Real Time Speech Enhancement using DeepFilterNet2 on low computation devices

Tan Cher Yang, Sophie Vilter, Tai Kai Jie, Chua Bok Leong

Introduction

Although deep learning has been employed for speech enhancement in the recent years, there still exists a trade-off between performance and speed. Speech denoising models that work at low signal to noise ratio (SnR) could not be employed in devices that have low computational power. We attempt to evaluate the feasibility of shrinking a popular neural network for speech enhancement DeepFilterNet2 and evaluate its performance.

Methodology

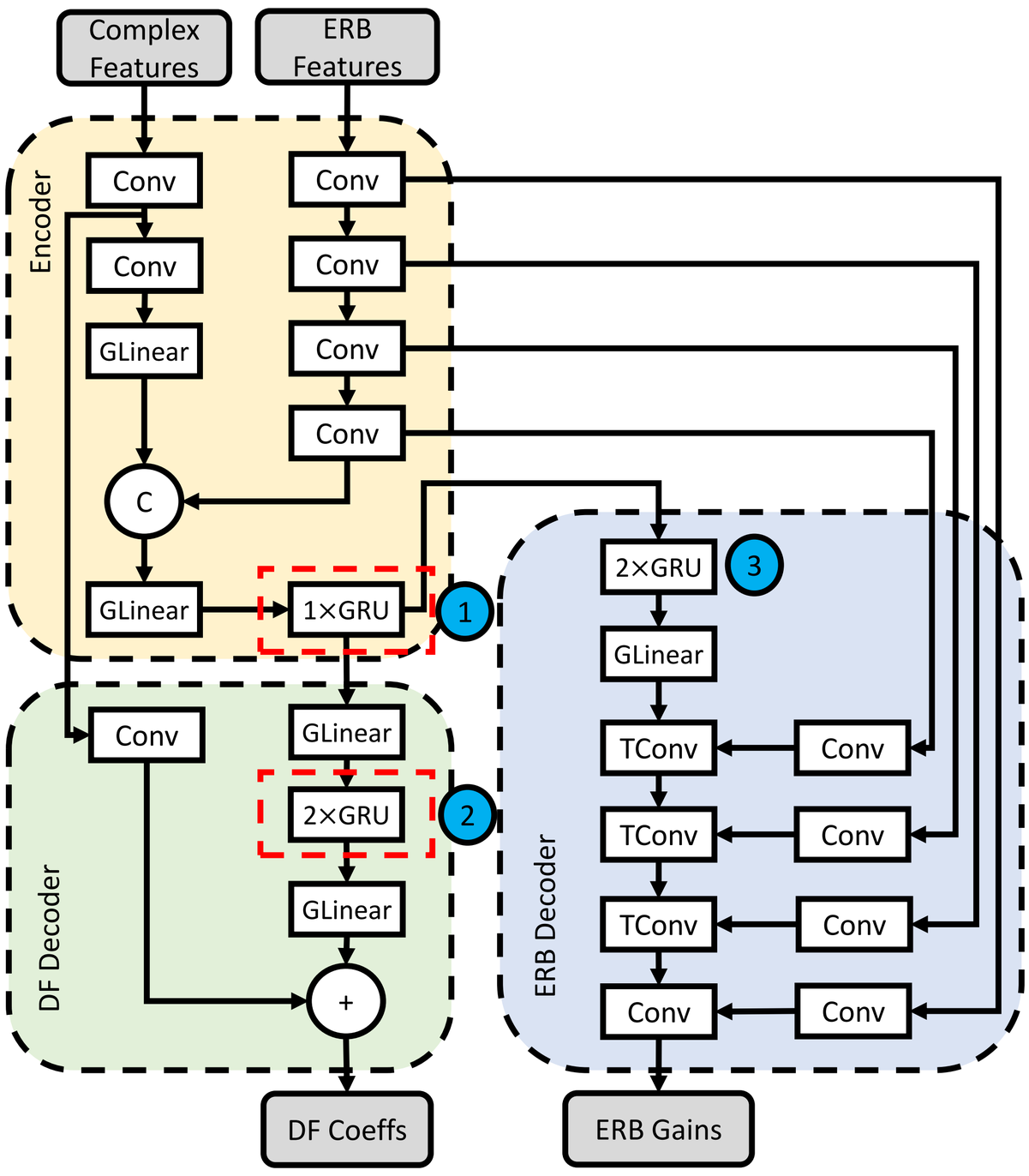

Figure 1: Architecture of DeepFilterNet2

Modules outlined in red ((1) and (2)) had their output dimensions reduced as shown in Figure 1. This would reduce the parameters of the other modules that depended on the reduced modules. Table 1 compares the total parameters of the original DeepFilterNet2 against the shrunk model and significant differences in parameters among the two models.

Table 1: Differences in number of parameters

Parameters	DeepFilterNet2	Shrunk DeepFilterNet2	AugNet
Module (1)	394,752	154,560	-
Module (2)	789,504	309,120	-
Module (3)	789,504	309,120	-
Total	2,306,236	1,006,772	1,000,000

Thereafter, the model was trained using the VCTK Speech Dataset and the DEMAND Noise Dataset. Thereafter, it is evaluated using noisy speeches generated from Microsoft's Deep Noise Suppression Challenge 4. Comparism was also made against Winer Filtering and AugNet noise suppression model provided by AugmentedHearing.io. This is illustrated in Figure 2.

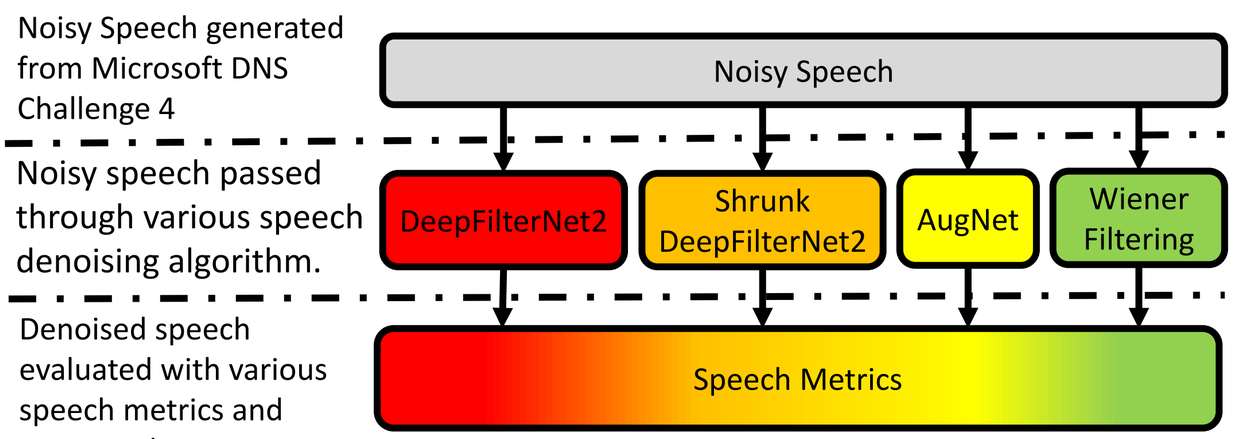

Figure 2: Experiment steps

Results

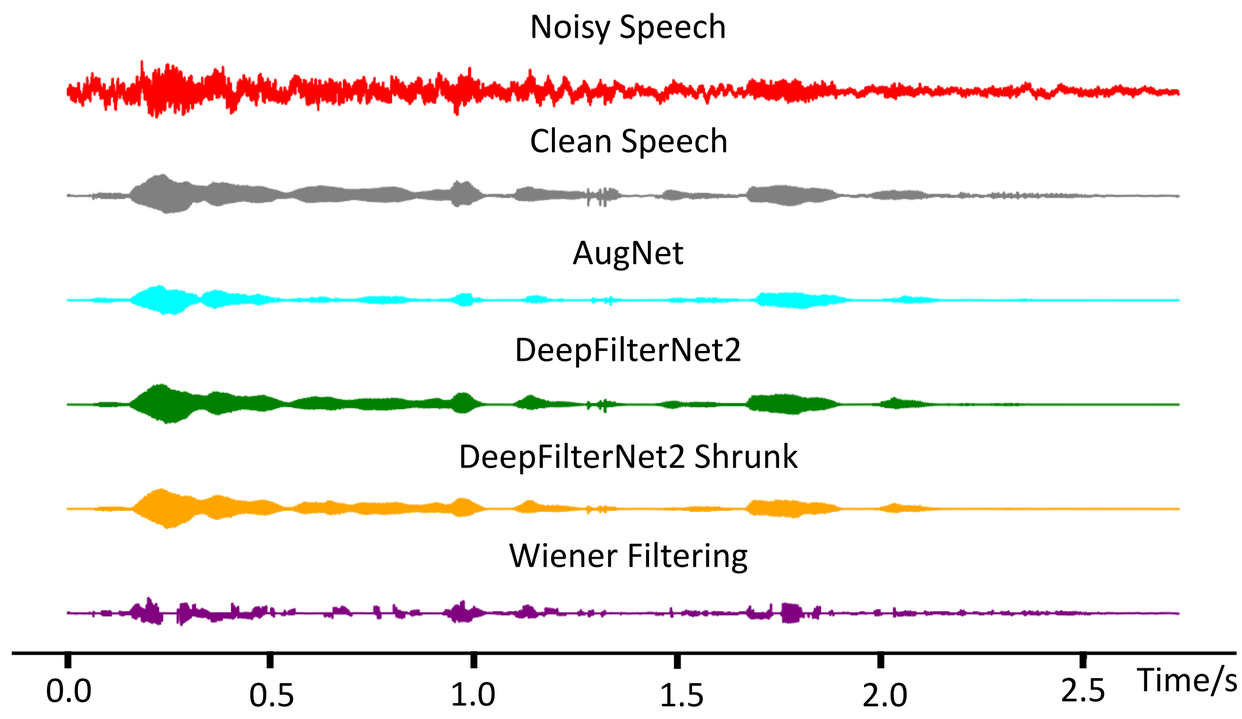

Figure 3: Comparison of denoised speech through various algorithms at -10dB SNR

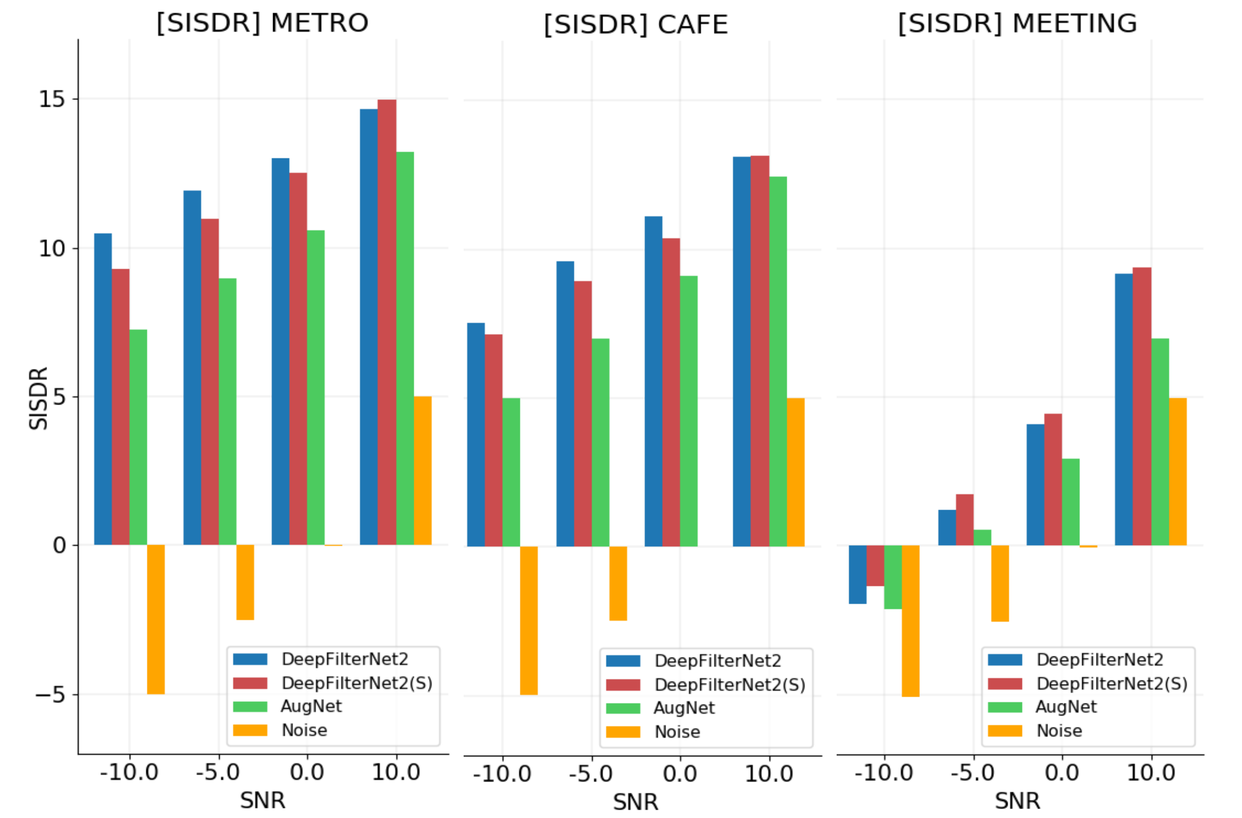

Figure 4: Scale-invariant signal-to-distortion ratio of different models at various environments

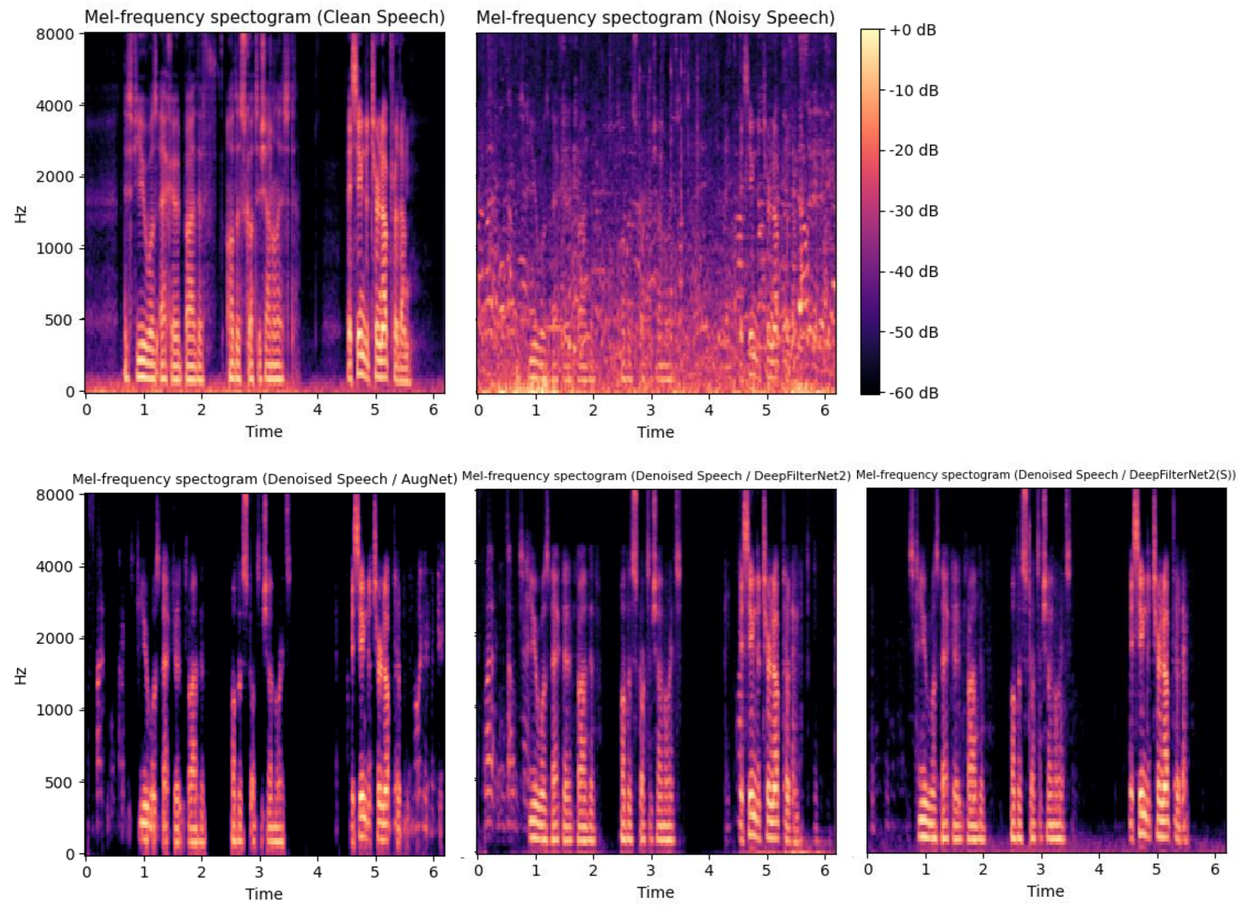

Figure 5: Mel-spectrogram of the various speeches at -10dB SNR

References

I. Schröter, H., Rosenkranz, T., & Maier, A. (2022). DeepFilterNet2: Towards Real-Time Speech Enhancement on Embedded Devices for Full-Band Audio. arXiv preprint arXiv:2205.05474.

II. Ge, X., Han, J., Long, Y., & Guan, H. (2022). PercepNet+: A Phase and SNR Aware PercepNet for Real-Time Speech Enhancement. *arXiv preprint arXiv:2203.02263*.

III. Valin, J. M. (2018, August). A hybrid DSP/deep learning approach to real-time full-band speech enhancement. In *2018 IEEE 20th international workshop on multimedia signal processing (MMSP)* (pp. 1-5). IEEE.