

図5：織物テクスチャの2つのデータセットにおける ℓ^2 -オートエンコーダーとSSIMオートエンコーダーの再構築、残差マップ、セグメンテーション結果の定性的比較。欠陥を含む真の領域は赤で囲まれ、緑の領域は各手法のセグメンテーション結果を示しています。

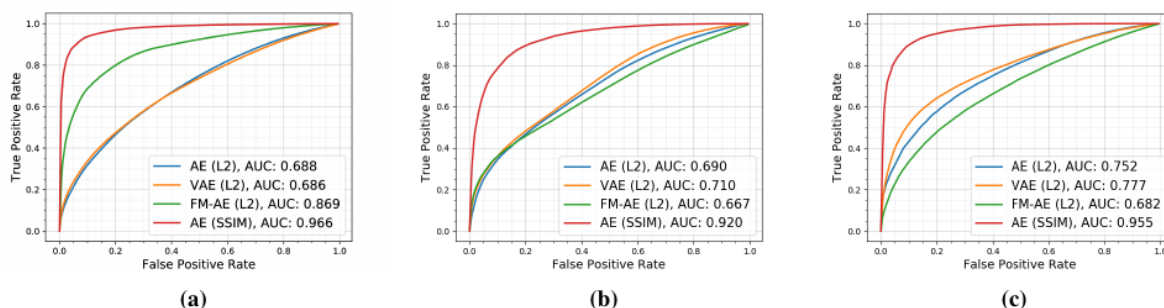


図6：提案するSSIMオートエンコーダー（赤線）の評価対象データセット（a）ナノファイバー材料と（b）、（c）2つのテクスチャデータセットにおけるROC曲線。 픽셀単位の損失関数を使用する他のオートエンコーダーアーキテクチャ（緑、オレンジ、青線）との比較。対応するAUC値は凡例に示されています。

4.3. Results

図5は、2つのテクスチャデータセットの画像における ℓ^2 -オートエンコーダーとSSIMオートエンコーダーの性能の定性的比較を示しています。両方のアーキテクチャは再構築における欠陥を削除しますが、SSIM残差マップのみが欠陥を明らかにし、正確なセグメンテーション結果を提供します。NanoTWICEデータセットにおいても同様の現象が観察され、図1に示されています。

数値結果により、この定性的挙動を確認します。図6は、SSIMを使用した当手法のROC曲線とそのAUC値を、 픽셀単位のアーキテクチャと比較しています。後者の性能は、各 픽셀をランダムに分類するのとはほぼ変わらない場合が多いです。VAEの場合、後方分布から異なる潜在サンプルで得られた再構築結果は大きく異なることはありません。したがって、決定論的フレームワークを改善できませんでした。特徴マッチングを採用しても、ナノファイバー材料のデータセットでのセグメンテーション結果のみが改善され、2つのテクスチャデータセットでは利益をもたらしませんでした。SSIMを損失関数および評価指標として使用すると、他のすべてのテスト済みアーキテクチャを大幅に上回ります。損失関数を変更するだけで、ナノファイバー材料のデータセットにおける達成されたAUCは0.688から0.966に改善され、Napoleto et al. (2018) の最新の研究結果（最大0.974の値が報告されている）とほぼ同等の性能を示しました。

この方法とは対照的に、オートエンコーダーは手動で設計された特徴量や事前学習済みネットワークなどのモデル事前知識に依存しません。2つのテクスチャデータセットにおいても、同様の性能向上が観察されます。

ナノファイバー材料のデータセットにはさまざまなサイズの欠陥が含まれており、すべての 픽셀を等しく加重した場合、小さな欠陥は全体の真陽性率への寄与が小さいため、このデータセットにおいて検出された異常領域と真の真実値の重なりをさらに評価し、 $p \in \{25\%, 50\%, 75\%\}$ におけるp-quantilesを図7に示します。偽陽性率が5%まで低下しても、欠陥の50%以上が真の値との重なりが91%を超える結果となりました。これは、Napoleto et al. (2018) が本設定で報告した最小重なり85%を上回る結果です。

さらに、SSIMオートエンコーダーの異なるハイパーパラメーター設定に対する感度をテストしました。潜在空間次元 d 、SSIMウィンドウサイズ k 、およびオートエンコーダーが訓練されたパッチのサイズを変化させました。表2は、潜在空間次元が十分に大きい場合、SSIMは異なるハイパーパラメーター設定に対して感度が低いことを示しています。 $d = 500$ 、 $k = 11$ 、パッチサイズ 128×128 の最適な設定を使用した場合、当アーキテクチャのフォワードパスはTesla V100 GPU上で2.23 msで完了します。NanoTWICEデータセットの全体画像に対するパッチごとの評価は、平均3.61秒で完了し、Napoleto et al. (2018) が報告した実行時間よりも大幅に高速です。彼らのアプローチでは、

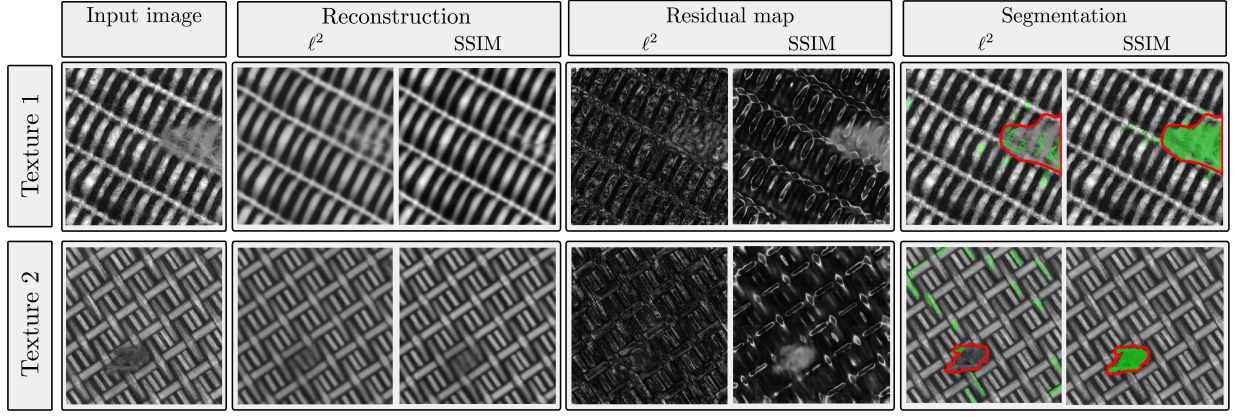


Figure 5: Qualitative comparison between reconstructions, residual maps, and segmentation results of an ℓ^2 -autoencoder and an SSIM autoencoder on two datasets of woven fabric textures. The ground truth regions containing defects are outlined in red while green areas mark the segmentation result of the respective method.

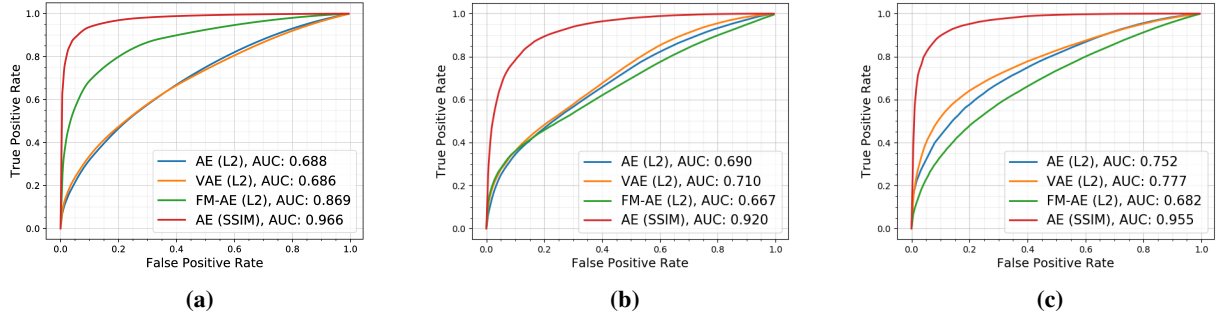


Figure 6: Resulting ROC curves of the proposed SSIM autoencoder (red line) on the evaluated datasets of nanofibrous materials (a) and the two texture datasets (b), (c) in comparison with other autoencoding architectures that use per-pixel loss functions (green, orange, and blue lines). Corresponding AUC values are given in the legend.

4.3. Results

Figure 5 shows a qualitative comparison between the performance of the ℓ^2 -autoencoder and the SSIM autoencoder on images of the two texture datasets. Although both architectures remove the defect in the reconstruction, only the SSIM residual map reveals the defects and provides an accurate segmentation result. The same can be observed for the NanoTWICE dataset, as shown in Figure 1.

We confirm this qualitative behavior by numerical results. Figure 6 compares the ROC curves and their respective AUC values of our approach using SSIM to the per-pixel architectures. The performance of the latter is often only marginally better than classifying each pixel randomly. For the VAE, we found that the reconstructions obtained by different latent samples from the posterior does not vary greatly. Thus, it could not improve on the deterministic framework. Employing feature matching only improved the segmentation result for the dataset of nanofibrous materials, while not yielding a benefit for the two texture datasets. Using SSIM as the loss and evaluation metric outperforms all other tested architectures significantly. By merely changing the loss function, the achieved AUC improves from 0.688 to 0.966 on the dataset of nanofibrous materials, which is comparable to the state-of-the-art by [Napoletano et al. \(2018\)](#), where values of up to 0.974 are reported. In contrast to this method, autoencoders do not rely on any model priors

such as handcrafted features or pretrained networks. For the two texture datasets, similar leaps in performance are observed.

Since the dataset of nanofibrous materials contains defects of various sizes and smaller sized defects contribute less to the overall true positive rate when weighting all pixel equally, we further evaluated the overlap of each detected anomaly region with the ground truth for this dataset and report the p -quantiles for $p \in \{25\%, 50\%, 75\%\}$ in Figure 7. For false positive rates as low as 5%, more than 50% of the defects have an overlap with the ground truth that is larger than 91%. This outperforms the results achieved by [Napoletano et al. \(2018\)](#), who report a minimal overlap of 85% in this setting.

We further tested the sensitivity of the SSIM autoencoder to different hyperparameter settings. We varied the latent space dimension d , SSIM window size k , and the size of the patches that the autoencoder was trained on. Table 2 shows that SSIM is insensitive to different hyperparameter settings once the latent space dimension is chosen to be sufficiently large. Using the optimal setup of $d = 500$, $k = 11$, and patch size 128×128 , a forward pass through our architecture takes 2.23 ms on a Tesla V100 GPU. Patch-by-patch evaluation of an entire image of the NanoTWICE dataset takes 3.61 s on average, which is significantly faster than the runtimes reported by [Napoletano et al. \(2018\)](#). Their approach requires between