

Pendahuluan

Latar Belakang

Music Source Separation mempunyai tujuan utama untuk membangun sebuah sistem yang dapat mengenali dan memisahkan sumber suara dari komposisi instrumen musik seperti bass, drum, vokal [23]. Untuk menyelesaikan masalah tersebut, sebelum ada banyaknya data yang dapat dipakai secara gratis dan terbuka, ilmuwan menggunakan metode sederhana atau signal processing "klasik" untuk pemisahan sumber suara musik, contohnya penggunaan Independent component analysis (ICA) [15], Fast ICA[8], Principal Component Analysis (PCA)[8], dan Non-Negative Matrix Factorization (NMF)[8]. Setelah adanya data yang cukup yang bersifat gratis dan terbuka, contohnya dataset DSD100 [12] dan populernya machine learning memunculkan banyaknya metode lain untuk menyelesaikan masalah ini, contohnya penerapan Deep Learning[3] dengan Convolutional Neural Network(CNN)[1], dan Recurrent Neural Network(RNN) [20] yang dapat menyelesaikan masalah pada pemisahan suara nyanyian pada suara musik [5]. Generative Adversarial Network(GAN) [7] yang merupakan perkembangan dari metode CNN, digunakan pada penelitian pemisahan sumber suara vokal pada tahun 2018 dan disebut dengan Singing Voice Separation Via Generative Adversarial Network(SVSGAN) [6]. Penelitian tersebut, Source to Distortion Ratio(SDR)[24] dipakai untuk pengukuran performa yang mendapatkan nilai SDR pada range -5 hingga 11 (dalam dB) yang dihitung menggunakan Blind Source Separation (BSS) Eval-toolbox v3.0[24] pada dataset DSD100. Penelitian lain juga menggunakan arsitektur CNN model U-Net [18] yang tujuan utama dibuatnya U-Net adalah untuk keperluan segmentasi citra medis, konsepnya diimplementasikan dan di modifikasi untuk memisahkan sumber suara yang dinamakan Wave-U-Net[22] dengan hasil rata-rata (mean) SDR dari vocal sebesar -2.1, SDR bass sebesar -0.3, SDR other sebesar 1.68 dan SDR drum sebesar 2.88 dengan menggunakan dataset MUSDB18[17]. Berdasarkan hasil dua penelitian sebelumnya, tugas akhir ini mencoba menggabungkan metode GAN dengan arsitektur generator yang menggunakan konsep U-net.

Topik dan Batasannya

Implementasi GAN untuk pemisahan sumber suara masih terbatas suara vocal, sehingga tugas akhir ini difokuskan dalam melakukan pemisahan sumber suara musik menggunakan GAN dengan mengimplementasikan konsep U-Net pada Generator yang optimal sehingga mendapatkan nilai SDR yang tinggi. Batasan pada tugas akhir ini adalah sebagai berikut: 1. Model dibuat dengan konsep Generative Adversarial Network(GAN) dengan konsep U-net pada generator. 2. Dataset yang digunakan adalah dataset DSD100. 3. Jumlah sumber suara yang dipisahkan sistem berjumlah empat yaitu vocal, bass, drum, dan other. 4. Skenario yang digunakan saat penelitian adalah sebagai berikut: • Menggunakan adversarial loss saat training. • Menggunakan adversarial loss dengan l1 loss saat training. Tujuan Tujuan dari tugas akhir ini adalah melakukan pemisahan sumber suara instrumen musik dan vokalis model GAN terbaik dan melakukan analisis terhadap skenario terbaik yang digunakan.

Organisasi Tulisan

Penelitian dibagi menjadi 5 bagian. Bagian pertama mengenai latar belakang mengapa penelitian dilakukan. Pada bagian kedua dijelaskan mengenai studi terkait dengan penelitian. Pada bagian ketiga dijelaskan sistem yang dibangun. Pada bagian keempat dijelaskan mengenai hasil dan evaluasi dari sistem yang telah dibangun. Pada bagian kelima, dibahas hasil penelitian yang dilakukan.

