



YouHi

이태훈

이주형

김성수

김민재

이인평



INDEX

01. 프로젝트 소개

02. 시스템 구조

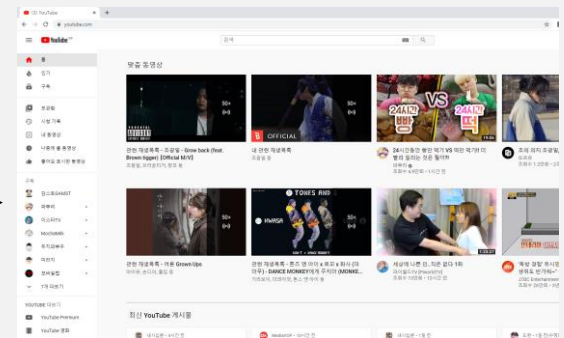
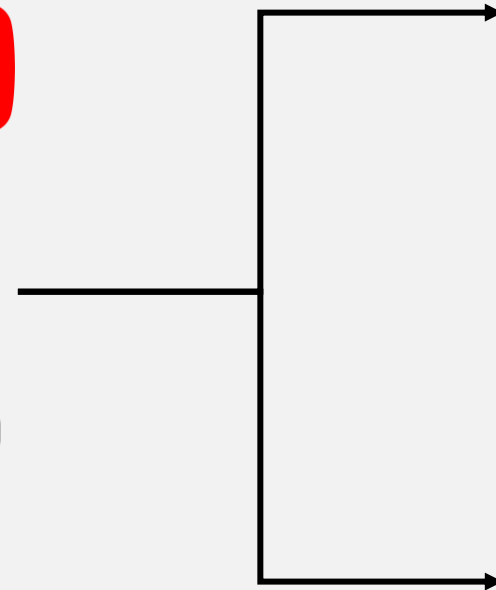
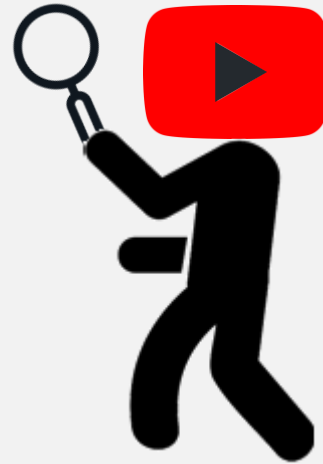
03. 주요 기능

04. 시연 영상



프로젝트 소개

YouHi 프로젝트





YouHi 프로젝트

홈 > AI 및 머신러닝 제품 > Cloud Vision API > 문서

사용해 보기

아래 애플리케이션을 사용하면 이미지 파일의 이미지 주석이 반환됩니다.

- 최대 파일 크기는 4MB입니다.
- 브라우저에서 자바스크립트를 사용 설정해야 합니다.

Try the API


Drag image file here or
Browse from your computer

Google Cloud Vision

YouHi is

YouHi는 영상분류 신경망 모델을 이용한 정확한 필터링을 제공함으로써 업로더의 동영상에서 청소년에게 부적합한 내용을 모두 걸러냅니다. 필터링의 기준은 YouTube 가이드라인의 연령 제한 콘텐츠에 명시되어 있는 연령 제한 적용 고려 사항이며 자세한 내용은 [여기](#)를 참고하세요.

Upload


동영상을 '여기' 끌어다 놓거나
'여기'를 클릭하세요.

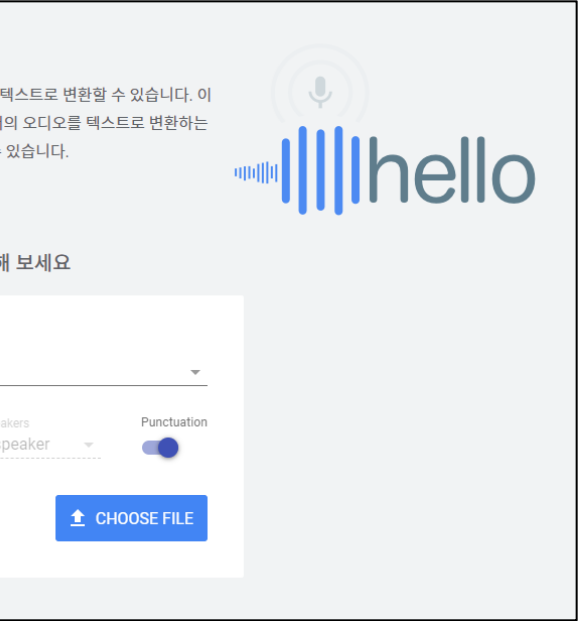
도움말 및 제안사항

- 필터링에 소요되는 시간은 약 10분이며, 사용자의 인터넷 환경에 따라 달라질 수 있습니다.
- 업로드 파일의 형식은 .avi와 .mp4로 제한합니다.
- **연령제한 옵션**을 선택하면 필터링없이 연령제한이 적용된 콘텐츠로 업로드됩니다.

업로드

필터

Service 제공 웹



API - demo



프로젝트 필요성

주요 OTT 월간 순이용자 (단위=명)			
구분	2018년 7월	2019년 7월	증감률 (%)
유튜브	2,466만	2,631만	6.7
넷플릭스	41만	185만	342
틱톡	87만	180만	107
옥수수	313만	336만	7.1
네이버TV	323만	193만	-40
폭	76만	139만	83

※ 자료=닐슨코리아클릭

유튜브 72%에 흡연 장면...청소년 노출 과다

입력 2019.04.22 (12:01) | 수정 2019.04.22 (13:01)

사회

일반 ▾

"선정적·폭력적 영상 잔뜩"...초등생 꿈 '유튜버' 악영향 받을라

2월 4시간 넘게 음란방송 송출 / 유튜브, 신고받고서 뒤늦게 삭제 / 국내법 적용 안 돼 관리 부실 지속 / 유해 콘텐츠 16.5%만 '자체 조치'

뉴스홈 | 최신기사

청소년 79% "폭력·선정적 콘텐츠 봤다"...SNS가 주경로

HOME > 뉴스 > 문화·연예

유해 콘텐츠 풍년인 유튜브, 안 좋은 씩은 잘라낸다

IT

'규제 사각지대' 유튜브...유해콘텐츠 '범람' 못 막나

유튜브로 음란물 보는 아이들 늘는데...해외 사이트, 규제 수단 없나

| 무분별한 유해 콘텐츠 계속 증가...정부 기관 "어쩔 수 없어...일일이 규제하기 쉽지 않아"

[토요워치] 유해 콘텐츠 천국...유튜브 '빨간맛' 방패 막이 없다

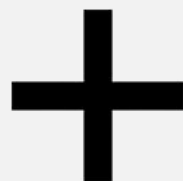
[마약중독자의 고백⑪] 마약 호기심 자극하는 유튜브 '빨간불'



프로젝트 필요성



운영자의
모니터링



YouTube
영상 자동 검열 장치



시청자의 신고



프로젝트 목표 - YouTube 연령 제한 기준

Video



폭력적이거나 노골적인 콘텐츠

Knife, Blood



유해하거나 위험한 콘텐츠

Smoke



과도한 노출 혹은 성적인 콘텐츠

Adult



Voice, Subtitle

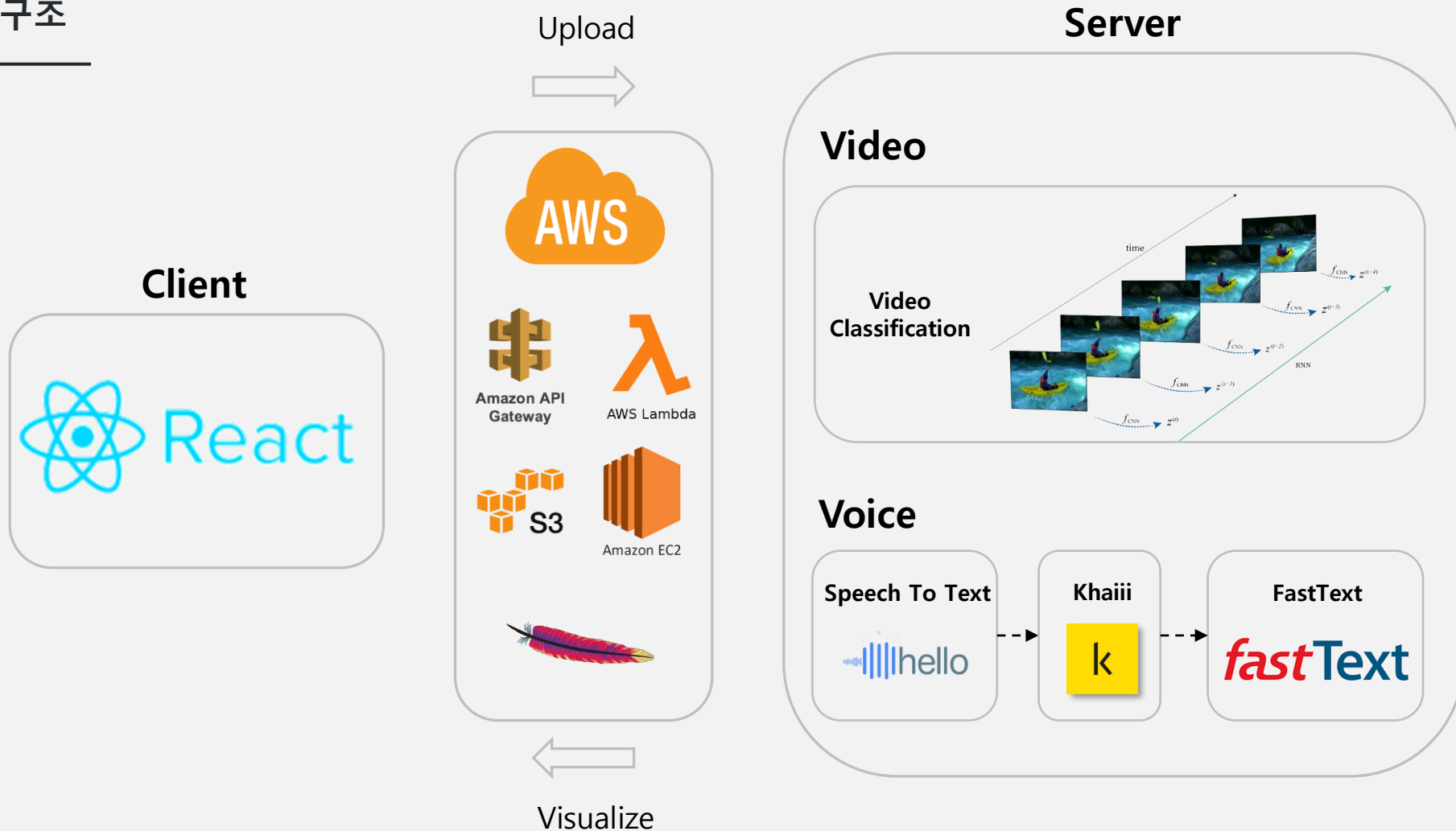


저속한 언어





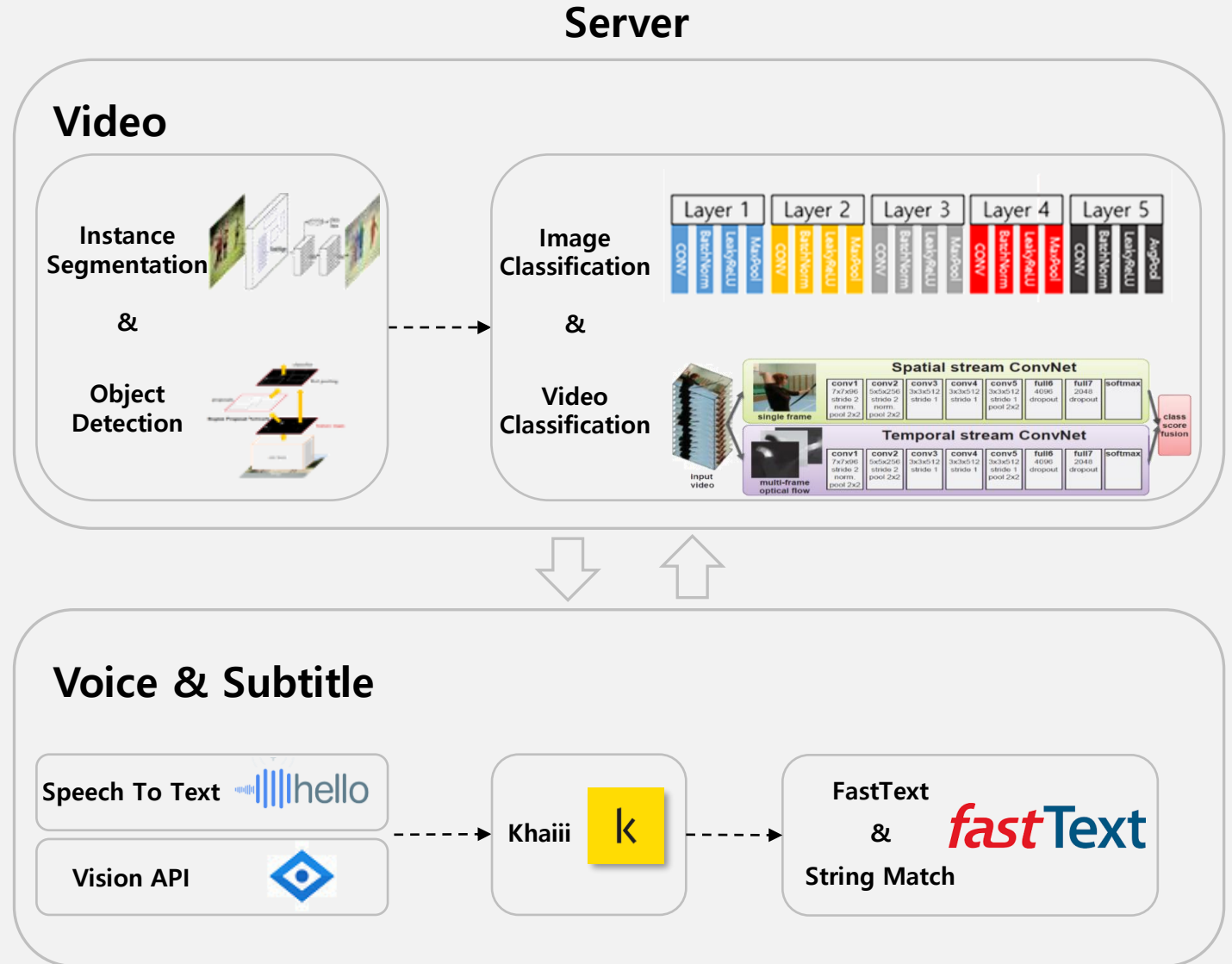
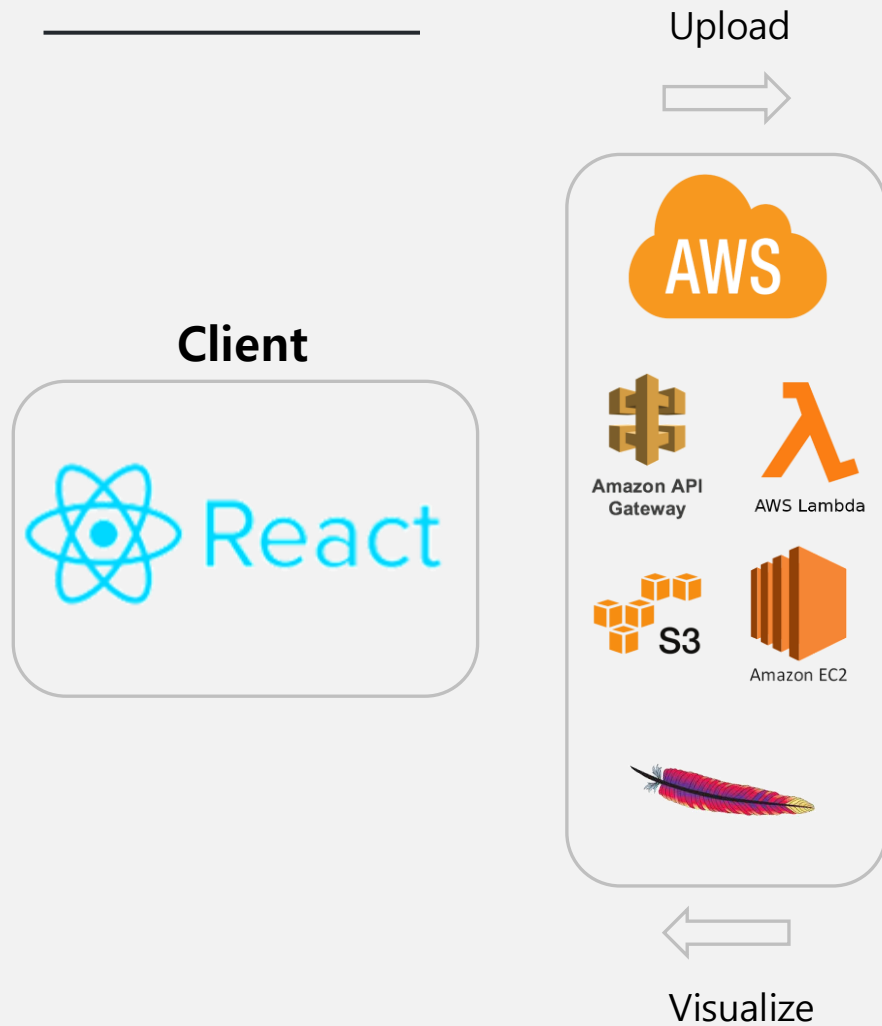
초기 시스템 구조





시스템 구조

최종 시스템 구조





Instance Segmentation & Object Detection



Knife



Smoke



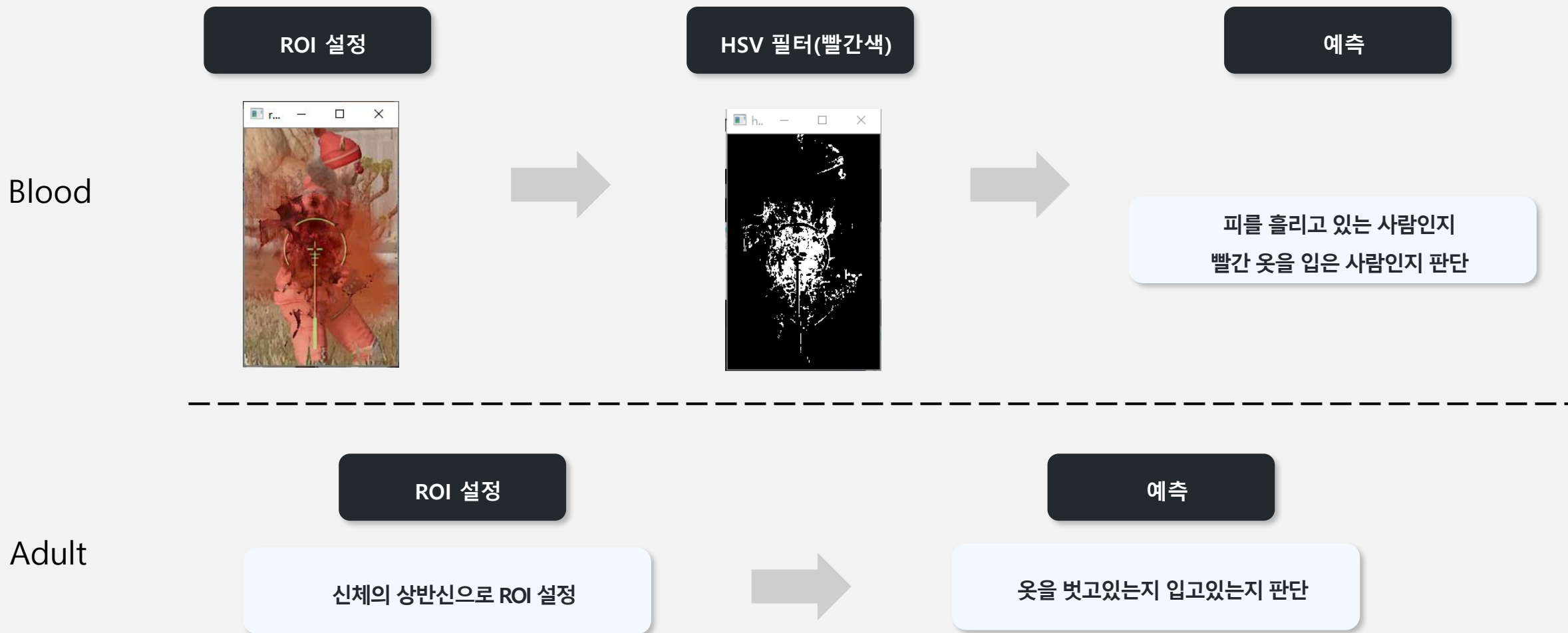
Gun



Blood

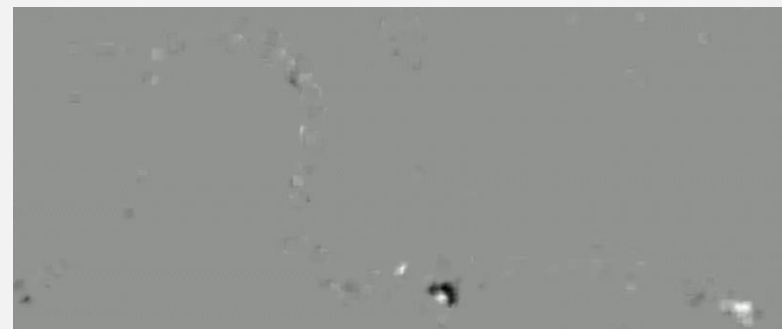
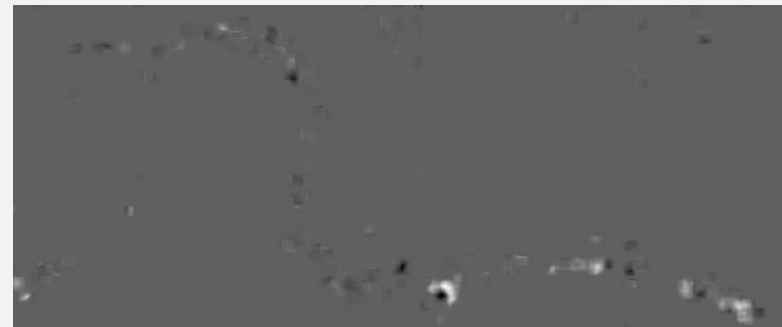


Image Classification





Video Classification



spatial stream

temporal stream

Class Score Fusion
(가중 평균값)



음성 및 자막

텍스트 변환

좋아요
구독
부탁드리겠습니다
안녕하세요
신대리입니다
다들
반갑습니다
오늘은
우리나라
대한민국을
욕을
했어
어떤
일본인비제이

형태소 변환

k

좋/VA 아요/EC
구/MAG 독/NNG
부탁/NNG 드리/VV 겠/EP 습니다/EC
안녕/NNG 하/XSA 세요/EC
신대리/NNG 이/VCP ㅂ니다/EC
다/MAG 들/XSN
반갑/VA 습니다/EC
오늘/NNG 은/JX
우리나라/NNG

옥설 검열

fastText

좋 0.07257053
아 요 0.4825888
구 0.018253189
독 0.26963842
부 탁 0.3488102
드 리 0.27959976
겠 0.04729164
습 니 다 0.40412754
새 끼 0.99999994
를 0.21808407
씨 발 새 끼 0.65116084
십 안 0.33072433

if (cosine similarity > 0.6):
String Match 진행



음성 및 자막

씨발새끼 / 0 : 21 / 0 : 22
새끼 / 0 : 24 / 0 : 24
씨발새끼야 / 0 : 27 / 0 : 28
개새끼야 / 0 : 28 / 0 : 29
 좇 / 0 : 29 / 0 : 29

Original audio



Beep sound

Voice

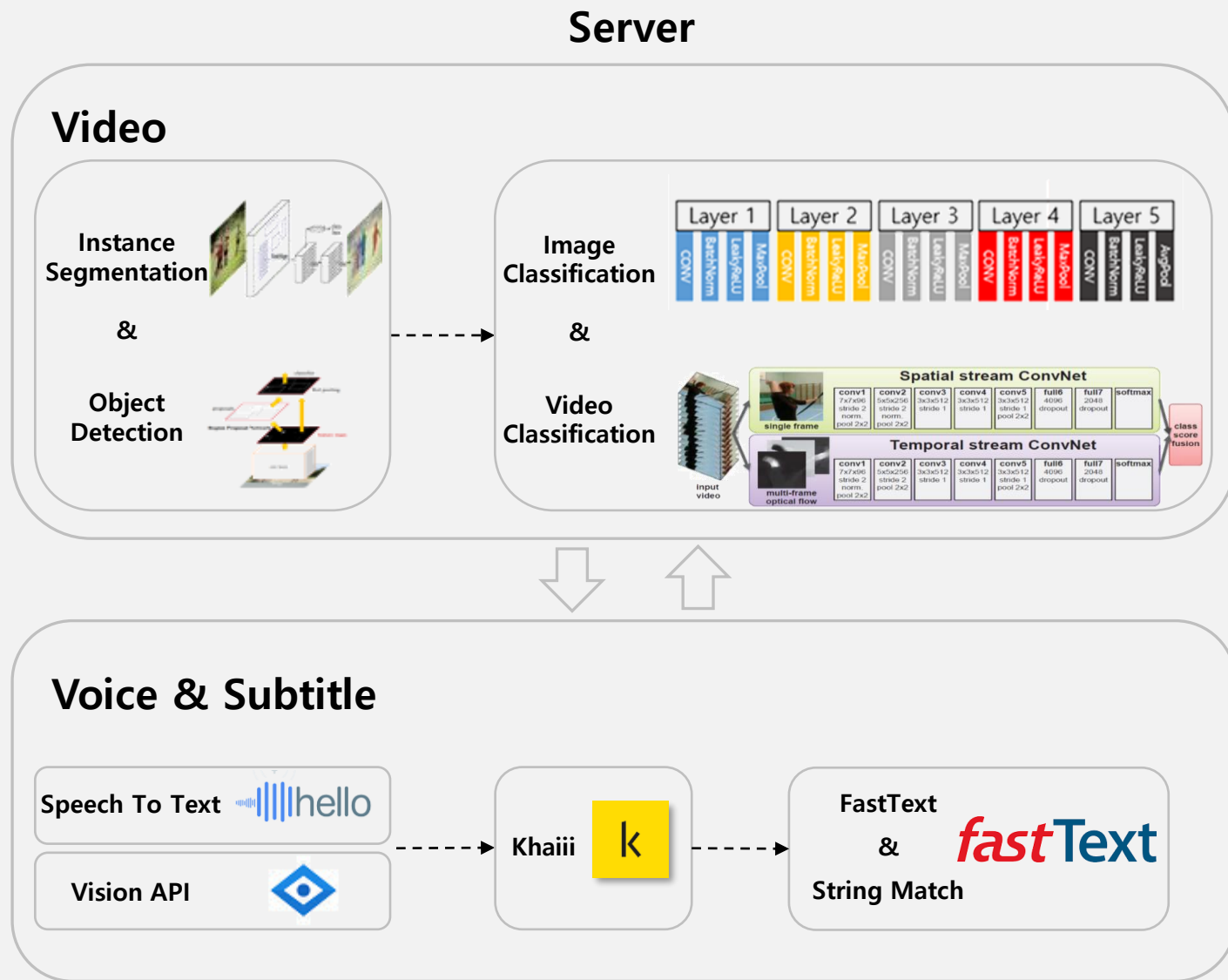


하 시발 이게 뭐야!



이 새끼는 쿵도 신경 안쓰고, 저 놈은 꿈이나 꾸고 앉아있고,
저 새끼는 타르타르 캐비아 화이트초콜릿 쓰레기나 만지작 거리고 있다

Subtitle





정확도 측정

True	영상 20개(각 라벨 당 영상 5개)
False	영상 30개

		실제 정답	
		True	False
분류 결과	True	14	2
	False	6	28

- True Positive(TP) : 실제 True인 정답을 True라고 예측 (정답)
- False Positive(FP) : 실제 False인 정답을 True라고 예측 (오답)
- False Negative(FN) : 실제 True인 정답을 False라고 예측 (오답)
- True Negative(TN) : 실제 False인 정답을 False라고 예측 (정답)

우리 프로젝트에 도입하게 되면 True는 유해한 것, False는 유해하지 않는 것

- **Precision(정밀도)** : 모델이 True라고 분류한 것 중에서 실제 True인 것의 비율

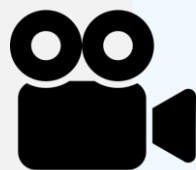
$$(Precision) = \frac{14}{14 + 2} = 87.5\%$$

- **Recall(재현율)** : 실제 True인 것 중에서 모델이 True라고 예측한 것의 비율

$$(Recall) = \frac{14}{14 + 6} = 70\%$$



주요 기능



영상 검열



자막 검열



음성 검열



이의 신청

Filter



이의신청 및 문의하기

안내

콘텐츠 규정

Team YouHi

Kookmin University Computer Science



Thank you.

들어주셔서 감사합니다.