World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:17, No:04, 2023

Extending Image Captioning to Video Captioning Using Encoder-Decoder

Authors : Sikiru Ademola Adewale, Joe Thomas, Bolanle Hafiz Matti, Tosin Ige

Abstract : This project demonstrates the implementation and use of an encoder-decoder model to perform a many-to-many mapping of video data to text captions. The many-to-many mapping occurs via an input temporal sequence of video frames to an output sequence of words to form a caption sentence. Data preprocessing, model construction, and model training are discussed. Caption correctness is evaluated using 2-gram BLEU scores across the different splits of the dataset. Specific examples of output captions were shown to demonstrate model generality over the video temporal dimension. Predicted captions were shown to generalize over video action, even in instances where the video scene changed dramatically. Model architecture changes are discussed to improve sentence grammar and correctness.

Keywords : decoder, encoder, many-to-many mapping, video captioning, 2-gram BLEU
Conference Title : ICIPACV 2023 : International Conference on Image Processing, Analysis and Computer Vision
Conference Location : Jerusalem, Israel
Conference Dates : April 24-25, 2023