



Tabular



Image



Video



Audio



Text

...



Vision-language



Vision-audio



Language-time series

Single Modality

Multi-Modality

Learns to reconstruct normal samples
Large reconstruction errors imply anomalies

Input

Generative Model
(Encoder)

Generative Model
(Decoder)

Output

Anomaly Score

Reconstruction-based
Methods

Generative
Anomaly
Detection

Density estimation-
based Methods

Learns the probability
distribution of normal data
Low likelihood values
indicate anomalies

Input

Generative
Model

Density
Estimation
 $P(x)$

Likelihood
 $-\log P(x')$

Anomaly Score

Representation-based
Methods

Learns compact embeddings of
data patterns
with anomalies identified via
deviations in the latent space
relative to normal representations

Input

Generative
Model

Embedding

Latent Space
Distance

Anomaly Score

$\text{dist}(z, Z)$