Table 1. Image-level anomaly detection accuracy on MVTec (Average ROCAUC %)

	Geom [11]	GANomaly [1]	AE_{L2}	ITAE [19]	SPADE
Average	67.2	76.2	75.4	83.9	**85.5**

Table 2. Sub-Image anomaly detection accuracy on MVTec (ROCAUC %)

	AE_{SSIM}	AE_{L2}	AnoGAN	CNN Dict	TI	VM	CAVGA-R_u	SPADE
Carpet	87	59	54	72	88	-	-	97.5
Grid	94	90	58	59	72	-	-	93.7
Leather	78	75	64	87	97	-	-	97.6
Tile	59	51	50	93	41	-	-	87.4
Wood	73	73	62	91	78	-	-	88.5
Bottle	93	86	86	78	-	82	-	98.4
Cable	82	86	78	79	-	-	-	97.2
Capsule	94	88	84	84	-	76	-	99.0
Hazelnut	97	95	87	72	-	-	-	99.1
Metal nut	89	86	76	82	-	60	-	98.1
Pill	91	85	87	68	-	83	-	96.5
Screw	96	96	80	87	-	94	-	98.9
Toothbrush	92	93	90	77		68	-	97.9
Transistor	90	86	80	66	-	-	-	94.1
Zipper	88	77	78	76	-	-	-	96.5
Average	87	82	74	78	75	77	89	**96.0**

lies. We evaluate our method using two established metrics. The first is per-pixel ROCAUC. This metric is calculated by scoring each pixel by the distance to its K nearest correspondences. By scanning over the range of thresholds, we can compute the pixel-level ROCAUC curve. The anomalous category is designated as positive. It was noted by several previous works that ROCAUC is biased in favor of large anomalies. In order to reduce this bias, Bergmann et al [6] propose the PRO (per-region overlap) curve metric. They first separate anomaly masks into their connected components, therefore dividing them into individual anomaly regions. By changing the detection threshold, they scan over false positive rates (FPR), for each FPR they compute PRO i.e. the proportion of the pixels of each region that are detected as anomalous. The PRO score at this FPR is the average coverage across all regions. The PRO curve metric computes the integral across FPR rates from 0 to 0.3. The PRO score is the normalized value of this integral.

In Tab. 2, we compare our methods on the per-pixel ROCAUC metric against state-of-the-art results reported by Bergmann et al. [5] as well as newer results by Venkataramanan et al. [31]. Most of the methods use different varieties of autoencoders, including the top-performer CAVGA-R_u. Our method significantly outperforms all methods. This attest to the strength of our pyramid based correspondence approach.