音频去噪的任务是从录制的语音中分离背景噪声，同时保留原语音数据的感知质量与清晰度。它在音频通话、语音识别、设备通信和生物监测等方面均有重要应用。然而，由于实际环境中存在噪声，不可避免的影响影响目标信号，环境噪声包括但不限于人类环境噪声、空气噪声、水下噪声等。

过去的几十年间，传统的信号处理方法，例如谱减法[[1]](#endnote-1)、维纳滤波[[2]](#endnote-2)、小波去噪[[3]](#endnote-3)等，已经被引入至音频去噪领域，但无法很好推广到非平稳或结构化噪声的情况，例如汽车喇叭、儿童玩耍、狗吠等。

早在20 世纪 80 年代，神经网络就被引入用于语音增强[[4]](#endnote-4)。后逐渐发展，取得了最先进的结果。这些模型通常在基于数据的监督环境中进行训练，在给定输入噪声音频的情况下，以干净音频信号为目标来学习预测干净的音频信号。这种音频信号监督学习的策略被称为noisy-clean training(NCT)。然而，这种基于数据的方法有一定的局限性，特别是在室外以及大自然等环境中，我们几乎无法获得干净的样本进行标注。在实际应用中，NCT这一类的监督去噪算法往往面临着成本高、数据少等问题。

为了克服这种限制，一些学者率先提出了Noisy-noisy strategy(NNT)[[5]](#endnote-5)[[6]](#endnote-6)与U-Net based deep audio prior[[7]](#endnote-7)，仅仅利用噪声语音数据就可以成功训练语音去噪深度神经网络。然而在现实世界中，前者很难满足其苛刻的噪声条件，后者受限于其过拟合问题。Consequently, researchers have focused on developing novel designs tailored to the specific audio of the target and the varying characteristics of environmental noise.

Wu et al.[[8]](#endnote-8)used the complex-valued speech denoising network DCUnet10 for noise reduction of vocal speech and introduced the complex-valued residual block cTSTM to model the correlation between amplitude and phase information.

Zhou et al.[[9]](#endnote-9) used the classical Unet network to estimate the self-noise spectrum of underwater AUVs and apply an improved spectral subtraction method to realize the self-noise suppression of AUVs, which significantly improves the signal-to-noise ratio.

Zhu et al.[[10]](#endnote-10) proposed the PriorDeNet, a prior-based denoising network, incorporating DenseNet's dense connection strategy to capture multi-scale features. Additionally, they enhanced the network's sensory field by integrating an expansive convolution kernel in the sense block

1. Steven Boll, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Transactions on acoustics speech and signal processing, 1979. [↑](#endnote-ref-1)
2. Jae Soo Lim and Alan V Oppenheim, "Enhancement and band-width compression of noisy speech", Proceedings of the IEEE, vol. 67, no. 12, pp. 1586-1604, 1979. [↑](#endnote-ref-2)
3. Priyadarshani N, Marsland S, Castro I, et al. Birdsong denoising using wavelets[J]. PloS one, 2016, 11(1): e0146790. [↑](#endnote-ref-3)
4. Shinichi Tamura and Alex Waibel, "Noise reduction using connectionist models", ICASSP, 1988. [↑](#endnote-ref-4)
5. Alamdari N, Azarang A, Kehtarnavaz N. Improving deep speech denoising by noisy2noisy signal mapping[J]. Applied Acoustics, 2021, 172: 107631. [↑](#endnote-ref-5)
6. Kashyap M M, Tambwekar A, Manohara K, et al. Speech denoising without clean training data: A noise2noise approach[J]. arXiv preprint arXiv:2104.03838, 2021. [↑](#endnote-ref-6)
7. Narayanaswamy, V.S., Thiagarajan, J.J., Spanias, A. (2021) On the Design of Deep Priors for Unsupervised Audio Restoration. Proc. Interspeech 2021, 2167-2171, doi: 10.21437/Interspeech.2021-1890 [↑](#endnote-ref-7)
8. Wu J, Li Q, Yang G, et al. Self-supervised speech denoising using only noisy audio signals[J]. Speech Communication, 2023, 149: 63-73. [↑](#endnote-ref-8)
9. Zhou W, Li J. Self-Noise Suppression for AUV without Clean Data: a Noise2Noise Approach[C]//2023 IEEE Underwater Technology (UT). IEEE, 2023: 1-5. [↑](#endnote-ref-9)
10. Zhu J, Cai W, Zhang M, et al. Self-supervised denoising model based on deep audio prior using single noisy marine mammal sound sample[J]. Applied Intelligence, 2023, 53(21): 25697-25714. [↑](#endnote-ref-10)