



# 감시 시스템을 위한 새로운 캡션 기법

나광호, 이진경, 전민성, 최경주

충북대학교 소프트웨어학과

## ABSTRACT

최근 딥러닝 기술의 발전으로 객체 탐지 성능이 향상되어, 지능형 감지 시스템의 수요가 증가하고 있다. 본 논문에서는 이미지 캡셔닝 기술을 사용한 감지 시스템을 제안한다. 보다 효과적인 감시 시스템을 만들기 위해 객체의 특징 정보, 공간 정보, 행동 정보 간 관계를 유기적으로 해석하는 캡션을 생성하는 방법을 제안한다. 감시 시스템 활용에 적절한 새로운 이미지와 캡션 쌍의 데이터 세트를 구축하여 BLP 모델을 새로 구축한 데이터셋으로 파인 튜닝을 진행하였다. 실험 결과 90% 성능으로 위험 상황, 사고에 대한 캡션을 생성함을 확인할 수 있었으며, 제안하는 방법이 위험 상황과 사고를 중심으로 캡션을 생성하며 객체뿐만 아니라 행동과 공간 정보를 포함한 구체적인 묘사가 가능함을 확인할 수 있었다.

## INTRODUCTION

### 연구 이유?

최근 컴퓨터비전 영역의 딥러닝 기술의 급속한 발전으로 인해 객체 탐지 및 추적 등의 성능이 향상되어, 지능형 CCTV는 차세대 CCTV로 인정받으며 그 수요가 증가하고 있다. 본 연구에서는 기존의 객체 탐지에 기반하지 않고, 이미지 캡셔닝 기술을 사용한 새로운 접근 방법의 감시 시스템을 제안하고자 한다. 이미지 캡셔닝 기술을 활용하면 장면의 복잡한 상황을 인간의 언어로 구체적으로 표현할 수 있으며, 이를 통해 감시자는 빠르게 상황을 파악할 수 있게 된다.

### 기존 데이터 세트의 문제점

첫째, 현재 공개된 대부분의 캡션 데이터 세트는 이미지 내 큰 객체에 중점을 두고 설명하므로, 항상 사람 중심의 해석을 보장하지 않는다.

둘째, 같은 행동을 포함한 객체 정보가 공간 정보를 활용하지 않을 경우 위험 상황을 검출하지 못할 수 있다.

셋째, 대규모 데이터를 사용하여 학습된 대규모 멀티모달 모델은 일반적인 상황의 이미지와 캡션이 대다수를 차지하기 때문에 감시 시스템이 검출하고자 하는 상황을 정확하게 설명하는데 한계가 있다.

### 제안

이러한 문제점을 해결하기 위하여 객체의 특징 정보, 공간 정보, 행동 정보 간 관계를 유기적으로 해석하는 감시 시스템에 적절한 캡션을 생성하는 방법을 제안한다. 이를 위해 감시 시스템에 적용 가능한 이미지와 캡션 쌍의 데이터 세트를 구축하였고, 이 데이터 세트를 활용하여 BLP 모델을 활용한 이미지 캡셔닝 기반의 감시 시스템을 제안한다.

## METHOD

### 이미지 캡션의 새로운 구조 제안

캡션의 문장 구조에는 **객체 정보**, **객체의 행동**, **행동이 발생한 공간 정보**가 모두 포함되도록 한다.



The **old man wearing a green jacket** is **sitting** on a **bench** in the **park**.



A **woman wearing a gray cardigan** is **sitting** on a **rooftop** **railing**.

좌측 그림은 같은 행동을 포함한 객체 정보가 공간 정보를 활용하지 않을 경우 위험 상황을 검출하지 못할 수 있음을 보여주는 예시이다.

상단 그림에서, 노인이 의자에 앉아 있는 행동은 정상적인 상황인 반면, 하단 그림의 난간 위에 걸터 앉은 여성의 행동은 위험 상황으로 볼 수 있다.

동일한 행동이라도 행동의 객체와 공간의 맥락에 따라 위험도가 크게 달라질 수 있음을 고려하였다.

## RESULTS

### 데이터 세트

총 2,193장의 이미지를 수집하였고, 각각의 이미지는 모두 제안한 구조에 맞춘 캡션을 가지도록 구축되었다.

클래스	싸움	쓰러짐	화재	교통사고	위험행동
개수	505	601	205	338	544

### 비교 실험 결과

기존의 BLP의 모델과 BLP 모델을 새로 구축한 데이터 세트로 파인 튜닝을 한 결과를 비교한 것이다. 제안 방법으로 구축한 데이터 세트는 기존 캡션 모델에서 수행하기 어려웠던 위험 상황에 대한 캡션 생성뿐만 아니라 객체와 특징, 행동, 공간에 대한 정보를 제안한 형식에 맞게 문장을 효과적으로 생성하는 것을 확인하였다.



BLP(기존) : two boys in a parking

제안방법 : a **man in black clothes** is **punching** a **man in white clothes** on **the road**.



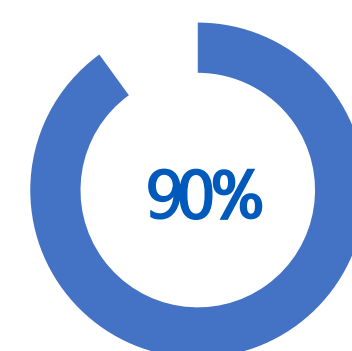
BLP(기존) : a woman is playing a video game on a tv

제안방법 : a **man wearing a black top** is **lying down** next to a **blue board** in the **office**.

### 정량적 평가

각 클래스 별로 20장씩 총 100장의 데이터 세트에 대한 모델의 위험 요소 검출 성공률 측정하였고, 90%의 성능으로 위험 상황, 사고에 대한 캡션을 생성하였다.

위험 요소 검출 성공률



## CONCLUSION

본 연구에서는 이미지 캡셔닝 기술을 사용한 새로운 접근 방법의 감시 시스템을 제안하였다. 이미지 캡셔닝 기술을 감시 시스템에 효과적으로 적용하기 위해 객체의 특징 정보, 공간 정보, 행동 정보 간 관계를 유기적으로 해석하는 캡션을 생성하는 방법을 제안하였고, 이 방법에 따라 이미지와 캡션 쌍의 새로운 데이터 세트를 구축하였다.

제안하는 방식으로 구축한 데이터 세트는 기존의 벤치마크 데이터 세트의 단점을 해결하여 감시 시스템에 적용할 수 있는 캡션 데이터 세트의 기준을 제시하였다.