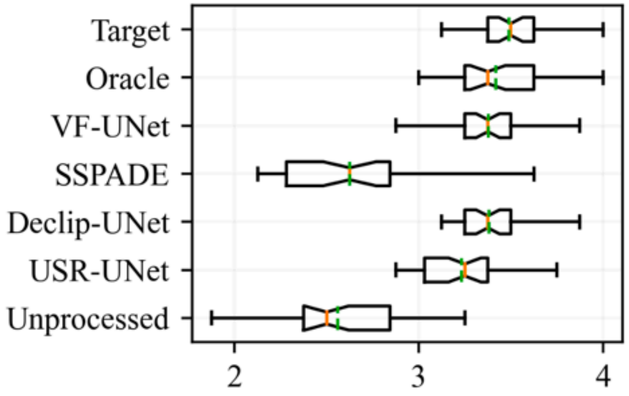
Experiments

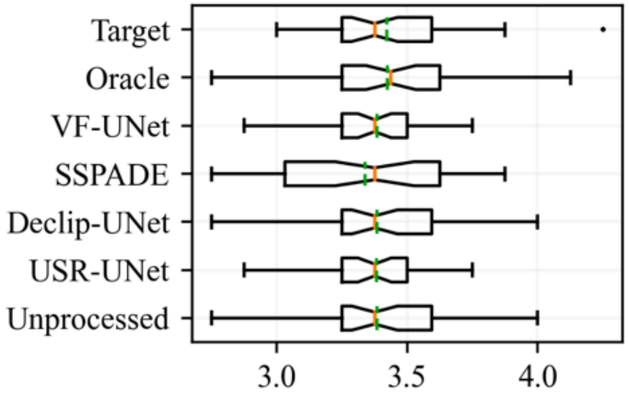
Speech Enhancement and Dereverberation

Table.13 Performance on declipping task

Clipping Level	0.25				0.10				Average			
Models	SiSNR	STOI	PESQ	MOS	SiSNR	STOI	PESQ	MOS	SiSNR	STOI	PESQ	MOS
Unprocessed	9.60	0.95	2.38	2.56	4.00	0.89	1.51	2.72	6.80	0.92	1.95	2.64
Oracle	-19.94	0.81	2.36	3.44	-19.94	0.81	2.36	3.42	-19.94	0.81	2.36	3.43
Target	/	/	/	3.42	/	/	/	3.49	/	/	/	3.46
GSR-UNet	11.01	0.97	3.54	3.38	7.47	0.94	2.89	3.23	9.24	0.95	3.21	3.31
Declip-UNet	12.45	**0.99**	**3.98**	3.38	8.43	**0.96**	**3.40**	3.38	10.44	**0.98**	3.69	3.38
SSPADE	**17.43**	0.98	3.55	3.34	**10.31**	0.92	2.12	2.63	**13.87**	0.95	2.84	2.98
VF-DNN	/	0.76	1.72	/	/	0.72	1.48	/	/	0.74	1.60	/
VF-BiGRU	/	0.81	2.09	/	/	0.79	1.82	/	/	0.80	1.95	/
VF-UNet-S	/	0.82	2.13	/	/	0.80	1.85	/	/	0.81	1.99	/
VF-UNet	/	0.82	2.21	3.38	/	0.80	1.93	3.38	/	0.81	2.07	3.38



(a) Declipping (0.1)



(d) Declipping (0.25)

Hard to distinguish!

Conclusion

- We proposed Generic Speech Restoration (GSR), to overcome the limitation of Single Task Speech Restoration (SSR).
- We proposed a novel GSR architecture VoiceFixer to alleviate the problem of regression based GSR/SSR model.
- We demonstrate that:
  - Model trained in a GSR way can achieve a comparable or even better performance than the SSR model.
  - *VoiceFixer* is an effective architecture for GSR. Especially for speech super resolution task.