Supplementary Material: Proposal-based Video Completion

In the supplementary material, we first provide a user study to compare our method with the state of the art for both object removal on DAVIS and fixed region inpainting on YouTube VOS. Subsequently we provide additional qualitative results and detailed analysis to discuss the advantages and drawbacks of our method.

6 User Study

We conduct a user study to assess the quality of video completion on both object removal and fixed region inpainting, since existing popular metrics (*i.e.*, SSIM, PSNR and LPIPS) may not reflect the real visual quality of inpainted results perceived by a human.

Fig. 9 and Fig. 10 illustrate the user interface we designed for the user study. Each time we show the inpainted results from two different algorithms together with the original input video. One is ours and the other is randomly chosen from Deepfill2 [42], FGI [40], OPN [25], CPN [20], VINet [18]. The two chosen results are randomly shuffled so that they can appear in any order. We ask the users to select a visually more plausible result from the two given inpainted videos. We compute the winning percentage of our method to each baseline and report them in Tab. 4 and Tab. 5. A winning percentage larger than 50% means the users chose our method to be better more often. On the DAVIS dataset as shown in Tab. 4, our method is the second best method among all, falling behind FGI with a winning percentage of 40% over FGI, *i.e.*, when compared to FGI, 40% of the time users suggested that the completion result by our method is better than FGI. Although our method falls behind FGI in terms of winning percentage, ours runs around 18 times faster than FGI in the object removal scenario. On the YouTube VOS dataset as shown in Tab. 5, our method is the second best method among all as well, falling behind OPN while running more than 10 times faster than OPN. Note that our method is better than OPN on object removal in Tab. 4, and better than FGI on fixed region inpainting in Tab. 5. In summary, we think our method is fast while providing appealing results for both object removal and fixed region inpainting. This message is consistent with the results reported in Fig. 4.

Table 4. User study with object removal scenario on DAVIS.

	Deepfill2 [42]	FGI [40]	OPN [25]	CPN [20]	VINet [18]	**Ours**
Runtime per frame	**0.09 s**	12.45 s	4.18 s	0.54 s	0.18 s	0.68 s
Winning percentage of our method	92.31%	**40.00%**	54.55%	57.14%	86.66%	N/A

Table 5. User study with fixed region inpainting scenario on YouTube VOS.

	Deepfill2 [42]	FGI [40]	OPN [25]	CPN [20]	VINet [18]	**Ours**
Runtime per frame	**0.32 s**	112.32 s	9.05 s	1.40 s	0.18 s	0.87 s
Winning percentage of our method	84.38%	52.00%	**42.85%**	64.29%	83.87%	N/A

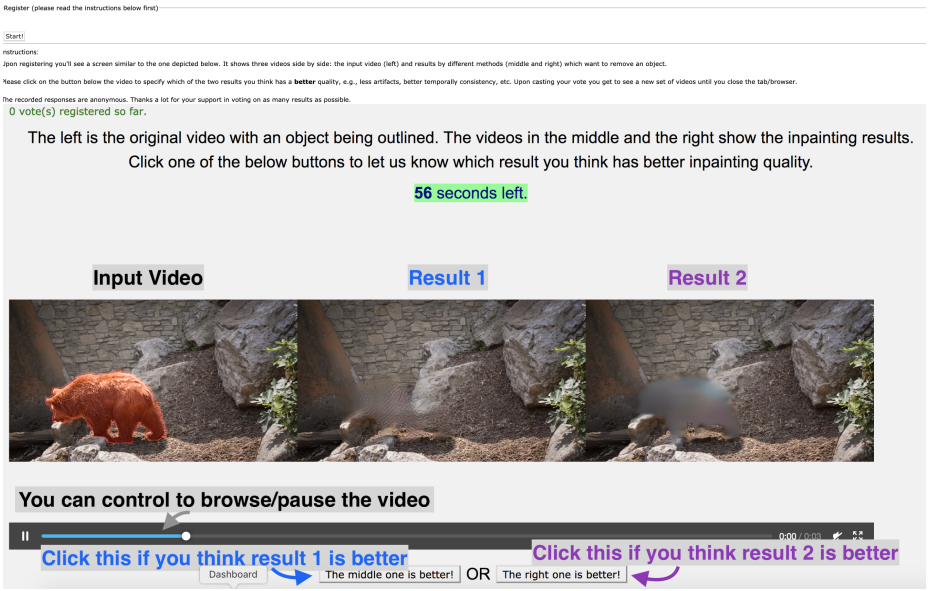

Fig. 9. The instruction page of the user study.

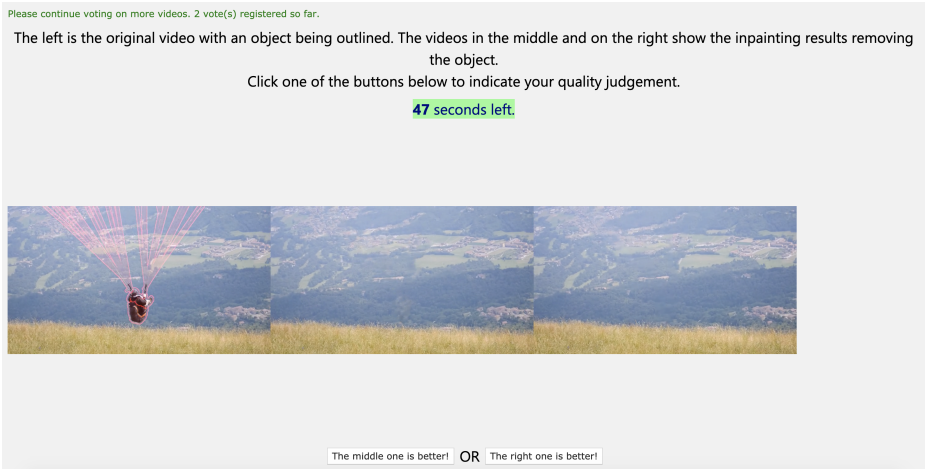

Fig. 10. A screenshot of the webpage for the user study.

7 Dataset Statistics

We use DAVIS [27] and YouTube VOS [39] datasets for evaluation. We provide the statistics of the two datasets for testing in Tab. 6. There are more videos with longer length and higher resolution in the YouTube VOS dataset compared to DAVIS.

Table 6. Dataset statistics for DAVIS and Youtube VOS test set..

Dataset	# of videos	# of frames	# of frames per video	Resolution
DAVIS [39]	90	6208 frames	68.9 ± 17.3 frames	$\sim 643^2$ pixels
Youtube VOS test [39]	541	75,784 frames	140 ± 41.5 frames	$\sim 947^2$ pixels

8 More qualitative results

In Fig. 11, Fig. 12 Fig. 13 Fig. 14 and Fig. 15, we show more qualitative results of our approach and compare to the baselines. In Fig. 11, CPN and OPN fail to smoothly connect the stick at the end of the video and FGI and VINet don't address the illumination changes in the video well and the completion results in the end are darker due to propagation from the beginning of the video. In Fig. 12, despite the missing area being large, *e.g.*, the last frame we show in Fig. 12, our method produces less artifacts compared to other methods.

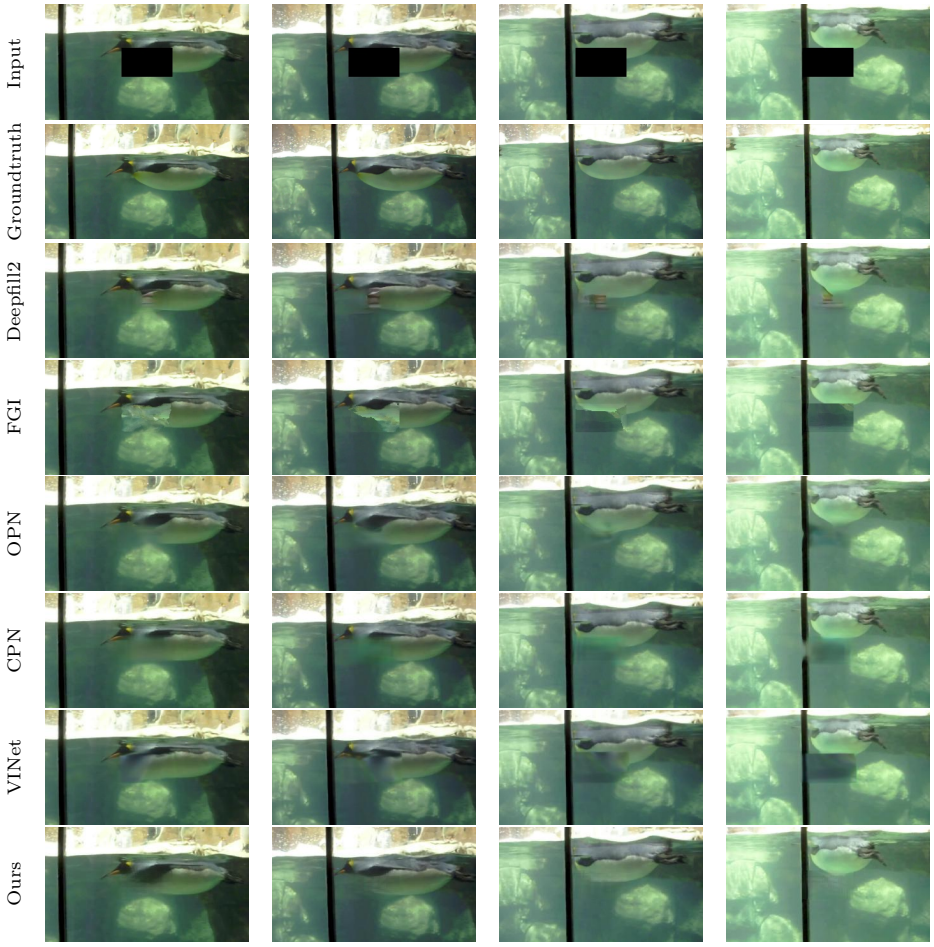

Fig. 11. Object removal results of our method compared to various baselines. Our method preserves much better the shape of different objects and generates much less artifacts.

Fig. 12. Object removal results of our method compared to various baselines. Our method preserves much better the shape of different objects and generates much less artifacts. Note that in this video the missing area is large in the end of the video, and our method is producing less artifacts.

Fig. 13. Object removal results of our method compared to various baselines. Our method preserves much better the shape of different objects and generates much less artifacts.

Fig. 14. Object removal results of our method compared to various baselines. Our method preserves much better the shape of different objects and generates much less artifacts.

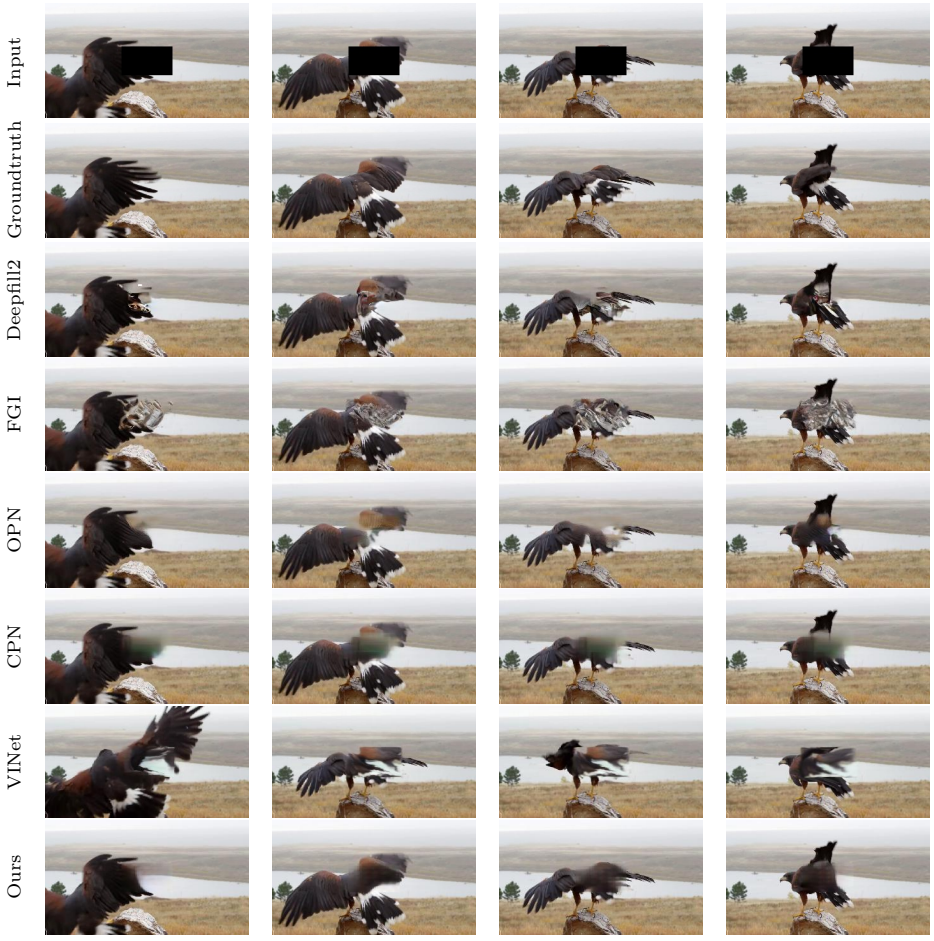

Fig. 15. Object removal results of our method compared to various baselines. Our method preserves much better the shape of different objects and generates much less artifacts.