

INTISARI

Pemanfaatan model klasifikasi dan pengenalan ekspresi wajah semakin merambah ke dalam banyak bidang. Pembuatan model menggunakan metode *deep learning Convolutional Neural Network* (CNN) yang biasa digunakan dalam pemrosesan gambar sampai pada pola kisi memiliki banyak jenis arsitektur yang dapat digunakan. Namun, tiap arsitektur memiliki perbedaan konsep dan parameter yang digunakan. Perbedaan ini dapat menghasilkan kinerja yang berbeda pula, disesuaikan dengan jenis tugas dan kompleksitas tugas tersebut. Maka dari itu, dengan menggunakan tugas untuk klasifikasi dan pengenalan ekspresi wajah, dilakukan perbandingan analisis kinerja untuk mengetahui arsitektur terbaik dalam pengenalan ekspresi wajah dan analisis penyebab yang memberikan perbedaan hasil. Metode yang digunakan adalah melakukan perbandingan dari beberapa arsitektur CNN, yaitu VGG16, Inception-V3, dan Resnet50 yang diterapkan pada beberapa jenis *dataset* ekspresi wajah, yaitu FER-2013, *Extended Cohn-Kanade* (CK+), dan AffectNet dengan menerapkan *transfer learning* untuk memanggil model yang sudah dilatih. Metrik yang digunakan untuk menganalisis kinerja arsitektur CNN adalah akurasi, *loss*, *precision*, dan *recall*. Proses *pra-training* dilakukan untuk mengetahui pengaruh keseimbangan jumlah masing-masing label yang menggunakan kombinasi *dataset* dari FER-2013 dan AffectNet. Jumlah seimbang pada masing-masing label tidak memberikan kinerja model dengan baik. Akurasi pada *test set* hanya mencapai maksimal 56,91% pada arsitektur Inception-V3. Pada pengujian jenis *dataset*, ketiga arsitektur memberikan hasil terbaik pada *dataset* CK+. Arsitektur VGG16 menghasilkan akurasi 100% dengan *loss* 2,7788. Arsitektur Inception-V3 memberikan hasil terbaik pada akurasi 100% dan *loss* 0,0790. Arsitektur Resnet50 menghasilkan akurasi sebesar 98,99% dan *loss* 0,0747. Sementara itu, *dataset* FER-2013 dan kombinasi CK+ dan FER-2013 menghasilkan akurasi tes maksimal 60,43% dan 71,07% pada arsitektur VGG16. Sedangkan nilai *loss* terendah adalah 1,8477 pada arsitektur Resnet50. Dari penelitian yang sudah dilakukan, pengembangan model klasifikasi dan pengenalan ekspresi wajah ke depannya dapat menggunakan arsitektur dengan pengaturan terbaik yang disesuaikan dengan jenis dan kompleksitas *dataset*.

Kata kunci: Pengolahan citra, Pengenalan ekspresi wajah, *Deep learning*, CNN, *Transfer learning*

ABSTRACT

The utilization of facial expression classification and recognition models is increasingly expanding across various fields. Creating models using deep learning Convolutional Neural Network (CNN) methods commonly applied in image processing, including grid patterns recognition, offers a lot of architecture options. However, each architecture possesses distinct concepts and parameters that yield varying performances, depending on the task type and its complexity. Hence, a performance analysis and comparison were conducted to determine the optimal architecture for facial expression recognition and evaluate the cause that leads to different outcomes. The method involved comparing several CNN architectures, namely VGG16, Inception-V3 and Resnet50, applied to different facial expression datasets, namely FER-2013, Extended Cohn-Kanade (CK+) and AffectNet, by implementing transfer learning to utilize pre-trained models. Performance metrics such as accuracy, loss, precision and recall were utilized to analyze CNN architecture's performance. Prior to training, a preprocessing step was performed to assess the influence of label balance, using a combination of the FER-2013 and AffectNet datasets. However, achieving a balanced distribution of labels did not result in satisfactory model performance. The maximum accuracy achieved on the test set was only 56.91% with the Inception-V3 architecture. Regarding the dataset types, all three architectures yielded the best result on CK+ dataset. The VGG16 architecture achieved 100% accuracy with a loss of 2.7788. The Inception-V3 architecture demonstrated the best performance with 100% accuracy and a loss of 0.0790. The Resnet50 architecture achieved an accuracy of 98.99% with a loss of 0.0747. Meanwhile, the FER-2013 and combination of CK+ and FER-2013 datasets produced the best test accuracy of 60.43% and 71.07% with the VGG16 architecture. While the lowest loss value is 1.8477 with a Resnet50 architecture. Based on the conducted research, future development and advancements of facial expression classification and recognition models can employ architectures with optimized settings and configurations that are tailored to the specific characteristics and complexity of the dataset.

Keywords: Image processing, Facial expression recognition, Deep learning, CNN, Transfer learning