the size of the dataset used for search which can be an issue when the dataset is very large or of high dimensionality. Our approach is designed to mitigate the complexity issues. First, we compute the initial image-level anomaly classification on global-pooled features which are 2048 dimensional vectors. Such kNN computation can be achieved very quickly for moderate sized datasets and different speedup techniques (e.g. KDTrees) can be used for large scale datasets. The anomaly segmentation stage requires pixel-level kNN computation which is significantly slower than image-level kNN. However, our method limits the sub-image kNN search to only the K nearest neighbors of the anomalous image significantly limiting computation time. We assume that the vast majority of images are normal, therefore only a small fraction of images require the next stage of anomaly segmentation. Additionally, the anomaly segmentation stage is required for explainability and trust building with the human operators, but in many cases it is not time-critical therefore putting a laxer requirement on computation time. Our method is therefore quite suitable for practical deployment from a complexity and runtime perspective.

Pre-trained vs. learned features: Previous sub-image anomaly detection methods have either used self-learned features or a combination of self-learned and pre-trained images features. Self-learned approaches in this context, typically train an autoencoder and use its reconstruction error for anomaly detection. Other approaches have used a combination of pre-trained and self-learned methods e.g. methods that use perceptual losses and [6] which uses a pre-trained encoder. Our numerical results have shown that our method significantly outperforms such approaches. We believe that given the limited supervision and small dataset size in normal-only training set as tackled in this work, it is rather hard to beat very deep pre-trained networks. We therefore use pre-trained features and do not attempt to modify them. The strong results achieved by our method attest to the effectiveness of this approach. We believe that future work should focus on methods for finetuning the deep pre-trained features for this particular task and expect it it improve over our method. That not-withstanding the ease of deployment and generality of our approach should make it a good choice in many practical settings.

6 Conclusion

We presented a novel alignment-based method for detecting and segmenting anomalies inside images. Our method relies on K nearest neighbors of pixel-level feature pyramids extracted by pre-trained deep features. Our method consists of two stages, which are designed to achieve high accuracy and reasonable computational complexity. Our method was shown to outperform the strongest current methods on two realistic sub-image anomaly detection datasets, while being much simpler. The ease of deployment enjoyed by our method should make it a good candidate for practitioners.