Note that many of these methods are not specific to images. There are three main classes of methods for image-level anomaly detection: reconstruction-based, distribution-based and classification-based.

Reconstruction-based methods learn a set of basis functions on the training data, and attempt to reconstruct the test image using a sparse set of these basis functions. If the test image cannot be faithfully reconstructed using the basis functions, it is denoted as anomalous, as it is likely that it came from a different basis from that of the normal training data. Different methods vary in terms of the set of basis functions and loss functions they use. Popular choices of basis functions include: K-means [15], K nearest neighbors (kNN) [9], principal component analysis (PCA) [20]. The loss functions used vary between simple vector metrics such as Euclidean or L_1 losses and can use more complex perceptual losses such as structural similarity (SSIM) [32]. Recently deep learning methods have broadened the toolbox of reconstruction-based methods. Principal components have been extended to non-linear functions learned by autoencoders [27], including both denoising as well as variational autoencoders (VAEs). Deep perceptual loss functions [34] significantly improve over traditional perceptual loss functions. The main disadvantages of reconstruction-based loss functions are: i) sensitivity to the particular loss measure used for evaluating the quality of reconstruction, making their design non-obvious and hurting performance ii) determining the correct functional basis.

The second class of methods is distribution-based. The main principle is to model the probability density function (PDF) of the distribution of the normal data. Test samples are evaluated using the PDF, and test samples with low probability density values are designated as anomalous. Different distribution-based methods differ by the distributional assumptions that they make, the approximations used to estimate the true PDF, and by the training procedure. Parametric methods include Gaussian or mixture of Gaussians (GMM). Kernel density estimation [21] is a notable non-parametric method. Nearest neighbors [9] can also be seen as a distributional (as it performs density estimation), but note that we also designated it a reconstruction-based method. Recently deep learning methods have improved performance, particularly by mapping high-dimensional data distributions into a lower and denser space. PDF estimation is typically easier in lower dimensional spaces. Learning the deep projection and distributional modeling can be done jointly as done by [36]. Another recent development, adversarial training, was also applied to anomaly detection e.g. ADGAN [8]. Although in principle distributional-methods are very promising, they suffer from some critical drawbacks: i) real image data rarely follows simple parametric distributional assumptions ii) non-parametric methods have high sample complexity and often require large training set that is often not available in practice.

Recently, classification-based methods have achieved dominance for image-level anomaly detection. One such paradigm is one-class support vector machines (OC-SVM) [28]. One of its most successful variants is support vector data description (SVDD) [30] which can be seen as a finding the minimal sphere which contains at least a given fraction of the data. These methods are very sensitive to