



국민대학교
전자정보통신대학
컴퓨터공학부

캡스톤 디자인 I

종합설계 프로젝트

프로젝트 명	동영상 연령제한 검열
팀 명	YouHi
문서 제목	중간보고서

Version	1.5
Date	2020-04-22

팀원	이태훈 (조장)
	이인평
	이주형
	김성수
	김민재
지도교수	임은진 교수

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	동영상 연령제한 필터링	
	팀 명	YouHi	
	Confidential Restricted	Version 1.5	2020-APR-22


CONFIDENTIALITY/SECURITY WARNING

이 문서에 포함되어 있는 정보는 국민대학교 전자정보통신대학 컴퓨터공학부 및 컴퓨터공학부 개설 교과목 캡스톤 디자인I 수강 학생 중 프로젝트 "You Hi"를 수행하는 팀 " You Hi "의 팀원들의 자산입니다. 국민대학교 컴퓨터공학부 및 팀 " You Hi "의 팀원들의 서면 허락없이 사용되거나, 재가공 될 수 없습니다.

문서 정보 / 수정 내역


Filename	중간보고서-동영상 연령제한 검열.doc
원안작성자	이태훈
수정작성자	이인평, 이주형, 김성수, 김민재

수정날짜	대표수정자	Revision	추가/수정 항목	내 용
2020-04-02	이태훈	1.0	최초 작성	
2020-04-10	김성수	1.1	내용 추가	WEB 관련 연구 내용 추가
2020-04-12	이인평	1.2	내용 추가	이미지 분류 관련 내용 추가
2020-04-14	김민재	1.3	내용 수정	오탈자 및 비문 수정
2020-04-17	이주형	1.4	내용 추가	음성 인식 및 STT 내용 추가
2020-04-22	김민재	1.5	내용 수정	프로젝트 목표 수정

	국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서	
		프로젝트 명	동영상 연령제한 필터링
		팀 명	YouHi
		Confidential Restricted	Version 1.5 2020-APR-22

목 차


1	프로젝트 목표	4
1.1	목표	4
1.2	목표에 따른 기대효과	5
2	수행 내용 및 중간결과	6
2.1	계획서 상의 연구내용	6
2.1.1	객체 탐지	6
2.1.2	영상 검열	7
2.1.3	음성 인식	9
2.1.4	형태소 분석	10
2.1.5	욕설 검열	11
2.1.6	Web Front End	13
2.1.7	Back End	14
2.2	수행내용	14
2.2.1	객체 탐지	14
2.2.2	영상 검열	15
2.2.3	음성 추출 및 분할	16
2.2.4	Speech To Text	16
2.2.5	형태소 분석	17
2.2.6	욕설 검열	18
2.2.7	Web Front End	20
2.2.8	Back End	22
3	수정된 연구내용 및 추진 방향	25
3.1	수정사항	25
3.1.1	동영상 업로드 방식 변경	25
3.1.2	영상 검열 방식 변경	26
3.1.3	웹페이지 설계 및 UI 변경	27
4	향후 추진계획	27
4.1	향후 계획의 세부 내용	27
4.1.1	영상검열	27
4.1.2	정확도 측정	27
4.1.3	Front End	28
4.1.4	Back End	28
5	고충 및 건의사항	29

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	동영상 연령제한 필터링	
	팀 명	YouHi	
	Confidential Restricted	Version 1.5	2020-APR-22

1 프로젝트 목표

1.1 목표


- YouTube 및 실시간 방송 플랫폼, 방송통신위원회 등의 유해 기준을 참고하여 동영상에 미성년자 시청 불가능 장면/욕설이 있는지 판단한다.
- 현재 YouTube <노란 딱지> 정책 기준의 모호성과 불공정성에 대한 문제를 해결하기 위해 사용자가 직접 자신의 영상 중 어떤 부분이 부적절한지 확인이 가능하게 한다. 영상 제작자(크리에이터)들의 영상 제작의 효율성을 높인다.(<노란 딱지>는 YouTube에 존재하는 기존 AI 영상 등급 검열 장치이다. 최근, <노란 딱지> 시스템의 필터링 기준이 모호하여 동일한 내용의 영상들 사이에서도 유해 영상 판정을 받는 영상과 그렇지 않은 영상들이 존재하며 영상 제작자가 자신의 영상 내용 중 유해 판정을 받은 부분을 직접 확인할 수 가 없어 사용자들 간 많은 불만을 사고 있다.)
- 동영상에 미성년자가 시청 불가능한 장면이 있는지 판단되면 해당 장면이 어떤 가이드라인을 위반했는지 알려준다.
- 현재 YouTube에 존재하는 다양한 가이드라인 중 선정성, 폭력성, 모방성(흡연, 욕설 등)에 대한 가이드라인을 충족시키는 지 중점적으로 확인하는 검열을 실시한다.
 - 1) "사람이 칼에 찔리는 장면이 명확하게 표현된 경우" - 폭력성
 - 2) "흡연하는 장면이 명확하게 표현된 경우" - 모방성
 - 3) "지나치게 특정 대상을 비방하거나 과도한 욕설이 포함된 경우" - 모방성
 - 4) "여성 혹은 남성의 노출 강도가 심한 경우" - 선정성
- 본 프로젝트의 최종 목표는 다양한 영상 플랫폼에서 이 시스템을 사용할 수 있도록 정확도를 높이고, 검열 과정에 너무 긴 시간이 소요되지 않도록 개선하는 것이다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	동영상 연령제한 필터링	
	팀 명	YouHi	
	Confidential Restricted	Version 1.5	2020-APR-22

1.2 목표에 따른 기대효과

검열 시스템은 Youtube 의 가이드 라인에 맞춰서 영상과 음성을 검열한다. 이는 현재 신고 기반으로 이루어지는 수작업 검열 과정보다 효율적으로 작동된다. 업로드 이전에 일어나는 자동 검열 시스템이므로 다양한 기대효과와 활용방안이 있다.

1. 검열 시스템을 적용함으로써 청소년에게 부적합한 영상들을 일차적으로 검열할 수 있다. 이는 동영상 업로드 플랫폼에서 실시하는 신고 기반 검열 시스템보다 더 많은 영상을 검열할 수 있고, 작업량과 비용 측면에서 효율적이다.
2. 추후 Youtube 이외의 다양한 실시간 스트리밍 서비스(Twitch, Affreca TV 등)에서도 효과적으로 사용할 수 있다. 각 스트리밍 서비스들은 운영진이 직접 실시간 모니터링과 시청자들의 신고를 통해 제재를 가하는데, 모든 과정이 수동 작업으로 이루어지기 때문에 효과적으로 이루어지지 않는다. 따라서 방송되고 있는 장면들을 실시간으로 딥러닝 모델에 넣어 검열할 수 있다.
4. 현재는 미성년자들을 대상으로 부적합한 영상들을 검열하는 시스템을 목표로 하지만, 데이터셋이 많아지고 다양한 라벨에 대해서 학습을 시킨다면, 교통사고, 자연 재해, 길거리 싸움, 집단 구타 및 학대 등 검열 대상을 확대시켜 활용이 가능하다. 또한 유해한 동영상의 업로드를 제한함으로써 사이버 범죄 예방에 도움을 준다.
5. 현 음성 검열 시스템은 단순 욕설 단어에 대해서만 진행되지만, 데이터셋을 확대시켜 학습시킨다면 욕설을 포함하지 않은 비방 목적의 문장 또한 검열이 가능하다.
6. 이 시스템이 적극적으로 활용된다면, 동영상 업로드 플랫폼에 대한 사람들의 신뢰도와 인식의 향상에 도움이 된다.

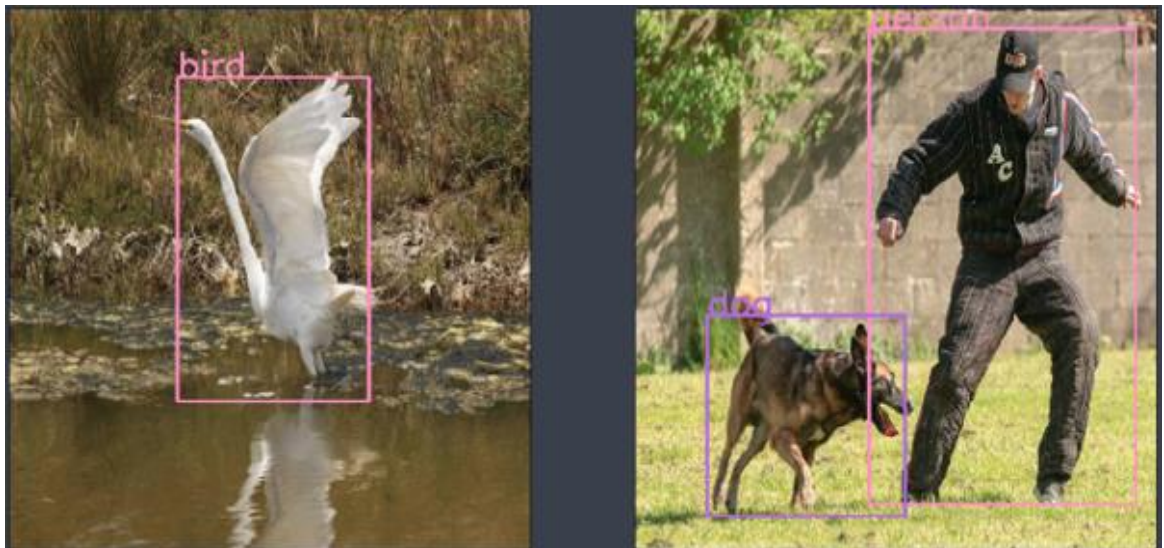
 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	동영상 연령제한 필터링	
	팀 명	YouHi	
	Confidential Restricted	Version 1.5	2020-APR-22

2 수행 내용 및 중간결과

2.1 계획서 상의 연구내용


2.1.1 객체 탐지

객체 탐지는 이미지에서 찾고 싶은 관심 객체를 배경과 구분해 식별하는 자동화 기법이다. 이 기술은 아래의 사진처럼 이미지에서 새와 사람 등의 객체를 자동으로 탐지해낸다. 이러한 객체 탐지는 딥러닝을 통해 이루어진다.

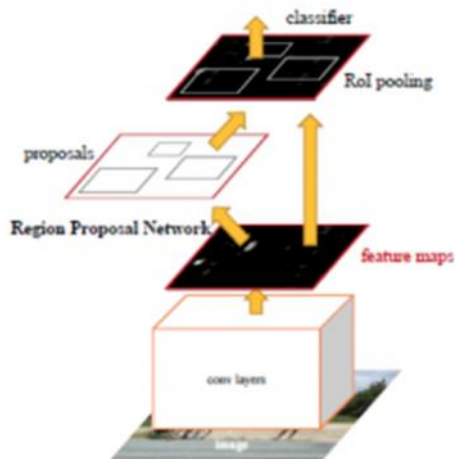


객체 탐지 기술은 2개 이상, 즉 N개의 객체를 탐지해 분류할 수 있어야 한다. 많은 객체를 탐지하는 데 한계가 있으므로 다수의 사각형 상자 위치와 크기를 가정해 컨볼루션 신경망을 변형한 후 이를 객체 분류(Object Classification)에 활용한다. 이러한 사각형 상자를 '윈도우(Window)'라고 부른다. 각 창의 크기와 위치는 객체의 존재 여부에 따라 결정될 수 있고 객체가 있는 경우에는 그 범주도 결정할 수 있다. 본 프로젝트에서는 다음과 같은 알고리즘을 사용한다.

2단계 방식의 객체 탐지 알고리즘, Faster RCNN 알고리즘 이름에 '빠른(Faster)'이라는 단어가 포함되어 있지만 이는 단일 단계 방식보다 빠른 처리가 된다는 뜻이 아닌 이전 버전이라 할 수 있는 RCNN 알고리즘과 Fast RCNN 알고리즘 보다 빠르다는 것을 뜻한다. 각 관심 영역(RoI; Region of Interest)에 대한 피쳐 추출의 계산을 공유하고 딥러닝 기반의

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	동영상 연령제한 필터링	
	팀 명	YouHi	
	Confidential Restricted	Version 1.5	2020-APR-22

RPN을 도입해 구현한다.



많은 CNN 레이어를 사용해 피쳐 맵 추출이 완료되면 RPN을 통해 개체를 포함하고 있을 가능성이 높은 윈도우가 다량으로 생성된다. 이 후 각 윈도우에 있는 피쳐 맵을 검색하고, 고정 크기로 조정된 뒤(RoI 풀링) 클래스 확률과 해당 객체에 대한 더 정확한 경계박스를 예측한다. RPN은 YOLO와 같은 방식의 앵커 박스를 사용한다. 하지만 YOLO 알고리즘과는 앵커 박스가 데이터로부터 생성되는 것이 아니라 고정된 크기와 형태로 생성된다는 차이가 있다. 이 앵커 박스는 이미지를 보다 조밀하게 커버할 수 있다. RPN은 여러 객체 카테고리들에 대한 분류 대신 윈도우의 객체 포함 유무에 대한 이진 분류(Binary Classification)만 수행한다.

2.1.2 영상 검열

영상 검열은 Two Stream Convolution Network 모델로 진행한다.

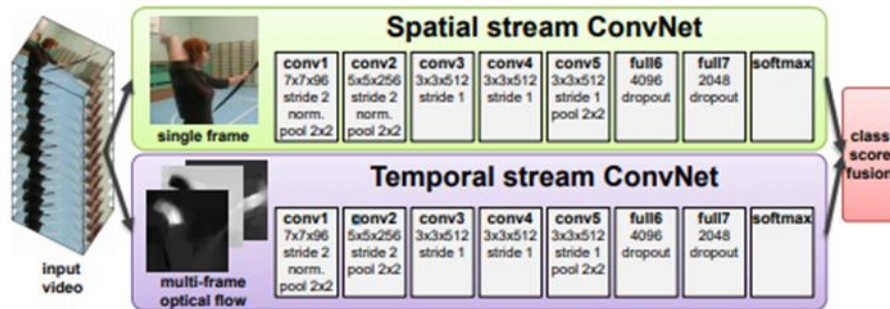



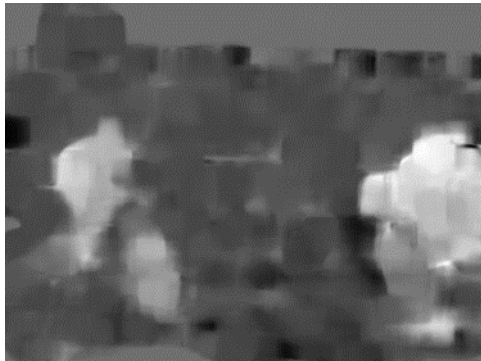
Figure 1: Two-stream architecture for video classification.

위 그림처럼 Spatial Stream ConvNet과 Temporal Stream ConvNet으로 각각 예측된 결과를 Average, Conv Fusion을 통해 결합한다. 먼저 Spatial Stream ConvNet는, 하나의 동영상 프레임에서 작동되며 해당 프레임에서 사람의 Action을 인식한다. 즉, 정적 이미지에서 결과를 예측하는 것이다. 반면 Temporal Stream ConvNet은 일정 길이의 이미지(프레임)를 취합한 내용을 통해 예측하는 것이다. Optical Flow ConvNet이라고도 불리는데, 그 이유는 RGB 이미지가 아닌 Optical Flow로 표현된 이미지로 학습과 예측을 진행하기 때문이다. Optical Flow는 Vertical Flow와 Horizontal Flow로 구성되어있다. 따라서 학습과 예측을 진행하기 전에 RGB이미지를 Vertical Flow, Horizontal Flow로 표현해야한다. Vertical Flow, Horizontal Flow는 다음과 같이 표현된다.

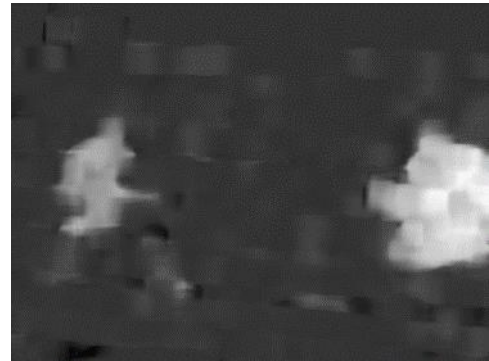


<Original Image>

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	동영상 연령제한 필터링	
	팀 명	YouHi	
	Confidential Restricted	Version 1.5	2020-APR-22



<Horizontal>

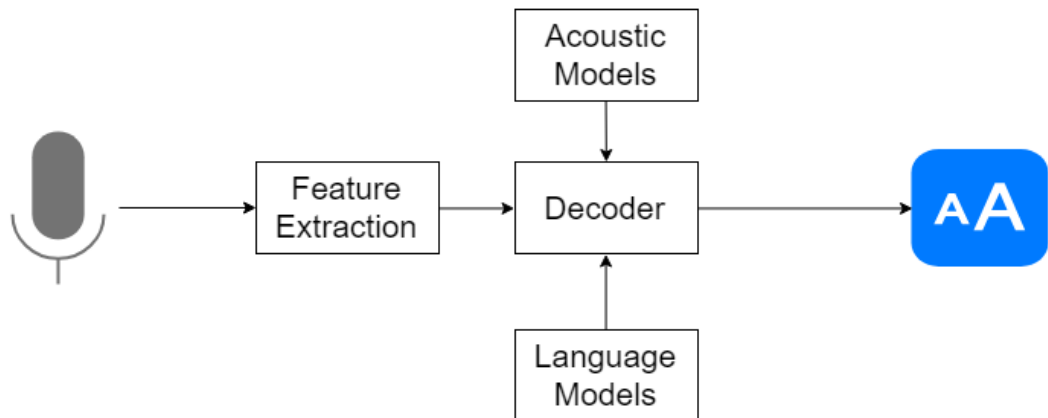


<Vertical>

이처럼 표현된 연속된 이미지들에서 행동을 추적하고 어떤 행동인지 예측한다. 그러나 이 두 가지 모델은 각각 단점이 있는데, Spatial의 경우 연속된 이미지를 고려하지 않다보니 단순 이미지 Classification의 결과와 크게 다르지 않다라는 것이다. 또한 Temporal의 경우 유사한 행동 패턴이 들어왔을 때 예측이 틀린 경우가 발생할 가능성이 높다. 따라서 Temporal과 Spatial을 Fusion을 진행하면 각각의 단점을 상호보완 해주기 때문에, 높은 정확도를 얻을 수 있다.

2.1.3 음성 인식

딥러닝 기반의 음성인식 기술은 크게 언어모델과 음향모델이라는 두 가지 중요한 지식원 (Knowledge source)을 사용해 음성 신호로부터 문자 정보를 출력한다. 언어모델은 단어 시퀀스에 확률을 할당(assign) 하는 일을 하는 모델이다. 이는 가장 자연스러운 단어 시퀀스를 찾는 역할을 한다. 음향모델은 언어의 소리단위를 딥러닝을 통해 학습하여 어떤 단어가 어떤 소리로 나는지를 확률적으로 변환한다. 이 두가지 모델이 동시에 작용하여 높은 확률을 보인 단어를 출력한다.



STT(Speech To Text) 는 잡음처리와 특징 추출과정을 거친 음성 데이터가 언어모델과 음향 모델이 결합된 디코더를 통과한 후 문장 형태로 출력되면 이를 텍스트로 저장한다.

2.1.4 형태소 분석


형태소 분석기는 텍스트 형태의 데이터를 형태소 단위로 분리한다. 형태소 단위로 분리된 데이터는 비속어 여부를 판단하기에 유리하다.

카카오에서 개발한 세 번째 형태소 분석기인 kharii는 형태소 분석 시 입력된 각 음절에 대해 하나의 출력 태그를 결정하는 분류 문제로 접근하게 된다. 일정 텍스트의 형태소 분석 결과는 다음 이미지와 같이 생성된다.

```

kmusw@kmusw-ThinkPad-T440: ~/바탕화면/kharii/build
kmusw@kmusw-ThinkPad-T440:~/바탕화면/kharii/build$ kharii --input input.txt
[2020-04-03 18:03:30.567] [Resource] [info] NN model loaded
[2020-04-03 18:03:30.567] [Preanal] [info] preanal dictionary opened
[2020-04-03 18:03:30.567] [ErrPatch] [info] errpatch dictionary opened
[2020-04-03 18:03:30.567] [Restore] [info] restore dictionary opened
[2020-04-03 18:03:30.567] [Resource] [info] PoS tagger opened
중 /VA + 있/EP + 던/ETM
기억 /NNG + 만/JK
그림 /VA + 는/ETM
마음 /NNG + 만/JK
나 /NP + 가/JKS
떠나 /V + 가/JKS
그 /MM
길 /NNG
에 /NNG + 에/JKB
이름 /VA + 게/EC
남 /VV + 아/EC
서 /VV + 어어/EC + 있/VX + 다/EC
인 /VV + 어/EC + 지/VX + 르/ETM
크 /NNG + 만/JK
할 /VA + 을/ETM
/NNB + 만/JK
/NNG
/VA + 고/EC
기다 /VV + 는/ETM
/NNG + 리/VV + □/ETN
/NNG + 에/JKB
나 /NP + 를/JKO
우리 /VV + 리라/EC
  
```

kharii는 기계학습 기반의 알고리즘을 이용하여 형태소를 분석한다. 형태소 분석은 자연어

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	동영상 연령제한 필터링	
	팀 명	YouHi	
	Confidential Restricted	Version 1.5	2020-APR-22

처리를 위한 가장 기본적인 전처리 과정이기 때문에 빠른 시간 내에 이루어져야 하므로 신경망 알고리즘들 중 빠른 속도로 진행되는 Convolutional Neural Network(CNN)을 사용하였다.

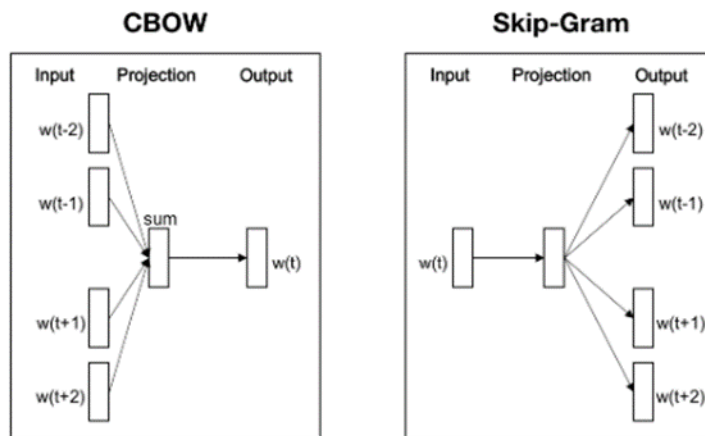
khaiii는 신경망 알고리즘의 앞단에 기분석 사전, 뒷단에 오분석 패치라는 두 가지 사용자 사전 장치를 제공한다. 기분석 사전은 단일 어절에 대해 문맥에 상관없이 일괄적인 분석 결과를 갖게 하고 싶을 경우 사용하고 오분석 패치는 여러 어절에 걸쳐 충분한 문맥과 함께 오분석을 바로잡아야 할 경우에 활용한다. 이러한 두 기능으로 형태소 오분석 확률을 줄일 수 있고 본 프로젝트에 맞게 형태소 분석을 진행할 수 있다.

2.1.5 욕설 검열

khaiii 형태소 분석기로 추출된 형태소들을 사전 훈련된 FastText 모델을 이용하여 욕설 데이터 리스트의 원소와 cosine similarity를 계산한다.

FastText는 단어를 벡터로 만드는 Word2Vec과 비슷한 매커니즘을 가진다. 하지만 Word2Vec은 어휘를 최소단위로 보는 반면에 FastText는 텍스트의 최소 단위를 어휘를 구성하는 글자 n-gram으로 하였다. 즉 Word2Vec은 임베딩 벡터를 어휘마다 하나씩 할당하고 이를 학습했다면, FastText는 어휘를 구성하고 있는 n-gram마다 하나씩 이를 할당하고, 어휘를 구성하는 모든 n-gram 벡터의 평균 벡터를 어휘 임베딩으로 보게 된다. 이러한 방법은 동일한 텍스트 데이터에서 더 많은 정보를 활용하기 때문에, 더 적은 양의 학습 데이터로도 높은 성능을 낼 수 있다.

FastText는 corpus 안에 존재하는 모든 단어를 지정해 놓은 윈도우 크기로 슬라이딩 하여 학습을 진행한다. 그리고 임베딩 기법에 따라 주변 단어를 보고 중심 단어가 무엇인지를 예측하는 CBOW(continuous bag-of-words) 모델과 중심 단어를 보고 어떤 주변 단어가 등장했는지를 맞추는 Skip-Gram 모델로 나눈다.




만약 윈도우의 크기를 2라고 가정하였을 때, CBOW 모델의 경우 중심 단어의 벡터는 주변 단어로부터 단 한번의 업데이트 기회를 갖게 된다. 하지만 Skip-gram 모델의 경우 중심 단어 벡터를 주변 단어 개수만큼, 즉 네 번을 업데이트할 수 있게 한다. 따라서 같은 corpus 크기를 갖더라도 학습량이 네 배 차이가 나게 되어 CBOW 모델에 비해 Skip-Gram 모델이 더 높은 성능을 보인다. 따라서 본 프로젝트에서는 FastText Skip-Gram 모델 방식을 채택하였다.

FastText 모델 학습 시 dictionary에 관한 주요 argument는 다음과 같다.

- 1) minCount : 주어진 값 이상으로 등장한 단어들만 임베딩을 실시할 수 있다.
- 2) wordNgrams : 위에서 말한 n-grams의 의미가 아닌 중심 어휘와 주변 어휘를 몇 개의 단어로 설정할 지에 대한 옵션이다.
- 3) bucket : 모델의 메모리 사용을 제한하기 위해 설정하는 n-grams이 hash된 bucket의 수이다.

FastText 모델 학습의 hyper parameter와 관련된 주요 argument는 다음과 같다.

- 1) lr : FastText 모델 학습 알고리즘의 learning rate를 설정한다. 0.1 ~ 1.0 사이의 값을 갖는 것이 좋다고 알려져 있다.
- 2) dim : 임베딩되는 단어 벡터들의 차원 수를 설정할 수 있다. dim이 큰 값을 가질수록,

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	동영상 연령제한 필터링	
	팀 명	YouHi	
	Confidential Restricted	Version 1.5	2020-APR-22

벡터들은 더 많은 정보를 가질 수 있지만 그만큼 더 많은 데이터를 학습해야 한다. 하지만 이 값이 너무 크게 되면, 학습 과정이 어려워지고 시간이 오래 걸린다. default 값으로 100을 가지지만 100부터 300사이의 값을 많이 이용한다.

3) ws : FastText 모델 학습 시 사용되는 window size의 크기를 설정할 수 있다.

4) epoch : FastText 모델 학습 시 수행하는 epoch 개수를 설정 가능하다.

5) loss : FastText 모델 학습 시 사용되는 loss function을 설정할 수 있다. 설정할수 있는 값으로 negative sampling을 뜻하는 ns, softmax의 softmax, hierarchical softmax을 뜻하는 hs가 있다.

FastText의 결과인 어휘 벡터들은 서로 의미가 유사할수록 그들이 갖는 방향 또한 같아진다. 이러한 점을 이용하여 욕설 검열 과정에서 형태소와 욕설의 유사도를 판단하기 위해 내적 공간의 두 벡터간 각도의 cosine값을 이용하여 측정된 벡터 간의 유사한 정도를 의미하는 cosine similarity를 이용하기로 했다.


2.1.6 Web Front End

React는 페이스북에서 개발한 유저인터페이스 라이브러리로 개발에게 재사용 가능한 UI를 생성 할 수 있도록 해준다. 이 라이브러리는 현재 페이스북, 인스타그램, 야후, 넷플릭스를 포함한 많은 서비스에서 사용되고 있다.

Web Page 특성상 기능별 컴포넌트들이 필요하고, 컴포넌트끼리의 연결도 용이해야한다. React Js는 컴포넌트 별 관리가 편하고, 디버깅 및 코드 수정이 비교적 간단하기 때문에 React Js로 구현을 진행한다.

React Js의 Rendering은 기존 HTML 문법과 상당히 유사하기 때문에, 많은 개발자에게 익숙하여 협업을 진행하기에 적합하다.

웹 구현에 있어 가장 중요한 것은 기능과 디자인이다. react는 다양한 라이브러리를 제공하는데 개발자는 이 라이브러리들을 쉽게 다운받아 쓸 수 있고 자체 커스텀하기도 편리하기 때문에 react js로 구현을 진행한다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	동영상 연령제한 필터링	
	팀 명	YouHi	
	Confidential Restricted	Version 1.5	2020-APR-22

2.1.7 Back End

딥러닝 모델 학습과 모델의 영상, 음성 검열 과정에 필요한 Amazon EC2 instance, Amazon S3, Amazon Lambda, Amazon Gateway API를 생성한다. 웹 페이지의 배포를 위해 웹서버를 구축한다. 웹서버, 웹페이지, AWS의 원활한 상호작용을 위해 Socket 통신을 활용한다.

동영상이 업로드 되는 공간, 업로드 된 동영상의 프레임 추출, 딥러닝 모델에 넣는 작업을 위한 Amzon S3, Amazon Lambda, Amazon Gateway API를 이용한다. 단순 S3에 업로드하는 방식을 채택할 경우, IAM 사용자의 Access Key 및 Secret Key가 노출될 가능성이 높으므로 Gateway Api와 Lambda를 통해 업로드를 진행한다. S3에 업로드가 완료되면 Lambda 함수에서는 프레임을 추출하고, 전처리를 진행하는 코드를 실행한다.

AWS EC2 instance는 AWS Deep Learning AMI를 채택하여 딥러닝 모델의 학습을 진행한다. 모델 학습에 있어 필요한 데이터셋은 Amazon S3에 저장한다.

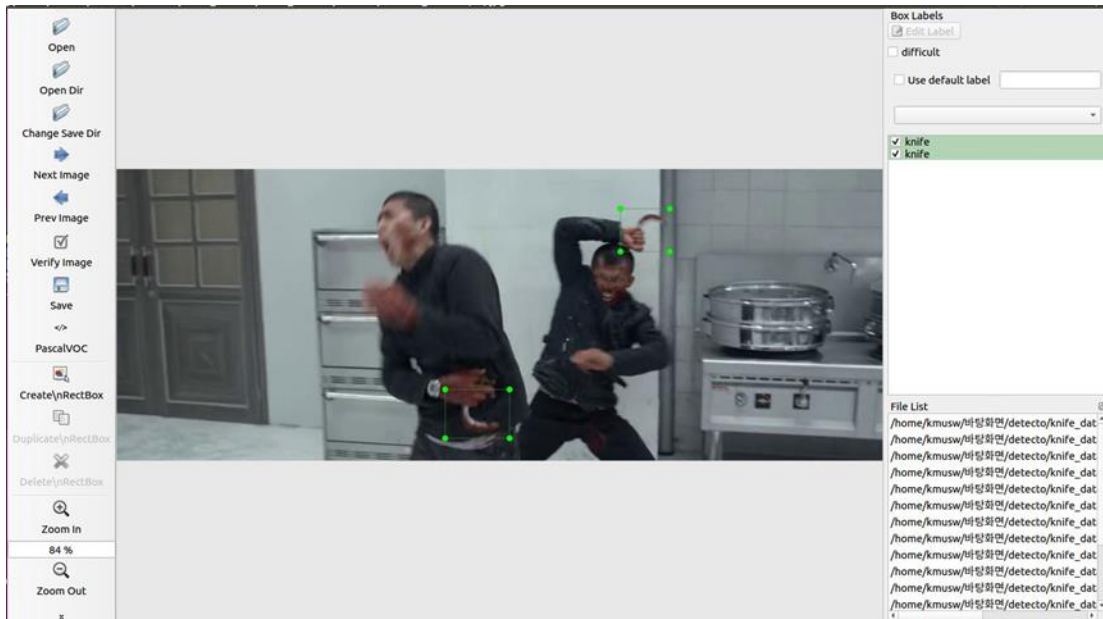
Apache 서버를 통해 웹페이지를 배포한다.

웹페이지 이용자와 AWS를 통해 얻어진 데이터의 주체가 동일인물인지 파악하고 Back End에서 일처리(검열 작업)의 종료를 웹페이지에 알려주기 위해 Socket 서버를 구축하여 통신한다. Socket은 Socketio 라이브러리를 활용한다.

2.2 수행내용

2.2.1 객체 탐지

먼저 Img와 annotation된 xml 데이터셋의 구축을 위해 labelImg 라이브러리를 활용한다. labelImg를 이용한 데이터셋 구축은 다음과 같이 구축했다.




각 클래스(라벨)는 약 200장의 이미지로 구성되었으며, xml 파일을 포함해 총 1600개의 데이터셋과 annotation셋을 구축했다. annotation은 매칭된 이미지에 표시된 박스의 위치와 좌표로 기록되어있으며, Object Detection 학습에 사용된다. 이후 구축이 완료된 데이터셋을 이용해 Faster R-CNN 딥러닝 모델이 적용된 detecto라는 라이브러리로 객체 탐지 학습을 진행한다. 각 라벨별로 객체 탐지 결과는 다음과 같다.



예측값이 높은 순서대로 출력되고 모든 예측에 윈도우가 그려진다. 이 중 가장 높은 예측값만을 사용하여 검열을 진행한다.

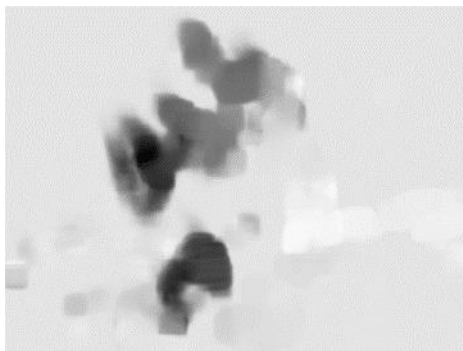
2.2.2 영상 검열

먼저 학습하기 전에, RGB로 표현된 이미지를 Optical Flow로 변환하는 과정을 수행한다. 각 Horizontal Flow와 Vertical Flow로 변환되며 총 10장의 프레임이 하나의 데이터로 학습되기때문에 (20 * 224 * 224) 차원에 각 Flow들을 넣는다.

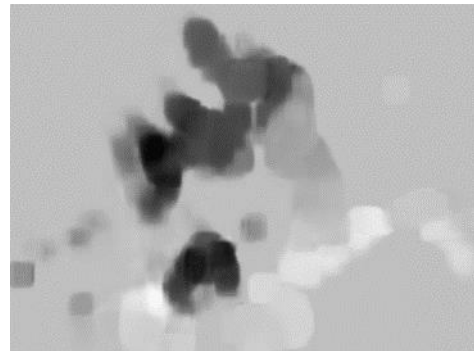
 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	동영상 연령제한 필터링	
	팀 명	YouHi	
	Confidential Restricted	Version 1.5	2020-APR-22



<Original Image>



<Horizontal>



<Vertical>


2.2.3 음성 추출 및 분할

영상에서 moviepy 라이브러리를 통해 음성 데이터를 추출한 후, pydub 라이브러리를 사용해 음성 데이터를 분할하여 저장한다. 이때 각 음성데이터의 길이가 5초 단위로 저장될 수 있도록 분할한다.

2.2.4 Speech To Text

분할된 음성 데이터를 Google Cloud Speech API 를 통해 텍스트 파일로 변환한 후 지정된 형식의 파일(csv, txt 등) 로 저장한다.

안녕하세요 오늘은 암기법 영상을 들고왔어요
 여러분들께서 정말 정말 많이 요청해주셔서 이렇게 찍어보게 되었습니다 사실 저는
 암기는 시간 투자와 반복이라고 생각을 해요 근데 이게 마치 다이어트는
 식이요법과 운동이다 이런 말이랑 똑같잖아요 왜냐면

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	동영상 연령제한 필터링	
	팀 명	YouHi	
	Confidential Restricted	Version 1.5	2020-APR-22

2.2.5 형태소 분석

저장된 텍스트파일은 khaiii api 를 이용해 모든 문장을 형태소 단위로 분리한다. khaiii 형태소 분석기를 빌드 및 설치한 후 Python 바인딩을 하여 Python 인터프리터에서 사용할 수 있도록 한다. 그 후 형태소 분석을 위해 khaiii 형태소 분석기를 이용하는 Python 파일을 실행한다. 인자는 Speech To Text로 추출된 텍스트가 저장되어 있는 텍스트 파일을 주도록 한다. 실행 결과는 다음과 같다.

```

kmsuw@kmsuw-ThinkPad-T440: ~/바탕화면/fastText/mywork/text_file
kmsuw@kmsuw-ThinkPad-T440:~/바탕화면/fastText/mywork/text_file$ cat test_tokenized khaiii.txt
안녕/NNG 하/XSA 세요/EC 오늘/NNG 은/JX 암기법/NNG 영상/NNG 을/JKO 들고오/VV 왔/EP 어요/EC
여러분/NP 들/XSN 께서/JKS 정말/MAG 정말/MAG 많이/MAG 요청/NNG 하/XSV 여/EC 주/VX 시/EP 어서/EC
이런/VA 게/EC 짝/VV 어/EC 보/VX 게/EC 되/VV 었/EP 습니다/EC 사실/MAG 저/NP 는/JX
암/VV 기/NNG 는/JX 시간/NNG 투자/NNG 와/JC 반복/NNG 이/VCP 라고/EC 생각/NNG 을/JKO 하/VV 여요/
EC 근데/MAJ 이것/NP 이/JKS 마치/MAG 다이어트/NNG 는/JX
식이/NNG 요법/NNG 과/JC 운동/NNG 이/VCP 다/EC 이런/MM 말/NNG 이/JX 랑/JKB 똑같/VA 지/EC 않/VX
아요/EC 왜냐면/MAG
kmsuw@kmsuw-ThinkPad-T440:~/바탕화면/fastText/mywork/text_file$

```

출력 결과는 분석된 형태소와 해당 형태소에 대한 품사를 나타내고 띄어쓰기로 구분되어 있다. 줄 바꿈은 인자로 받았던 텍스트 파일과 동일하게 적용된다. 위 출력결과에 있는 형태소들은 '/' 와 자신의 품사를 띄어쓰기 구분없이 붙여서 나타내기 때문에, 한글 데이터로만 학습된 FastText 모델에서는 테스트할 수 없는 데이터이다. 따라서 형태소 뒤에 붙은 '/품사'를 없애 주는 데이터 전처리 과정을 거쳐야 한다. 데이터 전처리 과정은 다음과 같이 실시한다. split 함수를 이용하여 하나의 line에 존재하는 형태소들을 각 하나의 원소로 갖는 List를 얻는다. 그리고 find 함수를 사용하여 형태소에서 '/'가 존재하는 위치를 파악한 후 그 위치의 앞부분들을 List의 원소로 저장한다. 각 line마다 이러한 전처리를 진행하여 한글로만 이루어진 형태소 데이터를 원소로 갖는 List를 출력하면 다음과 같다.



```
kmusw@kmusw-ThinkPad-T440: ~/바탕화면/fastText/mywork
kmusw@kmusw-ThinkPad-T440:~/바탕화면/fastText/mywork$ python3 voice_filter.py --input ./text_file/test_tokenized khaii.txt
[안녕, '하', '세요', '오늘', '은', '암기법', '영상', '을', '들고오', '았', '어요']
[여러분, '들', '께서', '정말', '정말', '많이', '요청', '하', '여', '주', '시', '어서', '이런', '게', '찍', '어', '보', '게', '되', '었', '습니다', '사실', '저', '는']
[암, '기', '는', '시간', '투자', '와', '반복', '이', '라고', '생각', '을', '하', '여요', '근', '데', '이것', '이', '마치', '다이어트', '는']
[식이', '요법', '과', '운동', '이', '다', '이런', '말', '이', '랑', '똑같', '지', '않', '아요', '왜냐면']
kmusw@kmusw-ThinkPad-T440:~/바탕화면/fastText/mywork$
```


2.2.6 욕설 검열

FastText 모델을 학습시키기 위한 텍스트 데이터셋은 한국어 위키디피아 텍스트와 웹 크롤링으로 수집한 욕설 텍스트로 이루어져 있다. 기존 데이터셋은 한글을 제외한 문자들이 다수 존재하기 때문에 한글을 제외한 문자들을 텍스트로부터 제외하는 전처리 과정을 진행한다. 그 결과로 생성되는 한글로만 이루어진 텍스트 데이터셋은 FastText skip-gram 모델의 학습 데이터로 이용된다. 데이터 전처리는 Python에서 제공하는 re 모듈을 이용하여 진행하였다.

학습이 완료된 FastText 모델은 분석된 형태소와 욕설 간의 코사인 유사도를 확인하는데 사용된다. 음성 검열에 있어 이 과정이 적합한지 판단하기 위해 3가지 경우에 대해 각각 코사인 유사도의 정확도를 판단하였는데, 첫 번째 경우는 욕설과 욕설이 아닌 형태소, 두 번째는 욕설과 욕설, 마지막 세 번째 경우는 욕설과 욕설과 비슷하지만 욕설이 아닌 형태소이다. 각 경우에 대해 계산한 코사인 유사도는 다음과 같다.

```
kmusw@kmusw-ThinkPad-T440:~/바탕화면/fastText/mywork$ python3 cs_sim.py
'안녕' | '시': 0.42002043
'병' | '시': 0.73076606
'시발점' | '시': 0.30771598
```

첫 번째 경우를 살펴보면, 욕설과 욕설이 아닌 형태소의 코사인 유사도는 0.420으로 0.5 이하인 값을 보였다. 그에 비해 욕설과 욕설의 코사인 유사도를 보면 0.730으로 매우 높은 값을 갖는데, 이는 다른 자음과 모음으로 이루어져 있다 하더라도 학습 데이터에 따라 욕설로 판단되었기 때문에 이와 같은 결과를 얻게 된 것이다. 마지막으로 세 번째 경우 '시

	국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서	
		프로젝트 명	동영상 연령제한 필터링
		팀 명	YouHi
		Confidential Restricted	Version 1.5 2020-APR-22

발점'이라는 단어는 '시발'이라는 욕설을 포함하고 있으나 첫 출발을 하는 지점이라는 의미를 갖고 있기 때문에 유사도는 0.307이라는 낮은 값을 갖게 되었다.

위의 결과에 따라 FastText 모델을 이용한 코사인 유사도 측정은 음성 검열에 있어 충분히 적합하다는 것을 알 수 있다. 욕설이나 일반 대화 내용 등 다양한 텍스트 데이터들이 학습 데이터셋에 추가될 수록 정확도는 높아질 것이기 때문에, 더 많은 욕설을 검열할 수 있도록 웹 크롤링 등을 통해 추가 데이터 셋을 구축해야 할 필요가 있다.

```

안녕 0.42002043
하 0.47822198
세요 0.5804804
오늘 0.48300332
은 0.1389905
암기법 0.2689774
영상 0.14835094
을 0.15537353
들고오 0.18225515
았 0.13777567
어요 0.5483629

여러분 0.47882926
들 0.13137506
꺼서 0.48440227
정말 0.60884464
정말 0.60884464
많이 0.3309979
요청 0.1979914
하 0.47822198
여 0.26458108
주 0.2204673
시 0.32391378
어서 0.39585653
이런 0.2586014
게 0.4637138
찍 0.081417896
어 0.44221303

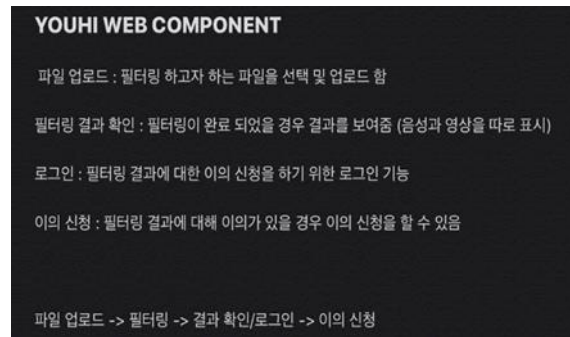
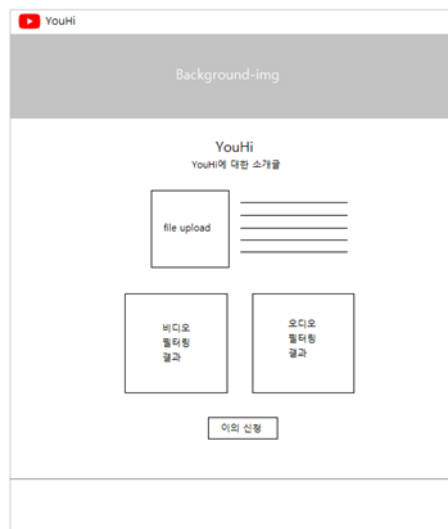
```

위 사진은 Khaiii 형태소 분석기 결과에 대한 전처리 적용 후 같은 문장에 속해 있는 형태소들끼리 욕설 '시발'과 코사인 유사도를 측정하여 출력한 결과이다. 몇몇 형태소들은 모델의 학습 데이터에서 욕설과 가까운 곳에 위치하고 있어 다른 형태소들보다 상대적으로 유사도가 높게 측정되었지만 대부분 욕설이 아닌 형태소들이기 때문에 매우 낮은 값을 갖게 되었다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	동영상 연령제한 필터링	
	팀 명	YouHi	
	Confidential Restricted	Version 1.5	2020-APR-22

2.2.7 Web Front End


아래 사진은 WEB 초기 구상 UI와 Component이며 해당 내용을 바탕으로 구현을 진행하였다.



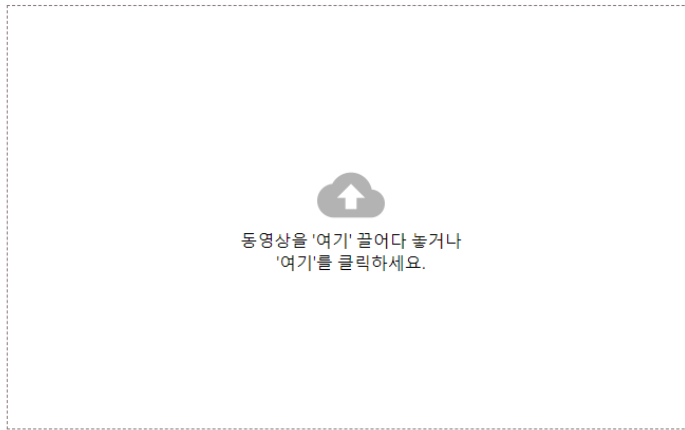
컴포넌트 종류는 다음과 같다.

1) 파일 업로드

파일 업로드의 경우 file type의 input form을 사용할 경우 직접 파일을 선택해 올리는 것만 가능하기 때문에 드래그 앤 드랍 형식의 파일 업로드를 추가하기 위해 react-dropzone 라이브러리를 사용해 구현했다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	동영상 연령제한 필터링	
	팀 명	YouHi	
	Confidential Restricted	Version 1.5	2020-APR-22

Upload



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.
☐ **연령제한 콘텐츠로 설정합니다.**

업로드

필터

파일을 선택한 후에는 하단 좌측 이미지와 같이 업로드 버튼 위의 파일명을 통해 파일이 정상적으로 선택되었는지 확인할 수 있다. 파일 선택을 완료했을 경우 업로드 버튼을 누르면 프로그래스 바가 나타나면서 파일이 어느정도 업로드 됐는지 확인 할 수 있다. 업로드가 완료되었다면 필터 버튼을 클릭해 필터링을 진행할 수 있다.


☐ **연령제한 콘텐츠로 설정합니다.**

☐ **연령제한 콘텐츠로 설정합니다.**



2) 필터링 확인

필터링 컴포넌트의 경우 영상 필터링과 음성 필터링 컨테이너를 각각 만들어 이미지를 클릭하면 새로운 팝업창을 띄워 각각의 필터링 결과를 확인할 수 있도록 할 것이다.

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	동영상 연령제한 필터링	
	팀 명	YouHi	
	Confidential Restricted	Version 1.5	2020-APR-22

Filter



각각의 영상, 음성 필터링 이미지를 클릭 시 새로운 팝업 창이 나오고 서버로 부터 받은 필터링 된 결과가 화면에 나타난다. 영상 업로더는 필터링 된 결과를 확인하고 문제가 있을 경우 이의신청 및 문의하기 버튼 클릭으로 문제를 제기할 수 있다.

이의신청 및 문의하기

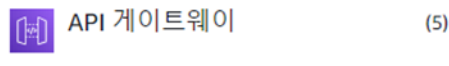
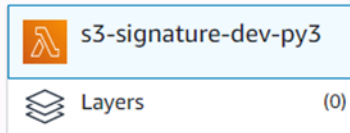
3) 로그인

영상을 업로드 하기 위해서는 본인 인증이 가능해야 하기 때문에 로그인 기능을 구현하고자 한다. Google login api를 사용해 Google 계정으로 로그인 할 수 있게 구현할 예정이다. 영상 업로드 기능은 로그인 후 이용할 수 있다.

2.2.8 Back End



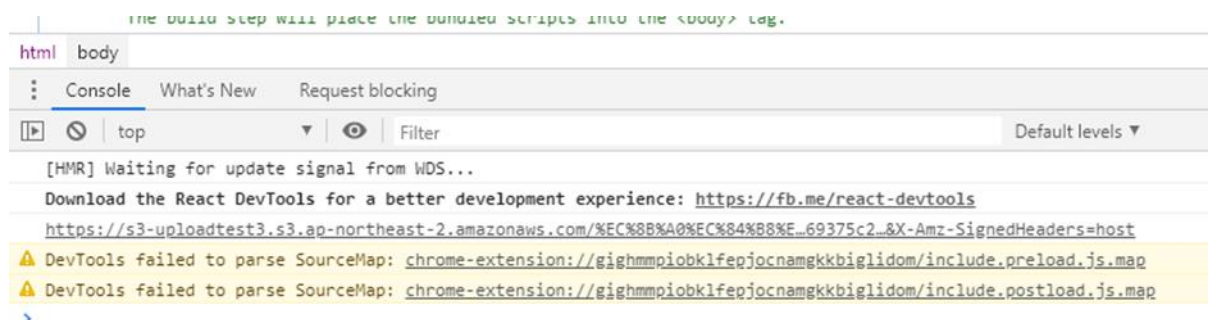
▼ Designer



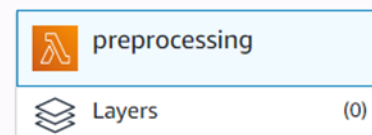
+ 대상 추가

+ 트리거 추가

Aws Gateway Api와 Lambda를 통해 pre-signed url을 GET Http 메소드를 통해 얻어온다.
하단에 https://s3-uploadtest3.s3.ap-northeast-2.amazonaws.com/%EC%88%A0%EC%84%B8%E_69375c2_8X-Amz-SignedHeaders=host로 시작하는 url이 pre-signed url이다.




▼ Designer



+ 대상 추가

+ 트리거 추가

	중간보고서		
	프로젝트 명	동영상 연령제한 필터링	
	팀 명	YouHi	
	Confidential Restricted	Version 1.5	2020-APR-22

s3-uploadtest3

개요 속성 권한 관리 **리소스** 액세스 지정

Q 검색하려면 접두사를 입력하고 Enter 키를 누릅니다. 지우려면 Esc 키를 누릅니다.

업로드 + 폴더 만들기 다운로드 작업

<input type="checkbox"/> 이름	마지막 수정	크기
<input type="checkbox"/> 신세계.avi	4월 2, 2020 11:39:54 오후 GMT+0900	5.9 MB

"신세계.avi"를 s3-uploadtest3에 업로드한 화면이다. 해당 업로드는 pre-signed url을 통해 업로드 하였다.

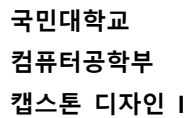
youhi-project

개요 속성 권한 관리 액세스 지정

Q 검색하려면 접두사를 입력하고 Enter 키를 누릅니다. 지우려면 Esc 키를 누릅니다.

업로드 + 폴더 만들기 다운로드 작업

<input type="checkbox"/> 이름
<input type="checkbox"/> 신세계



프로젝트 명

동영상 연령제한 필터링

YouHi

Confidential Restricted

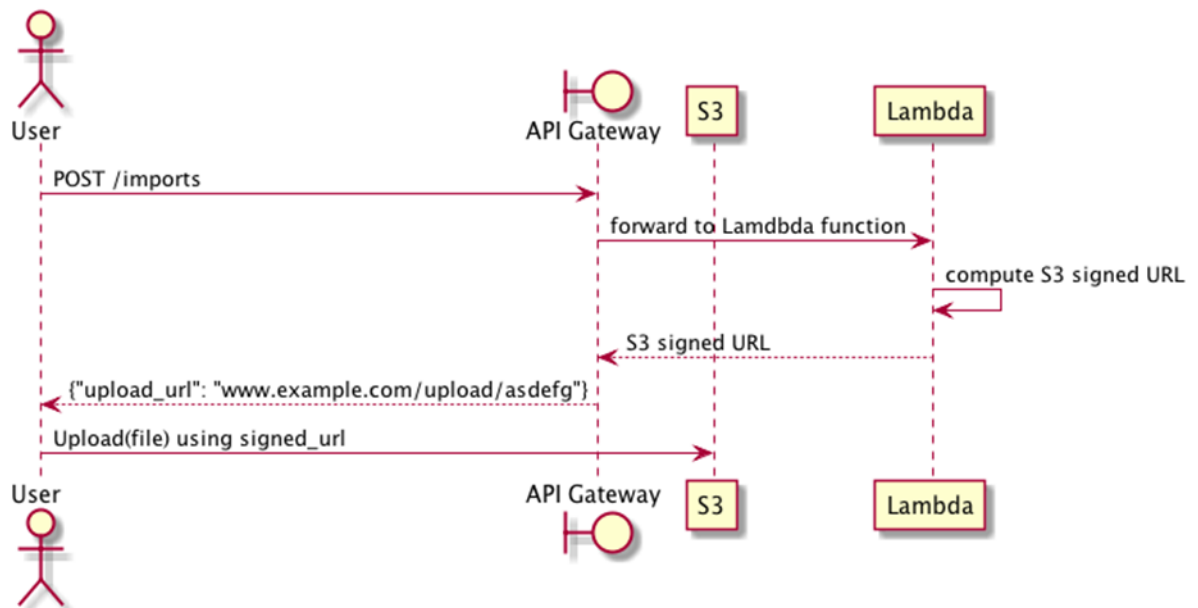
Version 1.5

2020-APR-22

s3-uploadtest3 버킷에 업로드 완료되는 이벤트에 따라 youhi-project 버킷에 파일명에 해당하는 폴더를 생성하고 5초에 하나의 프레임을 추출한다.

3.1.1 동영상 업로드 방식 변경


기존 동영상을 S3에 업로드할 때 Gateway Api와 Lambda를 통해 "업로드"를 바로 진행했으나, 5MB이상 파일을 업로드하지 못하는 문제점을 발견하였다. 또한 그 속도가 S3에 Direct하게 업로드 하는 것보다 느리다고 판단되었다. 따라서 다음과 같은 구조로 업로드 방식을 변경하였다.



User가 업로드를 시작하면 Gateway Api와 Lambda를 거쳐서 signed url을 전달받는다. 이 url은 S3에 Direct하게 업로드할 수 있는 S3 url이며 유효시간이 정해져있어, 일정 시간이 지난 이후에 이 url은 사용하지 못한다. 따라서 Direct하게 접근 가능하지만 기존과 마찬가지로 외부에 공개될 우려가 없어 보안을 안전하게 유지할 수 있다.

3.1.2 영상 검열 방식 변경

기존 Video Classification 단일 모델을 사용하는 것은 문제가 많은 것으로 판단되어 Object Detection를 결합하고 4개의 Video Classification 모델을 이용해서 영상 검열을 진행한다. 또한 모든 영상에 대한 프레임을 추출하여 Video Classification을 진행하는 것이 아닌 5초에 한 frame을 추출하여 Object Detection을 진행하고, Object Detection으로 칼, 총, 담배 등이 검출되면 Object에 따라 각 Video Classification이 진행되는 방식으로 변경했다. 예를 들어, Object Detection으로 칼이 검출되면, 해당 영상이 요리를 하는 영상이거나 칼에 대한 제품을 소개하는 영상 혹은 칼로 전투를 치르지만 15세 이상 적용 대상인 영상인지 Video Classification으로 최종 판단하는 과정을 거치게 된다. 이와 같은 방식을 사용하면, 오분류가 매우 줄어들 것으로 예상되며 또한 기존 모든 프레임 추출 방식에 비해 시간과 메모리가 극적으로 줄어들 것이다.

	중간보고서		
	프로젝트 명	동영상 연령제한 필터링	
	팀 명	YouHi	
	Confidential Restricted	Version 1.5	2020-APR-22

3.1.3 웹페이지 설계 및 UI 변경

기존 웹페이지에 YouTube등 동영상 플랫폼에서 이용할 수 있는 시스템처럼 보이기 위해 연령 제한 옵션과 해당 플랫폼에 대한 설명 등을 설정하고 표출시켰지만, 계획 발표의 피드백인 "YouTube에 적용시키겠다라는 목표가 현실적으로 불가능하니 새롭게 설정해야된다"를 반영하여 "YouTube 등 영상 플랫폼에서 활용 가능한 시스템을 개발하겠다"로 새롭게 설정했다. 그에 따라 웹페이지에 있던 연령 제한 옵션 등을 삭제했다.

4 향후 추진계획

4.1 향후 계획의 세부 내용

4.1.1 영상검열

영상 분류기 딥러닝 모델에 사용할 데이터셋을 추가로 구축한다. 기존 데이터셋에서 학습의 문제점을 발견하였다고 판단, 새로운 데이터셋을 추가한다.

선정적 요소와 폭력성이 포함된 게임 영상에 대한 유해 판단 기능을 구현한다. 현재 칼로 연출된 잔인한 영상과 흡연하는 영상에 대한 유해 판단 기능은 구현이 완료되었다.

영상 검열이 완료된 결과를 웹페이지에 띄우는 방식을 고안한다.

4.1.2 정확도 측정

본 프로젝트는 유해한 내용을 유해하다고 판단하는 것과 유해하지 않은 내용을 유해하지 않다고 판단해야 하기 때문에, 기술적 완성도와 시스템의 필요성을 평가 받으려면 정확도가 가장 중요하다. Precision / Recall 매트릭스를 통해 정확도를 측정하고 개선하는 방향으로 진행한다. 다음과 같이 Precision / Recall 매트릭스를 구성한다.

- 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)

		실제 정답	
		True	False
분류 결과	True	True Positive	False Positive
	False	False Negative	True Negative

True는 유해 요소, False는 유해하지 않은 요소로 설정하여 평가한다. 평가 방식은 Precision과 Recall을 이용한다.

$$(Precision) = \frac{TP}{TP + FP} \quad (Recall) = \frac{TP}{TP + FN}$$

4.1.3 Front End

영상과 음성 검열이 완료되면, 해당 결과를 웹페이지에 띄우는 기능을 구현한다. 현재는 검열 완료시 확인할 수 있는 화면을 버튼 클릭 이벤트로 처리해두었다.

Web 디자인 및 UI를 다양한 사용자들에게 보여주고, 지속적인 피드백을 통해 개선한다.

로그인 기능 및 문의 사항과 이의 신청 기능을 구현한다. 로그인은 Google 계정 연동으로 진행하고, 문의 사항과 이의 신청은 Database를 사용하는 것이 아닌, E-mail로 받아 관리할 예정이다.

4.1.4 Back End

여러 사용자가 동시에 영상 업로드 및 검열을 실시할 때 서버가 안정적으로 작동하는지

 국민대학교 컴퓨터공학부 캡스톤 디자인 I	중간보고서		
	프로젝트 명	동영상 연령제한 필터링	
	팀 명	YouHi	
	Confidential Restricted	Version 1.5	2020-APR-22

테스트하여 문제 발생 시, 이를 개선한다.

5 고충 및 건의사항

유튜브, 실시간 영상 송출 플랫폼, 한국 방송통신위원회에서는 각자 다른 영상등급 선정 기준을 사용하고 있습니다. 또한 영상 내용이 청소년 관람 불가 내용으로 판단하는 기준이 명확하게 설정되어있지 않기 때문에, 본 프로젝트에서는 각 기관의 선정 기준을 취합하고 각 기관의 모든 기준을 포괄하는 것을 목표로 자체적인 가이드라인을 설정했습니다. 따라서 저희의 가이드라인이 정확하지 않을 수 있지만 저희가 설정한 가이드라인에 충족되는 영상과 그렇지 않은 영상을 정확하게 검열하는 여부에 대해서 우선적으로 판단해주시면 감사하겠습니다.