



図2. (左から右) i) 異常画像 ii) 取得された最良の正常近傍画像 iii) SPADAで検出されたマスク iv) 予測された異常画像ピクセル。この例では、SPADAが最良の近傍画像との対応関係を検出することで異常画像領域を特定していることがわかります。異常部分は正常画像に対応関係がなく、そのため検出されました。

表5. STCにおけるピクセルレベル異常検出精度（平均ROCAUC%）

$AE_{L2}$	$AE_{SSIM}$	CAVGA- $R_u$	<span style="border: 1px solid green; padding: 0 2px;">31</span>	SPADA
74	76	85		<b>89.9</b>

活動（例：戦闘）および歩行者以外の移動物体（例：モーターバイク）を検出します。

まず、画像レベルのアノマリー検出における最初の段階を、他の最先端手法と比較評価しました。表4では、当社の最初の段階が最良の手法 [19] とほぼ同等の性能を有することを示しています。さらに興味深いのは、表5でピクセルレベルROCAUC性能を、最良の報告手法であるCAVGA- $R_u$  [31] と比較している点です。私たちの方法は、最良の報告方法に対して著しい差で優れています。注意すべき点は、私たちは異常監視を使用していないため、異常監視を使用しない最良の方法と比較した点です。実際、異常監視は実践ではしばしば利用できないためです。

4.3 アブレーション研究

私たちの手法の相対的な性能を評価するため、アブレーション研究を実施しました。表6では、特徴ピラミッドの異なるレベルを使用した場合の比較を行っています。高解像度（56×56）のアクティベーションを使用すると、文脈の制限により性能が大幅に低下することを確認しました。一方、



**Fig. 2.** (left to right) i) An anomalous image ii) The retrieved top normal neighbor image iii) The mask detected by SPADA iv) The predicted anomalous image pixels. We can see how in this example, SPADA detects the anomalous image region by finding the correspondence with the nearest-neighbor image. The anomalous parts did not have correspondences in the normal image and were therefore detected.

**Table 5.** Pixel-level anomaly detection accuracy on STC (Average ROCAUC %)

$AE_{L2}$	$AE_{SSIM}$	CAVGA- $R_u$ [31]	SPADA
74	76	85	<b>89.9</b>

tivities (e.g. fighting) as well as any moving object which is not a pedestrian (e.g. motorbikes).

We began by evaluating our first stage for detecting image-level anomalies against other state-of-the-art methods. We show in Tab. 4 that our first stage has comparable performance to the top performing method [19]. More interestingly, we compare in Tab. 5 the pixel-level ROCAUC performance with the best reported method, CAVGA- $R_u$  [31]. Our method significantly outperforms the best reported method by a significant margin. Note that we compared to the best method that did not use anomaly supervision, as we do not use it and as anomaly supervision is often not available in practice.

### 4.3 Ablation Study

We conduct an ablation study on our method in order to understand the relative performance of its different parts. In Tab. 6, we compare using different level of the feature pyramid. We experienced that using activations of too high resolution ( $56 \times 56$ ) significantly hurts performance (due to limited context) while using