

Abstract

The separation of sound sources in the decomposition of music has become a very interesting problem among scientists for the last 50 years. Separation of the sound source of this music has the main target of making it difficult for components in music such as vocals, bass, drums, etc. The results of the separation of sound sources have also been applied to many fields such as remixing, repanning, upmixing. In this thesis, the authors propose a method of separating the sound sources of musical instruments with vocals using Generative Adversarial Network (GAN) to rebuild the sound sources that exist in music. The GAN architecture built in this thesis, the author uses U-net with VGG19 as encoding block and mirror from VGG19 as encoder block on the generator and uses three times combination of Convolution, Batch Normalization and LeakyReLU blocks. The best scenario results (in dB) in this final project produce an average Source to Distortion Ratio (SDR) on test set are 7.03 for bass, 18.72 for drums, 20.20 for vocal and 12.73 for other.