MIT Professional Education

Convolutional Neural Network Autoencoder-based for anomaly detection

Detecting Defective Screws

Program: MITx Mastering Neural Networks

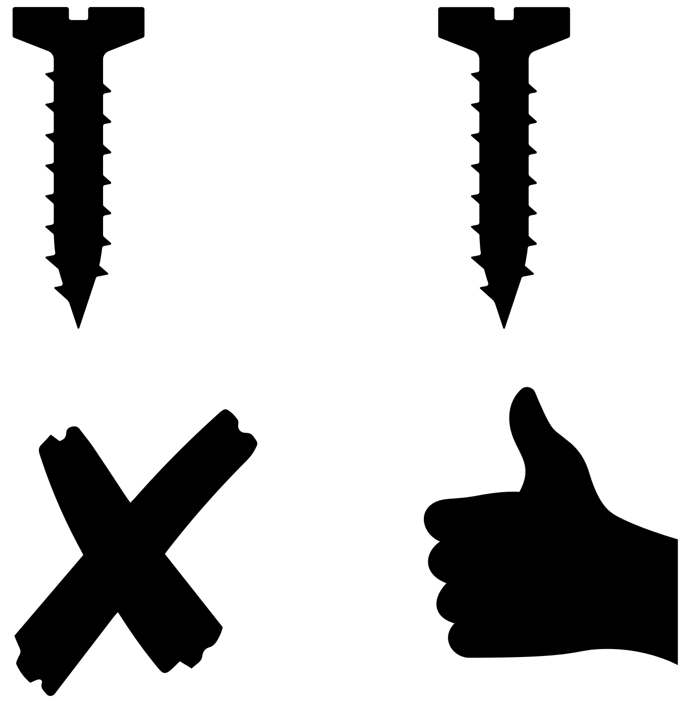
Kenny Daniel Herrera Cabrera | February 14th, 2026

Content

1. Problem statement
2. Dataset
3. Autoencoder architecture
4. Results
5. Conclusion
6. Reference to Jupyter notebook

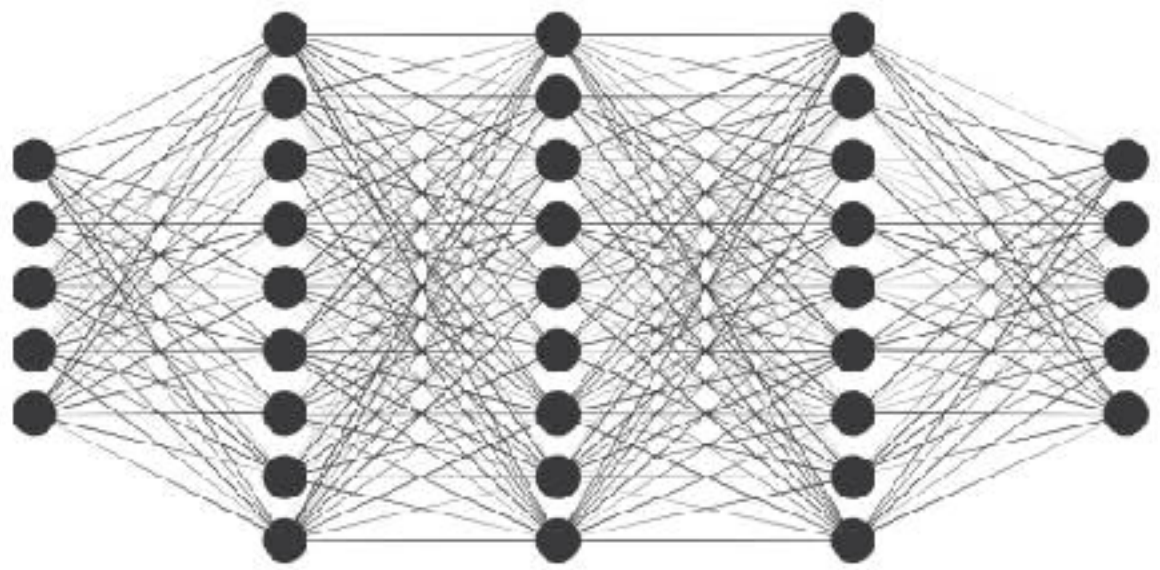
1. Problem Statement

Manufacturing Quality Control



Good or bad?

Goal



Manual inspection of screws for quality control is a labor-intensive, inconsistent and costly.

Build a deep learning system that is able to flag defect screws with high confidence.

2. Image Dataset

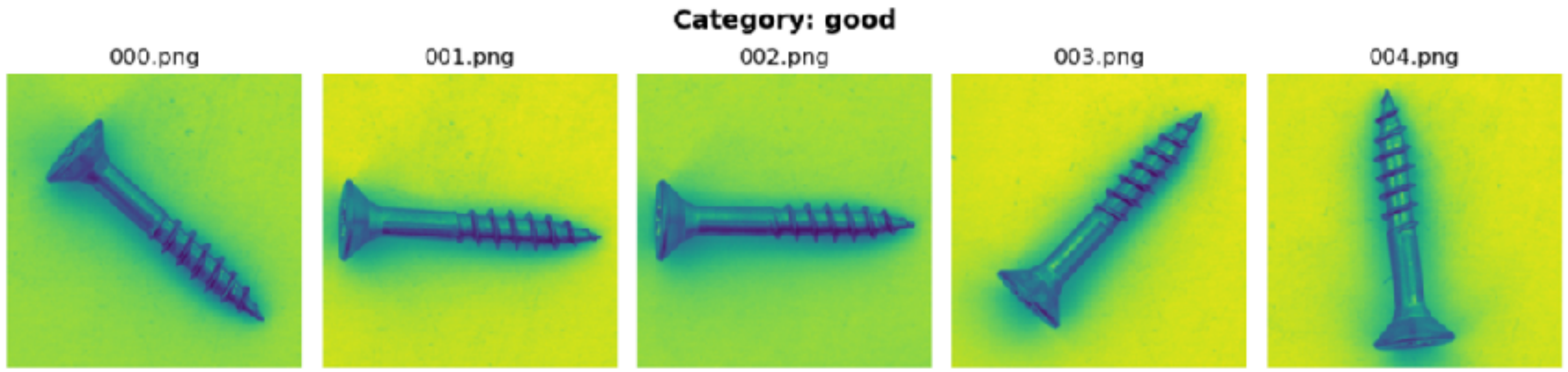
Training and test screws

MVTec screws dataset:

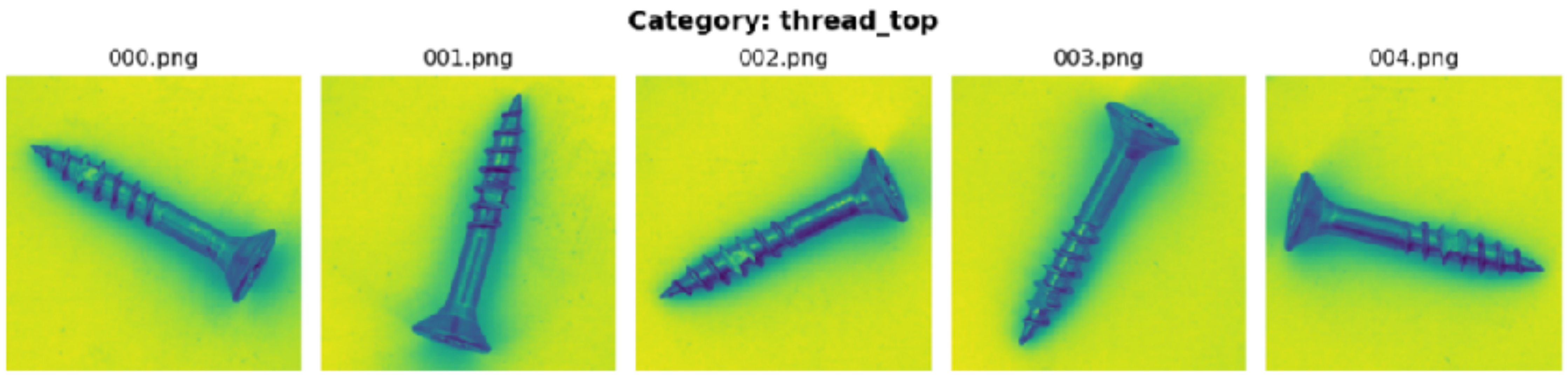
- 320 training images (good screws only)
- 160 test images -> 5 different types of defects

```
============================================================
Defect Types in Test Set
============================================================
  ✓ Normal: good                  - 41 samples
  ✗ Defect: manipulated_front     - 24 samples
  ✗ Defect: scratch_head          - 24 samples
  ✗ Defect: scratch_neck          - 25 samples
  ✗ Defect: thread_side           - 23 samples
  ✗ Defect: thread_top            - 23 samples
```

Example of good screws



Example of defective screw

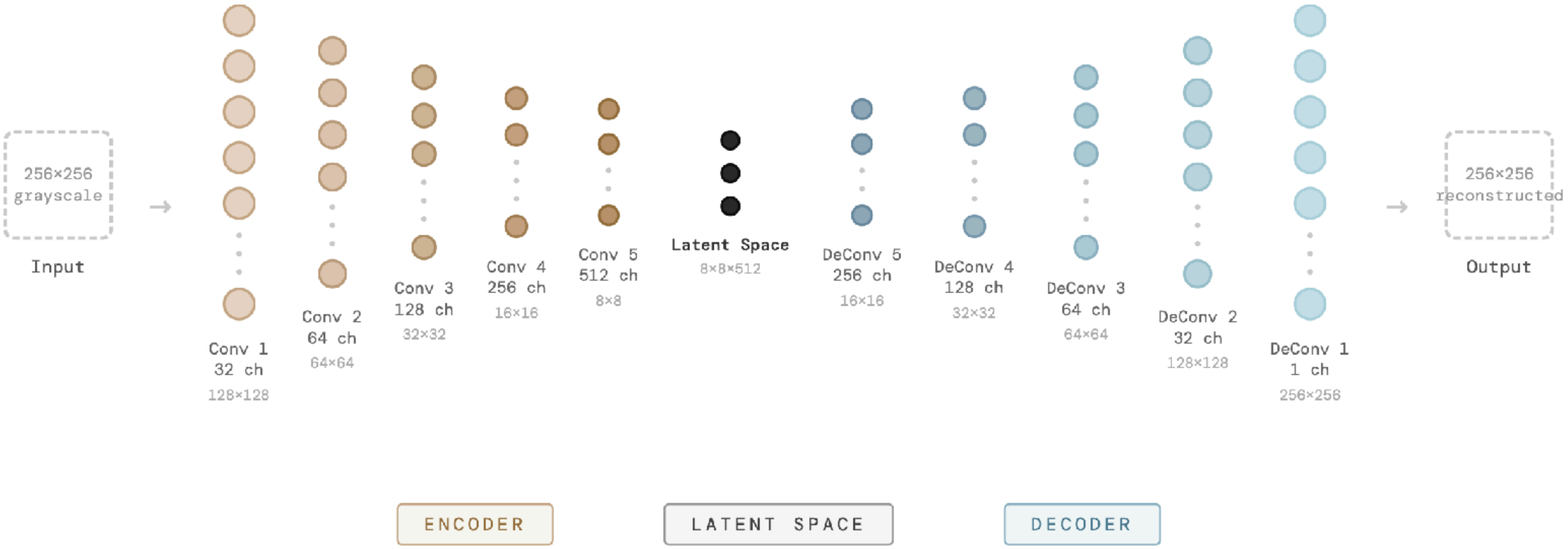


3. Autoencoder Architecture

CNN Autoencoder for Anomaly Detection

Screw Autoencoder - Architecture

256 × 256 input · 5-layer conv autoencoder · trained on normal screws only



3. Autoencoder Architecture

CNN Autoencoder for Anomaly Detection

Layer Details

Each conv	Conv2d (3×3, stride=2, pad=1) → BatchNorm2d → ReLU
Each deconv	ConvTranspose2d (3×3, stride=2, pad=1, out_pad=1) → BatchNorm2d → ReLU
Final layer	ConvTranspose2d → Sigmoid
Loss	MSE (Mean Squared Error)
Optimizer	Adam

Anomaly Scoring Pipeline

Original Image − Reconstructed Image → Per-Pixel Error Map → 4×4 Patch Max Pool → Anomaly Score

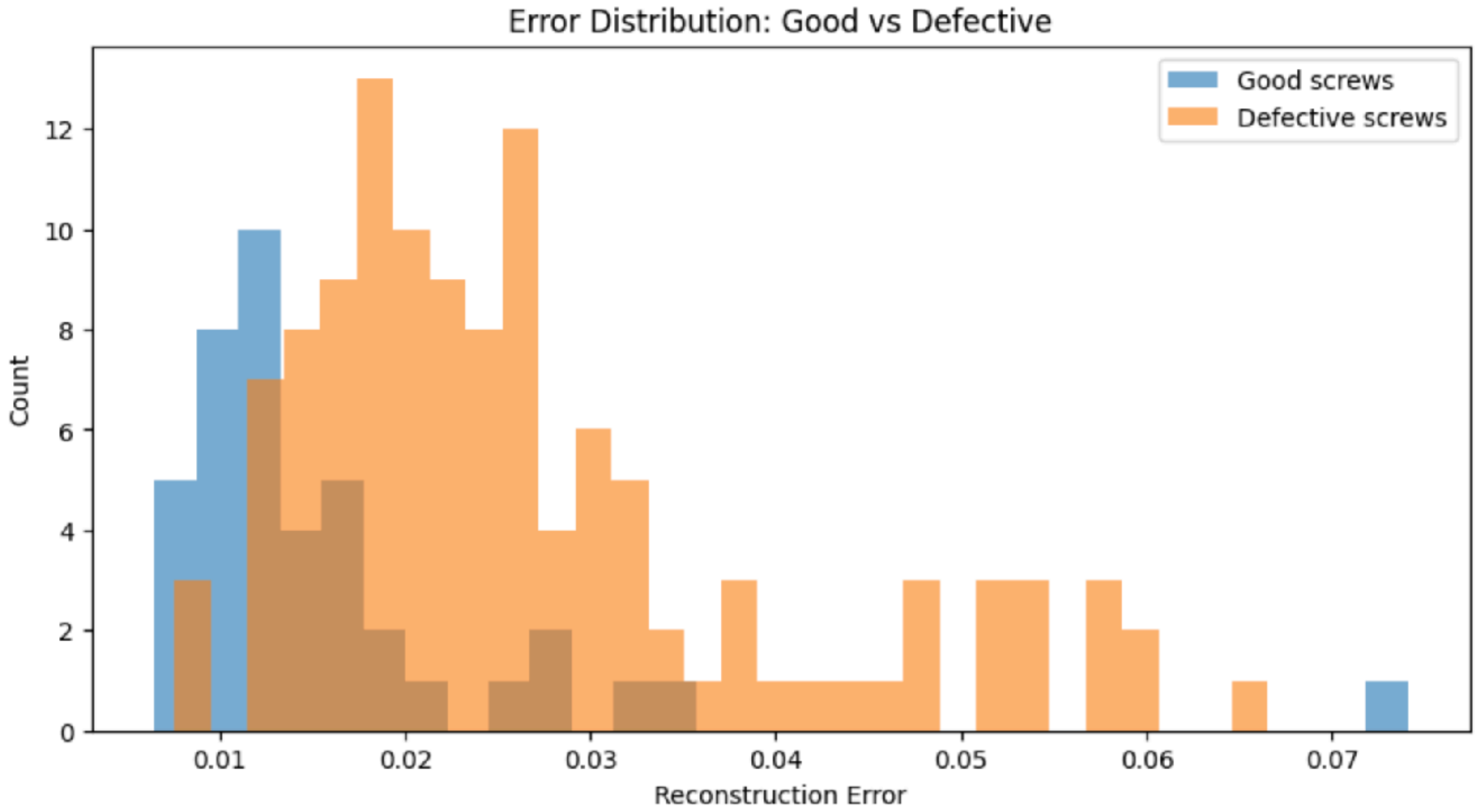
4. Results

Screw Reconstruction Error

Defect Type	Mean Error	Separation Ratio
Good (normal)	0.016039	1.00x
Thread top	0.033566	2.09×
Scratch Neck	0.027146	1.69×
Scratch Head	0.027053	1.69×
Manipulated Front	0.024414	1.52×
Thread Side	0.022231	1.39×

Thread top defects are detected with a **109% higher** reconstruction error than normal screws, while thread side defects are the hardest to distinguish, with only **39% higher** error than normal.

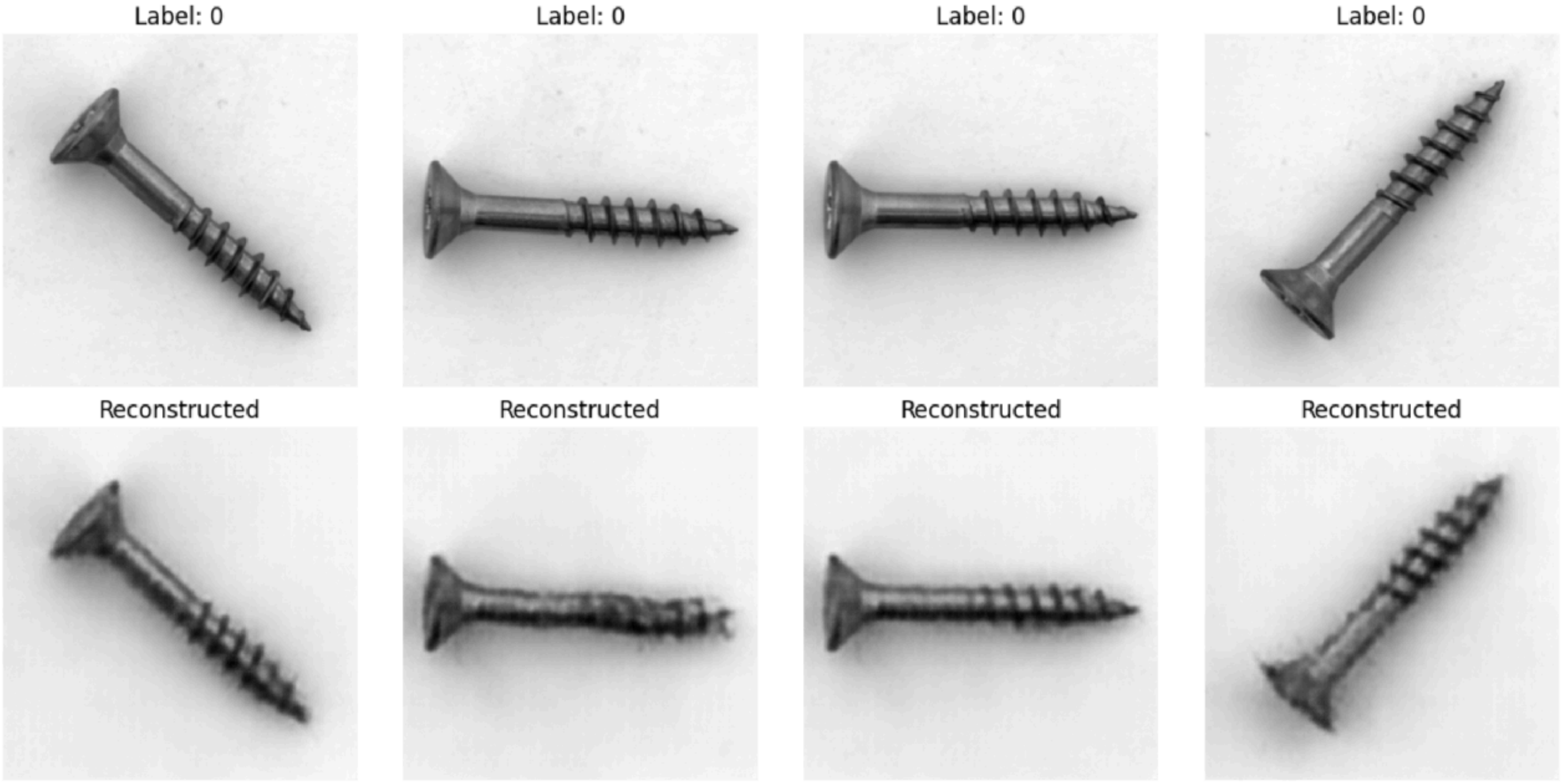
4. Results



Error Distribution: Good vs Defective
Good screws
Defective screws
Count
Reconstruction Error
0
2
4
6
8
10
12
0.01
0.02
0.03
0.04
0.05
0.06
0.07

4. Results

Real vs Reconstructed Images



Conclusion

- There is an overlap in reconstruction error between 0.01 and 0.02, where good and defective screws produce similar scores. In this range, the model cannot reliably classify a screw as normal or defective.
- The model performs best at detecting thread top defects (with a 2.09× error separation), as these involve structural deformation.
- Surface-level defects like scratches (1.69×) and thread side (1.39×) are harder, probably because their texture patterns remain similar to normal screws, producing lower reconstruction error
- This unsupervised approach, training only with normal samples with no defect labels, is practical for real manufacturing scenarios where defective examples are rare and expensive to produce.

6. Reference to Jupyter Notebook

Link to Google Colab

- https://colab.research.google.com/drive/1koB3IC4lZOU5BAzIYFnZnj2CYVvIC83x?usp=sharing