

Perbandingan Depthwise Separable Convolutional Neural Network dan Convolutional Neural Network Sebagai Multiclass Keyword Spotting Pada Edge Device = Comparison of Depthwise Separable Convolutional Neural Network and Convolutional Neural Network As Multiclass Keyword Spotting On Edge Devices

Bayu Satria Persada, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20516942&lokasi=lokal>

Abstrak

Perkembangan Artificial Intelligence (AI) sudah berkembang pesat. Dari ketiga arah pengembangan AI yakni computer vision, speech processing dan natural language processing. Speech processing memiliki tren paling rendah di antara ketiga pengembangan tersebut. Meskipun begitu pengembangan di bidang speech processing seperti speech recognition dan keyword spotting sudah banyak di implementasikan seperti model keyword spotting menggunakan Convolutional Neural Network (CNN) di microcontroller, mobile device dan perangkat lainnya. Namun CNN saja belum tentu menghasilkan akurasi yang tinggi maka dicoba Depthwise Separable Convolutional Neural Network (DSCNN) untuk mendapatkan hasil dengan akurasi yang lebih tinggi. Pengembangan model keyword spotting belum banyak diimplementasikan di edge device lainnya, yang dimaksud dengan edge device yaitu perangkat sederhana di sisi pengguna yang kemampuan komputasinya terbatas. Dengan menggunakan DSCNN menunjukkan nilai F1 score yang dibandingkan dengan model CNN. Model DSCNN menghasilkan model dengan nilai F1 score paling optimal dengan 4 layer konvolusi depthwise separable, menggunakan filter konvolusi sebanyak 256 dengan jumlah filter konvolusi depthwise 512 menggunakan optimizer RMSprop dan menggunakan batch size berukuran 126. Dari hasil pengujian dapat diketahui bahwa secara umum DSCNN menghasilkan F1 score yang lebih baik dibandingkan CNN yaitu sebesar 31,8% dengan CNN sebesar 28,35%. Namun DSCNN menggunakan sumber daya yang lebih banyak dan lebih lama waktu responsnya.

.....The development of Artificial Intelligence (AI) has grown rapidly. Of the three directions of AI development, namely computer vision, speech processing, and natural language processing. Speech processing has the lowest trend among the three developments. However, many developments in speech processing such as speech recognition and keyword spotting have been implemented, such as the keyword spotting model using the Convolutional Neural Network (CNN) in microcontrollers, mobile devices, and other devices. However, CNN alone does not necessarily produce high accuracy, so a Depthwise Separable Convolutional Neural Network (DSCNN) is used to get results with higher accuracy. The development of the keyword spotting model has not been widely implemented in other edge devices, which is meant by edge devices, namely simple devices on the user's side with limited computing capabilities. Using DSCNN shows the F1 score which is compared with the CNN model. The DSCNN model produces a model with the most optimal F1 score with 4 layers of convolution depthwise separable, using a convolution filter of 256 with a convolution depthwise filter of 512 using the RMSprop optimizer and using a batch size of 126. From the test results, in general DSCNN produces F1 score which is better than CNN, which is 31,8% with CNN at 28,35%. However, DSCNN uses more resources and a longer response time.