Self-Supervised Predictive Convolutional Attentive Block for Anomaly Detection – Supplementary

Nicolae-Cătălin Ristea[1,2], Neelu Madan[3], Radu Tudor Ionescu[4,5,*], Kamal Nasrollahi[3,6], Fahad Shahbaz Khan[2,7], Thomas B. Moeslund[3], Mubarak Shah[8]
[1]University Politehnica of Bucharest, Romania, [2]MBZ University of Artificial Intelligence, UAE, [3]Aalborg University, Denmark, [4]University of Bucharest, Romania, [5]SecurifAI, Romania, [6]Milestone Systems, Denmark, [7]Linköping University, Sweden, [8]University of Central Florida, US

	Location of SSPCAB			AUC		RBDC	TBDC
	Early	Middle	Late	Micro	Macro		
Plain auto-encoder				80.0	83.4	49.98	51.69
	✓			81.1	83.6	50.86	52.44
		✓		84.2	85.0	52.73	54.02
			✓	85.9	85.6	53.81	56.33
	✓	✓		82.7	83.8	50.54	52.70
	✓		✓	83.2	84.1	52.33	53.01
		✓	✓	**86.1**	**85.7**	**54.03**	56.07
	✓	✓	✓	85.3	85.4	53.11	**56.64**

Table 1. Micro-averaged frame-level AUC, macro-averaged frame-level AUC, RBDC, and TBDC scores (in %) on Avenue, while integrating SSPCAB into an auto-encoder, at different locations. SSPCAB improves the results regardless of the integration place or the number of blocks. The option highlighted in red is used throughout the experiments presented in the main article. Best results are highlighted in bold.

Size of M	AUC		RBDC	TBDC
	Micro	Macro		
	80.0	83.4	49.98	51.69
1×1	85.9	85.6	53.81	56.33
3×3	85.9	85.5	53.93	56.31

Table 2. Micro-averaged frame-level AUC, macro-averaged frame-level AUC, RBDC, and TBDC scores (in %) on Avenue, while varying the size of the masked kernel M.

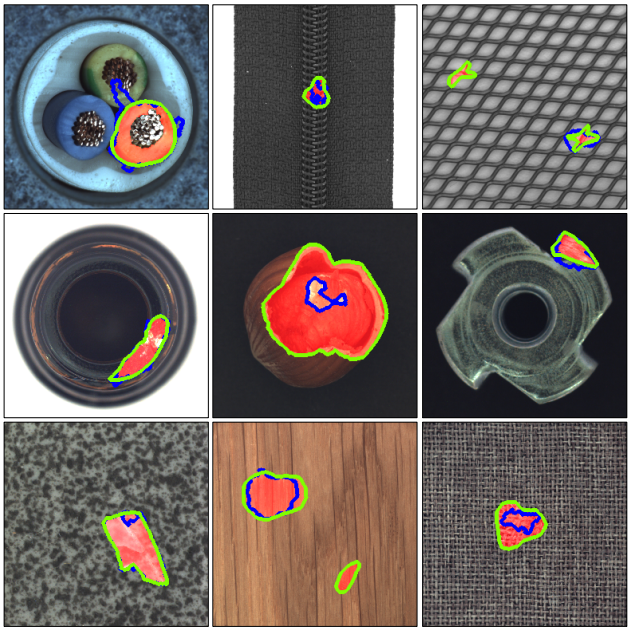

Figure 1. Additional anomaly localization examples of DRAEM [16] (blue) versus DRAEM+SSPCAB (green) on MVTec AD. The ground-truth anomalies are marked with a red mask. Best viewed in color.

1. Ablation Study

In the main article, we mention that we generally replace the penultimate convolutional layer with SSPCAB in underlying models [2, 5, 7, 8, 11, 16]. Ideally, for optimal performance gains, the integration place and the number of SSPCAB modules should be tuned on a validation set for each framework. However, anomaly detection data sets do not have a validation set and there is no way to obtain one from the training set, as the training contains only normal examples. In this context, to fairly demonstrate the generality and utility of SSPCAB, we only used a single configuration (one block, closer to the output) across all existing frameworks. However, adding more modules could be beneficial. To test various configurations, we perform an ablation study on the number of SSPCAB modules and the places where these modules can be integrated in a plain auto-encoder. In Table 1, we present the corresponding experiments on the Avenue data set. *We observe that SSPCAB improves the results, regardless of the place of integration or the number of blocks.* The improvements seem larger when SSPCAB is

*corresponding author: raducu.ionescu@gmail.com

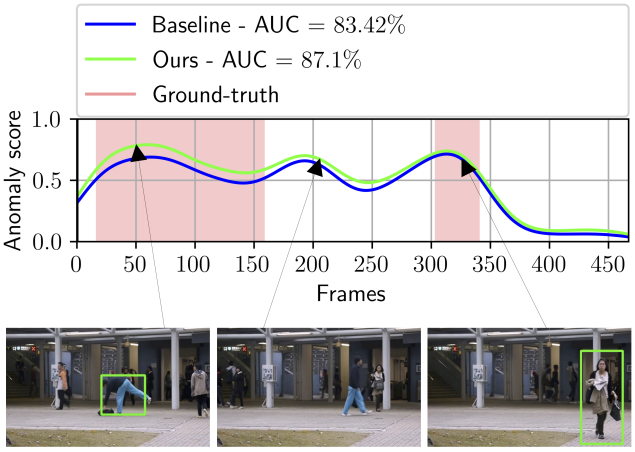

Figure 2. Frame-level anomaly scores for Liu *et al*. [7] before (baseline) and after (ours) integrating SSPCAB, for test video 10 from Avenue. Anomaly localization results correspond to the model based on SSPCAB. Best viewed in color.

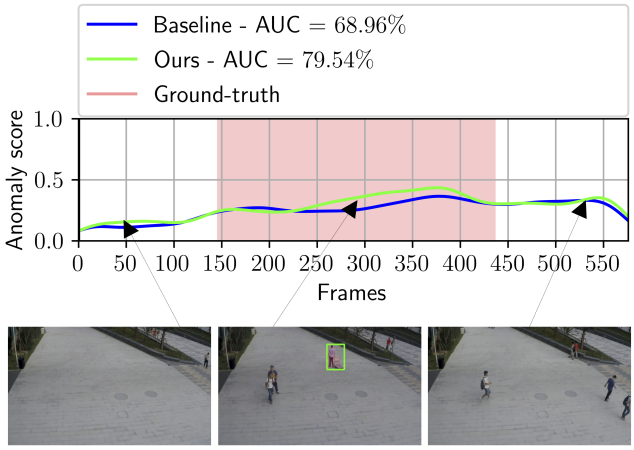

Figure 3. Frame-level anomaly scores for Georgescu *et al*. [2] before (baseline) and after (ours) integrating SSPCAB, for test video 01_0054 from ShanghaiTech. Anomaly localization results correspond to the model based on SSPCAB. Best viewed in color.

integrated closer to the output. Integrating more blocks can sometimes help.

Another hyperparameter that could be tuned is the size of the masked kernel M. In our experiments, we kept M to a size of 1×1 for simplicity and speed. To study the effect of increasing the size of M, we have tested the size of 3×3 with the plain auto-encoder on Avenue. We report the corresponding results in Table 2. When comparing the results with masked kernels of 1×1 or 3×3 components, we do not observe significant differences.

An additional aspect that can suffer multiple reconfigurations, given a validation set, is the pattern of the proposed kernel. In our experiments, we tried a simple pattern where the mask is placed in the center and the reception field is connected to the four corner sub-kernels denoted by K_i, $\forall i \in \{1, 2, 3, 4\}$. We designed this pattern while trying to extrapolate the idea from middle frame prediction (which was shown to provide somewhat better results than future frame prediction) to a 2D kernel. Of course, other patterns are possible and are likely to work equally well.

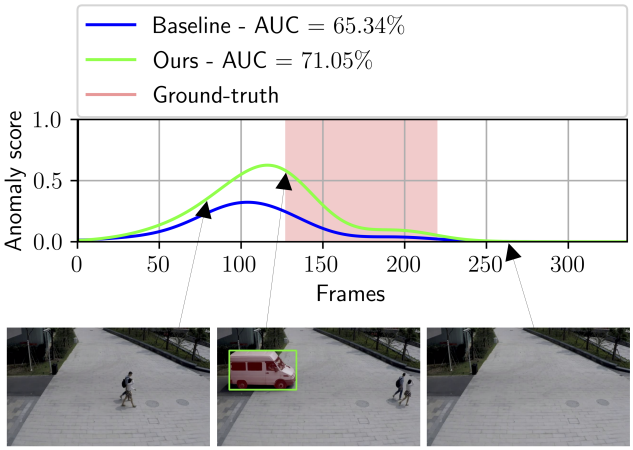

Figure 4. Frame-level anomaly scores for Georgescu *et al*. [2] before (baseline) and after (ours) integrating SSPCAB, for test video 01_0130 from ShanghaiTech. Anomaly localization results correspond to the model based on SSPCAB. Best viewed in color.

2. Qualitative Anomaly Detection Results

Anomaly detection in images. In Figure 1, we present additional qualitative results produced by DRAEM [16] on the MVTec AD benchmark. The displayed examples illustrate the benefit of integrating SSPCAB, which is much better at segmenting the anomalies compared to the baseline DRAEM. We show improvements in terms of the pixel-level annotation for both objects and textures.

Anomaly detection in videos. In Figure 2, we show a comparison of the frame-level anomaly scores on test video 10 from the Avenue data set, before and after integrating SSPCAB into the method of Liu *et al*. [7]. On this video, SSPCAB increases the AUC by nearly 4%. After introducing SSPCAB, we observe higher frame-level anomaly scores for the first abnormal event. The anomaly localization results depict *a person throwing a backpack* and *a person walking in the wrong direction*.

In Figures 3 and 4, we illustrate similar comparisons for test videos 01_0054 and 01_0130 from the ShanghaiTech data set, before and after adding SSPCAB into the frame-

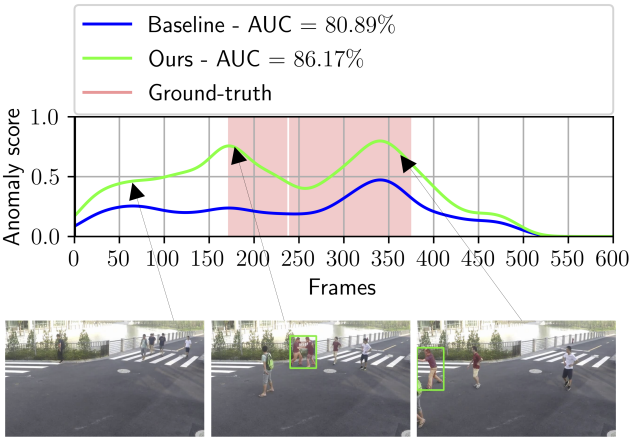

Figure 5. Frame-level anomaly scores for Georgescu *et al*. [2] before (baseline) and after (ours) integrating SSPCAB, for test video 07_0047 from ShanghaiTech. Anomaly localization results correspond to the model based on SSPCAB. Best viewed in color.

Method	Time (ms)		Relative (%)
	Baseline	+SSPCAB	
Liu *et al*. [7]	2.1	2.4	14.2
Georgescu *et al*. [2]	1.5	1.7	13.3

Table 3. Inference times (in milliseconds) and relative time expansions (in %) for two frameworks [2,7], before and after integrating SSPCAB. The running times are measured on an Nvidia GeForce GTX 3090 GPU with 24 GB of VRAM.

work of Georgescu *et al*. [2]. For test video 01_0054, SSPCAB increases the AUC by more than 10%. For test video 01_0130, the baseline framework seems to detect the abnormal event too early, but SSPCAB seems capable of shifting the detection towards the correct moment. As a result, SSPCAB increases the frame-level AUC score by almost 6%. We observe a similar AUC improvement from SSPCAB in Figure 5, where we compare the frame-level anomaly scores on test video 07_0047 from the ShanghaiTech data set. For this video, we underline that the frame-level scores are visibly more correlated to the ground-truth anomalies. Moreover, in all three ShanghaiTech videos, we observe that the approach based on SSPCAB can precisely localize and detect the abnormal events (*person pulling a lever cart*, *car inside pedestrian area*, *people fighting*, *people running*).

3. Inference Time

Regardless of the underlying framework [2,5,7,8,11,16], we add only one instance of SSPCAB, usually replacing the penultimate convolutional layer. As such, we expect the running time to increase. To assess the amount of extra time added by SSPCAB, we present the running times before and after integrating SSPCAB into two state-of-the-art frameworks [2, 7] in Table 3. The reported times show time expansions lower than 0.3 ms for both frameworks. Hence, we consider that the accuracy gains brought by SSPCAB outweigh the marginal running time expansions observed in Table 3.

4. Discussion

Although SSPCAB belongs to an existing family of anomaly detection methods, *i.e*. reconstruction-based frameworks [1, 3, 4, 6, 7, 9–15], we would like to underline that we are the first to integrate the reconstruction functionality at the block level. Unlike other reconstruction approaches, our contribution is more flexible, as it can be integrated in existing and future reconstruction methods. Moreover, SSPCAB can also be used to introduce reconstruction-based anomaly detection in other frameworks, which do not rely on reconstruction. We thus believe that our generic and effective approach will help ease future research in anomaly detection.

An important aspect that must be noted is that, due to the masked convolution, our block will not reconstruct the input exactly. Except for the degenerate case where the input is constant, this scenario should not occur in the real world, which means that the reconstruction performed by SSPCAB is not trivial. However, our foremost intuition about the usefulness of SSPCAB is different: our block provides a better reconstruction for normal convolutional features than for abnormal convolutional features. If the features representing normal versus abnormal examples are different at any layer of a neural architecture, it should result in greater differences at the final output of the architecture. This idea is also supported by the experiments presented in Table 1.

Further looking at the results shown in Table 1, we observe that SSPCAB does not bring significant gains when the block is placed near the input. We aim to further investigate this limitation in future work. Aside from this small issue, we did not observe other limitations of SSPCAB during our experiments.

References

[1] Ye Fei, Chaoqin Huang, Cao Jinkun, Maosen Li, Ya Zhang, and Cewu Lu. Attribute Restoration Framework for Anomaly Detection. *IEEE Transactions on Multimedia*, pages 1–1, 2020. 3

[2] Mariana Iuliana Georgescu, Radu Ionescu, Fahad Shahbaz Khan, Marius Popescu, and Mubarak Shah. A Background-Agnostic Framework with Adversarial Training for Abnormal Event Detection in Video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 1, 2, 3

[3] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton Van

Den Hengel. Memorizing Normality to Detect Anomaly: Memory-Augmented Deep Autoencoder for Unsupervised Anomaly Detection. In *Proceedings of ICCV*, pages 1705–1714, 2019. 3

[4] Mahmudul Hasan, Jonghyun Choi, Jan Neumann, Amit K. Roy-Chowdhury, and Larry S. Davis. Learning temporal regularity in video sequences. In *Proceedings of CVPR*, pages 733–742, 2016. 3

[5] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. CutPaste: Self-Supervised Learning for Anomaly Detection and Localization. In *Proceedings of CVPR*, pages 9664–9674, 2021. 1, 3

[6] Zhenyu Li, Ning Li, Kaitao Jiang, Zhiheng Ma, Xing Wei, Xiaopeng Hong, and Yihong Gong. Superpixel Masking and Inpainting for Self-Supervised Anomaly Detection. In *Proceedings of BMVC*, 2020. 3

[7] Wen Liu, Weixin Luo, Dongze Lian, and Shenghua Gao. Future Frame Prediction for Anomaly Detection – A New Baseline. In *Proceedings of CVPR*, pages 6536–6545, 2018. 1, 2, 3

[8] Zhian Liu, Yongwei Nie, Chengjiang Long, Qing Zhang, and Guiqing Li. A Hybrid Video Anomaly Detection Framework via Memory-Augmented Flow Reconstruction and Flow-Guided Frame Prediction. In *Proceedings of ICCV*, pages 13588–13597, 2021. 1, 3

[9] Weixin Luo, Wen Liu, and Shenghua Gao. A Revisit of Sparse Coding Based Anomaly Detection in Stacked RNN Framework. In *Proceedings of ICCV*, pages 341–349, 2017. 3

[10] Trong-Nguyen Nguyen and Jean Meunier. Anomaly Detection in Video Sequence With Appearance-Motion Correspondence. In *Proceedings of ICCV*, pages 1273–1283, 2019. 3

[11] Hyunjong Park, Jongyoun Noh, and Bumsub Ham. Learning Memory-guided Normality for Anomaly Detection. In *Proceedings of CVPR*, pages 14372–14381, 2020. 1, 3

[12] Mahdyar Ravanbakhsh, Moin Nabi, Enver Sangineto, Lucio Marcenaro, Carlo Regazzoni, and Nicu Sebe. Abnormal Event Detection in Videos using Generative Adversarial Nets. In *Proceedings of ICIP*, pages 1577–1581, 2017. 3

[13] Mohammadreza Salehi, Niousha Sadjadi, Soroosh Baselizadeh, Mohammad H. Rohban, and Hamid R. Rabiee. Multiresolution Knowledge Distillation for Anomaly Detection. In *Proceedings of CVPR*, pages 14902–14912, 2021. 3

[14] Yao Tang, Lin Zhao, Shanshan Zhang, Chen Gong, Guangyu Li, and Jian Yang. Integrating prediction and reconstruction for anomaly detection. *Pattern Recognition Letters*, 129:123–130, 2020. 3

[15] Shashanka Venkataramanan, Kuan-Chuan Peng, Rajat Vikram Singh, and Abhijit Mahalanobis. Attention guided anomaly localization in images. In *Proceedings of ECCV*, pages 485–503, 2020. 3

[16] Vitjan Zavrtanik, Matej Kristan, and Danijel Skocaj. DRAEM – A Discriminatively Trained Reconstruction Embedding for Surface Anomaly Detection. In *Proceedings of ICCV*, pages 8330–8339, 2021. 1, 2, 3