

## Introduction

- CRN typically employs consecutive downsampling and upsampling convolution for frequency modeling, which destroys the inherent structure of the signal over frequency.
- Hao et al. [1] propose FullSubNet, due to the introduction of full-band and sub-band model, these methods lead to a large number of parameters, increasing the complexity of the model.
- We propose an innovative module combing a State space model and Inplace Convolution [2] (SIC), and to replace the conventional convolution in CRN, called SICRN.
- The experimental results demonstrate notably high evaluation scores. Notably, this achievement is attained with less than 1/2 the parameters and 1/7 the computational complexity of FullSubNet.

## Proposed Method

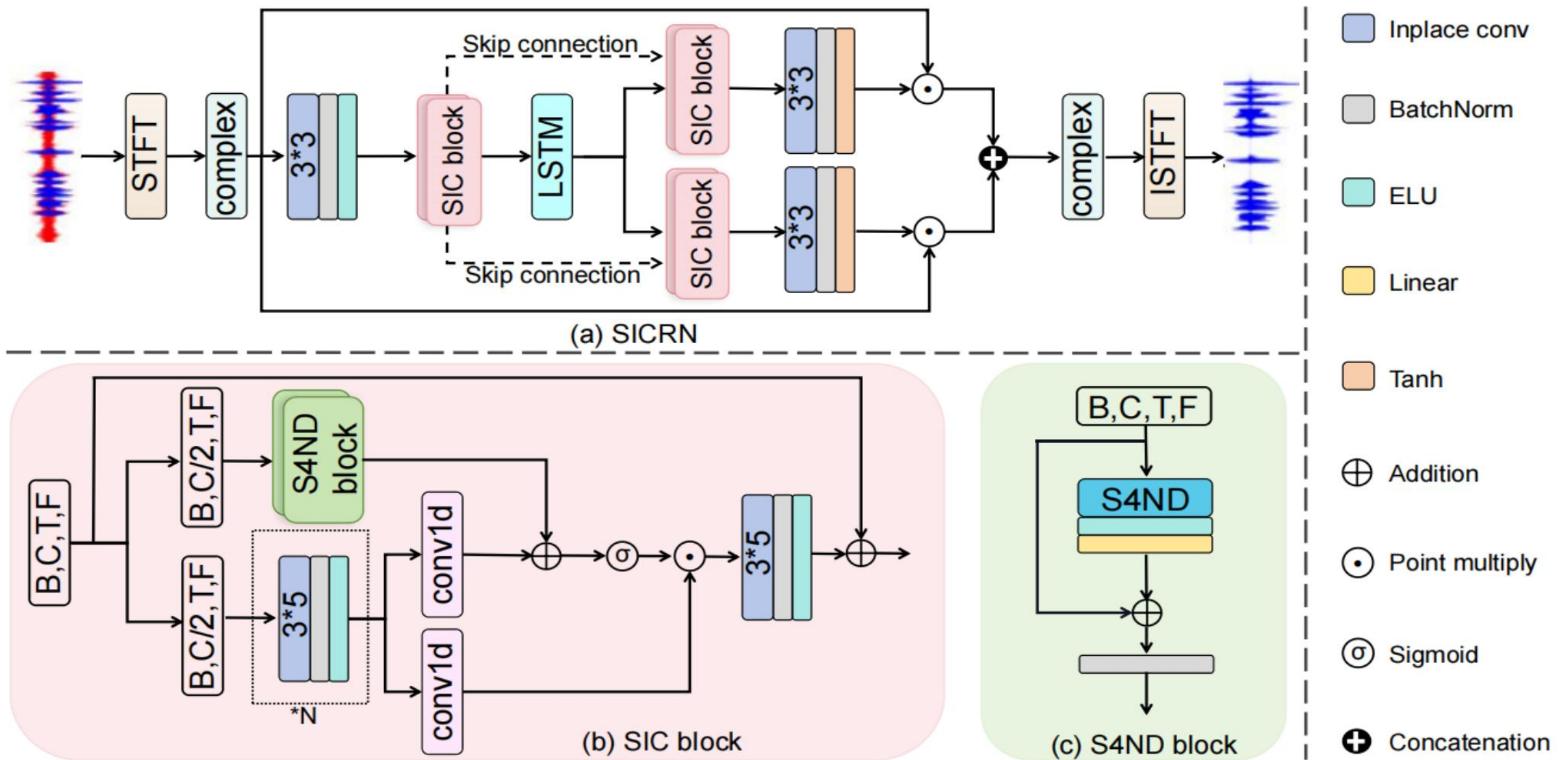


Fig. 1: Overview of the proposed SICRN system

## Experimental Results

Table 1: The performance in terms of WB-PESQ [MOS], NB-PESQ [MOS], STOI [%], and SI-SDR [dB] on the DNS challenge test dataset.

Method	#Para (M)	Look Ahead (ms)	With Reverb				Without Reverb			
			WB-PESQ	NB-PESQ	STOI	SI-SDR	WB-PESQ	NB-PESQ	STOI	SI-SDR
Noisy	-	-	1.822	2.753	86.62	9.033	1.582	2.454	91.52	9.071
NSNet [20]	5.1	0	2.365	3.076	90.43	14.721	2.145	2.873	94.47	15.613
DTLN [21]	1.0	-	-	2.700	84.68	10.530	-	3.040	94.76	16.340
Conv-TasNet [22]	5.08	33	2.750	-	-	-	2.730	-	-	-
DCCRN-E [23]	3.7	37.5	-	3.077	-	-	-	3.266	-	-
PoCoNet [24]	50	-	2.832	-	-	-	2.748	-	-	-
FullSubNet [11]	5.64	32	2.969	3.473	92.62	15.750	2.777	3.305	96.11	17.290
SICRN	2.16	0	2.891	3.433	92.59	15.137	2.624	3.233	95.83	15.998

Table 2: Comprehensive comparison with FullSubNet.

Method	#Para(M)	MACs(G/s)	Look Ahead(ms)
FullSubNet	5.64	30.84	32
SICRN	<b>2.16</b>	<b>4.24</b>	<b>0</b>

Table 3: Ablation experiment(With Reverb).

Method	WB-PESQ	NB-PESQ	STOI	SI-SDR
mixture	1.822	2.753	86.62	9.033
IICRN	2.797	3.378	91.71	14.929
SICRN	<b>2.891</b>	<b>3.433</b>	<b>92.59</b>	<b>15.137</b>

Table 4: Ablation experiment(Without Reverb).

Method	WB-PESQ	NB-PESQ	STOI	SI-SDR
Noisy	1.582	2.454	91.52	9.071
IICRN	2.596	3.218	95.56	15.795
SICRN	<b>2.624</b>	<b>3.233</b>	<b>95.83</b>	<b>15.998</b>

Replace the S4ND block in the SIC block with inplace convolution and name it IICRN.

## Conclusion

- We propose an innovative model combing a state space model and inplace convolution, called SICRN.
- This network avoids downsampling operations throughout its architecture.
- The experimental results demonstrate superior performance achieved with fewer parameters and computational resources.

## References

1. Hao X, Su X, Horaud R, et al. Fullsubnet: A full-band and sub-band fusion model for real-time single-channel speech enhancement[C]// ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021: 6633-6637.
1. Liu J, Zhang X. Iicrn: Inplace cepstral convolutional recurrent neural network for monaural speech enhancement[C]// ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.