DISENTANGLED CONDITIONAL VARIATIONAL AUTOENCODER FOR UNSUPERVISED ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Recently, generative models have shown promising performance in anomaly detection tasks. Specifically, auto-encoders learn representations of high-dimensional data, and their reconstruction ability can be used to assess whether a new instance is likely to be anomalous. However, the primary challenge of unsupervised anomaly detection (UAD) is in learning appropriate disentangled features and avoiding information loss, while incorporating known sources of variation to improve the reconstruction. In this paper, we propose a novel architecture of generative auto-encoders by combining the frameworks of β-VAE, conditional variational auto-encoder (CVAE), and the principle of total correlation (TC). We show that our architecture improves the disentanglement of latent features, optimizes TC loss more efficiently, and improves the ability to detect anomalies in an unsupervised manner with respect to high-dimensional instances, such as in imaging datasets. Through both qualitative and quantitative experiments on several benchmark datasets, we demonstrate that our proposed method excels in terms of both anomaly detection and capturing disentangled features. Our analysis underlines the importance of learning disentangled features for UAD tasks.