the feature space used giving rise to both kernel methods as well as deep methods [26] for learning features. Another set of methods is based on self-supervised learning. Golan and El-Yaniv [11] proposed a RotNet-based [10] approach, which performs geometric transformations on the input data and trains a network that attempts to recognize the transformation used. They use the idea that the trained classifier will generalize well to new normal images but not to anomalous images allowing it to be used as an anomaly detection criterion. Hendrycks et al. [18] improved the architecture and training procedure achieving strong performance. Bergman and Hoshen [4] combined this work with an SVDD type criterion and extended it to non-image data. Very recently Bergman et al. [3] showed that the features learned using such self-supervised methods are not competitive with generic ImageNet-based feature extractors. A simple method based on kNN (or efficient approximations) significantly outperformed such self-supervised methods.

Sub-image methods: The methods previously described tackled the task of classifying a whole image as normal or anomalous, and most of the techniques were not specific to images. The task of segmenting the particular pixels containing anomalies is special to images and has achieved far less attention from the deep learning community. Napoletano et al. [24] extracted deep features from small overlapping patches, and used a K-means based classifier over dimensionality reduced features. Bergmann et al. [5] evaluated both a ADGAN and autoencoder approaches on MVTech finding complementary strengths. More recently, Venkataramanan et al. [31] used an attention-guided VAE approach combining multiple methods (GAN loss [13], GRADCAM [29]). Bergmann et al. [6] used a student-teacher based autoencoder approach employing pre-trained ImageNet deep features (still requiring an expensive training stage). In this work, we present a novel sub-image alignment approach which is more accurate, faster, more stable than previous methods and does not require a dedicated training stage. To support research on sub-image anomaly detection, high quality datasets for evaluating this task have been introduced, such as: MVTech [5] - a dataset simulating an industrial fault detection where the objective is to detect parts of images of products that contain faults such as dents or missing parts. The ShanghaiTech Campus dataset [23] - a dataset simulating a surveillance setting where cameras observe a busy campus and the objective is to detect anomalous objects and activities such as fights. Hendrycks et al. [17] also proposed a new dataset containing anomalies such as road hazards. We evaluate our work on the two most used datasets, MVTech and ShanghaiTech Campus (STC).

3 Correspondence-based Sub-Image Anomaly Detection

We present our method for sub-image anomaly detection. Our method consists of several parts: i) image feature extraction ii) K nearest neighbor normal image retrieval iii) pixel alignment with deep feature pyramid correspondences.