Complexity-guided Slimmable Decoder for Efficient Deep Video Compression

Zhihao Hu[1], Dong Xu[2]

[1]School of Computer Science and Engineering, Beihang University, China
[2]Department of Computer Science,The University of Hong Kong, China

Motivation

- Recent learning-based video codecs outperform commercial codecs (*e.g.*, H.265).
- Current learning-based video compression systems[1,2,3]
  - Always inefficient due to computationally complex operations
- In practical application scenarios
  - It is desirable that the video codecs can decode the videos in real-time.
  - Decoders from different devices can afford different computational complexities under different scenarios.
    - Cloud server (higher computational resource), Smartphone (less computational resource)

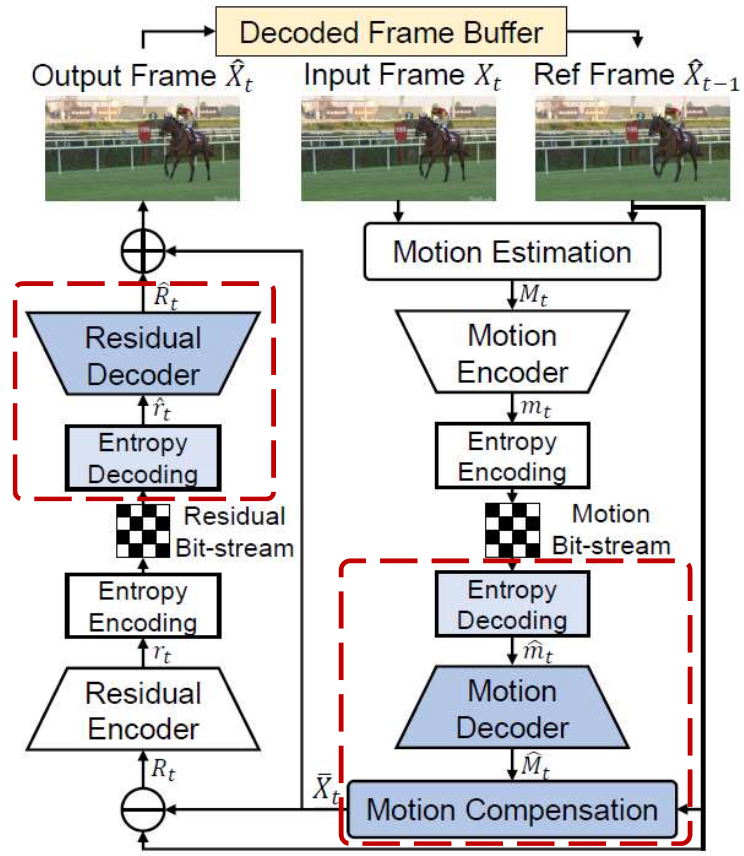
- [1] Lu, Guo, et al. "DVC: An end-to-end deep video compression framework", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.
- [2] Hu, Zhihao, et al. "FVC: A new framework towards deep video compression in feature space", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2021.
- [3] Li, Jiahao, et al. " Deep contextual video compression", Advances in Neural Information Processing Systems. 2021.

Overview

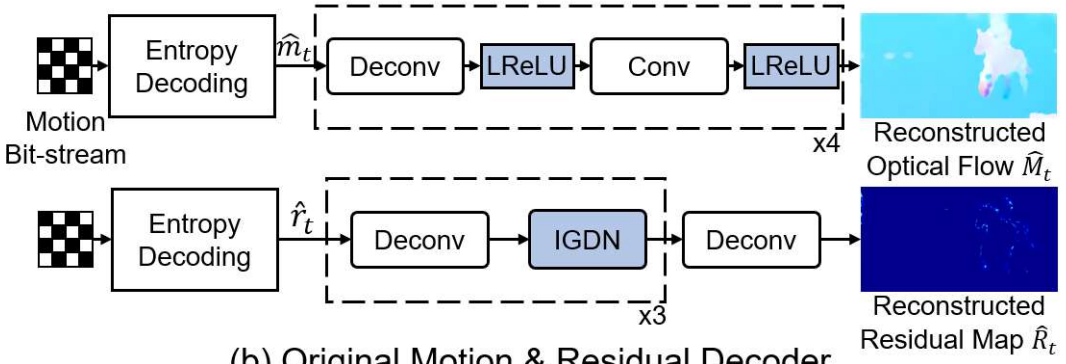
- Complexity-guided slimmable decoder (cgSlimDecoder)
  - For efficient video decoding
  - Support multiple complexity levels by simply using one learned decoder
  - Automatically allocate the optimal complexities for different modules

- Skip-adaptive Entropy Coding (SaEC)
  - For more efficient and effective entropy coding.

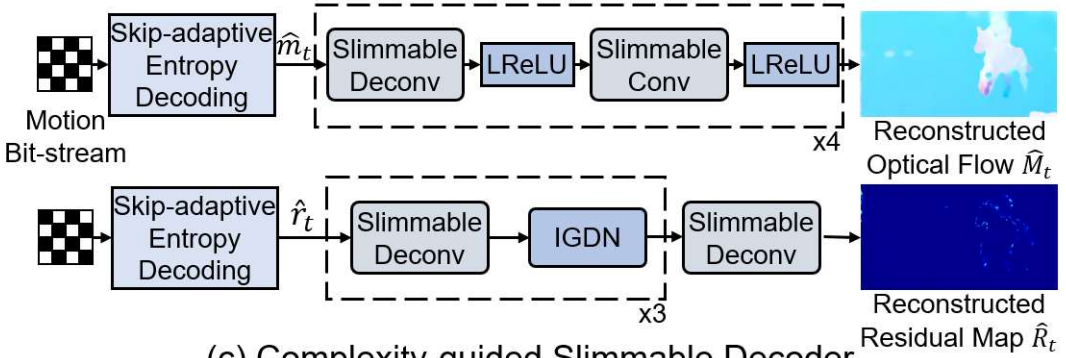
Complexity-guided Slimmable Decoder



(a) Mainstream Deep Video Coding Framework



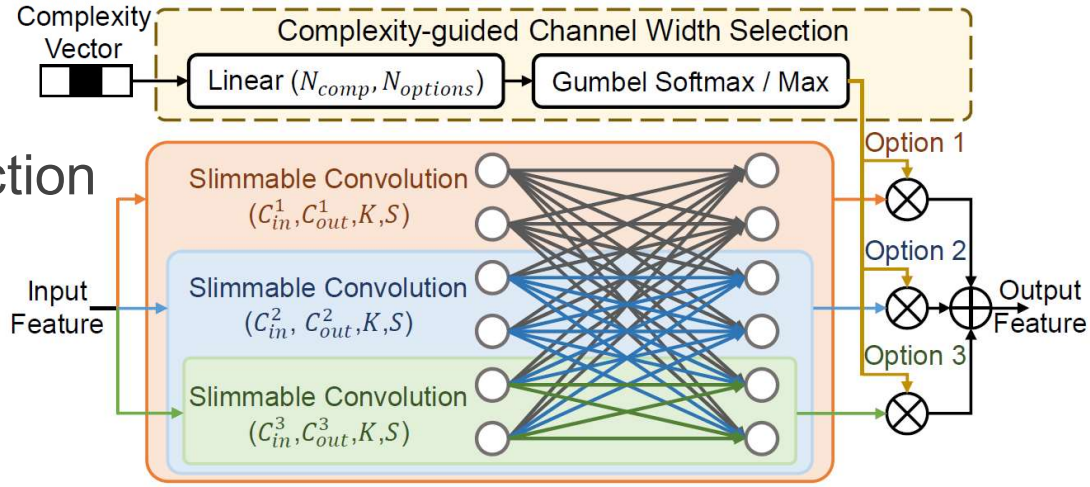
(b) Original Motion & Residual Decoder



(c) Complexity-guided Slimmable Decoder

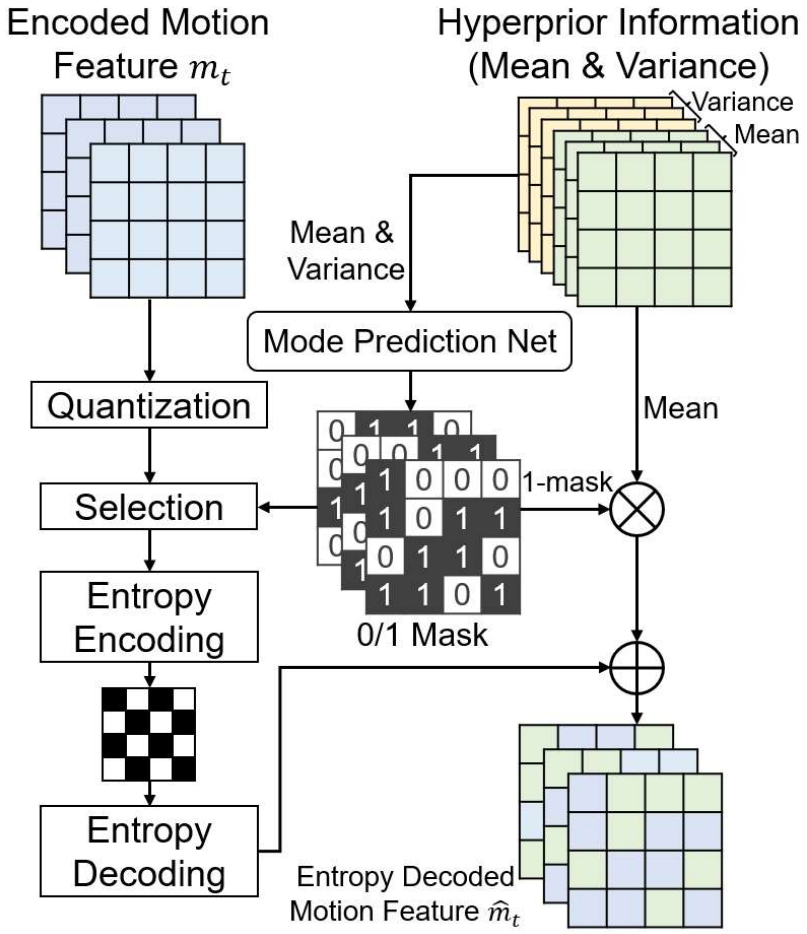
Complexity-guided Slimmable Convolution

- Complexity Vector
  - Current complexity constraint
- Complexity-guided channel width selection
  - Based on the Gumbel Softmax
  - Decide optimal channel width
- Computational resource is sufficient
  - Larger channel width & high-quality video sequences
- Computational resource is limited
  - Smaller channel width & more efficient decoding

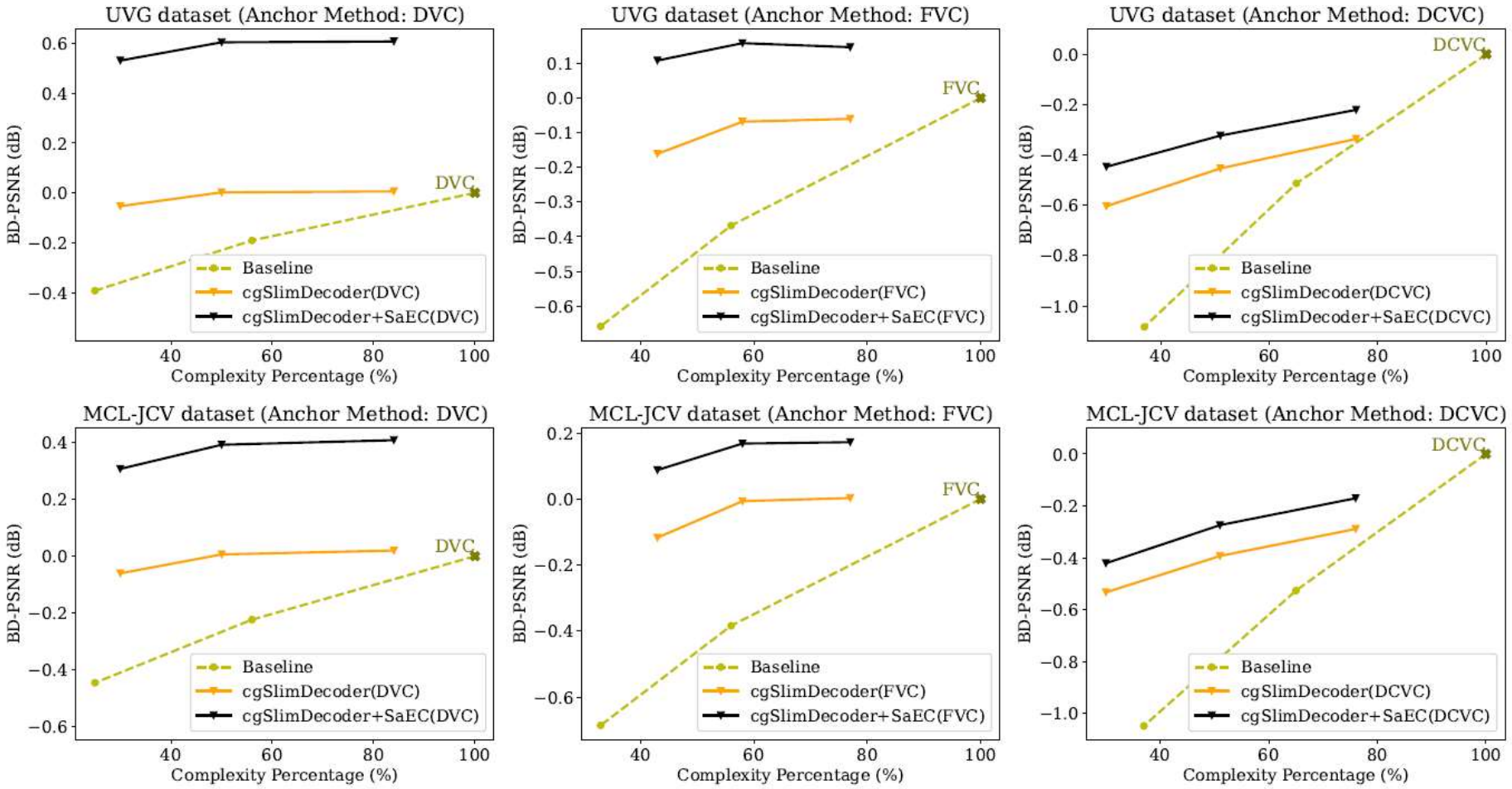


Skip-adaptive Entropy Coding (SaEC)

- Efficiency of entropy coding should be considered
- Automatically select the coding mode (*i.e.*, the skip mode)
  - For each element of the encoded motion feature
  - “1” : entropy coded
  - “0” : directly use the mean value



Performance



Complexity Percentage of Different Modules

- DCVC as an example

Complexity Levels	Original	Level 1	Level 2	Level 3
Motion Decoder	7%	1%	2%	3%
Motion Refinement	25%	27%	34%	18%
Feature Extraction	10%	13%	15%	25%
Context Refinement	15%	5%	8%	13%
Context Encoder	5%	3%	4%	4%
Contextual Decoder	39%	51%	38%	37%

Skip Percentage of our Skip-adaptive Entropy Coding (SaEC)

- FVC as an example
  - Motion
    - More than 99% are skipped
  - Residual
    - More than 85% are skipped

	λ=2048	λ=1024	λ=512	λ=256
Motion	99.60%	99.74%	99.79%	99.83%
Residual	85.90%	91.54%	94.27%	96.21%

Decoding Time

- Decoding Time on 1080p videos

	Original	Level 1	Level 2	Level 3
DVC	163ms	140ms	107ms	79ms
FVC	127ms	107ms	89ms	71ms
DCVC	283ms	237ms	189ms	144ms

Thank you for watching