

INTISARI

Pembangkitan Deskripsi Gambar dalam Bahasa Indonesia Menggunakan Pendekatan Berbasis Transformer

Oleh

Raphael Discky Zundria Putra
19/442487/PA/19236

Image captioning merupakan sistem yang menggabungkan bidang *computer vision* dengan *Natural Language Processing* (NLP) yang bertujuan untuk menghasilkan kalimat yang mendeskripsikan sebuah gambar. Sistem ini bermanfaat dalam membantu kegiatan manusia memahami konten visual pada gambar dan membantu mendeskripsikan gambar kepada tunanetra. Penelitian *image captioning* yang dihasilkan dalam bahasa Indonesia sebagian besar menggunakan gabungan *Convolutional Neural Network* (CNN) dan *Recurrent Neural Network* (RNN), sedangkan pada penelitian terkini menunjukkan bahwa menggunakan arsitektur *transformer* dapat menghasilkan performa model yang lebih baik.

Penelitian ini mengembangkan model *image captioning* yang mempunyai performa yang lebih baik dalam menghasilkan deskripsi gambar dalam bahasa Indonesia dibandingkan penelitian sebelumnya dengan menggunakan arsitektur berbasis *transformer*. Model yang dikembangkan adalah *transformer* dengan penambahan *pretrained Convolutional Neural Network* (CNN) yaitu EfficientNet dan *Residual Neural Network* (ResNet) pada lapisan *encoder* untuk mengekstraksi fitur gambar dan lapisan *decoder* untuk menghasilkan kalimat dalam bahasa Indonesia. Data diperoleh dari COCO *dataset* dengan pembagian menurut penelitian oleh Karpathy.

Hasil eksperimen menunjukkan bahwa model yang dikembangkan menggunakan penambahan EfficientNet mempunyai hasil yang lebih baik dibandingkan jika menggunakan ResNet dengan kenaikan sebesar 4.9% dari hasil rata-rata nilai metrik BLEU-1 hingga BLEU-4, METEOR, ROUGE-L, dan CIDEr.

Kata kunci: Pembangkitan Deskripsi Gambar, Bahasa Indonesia, *Transformer*, EfficientNet, ResNet, *Computer Vision*, *Natural Language Processing*

ABSTRACT

Image Caption Generation In Indonesian Language Using a Transformer-Based Approach

By

Raphael Discky Zundria Putra
19/442487/PA/19236

Image captioning is a system that combines the field of computer vision with Natural Language Processing (NLP) which aims to produce sentences that describe an image. This system is useful in helping human activities understand the visual content of images and helping describe images to the visually impaired. Research on image captioning in Indonesian mostly uses a combination of Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), while recent research shows that using a transformer architecture can produce better model performance.

This research develops an image captioning model based on transformer architecture which has good performance in producing image descriptions in Indonesian. The model developed is a transformer with the addition of pretrained Convolutional Neural Network (CNN) namely EfficientNet and Residual Neural Network (ResNet) at the encoder layer to extract image features and decoder layer to produce sentences in Indonesian. Data obtained from the COCO dataset with division according to research by Karpathy.

The experimental results show that the model developed using the addition of EfficientNet has better results than using ResNet with an increase of 4.9% from the average yield of BLEU-1 to BLEU-4, METEOR, ROUGE-L, and CIDEr metric values.

Keyword: Image Captioning, Indonesian Language, Transformer, EfficientNet, ResNet, Computer Vision, Natural Language Processing

