ADDIS ABABA INSTITUTE OF TECHNOLOGY
ADDIS ABABA UNIVERSITY

Computer Vision Seminar-2

Variational Autoencoders(VAEs): A Generative AI Technique for Computer Vision Applications

Ruth Tamiru(GSR/7345/16)
June 2024

Introduction

- Deep learning models excel in static image tasks but struggle with dynamic visual data like videos.
- Traditional models miss temporal relationships between frames.

Background

Challenges in Traditional Deep Learning for Sequential Data

- Difficulty capturing temporal relationships between frames
- High demand for extensive labeled video datasets, which are expensive and time consuming to collect

Variational Autoencoders(VAEs)

- VAEs generate new data resembling real-world distributions.
- Capable of capturing temporal dependencies in sequential data

VAE Architecture

Encoder

- Compresses input data into a lower-dimensional latent representation.

-Use CNNs for image data, and RNNs/LSTMs/GRUs for sequential data

Decoder

- Reconstructs original data from the latent representation

-Uses transposed convolutional layers for images and RNNs for sequential data

Loss Functions in VAEs

- Reconstruction Loss ensures accuracy of reconstructed data

- KL Divergence Loss regularizes the latent space to follow a standard normal distribution, promoting diverse data generation

Advantages of VAEs for Computer Vision

- Data Augmentation: Generates diverse variations of existing data

- Latent Space Exploration: Enables manipulation for tasks like anomaly detection and image editing

-Improved Representation Learning: Learns meaningful feature representations

State-of-the-Art Advancements in VAEs

-β-VAE: Balances reconstruction accuracy and latent space quality

-Conditional VAEs: Allows controlled data generation with additional information

-VAE-based Anomaly Detection: Identifies unusual patterns by analyzing deviations in latent space

Limitations and Challenges

-Posterior Collapse: Encoder maps all inputs to a single point in the latent space

-Computational Cost: High training costs for complex datasets

-Interpretability: Difficulty understanding the relationship between latent space and generated data

Future Directions

-Improved Training Methods: Addressing posterior collapse and optimizing training efficiency

-Hybrid Models: Combining VAEs with GANs for realistic data generation

-Interpretable VAEs: Developing techniques to understand latent space and generated data relationships

Conclusion

-VAEs enhance tasks like data augmentation, anomaly deection, and image editing

-Ongoing research aims to address current limitations and unlock the full potential of VAEs in computer vision


Thank
You