

表3. MVTEcにおけるサブ画像異常検出精度 (PRO %)

	Student	1-NN	OC-SVM	$\ell_2$ -AE	VAE	SSIM-AE	CNN-Dict	SPADE
Carpet	69.5	51.2	35.5	45.6	50.1	64.7	46.9	94.7
Grid	81.9	22.8	12.5	58.2	22.4	84.9	18.3	86.7
Leather	81.9	44.6	30.6	81.9	63.5	56.1	64.1	97.2
Tile	91.2	82.2	72.2	89.7	87.0	17.5	79.7	75.9
Wood	72.5	50.2	33.6	72.7	62.8	60.5	62.1	87.4
Bottle	91.8	89.8	85.0	91.0	89.7	83.4	74.2	95.5
Cable	86.5	80.6	43.1	82.5	65.4	47.8	55.8	90.9
Capsule	91.6	63.1	55.4	86.2	52.6	86.0	30.6	93.7
Hazelnut	93.7	86.1	61.6	91.7	87.8	91.6	84.4	95.4
Metal nut	89.5	70.5	31.9	83.0	57.6	60.3	35.8	94.4
Pill	93.5	72.5	54.4	89.3	76.9	83.0	46.0	94.6
Screw	92.8	60.4	64.4	75.4	55.9	88.7	27.7	96.0
Toothbrush	86.3	67.5	53.8	82.2	69.3	78.4	15.1	93.5
Transistor	70.1	68.0	49.6	72.8	62.6	72.5	62.8	87.4
Zipper	93.3	51.2	35.5	83.9	54.9	66.5	70.3	92.6
Average	85.7	64	47.9	79	63.9	69.4	51.5	<b>91.7</b>

表4. STCにおける画像レベル異常検出精度 (平均ROCAUC%)

TSC <a href="#">[23]</a>	StackRNN <a href="#">[23]</a>	AE-Conv3D <a href="#">[35]</a>	MemAE <a href="#">[12]</a>	AE(2D) <a href="#">[16]</a>	ITAE <a href="#">[19]</a>	SPADE
67.9	68.0	69.7	71.2	60.9	<b>72.5</b>	71.9

表3では、当社の手法をPROの観点から比較しています。上述の通り、これはピクセル単位の精度指標であり、少数のピクセルに及ぶ異常に対してより大きな重みを付与するものです。当社の手法は、Bregmannら [6] の事前学習済み特徴量に基づくオートエンコーダー手法および彼らの論文で提示されたベースライン手法と比較されています。当社の手法は、すべての従来手法よりも著しく優れた結果を達成しています。Bregmannらもより良い結果を示すアンサンブルアプローチを提案しています。当社の手法はアンサンブルを使用していませんが（おそらく当社の手法も改善されるでしょう）、アンサンブルアプローチよりも優れた性能を示しています。図1では、当社の手法が異常領域の正確なマスクを回復できることを示す定性的結果をさらに提示しています。

4.2 上海科技大学キャンパスデータセット

私たちは、上海科技大学キャンパスデータセットで私たちの手法を評価しました。このデータセットは監視環境を模擬しており、入力監視カメラで撮影された混雑したキャンパスを観察する動画から構成されています。データセットには12のシーンが含まれており、各シーンにはトレーニング動画と少数のテスト画像が含まれています。トレーニング動画には異常は含まれておらず、テスト動画には正常な画像と異常な画像が含まれています。異常は、歩行者が非標準的な行動を実行するものと定義されています。

**Table 3.** Sub-Image anomaly detection accuracy on MVTec (PRO %)

	Student	1-NN	OC-SVM	$\ell_2$ -AE	VAE	SSIM-AE	CNN-Dict	SPADE
Carpets	69.5	51.2	35.5	45.6	50.1	64.7	46.9	94.7
Grid	81.9	22.8	12.5	58.2	22.4	84.9	18.3	86.7
Leather	81.9	44.6	30.6	81.9	63.5	56.1	64.1	97.2
Tile	91.2	82.2	72.2	89.7	87.0	17.5	79.7	75.9
Wood	72.5	50.2	33.6	72.7	62.8	60.5	62.1	87.4
Bottle	91.8	89.8	85.0	91.0	89.7	83.4	74.2	95.5
Cable	86.5	80.6	43.1	82.5	65.4	47.8	55.8	90.9
Capsule	91.6	63.1	55.4	86.2	52.6	86.0	30.6	93.7
Hazelnut	93.7	86.1	61.6	91.7	87.8	91.6	84.4	95.4
Metal nut	89.5	70.5	31.9	83.0	57.6	60.3	35.8	94.4
Pill	93.5	72.5	54.4	89.3	76.9	83.0	46.0	94.6
Screw	92.8	60.4	64.4	75.4	55.9	88.7	27.7	96.0
Toothbrush	86.3	67.5	53.8	82.2	69.3	78.4	15.1	93.5
Transistor	70.1	68.0	49.6	72.8	62.6	72.5	62.8	87.4
Zipper	93.3	51.2	35.5	83.9	54.9	66.5	70.3	92.6
Average	85.7	64	47.9	79	63.9	69.4	51.5	<b>91.7</b>

**Table 4.** Image-level anomaly detection accuracy on STC (Average ROCAUC %)

TSC [23]	StackRNN [23]	AE-Conv3D [35]	MemAE [12]	AE(2D) [16]	ITAE [19]	SPADE
67.9	68.0	69.7	71.2	60.9	<b>72.5</b>	71.9

In Tab. 3, we compare our method in terms of PRO. As explained above, this is another per-pixel accuracy measure which gives larger weight to anomalies which cover few pixels. Our method is compared with the auto-encoder with pre-trained features based approach of Bregmann et al. [6] and the baselines presented in their paper. Our approach achieves significantly better results than all previous methods. We note that Bregmann et al also presented an ensemble approach with better results. While our method does not use ensembles (which will probably improve our method too), we outperform the ensemble approach as well. We present more qualitative results of our method in Fig. 1 that show that our method is able to recover accurate masks of the anomalous regions.

## 4.2 Shanghai Tech Campus Dataset

We evaluate our method on the Shanghai Tech Campus dataset. It simulates a surveillance setting, where the input consists of videos captured by surveillance cameras observing a busy campus. The dataset contains 12 scenes, each scene consists of training videos and a smaller number of test images. The training videos do not contain anomalies while the test videos contain normal and anomalous images. Anomalies are defined as pedestrians performing non-standard ac-