

시간 영역 오토인코더의 성능 개선을 위한 다중 대역 손실 함수

*임유진 유정찬 서은미 박호종

광운대학교

*dbwls553@naver.com

A Multi-band Loss Function for Improving Time-Domain Autoencoder

*Lim, Yujin Yu, Jeongchan Seo, Eunmi Park, Hochong

Kwangwoon University

요약

본 논문에서는 시간 영역 오토인코더의 성능 개선을 위한 다중 대역 손실 함수를 제안한다. 기존의 시간 영역 오토인코더를 사용하는 압축 및 복원 모델은 저 대역 손실에 치중되어 고 대역 신호를 생성하지 못하고 다운 샘플링된 신호를 결과로 출력하는 문제점을 가진다. 이를 해결하기 위해 대역별로 손실을 분리하여 가중치를 조절할 수 있는 다중 대역 손실 함수를 제안한다. 제안하는 손실 함수가 적용된 오토인코더에 음성 신호를 입력하여 학습을 진행한 결과, 다운 샘플링이 발생하지 않으며 고 대역 신호가 복원되는 것을 스펙트로그램을 통해 확인하였다.

1. 서론

머신러닝을 이용한 신호의 압축 및 복원에 대한 연구가 활발히 진행되고 있다. 그 중 가장 대표적인 방법은 오토인코더 구조를 이용하는 것이다[1]. 최근에는 위상을 복원할 수 있는 시간 영역 오토인코더가 많이 연구되고 있는데, 이때 손실 함수로 mean absolute error (MAE) 또는 mean square error (MSE)를 주로 사용한다[2]. 이러한 손실 함수는 샘플간의 차이를 최소화 하는 방식으로 직관적이고 사용이 편리하지만, 큰 차이를 우선하여 미세한 차이를 무시할 수 있다는 문제를 가진다. 음성, 음악, 사운드 등과 같은 신호는 주로 고 대역보다 저 대역이 큰 에너지를 가지는데, 이러한 손실 함수를 사용할 경우 저 대역 신호가 우선되어 고 대역 신호가 무시될 수 있다[3]. 특히, 병목 (bottleneck) 을 가진 구조에서는 정보량을 줄이면서 손실 함수를 최소화하기 위해 입력 신호를 다운 샘플링 (down sampling)하여 저 대역 신호만 남기는 방식으로 쉽게 학습된다. 경험적으로, 다운 샘플링으로 학습 방향이 결정된 후에는 고 대역 신호를 생성하지 못하고 다운 샘플링 된 신호만을 출력하는 것을 확인하였다. 이는 모델의 심각한 성능 저하를 야기한다.

제안하는 방법은 오토인코더 구조에 다중 대역 손실 함수를 적용함으로써 이 문제를 해결하고자 한다. 대역을 분리하여 상대적으로 작은 에너지를 가지는 대역의 손실 값에 적절한 가중치를 부여하여 모든 대역이 동일한 손실 규모를 가지도록 한다. 이를 통해 특정 대역이 무시되지 않도록 하여 다운 샘플링 되는 것을 방지하였다.

다양한 신호에 대해 광범위하게 적용할 수 있으나, 본 논문에서는 음성신호로 국한하여 성능을 평가하였다. 그 결과, 동일한 오토인코더 구조에 손실 함수만 변경한 것임에도 불구하고 다운 샘플링이 발생하지 않으며 고 대역 신호가 복원되는 것을 스펙트로그램을 통해 확인하였다.

2. 제안하는 방법

2.1 다중 대역 손실 함수

제안하는 방법은 오토인코더의 출력을 대역별로 분리하여 손실을 계산하고, 적절한 가중치를 곱하여 모든 대역이 같은 손실 규모를 가지도록 하는 것이다.

일반적으로 대역을 분리하기 위해서는 푸리에 변환을 통한 대역 통과 필터링이 필요하다. 하지만 푸리에 변환을 통해 주파수 영역으로 이동할 경우 위상에 대한 정보를 학습시키기 어렵다. 제안하는 방법은 시간 영역에서 대역 통과 필터를 설계하고 컨볼루션 (convolution)을 통해 필터링함으로써, 대역을 분리함과 동시에 위상정보를 쉽게 학습할 수 있도록 하였다. 이 과정은 그림 1과 같이 K 개의 시간 영역 대역 통과 필터를 가지는 K 채널 컨볼루션으로 쉽게 나타낼 수 있다.

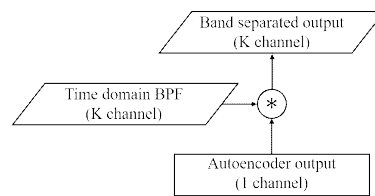


그림 1. 다중 대역 통과 필터의 적용 방법

Fig. 1. Structure of applying multiple band-pass filter

대역별로 분리된 신호의 손실 값을 MAE 또는 MSE와 같은 손실 함수로 계산하고 적절한 가중치를 곱함으로써 손실 규모가 조정된 대역별 손실 값을 구할 수 있다. 규모가 조정된 대역별 손실 값을 합하여 최종 손실 함수를 구할 수 있다. 이 과정은 식 (1)과 같이 나타낼 수 있으며 제안하는 손실 함수이다. 여기서, x 는 입력 신호, y 는 목표 신호, K 는

대역의 개수, $H_k(\cdot)$ 는 k 번째 대역 통과 필터와의 컨볼루션 연산, w_k 는 k 번째 대역의 가중치, L 은 각 대역에 적용할 손실 함수, $AE(\cdot)$ 는 오토인코더의 출력이다.

$$\sum_{k=0}^{K-1} \{L(H_k(y), H_k(AE(x))) \cdot w_k\} \quad (1)$$

2.2 대역별 가중치 설정

모든 대역이 동일한 손실 규모를 가지게 하기 위해서는 대역별 손실 값이 같아지도록 가중치를 계산하여야 한다. 하지만 실제 손실 값은 미리 계산할 수 없다. 따라서 손실의 크기가 각 대역별 신호 크기에 동일한 비율로 비례한다고 가정하여 손실 규모를 계산한다.

예를 들어 대역이 2개이고 대역별 손실 함수로 MAE를 사용 할 경우 각 대역의 손실 값은 식 (2)와 식 (3)과 같이 나타낼 수 있다. 이때 α 는 신호 크기와 손실의 비율이다. 이때 모든 학습 프레임에 대해 L 을 구하고 평균을 취하여 사용한다. 최종적으로 손실 규모가 같아지도록 각각의 대역에 곱해져야 하는 가중치는 식 (4)을 통해 구할 수 있다.

$$L_1 = \sum |y_1 - \hat{y}_1| = \alpha \sum |y_1| \quad (2)$$

$$L_2 = \sum |y_2 - \hat{y}_2| = \alpha \sum |y_2| \quad (3)$$

$$w_1 = \frac{L_1}{L_2} = 1, \quad w_2 = \frac{L_2}{L_1} = \frac{\sum |y_1|}{\sum |y_2|} \quad (4)$$

3. 성능 평가

학습과 성능평가를 위한 신호는 16 kHz로 샘플링 된 화자 1 명의 한국어 음성신호이다. 데이터의 크기는 총 18분이며 학습에 15분, 성능평가에 3분을 사용하였다.

오토인코더는 시간 영역에서 동작하는 convolutional neural network (CNN)을 사용하였으며, 입력과 출력의 크기는 1024 샘플, 병목 지점의 크기는 입력의 1/4 인 256 샘플로 하였다. Adam을 사용하여 학습을 진행하였고, 손실 함수로는 MAE만 사용하였을 때와 대역별 손실 함수를 적용하였을 때를 비교하였다.

그림 2는 원본신호와 기존의 손실 함수를 사용하였을 때, 다중 대역 손실함수를 사용하였을 때의 스펙트로그램이다. 기존의 손실 함수를 사용한 결과를 보면 일부를 제외하면 모두 저 대역 신호만을 복원한 것을 확인할 수 있다. 복원된 일부분의 경우 무성음 구간으로 저 대역 에너지보다 고 대역 에너지가 매우 커 예외적으로 해당 부분만 복원이 된 것으로 보인다. 하지만 그 부분을 제외한 중간 대역과 고 대역 부분은 복원 성능이 매우 떨어지는 것을 확인할 수 있다. 반면 다중 대역 손실 함수를 사용하였을 때는 중간 대역의 신호를 복원할 수 있으며 기존의 손실 함수와 비교하였을 때 보다 선명한 고주파 신호를 복원하는 것을 확인할 수 있다. 따라서 다중 대역 손실 함수가 대역별 가중치를 적절하게 조절하여 다운 샘플링이 되지 않도록 모든 대역의 신호를 균등하게 복원하도록 하였음을 확인할 수 있다.

그림 3은 원본신호와 기존의 손실 함수를 사용하였을 때, 다중 대역 손실함수를 사용하였을 때의 스펙트럼이다. 마찬가지로 다중 대역 손실 함수를 사용한 경우 저 대역을 잘 복원 하고 포락선의 형태 또한 잘 복원

하는 것을 확인할 수 있다. 반면 기존의 손실 함수를 사용한 경우는 저 대역 복원에 치중되어 저 대역은 매우 잘 복원하지만 타 대역의 복원 성능이 매우 떨어지는 것을 확인할 수 있다.

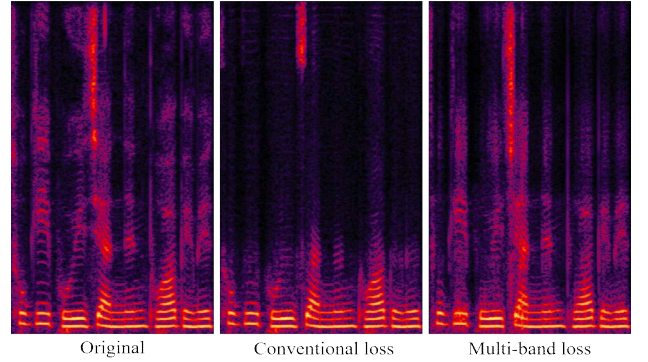


그림 2. 손실 함수에 따른 스펙트로그램

Fig. 2. Spectrogram for each loss function

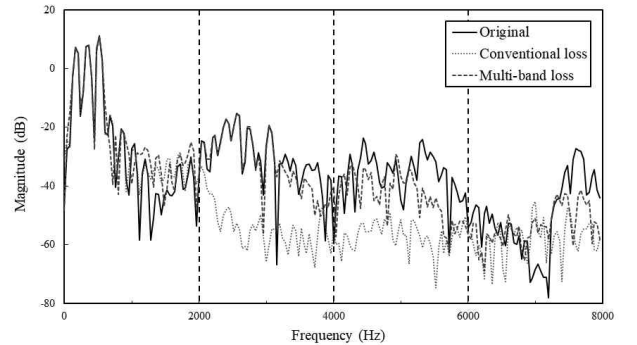


그림 3. 손실 함수 따른 스펙트럼

Fig. 3. Spectrum for each loss function

4. 결론

본 논문에서는 시간 영역 오토인코더의 성능 개선을 위한 다중 대역 손실 함수를 제안하였다. 제안하는 손실 함수가 적용된 오토인코더에 음성 신호를 입력하여 학습을 진행한 결과, 다운 샘플링이 발생하지 않으며 고 대역 신호가 복원되는 것을 스펙트로그램을 통해 확인하였다.

감사의 글

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2021R1F1A1059233).

참고문헌

- [1] L. Theis, W. Shi, A. Cunningham, and F. Huszár, "Lossy Image Compression with Compressive Autoencoders," arXiv:1703.00395, 2017
- [2] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," arXiv:2003.05991, 2020
- [3] B. Monson, E. Hunter, A. Lotto, and B. Story, Frontiers in Psychology, pp.1-11, 2014 dio:10.3389/fpsyg.2014.00587