

Short-time Fourier transform 소음맵을 이용한 컨볼루션 기반 BSR (Buzz, Squeak, Rattle) 소음 분류

BSR (Buzz, Squeak, Rattle) noise classification based on convolutional neural network with short-time Fourier transform noise-map

부석준,¹ 문세민,² 조성배[†]

(Seok-Jun Bu,¹ Se-Min Moon,² and Sung-Bae Cho^{1†})

¹연세대학교 컴퓨터과학과, ²현대자동차

(Received June 1, 2018; accepted July 27, 2018)

초 록: 차량 내부에는 BSR(Buzz, Squeak, Rattle) 세 가지 유형의 소음이 발생한다. 본 논문에서는 심층 컨볼루션 신경망으로 추출한 소음 특징에 기반하여 자동으로 차량 내부의 BSR 소음을 분류하는 분류기를 제안한다. 차량 내부의 소음은 전처리 단계에서 STFT(Short-time Fourier Transform) 알고리즘을 사용하여 소음 맵으로 표현된다. 생성된 소음 맵 내부에서 실제 소음의 위치를 정확하게 파악하기 어려운 문제에 대처하기 위해서 슬라이딩 윈도우 방법으로 분할하였다. 본 논문에서는 t-SNE(t-Stochastic Neighbor Embedding) 알고리즘을 사용하여 심층 컨볼루션 신경망 내부 파라미터를 시각화하고 정성적인 방식으로 오분류데이터를 분석하였다. 분류된 데이터의 정량적인 분석을 위해 소음의 종류별 유사도를 SSIM(Structural Similarity Index) 수치에 기반하여 정량화하여 리트랙터의 떨림음이 정상주행음과 가장 유사하다는 것을 밝혔다. 제안하는 방법의 분류기는 기타 기계학습 알고리즘 대비 최고 분류 정확도를 달성하였다(99.15%).
핵심용어: 차량 소음, BSR (Buzz, Squeak, Rattle) 분류, 딥 러닝, 컨볼루션 신경망

ABSTRACT: There are three types of noise generated inside the vehicle: BSR (Buzz, Squeak, Rattle). In this paper, we propose a classifier that automatically classifies automotive BSR noise by using features extracted from deep convolutional neural networks. In the preprocessing process, the features of above three noises are represented as noise-map using STFT (Short-time Fourier Transform) algorithm. In order to cope with the problem that the position of the actual noise is unknown in the part of the generated noise map, the noise map is divided using the sliding window method. In this paper, internal parameter of the deep convolutional neural networks is visualized using the t-SNE (t-Stochastic Neighbor Embedding) algorithm, and the misclassified data is analyzed in a qualitative way. In order to analyze the classified data, the similarity of the noise type was quantified by SSIM (Structural Similarity Index) value, and it was found that the retractor tremble sound is most similar to the normal travel sound. The classifier of the proposed method compared with other classifiers of machine learning method recorded the highest classification accuracy (99.15 %).

Keywords: Automotive noise, BSR (Buzz, Squeak, Rattle) classification, Deep learning, Convolutional neural network

PACS numbers: 43.60.Bf, 43.60.Lq

1. 서 론

차량 내부에서 발생하는 소음은 고객이 차량의 수

준을 평가하는 매우 중요한 요소로 업계에서 중요한 평가 항목으로 사용되고 있다. 차량 내 소음으로 인한 수리비는 전체 보증 수리비의 50% 이상을 차지하기 때문에 소음을 분석하고 분류할 수 있는 기술이 업계에서 주된 차세대 기술로서 조명받고 있다.

기존의 소음 분석결과 차량에서 발생하는 소음의

[†]Corresponding author: Sung-Bae Cho (sbcho@yonsei.ac.kr)
Department of Computer Science, Yonsei University, 50 Yonsei-ro
Seodaemun-gu, Seoul, 03722, Republic of Korea
(Tel: 82-2-2123-2720, Fax: 82-2-365-2579)

종류를 크게 3가지로 분류하여 각각 Buzz(부품의 판넬이 자체적으로 복처럼 내는 소음), Squeak(부품간 전단방향으로 마찰에 의해 발생하는 소음), Rattle(부품간 수직방향으로 부딪혀 발생하는 소음)으로 정의하였다.^[1]

BSR 소음을 자동적으로 분류하는 것은 아직 많은 한계점을 가지고 있다. 우선 다른 소리를 배제한 정확한 소음 데이터를 확보하기 어렵고, 차량의 종류가 다르면 소음의 종류가 같아도 그 주파수 대역이 달라진다. 소음 자체가 음향 데이터로 매우 높은 차원을 가지는 것과 소음의 정확한 시작과 끝을 알기 어려운 점 또한 BSR 소음의 자동적인 분류를 어렵게 만드는 원인이다.

구체적으로 BSR 소음은 시계열 특성을 가지는 음향 데이터로 기계학습 방법을 사용하여 분류하기에는 데이터의 차원이 높고, 정확한 소음의 시작과 끝을 알기 어려운 문제가 있다. 따라서 본 논문에서는 소음의 분류에 사용할 특징을 추출하기 위해 STFT(Short-time Fourier Transform) 알고리즘을 사용하여 소음별로 고유한 소음 맵을 구성하였다. Fig. 1에서 BSR 소음을 소음 맵을 변환하는 과정을 도식화하였다.

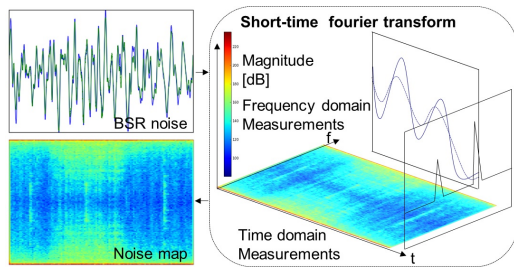


Fig. 1. BSR noise map generated using STFT algorithm.

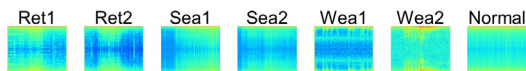


Fig. 2. Comparison of noise map by type of noise.

Table 1. Noise dataset specification.

Class	Number of instances	Average length	Number of windows
Normal	1	28.58s	4,914
Retractor	12	0.73s	1,402
Seat rattle	16	1.72s	5,017
Weatherstrip	59	0.67s	6,289

분류한 소음의 종류는 네가지로, Table 1에 소음의 종류와 개수, 평균 음원 길이를 정리하였다. 시간도메인 데이터의 특징과 주파수도메인 데이터의 특징을 동시에 고려하기 위해 데이터의 구간별로 Fourier transform을 적용하였고, Fig. 2에서 소음의 종류별 소음 맵을 비교하였다. 각 소음 맵은 육안으로 보았을 때 식별이 불가능할 정도로 유사하다.

따라서 본 논문에서는 복잡하고 유사한 소음 데이터를 모델링하고 분류하기 위해서 기계학습 방법 중 심층 컨볼루션 신경망을 사용하는 BSR 분류를 제안한다. 컨볼루션 신경망은 입력된 데이터에 컨볼루션과 풀링연산을 적용하며 자동적으로 분류에 필요한 특징을 추출, 학습하는 방법으로 타 기계학습 방법을 사용한 BSR 분류기와 비교하여 가장 높은 분류정확도(99.20%)를 기록하였다.

II. 관련 연구

BSR 소음을 분석하고 분류하기 위해 여러 시도가 있었다. 차량 소음을 모델링하고 정의하려는 연구에서는 차량의 소음으로부터 특징을 추출하고 분석하였다.^[2] 소음원의 추적시스템을 개발하려는 연구에서는 BSR 소음의 유형을 체계적으로 분류하고 정량화하였다.^[1] BSR 소음을 응용하려는 목적으로 주행시 발생하는 여러 외부요인을 고려하여 차체 강건설계를 목적으로 한 연구에서는 BSR 측정과 분석을 위한 절차를 제안하였다.^[3] Table 2에서 이러한 연구들을 모델 및 재현 단계, 측정 및 예측 단계, 분석 단계로 정리하였다.

BSR 소음을 기계학습 방법에 기반하여 자동적으

Table 2. Noise dataset specification.

Domain	Author	Description
Modeling	Wu <i>et al.</i> ^[2]	Automotive noise modeling, recognition
	Mog <i>et al.</i> ^[1]	BSR categorization, quantification
Prediction	Choi <i>et al.</i> ^[4]	Noise prediction at the drawing stage
	Woo and Park ^[5]	Measurement procedure proposal
Analysis	Cerrato ^[6]	Quantitative analysis by noise type
	Wang <i>et al.</i> ^[7]	BSR classification
	Lee <i>et al.</i> ^[8]	BSR classification

로 특징을 추출하고 분류하려는 시도로 인공지능망을 사용한 사례가 있다.^[7] 차원이 매우 높은 BSR 소음의 경우 불필요한 정보가 모델링에 사용되어 성능을 하락시키는 문제가 있으나, 최근 신경망을 이용한 딥러닝 학습방법들이 고해상도 이미지 분야에서 불필요한 정보를 포함한 고차원 입력의 특징을 성공적으로 추출해내고 있다.^[9]

III. 방 법

3.1 분류기 아키텍처 및 BSR 소음 맵 전처리

Fig. 3에서 BSR 소음 분류기의 아키텍처를 도식화하였다. 컨볼루션 신경망을 사용하는 분류기 모델은 이미지 분류에서 주로 사용되며 특히 2012년 이후로 각종 이미지 분류 경연대회에서 최고의 성능을 거두고 있다.^[10] 하단부터 2회의 컨볼루션-풀링 층으로 구성되어 있고, 각 컨볼루션, 풀링층의 필터 크기는 2×2 벡터이다. 일반적인 이미지 분류를 위한 컨볼루션 신경망이 하단에서 점, 선 등의 저수준 특징을 추출하기 위해 필터 크기를 크게 하는 것에 비해 BSR 소음 맵은 구조가 복잡하지 않아 필터의 크기를 작게 설정하였다. 분류기의 상단에서는 학습데이터에 대한 과적합 문제에 대처하기 위해서 이전 층 출력의 일부를 버리는 Dropout 층을 사용하였다.^[11]

각 BSR 소음으로부터 STFT 알고리즘을 사용하여 추출한 소음 맵의 일부 중 실제 소음의 시작과 끝을

알 수 없는 문제에 대처하기 위해서 슬라이딩 윈도우 방법을 사용하여 일정 구간을 샘플링하였다. 슬라이딩 윈도우 방법은 정해진 크기의 윈도우를 시간 축을 따라 이동시키며 데이터의 일정 구간을 샘플링하는 방법이다.^[12] 본 논문에서 사용한 윈도우는 STFT 알고리즘에 의해 픽셀공간으로 사상된 소음맵의 일부 열(시간)과 전체 행(주파수)으로 정의한다.

녹음된 음원의 초당 샘플수는 44,100 프레임으로, 윈도우의 크기는 추출한 프레임의 수로 정의한다. 윈도우의 크기에 따라 학습 데이터의 수가 결정되며, 윈도우에 소음이 포함되었는지의 여부에 따라 분류기의 성능을 변동시키기 때문에 최적의 윈도우 크기를 찾는 것이 중요하다. 실험적 결과에 의해 최적의 윈도우 크기로 도출된 크기 9 프레임(약 0.0002 s)를 사용하였고, 이때 생성된 윈도우 집합의 분포를 Fig. 4에서 t-SNE(t-Stochastic Neighbor Embedding) 알고리즘을 사용하여 시각화하였다.

t-SNE 알고리즘은 고차원상 데이터의 관계를 저차원상에서 유지하는 비선형적인 차원 축소 방법이다.^[13] 소음원이 다름에도 불구하고 가까운 거리에서 뒤섞인 데이터들은 소음의 특성이 비슷하여 오분류될 확률이 높다. 하단에 정상주행음과 리트랙터 떨림음이 하단에 뒤섞여 있는 것으로 보아 두 소음으로부터 생성해낸 윈도우가 서로 비슷한 특성을 가지고 있어 분류하기 어렵다. 반면에 웨더스트립 비범음과 시트 래틀음은 우측 하단과 좌측 상단에 각각 군집화되어 소음의 특성이 다른 소음과 명확하게 구별됨을 알 수 있다.

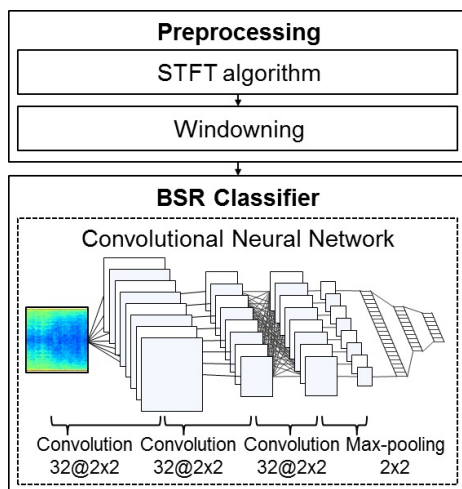


Fig. 3. CNN based BSR noise classifier architecture.

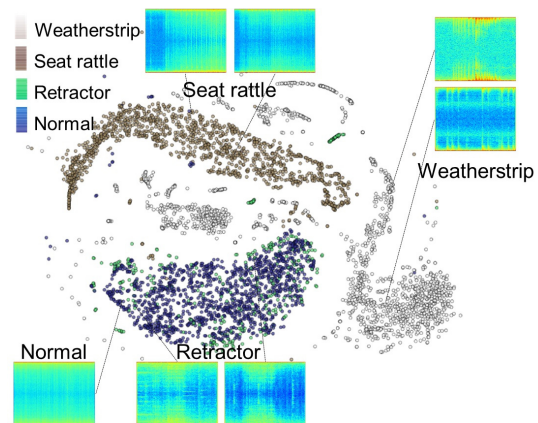


Fig. 4. BSR noise data distribution at window size 9.

3.2 심층 컨볼루션 신경망

BSR 소음을 벡터화하는 경우 차원이 높아 기존 기계학습에 기반한 분류기가 모델링 하기에 적절하지 못한 어려움이 있으나, 자동으로 분류에 중요한 특징을 추출하는 심층 컨볼루션 신경망을 사용해서 해당 어려움에 대처할 수 있다.

컨볼루션 신경망(Convolutional Neural Networks, CNN)은 컨볼루션, 풀링 연산을 수행하여 이미지로부터 퇴적의 특징을 추출해내고 학습하는 방법이다.^[14] l 번째 컨볼루션층에서 출력되는 벡터의 원소 c^l 는 이전 층의 출력벡터 y^{l-1} 와 $m \times m$ 크기의 벡터인 필터 w 에 의해 Eq. (1)의 컨볼루션 연산을 수행한다. 필터는 이미지보다 작은 크기의 벡터로 BSR 소음 벡터와 곱해져서 일부를 강조하거나 왜곡하는 역할을 한다.

$$c_{xy}^l = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} w_{ab} y_{(x+a)(y+b)}^{l-1}. \quad (1)$$

l 번째 풀링층에서는 이전 층에서 입력된 $N \times N$ 크기의 벡터 내 $k \times k$ 영역으로부터 하나의 최대값을 대표값으로 고르는 Eq. (2)의 최대값-풀링 연산을 수행한다. τ 는 풀링 거리로 풀링 연산을 수행할 영역의 거리이다.

$$p_{xy}^l = \max_{r \in R} c_{xy \times \tau}^{l-1}. \quad (2)$$

컨볼루션, 풀링 연산을 반복하여 BSR 소음 벡터가 분해되고 분류기 하단의 l 번째 완전-연결층에 도달하게 되면 Eq. (3)에 의해 각 노드의 값이 결정된다. W_{ji} 는 해당 층의 i 번째 노드와 j 번째 노드 사이에 곱해지는 가중치값이고, 마지막 풀링층에서 출력한 벡터 P 는 비선형 시그모이드 함수 σ 에 의해 출력된 뒤 바이어스 b 와 더해지고 가중치와 곱해진다.

$$h_j^l = \sum_i w_{ji}^{l-1} (\sigma(p_i^{l-1}) + b_i). \quad (3)$$

마지막으로 완전-연결층 중 최하단 L 층의 소프트맥스 층에서는 실제 BSR 소음의 종류 c 와 소음의 개수 N_c 에 대해 Eq. (4)에 의해 BSR 소음의 종류를 출력한다.

$$P(c|p) = \underset{c \in C}{\operatorname{argmax}} \frac{\exp(p^{L-1} w^L + b^L)}{\sum_{k=1}^{N_c} \exp(p^{L-1} w_k)}. \quad (4)$$

이어서 역전파 알고리즘에 의해 분류되어야 하는 소음의 종류와 BSR 소음 맵 벡터가 사상되도록 분류기 각 층의 가중치 w 를 수정하며 학습과정이 진행된다.^[15]

IV. 실험 결과

4.1 오분류 데이터 분석

Table 3에서 최적의 윈도우 크기를 결정하기 위해 다른 변수를 고정하고 윈도우 크기 별 분류 정확도를 정리하였다. BSR 소음 윈도우 크기 6에서 87.31%로 분류 정확도가 가장 낮았고 크기 9에서 99.20%로 거의 모든 BSR 소음 윈도우를 분류하였다. Table 4에서 윈도우 크기 9로 설정한 경우의 분류결과와 혼동행렬분석을 정리하였다.

t-SNE 알고리즘으로 시각화한 분포와 마찬가지로 주로 정상주행음과 리트랙터 떨림음이 오분류되었

Table 3. Classification accuracy by window size.

Window size	6	7	8	9	10
Accuracy	0.8731	0.9455	0.9265	0.9920	0.9657

Table 4. confusion matrix analysis of CNN based BSR classifier.

Confusion matrix		Predicted			
		Nor	Ret	Sea	Wea
Actual	Nor	469	3	3	0
	Ret	6	132	0	0
	Sea	1	0	529	2
	Wea	0	0	0	618

Table 5. SSIM value by noise type.

Class	Misclassified	Similarity [SSIM]
Nor-Ret	9	0.0838
Nor-Sea	4	0.0403
Nor-Wea	0	0.0012
Ret-Sea	0	0.0262
Ret-Wea	0	0.0086
Sea-Wea	2	0.0307

다. 전체 오분류 데이터 중 리트랙터 펼침음을 정상 주행음으로 오분류한 경우가 6건으로 가장 많았고, 정상주행음을 리트랙터 펼침음으로 분류한 경우가 3건으로 정상 주행음과 리트랙터 펼침음의 오분류가 총 9건 발생하였다. 반면에 웨더스트립 비빔음은 오분류되는 경우가 전혀 없었다.

주로 오분류되는 정상주행음과 리트랙터 펼침음의 정량적인 비교를 위해 SSIM(Structural Similarity Index)수치를 가지고 비교분석하였다. SSIM 수치는 두 이미지 사이의 유사성을 측정하는 척도로, 이미지의 구조적인 속성에 기반하여 MSE(Mean Squared Error), PSNR(Peak Signal-to-Noise Ratio) 척도대비 정확한 비교수치를 제공한다.^[16] Table 5에서 각 소음의 종류별로 오분류 수와 유사도를 정리하였다. 정상 주행음과 리트랙터 펼침음의 유사도가 0.0838으로 다른 소음 종류 사이의 유사도보다 최대 80배가량 높다.

4.2 BSR 분류기 분석

Fig. 5에서 제안하는 분류기로부터 출력된 데이터 특징의 분포를 t-SNE 알고리즘을 사용하여 시각화하였다. 데이터가 인공신경망에 입력되면 심층으로 갈 수록 각 층의 가중치와 곱해지고 비선형함수를

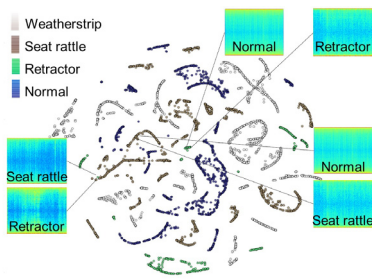


Fig. 5. Visualizing distribution of data represented within CNN.

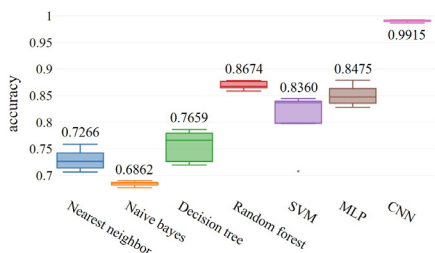


Fig. 6. 10-fold CV with other machine learning methods.

통해 출력되는데, 소음의 종류를 출력하는 출력 직전 층의 활성화함수값을 t-SNE 알고리즘을 사용하여 시각화하였다. 분류기의 심층에서 종류별로 군집화하는 것으로 분류기의 학습 정도를 정성적으로 분석할 수 있다. 분류기에 입력되기 이전의 데이터 분포에 비해 깔끔하게 군집화하였고, 중단과 좌측에서 서로 유사도가 높아 오분류된 데이터를 확인할 수 있다.

제안하는 컨볼루션 신경망을 사용한 BSR 분류기와 기존의 기계학습 기반 분류기의 성능을 비교하기 위해서 10겹 교차검증(10 fold Cross validation, CV)하였다. Fig. 6에서 기존의 기계학습 기반 BSR 분류기와 제안하는 컨볼루션 신경망을 사용한 BSR 분류기의 비교실험 결과를 Box-plot으로 정리하였다.

기존의 기계학습에 기반한 분류기는 Naive Bayes, Nearest Neighbor, Decision Tree, SVM, MLP, Random Forest 순으로 분류 정확도가 높았다. 기존 방법은 소음 맵의 복잡하고 높은 차원으로 인해 소음의 특징을 완전히 학습하지 못하였으나, 제안하는 심층 컨볼루션 신경망 방법의 분류기는 10겹 교차검증 결과 평균 99.15% 테스트 분류정확도를 기록하였다.

V. 결론

본 논문에서는 BSR 소음데이터로부터 STFT 알고리즘을 사용하여 특징을 추출하고 슬라이딩 윈도우 방법을 사용하여 전처리하였다. 각 BSR 소음은 고유한 소음 맵으로 사상되어 심층 컨볼루션 신경망을 사용하여 분류하였다.

- 1) BSR 소음맵과 해당 소음의 종류를 사상하도록 훈련시킨 심층 컨볼루션 신경망은 BSR 소음의 높은 차원에 대처하였고, 10겹 교차검증 결과 기존의 기계학습에 기반한 분류기 중 최고의 성능 (99.15% 테스트 정확도)을 기록하였다.
- 2) 벡터의 유사성을 비교하는 SSIM 수치로 오분류된 데이터를 정량적으로 분석하여, 정상 주행음과 리트랙터 펼침음의 유사도가 다른 소음 종류 사이의 유사도보다 최대 80배가량 높다는 것을 보였다. 추가적인 연구로 차량의 종류와 외부의 요인에 대해 대처할 수 있도록 소음 데이터의 개수를 늘리고 잡음-강건성을 확보하는 것이 필요하다. 특히 정상

주행음과 리트랙터 펄림음의 명확한 구분을 위해서 추가적인 해당 데이터의 샘플을 확보하고 특징에 대해 분석하는 것이 필요하다.

References

1. M. G. Mog, B. H. Min, S. W. Choi, and H. J. Lee, "Development of the reproduction test method of automobile buzz, squeak, rattle noise and the tracking system" (in Korean), Proc. KSAE 2010 Conference, 1475-1481 (2010).
2. H. Wu, M. Siegel, and P. Khosla, "Vehicle sound signature recognition by frequency vector principal component analysis," IMTC, IEEE, **1**, 429-434 (1998).
3. K. C. Kim, S. W. Lee, J. M. Choi, and Y. J. Kim, "Study on the simulation process for the buzz, squeak and rattle incorporating vehicle to vehicle variance" (in Korean), Proc. KSNVE 2014 Conference, 480-481 (2014).
4. S. U. Choi, M. G. Mog, S. J. Kang, K. D. Yee, and M. W. Suh, "A study on the BSR performance prediction technology" (in Korean), Proc. KSAE 2013 Conference, 2040-2044 (2013).
5. C. Woo and H. Park, "Measurement of BSR noise of cockpit module materials for automobiles," Proc. AIP Conference, 1727 (2016).
6. G. Cerrato, "Automotive sound quality-accessories, BSR and brakes," Sound and Vibration, **43**, 10, (2009).
7. Y. S. Wang, C. M. Lee, D. G. Kim, and Y. Xu, "Sound-quality prediction for nonstationary vehicle interior noise based on wavelet pre-processing neural network model," J. Sound and Vibration, **299**, 933-947 (2007).
8. S. Lee, Y. Kwak, B. Kim, J. Lee, and J. Park, "Parameter quantification for evaluation of vehicle's impulsive bsr noise," Proc. Inter-noise Conference, 3361-3363 (2014).
9. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, 1097-1105 (2012).
10. M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," Proc. European Conference on Computer Vision, 818-833 (2014).
11. N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: simple way to prevent neural networks from overfitting," J. Machine Learning Research, **15**, 1929-1958 (2014).
12. T. G. Dietterich, "Machine learning for sequential data: a review," Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition and Structural and Syntactic Pattern Recognition, 15-30 (2002).
13. L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE," J. Machine Learning Research, **9**, 2579-2605 (2008).
14. T. N. Sainath, A. R. Mohamed, B. Kingbury, and B. Ramabhadran, "Deep convolutional neural networks for LVCSR," Proc. ICASSP, 8614-8618 (2013).
15. S. J. Bu and S. B. Cho, "A hybrid system of deep learning and learning classifier system for database intrusion detection," Proc. HAIS, 615-625 (2017).
16. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. on Image Processing, **13**, 600-612 (2004).

저자 약력

▶ 부 석 준 (Seok-Jun Bu)



2016년 2월: 한양대학교 ERICA 컴퓨터공학과 학사
2016년 3월 ~ 현재: 연세대학교 컴퓨터과 학과 석박통합과정

▶ 문 세 민 (Se-Min Moon)



2008년: 서강대학교 기계공학과 학사
2008년 ~ 현재: 현대자동차 책임연구원

▶ 조 성 배 (Sung-Bae Cho)



1988년: 연세대학교 전산학과 학사
1990년: KAIST 전산학과 석사
1993년: KAIST 전산학과 박사
1993년: 일본 ATR 인간정보통신연구소 객원 연구원
1998년: 호주 Univ. of New South Wales 초청 연구원
1995년 ~ 현재: 연세대학교 컴퓨터과 학과 정교수