Paper ID: mmfp1683

Structure-Preserving Motion Estimation for Learned Video Compression

Han Gao, Jinzhong Cui, Mao Ye, Shuai Li, Yu Zhao, Xiatian Zhu
University of Electronic Science and Technology of China, Shandong University, University of Surrey

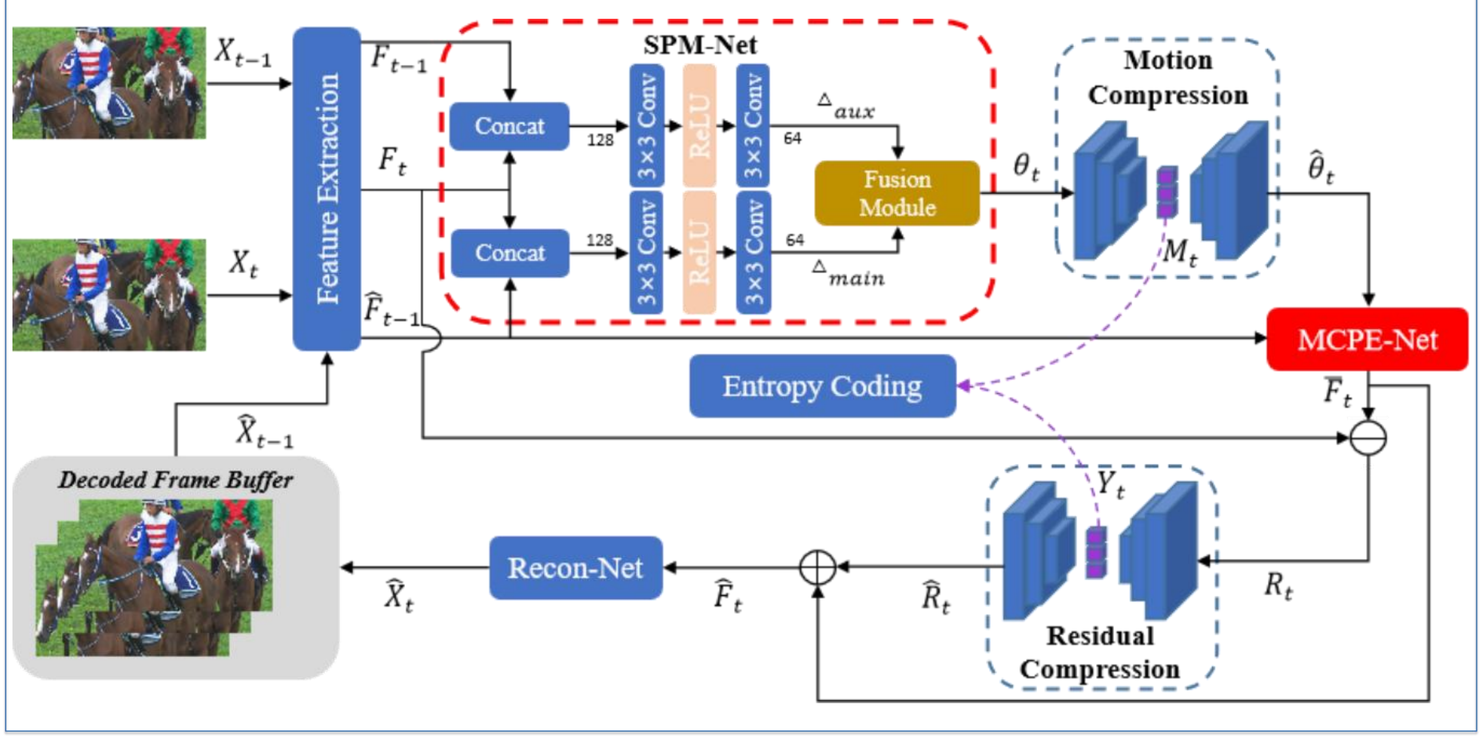
Problem & Proposed method

» Problem statement

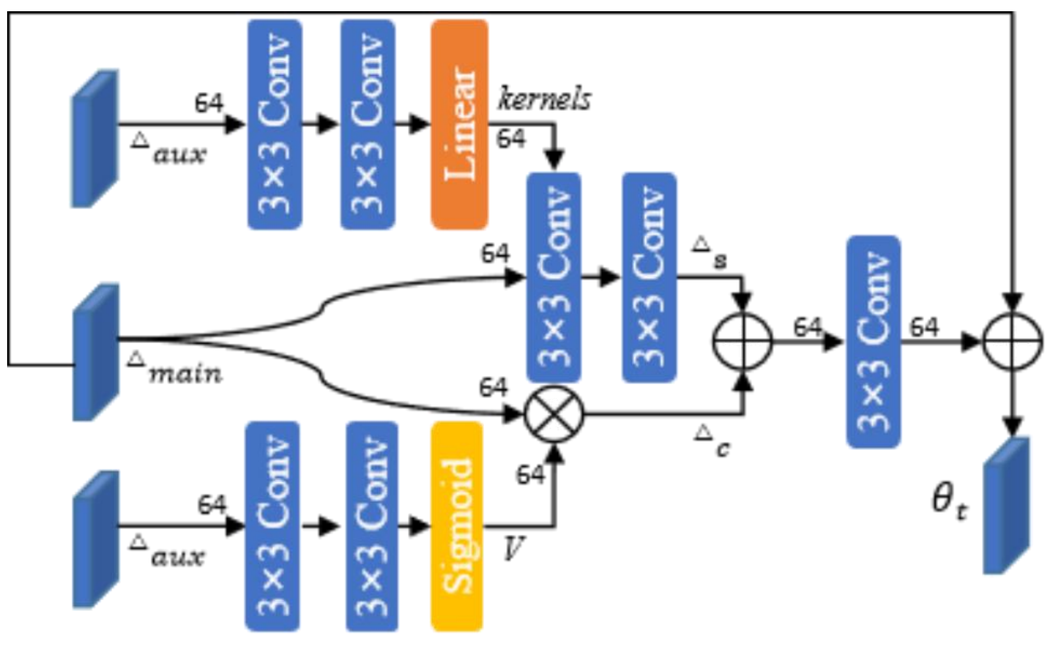
Method	Illustration	Characteristic
Previous & conventional	X_t, $\hat{X}_{t-1}$ → Motion Estimation	• Damage both the spatial structure of motion inferred and the corresponding residual; • Break the consistent nature across frames since the estimated motion is no longer consistent with the movement in the original video due to the distortion.
Ideal but not corresponding	X_t, X_{t-1} → Motion Estimation	• Spatial structure is preserved when estimating motion; • Non-correspondence at motion compensation stage.
Proposed	$\hat{X}_{t-1}$, X_t, X_{t-1} → Motion Estimation	• Spatial structure is preserved when estimating motion; • Without any non-correspondence at motion compensation stage.

How to use both of them efficiently?

» Overview of proposed method



» Architecture of *Fusion Module*



$$\begin{cases} \Delta_s = Conv_{3\times3} \circ Conv_k(\Delta_{main}),\ kernels = Linear \circ (Conv_{3\times3})^2(\Delta_{aux}), \\ \Delta_c = V \otimes \Delta_{main},\ \ V = Sigmoid \circ (Conv_{3\times3})^2(\Delta_{aux}), \end{cases}$$

$$\theta_t = \Delta_{main} + Conv_{3\times3}(\Delta_s + \Delta_c).$$

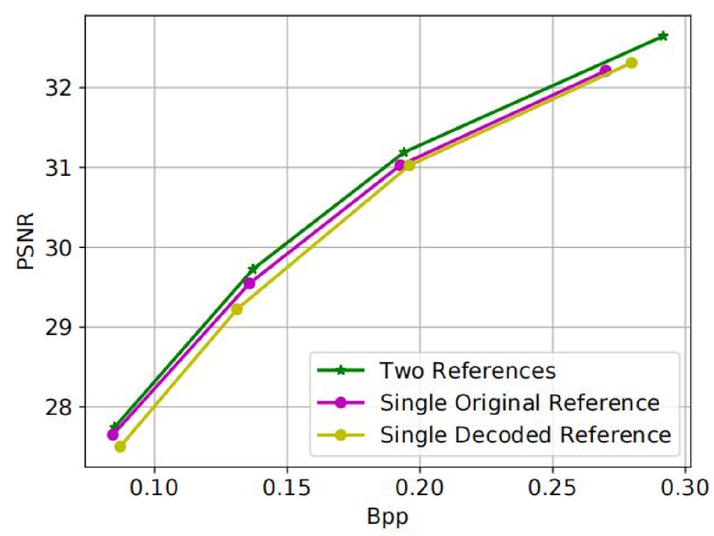
Experiments

» Experimental results

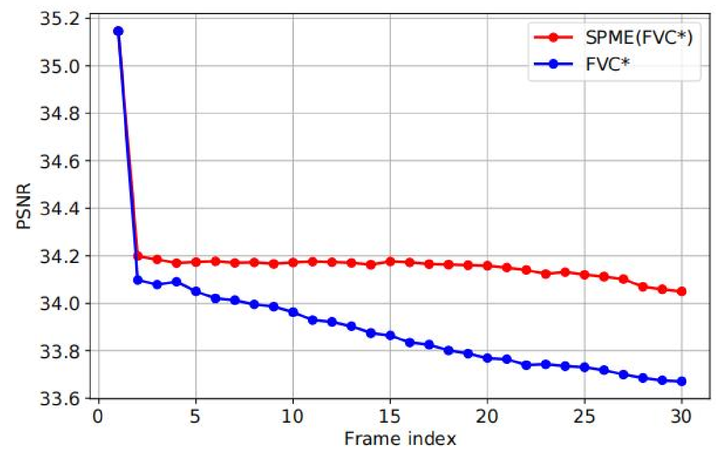
	FVC*	SPME (FVC*)	DCVC	SPME (DCVC)
HEVC Class B	-21.45	**-26.78**	-35.59	**-39.53**
HEVC Class C	-2.14	**-9.06**	-14.88	**-18.93**
HEVC Class D	-16.55	**-21.60**	-26.26	**-29.98**
HEVC Class E	8.31	**-11.08**	-17.69	**-21.04**
MCL-JCV	16.12	**-3.16**	-28.78	**-31.70**
UVG	-12.82	**-16.81**	-37.74	**-41.30**

BDBR results in terms of PSNR.

» Ablation studies



Effect of different reference frames.



The change of PSNR with time inside a GoP.

Conclusion and others

» Conclusion

- Identify a generic limitation of motion estimation in learned video compression;
- Propose a plug-and-play method to use the original previous frame as auxiliary data for motion estimation.

» Citation

Han Gao, Jinzhong Cui, Mao Ye, Shuai Li, Yu Zhao, and Xiatian Zhu. 2022. Structure-Preserving Motion Estimation for Learned Video Compression. In Proceedings of the 30th ACM International Conference on Multimedia (MM'22), October 10–14, 2022, Lisbon, Portugal. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3503161.3548156

» Reproducibility

Code and Appendix are available at: https://github.com/gaohan-12/SPME.

» Contact

Han Gao: han.gao@std.uestc.edu.cn