VFHQ: A High-Quality Dataset and Benchmark for Video Face Super-Resolution

Liangbin Xie[*] [1,2,3] Xintao Wang[3] Honglun Zhang[3] Chao Dong[†1] Ying Shan[3]

[1]Shenzhen Key Lab of Computer Vision and Pattern Recognition,
Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences

[2]University of Chinese Academy of Sciences [3]ARC Lab, Tencent PCG

{lb.xie, chao.dong}@siat.ac.cn {xintaowang, honlanzhang, yingsshan}@tencent.com

Abstract

In this supplementary file, we provide more quantitative results (Section 1) and qualitative results (Section 2) of the benchmarking study in bicubic degradation and blind degradation settings. Specifically, we report the results for scale $\times 4$, $\times 8$ in bicubic degradation setting and $\times 4$ in blind degradation setting.

1. Quantitative Results

As mentioned in the main text, in order to comprehensively evaluate the performance of existing methods towards different levels of face motion, we test them with different sampling intervals (i.e, $\{1, 3, 5, 10, 15\}$).

In Tab. 1 and Tab. 2, we evaluate the performance of existing methods in bicubic degradation with scale $\times 4$ and $\times 8$. For different sampling intervals, we can observe that BasisVSR achieves the best performance in both PSNR and SSIM metrics.

In Tab. 3, we list the results of selected algorithms in blind degradation with a scale $\times 4$. We can find that in the blind degradation setting, the performance gap between EDVR [3] and BasicVSR [1] are smaller than bicubic degradation.

2. Qualitative Results

The qualitative results of bicubic degradation with scale $\times 4$ and $\times 8$ are shown in Fig. 1 and Fig. 2, respectively. It can be found that in the $\times 4$ bicubic degradation setting, current methods are capable of restoring high-quality face videos. For the $\times 8$ bicubic degradation setting, there is still a clear gap between the output of BasicVSR and GT, which indicates that VFSR with large scale ratio in bicubic degradation setting (e.g, ×8, ×16) is a challenge for further investigation.

Fig. 3 and Fig. 4 show the results of four state-of-the-art methods in slight and severe blind degradation settings. As shown in Fig. 3, when the degradation contained in the input sequence is slight, BasicVSR-GAN can restore more visual-pleasing results than the other three methods. There are two reasons, 1) Although the adopted blind degradation model is implemented by following the practice in GFPGAN [4], there still exists bias due to the different compression types between video and image. 2) BasicVSR-GAN can use the temporal information between neighboring frames, which helps to mitigate the inconsistency in the restored videos.

However, when the degradation of the input video is relatively severe (Fig. 4), BasicVSR-GAN can not restore realistic faces. For DFDNet [2], we find that the restored faces of this method contain strange artifacts. Although GPEN [5] and GFPGAN can output better result for each input frames, the neighboring frames among the restored video are inconsistent (e.g, face identity, eye). This phenomenon of inconsistency is severe in videos with large motion. All these observations indicate that VFSR in blind degradation setting needs further investigation, especially for videos with large motion, video compression and large pose.

References

[1] Kelvin CK Chan, Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Basicvsr: The search for essential components in video super-resolution and beyond. In *CVPR*, 2021.

[2] Xiaoming Li, Chaofeng Chen, Shangchen Zhou, Xianhui Lin, Wangmeng Zuo, and Lei Zhang. Blind face restoration via deep multi-scale component dictionaries. In *ECCV*. Springer, 2020.

[3] Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *CVPR Workshops*, 2019.

[4] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. Towards real-world blind face restoration with generative facial prior. In *CVPR*, 2021.

[5] Tao Yang, Peiran Ren, Xuansong Xie, and Lei Zhang. Gan prior embedded network for blind face restoration in the wild. In *CVPR*, 2021.

[*]Liangbin Xie is an intern in ARC Lab, Tencent PCG.

[†]Corresponding author.

Table 1. Benchmarking results with **bicubic** degradation model (evaluated on VFHQ-Test). Average PSNR/SSIM values for scaling factor $\times\mathbf{4}$. **Red** and blue indicates the best and second best performance. The selected sampling intervals are $\{1, 3, 5, 10, 15\}$.

Interval	Metrics	MSE-based				GAN-based		
		Bicubic	RRDB	EDVRM	BasicVSR	ESRGAN	EDVRM-GAN	BasicVSR-GAN
1	PSNR	31.959	35.317	36.259	**36.391**	32.790	33.663	32.315
	SSIM	0.8938	0.9301	0.9416	**0.9429**	0.8960	0.9100	0.8868
3	PSNR	31.955	35.319	36.207	**36.364**	32.795	33.664	32.317
	SSIM	0.8939	0.9302	0.9412	**0.9425**	0.8961	0.9102	0.8869
5	PSNR	31.964	35.332	36.090	**36.258**	32.803	33.592	32.327
	SSIM	0.8939	0.9302	0.9399	**0.9412**	0.8961	0.9089	0.8869
10	PSNR	31.960	35.353	35.885	**36.135**	32.813	33.461	32.334
	SSIM	0.8944	0.9308	0.9378	**0.9399**	0.8969	0.9070	0.8876
15	PSNR	32.004	35.389	35.846	**36.068**	32.862	33.450	32.369
	SSIM	0.8946	0.9308	0.9365	**0.9386**	0.8969	0.9058	0.8878

Table 2. Benchmarking results with **bicubic** degradation model (evaluated on VFHQ-Test). Average PSNR/SSIM values for scaling factor $\times\mathbf{8}$. **Red** and blue indicates the best and second best performance. The selected sampling intervals are $\{1, 3, 5, 10, 15\}$.

Interval	Metrics	MSE-based				GAN-based		
		Bicubic	RRDB	EDVRM	BasicVSR	ESRGAN	EDVRM-GAN	BasicVSR-GAN
1	PSNR	28.125	31.210	31.913	**32.014**	28.113	29.311	28.861
	SSIM	0.8182	0.8728	0.8817	**0.8838**	0.8055	0.8208	0.8152
3	PSNR	28.12	31.204	31.963	**32.129**	28.102	29.360	28.953
	SSIM	0.8182	0.8729	0.8829	**0.8858**	0.8056	0.8249	0.8187
5	PSNR	28.124	31.203	31.888	**32.095**	28.113	29.360	28.993
	SSIM	0.8183	0.8730	0.8820	**0.8853**	0.8058	0.8260	0.8200
10	PSNR	28.119	31.213	31.747	**31.992**	28.108	29.366	29.014
	SSIM	0.8186	0.8735	0.8800	**0.8842**	0.8062	0.8275	0.8212
15	PSNR	28.146	31.255	31.730	**31.964**	28.150	29.421	29.063
	SSIM	0.8190	0.8736	0.8789	**0.8831**	0.8068	0.8280	0.8216

Table 3. Benchmarking results with **blind** degradation model (evaluated on VFHQ-Test). Average PSNR/SSIM/LPIPS values for scaling factor $\times\mathbf{4}$. **Red** and blue indicates the best and second best performance. The selected sampling intervals are $\{1, 3, 5, 10, 15\}$.

Interval	Metrics	MSE-based			GAN-based			GAN-prior based	
		Bicubic	EDVRM	BasicVSR	EDVRM-GAN	BasicVSR-GAN	DFDNet	GFPGAN	GPEN
1	PSNR	26.482	29.283	**29.356**	26.008	25.740	25.013	25.936	26.503
	SSIM	0.7868	0.8409	**0.8423**	0.7435	0.7486	0.7521	0.7704	0.7742
	LPIPS	0.4121	0.3289	0.3306	**0.3186**	0.3252	0.4006	0.3439	0.3634
3	PSNR	26.690	29.383	29.425	26.311	25.940	25.220	25.931	26.502
	SSIM	0.7915	0.8436	**0.8444**	0.7593	0.7560	0.7561	0.7704	0.7742
	LPIPS	0.4053	0.3277	0.3301	**0.3090**	0.3217	0.3979	0.3439	0.3637
5	PSNR	26.842	29.457	**29.472**	26.682	25.813	25.178	25.978	26.672
	SSIM	0.7909	0.8428	**0.8430**	0.7638	0.7410	0.7560	0.7723	0.7768
	LPIPS	0.4098	0.3288	0.3309	**0.3076**	0.3214	0.4008	0.3446	0.3607
10	PSNR	26.342	28.988	**29.014**	26.301	25.658	25.144	25.913	26.500
	SSIM	0.7827	0.8365	**0.8370**	0.7617	0.7498	0.7528	0.7697	0.7743
	LPIPS	0.4235	0.3371	0.3396	**0.3119**	0.3265	0.4090	0.3406	0.3603
15	PSNR	26.433	29.052	**29.060**	26.274	25.664	25.038	25.949	26.532
	SSIM	0.7839	0.8369	**0.8374**	0.7621	0.7508	0.7516	0.7701	0.7745
	LPIPS	0.4148	0.3354	0.3390	**0.3112**	0.3257	0.4069	0.3405	0.3603

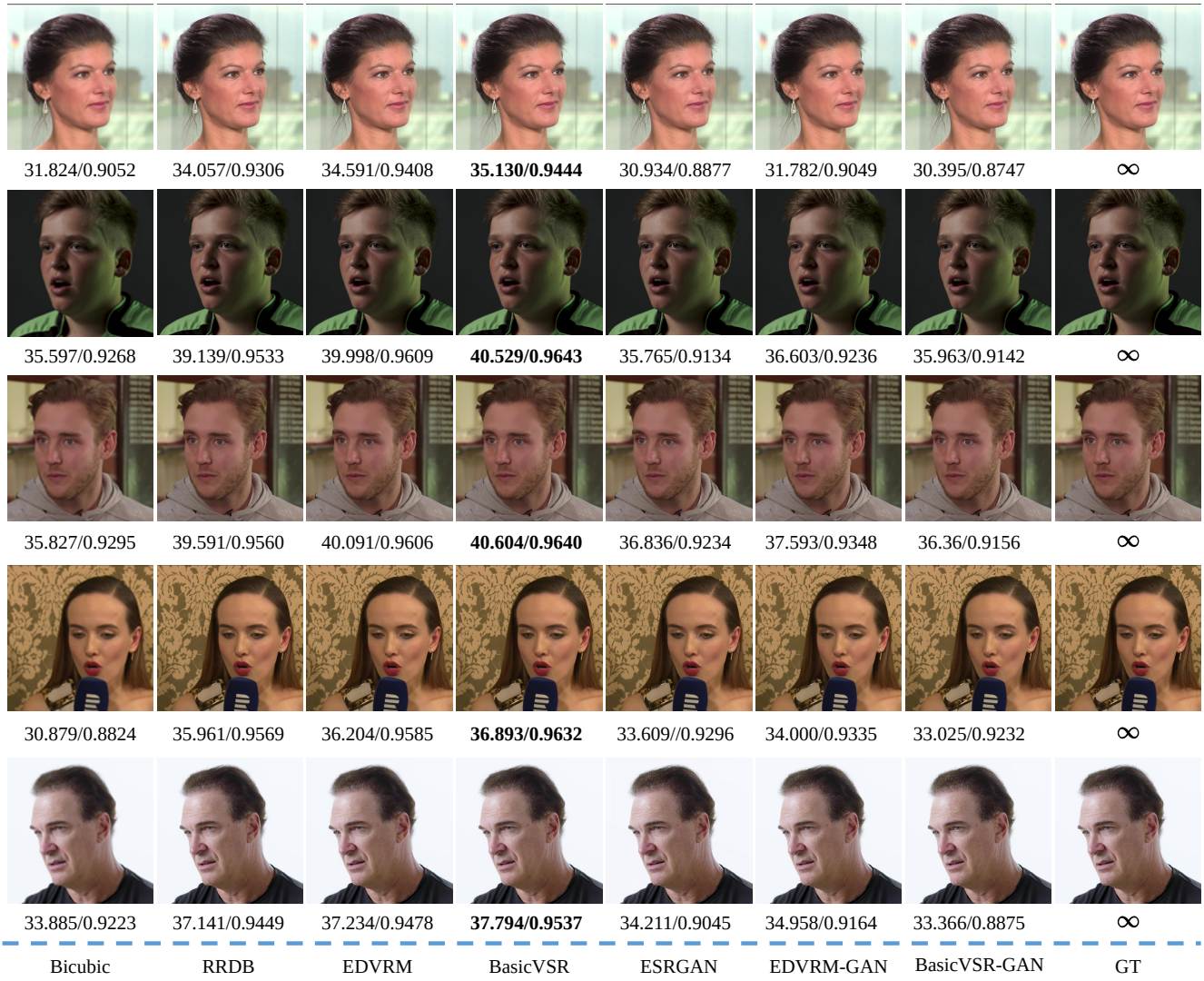


Figure 1. Qualitative comparison by different models in $\times$**4 bicubic** degradation setting. From top to bottom, the sampling intervals are 1, 3, 5, 10, 15. **Zoom in for best view**.

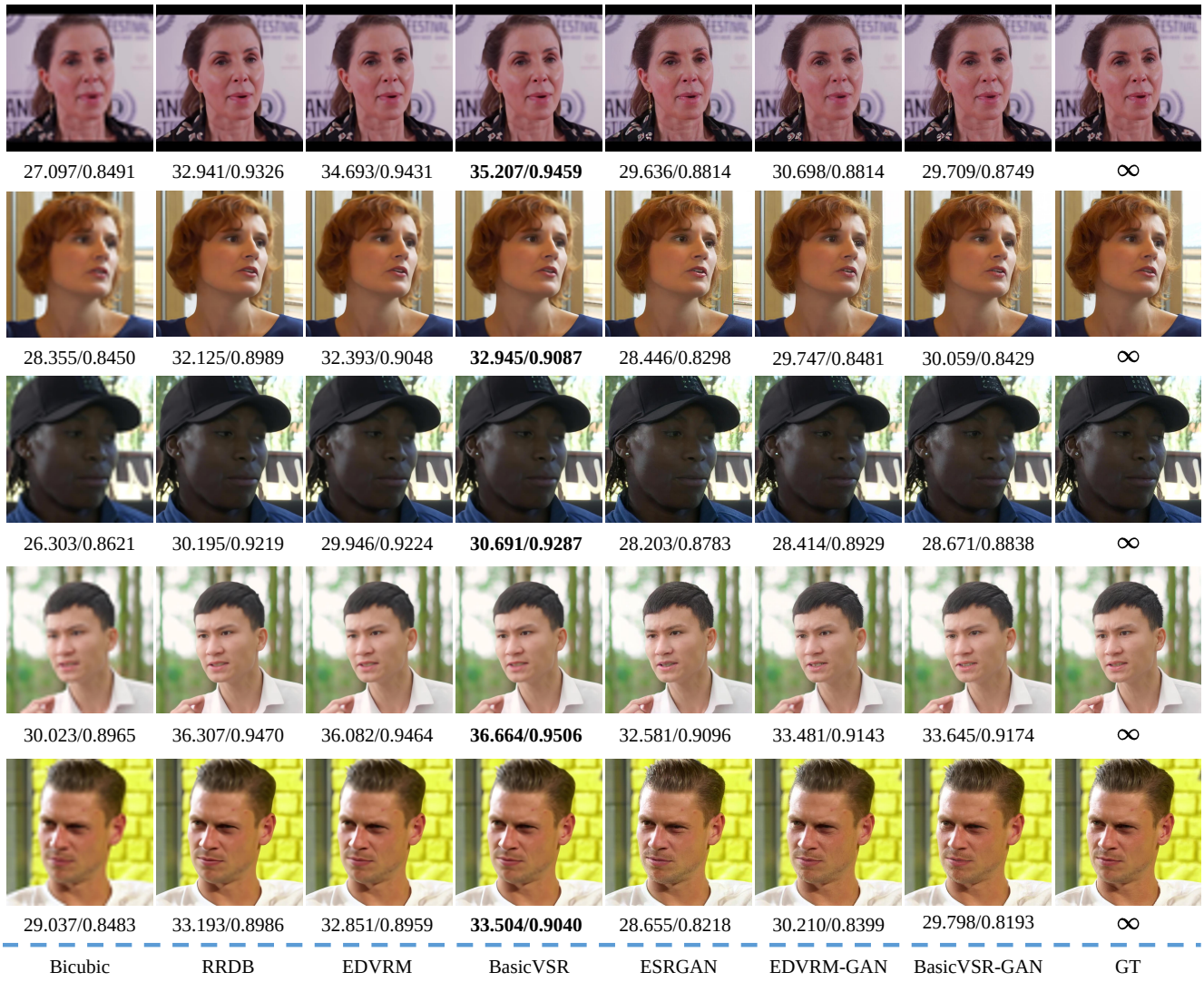


Figure 2. Qualitative comparison by different models in $\times$**8 bicubic** degradation setting. From top to bottom, the sampling intervals are 1, 3, 5, 10, 15. **Zoom in for best view**.

Bicubic

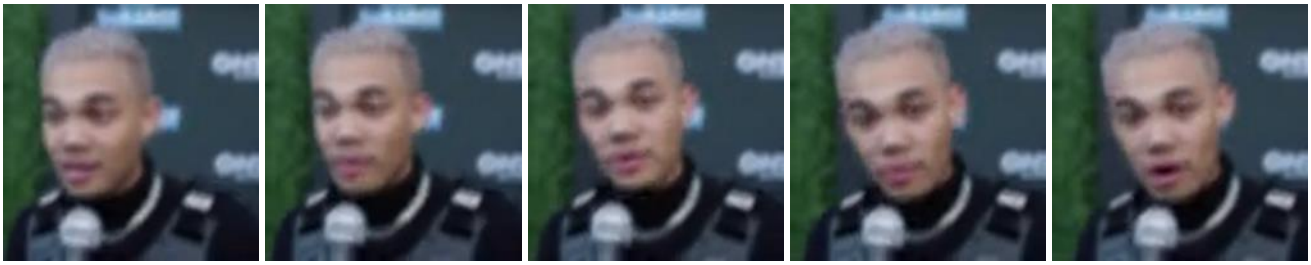
BasicVSR-GAN

DFDNet

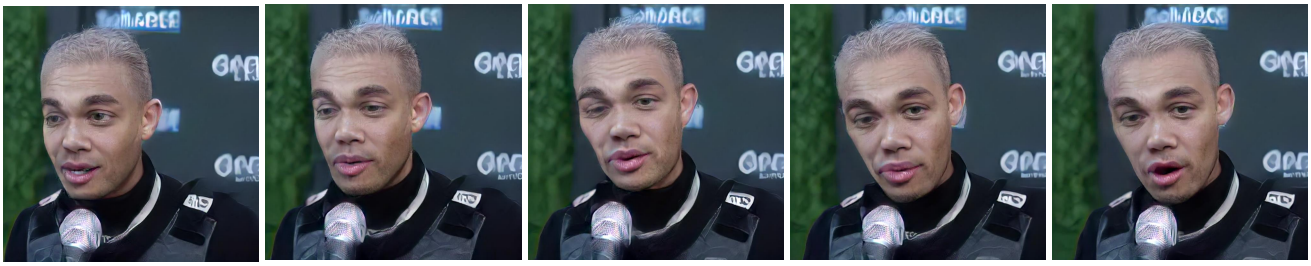
GPEN

GFPGAN

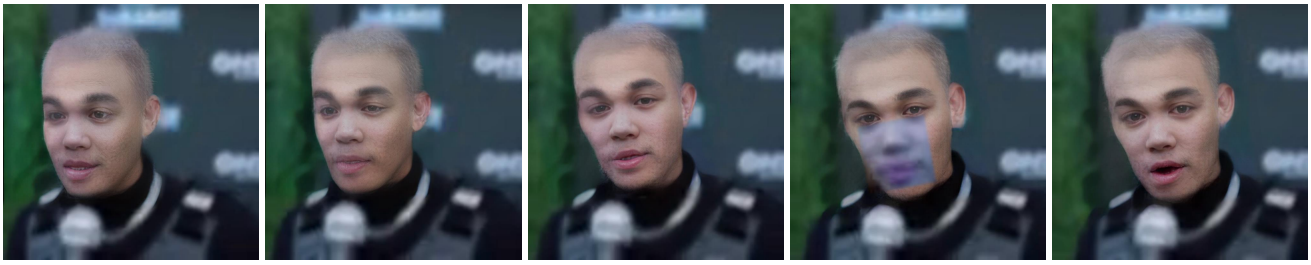
Figure 3. Qualitative comparison by different models in $\times$**4 blind** degradation setting. The degradation contained in the input sequence is slight. From top to bottom, the sampling intervals are 1, 3, 5, 10, 15. **Zoom in for best view**.



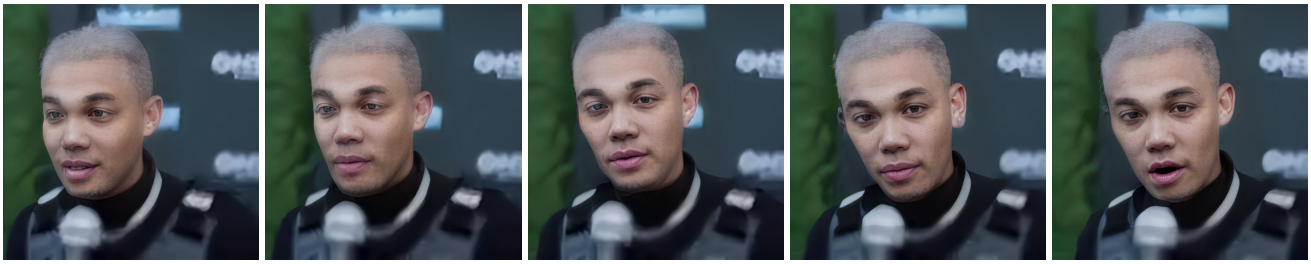
Bicubic



BasicVSR-GAN



DFDNet



GPEN

GFPGAN

Figure 4. Qualitative comparison by different models in $\times$**8 blind** degradation setting. The degradation contained in the input sequence is severe. From top to bottom, the sampling intervals are 1, 3, 5, 10, 15. **Zoom in for best view**.