Supplementary Material

1 Parameter settings

Table 1: Settings and hyperparameters for our experiments.

	Settings or hyperparameters	*Value*
Common settings	Pyramid level L	3
	Number of neighboring frames N	1
	Number of sub-bands in (5) n_w	4
	Stepsize of Adam optimizer	2.5×10^{-4}
	Minibatch size	32
	Scaling parameter ϵ_m in (4)	10^{-3}
	Scaling parameter ϵ_w in (5)	10^{-3}
	Weight matrix $\mathbf{W}$ in (5)	$\{0.1, 1, 1, 1\}$
Training stage 1	Input size (H × W)	256×256
	Learning rate	2×10^{-4} (decay 2 per 10^5 iterations)
	Iteration number	3×10^5
	Motion loss weight α in (3)	10 (decay 10 per 5×10^4 iterations)
	Adversarial loss weight β in (3)	0
Training stage 2	Input size (H × W)	128×128
	Learning rate	10^{-4} (decay 2 per 10^5 iterations)
	Iteration number	6×10^5
	Motion loss weight α in (3)	10^{-2}
	Adversarial loss weight β in (3)	5×10^{-3}

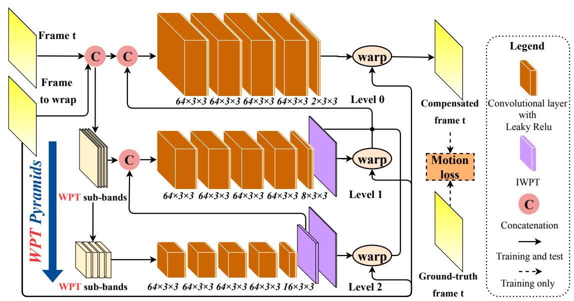
Due to the space limitation, we move the table of settings and hyperparameters for our experiments to our supplementary material. The settings and hyperparameters referred to in Section 5.1 are summarized in Table 1.

2 Details about the framework

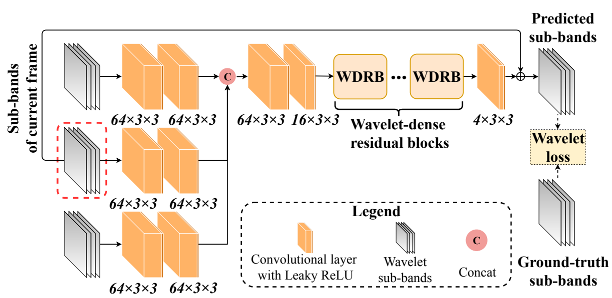
Here, we further provide the details about the motion compensation and the wavelet reconstruction network applied in MW-GAN, as shown in Figure 1.

3 Visualization in Wavelet Domain

To further verify the ability of our MW-GAN in high-frequency recovery, we show the result of our MW-GAN in wavelet domain in Figure 2. As illustrated in Figure 2, our MW-GAN can effectively enrich the high-frequency sub-bands, thus obtaining high perceptual quality, while the state-of-the-art method [1]



(a)



(b)

Fig. 1: Details about our framework. (a) Architecture of the motion compensation network in the generator; (b) Architecture of the wavelet reconstruction network in MW-GAN.

obtains low perceptual quality, due to the failure of enhancing high-frequency sub-bands. Similar results can be found for other state-of-the-art methods [9, 2, 8, 6].

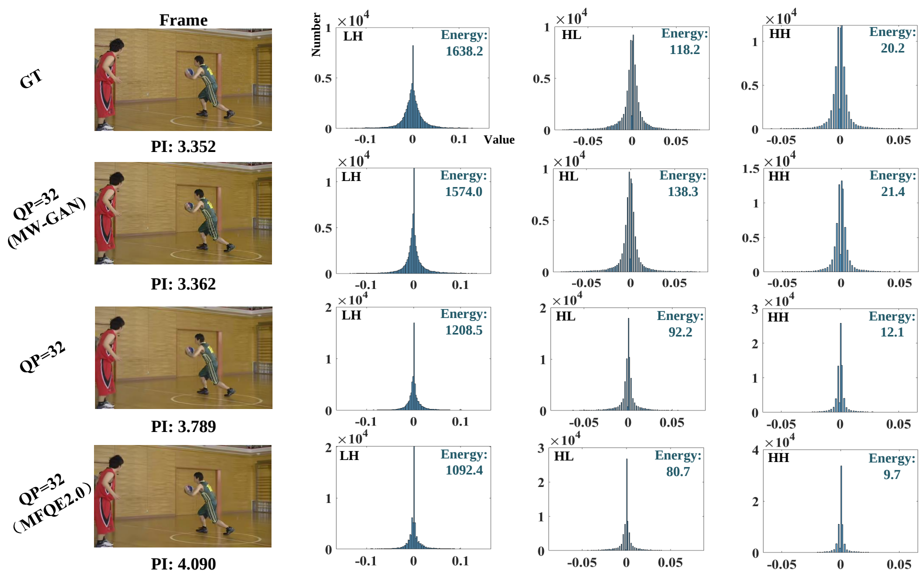


Fig. 2: Histograms of the wavelet sub-bands (zoom in for details). As can be seen, our MW-GAN can enrich the high-frequency sub-bands, thus obtaining high perceptual quality (low PI). Howerver, the state-of-the-art method [1] obtains low perceptual quality, due to the failure of enhancing high-frequency sub-bands. Similar results can be found for other state-of-the-art methods [9, 2, 8, 6].

4 Additional Metric Evaluation

In this section, we further provide results of SSIM [7], Ma [3] and NIQE [4] referred to in our paper. Table 2 represents ΔPSNR and ΔSSIM between enhanced and compressed frames, where $\Delta\text{PSNR} > 0$ and $\Delta\text{SSIM} > 0$ indicate improvement in objective quality. Besides, Table 2 represents ΔMa and ΔNIQE between enhanced and compressed frames, where $\Delta\text{Ma} > 0$ and $\Delta\text{NIQE} < 0$ indicate improvement in perceptual quality. As shown in these tables, although existing methods [2, 6, 8, 9, 1] have positive ΔPSNR and ΔSSIM values, all of them but MFQE2.0 [1] also gain negative ΔMa and positive ΔNIQE. Besides, although MFQE2.0 [1] has positive ΔMa, it also reaches highest ΔNIQE. Thus, all the methods above fail to enhance perceptual quality of compressed video.

Furthermore, our MW-GAN gains the largest increase of ΔMa and decrease of ΔNIQE in both QP = 32 and QP = 37, verifying the generalization ability of our MW-GAN for perceptual quality enhancement of compressed video.

Table 2: Overall ΔPSNR (dB) and ΔSSIM between enhanced and compressed frames on the test set of JCT-VC [5] at QP = 32 and QP = 37.

QP		Video sequence	Li *et al.* [2]		DCAD [6]		DS-CNN [8]		MFQE [9]		MFQE 2.0 [1]		Ours	
			ΔPSNR	ΔSSIM	ΔPSNR	ΔSSIM	ΔPSNR	ΔSSIM	ΔPSNR	ΔSSIM	ΔPSNR	ΔSSIM	ΔPSNR	ΔSSIM
32	A	*Traffic*	0.318	0.0037	0.320	0.0040	0.294	0.0040	0.536	0.0062	**0.599**	**0.0067**	-1.033	-0.032
		PeopleOnStreet	0.457	0.0045	0.461	0.0052	0.410	0.0045	0.718	0.0082	**0.725**	**0.0086**	-0.467	-0.0100
	B	*Kimono*	0.269	0.0048	0.278	0.0048	0.251	0.0048	**0.448**	**0.0070**	0.439	0.0070	-1.270	-0.0039
		ParkScene	0.144	0.0028	0.201	0.0038	0.151	0.0028	0.261	0.0050	**0.471**	**0.0080**	-1.232	-0.0291
		Cactus	0.210	0.0038	0.230	0.0038	0.195	0.0038	0.378	0.0060	**0.424**	**0.0061**	-0.534	-0.0238
		BQTerrace	0.166	0.0018	0.199	0.0028	0.174	0.0018	0.228	0.0025	**0.316**	**0.0043**	-0.928	-0.0268
		BasketballDrive	0.238	0.0035	0.266	0.0038	0.228	0.0038	0.316	0.0045	**0.351**	**0.0054**	-0.716	-0.0256
	C	*RaceHorses*	0.244	0.0045	0.272	0.0048	0.236	0.0045	0.304	0.0045	**0.349**	**0.0062**	-0.919	-0.0383
		BQMall	0.299	0.0048	0.368	0.0048	0.316	0.0048	0.491	0.0063	**0.578**	**0.0072**	-0.893	-0.0262
		PartyScene	0.152	0.0030	0.160	0.0038	0.155	0.0038	0.431	0.0075	**0.433**	**0.0080**	-0.793	-0.0326
		BasketballDrill	0.321	0.0038	0.410	0.0058	0.329	0.0038	0.428	0.0053	**0.508**	**0.0070**	-1.007	-0.0291
	D	*RaceHorses*	0.309	0.0048	0.351	0.0055	0.308	0.0048	0.464	0.0070	**0.481**	**0.0084**	-0.679	-0.0320
		BQSquare	0.108	0.0018	0.171	0.0014	0.154	0.0010	0.009	-0.0011	**0.242**	**0.0052**	-0.319	-0.0317
		BlowingBubbles	0.183	0.0032	0.213	0.0038	0.130	0.0018	0.054	0.0025	**0.342**	**0.0011**	-2.019	-0.0424
		BasketballPass	0.356	0.0055	0.408	0.0065	0.367	0.0058	0.592	0.0095	**0.653**	**0.0102**	-0.793	-0.0251
	E	*FourPeople*	0.405	0.0040	0.457	0.0040	0.413	0.0040	0.632	0.0055	**0.741**	**0.0060**	-1.206	-0.0265
		Johnny	0.349	0.0031	0.389	0.0040	0.343	0.0031	0.474	0.0038	**0.559**	**0.0050**	-1.441	-0.0315
		KristenAndSara	0.406	0.0030	0.453	0.0040	0.390	0.0031	0.568	0.0045	**0.698**	**0.0050**	-1.823	-0.0384
	Average		0.275	0.0037	0.316	0.0044	0.273	0.0038	0.431	0.0058	**0.516**	**0.0070**	-1.040	-0.0303
37	Average		0.299	0.0066	0.322	0.0067	0.300	0.0063	0.455	0.0088	**0.562**	**0.0109**	-0.651	-0.0204

5 Performance on other sequences

In addition to the common-used test sequences of JCT-VC, we also test the performance of our and other methods over the test set in [9], which is different from that in the paper. The results are fully shown in Table 4, which still shows the superiority of our MW-GAN according to the perceptual quality enhancement.

Table 3: Overall ΔMa and ΔNIQE between enhanced and compressed frames on the test set of JCT-VC [5] at QP = 32 and QP = 37.

QP	Video sequence		Li *et al.* [2]		DCAD [6]		DS-CNN [8]		MFQE [9]		MFQE 2.0 [1]		Ours	
			ΔMa	ΔNIQE	ΔMa	ΔNIQE	ΔMa	ΔNIQE	ΔMa	ΔNIQE	ΔMa	ΔNIQE	ΔMa	ΔNIQE
32	A	*Traffic*	-0.683	0.319	-0.611	0.228	-0.536	0.211	-0.604	0.277	-0.574	0.287	**1.395**	**-0.852**
		PeopleOnStreet	-0.892	0.838	-0.697	0.639	-0.643	0.684	-0.697	0.917	-0.765	0.823	**0.822**	**-0.294**
	B	*Kimono*	-0.373	0.585	-0.277	-0.530	-0.190	0.475	-0.296	0.584	-0.278	0.609	**1.823**	**-1.299**
		ParkScene	-0.366	0.506	-0.327	0.434	-0.235	0.458	-0.310	0.507	-0.336	0.489	**0.660**	**-0.870**
		Cactus	-0.099	0.429	-0.091	0.437	-0.099	0.372	-0.063	0.433	-0.079	0.469	**1.096**	**-0.810**
		BQTerrace	-0.305	0.463	-0.377	0.449	-0.327	0.397	-0.345	0.492	-0.402	0.492	**0.377**	**-0.274**
		BasketballDrive	-0.243	0.824	-0.202	0.800	-0.192	0.732	-0.294	0.896	-0.283	0.953	**1.254**	**-0.588**
	C	*RaceHorses*	-0.310	0.699	-0.308	0.764	-0.269	0.669	-0.301	0.677	-0.325	0.808	**0.398**	**-0.761**
		BQMall	-0.391	0.523	-0.364	0.651	-0.320	0.532	-0.353	0.582	-0.369	0.628	**0.973**	**-0.844**
		PartyScene	-0.162	0.435	-0.174	0.579	-0.135	0.447	-0.140	0.337	-0.188	0.504	**0.106**	**-0.213**
		BasketballDrill	-0.783	0.781	-0.696	1.156	-0.759	0.924	-0.731	0.812	-0.767	1.090	**0.516**	**-0.549**
	D	*RaceHorses*	-0.376	0.881	-0.367	0.947	-0.321	0.842	-0.413	0.901	-0.415	0.956	**0.267**	**-0.132**
		BQSquare	-0.048	0.523	-0.084	1.114	-0.063	0.700	-0.067	0.621	-0.083	1.057	**0.068**	**-0.981**
		BlowingBubbles	-0.143	0.841	-0.166	1.004	-0.133	0.783	-0.143	0.645	-0.164	0.780	**0.168**	**-0.315**
		BasketballPass	-0.410	0.664	-0.385	0.743	-0.360	0.675	-0.485	0.662	-0.493	0.695	**0.697**	**-0.890**
	E	*FourPeople*	-0.088	0.443	-0.028	0.434	-0.038	0.429	-0.039	0.438	-0.015	0.424	**1.468**	**-0.790**
		Johnny	-0.440	0.176	-0.415	0.264	-0.133	0.297	-0.409	0.318	-0.262	0.256	**1.462**	**-0.881**
		KristenAndSara	-0.042	0.515	0.030	0.661	-0.080	0.641	-0.025	0.650	0.040	0.745	**1.582**	**-1.027**
	Average		-0.342	0.580	-0.308	0.657	-0.269	0.570	-0.318	0.597	-0.320	0.670	**0.841**	**-0.687**
37	Average		-0.300	0.782	-0.324	0.917	-0.286	0.838	-0.363	0.872	-0.314	0.913	**1.145**	**-0.842**

Table 4: Overall ΔLPIPS and ΔPI between enhanced and compressed frames on the test set of MFQE [9] at QP = 32.

QP	Video sequence	Li *et al.* [2]		DCAD [6]		DS-CNN [8]		MFQE [9]		MFQE 2.0 [1]		Ours	
		ΔLPIPS	ΔPI	ΔLPIPS	ΔPI	ΔLPIPS	ΔPI	ΔLPIPS	ΔPI	ΔLPIPS	ΔPI	ΔLPIPS	ΔPI
32	*PeopleOnStreet*	0.020	0.865	0.020	0.668	0.019	0.663	0.017	0.807	0.017	0.794	**-0.020**	**-0.558**
	TunnelFlag	0.013	0.151	0.013	0.120	0.013	0.200	0.013	0.103	0.013	0.082	**-0.029**	**-0.832**
	Kimono	0.038	0.479	0.034	0.403	0.033	0.332	0.034	0.440	0.036	0.443	**-0.069**	**-1.561**
	BarScene	0.021	0.309	0.018	0.342	0.020	0.346	0.020	0.280	0.018	0.284	**-0.064**	**-2.024**
	Vidyo1	0.017	0.599	0.013	0.513	0.017	0.441	0.014	0.460	0.012	0.443	**-0.024**	**-1.339**
	Vidyo3	0.007	0.376	0.004	0.384	0.006	0.421	0.004	0.393	0.003	0.343	**-0.038**	**-0.877**
	Vidyo4	0.024	0.271	0.021	0.282	0.023	0.352	0.021	0.330	0.020	0.243	**-0.035**	**-1.531**
	BasketballPass	0.020	0.537	0.016	0.564	0.016	0.517	0.020	0.574	0.019	0.594	**-0.021**	**-0.794**
	RaceHorses	0.023	0.504	0.022	0.536	0.021	0.469	0.025	0.489	0.027	0.566	**-0.021**	**-0.579**
	MaD	0.013	0.436	0.012	0.374	0.011	0.272	0.010	0.313	0.009	0.300	**-0.074**	**-0.483**
	Average	0.020	0.453	0.017	0.419	0.018	0.401	0.018	0.419	0.017	0.409	**-0.039**	**-1.058**

References

1. Guan, Z., Xing, Q., Xu, M., Yang, R., Liu, T., Wang, Z.: Mfqe 2.0: A new approach for multi-frame quality enhancement on compressed video. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) pp. 1–1 (2019)
2. Li, K., Bare, B., Yan, B.: An efficient deep convolutional neural networks model for compressed image deblocking. In: 2017 IEEE International Conference on Multimedia and Expo (ICME). pp. 1320–1325. IEEE (2017)
3. Ma, C., Yang, C.Y., Yang, X., Yang, M.H.: Learning a no-reference quality metric for single-image super-resolution. Computer Vision and Image Understanding (CVIU) **158**, 1–16 (2017)
4. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a completely blind image quality analyzer. IEEE Signal Processing Letters **20**(3), 209–212 (2012)
5. Ohm, J.R., Sullivan, G.J., Schwarz, H., Tan, T.K., Wiegand, T.: Comparison of the coding efficiency of video coding standardsincluding high efficiency video coding (hevc). IEEE Transactions on circuits and systems for video technology (TCSVT) **22**(12), 1669–1684 (2012)
6. Wang, T., Chen, M., Chao, H.: A novel deep learning-based method of improving coding efficiency from the decoder-end for hevc. In: 2017 Data Compression Conference (DCC). pp. 410–419. IEEE (2017)
7. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., et al.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing (TIP) **13**(4), 600–612 (2004)
8. Yang, R., Xu, M., Wang, Z.: Decoder-side hevc quality enhancement with scalable convolutional neural network. In: 2017 IEEE International Conference on Multimedia and Expo (ICME). pp. 817–822. IEEE (2017)
9. Yang, R., Xu, M., Wang, Z., Li, T.: Multi-frame quality enhancement for compressed video. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6664–6673 (2018)