Abstrak

Image captioning is one of the most widely discussed topic nowadays. However, most research in this area generate English caption while there are thousands of language exist around the world. With their language uniqueness, there's a need of specific research to generate captions in those languages. Indonesia, as the largest Southeast Asian country, has its own language, which is Bahasa Indonesia. Bahasa Indonesia has been taught in various countries such as Vietnam, Australia, and Japan. In this research, we propose the attention-based image captioning model using ResNet101 as the encoder and LSTM with adaptive attention as the decoder for the Indonesian image captioning task. Adaptive attention used to decide when and at which region of the image should be attended to produce the next word. The model we used was trained with the MSCOCO and Flick30k datasets besides. Both datasets are translated manually into Bahasa by human and by using Google Translate. Our research resulted in 0.678, 0.512, 0.375, 0.274, and 0.990 for BLEU-1, BLEU-2, BLEU-3, BLEU-4, and CIDEr scores respectively. Our model also produces a similar score for the English image captioning model, which means our model capable of being equivalent to English image captioning. We also propose a new metric score by conducting a survey. The results state that 76.8\% of our model's caption results are better than validation data that has been translated using Google Translate.