Introduction

Anomaly detection in images is crucial for quality control in industries where identifying defects early is essential. There are three main approaches to this task: supervised, unsupervised, and the newer zero-shot anomaly segmentation (ZSAS).

Supervised methods require labeled data, where both normal and defective examples are provided during training. These models tend to be highly accurate for defects they've been trained on. However, gathering enough labeled data, especially for rare or new types of defects, is time consuming and costly, limiting their scalability in dynamic environments.

Unsupervised methods, on the other hand, don't need labeled data. Instead, they learn to recognize what a "normal" image looks like, and anomalies are identified as deviations from this norm. While these methods are more flexible, they can sometimes struggle with precision, leading to more false positives compared to supervised models.

Zero-shot anomaly segmentation (ZSAS) is a more recent approach that bypasses the need for specific training on a dataset. Instead, it uses pre-trained models to detect unseen anomalies. This makes ZSAS faster to deploy and more adaptable to diverse or unpredictable defects, offering a flexible solution where data collection is challenging.

Each approach has its advantages and limitations. In this section, we will review different supervised and unsupervised, Zero-shot anomaly segmentation approaches

1 - PaDiM

The PaDiM [1] uses a pretrained CNN to extract patch embeddings from normal images during training, modeling each patch with a multivariate Gaussian distribution. During testing, for each patch in a test image, PaDiM computes the Mahalanobis distance between the test patch's embedding and the corresponding learned Gaussian distribution. This distance provides an anomaly score, and high scores indicate anomalies. PaDiM localizes defects by generating an anomaly map, where regions with high scores mark defective parts. It distinguishes between normal and anomalous patches but does not classify different types of anomalies.

You can see the model's structure below.
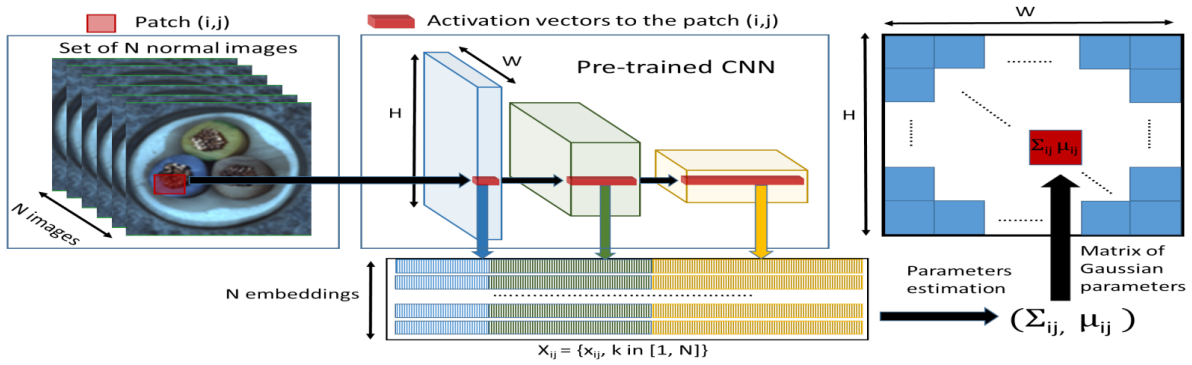
In the test phase, the model efficiently computes anomaly scores by comparing each test patch's embedding to the learned Gaussian distributions, without the need to store or compare all training data. The key advantage of PaDiM is its efficiency at test time, as it relies on parametric modeling instead of computationally expensive approaches like K-NN, making PaDiM scalable and suitable for industrial applications.

2 - PatchCore

The PatchCore focuses on extracting mid level patch features from images using a pretrained CNN, such as ResNet. During the embedding extraction phase, these patch features, representing normal images, are stored in a memory bank. To optimize efficiency, PatchCore uses greedy coreset subsampling during the learning phase, reducing the size of the memory bank while retaining the most representative patch features, which minimizes redundancy and improves inference speed. During the inference phase, PatchCore compares each patch in a test image to its nearest neighbor in the memory bank, calculating an anomaly score based on the distance. High anomaly scores indicate defective areas, which are localized using an anomaly map. The system operates in a binary classification mode and offers fast, scalable detection and localization due to the coreset subsampling.

Below is the complete structure of the model, which resembles PaDim.
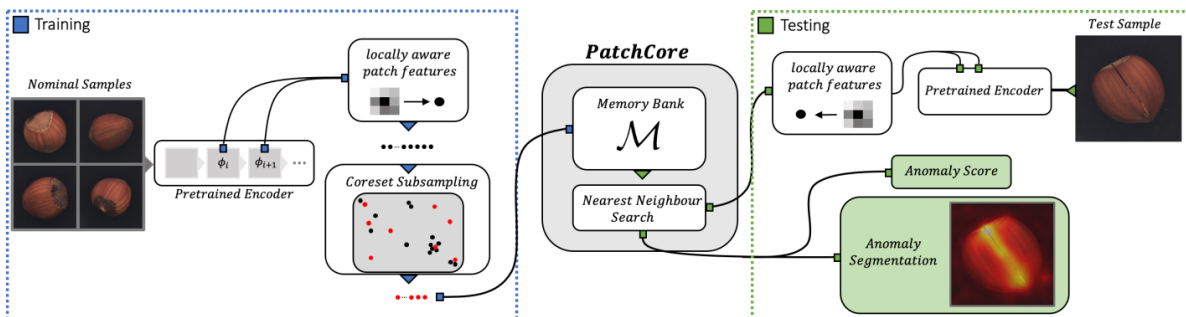
Its key advantage lies in the use of coreset subsampling, which significantly reduces the memory bank size while retaining the most representative patch features, thereby reducing computational costs.

3 - Skip-GANomaly

Skip-GANomaly [3] combines an encoder-decoder architecture with skip connections and adversarial training. The encoder compresses the input image into a low dimensional latent representation, while the decoder reconstructs the image. Skip connections between corresponding layers in the encoder and decoder help preserve spatial information and improve reconstruction quality.

The model is trained on normal images only, with the generator (encoder-decoder) learning to reconstruct normal images, while the discriminator distinguishes real from generated images. The training objective minimizes three losses: adversarial loss, contextual loss (similarity between input and output images), and latent loss (similarity in latent representations).

To detect anomalies, it computes an anomaly score by comparing the input image with its reconstruction in both image and latent spaces. A higher reconstruction error or latent space discrepancy indicates an anomaly. This method performs binary classification.

The key advantage of this model is its use of skip connections, which enhance reconstruction by preserving both local and global image details. Its dual-space anomaly detection in both image and latent spaces increases accuracy and robustness. The adversarial training further refines the model's ability to detect subtle anomalies in imbalanced datasets.

4 - Auto-Classifier

The Auto-Classifier [4] method is a two-part approach designed to enhance defect detection by integrating multiple Convolutional Neural Networks (CNNs) with automated machine learning (AutoML).

4 – 1 Weighted CNN Ensemble

The first part of the method leverages multiple state-of-the-art CNN architectures, such as VGG, ResNet, and DenseNet. Each CNN is independently trained to extract features

from input images, and their predictions are combined to create a final output. The key innovation here is the use of a weighted fusion, where each CNN's prediction is weighted based on its performance, measured by the Area Under the Curve (AUC) score from the validation set. CNNs with higher AUC scores contribute more to the final prediction, ensuring that the most accurate models have a greater influence.

4 – 2 Replacing the Classification Layer and AutoML Optimization

In the second part, the best-performing CNN (based on validation results) has its final classification layer removed and replaced. Specifically, the last fully connected layers of the CNN are replaced with a new classifier generated through H2O AutoML. The AutoML process evaluates several machine learning models (such as XGBoost, Random Forest, and Neural Networks) and optimizes the new classifier by performing hyperparameter tuning and model stacking. This allows the Auto-Classifier to build on the robust feature extraction capabilities of the CNN while optimizing the classification stage for better accuracy and faster inference.

By combining both approaches, an ensemble of weighted CNNs and the replacement of the classification layer with an AutoML improves overall detection performance. This method is particularly effective for real-world industrial applications like surface defect detection,

5- Baru-Net

Baru-Net [5] is a supervised deep learning model based on the U-Net architecture, designed for multi-class classification in defect detection tasks. It integrates two key modules: the Convolutional Block Attention Module (CBAM) and Atrous Spatial Pyramid Pooling (ASPP), along with bilinear interpolation for efficient upsampling. The model enhances both speed and accuracy, particularly for detecting multiple types of surface defects in highly reflective parts.

5-1 CBAM (Convolutional Block Attention Module)

CBAM is designed to refine feature maps by applying attention both across channels and spatial dimensions. It consists of two submodules:

5-1-1 CAM (Channel Attention Module):

Identifies crucial feature channels using global pooling and generates attention weights to emphasize relevant channels.

5-1-2 SAM (Spatial Attention Module):

Generates attention maps for key spatial regions, focusing on critical areas in the image for better defect detection.

5-2 ASPP (Atrous Spatial Pyramid Pooling)

ASPP uses atrous convolutions with different dilation rates to capture features at multiple scales. This allows the model to detect defects of varying sizes without reducing resolution.

Baru-Net applies CBAM at each stage of the encoder to enhance feature extraction, while ASPP processes the final feature maps for multi-scale detection.

Overall , The key advantage of Baru-Net is its ability to combine attention mechanisms with multi-scale feature extraction, achieving a high detection accuracy for real-world industrial applications.

For example, you can see the result of applying this method to ceramic for segmenting the defective parts.

(b)

(e)
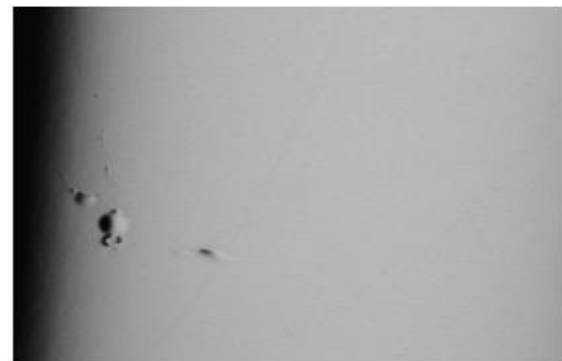
6 - UzADL

The UzADL [6] model is an unsupervised anomaly detection and localization framework that uses a graph Laplacian matrix for defect detection without labeled data. The method begins with a pre-processing stage, where images are standardized and converted into tensors for uniform data handling.
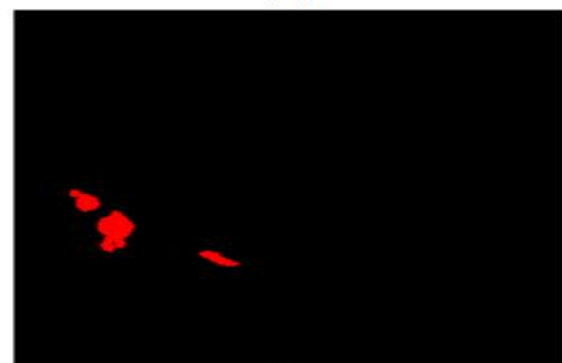
Then in the pseudo-labeling stage, the model constructs a graph Laplacian matrix from adjacency and degree matrices, and then uses eigenvalues and eigenvectors to partition the dataset into normal and defective samples. Next, during the training process, the model fine-tunes a ResNet-18 architecture pre-trained on ImageNet,

transforming the anomaly detection task into a binary classification problem, which classifies images as normal or defective.

After training, in the defect visualization stage, the model uses global average pooling (GAP) and feature map activations to generate heatmaps, highlighting the defective regions for interpretability.

The key advantage of UzADL is its ability to detect anomalies without labeled data while maintaining high accuracy, fast convergence, and interpretable visual feedback.

7 - Segment Any Anomaly plus

The *SAA+* [7] method is a system that detects defects in images without needing specific training for each type of defect. It uses pre-trained models like *GroundingDINO* and *SAM* (Segment Anything Model). First, an anomaly region generator finds rough areas in the image that may have defects using a simple prompt like "anomaly." These areas are then refined into accurate, pixel-level masks using *SAM*.

Unlike unsupervised or self-supervised learning methods, which need a lot of training on normal data to understand what is "normal" and find defects, *Zero-Shot Anomaly Segmentation* (ZSAS) can detect new types of defects without any specific training. This makes it faster and more flexible, especially when it's hard or costly to collect data.

In the *SAA* method, prompts like "anomaly" or "defect" are used, but these simple prompts often give inaccurate results. To fix this problem, *SAA+* uses hybrid prompts, combining expert knowledge with information from the image. Also, image features like saliency maps help the model focus on the most important areas, improving accuracy.

We mention the difference between SAA and SAA+ in the following image.
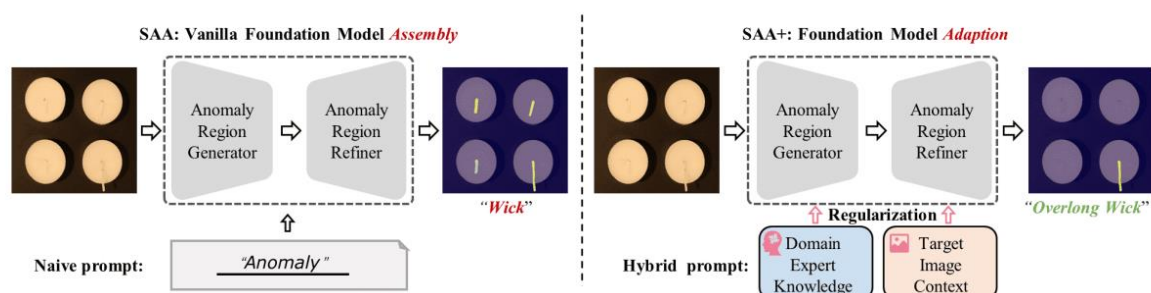
The key advantage of the SAA+ method is in its hybrid prompt regularization, which enhances the accuracy of zero-shot anomaly segmentation without the need for

training. By combining expert knowledge with target image context, SAA+ refines basic prompts and increases accuracy.

Reference

[1] Defard, T., Setkov, A., Loesch, A., & Audigier, R. (2021). PADIM: A patch Distribution Modeling Framework for Anomaly Detection and Localization. In *Lecture notes in computer science* (pp. 475–489). https://doi.org/10.1007/978-3-030-68799-1_35

[2] Roth, K., Pemula, L., Zepeda, J., Scholkopf, B., Brox, T., & Gehler, P. (2022b). Towards Total Recall in Industrial Anomaly Detection. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr52688.2022.01392

[3] Akcay, S., Atapour-Abarghouei, A., & Breckon, T. P. (2019). Skip-GANomaly: Skip Connected and Adversarially Trained Encoder-Decoder Anomaly Detection. https://doi.org/10.1109/ijcnn.2019.8851808

[4] Lopes, V., & Alexandre, L. A. (2020). Auto-Classifier: A Robust Defect Detector Based on an AutoML Head. In *Lecture notes in computer science* (pp. 137–149). https://doi.org/10.1007/978-3-030-63830-6_12

[5] Chen, J., Zhang, B., Jiang, Q., & Chen, X. (2023). Baru-Net: Surface Defects Detection of Highly Reflective Chrome-Plated Appearance Parts. *Coatings*, *13*(7), 1205. https://doi.org/10.3390/coatings13071205

[6] Olimov, B. a. U., Veluvolu, K. C., Paul, A., & Kim, J. (2022). UzADL: Anomaly detection and localization using graph Laplacian matrix-based unsupervised learning method. *Computers & Industrial Engineering*, *171*, 108313. https://doi.org/10.1016/j.cie.2022.108313

[7] Cao, Y., Xu, X., Sun, C., Cheng, Y., Du, Z., Gao, L., & Shen, W. (2023). Segment Any

Anomaly without Training via Hybrid Prompt Regularization. *Preprint*.

https://doi.org/10.48550/arXiv.2305.10724