

# HOTR을 활용한 유해성 감지 모델 개발

이두진 전민주 임청수 정세웅

# Contents

1. Background
2. Project Overview
3. Dataset
4. Modeling
5. Implementation

# 1. Background

시사 > 전제기사

## ‘신림동 칼부림’ 피해자 1명 사망...나머지 3명은 부상

입력 : 2023-07-21 15:30 / 수정 : 2023-07-21 20:04



21일 오후 신림역 4번 출구 인근 현장에 폴리스라인이 구 신림동 신림역 4번 출구 인근에서 괴한이 휘두른 칼부림 난동을 부려 시민 1명이 숨지고 3명이 다쳤다. 연합뉴스

서울 관악구 신림동에서 30대 남성이 칼부림 난동을 부리고 있다.

21일 경찰에 따르면 이날 오후 2시쯤 신림역 4번 출구 인근에서 30대 남성이 칼부림 난동을 부려 시민 1명이 숨지고 3명이 다쳤다.

## 신림동 이어 이번에는 서현역 문지마 칼부림 사건

차로 인도 돌진 후, 흥기로 행인들 찔러

김경민 기자



경찰이 3일 오후 문지마 흥기난동이 발생한 경기도 성남시 분당구 서현역 한 백화점에서 사건을 처리하고 있다. 피의자 A씨는 경찰차를 끌고 서현역 앞 인도로 돌진, 보행자 다수를 충격한 뒤 내려 역사 안으로 도주한 것으로 전해졌다. photo 뉴시스

경기 성남시 서현역 인근에서 또 다시 '문지마 칼부림' 사건이 3일 발생했다. 신림역 사건과 유사한 유형의 사건이 반복돼 시민들에게 충격을 던져주고 있다.

경찰과 소방 당국에 따르면 오후 5시 59분쯤 성남시 분당구 이매동 서현역 AK프라자 상용차를 타고 인도에 돌진한 후 칼을 들고 백화점으로 들어가 흥기를 휘둘렀다. 이

사회

## 살인예고 하루새 20건 ..."주말 약속 취소했어요"

권선미 기자 kwon.seonmi@mk.co.kr · 최예빈 기자 ybl2@mk.co.kr

입력 : 2023-08-04 17:34:26

가 0 0 0 0

강남·부산·분당 등 전국서  
모방범죄 늘고 예고글 봇물  
"공권력도 못믿어" 불안확산  
주요역·대형몰 순찰 늘리고  
폭발물 탐지반 투입되기도



경기 성남시 분당 서현역 흥기 난동 사건 이후 비슷한 범행을 저지르겠다는 예고성 글이 인터넷에 잇달아 올라오고 있는 가운데 4일 범행 예고 장소 중 한 곳인 성남 오리역에 경찰이 배치돼 있다(왼쪽). 오른쪽은 흥기 난동 예고글이 올라온 서울 잠실역 주변에 경찰이 출동해 현장 상황을 점검하고 있는 모습. 이충우 기자

4일 오전 서울 지하철 2호선 잠실역 일대에는 긴장감이 감돌았다. 전날 모 인터넷 사이트에 "내일 아침 잠실역에서 20명 죽일 것"이라는 게시글이 올라왔기 때문이다. 경찰 측은 오전 7시 50분부

# 1. Background

시사  
'신림동 칼부림' 피해자 15

입력 : 2023-07-21 15:30 / 수정 : 2023-07-21 20:04



21일 오후 신림역 4번 출구 인근 현장에 출구 신림동 신림역 4번 출구 인근에서 피해자가 이송됐다. 연합뉴스

서울 관악구 신림동에서 30대 남성이 칼부림을 받고 있다.

21일 경찰에 따르면 이날 오후 2시쯤 신림역 4번 출구 인근에서 피해자가 이송됐다. 연합뉴스



82Cook

<https://www.82cook.com> > entiz > read

## 신림동 칼부림영상 절대보지마세요

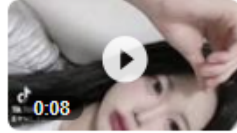
신림동 칼부림영상 절대보지마세요. 무서워 조회수 : 28,497 ... 엄청 잔인하다해서 보지말라고 트라우마 생긴다고해서 안봤습니다 펜타닐이 무서워요 ...



TikTok

<https://www.tiktok.com> > video

## 신림동 칼부림 영상 진짜 보지마세요ππ #신림동 ... - TikTok



좋아요 9.2K개, 댓글 119개. 은혜 (@163\_\_e.u) 님의 TikTok (틱톡) 동영상: "신림동 칼부림 영상 진짜 보지마세요ππ #신림동 #신림동칼부림 #fyp ...

TikTok · 163\_\_e.u · 1주 전



네이트

<https://m.news.nate.com> > view

## '신림동 문지마 칼부림' 영상 확산 우려..."절대 보지 마세요"

2023. 7. 22. — 21일 발생한 '신림동 문지마 칼부림' 폐쇄회로(CC)TV 영상이 확산돼 우려를 낳고 있다. 이를 접한 네티즌들은 "잔인하다", "속이 울렁거린다", "절대 보지 ...

<https://news.nate.com> > view

## "소름 끼치게 잔혹"... '신림동 칼부림' CCTV 영상 보지마세요

2023. 7. 23. — [파이낸셜뉴스] 서울 신림역 인근에서 30대 남성이 행인에게 흉기를 휘둘러 1명이 숨지고 3명이 중상을 입은 가운데 범행 당시의 모습이 담긴 영상이 ...



pikicast.com

<https://issue.pikicast.com> > Uncategorized

## 신림역 칼부림 CCTV 영상 급속 확산...네티즌들 "절대 보지 ...

2023. 7. 21. — 신림역 칼부림 사건 CCTV 영상이 21일 온라인 공간에서 빠르게 확산되고 있다. ... 네티즌들 "절대 보지 마세요" 경악. 위키트리.



위키트리

<https://m.wikitree.co.kr> > articles

## 신림역 칼부림 CCTV 영상 급속 확산...네티즌들 "절대 보지 ...

신림역 칼부림 사건 당시 영상 확산 중, 범행 장면 적나라하게 담겨 있어 논란. ... 칼부림 CCTV 영상 급속 확산...네티즌들 "절대 보지 마세요" 경악.



Daum

주말 약속

mk.co.kr



고성 금이 인터넷에 잇달아 올라오고 있는 가운데 4일 범행 당시 난동 예고글이 올라온 서울 잠실역 주변에 경찰이 출

감돌았다. 전날 모 인터넷 사이트에 "내일 기 때문이다. 경찰 측은 오전 7시 50분부

# 1. Background

최근 발생한 **"문지마 칼부림 사건"**들에 대하여

**CCTV 및 촬영 영상이 무분별하게 유포** 되었고,

**잔혹성**이 높고 **모방 범죄의 우려를 요하는** 영상에 대한

개인의 접근성이 매우 높아졌다



# 1. Background



- 유튜브 숏츠, 인스타그램 릴스, 틱톡 등 개인이 특별한 검수 없이 업로드 할 수 있는 다양한 플랫폼을 통해 **무분별하게 자극적이고 유해하며, 모방성이 높은 영상들이 업로드** 되고 있고,

- 현재 이러한 **유해성을 검수 및 관리하는 과정은 자동화 되어 있지 않다.** 급증하고 있는 영상 콘텐츠의 **유해성을 수작업을 통해 판별하는 것에는 한계가 존재하고 매우 비효율적이다.**



- **영상의 유해성을 판단하고, 검수 및 관리를 자동화하는 모델이 필요하다!**

# HOW?

영상 내 유해성을 판단하는 기준을 어떻게 정해야 할까?

# HOW?

영상 내 유해성을 판단하는 기준을 어떻게 정해야 할까?

## Object Detection?

## Pose Estimation?



## 2. Project Overview

### ◆ Object Detection

- Localization : 객체의 위치 도출
- Classification : 해당 객체의 클래스 도출



### ◆ Limitation

- Threaten 과 Cook처럼 사람과 물체 사이의 상호작용을 판단할 수 없다.
- 즉, 유해성 판단에 활용하기 어렵다.

## 2. Project Overview

### ◆ Pose estimation

- Localization : 객체의 위치 도출
- Pose estimation : 사람의 동작 감지

Punch



### ◆ Limitation

- 샌드백을 때리는 (punch) 행위와 사람을 때리는 (punch) 행위는 동일한 pose 로 인식되나, 행위를 당하는 객체에 따라 그 유해성의 유무가 달라진다.
- 따라서 pose estimation 만으로 유해성을 판단하기에 무리가 있다.

## 2. Project Overview

### ◆ HOI Detection

- <human, object, action> triplets를 탐지한다.
- Human Detection, Object Detection, Object Classification, Interaction Classification  
4개의 task를 포함하고 있다.



<Human, Knife, Cook>

유해성 X



<Human, Knife, Threaten>

유해성 O

## 2. Project Overview

### ◆ HOI Detection의 한계

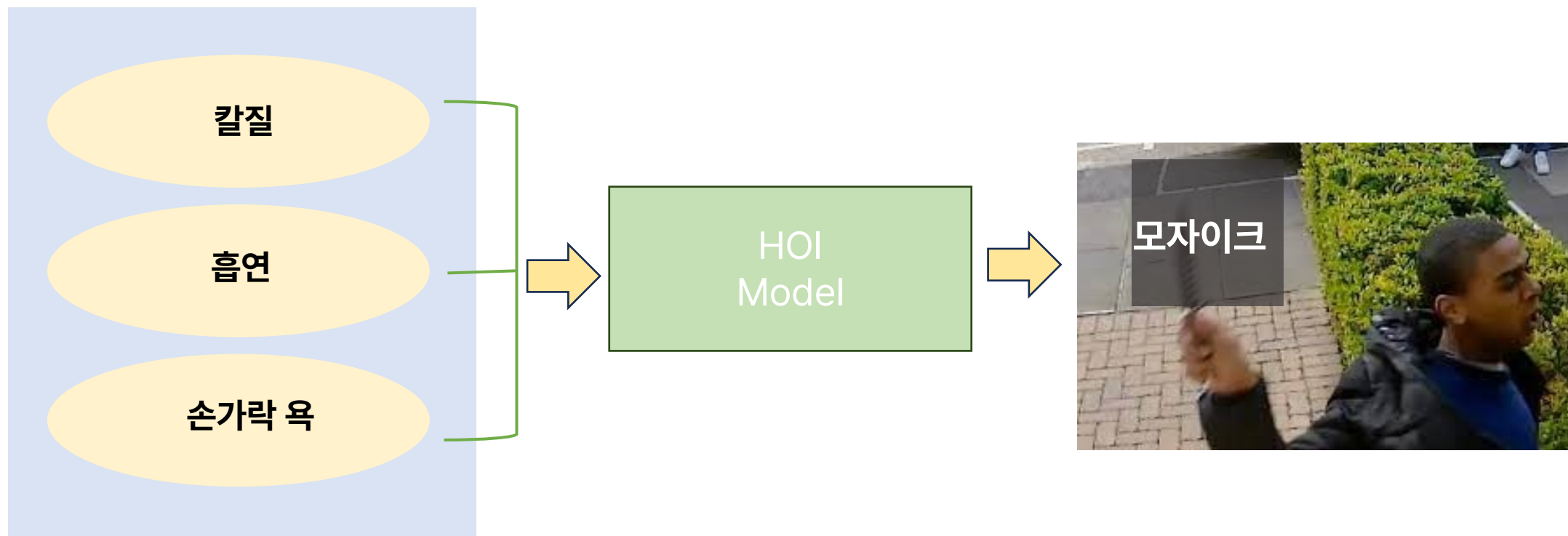
- 폭력은 Human – Human 사이 Interaction을 연속적으로 확인해야 정확도가 높아진다.
- 하지만 HOI 모델에서는 프레임 별 이미지를 기준으로 위험성을 판별하기 때문에 정확한 폭력 판단이 어렵다.



[violation video]

## 2. Project Overview

### ◆ 프로젝트 파이프라인





### 3. Dataset

#### ◆ 수집 데이터셋

- 칼질, 흡연, 손가락욕

※ 기존 HOI 모델에 사용되는 HICO-DET dataset 의 annotation 을 확인해 보니,

※ 흡연 및 손가락 욕에 대한 라벨링이 되어 있지 않아, 직접 dataset을 만드는 과정이 필요할 것이라 판단

#### Hico-det object class

id	object
001	airplane
002	apple
003	backpack
004	banana
005	baseball_bat
006	baseball_glove
007	bear
008	bed
009	bench
010	bicycle
011	bird
012	boat
013	book
014	bottle
015	bowl
016	broccoli
017	bus
018	cake
019	car
020	carrot
021	cat
022	cell_phone
023	chair
024	clock
025	couch
026	cow
027	cup
028	dining_table
029	dog
030	donut
031	elephant
032	fire_hydrant
033	fork
034	frisbee
035	giraffe
036	hair_drier
037	handbag
038	horse
039	hot_dog
040	keyboard
041	kite
042	knife
043	laptop
044	microwave
045	motorcycle
046	mouse
047	orange
048	oven
049	parking_meter
050	person
051	pizza
052	potted_plant

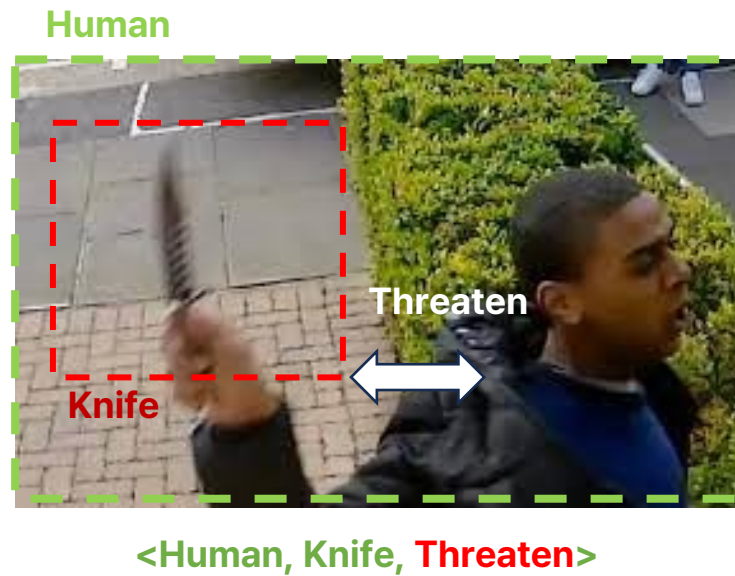
\* "칼질" 에 대한 knife class 는 존재하나, 흡연 및 손가락 욕에 대한 class 는 존재 x

### 3. Dataset

#### ◆ 수집 데이터셋

- 칼질, 흡연, 손가락욕

※ 따라서, 직접 크롤링 후 HOI model 에 적용하기 위해 <Human, Object, Interaction> 에 대한 라벨링 수행



※ 칼질 / 흡연 / 손가락욕 class 에 대하여 각각 약 200 장의 Image를 크롤링 후 labelling tool 활용해 직접 이미지를 <Human, Object, Interaction> set에 맞게 라벨링을 수행해 해당 data를 모델의 input 으로 활용했다.



## 4. Modeling

### ◆ HOTR 모델 소개

- HOI Detection에 처음으로 트랜스포머 구조를 적용한 모델이다.
- DETR 기반으로 인코더와 디코더 구조를 활용하였다.
- 출력은 사람과 객체를 각각 탐지하고 서로간의 상호작용을 분류 예측하도록 구현되었다.

### ◆ 모델의 장점

1. HOI detection 처음으로 트랜스포머 구조를 통해 이미지의 특징을 더 잘 추출해낼 수 있다.
2. 상호작용 정의할 때 수작업 후처리 단계를 제거하면서 속도를 높였다.

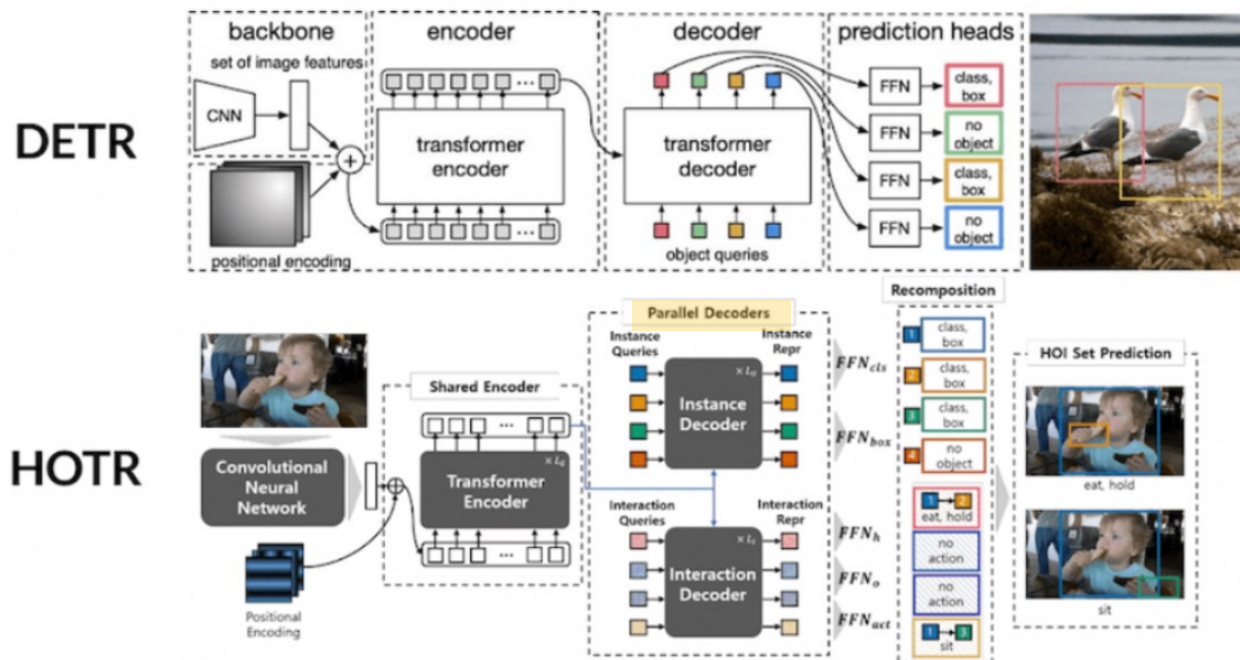
## 4. Modeling

### ◆ HOTR 모델 구조

Instance Decoder는 객체 탐지를 위한 출력을 생성한다.

Interaction Decoder는 상호작용 검출을 위한 출력을 생성한다.

출력은 HO pointer 형태로 최종적으로는 HOI triplets을 생성한다.

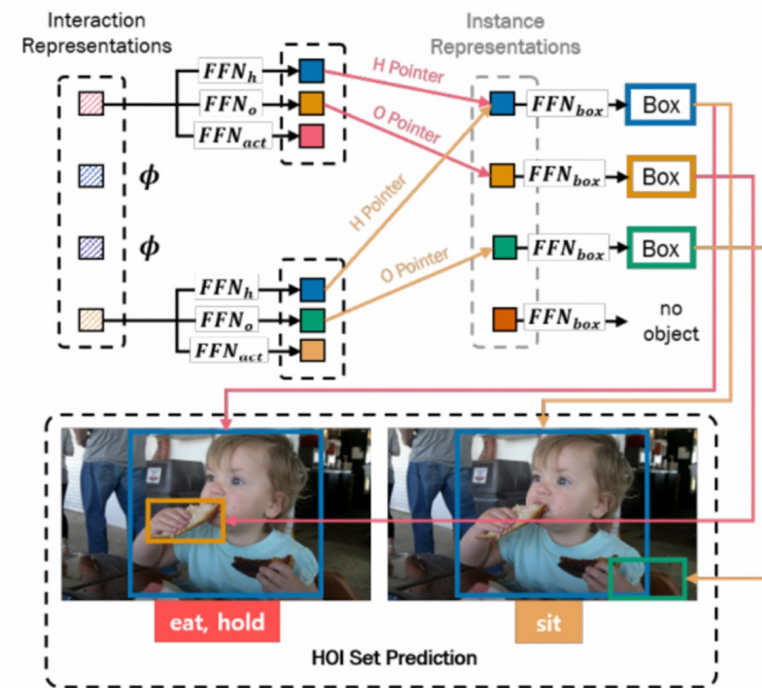


DETR과 HOTR의 Architecture 비교

## 4. Modeling

### ◆ HOTR 모델 구조

1. interaction representation에 3가지 FFN을 적용하여 세 개의 output(human pointer, object pointer, interaction type)을 얻는다.
2. pointer는 모든 instance representation과 유사도를 비교하여 가장 높은 유사도를 가지면 interaction으로 판단하고 localization을 수행한다.
3. 동일한 대상에 대해 중복학습을 할 필요가 없어졌고 기존에 연구된 모델에 비해 Inference 타임을 줄일 수 있었다.



HO pointers와 HOI Set Prediction

## 4. Modeling

### ◆ 커스텀 데이터셋을 활용한 모델의 문제점

※ 직접 수집한 세 가지 class 에 대한 이미지  
(칼질, 흡연, 손가락욕) 에 대해 HOCR 모델을 train

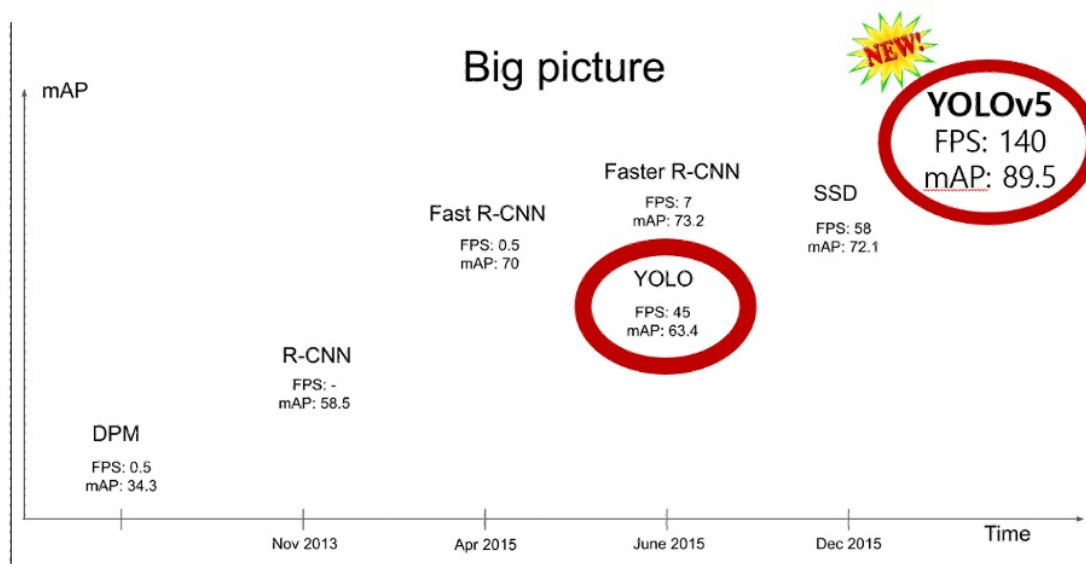
※ 하지만 실제 성능이 좋지 않았고, 목적인 object를  
거의 잡지 못하는 모습을 보임



## 4. Modeling

### ◆ yolov5

1. Bounding box와 classification을 신경망 구조를 통해 한번에 구현한 one stage 모델이다.
2. 높은 성능, 빠른 연산속도, 쉬운 환경 설정의 특징을 가지고 있다.



## 5. Implementation

---

### ◆ HOI Detection 구현

<https://youtu.be/U7idwc3WtNc>

## 5. Implementation

### ◆ 결론

1. HOI Detection Task에 대해 직접 데이터 수집부터 라벨링, 모델 구현까지 경험
2. Custom dataset을 활용한 모델 성능 확인

### ◆ 한계 및 제언

1. 저화질의 영상에서 HOI Detection을 잘 수행하지 못함
2. 데이터셋 부족으로 custom dataset HOCR 모델 구현에 한계
3. 적은 데이터로 고성능을 보이는 모델을 활용하면 프로젝트 목적을 달성할 수 있을 것으로 판단



**END**