

Video Coding using Deep Learning

Focus: Deep Contextual Video Compression

Gerald Schuller

TU Ilmenau

July 10, 2024

Outline

Introduction to Video Coding

Deep Learning for Video Coding

Deep Contextual Video Compression (DCVC)

Components of Deep Contextual Video Compression

- Autoencoders

- Encoder

- Decoder

- Context Model

Training Deep Video Compression Models

Evaluation Metrics

Case Study: Deep Contextual Video Compression

Future Directions

Conclusion

Introduction to Video Coding

- ▶ Definition of video coding
- ▶ Importance of video compression
- ▶ Traditional video coding techniques (e.g., MPEG, H.264)

Deep Learning for Video Coding

- ▶ Overview of deep learning
- ▶ Application of deep learning in video compression
- ▶ Advantages over traditional methods

Deep Contextual Video Compression

- ▶ Introduction to Deep Contextual Video Compression
- ▶ Key features and innovations
- ▶ Comparison with traditional methods

DCVC Comparison with traditional methods

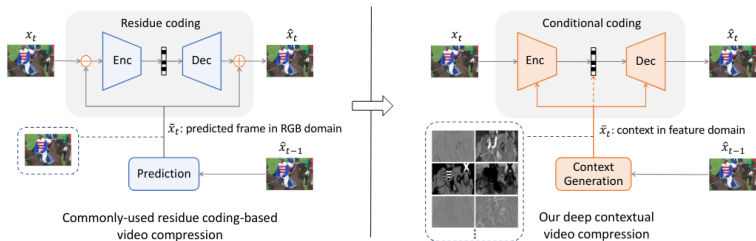


Figure 1: Paradigm shift from residue coding-based framework to conditional coding-based framework. x_t is the current frame. \hat{x}_t and \hat{x}_{t-1} are the current and previous decoded frames. The orange dashed line means that the context is also used for entropy modeling.

Figure: From: Jiahao Li, Bin Li, Yan Lu: "Deep Contextual Video Compression"

DCVC Detailed Structure

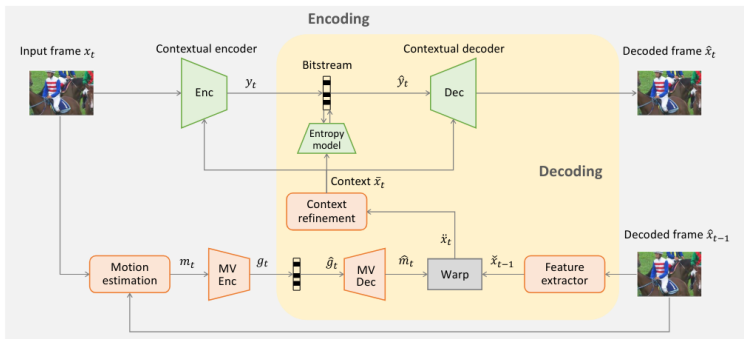


Figure: From: Jiahao Li, Bin Li, Yan Lu: "Deep Contextual Video Compression"

DCVC Bit Per Pixel vs Quality in MS-SSIM

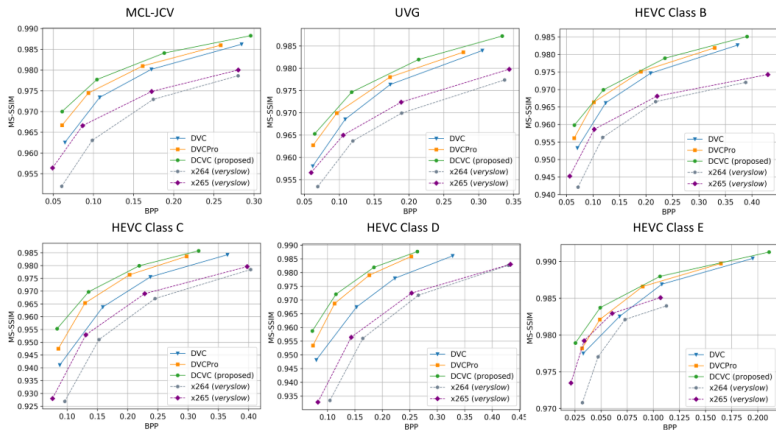


Figure 6: MS-SSIM and bitrate comparison. The DL-based codecs are fine-tuned for MS-SSIM.

Figure: DCVC gets comparable quality at roughly half the bitrate of X264 and x265. From: Jiahao Li, Bin Li, Yan Lu: "Deep Contextual Video Compression"

Autoencoders as Core Component of DCVC

- ▶ Definition and function of autoencoders
- ▶ Types of autoencoders (e.g., Variational Autoencoders, Convolutional Autoencoders)
- ▶ Role of autoencoders in video compression

Autoencoders: Definition

► Definition

- An autoencoder is a type of artificial neural network used to learn efficient codings of input data.
- It consists of two main parts: an encoder that compresses the input into a latent-space representation, and a decoder that reconstructs the input from this representation.

Autoencoders: Function

▶ **Function**

▶ **Dimensionality Reduction**

- ▶ Autoencoders reduce the dimensionality of data, making it easier to process and analyze.

▶ **Feature Learning**

- ▶ They automatically learn the most relevant features of the data without human intervention.

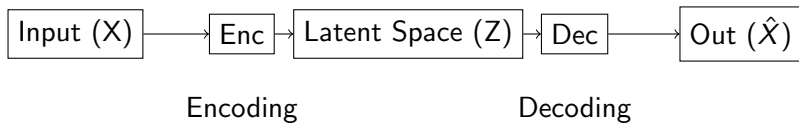
▶ **Data Reconstruction**

- ▶ By reconstructing the input data from the encoded representation, autoencoders can be used for data denoising and anomaly detection.

▶ **Compression in Video Coding**

- ▶ In video compression, autoencoders help in reducing the size of video frames by learning efficient representations, which are then used for reconstructing high-quality frames.

Structure of an Autoencoder



- ▶ **Input (X):** Original data.
- ▶ **Encoder:** Compresses the input data into a latent-space representation (Z).
- ▶ **Latent Space (Z):** Compressed representation of the data.
- ▶ **Decoder:** Reconstructs the data from the latent-space representation.
- ▶ **Output (\hat{X}):** Reconstructed data.

Autoencoder Audio Example Implementation

- ▶ A Colab notebook example for an autoencoder for audio:
- ▶ https://github.com/TUilmenauAMS/AES_Tutorial_2021
- ▶ Chapter: "The Convolutional Autoencoder Network"

Autoencoder Image Example Implementation

- ▶ A Colab notebook example for an autoencoder for images:
- ▶ <https://github.com/TUilmenauAMS/Videocoding/blob/main/imageVAE.ipynb>

DCVC Github

- ▶ DCVC can be installed from the following repository,
- ▶ It includes several versions and links to training checkpoints for psnr and msssim
- ▶ <https://github.com/microsoft/DCVC>

Encoder

- ▶ Structure of the encoder
- ▶ Encoding process
- ▶ Techniques for improving encoding efficiency

Decoder

- ▶ Structure of the decoder
- ▶ Decoding process
- ▶ Techniques for improving decoding accuracy

Context Model

- ▶ Definition of context in video compression
- ▶ Building a context model
- ▶ Using context to enhance compression

Training Deep Video Compression Models

- ▶ Data preparation
- ▶ Training process
- ▶ Loss functions and optimization

Evaluation Metrics

- ▶ Common metrics (e.g., PSNR, SSIM)
- ▶ Subjective evaluation
- ▶ Benchmark datasets and comparisons

Case Study: Deep Contextual Video Compression

- ▶ Overview of a specific implementation
- ▶ Results and performance analysis
- ▶ Lessons learned and future directions

Future Directions

- ▶ Emerging trends in video compression
- ▶ Potential improvements with deep learning
- ▶ Research challenges and opportunities

Conclusion

- ▶ Summary of key points
- ▶ Impact of deep learning on video compression
- ▶ Final thoughts

Questions?