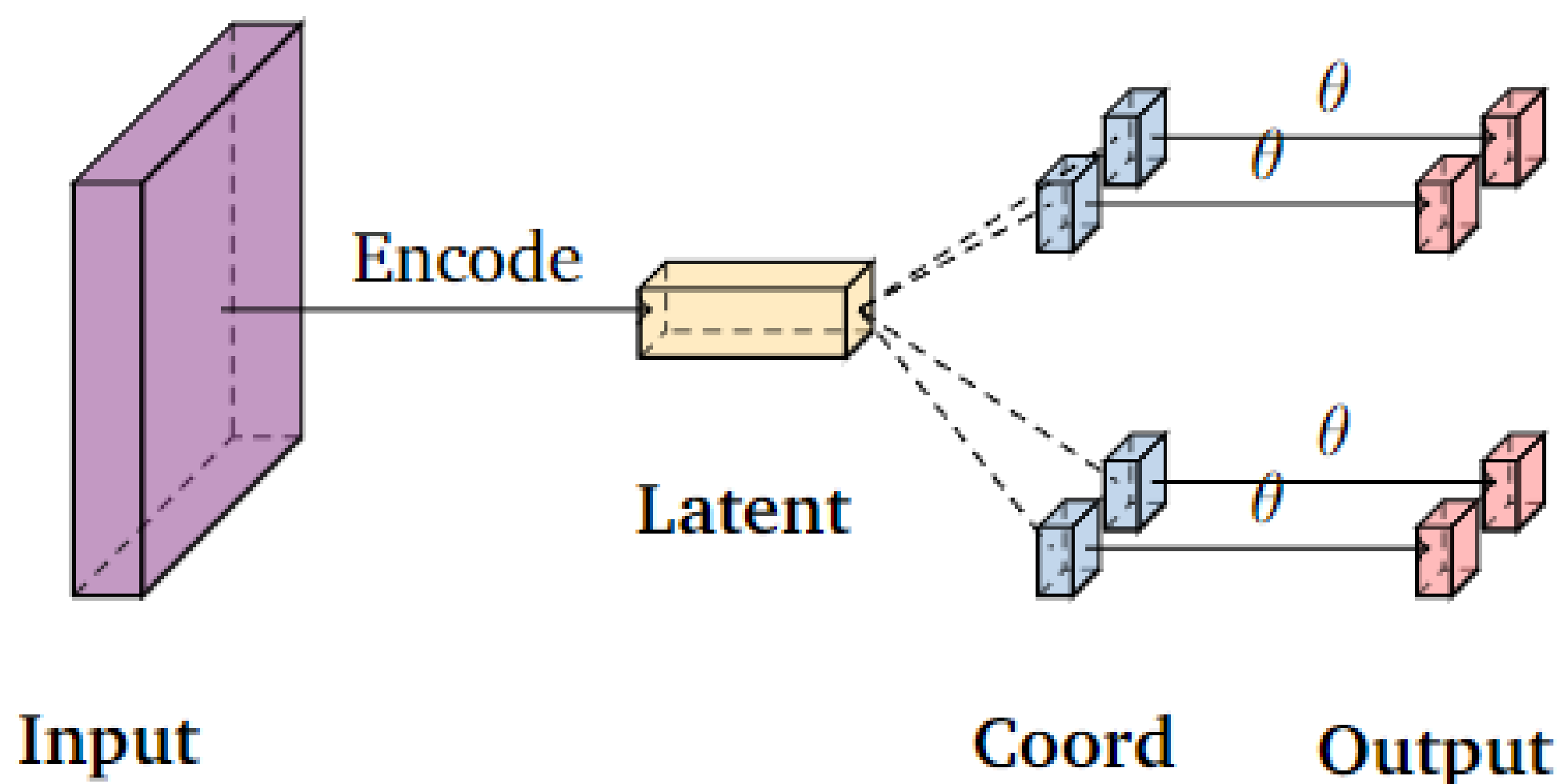


Likelihood Oriented Variational AutoEncoder for Representation Learning

Ningzhi Wang

2022



Finding

Coordinate VAE eliminates spatial structure in the latent space and employs manually defined coordinates to capture inter-pixel correlations in the decoder. We observed that this representation method struggled with pitch-related tasks, as pitch information relies on combining adjacent frequency bins, which are missing in the latent representation.

Question

What information should we extract from Mel-Spectrograms, and in what situations will it be useful?