

DIGIT SPEECH RECOGNITION MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK

Ardian Kurniawan
(16.04.1.1.1.00024)

Pembimbing I : Dr. Wahyudi Setiawan, S.Kom., M.Kom.
Pembimbing II : Dr. Fika Hastarita Rachman, S.T., M.Eng.

ABSTRAK

Salah satu disiplin ilmu dalam pengolahan sinyal digital ialah *speech recognition*. Implementasinya dapat digunakan untuk sistem absensi, keamanan dan entri PIN. Penelitian ini membahas tentang *speech recognition* berupa digit 0 sampai 9 dalam Bahasa Inggris. *Dataset* ujicoba diambil dari *storage googleapis speech command*. Penelitian ini menggunakan metode *Convolutional Neural Network* (CNN) 24 *layer*. Proses penelitian terdiri dari pengambilan suara (.wav) di *dataset* yang diubah dari sinyal digital menjadi spektogram, selanjutnya proses ekstraksi fitur menggunakan CNN pada *Feature Extraction Layer*, hasil ekstraksi ini dilatih dan diuji oleh *layer* CNN berikutnya yaitu *Classification Layer*. Terakhir, pengenalan dilakukan dengan menginputkan suara ke sistem, sehingga sistem dapat mengenali suara tersebut. Hasil ujicoba menunjukkan pengukuran kinerja terbaik menggunakan inisialisasi parameter *Learning Rate* 0.001, *Max-epoch* 30, *Validation Frequency* 300, dan *Minibatch-size* 64 diperoleh nilai *Precision* 97.46%, *Recall* 97.47% dan Akurasi 97.45%. Kemudian hasil *recognition* suara akurasi terbaik diperoleh menggunakan data augmentasi sebesar 92% dengan 10 data setiap kelas dengan *spoken digit dataset*.

Kata Kunci : *Spektogram, Speech Recognition, Convolutional Neural Network*



DIGIT SPEECH RECOGNITION USING CONVOLUTIONAL NEURAL NETWORK

Ardian Kurniawan
(16.04.1.1.1.00024)

Pembimbing I : Dr. Wahyudi Setiawan, S.Kom., M.Kom.
Pembimbing II : Dr. Fika Hastarita Rachman, S.T., M.Eng.

ABSTRACT

One of the disciplines in digital signal processing is speech recognition. Its implementation can be used for attendance, security and PIN entry systems. This study discusses speech recognition in the form of digits 0 to 9 in English. The test dataset is taken from the googleapis speech command storage. This study used a 24 layer Convolutional Neural Network (CNN) method. The research process consists of voting (.wav) in the dataset which is converted from a digital signal to a spectrogram, then the feature extraction process uses CNN on the Feature Extraction Layer, the extraction results are trained and tested by the next CNN layer, namely the Classification Layer. Finally, recognition is done by inputting a voice into the system, so that the system can recognize the voice. The test results showed that the best performance measurement using the initialization parameters Learning Rate 0.001, Max-epoch 30, Validation Frequency 300, and Minibatch-size 64 obtained Precision 97.46%, Recall 97.47% and accuracy 97.45%. Then the best accuracy voice recognition results were obtained using augmentation data of 92% with 10 data per class with the spoken digit dataset.

Keywords : Spectrograms, Speech Recognition, Convolutional Neural Network

