on a normal frame (d) Pred on an abnormal frame

Figure 1: Failure cases for normal and abnormal frames when imposing larger reconstruction errors for anomalies ($-\|\hat{I}^p_{t+T} - I^p_{t+T}\|_1$) in the prediction loss. GT denotes Ground-Truth and Pred denotes Prediction. We can see that such term also makes the prediction of normal frames impossible, thus reduces the performance.

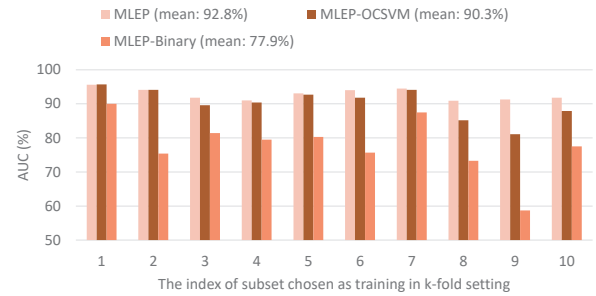
1 Evaluation of different components in MLEP

The losses in the future prediction module. Since we have anomalies in training set, it seems we should enforce large reconstruction errors for those anomalies for predictor, in other words, we also need to minimize the term $-\|\hat{I}^p_{t+T} - I^p_{t+T}\|_1$ in the prediction loss for anomalies 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%. The reason is that normal and abnormal frames share the same background in the scene, as shown in Figure 1, meanwhile the area of background is much larger than the foreground, so 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. Even though we don't enforce larger reconstruction errors for anomalies, since we enforce the normal and abnormal distances to be larger in the margin learning module, our solution also can push the anomalies far from normal data, thus also makes open-set anomalies possible and also makes the anomalies and normal data well separated.

Comparison of 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. We denote these baselines as Prediction-OCSVM, Margin Learning-OCSVM,



(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 both feature and classifier. Best view in color.

and MLEP-OCSVM, respectively. The 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-fold setting 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-fold setting on the Avenue dataset. It can be seen that both prediction and margin learning modules are important for open-set supervised anomaly detection, and their combination further improves the performance.

Comparison of MLEP 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 as well as the unknown anomalies distribution for unseen anomalies in open-set setting. The observed types of anomalies are limited in the training set, and the binary classifier cannot handle the unseen types of anomalies in testing phase in open-set supervised anomaly detection setting.

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.