

Abstract

This research contributes to integrating Selective Kernel (SK) and Convolutional Block Attention Module (CBAM) methods for adapting perception field and enriched feature extraction with basis architecture Visual Geometry Group (VGG). Furthermore, Batch Normalization (BN) will be applied to the model to improve generalization. CBAM is an attention method that combines two modules, channel and spatial to enrich feature extraction. SK is an attention method that focuses on selecting kernel adaptively. VGG is a model architecture based on a Convolutional Neural Network (CNN). VGG architecture utilizes convolution filters (3x3) that improve significantly compared to previous configurations. This architecture has won the ImageNet Challenge 2014 competition. The metrics for this research are accuracy, precision, recall, F-1 score, time, and sparse top-k categorical accuracy. Furthermore, the benchmark datasets are CIFAR-10 and CIFAR-100. According to the experiments, the best model got an accuracy of 86%, sparse top-k categorical accuracy of 99%, precision of 86%, recall of 86%, and F-1 score of 86% on the CIFAR-10 dataset. Based on the result, the model improved accuracy, precision, recall, and F-1 score by 7% compared to the CNN model at the beginning. However, there is an increase in training time by 34,5 minutes. Additionally, on the CIFAR-100 dataset, the best model got an accuracy of 60%, sparse top-k categorical accuracy of 85%, precision of 60%, recall of 60%, and F-1 score of 59%. Based on these results, the model improved by an accuracy of 21%, sparse top-k categorical accuracy of 16%, precision of 20%, recall of 21%, and F-1 score of 21% compared with the CNN model at the beginning. However, there is an increase in training time by 62,1 minutes. Suggestions for further research include experimenting with different datasets, using the CNN model architecture, and applying more efficient convolution methods. This concludes a significant increase in the model performance. However, there is an increase in training time.

Keywords - *Selective Kernel, Convolutional Block Attention Module, Visual Geometry Group, Batch Normalization, CIFAR Datasets*

Abstrak

Penelitian ini berkontribusi pada pengintegrasian metode *Selective Kernel* (SK) dan *Convolutional Block Attention Module* (CBAM) untuk mengadaptifkan bidang persepsi dan memperkaya ekstraksi fitur dengan basis arsitektur *Visual Geometry Group* (VGG). Selain itu, model diterapkan *Batch Normalization* (BN) untuk meningkatkan generalisasi model. CBAM merupakan sebuah metode *attention* dengan menggabungkan dua modul yaitu kanal dan spasial untuk memperkaya ekstraksi fitur. Selain itu, CBAM dapat mudah diintegrasikan pada model basis *Convolutional Neural Network* (CNN). SK merupakan sebuah metode *attention* yang berfokus pada pemilihan *kernel* secara adaptif. VGG merupakan sebuah arsitektur dengan basis CNN untuk mengklasifikasi gambar. Arsitektur VGG memanfaatkan filter konvolusi (3x3), yang memberikan peningkatan signifikan dibandingkan konfigurasi sebelumnya. Arsitektur tersebut memiliki prestasi dengan memenangkan kompetisi *ImageNet Challenge* 2014. Metrik utama pada penelitian ini diantaranya akurasi, presisi, *recall*, *f-1 score*, waktu, dan *sparse top-k categorical accuracy*. Selain itu, *dataset* tolok ukur yang digunakan pada penelitian ini yaitu CIFAR-10 dan CIFAR-100. Berdasarkan hasil percobaan model terbaik mendapatkan akurasi 86%, *sparse top-k categorical accuracy* 99%, presisi 86%, *recall* 86%, dan *f-1 score* 86% pada *dataset* CIFAR-10. Hasil tersebut mendapatkan peningkatan pada metrik akurasi, presisi, *recall*, dan *f-1 score* sebesar 7% dibandingkan dengan model CNN awal. Namun, terjadi peningkatan waktu pelatihan sebanyak 34,5 menit. Selanjutnya, pada *dataset* CIFAR-100 model terbaik mendapatkan akurasi 60%, *sparse top-k categorical accuracy* 85%, presisi 60%, *recall* 60%, dan *f-1 score* 59%. Hasil tersebut terjadi peningkatan akurasi 21%, *sparse top-k categorical accuracy* 16%, presisi 20%, *recall* 21%, dan *f-1 score* 21% dibandingkan model CNN awal. Namun, terjadi peningkatan waktu pelatihan sebanyak 62,1 menit. Saran penelitian selanjutnya dapat bereksperimen pada *dataset* yang berbeda, arsitektur CNN yang berbeda, dan menerapkan metode untuk mengefisienkan konvolusi. Dapat disimpulkan dengan hasil tersebut model mendapatkan peningkatan performa secara signifikan. Namun, terjadi peningkatan waktu pelatihan.

Kata Kunci - *Selective Kernel*, *Convolutional Block Attention Module*, *Visual Geometry Group*, *Batch Normalization*, *Dataset* CIFAR