Image Anomaly Detection Using Normal Data Only by Latent Space Resampling

Introduction

- Anomalous instances are generally rare
- Normal instances account for a significant proportion
- It is almost impossible to capture a large number of abnormal data containing all anomaly types

Related Work

- **Feature Extraction Based Methods**
  - They generally map images to the appropriate feature space and detect anomalies based on distance;
  - Generally, feature extraction and anomaly detection are disjointed;
  - There are two common practices for applying CNN for feature extraction, one is to use a pre-trained network:
    - Pre-trained network, such as VGG or ResNet
    - Deep feature extraction model specifically trained for the purpose

Related Work

- **Probability Based Method**
  - These methods assume that anomalies occur in low probability regions of the normal data;
  - Establish the probability density function (PDF) of normal data;
  - Evaluate the test samples by PDF and low probability density values are most likely to be abnormal;
  - *Gaussian*, *Gaussian Mixture Model* (GMM) or *Markov Random Fields* (MRF)

Related Work

- **Reconstruction Based Method**
  - Normal images can be reconstructed from latent space better than anomalous images
  - AE, VAEs, GANs trained on normal images

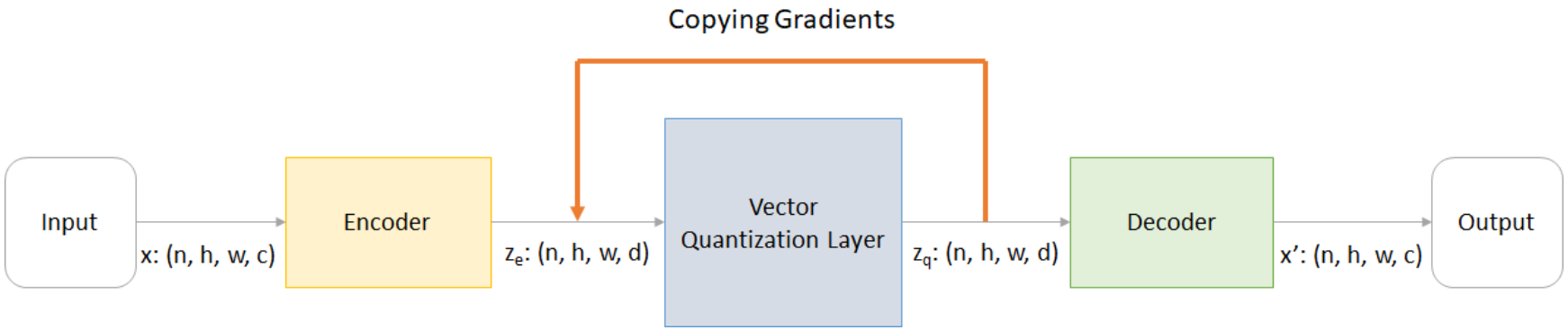
Ingredients

- **Vector Quantized Variational AutoEncoder** (VQ-VAE)
  - *Neural Discrete Representation Learning* (DeepMind, NIPS 2017)
  - https://blog.usejournal.com/understanding-vector-quantized-variational-autoencoders-vq-vae-323d710a888a

- **PixelSNAIL**
  - *PixelSNAIL: An Improved Autoregressive Generative Model* (UC Berkeley, PMLR 2018)

VQ-VAE

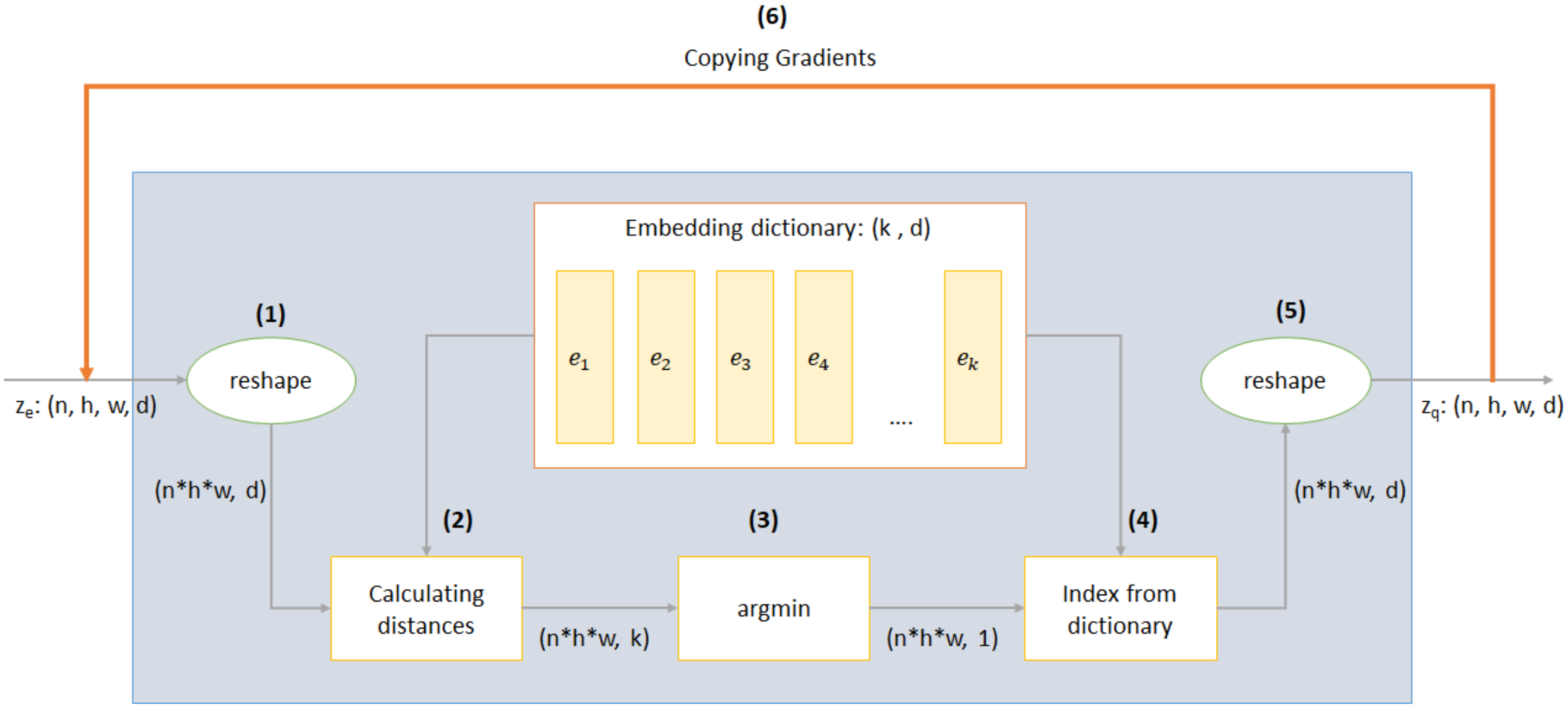
- VAE that uses a **vector quantization** to obtain a **discrete** latent representation
  - The encoder outputs a discrete code (k categories)
  - A variable can take one of the k categories
  - Each category has a separated probability



Steps to quantization

1. Reshape (b, h, w, d) → (b x h x w, d)
2. Compute the matrix of distances (v, k)
3. Index of closest of the k vectors (argmin)
4. Create the index table
5. Reshape to (b, h, w, d)
6. Copying gradient (argmin is not derivable)

Overview of Quantization



Loss Functions

- **Reconstruction loss**
  - To optimize the Encoder and the Decoder
  - To encourages the output to be as close to the input as possible

- **Codebook loss**
  - L2-based error loss to move embedding vectors towards the encoder output

- **Commitment loss**
  - To prevent z_e from fluctuating too frequently

PixelSNAIL

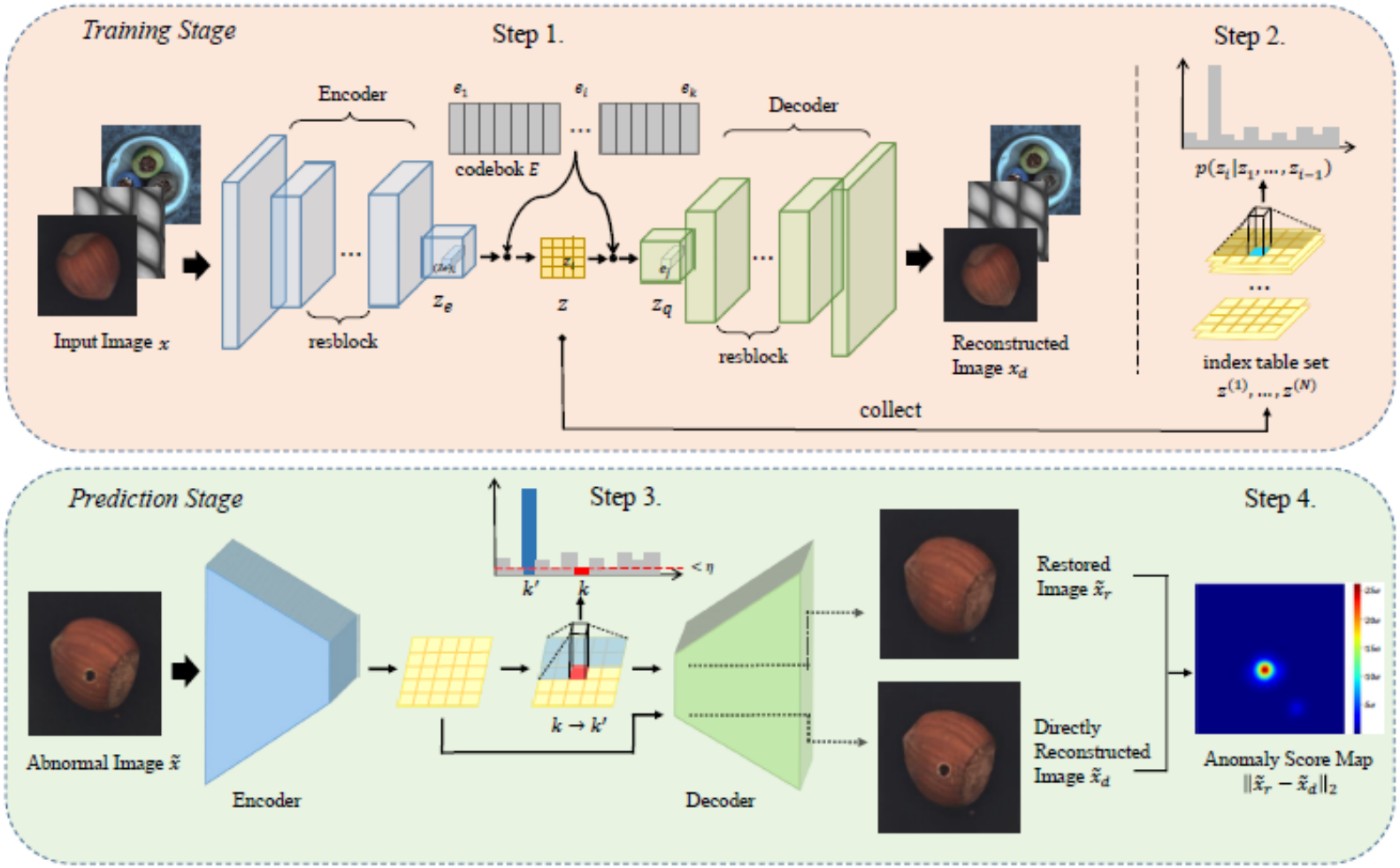
- Autoregressive generative model
- Best results in density estimation with high dimensional data
- != GANs
  - It provides likelihood
  - The training is much more stable
  - It handles both discrete and continuous data
  - Forced to capture the entire data distribution
- It combines masked convolution + self-attention

$$p(\mathbf{x}) = p\left(x_1, \ldots, x_n\right) = \prod_{i=1}^{n} p\left(x_i \mid x_1, \ldots, x_{i-1}\right)$$

Previous methods for DE

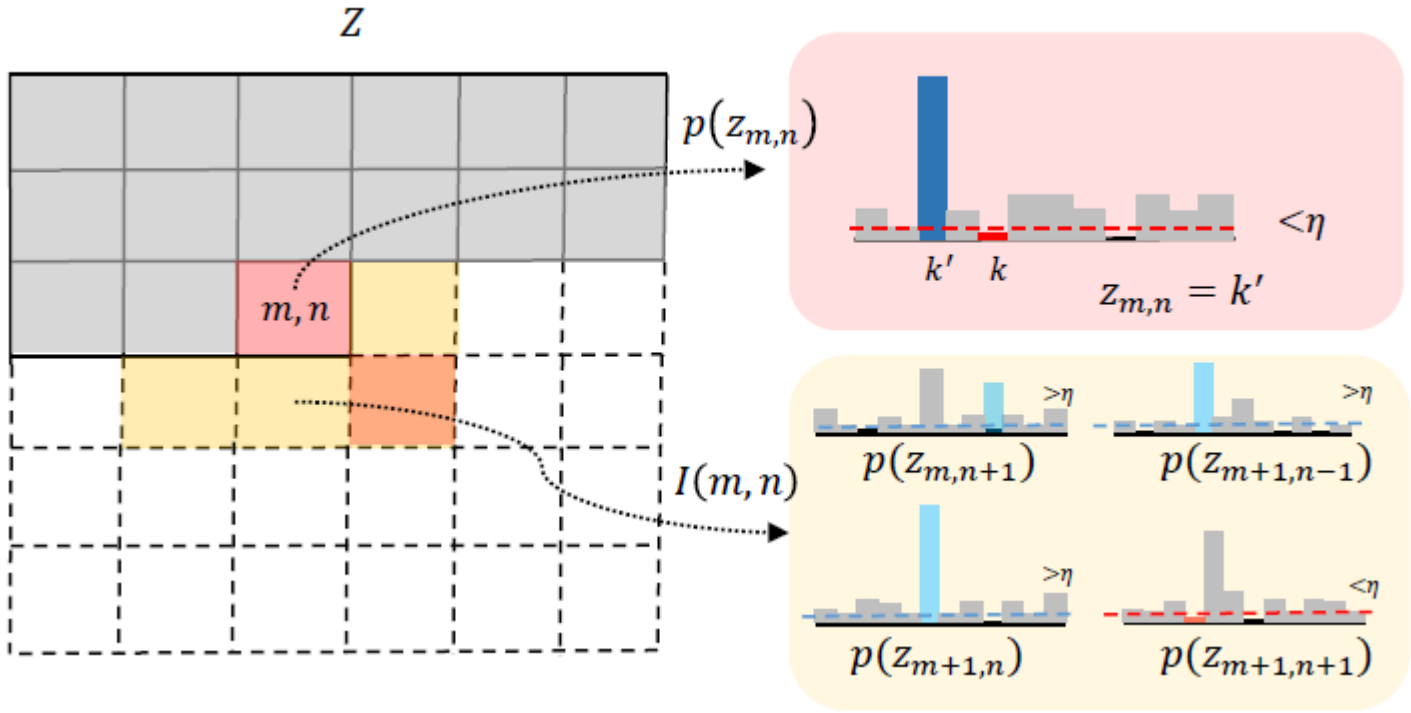
- **PixelRNN**
  - LSTM layers to capture information
- **PixelCNN**
  - Masked convolutions
- **PixelSNAIL**
  - An enhanced version of PixelCNN
  - Masked convolutions + self-attention

Proposed method



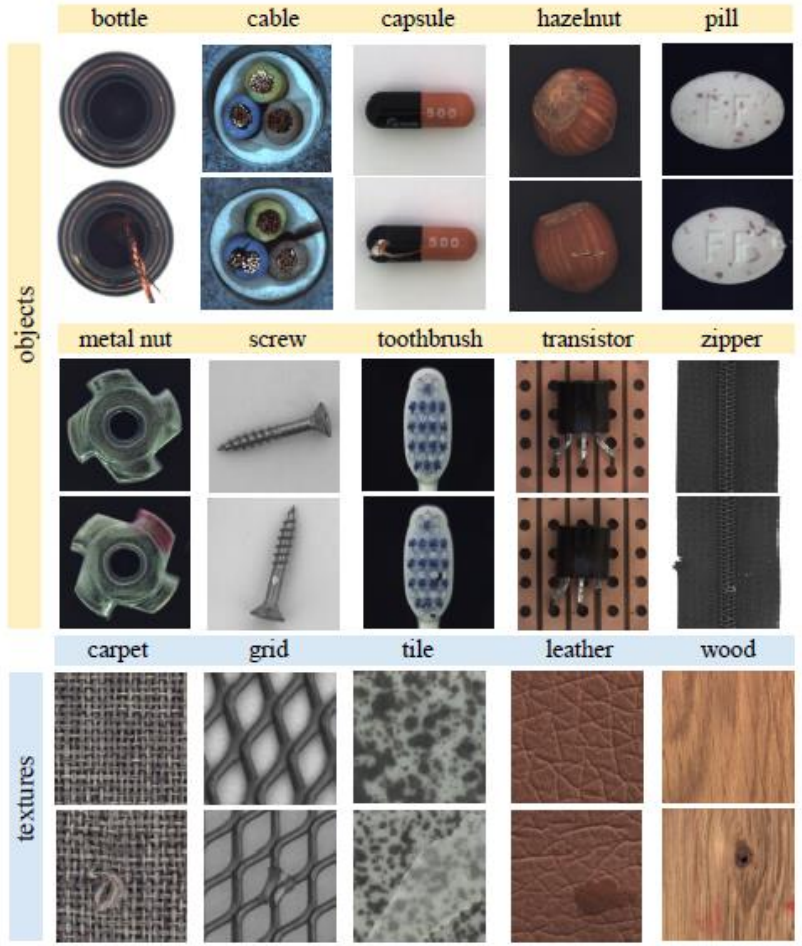
Resampling operation

- We introduce the likelihood about the bottom half (yellow zone) of the current component's 8-neighbor to avoid noise
- Resampling operation is applied when the likelihood of current component and at least one component in the bottom half is less than the threshold

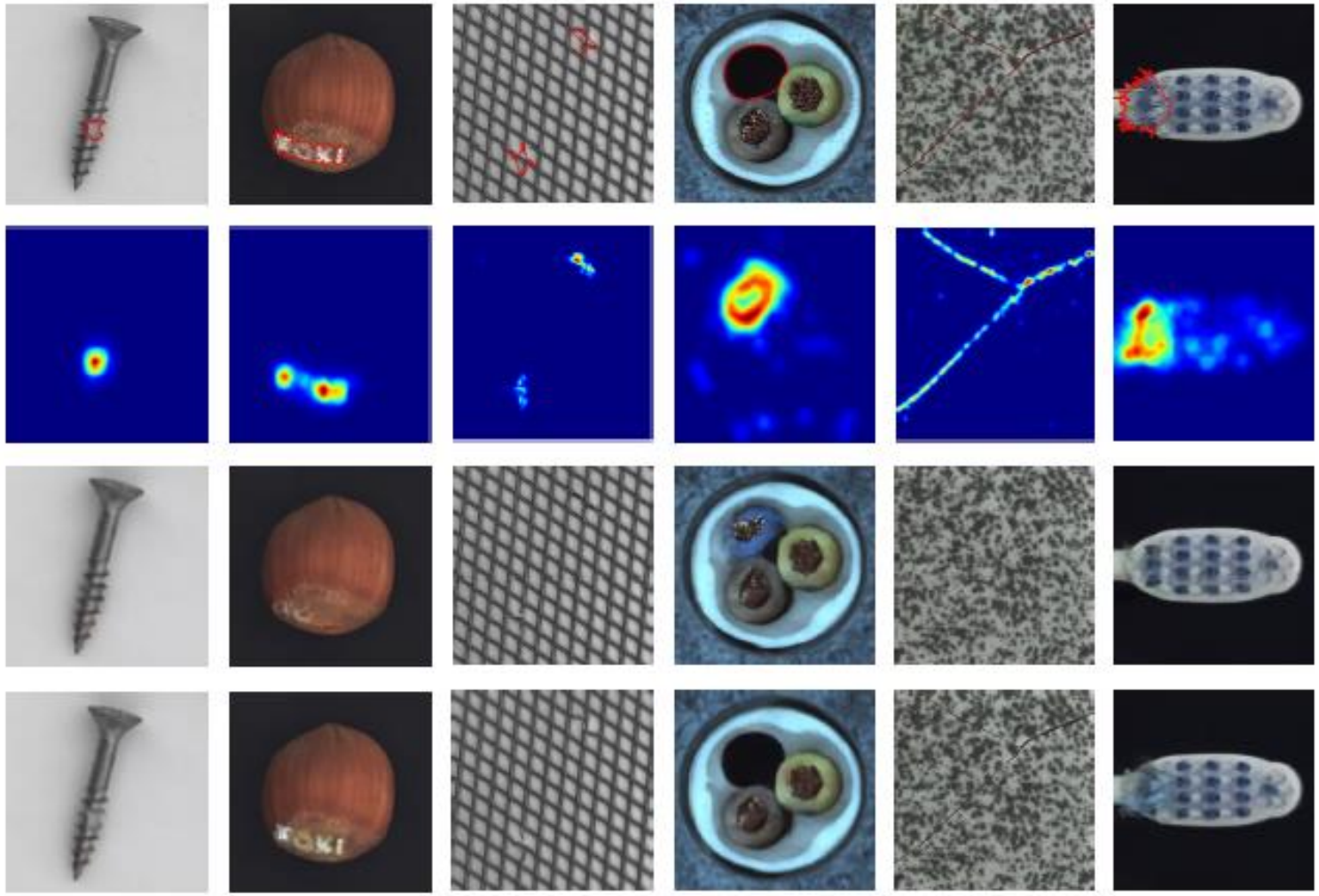


Dataset

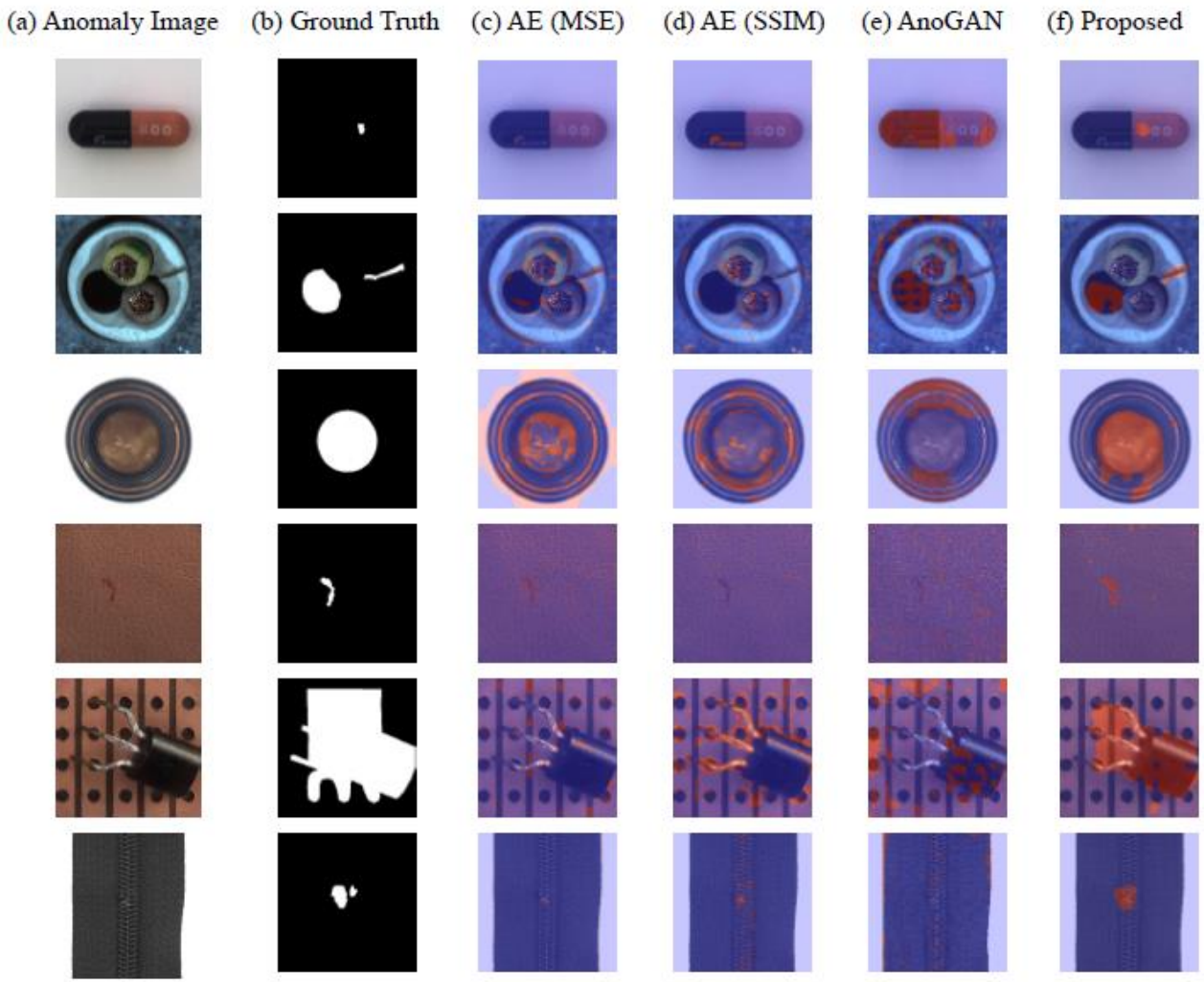
The MVTec AD dataset provides over 5000 high-resolution images divided into five texture and 10 object categories



Results



Results

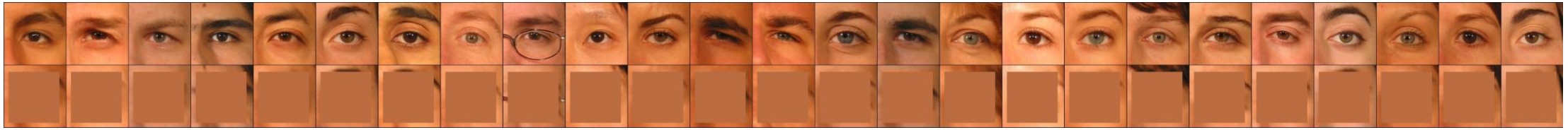

(a) Anomaly Image
(b) Ground Truth
(c) AE (MSE)
(d) AE (SSIM)
(e) AnoGAN
(f) Proposed

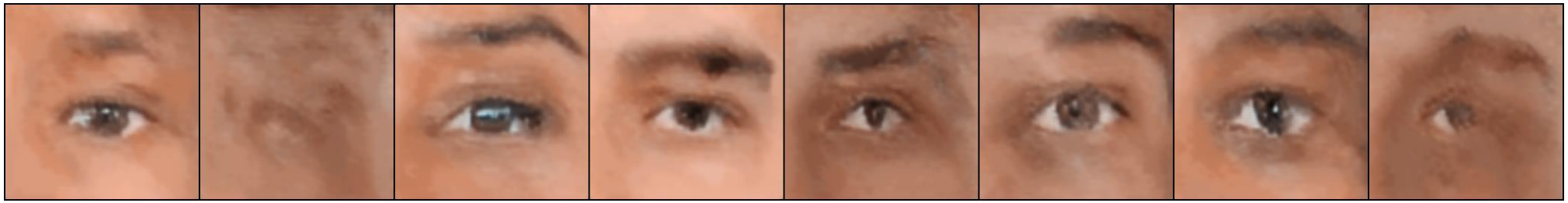
Conclusion

- Attention Map
- S-MAD
- D-MAD

- Limitations:
  - Strong computational load (PixelSNAIL)
    - GPUs
    - Video memory allocation

First results





First Results