한국음향학회지 제38권 제6호 pp. 697~702 (2019)
The Journal of the Acoustical Society of Korea Vol.38, No.6 (2019)
https://doi.org/10.7776/ASK.2019.38.6.697

pISSN : 1225-4428
eISSN : 2287-3775

배경음악 분리를 위한 확장된 합성곱을 이용한 멀티 밴드 멀티 스케일 DenseNet

Multi-band multi-scale DenseNet with dilated convolution for background music separation

허운행,[1] 김혜미,[2] 권오욱[3†]

(Woon-Haeng Heo[1], Hyemi Kim[2], and Oh-Wook Kwon[3†])

[1]충북대학교 일반대학원 제어로봇공학전공, [2]한국전자통신연구원 차세대콘텐츠연구본부, [3]충북대학교 전자공학부

(Received September 10, 2019; accepted September 20, 2019)

초 록: 방송 콘텐츠의 혼합 신호에서 배경음악 신호를 분리하는 확장된 합성곱을 이용한 멀티 밴드 멀티 스케일 DenseNet을 제안한다. 확장된 합성곱은 스펙트로그램의 다양한 스케일 문맥 정보를 학습하기 용이하도록 한다. 컴퓨터 모의실험 결과, 제안한 구조는 신호대잡음비(Signal to Noise Ratio, SNR) 0 dB, -10 dB의 환경에서 각각 0.15 dB, 0.27 dB의 신호대왜곡비(Signal to Distortion Ratio, SDR)를 개선하였다.
핵심용어: 방송 콘텐츠, 배경음악 분리, 확장된 합성곱, DenseNet

ABSTRACT: We propose a multi-band multi-scale DenseNet with dilated convolution that separates background music signals from broadcast content. Dilated convolution can learn the multi-scale context information represented by spectrogram. In computer simulation experiments, the proposed architecture is shown to improve Signal to Distortion Ratio (SDR) by 0.15 dB and 0.27 dB in 0dB and −10 dB Signal to Noise Ratio (SNR) environments, respectively.
Keywords: Broadcast content, Background music separation, Dilated convolution, DenseNet

PACS numbers: 43.60.Uv, 43.75.Zz

I. 서 론

방송 콘텐츠에서 배경음악은 저작권과 관련하여 민감한 문제를 가진다. 방송물에서 배경음악으로 쓰이는 음악 제목, 음악 구간 등의 정보는 사람이 입력한다. 사람이 직접 입력하기 때문에 정보가 정확하지 않고, 시간과 노동이 많이 들어간다. 이러한 문제점을 해결하기 위하여 자동 음악 검색 기술이 필요하다. 방송 콘텐츠에서는 음악과 음성이 혼합된 구간이 많이 존재하고, 보통 음악보다 음성이 더 큰 소리로 혼합되는 특징이 있다. 이러한 이유로 자동 음악 검색 기술을 위한 배경음악 분리 기술이 필요하다. 본 연구에서는 방송 콘텐츠의 음악과 음성이 혼합된 신호에서 배경음악을 분리하였다.

혼합된 신호에서 음악 신호를 분리하기 위해서 신호 분리 태스크에서 사용하는 모델을 이용하였다. 기존부터 혼합된 신호에서 원하는 신호를 분리하는 기술은 많은 연구가 이루어지고 있다. 기존에는 Non-negative Matrix Factorization(NMF)을 이용하여 원하는 신호를 얻었지만, 성능이 좋지 않았다.[1] 최근 딥러닝 기술이 적용된 이후부터 좋은 성능을 보여주어 딥러닝 구조에 대한 연구가 이루어지고 있다.[2,3]

보통 음향 신호의 스펙트로그램을 입력으로 사용하는데, 스펙트로그램의 크기를 이미지로 생각할 수 있다. 이미지에서 좋은 성능을 보이는 딥러닝 구조

†**Corresponding author:** Oh-Wook Kwon (owkwon@cbnu.ac.kr)
School of Electronics Engineering, Chungbuk National University, Chungdae-ro 1, Seowon-gu, Cheongju 28644, Republic of Korea
(Tel: 82-43-261-3374, Fax: 82-43-268-2386)

를 응용한 연구들도 많이 있다. 음악 신호 분리 태스크에서 사용한 U-Net,[4] Densely connected convolutional network(DenseNet)[5] 구조는 이미지 분할이나 분류 태스크에서 이미 좋은 성능을 보여주었다.[6,7]

Convolutional Neural Network(CNN) 기반 U-Net은 수용범위(receptive field)를 효과적으로 늘이기 위하여 다운 샘플링 과정인 인코더와 업 샘플링 과정인 디코더 구조를 가진다. 또한, 정보 전달과 오류 전파에 용이하도록 다운 샘플링과 업 샘플링 과정에서 같은 크기의 특징맵(feature map)을 연결하였다.

스펙트로그램 도메인에서 동작하는 U-Net과 다르게 시간 도메인에서 동작하는 Wave-U-Net[8]은 더 좋은 성능을 보인다. 일반적으로 스펙트로그램 도메인에서 동작하는 시스템은 시간 도메인의 신호로 복원할 때 혼합 신호의 위상을 이용한다. 혼합 신호의 위상을 이용하기 때문에 추정된 신호에 오류가 생긴다. 이러한 단점을 해결하기 위해서 시간 도메인의 신호에서 바로 원하는 신호를 추정한다. Wave-U-Net이 U-Net보다 좋은 성능을 보이지만, Bidirectional Long Short-Term Memory(BLSTM)과 DenseNet 구조보다는 낮은 성능을 보인다.[9]

DenseNet 구조는 최근 음악 소스 분리에서 가장 좋은 성능을 보인 구조이다.[5,10,11] Multi-scale Multi-band DenseNet(MMDenseNet),[5] MMDenseLSTM[11]은 스펙트로그램 도메인에서 DenseNet 구조를 이용하여 신호를 분리하였다. 두 구조는 모두 CNN 기반이고, 인코더와 디코더 구조를 가진다. DenseNet은 각 합성곱(convolution)의 입력과 출력을 매번 연결하여 정보 전달에 용이한 장점이 있다.

본 연구에서는 dilated convolution[12]을 DenseNet 구조에 추가하여 dilated dense block을 만들었다. Dilated convolution은 해상도를 잃지 않고 문맥 정보를 효과적으로 학습할 수 있다.

II. 베이스라인

2.1 DenseNet

DenseNet[5]은 Fig. 1처럼 여러 개의 composite function으로 이루어져 있다. Ⓒ는 입력과 출력 특징맵의 연결을 의미한다. Composite function은 Fig. 2와 같이 Batch

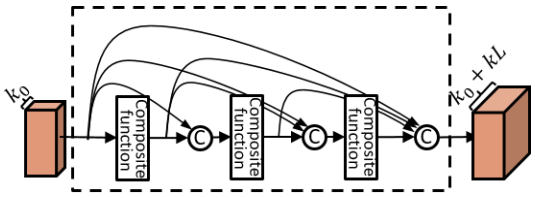

Fig. 1. Dense block.

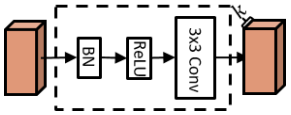

Fig. 2. Composite function.

Normalization(BN),[13] Rectified Linear Units(ReLU),[14] 그리고 3 × 3 커널의 합성곱("Conv")의 연속적인 함수들로 구성된다.

일반적인 심층 신경망에서 l layer에서의 출력을 아래의 Eq. (1)처럼 표현할 수 있다.

$$x_l = H_l(x_{l-1}), \tag{1}$$

여기서 x_{l-1}은 $l-1$ layer의 출력이자 l layer의 입력이다. 비선형 변환함수 H_l는 composite function을 말한다.

DenseNet은 composite function의 입력과 출력의 특징맵을 연결하는 방법을 제안하였다. Eq. (2)은 DenseNet의 입력, 출력을 식으로 표현한 것이다.

$$x_l = H_l([x_0, x_1, ..., x_{l-1}]), \tag{2}$$

여기서 $[x_0, x_1, ..., x_{l-1}]$은 0부터 $l-1$ 층까지의 출력 특징맵이 연결된 것을 말한다. 이러한 연결 방식은 오류가 각 층으로 바로 전달되어 효과적이고, feed forward 과정에서 이전 층의 출력이 모두 입력으로 사용되므로 층이 깊어질수록 정보가 약해지는 단점을 보완할 수 있다.

Composite function을 넘어가는 선들은 각 층의 입력과 출력 특징맵이 연결되는 것을 나타낸다. Composite function의 출력 특징맵 개수를 growth rate k로 나타낸다. Dense block의 최종 특징맵 개수는 $k_0 + k \times L$로 나타낼 수 있다. k_0는 dense block 입력의 특징맵 개수이고 composite function을 지나갈 때, 입력

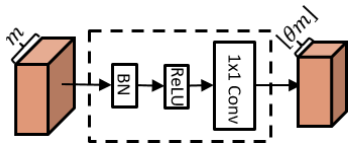

Fig. 3. Compression.

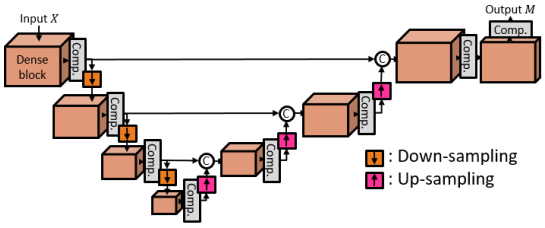

Fig. 4. MDenseNet architecture.

되었던 특징맵이 연결되므로 growth rate k와 composite function 개수 L을 이용하여 위와 같은 식으로 나타낼 수 있다.

DenseNet의 연결 특성상 growth rate, composite function 개수와 dense block의 개수에 따라 특징맵이 아주 많아진다. 이러한 구조의 확장성을 줄이고자 Fig. 3과 같은 특징맵 개수를 줄이는 compression을 dense block 뒷단에 추가한다. 만약 dense block의 출력 특징맵이 $m = k_0 + k \times L$이면, compression에서 $\theta \times m$의 정수로 특징맵 개수를 줄여준다. Compression rate θ는 $0 < \theta \leq 1$범위의 값이다.

2.2 Multi-scale DenseNet(MDenseNet)

MDenseNet[5]은 음악 신호 분리 태스크에 DenseNet 구조를 적용시킨 구조이다. Fig. 4와 같이 다운 샘플링과 업 샘플링 과정을 통하여 멀티 스케일 특징을 얻는 구조로 만들었다.[5] Dense block 옆의 $Comp.$는 compression을 의미한다. 입력 X은 스펙트로그램이 되고, 출력 $\hat{M}$은 입력과 곱해질 마스크가 추정된다.

회귀 태스크인 신호 분리에서는 네트워크의 출력이 입력 크기와 같아야 하므로 다운 샘플링이 된 보틀넥 특징[4]을 입력 크기로 복원하는 업 샘플링 과정이 필요하다. 다운 샘플링은 2 × 2 커널의 average pooling을 이용하였고, 업 샘플링은 2 × 2 커널의 transposed convolution[15]을 이용하였다. Dense block에서 composite function의 입력, 출력 특징맵들을 연결한 것과 같이 블록간의 출력 특징맵을 연결하였다.

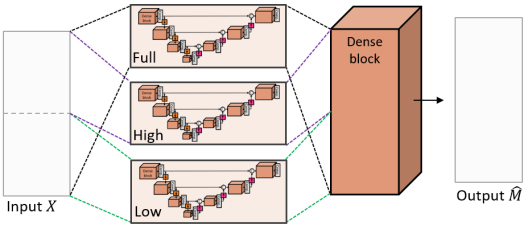

Fig. 5. MMDenseNet architecture.

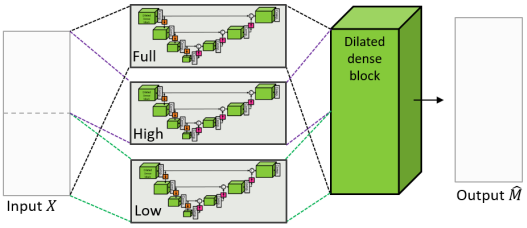

Fig. 6. DMMDenseNet architecture.

2.3 MMDenseNet

MMDenseNet은 Fig. 5와 같이 스펙트로그램의 주파수에 따라 서로 다른 특성의 패턴을 가지는 것을 반영한 구조이다. 이러한 특성을 딥러닝 구조에 적용하기 위하여 입력의 주파수 축을 반으로 나눈 각 밴드 "Low", "High"와 전체 주파수 "Full"에 대하여 서로 다른 MDenseNet 구조를 삽입하였다.[5]

III. 제안 구조

수용범위를 효과적으로 늘이는 또 다른 방법은 dilated convolution[12]을 사용하는 것이다. 이 방법은 시맨틱 분할 태스크에서 좋은 결과를 보여주었다. 본 연구에서는 dilated convolution을 dense block에 추가한 소스 분리 구조 dilated MMDenseNet(DMMDenseNet)을 제안한다.

전체적인 구조도는 Fig. 6과 같다. 기존 MMDenseNet의 모든 dense block 앞단에 Fig. 7과 같이 dilated block을 삽입하였다. Dilated block의 구조는 BN-ReLU 함수 다음에 Dilated Convolution("DConv")과 일반적인 합성곱이 병렬적으로 배치된다. 각 합성곱은 k개의 특징맵을 출력하고, 입력 특징맵과 연결한다.

Fig. 8은 [2, 1], [1, 2], [2, 2]의 dilation ratio를 가지는 5 × 3, 3 × 5, 5 × 5 커널의 합성곱 과정을 보여준다. 본 연구에서는 위의 세 가지 dilated convolution에 대하

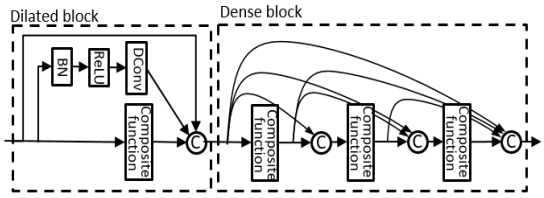

Fig. 7. Dilated dense block.

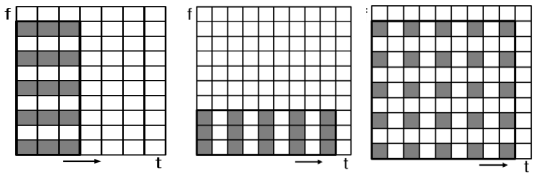

Fig. 8. [2, 1], [1, 2], [2, 2] dilated convolution (5 × 3, 3 × 5, 5 × 5 kernel size).

여 각각 실험하였다. 합성곱의 dilation ratio에 따라 [2 ,1]은 frequency dilated convolution, [1, 2]는 time dilated convolution이라 한다.

제안한 신경망의 출력은 입력 스펙트로그램에 곱해질 마스크이다. 입력 스펙트로그램을 X, 신경망에서 추정된 마스크는 $\hat{M}$, 정답 신호의 스펙트로그램의 크기를 Y라고 한다면, 아래와 같은 수식으로 목적함수를 나타낼 수 있다.

$$L = \| Y - X \odot \hat{M} \|_1, \tag{3}$$

여기서 $\odot$는 요소별 곱(element-wise multiplication)을 나타낸다. $\| . \|_1$은 L1 norm으로서, 행렬 각 요소의 절댓값의 합을 나타낸다. 잡음과 같은 이상치(outlier)가 많이 존재하지 않는 신호에서 실험을 하여서 이상치에 영향을 많이 받지 않는 목적함수를 사용하였다.

IV. 실 험

4.1 데이터베이스

배경음악 신호 분리 실험을 위하여 음악과 음성 데이터셋을 수집하였다.[16] 음악 데이터는 다양한 장르의 대중가요 3,646곡에서 각 곡당 12 s씩 임의로 선택하였다. 음성 데이터는 다양한 장르의 방송 콘텐츠 90 h에서 순수 음성만 존재하는 구간만 추출하였다. 추출된 음성을 12 s씩 구간을 나누어 총 3,646개의 약 12시간에 해당하는 음성 데이터를 만들었다.

Table 1. Details of the proposed architecture.

Layers	Frequency band		
	Low	High	Full
Conv ($f \times t$, ch)	3 × 4, 32	3 × 3, 32	3 × 4, 32
Dil. dense 1 (k, L)	16, 4	10, 3	6, 2
Compression 1 (θ)	0.13	0.20	0.16
Down-sampling 1, 2, 3	2 × 2 AP	2 × 2 AP	2 × 2 AP
Dil. dense 2,3,4 (k, L)	16, 4	10, 3	6, 2
Compression 2, 3, 4 (θ)	0.13	0.20	0.16
Up-sampling	2 × 2 TC	2 × 2 TC	2 × 2 TC
Concatenation	Comp. 3	Comp. 3	Comp. 3
Dil. dense 5 (k, L)	16, 4	10, 3	6, 2
Compression (θ)	0.13	0.20	0.16
Up-sampling	2 × 2 TC	2 × 2 TC	2 × 2 TC
Concatenation	Comp. 2	Comp. 2	Comp. 2
Dil. dense 6 (k, L)	16, 4	10, 3	6, 2
Compression (θ)	0.13	0.20	0.16
Up-sampling	2 × 2 TC	2 × 2 TC	2 × 2 TC
Concatenation	Comp. 1	Comp. 1	Comp. 1
Dil. dense 7 (k, L)	16, 4	10, 3	6, 2
Compression (θ)	0.13	0.20	0.16
Concatenation (axis)	Frequency		-
Concatenation (axis)	Channel (feature map)		
Dil. dense 8 (k, L)	4, 2		
Compression (θ)	0.25		
Conv ($f \times t$, ch)	2 × 1, 1		

3,646개의 음악, 음성 데이터들을 반으로 나누어 각 1,823개씩 학습 및 테스트 데이터로 사용하였다. 학습 데이터는 음성의 볼륨이 음악의 볼륨보다 큰 방송물의 특성을 적용하여 음악을 기준으로 −30 dB ~ 0 dB의 Signal to Noise Ratio(SNR)로 임의로 혼합하여 혼합 신호를 만든다. 테스트 데이터는 0 dB, −10 dB의 SNR로 혼합하여 각 SNR에 따른 분리 성능을 측정하였다.

4.2 실험환경

음향 신호는 16 kHz의 샘플링 주파수를 가지고, 모노 환경에서 실험하였다. 입력으로 사용하는 스펙트로그램은 1,024 윈도우 사이즈와 256 이동 사이즈를 사용하여 얻는다.

Table 1은 제안한 구조의 세부사항을 나타낸다. 합성곱의 $f \times t$는 커널 크기, ch는 합성곱의 출력 특징맵 개수를 말한다. k는 growth rate, L은 composite layer

Table 2. Experimental results of background music separation in broadcast contents.

SNR	0 dB			-10 dB		
Measure / Architecture	SDR	SIR	SAR	SDR	SIR	SAR
U-Net [5]	6.26	12.28	8.11	3.18	10.11	4.98
Wave-U-Net [9]	6.67	**15.26**	7.19	3.48	**14.53**	4.29
MDenseNet [6]	6.91	13.49	8.43	3.74	11.19	5.19
MMDenseNet [6]	7.19	13.71	8.72	3.99	11.33	5.49
FDMMDenseNet	7.27	14.10	8.70	**4.26**	12.04	5.61
TDMMDenseNet	7.28	13.66	8.86	4.12	11.15	**5.69**
2DMMDenseNet	**7.34**	13.44	**9.04**	4.13	11.44	5.61

개수, θ은 compression rate이다. 다운 샘플링은 Average Pooling(AP)을 사용하였고, 업 샘플링은 Transposed Convolution(TC)을 사용하였다.

Graphics Processing Unit(GPU)은 NVIDIA Titan RTX를 사용하였고, batch size는 15, epoch는 90으로 설정하였다. 6시간의 학습 데이터를 학습하는 데 대략 13시간 정도 소요되었다. 12 s의 wav 파일을 테스트하는 데 약 0.11 s 정도 시간이 소요되었다.

4.3 실험 결과

평가 지표로 분리된 배경음악 신호의 Signal to Distortion Ratio(SDR), Signal to Interference Ratio(SIR), Signal to Artifact Ratio(SAR) 3개의 지표를 측정하였다.[17] 보통 분리 성능을 비교할 때 SDR을 비교한다. SIR의 interference 오류와 SAR의 artifact 오류의 합을 SDR의 distortion 오류로 정의하기 때문에, SDR은 SIR과 SAR을 모두 고려한 분리 성능 결과이다.

Table 2는 실험 결과를 나타낸다. 실험은 U-Net, Wave-U-Net, MDenseNet, MMDenseNet과 제안한 구조의 성능을 비교하였다. Dilated block에 [2, 1], [1, 2], [2, 2] dilation rate의 dilated convolution을 적용하여 세 개의 제안구조 실험을 하였다. Table 2에는 위의 dilation rate 순서대로 FDMMDenseNet(Frequency DMMDenseNet), TDMMDenseNet(Time DMMDenseNet), 2DMMDenseNet([2,2] DMMDenseNet)로 표기하였다. U-Net, Wave-U-Net은 공개된 코드[1][8]를 사용하였고, MDenseNet, MMDenseNet 구조는 직접 구현하여 실험하였다.

U-Net은 실험한 구조 중에 가장 낮은 성능을 보이고, Wave-U-Net은 U-Net과 비교하여 혼합 SNR 0 dB에서 0.41 dB, 혼합 SNR −10 dB에서 0.30 dB SDR을 개선하였다. 특히 SIR은 Wave-U-Net이 가장 높은 성능을 보인다. SIR은 분리된 배경음악에 음성이 얼마나 섞여 있는지를 나타낸다. Wave-U-Net을 제외한 다른 구조들은 스펙트로그램 도메인에서 배경음악의 스펙트로그램 크기를 추정하고 혼합 신호의 위상 정보를 이용하여 신호로 복원한다. 혼합 신호의 위상 정보를 사용함으로써 SIR이 낮아지는 원인이 된다. 이러한 오류를 피하고자 Wave-U-Net은 시간 도메인에서 신호를 바로 추정한다. 위와 같은 이유로 SIR은 Wave-U-Net이 가장 높다. 하지만 SDR에서는 Wave-U-Net보다 MDenseNet, MMDenseNet과 제안한 구조가 더 높은 성능을 보인다.

MDenseNet은 Wave-U-Net보다 혼합 SNR 0dB에서 0.24 dB, 혼합 SNR −10 dB에서 0.26 dB SDR을 개선하였다. MMDenseNet은 MDenseNet보다 혼합 SNR 0 dB에서 0.28 dB, 혼합 SNR −10 dB에서 0.25 dB SDR을 개선하였다. 혼합 SNR 0 dB에서 제안 구조인 2DMM DenseNet은 MMDenseNet보다 0.15 dB SDR을 개선하여 제안한 구조 중 가장 높은 성능을 보였다. 혼합 SNR −10 dB에서 제안 구조인 FDMMDenseNet은 MM DenseNet보다 0.27 dB SDR을 개선하여 제안한 구조 중 가장 높은 성능을 보였다.

V. 결 론

본 연구에서는 방송 콘텐츠의 배경음악을 분리하기 위한 딥러닝 구조를 제안하였다. 딥러닝 구조의 수용범위를 효과적으로 늘이기 위해서 dilated convolution을 추가하였다. 분리한 배경음악에 음성이 가장 적게 남아 있는 구조는 Wave-U-Net이였다. 스펙트로그램 도메인에서는 혼합 신호의 위상 정보를 이용할 수밖에 없는 시스템 구조적 한계가 있다. 이러한 단점에도 불구하고 SNR 0 dB 테스트 환경에서 2DMMDenseNet이, SNR −10 dB 테스트 환경에서 FD MMDenseNet이 가장 높은 성능을 보였다.

1) github.com/Xiao-Ming/UNet-VocalSeparation-Chainer

감사의 글

본 연구는 문화체육관광부 및 한국저작권위원회의 2019 년도 저작권기술개발사업의 연구결과로 수행되었다[2018-micro-9500, 음악 및 동영상 모니터링을 위한 지능형 마이크로 식별 기술개발].

References

1. D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” Proc. NIPS, 556-562 (2001).
2. J. Le Roux, J. Hershey, and F. Weninger, “Deep NMF for speech separation,” Proc. IEEE Int. Conf. Acoust., Speech Signal Process, 66-70 (2015).
3. A. A. Nugraha, A. Liutkus, and E. Vincent, “Multi-channel music separation with deep neural networks,” Proc. EUSIPCO. 1748-1752 (2015).
4. A. Jansson, E. Humphrey, N. Montecchio, R. Bittner, A. Kumar, and T. Weyde, “Singing voice separation with deep U-Net convolutional Networks,” Proc. ISMIR, 323-332 (2017).
5. N. Takahashi and Y. Mitsufuji, “Multi-scale multi-band DenseNets for audio source separation,” Proc. WASPAA. 261-265 (2017).
6. O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention, 234-241 (2015).
7. G. Huang, Z. Liu, K. Q. Weinberger, and L. Maaten, “Densely connected convolutional networks,” Proc. CVPR. 4700-4708 (2017).
8. D. Stoller, S. Ewert, and S. Dixon, “Wave-u-net: A multi-scale neural network for end-to-end audio source separation,” Proc. ISMIR. (2018).
9. D. Ward, R. D. Mason, R. C. Kim, F.-R. Stöter, A. Liutkus, and M. D. Plumbley, “SISEC 2018: State of the art in musical audio source separation-subjective selection of the best algorithm,” Proc. 4th Workshop on Intelligent Music Production, (2018).
10. N. Takahashi, P. Agrawal, N. Goswami, and Y. Mitsufuji, “PhaseNet: Discretized phase modeling with deep neural networks for audio source separation,” Proc. Interspeech, 2713-2717 (2018).
11. N. Takahashi, N. Goswami, and Y. Mitsufuji, “MM DenseLSTM: An efficient combination of convolutional and recurrent neural networks for audio source separation,” Proc. IWAENC. 106-110 (2018).
12. F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” Proc. Int. Conf. Learn. Representations, (2016).
13. S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” Proc. ICML. 448-456 (2015).
14. X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural networks,” Proc. AISTATS. 315-323 (2011).
15. V. Dumoulin and F. Visin, “A guide to convolution arithmetic for deep learning,” arXiv preprint arXiv: 1603.07285 (2016).
16. H. Kim, J. Kim, and J. Park, “Music-speech separation based background music identification in TV programs” (in Korean), Proc. HCI KOREA, 1158-1161 (2019).
17. A. Liutkus, F. Stöter, Z. Rafii, D. Kitamura, B. Rivet, N. Ito, N. Ono, and J. Fontecave, “The 2016 Signal separation evaluation campaign,” Proc. LVA/ICA. 66-70 (2017).

저자 약력

▸ 허 운 행 (Woon-Haeng Heo)

2015년 2월 : 충북대학교 전자공학부 학사
2017년 2월 : 충북대학교 제어로봇공학전공 석사
2017년 3월 ~ 현재 : 충북대학교 제어로봇공학전공 박사 과정

▸ 김 혜 미 (Hyemi Kim)

2004년 2월 : 부산대학교 전자전기컴퓨터공학부 학사
2006년 2월 : 한국과학기술원 전기 및 전자공학부 석사
2006년 3월 ~ 현재 : 한국전자통신연구원 선임연구원

▸ 권 오 욱 (Oh-Wook Kwon)

1986년 2월 : 서울대학교 전자공학과 학사
1988년 2월 : 한국과학기술원 전기및전자공학과 석사
1997년 2월 : 한국과학기술원 전기및전자공학과 박사
1986년 3월 ~ 2000년 4월 : 한국전자통신연구원 책임연구원
2000년 5월 ~ 2001년 3월 : 한국과학기술원 연구교수
2001년 3월 ~ 2003년 8월 : UCSD 박사후 연구원
2003년 9월 ~ 현재 : 충북대학교 전자공학부 교수