

CC-GAN 과 VAW-GAN 을 이용한 한국어 음성 변환 실험

고봉구, 이건녕, 유인철, 육동석
고려대학교 컴퓨터학과 인공지능연구소

KOREAN VOICE CONVERSION EXPERIMENTS USING CC-GAN AND VAW-GAN

Bonggu Ko, Keonnyeong Lee, In-Chul Yoo, and Dongsuk Yook

Artificial Intelligence Laboratory, Department of Computer Science and Engineering, Korea University

중심어: 음성 변환, DNN, GAN, VAE, CC-GAN, VAW-GAN

요약

음성 변환은 발화의 내용을 유지하면서 화자의 목소리를 변환하는 것으로 익명화, 언어 교육 등에 활용될 수 있다. 최근 활발히 연구되고 있는 DNN(deep neural network)을 이용한 음성 변환 중, GAN(generative adversarial network)과 VAE(variational autoencoder) 등을 기반으로 한 모델이 비지도(non-parallel) 학습 데이터를 이용한 음성 변환에 활용되고 있다. GAN 의 일종인 CycleGAN(cycle-consistent adversarial network)[1]을 이용한 음성 변환이 가장 좋은 성능을 보이고 있으나 1 대 1 음성 변환만 가능하다. VAE 기반 음성 변환 방법은 다수 화자 사이의 many-to-many 음성 변환이 가능하지만 CycleGAN 에 비해 다소 낮은 성능을 보인다. 본 연구에서는 many-to-many 음성 변환이 가능하도록 CycleGAN 을 변형한 CC-GAN(conditional cycle-consistent adversarial network)[2]과 VAE 의 음질을 개선한 VAW-GAN(variational autoencoding Wasserstein GAN)[3]을 이용하여 한국어 음성 변환 실험을 수행하였다. 서울말 낭독체 코퍼스에서 남녀 각각 두 명씩 화자당 200 문장을 임의로 선택하여 모델을 학습하는데 사용하였다.

연구 지원 기관: 한국연구재단(NRF-2017R1E1A1A01078157), 정보통신기획평가원(2018-0-00269, 2018-0-01405).

참고문헌

1. T. Kaneko, *et al.*, "CycleGAN-VC2: Improved CycleGAN-based non-parallel voice conversion," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 6820-6824, 2019.
2. D. Yook, *et al.*, "Voice conversion using conditional CycleGAN," *International Conference on Computational Science and Computational Intelligence*, pp. 1460-1461, 2018.
3. C. Hsu, *et al.*, "Voice conversion from unaligned corpora using variational autoencoding Wasserstein generative adversarial networks," *Interspeech*, pp., 3364-3368, 2017.