Video Compression with Arbitrary Rescaling Network

Mengxi Guo, Shijie Zhao*, Hao Jiang, Junlin Li and Li Zhang

Bytedance Inc.
San Diego, CA, 92122 USA and Shenzhen, China
{*guomengxi.qoelab*, *zhaoshijie*.0526, *jianghao*, *lijunlin.li*, *lizhang.idm* }*@bytedance.com*

We propose a practical downsampling compression scheme, as shown in Fig 1, to use a neural network as a pre-processing module for traditional codecs to improve compression performance. Specifically, we propose a neural network-based pre-processing method to improve video compression performance by downsampling high-resolution videos. Our method, called the rate-guided arbitrary rescaling network (RARN), uses a pre-trained entropy module [1] to estimate bitrate and guide the sampling process. We also use a transformer-based virtual codec (TVC) to simulate the performance of traditional codecs, using a swing-transformer-based invertible neural network to learn distortion from standard codecs, and the cyclic shift attention [2] in the network can approximate the prediction modes of HEVC. Our method, which is compatible with standard codecs (HEVC and VVC), performs well in various sampling ratios according to experimental results.
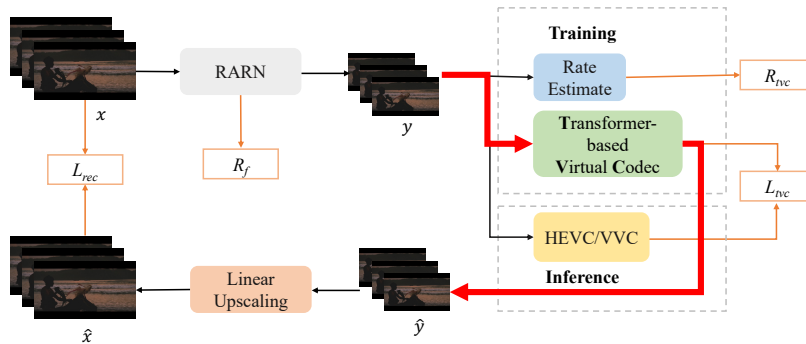

Figure 1: The architecture of the proposed framework. During training, the downsampled frames y are fed into the Transformer base Virtual Codec along the red arrow direction.

[1] Johannes Ballé, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston, "Variational image compression with a scale hyperprior," *arXiv preprint arXiv:1802.01436*, 2018.

[2] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10012–10022.

---

*Corresponding author.