Supervised speech denoising methods are typically neural networks that map a noisy speech to an estimate of the clean speech. Supervised denoising methods are typically trained on pairs of clean speech .and noisy measurements **, where e is noise. We refer to supervised denoising as noise2clean (N2C). traditional N2C approaches[参考文献待补充] have access to clean training audio targets, and commonly employ a**  loss function to solve the following optimization:

**However, neural networks can also be trained on different noisy observations of the same clean speech. Noise2noise (N2N) assumes access to a set of pairs of noisy speech** , , where , are independent noise vectors. Specially, it employs noisy inputs and noisy targets during the training stage.

=

=

Where due to condition x.x. This causes the second term in Eqn X and the third tern in Eqn X to equal 0. Mathematically, the expectation of the is equal to the variance of plus the square of the expectation of . This fact is used to expand the third term in Eqn 6. The variance of the sample distribution Var() is equal to the variance of the population divided by the sampling size. Hence as the size of the noisy training dataset increases, the Noise2Noise loss value tends to equal the Noise2Clean loss value.

In theory N2N training reaches the same performance as N2C training if the dataset is

infinitely large. In practice, since the training set is limited in size, N2N falls slightly short of

N2C. For white noise, N2C performs slightly better than N2N in terms of SNR, SSNR, narrowband PESQ score (PESQ-NB), wideband PESQ score (PESQ-WB), and STOI[[1]](#endnote-1).

N2N outperforms N2C in some categories (e.g., sirens) of UrbanSound8K due to the network's better generalization ability to avoid falling into local optimum[[2]](#endnote-2)[[3]](#endnote-3).

Although Noise2Noise (N2N) demonstrates impressive performance, its practical application is frequently constrained. This limitation arises due to the challenge in acquiring pairs of noisy speech recordings from identical static scenes. For instance, ambient noise during audio recording can be variable and subject to rapid changes.

Therefore scholars targeted innovative designs for different characteristics of the target audio respectively. Wu et al.[[4]](#endnote-4)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.[[5]](#endnote-5) use 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.[[6]](#endnote-6) proposed a prior-based denoising network PriorDeNet and introduced the dense connection strategy in DenseNet to obtain multi-scale features, and introduced an expansion convolution kernel in the sense block to increase the sensory field of the network.

1. 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-1)
2. Alamdari N, Azarang A, Kehtarnavaz N. Improving deep speech denoising by noisy2noisy signal mapping[J]. Applied Acoustics, 2021, 172: 107631. [↑](#endnote-ref-2)
3. Zhou M, Liu T, Li Y, et al. Toward understanding the importance of noise in training neural networks[C]//International Conference on Machine Learning. PMLR, 2019: 7594-7602. [↑](#endnote-ref-3)
4. 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-4)
5. 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-5)
6. 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-6)