

- Video Encoder
- Audio Encoder
- Text Encoder
- Bottleneck
- Decoder
- Generic Layer

Network Architecture (55 layers, 2,043,556,210 parameters)

