Regular paper | Journal of the KIECS. pp. 1103-1108, vol. 18, no. 6, Dec. 31. 2023, t. 122, pISSN 1975-8170 | eISSN 2288-2189
http://dx.doi.org/10.13067/JKIECS.2023.18.6.1103

# 이산 Wavelet 변환을 이용한 딥러닝 기반 잡음제거기

이행우*

# Noise Canceler Based on Deep Learning Using Discrete Wavelet Transform

Haeng-Woo Lee*

요 약

본 논문에서는 음향신호의 배경잡음을 감쇠하기 위한 새로운 알고리즘을 제안한다. 이 알고리즘은 이산 웨이블릿 변환(DWT: Discrete Wavelet Transform) 후 기존의 적응필터를 대신 FNN(: Full-connected Neural Network) 심층학습 알고리즘을 이용하여 잡음감쇠 성능을 개선하였다. 입력신호를 단시간 구간별로 웨이블릿 변환한 다음 1024-1024-512-neuron FNN 딥러닝 모델을 이용하여 잡음이 포함된 단일입력 음성신호로부터 잡음을 제거한다. 이는 시간영역 음성신호를 잡음특성이 잘 표현되도록 시간-주파수영역으로 변환하고 변환 파라미터에 대해 순수 음성신호의 변환 파라미터를 이용한 지도학습을 통하여 잡음환경에서 효과적으로 음성을 예측한다. 본 연구에서 제안한 잡음감쇠시스템의 성능을 검증하기 위하여 Tensorflow와 Keras 라이브러리를 사용한 시뮬레이션 프로그램을 작성하고 모의실험을 수행하였다. 실험 결과, 제안한 심층학습 알고리즘을 사용하면 기존의 적응필터를 사용하는 경우보다 30%, STFT(: Short-Time Fourier Transform) 변환을 사용하는 경우보다는 20%의 평균자승오차(MSE: Mean Square Error) 개선효과를 얻을 수 있었다.

ABSTRACT

In this paper, we propose a new algorithm for attenuating the background noises in acoustic signal. This algorithm improves the noise attenuation performance by using the FNN(: Full-connected Neural Network) deep learning algorithm instead of the existing adaptive filter after wavelet transform. After wavelet transforming the input signal for each short-time period, noise is removed from a single input audio signal containing noise by using a 1024-1024-512-neuron FNN deep learning model. This transforms the time-domain voice signal into the time-frequency domain so that the noise characteristics are well expressed, and effectively predicts voice in a noisy environment through supervised learning using the conversion parameter of the pure voice signal for the conversion parameter. In order to verify the performance of the noise reduction system proposed in this study, a simulation program using Tensorflow and Keras libraries was written and a simulation was performed. As a result of the experiment, the proposed deep learning algorithm improved Mean Square Error (MSE) by 30% compared to the case of using the existing adaptive filter and by 20% compared to the case of using the STFT(: Short-Time Fourier Transform) transform effect was obtained.

키워드

Noise Canceler, Deep Learning, Discrete Wavelet Transform, Fully-Connected Neural Network
잡음 제거기, 심층 학습, 이산 웨이블릿 변환, FNN

*남서울대학교 지능정보통신공학과 교수
• 접 수 일 : 2023. 10. 11
• 수정완료일 : 2023. 11. 11
• 게재확정일 : 2023. 12. 27
• Received : Oct. 11, 2023, Revised : Nov. 11, 2023, Accepted : Dec. 27, 2023
• Corresponding Author : Haeng-Woo Lee
Dept. Intelligent Information Communication Engineering, Namseoul University
Email : hwlee@nsu.ac.kr

## Ⅰ. 서 론

잡음에 의해 오염된 음성신호에서 잡음을 제거하는 기술인 음성개선기술은 음성인식 시스템의 전처리기에서 배경잡음의 제거 및 오래된 음반의 음질 향상과 보청기 등 여러 종류의 시스템에 활용되고 있으며 이에 관한 연구는 지속적으로 이루어지고 있다[1]. 잡음의 통계적 특성을 알지 못하는 경우 잡음 경감을 위한 기술로서 적응필터방법[2-3]이 있다. 이 방법은 음성신호의 준주기적 특성을 이용하며 필터의 계수를 자동적으로 조정하여 음성을 추정하는 기능을 가지고 있다. 그러나 음성신호와 같이 자기상관 행렬의 고유치 분포가 큰 경우 수렴속도가 늦어지는 단점이 있다.

일반적으로 시간에 따라 변하는 신호의 주파수 해석을 위하여 단구간 푸리에 변환(STFT: Short-Time Fourier Transform)이 이용되고 있으나 음성신호와 같이 비정재성(nonstationary)을 갖는 신호의 특성을 표현하는 것은 적합하지 않다. 이러한 문제를 극복하기 위해 널리 사용되고 있는 웨이블릿 변환은 다중해상도(multi-resolution)를 갖는 신호해석방법으로서 시간 및 주파수 영역에서 국부성(localization)을 가지므로 통계적 특성을 알지 못하거나 시간적으로 예측하기 어려운 신호해석에 유용하다. 웨이블릿 변환은 변환된 신호의 자기상관 행렬이 거의 대각요소로 집중되므로 신호의 고유치 비를 작게 할 수 있고 이 변환을 변환영역 적응필터에 이용하면 수렴속도가 향상된다. 이러한 웨이블릿 변환의 성질을 이용하여 신호의 압축 및 부호화, 잡음제거 등에 널리 이용되고 있다[4-6]. 웨이블릿 변환을 푸리에 변환과 비교해 보면 다음과 같은 이점이 있다. 첫째, 웨이블릿 변환은 신호의 저주파 부분에서 좋은 주파수 분해능을 얻을 수 있다. 둘째, 웨이블릿 변환은 고주파 부분에서 좋은 시간 분해능을 얻을 수 있다[7]. 이러한 장점으로 인해 웨이블릿 변환을 이용한 특징벡터를 사용하면 파열음이나 파찰음에서와 같이 시간-주파수 상에서 갑자기 튀는 국부적 특성을 잘 반영할 수 있기 때문에 더 나은 음성인식 성능을 얻을 수 있다. 그리고 웨이블릿 변환은 실수 웨이블릿을 이용하면 STFT와는 달리 복소수 연산없이 신호의 변환이 가능하고 모 웨이블릿(mother wavelet)으로 Haar 웨이블릿을 이용하면 곱셈 연산없이 변환이 가능하므로 회로 구조가 크게 간략화된다. 또한 변환영역에서 적응필터의 가중치 갱신은 입력신호를 일정시간 블록으로 나누어 변환한 후 변환된 각각의 값에 독립적으로 갱신된 각각의 가중치를 곱한 후 역변환한다. 이 방법은 한 블록의 신호를 받아들인 후 처리하므로 계산량이 적으며 회로 구현시 병렬처리 구조를 적용하기가 용이하다.

딥러닝은 신경망을 기반으로 많은 수의 은닉층을 사용하는 복잡한 머신러닝 모델이다[8]. 최근 딥러닝 모델이 여러 분야에서 큰 성과를 내고 있는 것은 많은 레이어로 이루어진 다층 신경망을 학습할 수 있는 기술이 개발되었기 때문이다. 다층 신경망을 학습시키는 오차 역전파(back propagation) 알고리즘[9]이 상위층을 학습하기 전에 먼저 하위층의 가중치(weight)를 미리 학습시킴으로서 많은 층으로 구성된 심층 신경망도 학습이 가능하게 되었다[10]. 현재 가장 많이 사용되는 딥러닝 모델은 Fully-connected Neural Network (FNN)[11]이다. 본 연구에서는 이산 웨이블릿 변환을 이용한 적응 잡음감쇠시스템에서 적응필터 대신에 신경망 필터의 심층학습(deep learning) 알고리즘을 이용하여 잡음을 감쇠시키는 방법을 제안한다.

논문의 내용은 Ⅱ절에서 이산 웨이블릿 변환에 대해 살펴보고, Ⅲ절에서는 잡음제거를 위한 변환영역 블록 딥러닝 모델을 제안하였다. 그리고 Ⅳ절에서 이 시스템에 대한 시뮬레이션 및 그 결과에 대하여 기술하였고, 끝으로 Ⅴ절에서 결론을 도출하였다.

## Ⅱ. 이산 Wavelet 변환

웨이블릿 변환은 기저함수들의 집합에 의한 신호의 분해로서 이해할 수 있다. 이때 웨이블릿 변환에서 하나의 기저함수를 웨이블릿이라 하며 웨이블릿은 하나의 대역통과필터라고 할 수 있다. 모든 주파수에 대해 균일한 시간 분해능을 제공하는 STFT와 달리 웨이블릿 변환은 고주파수에 대해서는 높은 시간 분해능과 낮은 주파수 분해능을 제공하고 저주파수에 대해서는 높은 주파수 분해능과 낮은 시간 분해능을 제공한다. 이는 유사한 시간-주파수 분해능 특성을 나타내는 인간의 귀와 매우 흡사하다.

웨이블릿 변환은 모 웨이블릿을 시간축으로 이동(shifting)시키고 스케일링(scaling)한 여러 웨이블릿

기저(basis)들을 이용하여 신호를 분석한다. 다해상도 신호해석에서 주어진 함수는 다른 해상도를 가진 연속적인 추정치들의 합계로써 표현되며 스케일링 함수라고 하는 저주파 통과 커널과 콘볼루션함으로써 이루어진다. DWT는 식 (1)로 정의되며 시간 이동 및 스케일링 파라미터가 이산적인 값을 갖는다. 이 식에서 $\psi^*$는 원형 웨이블릿이며, 이를 신호의 주파수에 따라 스케일링 파라미터 j와 시간 이동 파라미터 k에 의해 변형되어 적용된다. 즉, 푸리에 변환과 같이 고정된 크기의 창함수를 사용하지 않고 짧은 지속시간을 갖는 고주파신호에 대해서는 짧은 창함수를 사용하고 긴 지속시간을 갖는 저주파신호에 대해서는 긴 창함수를 이용함으로써 주파수 영역에 따른 다중 해상도를 갖는다.

$$y[n] = \sum_{k=0}^{K-1} x[k] \cdot \psi^*[n-k] \quad \cdots (1)$$

실제 응용에서는 a와 b를 2의 지수 형태로 나타낸 dyadic 웨이블릿 변환(DyWT: Dyadic Wavelet Transform)이 많이 이용된다. 원 이산신호는 다해상도 분석의 다운 샘플링을 통해 주파수가 다른 여러 개의 부대역(sub-band)으로 분해되고, 업 샘플링을 통해 원 이산신호로 합성된다. 각 레벨에서 시간영역 이산신호는 저역통과필터 $H(z)$를 통해 근사(approximation) 성분과 고역통과필터 $G(z)$를 통해 상세(detail) 성분으로 분해된다.

$$\begin{aligned} x(t) &= \sum_{j,k} a_j(k)\Phi_{j,k}(t) + \sum_{j,k} d_j(k)\Psi_{j,k}(t) \quad (2) \\ &= H(z) + G(z) \end{aligned}$$

근사성분은 근사계수 $a_j(k)$와 척도함수(Scale function) ϕ(t)의 곱으로 표현되고 상세성분은 상세계수 $d_j(k)$와 상세함수(Detail function) ψ(t)의 곱으로 구성되며 척도함수와 상세함수는 서로 직교한다. 여기서 j와 k는 각각 이산 변환에서의 스케일과 시간영역에서의 이동을 나타낸다.

$$\Phi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \Phi\left(\frac{t-k2^j}{2^j}\right) \quad \cdots (3)$$

$$\Psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \Psi\left(\frac{t-k2^j}{2^j}\right) \quad \cdots (4)$$

각 레벨의 근사성분이 2배 간축된 출력은 그 다음 레벨의 입력신호가 된다. 각 레벨에서 근사계수는 입력신호와 척도함수가 결합되고 상세계수는 입력신호와 상세함수가 결합된 형태로 표현된다.

$$a_j(k) = \int_{-\infty}^{\infty} x(t) \cdot \phi_{j,k}(t)\, dt \quad \cdots (5)$$

$$d_j(k) = \int_{-\infty}^{\infty} x(t) \cdot \psi_{j,k}(t)\, dt \quad \cdots (6)$$

웨이블릿 변환은 신호처리 관점에서 대역통과필터 뱅크의 출력으로 볼 수 있으며, 신호를 분할하기 위해서 그림 1과 같이 일반적인 트리 형태의 웨이블릿 분해 필터뱅크를 구성한다. 입력신호가 저역통과필터와 고역통과필터를 거치고 2배의 간축(decimation) 과정을 거치게 되면 한 번의 웨이블릿 변환이 수행되며, 이러한 과정을 원하는 스케일까지 반복적으로 수행하면 웨이블릿 변환된 신호를 얻을 수 있다. 그리고 일반적인 웨이블릿 변환에서 각 스케일은 상위 스케일에서 2배로 간축하여 구해지므로 각 스케일의 샘플 수가 상위 스케일의 절반이 된다.

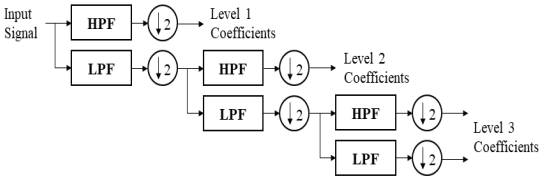



그림 1. Dyadic 웨이블릿 변환의 필터뱅크 구조
Fig. 1 Filter bank of dyadic wavelet transform

## III. 잡음제거를 위한 변환영역 블록 딥러닝 알고리즘

본 논문에서는 기존의 적응필터 대신에 딥러닝 기술을 적용하여 잡음감쇠 성능을 개선하고자 한다. 변

환영역 딥러닝 방법은 입력음성의 웨이블릿 변환계수를 목표값인 순수 음성의 변환계수와 같아지도록 신경망의 가중치를 조정해 나간다. 그림 2는 웨이블릿 변환을 이용하여 기준신호의 변환계수를 추정하는 적응 잡음감쇠시스템이다.

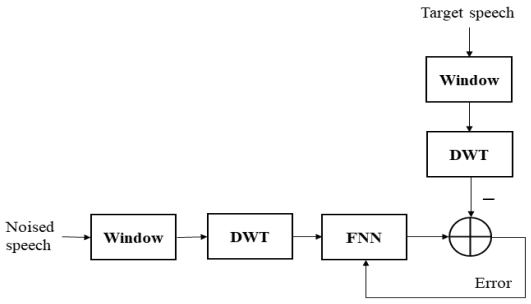



그림 2. 새로운 딥러닝기반 잡음감쇠시스템
Fig. 2 Noise reduction system based deep learning

이 시스템에는 잡음이 포함된 음성신호 뿐만 아니라 순수한 음성신호도 입력되어 딥러닝 학습의 목표값으로 사용된다. 각 입력신호의 이산 웨이블릿 변환을 구하기 위해 신호를 음성의 통계적 특성이 변하지 않는 32ms 구간으로 나누고 512 샘플에 대해 해밍(Hamming) 윈도우 함수와 곱한다. 그 다음 이 블록에 대해 9 레벨의 웨이블릿 변환계수를 산출하고 변환계수들을 1차원 배열로 정리한다. 이 배열이 딥러닝의 입력 데이터로 사용된다.

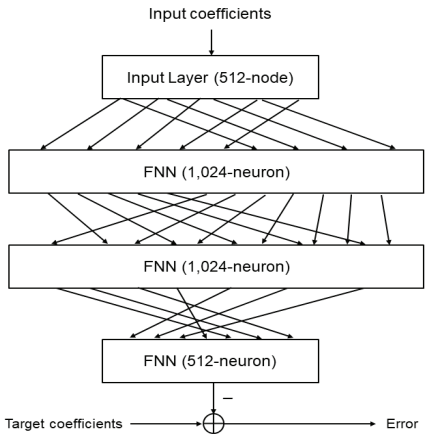



그림 3. 딥러닝 모델 구조
Fig. 3 Structure of deep learning model

딥러닝 모델은 그림 3과 같이 FNN 구조를 이용한다. 입력층과 출력층은 512-neuron으로 구성되고 2개의 1,024-neuron으로 이루어진 은닉층이 중간에 배치되어 있다. 따라서 총 파라미터는 2,097,152 개의 weight 및 2,560 개의 bias가 존재한다. 이 파라미터들은 주어진 데이터를 사용하여 오차 역전파(back propagation) 알고리즘과 Adam 적응 알고리즘으로 지도학습을 통해 샘플마다 업데이트된다. 학습과정에서 매 샘플마다 MSE(: Mean Square Error)와 MAE(: Mean Absolute Error)를 모니터링할 수 있다.

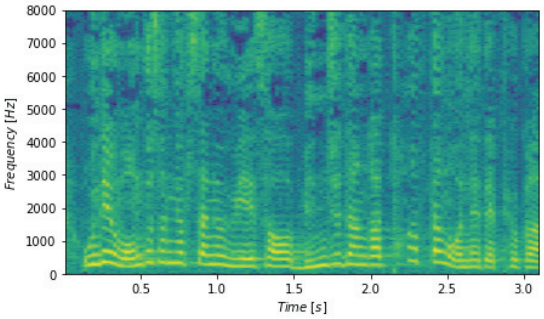
표 1. 잡음감쇠시스템의 주요 사양
Table 1. Specifcations of the noise reduction system

| Items | Values |
|---|---|
| Sampling frequency | 16 kHz |
| Data resolution | 16-bit |
| Window type | Hamming |
| Window size | 512 |
| Window overlap size | 256 |
| Wavelet type | Haar |
| Transform level depth | 9 levels |
| Batch size | 96 |
| Optimization algorithm | Adam |

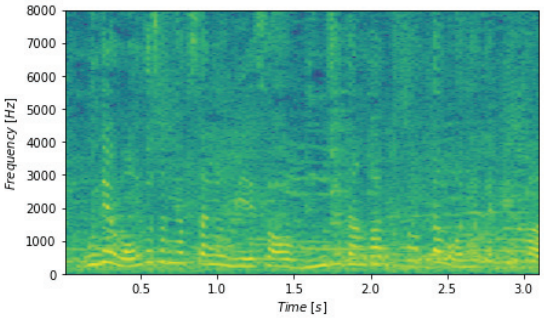
본 잡음감쇠시스템을 구현하는데 가장 적합한 주요 사양이 표 1에 나열되어 있다. 이 값들은 여러 시뮬레이션을 통해 가장 성능이 우수한 것으로 선정되었다.

## Ⅳ. 모의실험 및 분석

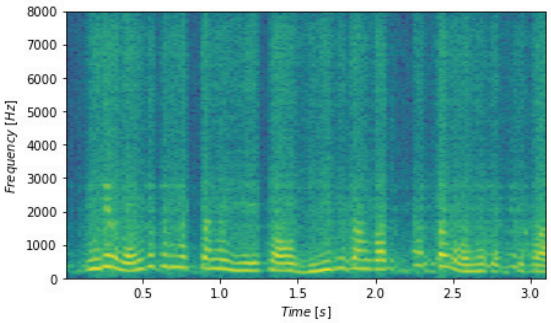
본 논문에서 제안한 음성잡음감쇠시스템의 성능을 검증하기 위해 Tensorflow와 Keras 라이브러리를 이용하여 시뮬레이션 프로그램을 작성하였다. 입력신호는 음성과 잡음의 혼합신호와 순수 음성신호가 사용되며 16-bit, 16kHz로 샘플링된 500,000 샘플(28.125 sec) 데이터가 제공된다. 이 시스템은 지도학습에 해당되며 입력데이터는 내부적으로 512×(500,000-511) 샘플의 입력배열과 (500,000-511) 샘플의 목표값으로 구성된다.




(a) 입력 음성신호의 스펙트로그램




(b) 잡음이 포함된 입력신호의 스펙트로그램




(c) 출력신호의 스펙트로그램

그림 4. 입출력신호의 스펙트로그램
Fig. 4 Spectrograms of input and output signal

그림 4는 음성과 잡음 혼합신호 및 음성 추정신호의 스펙트로그램을 보여준다. 잡음에 의해 주파수 특성이 크게 훼손된 상태에서 출력신호는 상당히 복원된 모습을 볼 수 있다.

구현방법들의 성능을 평가하기 위하여 평균제곱오차(MSE: Mean Square Error)와 평균절대오차(MAE: Mean Absolute Error)를 사용하였다. MSE는 목표값인 기준신호의 변환계수와 입력신호의 변환계수 간 오차의 제곱에 대한 평균값, MAE는 오차의 절대값에 대한 평균값을 나타낸다. 그림 5에서 3가지 구현방법에 대한 MSE 특성을, 그림 6에서는 MAE 특성을 보여주고 있다.

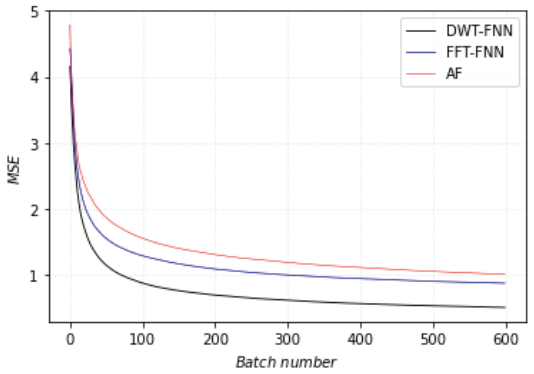



그림 5. 알고리즘별 평균제곱오차의 특성 비교
Fig. 5 Comparison of MSE for algorithms

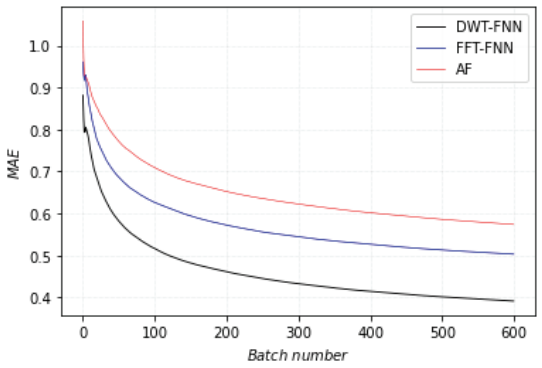



그림 6. 알고리즘별 평균절대오차의 특성 비교
Fig. 6 Comparison of MAE for algorithms

기존의 적응필터(AF)와 푸리에 변환(FFT-FNN) 및 웨이블릿 변환(DWT-FNN) 등을 사용한 경우 MSE 성능은 웨이블릿 변환을 사용한 딥러닝 모델이 가장 우수한 것으로 나타났다. 시뮬레이션 결과, 웨이블릿 변환을 사용했을 때 푸리에 변환 대비 20%, 적응필터 대비 30%의 MSE가 개선되는 것을 보여주었다. 그 이유는 적응필터는 시간영역에서 선형 잡음만을 감쇠시키지만 푸리에 변환방법은 주파수 영역, 그리고 웨이블릿 변환

방법은 음성신호에 적합한 시간 및 주파수 영역에서 잡음을 감쇠시키기 때문이다. 또한 MAE 성능도 MSE 성능과 마찬가지로 웨이블릿 변환을 사용하는 경우에 가장 우수한 성능을 달성하는 것으로 나타났다.

## V. 결 론

보청기의 음성청취 성능을 개선하기 위하여 우수한 잡음감쇠기의 개발이 요구되고 있다. 본 논문에서는 웨이블릿 변환과 딥러닝 기술을 적용한 새로운 잡음감쇠시스템을 제안하였다. 웨이블릿 변환계수에 대하여 기존의 적응필터 대신 FNN 신경망을 이용한 심층학습으로 잡음감쇠 성능을 향상시킬 수 있다.

본 잡음감쇠시스템은 입력신호의 웨이블릿 변환 후 1024-1024-512-neuron FNN 딥러닝 학습을 통하여 상당한 성능 개선을 달성하였다. 연구 결과, 제안한 시스템은 푸리에 변환방법 대비 20%, 적응필터 대비 30%의 MSE 감쇠효과를 얻었고 MAE도 비슷한 성능개선을 보였다.

감사의 글

이 논문은 2023년도 남서울대학교 학술연구비 지원에 의해 연구되었음.

## References

[1] H. Lee, "Nonlinear noise attenuator by adaptive Wiener filter with neural network," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 18, no. 1, 2023, pp. 71-76.

[2] S. F. Boll and D. C. Pulsipher, "Suppression of acoustic noise in speech using two microphone adaptive noise cancellation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, no. 6, Dec. 1989, pp. 752-753.

[3] W. A. Harrison, J. S. Lim, and E. Singer, "A new application of adaptive noise cancellation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, Feb. 1986, pp. 21-27.

[4] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. on Information Theory*, vol. 38, no. 2, 1992, pp. 617-643.

[5] M. J. Shensa, "The Discrete Wavelet Transform: Wedding the A Trous and Mallat Algorithms," *IEEE Trans. on Signal Processing*, vol. 40, no. 10, 1992, pp. 2464-2482.

[6] D. L. Donoho, "De-Noising by Soft-Thresholding," *IEEE Trans. on Information Theory*, vol. 41, no. 3, 1995, pp. 613-627.

[7] I. Daubechies, "The Wavelet Transform Time-Frequency Localization and Signal Analysis," *IEEE Trans. on Information Theory*, vol. 36, no. 5, 1990, pp. 961-1005.

[8] H. Lee, "Optimization of the number of filter in CNN noise attenuator," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 16, no. 4, 2021, pp. 625-632.

[9] D. Rumelhart, G. Hinton, and R. Williams, "Learning representations by back-propagating errors," *Cognitive modeling*, vol. 5, 1988, pp. 3.

[10] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, 2015, pp. 85-117.

[11] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE*, vol. 86, no. 11, Nov. 1998, pp. 2278-2324.

## 저자 소개

**이행우(Haeng-Woo Lee)**

1985 광운대학교 전자공학과(학사)
1987 서강대학교 대학원 전자공학과(석사)
2001 전북대학교 대학원 전자공학과(박사)
1987~1998 한국전자통신연구원 선임연구원
2001~ 남서울대학교 지능정보통신공학과 교수
※관심분야 : VLSI 설계, 딥러닝, 배경잡음 감쇠