

주 객체 검출 기반의 이미지 캡셔닝 연구

배주원*, 이수환*, 성주현*, 서동환**

*한국해양대학교 전기전자공학과

**한국해양대학교 전기전자공학부

e-mail : wjb0406@kmou.ac.kr

A study on Image Captioning based on main object detection

Ju-Won Bae*, Soo-Hwan Lee*, Ju-Hyeon Seong*, Dong-Hoan Seo**

*Dept of Electrical and Electronics Engineering, Korea Maritime and Ocean University

**Division of Electronics and Electrical Engineering, Korea Maritime and Ocean University

1. 연구 필요성 및 문제점

이미지 내의 객체나 상황에 대해 인간의 언어로 표현하는 이미지 캡셔닝(Image Captioning) 기술은 딥러닝 기술 발전의 영향으로 CNN(Convolution Neural Network)을 이용해 입력 이미지를 가공해 RNN(Recurrent Neural Network)으로 캡션을 만들어내는 Encoder-Decoder 방식 [1]이나 지도 학습(Supervised Learning) 방식을 이용한 이미지 캡셔닝 등 다양한 접근법들로 연구되고 있다.

이미지 캡셔닝의 목표는 이미지 내의 객체나 상황에 대한 정확한 이해와 그에 대해 인간이 사용하는 문장의 수준과 거의 근접하게 묘사되는 것이다. 많은 연구들로 여러 성과를 냈음에도 불구하고 아직 출력되는 문장이 딱딱하게 느껴지거나 이미지 내의 상황이나 객체에 대해 모두 서술하지만 이미지를 만들어낸 사람이 원하는 주제에 부합하거나 이미지 생성 당시의 상황에 맞는 설명이 제대로 생성되지 않는 한계가 있다.

2. 연구내용과 방법

특정 상황을 나타내고 있는 이미지 내에는 객체와 상황과 같은 주요 요소들이 있고 이 중에는 이미지가 나타내는 정보에서 주인공이 되는 주 객체가 존재한다. 주 객체는 주로 이미지를 설명하는 문장에서 주어로 사용되고 이를 중심으로 설명을 만들어내는 중심의 역할을 한다. 그러므로 이미지 내의 주 객체를 찾아내 캡션을 만들면 더 풍부하고 이미지의 주제에 맞는 캡션을 생성할 수 있을 것이다. 본 연구에서는 이미지에서 주 객체를 검출하는 방법을 사용해 캡션을 생성한다.

본 연구는 Encoder-Decoder기반의 이미지 캡셔닝 알고리즘을 사용한다. 알고리즘의 전개 과정은 그림 1과 같다. 먼저 Encoder에서 이미지를 읽어 들여 이미지 내에 객체나 상황 등을 나타내는 특징들을 그림 1의 원처럼 찾아낼 때 주 객체를 찾아 주 객체에 더 비중을 주는 작업을 한다. 그 다음, Decoder에서 주 객체를 더 강조한 데이터를 받아 캡션을 만들어낸다. 본 연구는 문장과 이미지가 함께 있는 flickr 8K 데이터셋을 사용했고 출력된 문장이 데이터셋이 가지고 있는 문장과 일치한지 비교한다.

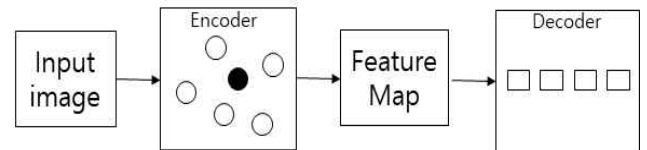


그림 1. 주 객체 검출 기반 이미지 캡셔닝 알고리즘의 구조

이미지에서 주 설명이 되는 주 객체들을 검출하는 기준은 다음과 같다. 이미지 내 객체들 중 가장 큰 객체이거나, 정중앙의 위치에 있는 객체, 객체들 간의 분포에서 중간에 있는 객체를 주 객체로 정의하여 연구를 수행한다. 본 연구에서는 Encoder에서 위의 기준을 만족하는 객체를 검출하는 알고리즘을 제시한다. 그리고 특징 맵 시 가중치를 더 주는 방식을 이용해 Decoder에서 선정된 주 객체를 중심으로 문장을 생성할 수 있게 한다.

3. 결론 및 향후 연구

본 논문에서는 주 객체 검출 기반의 이미지 캡셔닝 알고리즘을 연구하였다. 기존의 이미지에서 검출된 특징들을 이용해 캡션을 만들어내는 구조와는 다르게 본 논문에서 사용한 알고리즘은 캡션을 생성하기 이전인 Encoder단계에서 이미지의 설명에서 주어로 사용할 주 객체를 먼저 검출하고 Decoder에서는 이를 우선적으로 활용하여 캡션을 만들어낸다. 그리고 생성된 캡션과 데이터셋이 가진 캡션과 정확도를 비교했다.

향후 본 논문에서 제시한 관계를 중심으로 한 이미지 캡셔닝 기술을 사용해 만들어낸 캡션으로 더 풍부하고 목적에 맞는 정보전달을 위한 방법을 연구할 예정이다.

후 기

이 논문은 2016년도 정부(교육부)의 재원으로 한국연구재단 기본연구지원사업의 지원을 받아 수행된 기본연구임(No.2016R1D1A1B03934812)

참고문헌

- [1] O. Vinyals, A. Toshev, S.Benigio, D.Erhan "Show and Tell: A neural image caption generator", IEEE conference on computer vision and pattern recognition, pp. 3156-3164, 2015