TABLE III
COMPARISON OF OUR PADIM MODELS WITH THE STATE-OF-THE-ART FOR THE ANOMALY LOCALIZATION ON THE MVTEC AD. RESULTS ARE DISPLAYED AS TUPLES (AUROC%, PRO-SCORE%)

Type	Reconstruction-based methods			Embedding similarity based methods			Our methods	
Model	AE simm [1], [2], [9]	AE L2 [1], [2]	VAE	Student [2]	Patch SVDD [4]	SPADE [5]	PaDiM-R18-Rd100	PaDiM-WR50-Rd550
Carpet	(87, 64.7)	(59, 45.6)	(59.7, 61.9)	(-, 69.5)	(92.6, -)	(97.5, 94.7)	(98.9, 96.0)	(**99.1**, **96.2**)
Grid	(94, 84.9)	(90, 58.2)	(61.2, 40.8)	(-, 81.9)	(96.2, -)	(93.7, 86.7)	(94.9, 90.9)	(**97.3**, **94.6**)
Leather	(78, 56.1)	(75, 81.9)	(67.1, 64.9)	(-, 81.9)	(97.4, -)	(97.6, 97.2)	(99.1, 97.9)	(**99.2**, **97.8**)
Tile	(59, 17.5)	(51, 89.7)	(51.3, 24.2)	(-, 91.2)	(91.4, -)	(87.4, 75.9)	(91.2, 81.6)	(**94.1**, **86.0**)
Wood	(73, 60.5)	(73, 72.7)	(66.6, 57.8)	(-, 72.5)	(90.8, -)	(88.5, 87.4)	(93.6, 90.3)	(**94.9**, **91.1**)
All texture classes	(78, 56.7)	(70, 69.6)	(61.2, 49.9)	(-, 79.4)	(93.7, -)	(92.9, 88.4)	(95.6, 91.3)	(**96.9**, **93.2**)
Bottle	(93, 83.4)	(86, 91.0)	(83.1, 70.5)	(-, 91.8)	(98.1, -)	(**98.4**, **95.5**)	(98.1, 93.9)	(98.3, 94.8)
Cable	(82, 47.8)	(86, 82.5)	(83.1, 77.9)	(-, 86.5)	(96.8, -)	(**97.2**, **90.9**)	(95.8, 86.2)	(96.7, 88.8)
Capsule	(94, 86.0)	(88, 86.2)	(81.7, 77.9)	(-, 91.6)	(95.8, -)	(**99.0** ,**93.7**)	(98.3, 91.9)	(98.5, 93.5)
Hazelnut	(97, 91.6)	(95, 91.7)	(87.7, 77.0)	(-, 93.7)	(97.5, -)	(**99.1**, **95.4**)	(97.7, 91.4)	(98.2, 92.6)
Metal Nut	(89, 60.3)	(86, 83.0)	(78.7, 57.6)	(-, 89.5)	(98.0, -)	(**98.1**, **94.4**)	(96.7, 81.9)	(97.2, 85.6)
Pill	(91, 83.0)	(85, 89.3)	(81.3, 79.3)	(-, 93.5)	(95.1, -)	(**96.5**, **94.6**)	(94.7, 90.6)	(95.7, 92.7)
Screw	(96, 88.7)	(96, 75.4)	(75.3, 66.4)	(-, 92.8)	(95.7, -)	(**98.9**, **96.0**)	(97.4, 91.3)	(98.5, 94.4)
Toothbrush	(92, 78.4)	(93, 82.2)	(91.9, 85.4)	(-, 86.3)	(98.1, -)	(97.9, **93.5**)	(98.7, 92.3)	(**98.8**, 93.1)
Transistor	(90, 72.5)	(86, 72.8)	(75.4, 61.0)	(-, 70.1)	(97.0, -)	(94.1, **87.4**)	(97.2, 80.2)	(**97.5**, 84.5)
Zipper	(88, 66.5)	(77, 83.9)	(71.6, 60.8)	(-, 93.3)	(95.1, -)	(96.5, 92.6)	(98.2, 94.7)	(**98.5**, **95.9**)
All object classes	(91, 75.8)	(88, 83.8)	(81.0, 71.4)	(-, 88.9)	(96.7, -)	(97.6, **93.4**)	(97.3, 89.4)	(**97.8**, 91.6)
All classes	(87, 69.4)	(82, 79.0)	(74.4, 64.2)	(-, 85.7)	(95.7, -)	(96.5, 91.7)	(96.7, 90.1)	(**97.5**, **92.1**)

It can also be noted from Table II that randomly reducing the embedding vector size to only 100 dimensions has a very little impact on the anomaly localization performance. The results drop only by 0.4p.p in the AUROC and 0.3p.p in the PRO-score. This simple yet effective dimensionality reduction method significantly reduces PaDiM time and space complexity as it will be shown in Section V-D.

B. Comparison with the state-of-the-art

Localization. In Table III, we show the AUROC and the PRO-score results for anomaly localization on the MVTec AD. For a fair comparison, we used a Wide ResNet-50-2 (WR50) as this backbone is used in SPADE [5]. Since the other baselines have smaller backbones, we also try a ResNet18 (R18). We randomly reduce the embedding size to 550 and 100 for PaDiM with WR50 and R18 respectively.

We first notice that PaDiM-WR50-Rd550 outperforms all the other methods in both the PRO-score and the AUROC on average for all the classes. PaDiM-R18-Rd100 which is a very light model also outperforms all models in the average AUROC on the MVTec AD classes by at least 0.2p.p. When we further analyze the PaDiM performances, we see that the gap for the object classes is small as PaDiM-WR50-Rd550 is the best only in the AUROC (+0.2p.p) but SPADE [5] is the best in the PRO-score (+1.8p.p). However, our models are particularly accurate on texture classes. PaDiM-WR50-Rd550 outperforms the second best model SPADE [5] by 4.8p.p and 4.0p.p in the PRO-score and the AUROC respectively on average on texture classes. Indeed, PaDiM learns an explicit probabilistic model of the normal classes contrary to SPADE [5] or Patch-SVDD [4]. It is particularly efficient on texture images because even if they are not aligned and centered like object images, PaDiM effectively captures their statistical similarity accross the normal train dataset.

Additionally, we evaluate our model on the STC dataset. We compare our method to the two best reported models performing anomaly localization without temporal information, CAVGA-RU [3] and SPADE [5]. As shown in Table IV, the best result (AUROC) on the STC dataset is achieved with our simplest model PaDiM-R18-Rd100 by a 2.1p.p. margin. In fact, pedestrian positions in images are highly variable in this dataset and, as shown in Section V-C, our method performs well on non-aligned datasets.

Detection. By taking the maximum score of the anomaly maps issued by our models (see Section III-C) we give anomaly scores to entire images to perform anomaly detection at the image level. We test PaDiM for anomaly detection with a Wide ResNet-50-2 (WR50) [28] used in SPADE and an EfficientNet-B5 [29]. The Table V shows that our model PaDiM-WR50-Rd550 outperforms every method except MahalanobisAD [23] with their best reported backbone, an EfficientNet-B4. Still our PaDiM-EfficientNet-B5 outperforms every model by at least 2.6p.p on average on all the classes in the AUROC. Besides, contrary to the second best method for anomaly detection, MahalanobisAD [23], our model also performs anomaly segmentation which characterizes more precisely the anomalous areas in the images.

C. Anomaly localization on a non-aligned dataset

To estimate the robustness of anomaly localization methods, we train and evaluate the performance of PaDiM and several

TABLE IV
COMPARISON OF OUR PADIM MODEL WITH THE STATE-OF-THE-ART FOR THE ANOMALY LOCALIZATION ON THE STC IN THE AUROC%.

Model	CAVGA-RU [3]	SPADE [5]	PaDiM-R18-Rd100
AUROC score%	85	89.9	**91.2**