3.1 Feature Extraction

The first stage of our method is the extraction of strong image level features. The same features are later used for pixel-level image alignment. There are multiple options for extracting features. The most commonly used option is self-supervised feature learning, that is, learning features from scratch directly on the input normal images. Although it is an attractive option, it is not obvious that the features learned on small training datasets will indeed be sufficient for serving as high-quality similarity measures. The analysis performed in Bergman et al. [4] illustrates that self-supervised features underperform ImageNet-trained ResNet features for the purposes of anomaly detection. We therefore used a ResNet feature extractor pre-trained on the ImageNet dataset. As image-level features we used the feature vector obtained after global-pooling the last convolutional layer. Let us denote the global feature extractor F, for a given image x_i, we denote the extracted features f_i:

$$f_i = F(x_i) \tag{1}$$

At initialization, the features for all training images (which are all normal) are computed and stored. At inference, only the features of the target image are extracted.

3.2 K Nearest Neighbor Normal Image Retrieval

The first stage in our method is determining which images contain anomalies using DN2 [4]. For a given test image y, we retrieve its K nearest normal images from the training set, $N_k(f_y)$. The distance is measured using the Euclidean metric between the image-level feature representations.

$$d(y) = \frac{1}{K} \sum_{f \in N_K(f_y)} \|f - f_y\|^2 \tag{2}$$

Images are labelled at this stage as normal or anomalous. Positive classification is determined by verifying if the kNN distance is larger than a threshold τ. It is expected that most images are normal, and only few images are designated as anomalous.

3.3 Sub-image Anomaly Detection via Image Alignment

After being labelled as anomalous at the image-level stage, the objective is to locate and segment the pixels of one or multiple anomalies. In the case that the image was falsely classified as anomalous, our objective would be to mark no pixels as anomalous.

As a motivational idea, let us consider aligning the test image to a retrieved normal image. By finding the differences between the test and normal image, we would hope to detect the anomalous pixels. This naive method has several flaws i) assume that there are multiple normal parts the object may possibly