

Building a program to generate descriptive captions for images using CNN & LSTM.

Ngo Anh Tung

Abstract: This topic focuses on building an automated system for generating descriptive captions for images by combining deep learning methods such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. The practical context has driven this research to meet the urgent need for automating the caption generation process. The objective is to develop a system capable of analyzing images, recognizing image content, and converting them into accurate descriptive text. The method uses CNN for image processing and LSTM to capture the temporal context of terms in the captions. The expected outcome of this research is to provide a useful tool for the community, with broad potential applications in education, media, and assistive services for people with disabilities.

Keywords: deep learning, convolutional neural network (CNN), long short-term memory (LSTM), image captioning