SVSGAN: SINGING VOICE SEPARATION VIA GENERATIVE ADVERSARIAL NETWORK

Zhe-Cheng Fan, Yen-Lin Lai, and Jhy-Shing Roger Jang
{lambert.fan, andy.lai, jang}@mirlab.org
Dept. Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan

Singing Voice Separation (SVS)

- Goal: Extract singing voice from the polyphonic audio music
- Restriction: Only one channel for analysis
- Approach: Deep neural network (DNN)

![Block diagram of SVS](image)

SVS via Generative Adversarial Network

- Parameters are initialized in a supervised setting
- Performance is optimized during adversarial learning
- Framework: Two conventional DNNs, G and D

![Block diagram of SVSGAN](image)

Experimental results

![Supervised Learning](image)

![Adversarial Learning](image)

Table 1. Vocal results (in dB) of conventional DNN and SVSGANs

![Table 1](image)

Fig. 2. Block diagram of SVSGAN

Fig. 3. The proposed SVSGAN framework. Each spectrum is considered to be a sample vector coming from a distribution of spectra.

Fig. 4. SVSGAN training process

Fig. 5. Vocal results on the Test part of the DSD100 dataset