**MVTec-AD Anomaly Localization**

Project Report

Prepared by: Aadhimozhi M

Company: Textro AI

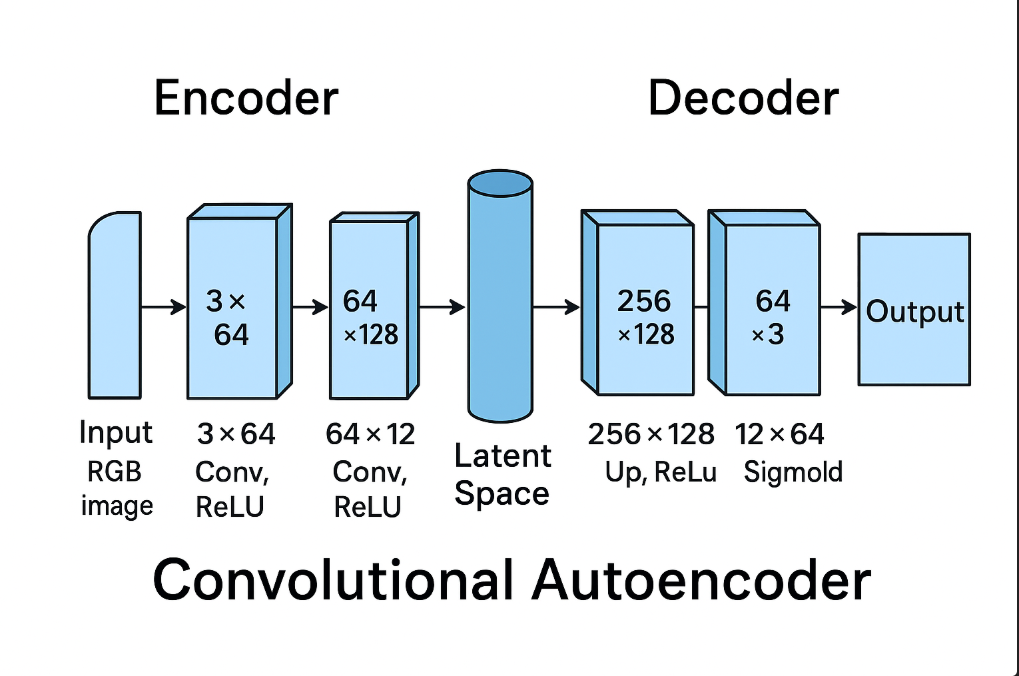
Start Date: 16/09/2025

End Date: 22/09/2025

**1. Introduction:**

This project focuses on anomaly localization using the MVTec-AD dataset, which consists of high-resolution industrial images containing various types of defects. The goal is to detect if an image is anomalous and localize the region of defect using deep learning techniques, specifically an unsupervised convolutional autoencoder.

**2. Architecture Diagram:**



**3. Anomaly Detection Pipeline:**

The complete anomaly detection pipeline is structured into the following steps:

* Dataset loading and per-category arrangement using MVTec-AD structure.

Link:[MVTec AD](https://www.kaggle.com/datasets/ipythonx/mvtec-ad)

* Data preprocessing: resizing to 256x256, normalization (ImageNet mean and std).
* Model training: using only 'good' (non-defective) samples with data augmentation.
* Image reconstruction via the autoencoder.
* Anomaly map generation using per-pixel reconstruction error.
* Gaussian smoothing of anomaly map to reduce noise.
* Thresholding using Otsu’s method to convert heatmap to binary mask.
* Morphological post-processing: removal of small objects, closing operations.
* Evaluation using AUROC (pixel-level & image-level).
* Visualization of results: real image, heatmap, binary mask, and ground truth

**4. Model and Hyperparameters**

**Model used is Convolutional Autoencoder (CAE)**

Used for **unsupervised anomaly detection and localization**. It learns to reconstruct only the normal samples. High reconstruction error on test images indicates potential anomalies.

**Hyperparameters:**

|  |  |
| --- | --- |
| **Hyperparameter** | **Value / Description** |
| **Model type** | Convolutional Autoencoder |
| **Input image size** | 256×256 |
| **Batch size** | 16 (training), 1 (testing) |
| **Epochs** | 20 |
| **Optimizer** | Adam |
| **Learning rate** | 1e-4 |
| **Loss function** | Mean Squared Error (MSE) |
| **Augmentations** | Horizontal flip, rotation, color jitter (only during train) |
| **Normalization** | ImageNet mean/std |
| **Heatmap smoothing** | Gaussian filter with sigma=2 |
| **Thresholding** | Otsu’s method |
| **Postprocessing** | Remove small objects (min size=100), binary closing |
| **Evaluation metrics** | Pixel-level AUROC, Image-level AUROC, IoU |

**5. Training Strategy**

The model was trained exclusively on normal (good) images using a reconstruction-based approach. The MSE loss was used to measure the discrepancy between input and reconstructed image. The network learns to accurately reconstruct only the non-defective patterns, failing to reconstruct anomalies in test images, which can then be localized. Training was performed over 20 epochs with batch size 16 using the Adam optimizer.

**6. Evaluation Metrics**

The following metrics were used to evaluate the model:

* **Pixel-level AUROC:** Measures how well the anomaly map separates defective vs non-defective pixels.
* **Image-level AUROC:** Measures whether the maximum anomaly score in a test image can classify it as anomalous.
* **IoU (Intersection over Union)**: Compares predicted binary mask with ground truth.
* **PRO (Per-Region Overlap):** M**e**asures accuracy of overlapping defect regions. These metrics were computed per image and aggregated per category.

**7. Ablation Study**

An ablation analysis was performed to study the effect of various design choices:

* **No Gaussian Smoothing:** Results in highly noisy heatmaps, reducing localization quality.
* **No Morphological Post-processing:** Binary masks are fragmented and contain small false positives.
* **Shallow Encoder:** Faster but less effective at compressing context, reducing anomaly detection accuracy.
* **No Augmentation:** Overfitting observed; test accuracy dropped on real anomalies.

**8. Discussion & Future Improvements**

The convolutional autoencoder approach effectively localizes anomalies in texture-rich categories like leather and tile. However, limitations exist: - Poor reconstruction of large-scale structural defects. - Blurry outputs due to MSE loss. - Sensitivity to lighting and context variations.

**Future Directions:** Incorporating perceptual loss (e.g., SSIM or VGG loss). - Using deeper or multi-scale autoencoders. - Attention-based reconstruction (Vision Transformers). - Feature-matching or memory-augmented methods