Reproducibility Report ML (Machine Learning) Reproducibility Challenge 2021

Scope of Reproducibility

This work is intended to reproduce a paper named “ADASPEECH: ADAPTIVE TEXT TO SPEECH FOR CUSTOM VOICE” authored by Mingjian Chen, Xu Tan, Bohan Li, Yanqing Liu, Tao Qin, Sheng Zhao, and Tie-Yan Liu which was originally in ICLR 2021.

The above paper introduced AdaSpeech, which is primarily derived from Adaptive Text to Speech System. It claims to introduce a better version of the TTS system to handle the different acoustic conditions and trade off voice quality and adaption parameters.

Methodology

The code of the original author was used as a primary resource of this reproducibility attempt. custom voice, Adaspeech used three staged pipelines including pretraining, fine-tuning, and inference. In the first stage, it was a large multi-speaker dataset (LibriTTS (Zen et al., 2019) dataset). In the second stage, the model attempted to adapt to a new voice with diverse acoustic conditions using VCTK (Veaux et al., 2016) and LJSpeech (Ito, 2017) datasets. In the final stage, both unadapted and adapted parts were targeted to inference request.

Results

Experiments done during this reproducibility attempt showed that Adaspeech was successfully able to produce high-quality speech in handling diverse acoustic conditions very well. During different experiments, removing and adding the different layers of Adaspeech, it was concluded that each component is vital to optimal performance of AdaSpeech and removing any layer affects the sound quality.

What was easy

It was easy to run the authors code and re-implement the method as described by the authors in the original paper.

What was difficult

This work was not difficult to reproduce, the author’s code and methods described were explained in a concise manner. It has been proved an exceptionally fine piece of work to reproduce.

Communication with original authors

No contact was made with the original authors during this reproducibility attempt.