1. Introduction

- **Brief Overview of Anomaly Detection**:
- **MVTec-AD Dataset**
- **Objective**

2. Methods

- **Dataset Preparation**
- **Model Descriptions**:
    1. PatchCore
    2. EfficientAd
- **Training and Evaluation**

3. Results

4. Discussion

- **Model Strengths and Weaknesses**
- **Usefulness of AUROC for Anomaly Detection**
- **Potential Improvements**

5. Conclusion

- **Summary of Findings**
- **Implications for Industrial Applications**
- **Future Directions**

1. Introduction

Brief Overview of Anomaly Detection

Anomaly detection is the process of identifying unusual patterns or behaviors that deviate from the norm. In industrial settings, detecting such anomalies is essential for quality assurance, safety, and cost-effectiveness, as defects in products can lead to failures and increased costs.

MVTec-AD Dataset

The MVTec Anomaly Detection (MVTec-AD) dataset is a benchmark dataset used in the computer vision community to evaluate anomaly detection models. It includes 15 object and texture categories, such as tile, leather, and grid, each with defect-free samples and samples containing anomalies. These anomalies include both structural defects, like scratches or dents, and textural defects, such as discoloration. The high variability of normal samples and subtlety of certain defects make this dataset challenging for unsupervised anomaly detection models.

Objective

This project aims to benchmark the performance of two anomaly detection models, PatchCore and EfficientAd, on selected categories of the MVTec-AD dataset. By training and evaluating these models on flat surface categories—tile, leather, and grid—we seek to assess their effectiveness in detecting anomalies in different contexts. The performance of each model will be evaluated using AUROC scores, and we will also implement a similarity search feature using Qdrant to retrieve visually similar anomalous cases

2. Methods

Dataset Preparation

The dataset preparation process involves downloading and setting up the MVTec-AD dataset using the anomalib library. We focused on three flat surface categories: tile, leather, and grid. Each category's data is split into training and testing sets, with data transformations applied to ensure consistency and compatibility with the models.

Model Descriptions

PatchCore is designed for efficient anomaly detection through coresets, which are subsets of feature vectors that represent critical data points. By storing only a fraction of features, PatchCore can perform anomaly detection with less memory usage, making it well-suited for detecting subtle differences between normal and defective samples.

EfficientAd is a segmentation-based anomaly detection model that focuses on highlighting defect regions in an image. It applies segmentation techniques to identify anomalies in texture and structure, providing a detailed analysis of potential defects.

Training and Evaluation

Each model was trained on the MVTec-AD dataset using the anomalib library. Training parameters, such as batch size and number of epochs, were adjusted to optimise performance. After training, models were evaluated on a test set to calculate the AUROC (Area Under the Receiver Operating Characteristic curve) for each category, which provides a measure of the model's ability to distinguish between normal and defective samples.

3. Results

The performance of both PatchCore and EfficientAd models was evaluated on the three selected flat surface categories from the MVTec-AD dataset: tile, leather, and grid. The evaluation focused on two primary metrics, Image AUROC and Pixel AUROC, as well as Image F1Score and Pixel F1Score for a more detailed analysis of each model's strengths and weaknesses at both the image level and the pixel (localized anomaly) level.

PatchCore Model Results

PatchCore’s approach of using coresets allowed it to perform well across all three categories, demonstrating its efficiency in identifying anomalies with high precision at the image level. Below are the detailed results for each category:

Tile:

Image AUROC: 0.987

Image F1Score: 0.981

Pixel AUROC: 0.947

Pixel F1Score: 0.620

Interpretation: PatchCore achieved an Image AUROC of 0.987, indicating that it is highly effective at distinguishing normal tile surfaces from defective ones. The high Image F1Score of 0.981 further supports this, showing that PatchCore provides consistent predictions at the image level. However, at the pixel level, while it achieved a strong Pixel AUROC of 0.947, the Pixel F1Score of 0.620 suggests some challenges in precisely localizing anomalies within the image. This may be due to the subtle texture variations within tile surfaces that can complicate pixel-level segmentation.

Leather:

Image AUROC: 1.000

Image F1Score: 0.994

Pixel AUROC: 0.990

Pixel F1Score: 0.442

Interpretation: PatchCore showed excellent performance on leather surfaces, achieving a perfect Image AUROC of 1.000 and an Image F1Score of 0.994. This indicates that PatchCore can reliably distinguish between normal and anomalous leather surfaces at the image level. However, the Pixel F1Score of 0.442 is notably lower than its Image F1Score. This could be due to the natural variability in leather textures, which can cause the model to struggle with precise anomaly localization despite its high overall classification accuracy.

Grid:

Image AUROC: 0.980

Image F1Score: 0.964

Pixel AUROC: 0.979

Pixel F1Score: 0.376

Interpretation: For grid surfaces, PatchCore achieved an Image AUROC of 0.980 and an Image F1Score of 0.964, both indicating strong classification performance at the image level. At the pixel level, it achieved a Pixel AUROC of 0.979, suggesting it can reasonably identify anomalous regions. However, the Pixel F1Score was relatively low at 0.376, indicating that while PatchCore can detect the presence of anomalies in grid images, it may struggle to pinpoint exact locations. This could be due to the repetitive patterns in grid textures that make it difficult for the model to discern subtle defects at a fine-grained level.

In summary, PatchCore displayed strong image-level anomaly detection across all categories, as reflected by its high Image AUROC and Image F1Score values. However, its pixel-level performance varied, indicating challenges in precise anomaly localization, particularly in categories with intricate textures like leather and grid.

EfficientAd Model Results

EfficientAd, designed to perform pixel-level segmentation for more detailed anomaly localization, demonstrated a different performance pattern across the three categories. The model's results highlight its effectiveness at localizing defect regions, although it also showed variability in certain cases.

Tile:

Image AUROC: 0.996

Image F1Score: 0.976

Pixel AUROC: 0.883

Pixel F1Score: 0.681

Interpretation: EfficientAd achieved an Image AUROC of 0.996 and an Image F1Score of 0.976 on tile surfaces, indicating its high accuracy in detecting anomalies at the image level. At the pixel level, it scored a Pixel AUROC of 0.883 and a Pixel F1Score of 0.681, suggesting that EfficientAd is proficient at localizing defects within tile images. This result implies that EfficientAd's segmentation approach is well-suited for tile textures, as it can identify and highlight defective regions effectively, aiding in visual inspection.

Leather:

Image AUROC: 0.622

Image F1Score: 0.847

Pixel AUROC: 0.778

Pixel F1Score: 0.232

Interpretation: On leather surfaces, EfficientAd's performance was less consistent, with an Image AUROC of 0.622 and a relatively high Image F1Score of 0.847. The low Image AUROC suggests challenges in distinguishing defective leather samples from normal ones at the image level, potentially due to the high variability of leather textures. At the pixel level, EfficientAd achieved a Pixel AUROC of 0.778 but a significantly lower Pixel F1Score of 0.232, indicating difficulty in precisely localizing anomalies within the leather texture. This might be due to the model's sensitivity to normal variations in leather textures, which can lead to misclassification of natural variations as defects.

Grid:

Image AUROC: 0.720

Image F1Score: 0.846

Pixel AUROC: 0.732

Pixel F1Score: 0.127

Interpretation: For grid surfaces, EfficientAd's Image AUROC of 0.720 indicates that it had some difficulty in accurately classifying images as anomalous or normal. The Image F1Score of 0.846, however, shows that EfficientAd still maintained a fair level of consistency at the image level. The Pixel AUROC of 0.732 and Pixel F1Score of 0.127 reveal that EfficientAd faced challenges with precise anomaly localization in grid images. The repetitive pattern of grid textures may have contributed to this lower performance, as EfficientAd's segmentation-based approach might struggle with such structured patterns that often have subtle defects.

Comparison and Summary

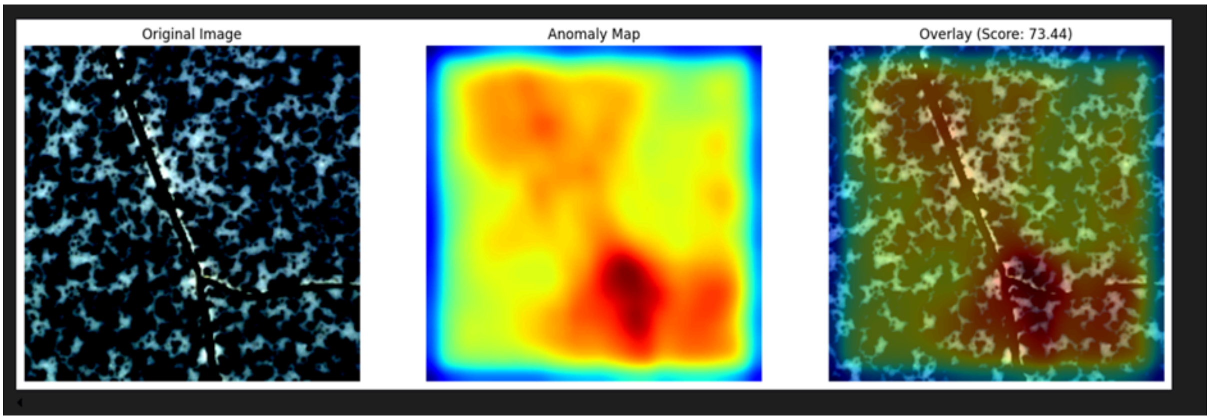
The average metrics across the three categories provide an overall summary of each model's performance:

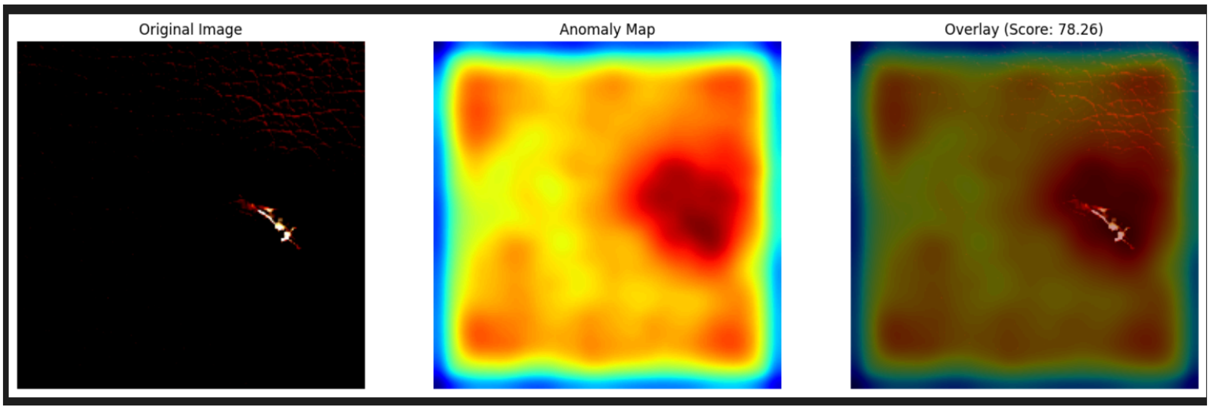
These results show that PatchCore consistently outperformed EfficientAd at the image level, achieving an average Image AUROC of 0.9893 compared to EfficientAd's 0.7792. This suggests that PatchCore is highly effective in distinguishing between normal and defective samples, making it a reliable choice for high-level anomaly detection tasks where detailed localization may not be critical.

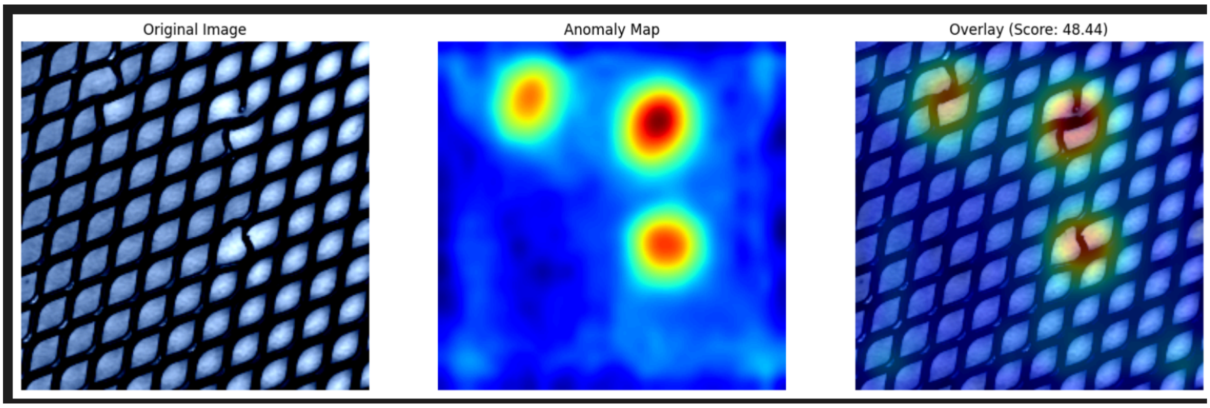
At the pixel level, both models faced challenges with precise localization, though PatchCore had a higher average Pixel AUROC (0.9724) compared to EfficientAd's 0.7975. EfficientAd's approach was beneficial for cases requiring detailed defect visualization, as seen in its performance with tile surfaces, where it provided a relatively high Pixel F1Score of 0.681. However, in categories with complex textures like leather and grid, EfficientAd struggled, as evidenced by its lower Pixel F1Scores.

Overall, these results indicate that PatchCore is more suitable for high-level classification tasks, while EfficientAd offers valuable insights when detailed defect localization is required, albeit with some limitations in highly variable textures. The evaluation highlights how the choice of model may depend on specific industrial requirements, such as the need for high-level anomaly detection versus detailed defect segmentation.

Also these images show the anomaly identification heatmap for Patchcore




Original Image
Anomaly Map
Overlay (Score: 78.26)


Original Image
Anomaly Map
Overlay (Score: 48.44)

4. Discussion

Model Strengths and Weaknesses

PatchCore performed well on categories with high textural variability, thanks to its use of coresets for efficient feature representation. However, it sometimes struggled with subtle anomalies. EfficientAd, while effective at segmenting defect regions, occasionally misidentified normal variations as defects in categories with high intra-class variability

Usefulness of AUROC for Anomaly Detection

AUROC is a key metric for anomaly detection, as it measures the model's ability to differentiate between normal and defective samples. High AUROC values indicate strong discriminatory power, which is crucial for applications where detecting even subtle defects is critical.

Potential Improvements

Future work could explore additional anomaly detection methods, fine-tune model hyperparameters, or incorporate feature extraction techniques to further improve detection accuracy, especially for challenging defect types.

5. Conclusion

Summary of Findings

This study evaluated the PatchCore and EfficientAd models on the MVTec-AD dataset, focusing on the flat surface categories of tile, leather, and grid. Both models demonstrated high AUROC scores, with PatchCore showing strength in textural analysis and EfficientAd excelling in segmentation.

Implications for Industrial Applications

The findings suggest that both PatchCore and EfficientAd could be valuable tools in industrial quality control, where accurate and efficient anomaly detection is essential for maintaining high standards and reducing costs associated with defective products.

Future Directions

Future work could expand on this study by testing these models on different datasets or real-time data. Enhancements to the similarity search functionality could also enable more robust anomaly clustering, aiding in faster defect identification.