Flow-Guided Temporal-Spatial Network for HEVC Compressed Video Quality Enhancement

Xiandong Meng[1], Xuan Deng[2], Shuyuan Zhu[2], Shuaicheng Liu[2] and Bing Zeng[2]

[1] Hong Kong University of Science and Technology

[2] University of Electronic Science and Technology of China

In this work, a flow-guided temporal-spatial network (FGTSN) is proposed to enhance the quality of HEVC compressed video. The framework of our proposed FGTSN method is shown in Fig. 1. Specially, we first employ a motion estimation subnet via trainable optical flow module to estimate the motion flow between current frame and its adjacent frames. Guiding by the predicted motion flow, the adjacent frames are aligned to current frame. Then, a temporal encoder is designed to discover the variations between current frame and its warped frames. Finally, the reconstruction frame is generated by training the model in a multi-supervised fashion. Our proposed method takes advantage of temporal-spatial information to enhance the Quality Of HEVC compressed video. Experimental results demonstrate the superior performance of our FGTSN method.

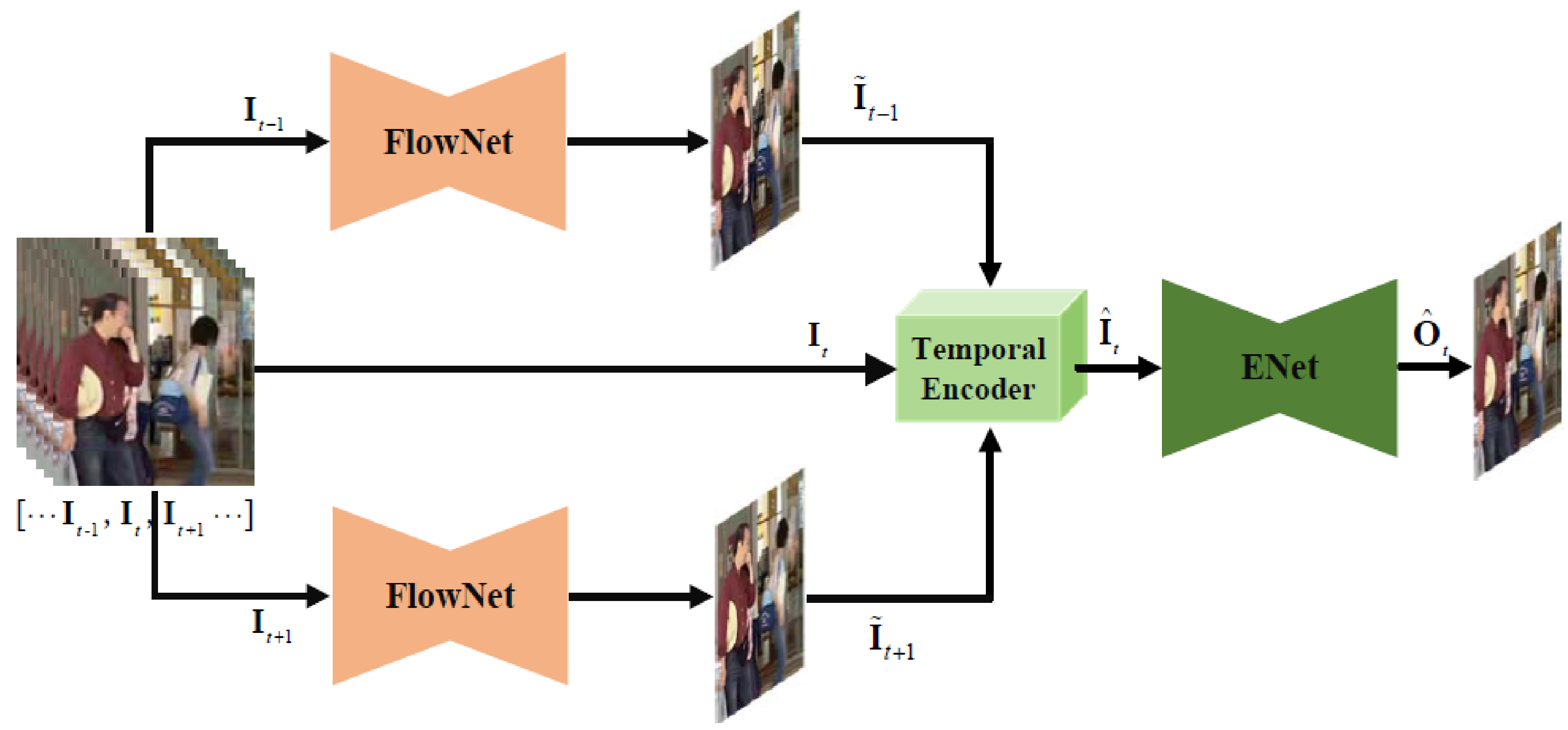

Fig. 1. The framework of our proposed FGTSN method

As shown in Fig. 2, a quality enhancement subnet (ENet) with multi-scale encoder-decoder structure is designed to explore the spatial information. The encoder consists of four convolutional layers with stride equal to 1 and four convolutional layers with stride equal to 2. Convolutions are directly followed down-sampling steps double the number of feature channels at each subsequent scale. The decoder consists of repeated application of a deconvolution that halves the number of feature channels, then concatenation with the cropped encoder feature map at corresponding resolution. Each set of decoder activations is passed through another depth-wise convolution layer to generate an intermediate prediction at its resolution.

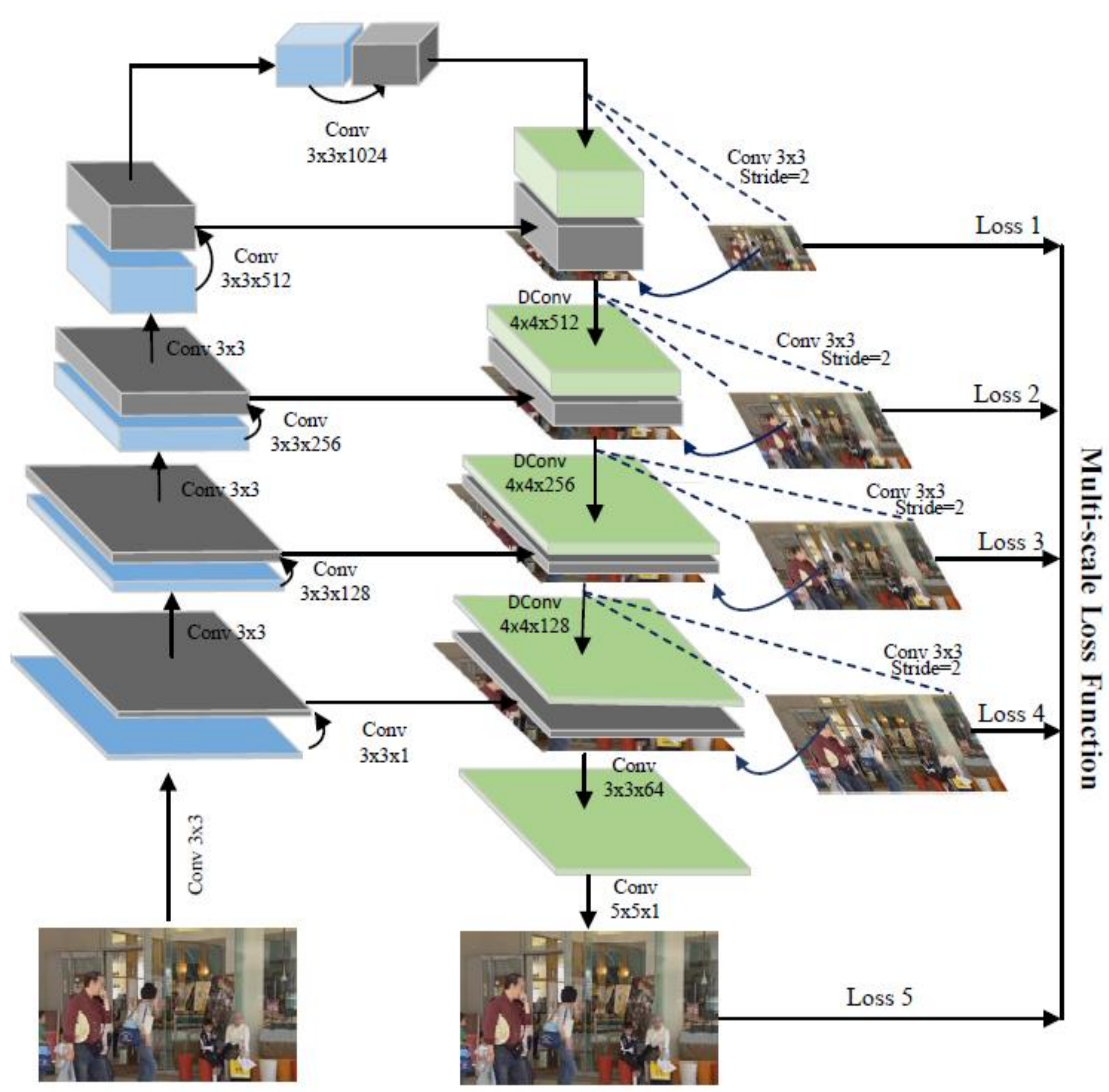

Fig. 2. The proposed multi-scale ENet structure

We implement our framework on a PC with Intel Xeon E5 CPU and NVidia GeForce GTX 1080Ti GPU. For fairness, all experiments are conducted on the same dataset, and the comparison methods are also retrained over the same training dataset according to authors' recommended parameters. 18 sequences of Classes A~E with different resolutions proposed by the Joint Collaborative Team on Video Coding (JCT-VC) are used in the experiments. The data set is compressed by HEVC reference software, HM16.9, under LD configuration. We set Quantization Parameters (QPs) to 37 and 42. The results are shown in Table 1.

Table 1. Overall ΔPSNR (dB) results of different methods under LD

QP	Class	Seq	ARCNN [3]	QECNN [4]	DCAD [5]	MFQE [6]	SDTS [7]	FGTSN (ours)
37	A	PeopleOnStreet	0.4579	0.5517	0.5286	0.7526	0.7818	**0.8223**
		Traffic	0.2864	0.3213	0.3208	0.4215	0.4636	**0.4684**
	B	Kimono1	0.2439	0.2502	0.2557	0.3927	0.4066	**0.7537**
		ParkScene	0.1658	0.1769	0.1644	0.3135	0.3568	**0.4548**
		Cactus	0.1886	0.2334	0.2659	0.3263	0.3709	**0.4823**
		BasketballDrive	0.1753	0.2164	0.2540	0.1738	0.1932	**0.3951**
		BQTerrace	0.1148	0.1617	0.2527	0.2394	0.2303	**0.3298**
	C	BasketballDrill	0.1313	0.2037	0.2844	0.2656	0.2761	**0.4361**
		BQMall	0.0673	0.1194	0.1962	0.2670	0.2807	**0.3828**
		PartyScene	-0.0337	0.0062	0.1083	0.0413	0.0624	**0.3026**
		ParkScene	0.1909	0.2596	0.2643	0.2414	0.2602	**0.4641**
	D	BasketballPass	0.1315	0.2456	0.2726	0.4687	0.4691	**0.5084**
		BQSquare	-0.1283	-0.0607	0.1527	-0.0275	0.0131	**0.3162**
		BlowingBubbles	0.0774	0.1416	0.1347	0.2976	0.3474	**0.3855**
		RaceHorses	0.3166	0.3590	0.3887	0.4770	0.5247	**0.7248**
	E	FourPeople	0.3813	0.4686	0.5140	0.5741	0.6085	**0.6313**
		Johnny	0.2635	0.3496	0.3736	0.4792	0.4468	**0.5076**
		KristenAndSara	0.3158	0.4080	0.4837	0.5055	0.5461	**0.6838**
	Average		0.1859	0.2451	0.2906	0.3416	0.3688	**0.5028**
42	**Average**		0.2009	0.2417	0.2622	0.3747	0.4163	**0.5167**

* All comparison methods are retrained over the same training dataset according to authors' recommended parameters.

The experimental results show that our proposed FGTSN method significantly improves the quality of HEVC compressed video at different QPs under LD configuration, much better than the state-of-the-art quality enhancement methods.