# Supplementary Material for Learning to Adapt to Unseen Abnormal Activities under Weak Supervision

Jaeyoo Park*, Junha Kim*, and Bohyung Han

ECE & ASRI, Seoul National University, Korea
{bellos1203,junha.kim,bhhan}@snu.ac.kr

## 1 Supplementary

### 1.1 Multiple classes for meta-test

To verify that our algorithm works well for the meta-test task consisting of multiple abnormal subclasses, we conduct an additional experiment. We randomly split 13 subclasses of the UCF-Crime dataset into 7 for meta-train and 6 for meta-test, and vice versa (6 for meta-train and 7 for meta-test). We aggregate the results and report the average performance. We conduct the experiments for two randomly sampled subclass splits. Table. A describes the results. The results show that meta-learning performance is still better than the others, even with multiple meta-test subclasses, and the overall results are similar to Table 2. in the main paper.

### 1.2 Additional Qualitative Results

Fig. A illustrates the additional qualitative results from three methods on three test videos in the UCF-Crime and ShaghaiTech. Following Fig. 3 in the main paper, the shaded regions in the graphs correspond to the ground-truth intervals of abnormal events. The area under the ROC curve (AUROC or AUC) for each video and model is also reported in the graph. The scores given by the meta-trained model are much more discriminative than the other two methods. In (a), the scratch model does not capture the abnormalities well compared to the other two models, which take advantage of prior knowledge. Additionally, we observed that the pretrained and meta-trained models are activated by the frames containing the logo with the background colored in black, while the models still distinguish well those frames from the ones containing the ground truth abnormalities. In (b), the figure illustrates that the models trained from scratch or pretrained models are prone to suffer from mis-detections and/or false alarms while the meta-trained models maintain a better balance between positive and negative scores. We attached four sample videos, including the main paper's videos, with the three methods' scores.

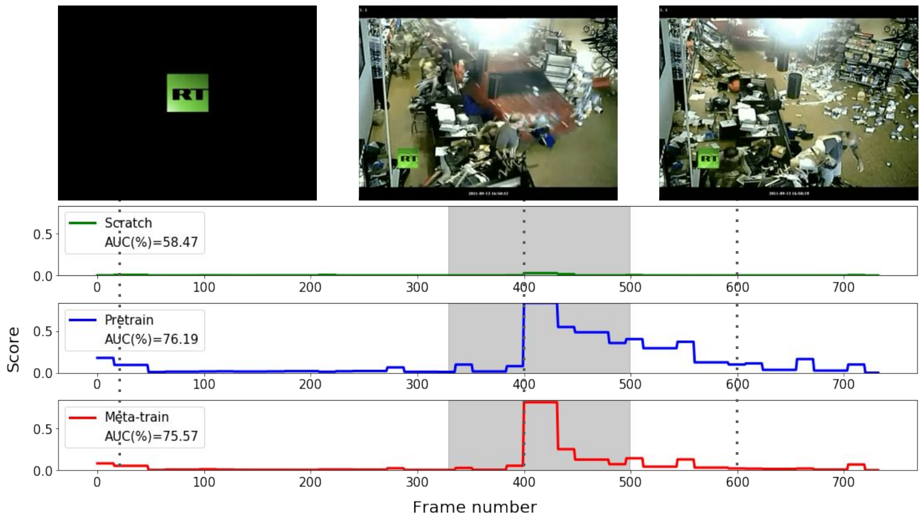
* These authors contributed equally.

Table A: Quantitative results where meta-test task contains multiple subclasses. The different split number denotes the different split of 7 and 6 subclasses while, in Table 2. of the main paper, it indicates different split of data samples. In this experiment, we use the first data sample split to construct the support set and query set of the meta-test.

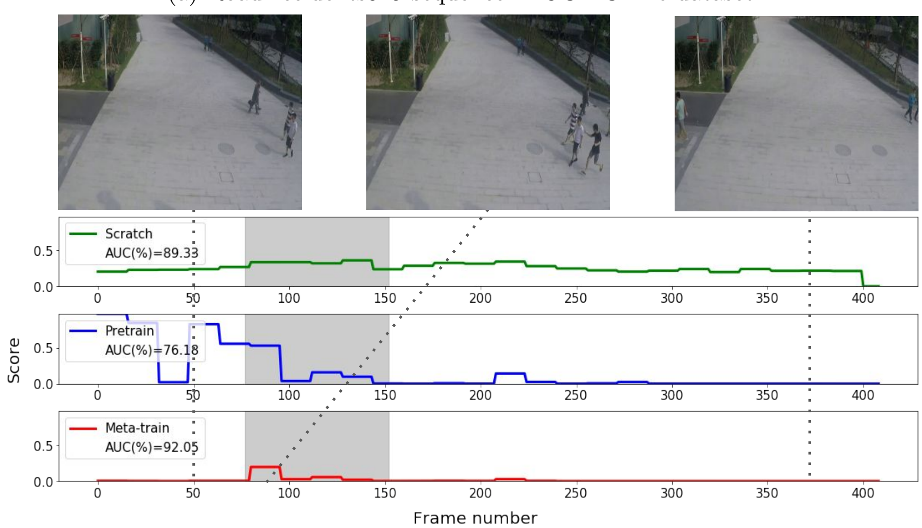
| Split | Algorithm | AUC (%) |
|---|---|---|
| 1 | S | 71.30 |
| | P | 70.75 |
| | $M_S$ | 72.69 |
| 2 | S | 70.32 |
| | P | 70.36 |
| | $M_S$ | 72.90 |

### 1.3 Fine-tuning Curve Results for All Classes

We attach additional fine-tuning graphs for each scenario in this supplementary material, including the graphs for the subclasses omitted in the main paper due to space. The results in Fig. B show the final performance curves on the test dataset during the fine-tuning phase. Although there are some cases where the meta-trained model is not the best, the performance of meta-trained cases shows better than or comparable to that of the other cases, i.e., the scratch and pre-trained model. In failure cases like (c) or (j), the margin between the best and meta-trained models is quite negligible, given the margin of successful cases. In addition, the performance of the meta-trained model is still better than the scratch model, even in the failure cases. Please refer to Section 4.6 of the main paper for the details.

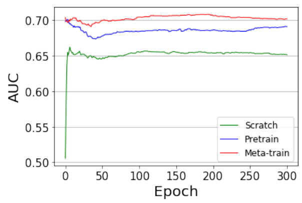



(a) RoadAccidents049 sequence in UCF-Crime dataset

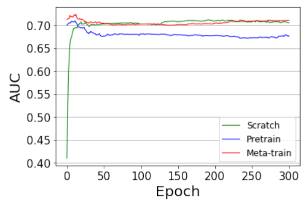



(b) 01_0027 sequence in ShanghaiTech dataset

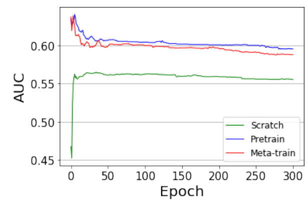
Fig. A: Qualitative results from UCF-Crime and ShanghaiTech datasets. The scores of three different methods are presented together with the ground-truth represented by the shaded regions correspond to the ground-truth.



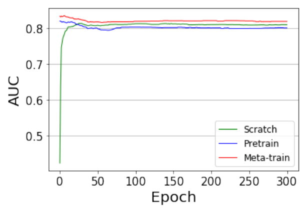

(a) Abuse



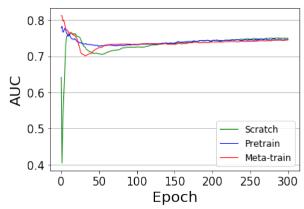

(b) Arrest



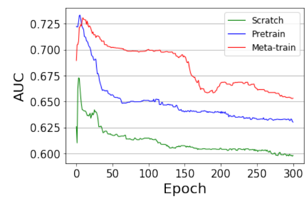

(c) Arson



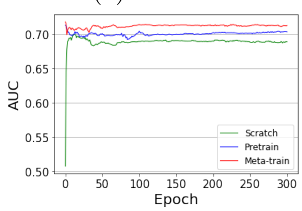

(d) Assault



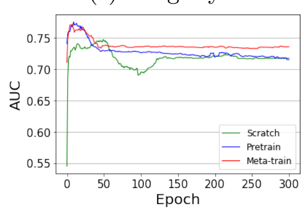

(e) Burglary



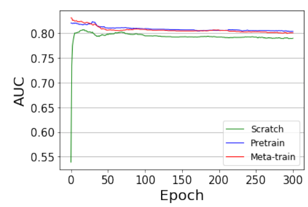

(f) Explosion



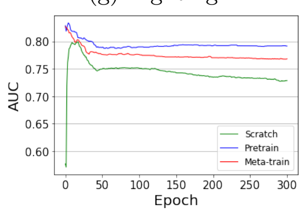

(g) Fighting



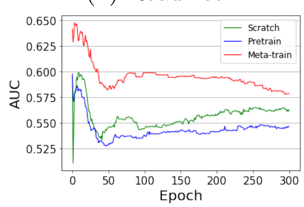

(h) RoadAcc



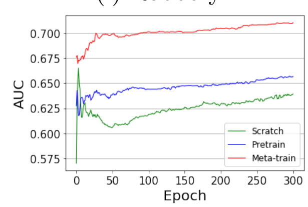

(i) Robbery



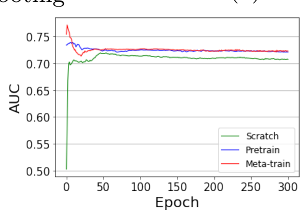

(j) Shooting



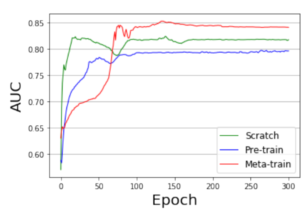

(k) Shoplifting




(l) Stealing




(m) Vandalism




(n) S-Tech

Fig. B: Comparison of subclass-wise fine-tuning curve for each scenario