

캡스톤 디자인 2

2015104199 이수인

0. 목차

- 개요
- 기존 연구
- 주제 설명
- 프로젝트 진행 과정
- 데모 영상

1. 개요

- 대주제 : 공공데이터 분석을 위한 딥러닝 기반의 분석

- 연구 배경

대형복합재난 등의 경우 모든 현장들이 사람의 접근이 가능한 현장이라고 확신하기가 어려움. 그러한 상황에서 더욱 더 신속한 인명구조가 이루어질 수 있도록 드론에서 취득된 음성 데이터에서 구조 요청의 발원지를 추정하고자 함

- 프로젝트 주제 : 드론에서 취득된 스테레오 타입의 음성데이터에서 발화 방향 및 성별, 연령 추정
 - 스테레오 타입 : 두 개의 채널로 이루어져 있는 음성 데이터

2. 기존 연구 (1)

[DRONAIID]

사람을 탐지할 수 있는 실시간 자율 무인 기술 시스템인 DRONAIID에 관한 연구

무인 항공기 기반 시스템이기 때문에 쉽게 동원과 제어 가능

카메라 모듈 및 센서 장치가 포함되어 있어 잔해 아래에 묻혀있는 인간의 존재 역시 확인이 가능

PIR 센서를 탑재하여 사람의 방사선을 감지, 위치 파악이 가능

2. 기존 연구 (1)

[DRONAIID] - 문제점

PIR 센서의 경우 일정한 적외선을 가진 물체의 움직임을 감지

움직임이 없거나, 미세한 움직임은 감지를 하지 못함

희생자가 움직이기 여의치 않은 경우에는 성능을 기대하기가 어려움

센서 자체가 온도 변화와 같은 주변 환경의 영향을 민감하게 받아들이기 때문에 야외에서 사용할 때에는 정확한 탐지가 어려움

2. 기존 연구 (2)

[동일 연구 (캡스톤 디자인 1)]

저번 학기에 진행을 했었던 동일한 연구

드론으로 추출된 스테레오 타입의 음성 데이터를 사용하여 발화자가 어디서부터 구조요청을 했는지를 추정하는 연구

드론의 마이크로 수집된 음성정보로 추정하기 때문에 야외에서도 당연히 사용이 가능하며, 희생자가 움직이기 여의치 않은 상황에서도 구조요청만 한다면 해당 발화 방향을 추정할 수 있음

2. 기존 연구 (1)

[동일 연구 (캡스톤 디자인 1)] - 문제점

만족할 만한 성과를 거두지 못함

해당 목소리의 발원 방향을 추정하는 것까지는 성공하였지만, 해당 추정의 범위는 60도 정도로 실제 현장에 해당 각도를 적용시켜본다면 꽤나 넓은 범위가 선택되어 희생자의 정확한 위치 파악이 어려움

발화자의 성별, 연령대의 정보도 학습시키지 않았기 때문에 구조요청 대상에 따른 조속한 처치를 기대할 수 없음

3. 주제 설명



3. 주제 설명

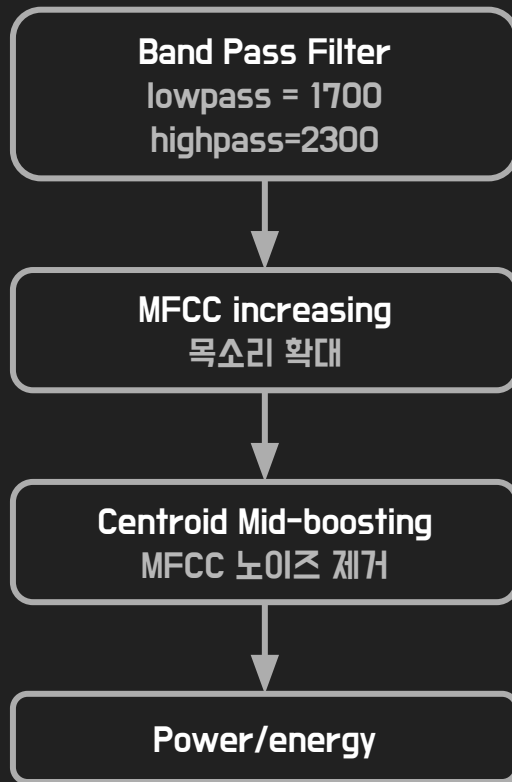
- 목표 :
 - 노이즈가 포함된 스테레오 타입의 음성 데이터에서 구조요청 소리, 즉 사람의 목소리만을 분리
 - 20도 간격으로 해당 발화 방향 추정
 - 남자 / 여자 / 어린이로 나누어 해당 발화자의 정보 추정

4. 프로젝트 진행 과정

- 크게 네 단계로 나누어 볼 수 있음 :
 - 음성 데이터에서 사람의 목소리만을 추출해내기 위한 필터링
 - 노이즈가 제거된 데이터로부터 특징 추출
 - 추출된 발화 방향 관련 특징을 통해 20도 간격으로 발화 방향을 추정하는 모델 설계
 - 추출된 특징을 통해 남자/여자/어린이 세 개의 클래스로 구분하는 모델 설계

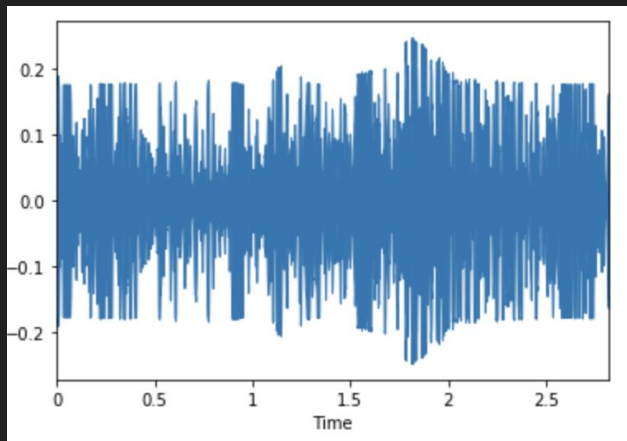
4. 프로젝트 진행 과정 - (1) 노이즈 제거

- 노이즈 제거 필터
 - Band Pass Filter
- 음성 강화 필터
 - using Centroid
 - using MFCC
 - using Power/energy

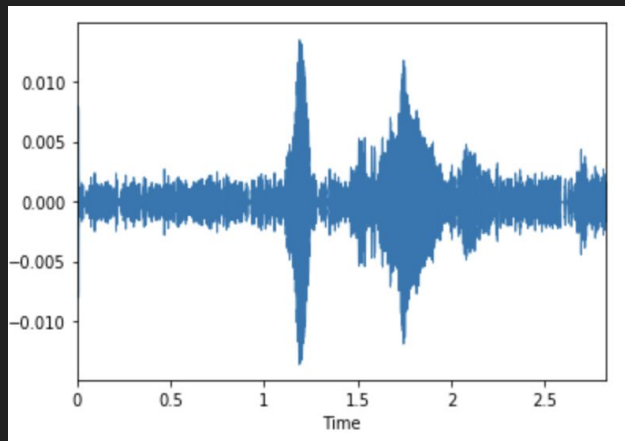


4. 프로젝트 진행 과정 - (1) 노이즈 제거

- 노이즈 제거 전



- 노이즈 제거 후



4. 프로젝트 진행 과정 - (2) 특징 추출

- 각도 추정에 이용되는 특징

MFCC

- 일정 구간 단위의 스펙트럼 분석

GCC-PHAT

- 다채널 마이크의 경우 TDOA(Time Difference of Arrival)추정에 사용되는 특징

PLP Cepstrum

- 시간적 변화에 민감한 특징 (채널 간의 시간 차이에 요긴할 것으로 예상)

STFT

- 주파수 정보

4. 프로젝트 진행 과정 - (2) 특징 추출

- CNN 학습을 위해 2차원 배열의 Feature Map 형태로 생성
- MFCC + STFT + GCC-PHAT + PLP

⇒ (120, 150) 크기의 벡터

- 데이터 손실을 최소화 하고자 처음에는 (2065, 150) 크기의 벡터로 학습을 진행하였음

⇒ 학습 속도가 너무 느림

⇒ 학습 정확도가 높지 않음 (유용하지 않은 정보가 포함되었기 때문으로 예측)

4. 프로젝트 진행 과정 - (2) 특징 추출

- 성별/나이 추정에 이용되는 특징

10차 LPC Cepstrum 계수

12차 LPC Cepstrum 계수

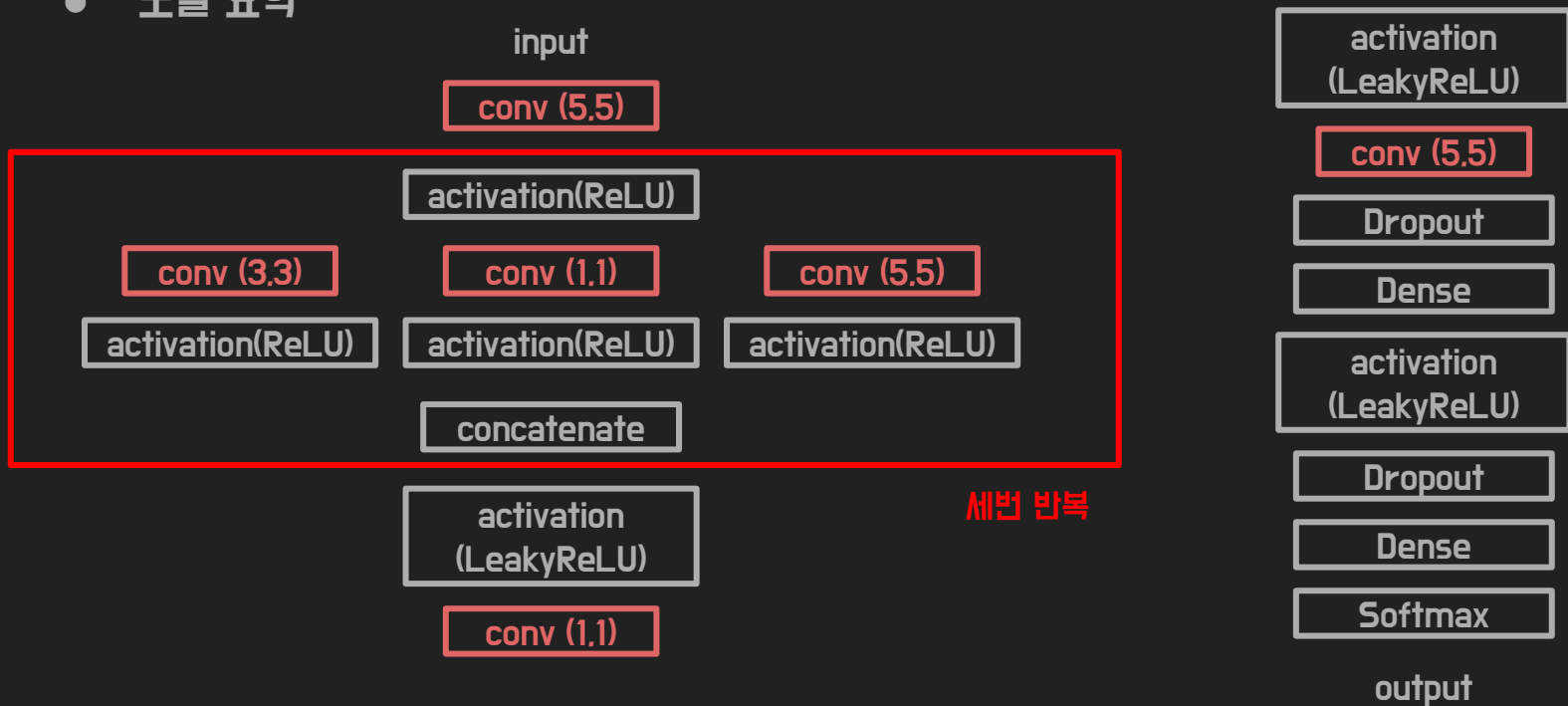
12차 FTT Cepstrum

8차 FTT Cepstrum

+ MFCC (정확도 향상을 위해 추가)

4. 프로젝트 진행 과정 - (3) 방향 추정 모델 설계

- 2차원 벡터의 형태로 특징을 추출하였기 때문에 CNN 사용가능
- 모델 요약



4. 프로젝트 진행 과정 - (4) 성별/나이 모델 설계

- 2차원 벡터의 형태로 특징을 추출하였기 때문에 CNN 사용가능
- 모델 요약

Layer (type)	Output Shape	Param #
=====		
conv2d_54 (Conv2D)	(None, 40, 150, 64)	640
batch_normalization_42 (Batch Normalization)	(None, 40, 150, 64)	256
conv2d_55 (Conv2D)	(None, 38, 148, 64)	36928
batch_normalization_43 (Batch Normalization)	(None, 38, 148, 64)	256
max_pooling2d_47 (Max Pooling)	(None, 19, 74, 64)	0
conv2d_56 (Conv2D)	(None, 19, 74, 128)	73856
batch_normalization_44 (Batch Normalization)	(None, 19, 74, 128)	512
conv2d_57 (Conv2D)	(None, 17, 72, 128)	147584
batch_normalization_45 (Batch Normalization)	(None, 17, 72, 128)	512
max_pooling2d_48 (Max Pooling)	(None, 8, 36, 128)	0

conv2d_58 (Conv2D)	(None, 8, 36, 256)	295168
batch_normalization_46 (Batch Normalization)	(None, 8, 36, 256)	1024
conv2d_59 (Conv2D)	(None, 6, 34, 256)	590080
batch_normalization_47 (Batch Normalization)	(None, 6, 34, 256)	1024
max_pooling2d_49 (Max Pooling)	(None, 3, 17, 256)	0
flatten_5 (Flatten)	(None, 13056)	0
dense_11 (Dense)	(None, 1024)	13370368
dense_12 (Dense)	(None, 512)	524800
dense_13 (Dense)	(None, 3)	1539
=====		
Total params: 15,044,547		
Trainable params: 15,042,755		
Non-trainable params: 1,792		

4. 프로젝트 진행 과정 - (3, 4) 모델 설계

- Overfitting

Batch Normalization

L2 Norm

Dropout

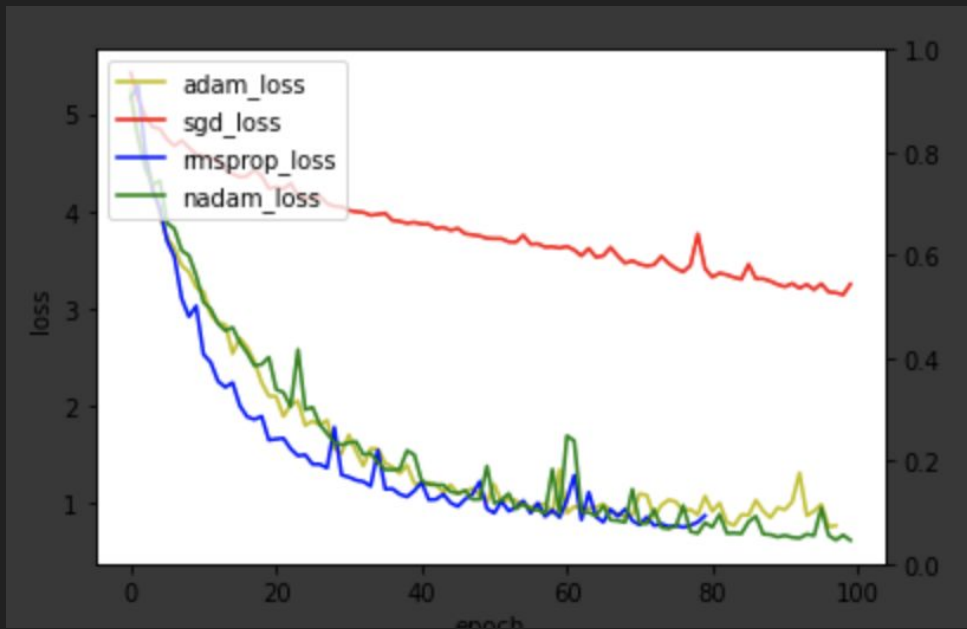
Early Stopping (validation loss = min, patience = 30)

4. 프로젝트 진행 과정 - (3, 4) 모델 설계

- Dead Relu 문제 \Rightarrow Leaky Relu 혼합 사용
- Activation : Nadam
- Train : Valid : Test

= 0.64 : 0.16 : 0.2

- loss = categorical cross entropy
- batch size = 256



4. 프로젝트 진행 과정 - (3, 4) 모델 설계

- 정확도

K-Fold Validation (K=5)

"방향 추정 정확도 : 91.89 %"

"성별/나이 추정 정확도 : 97.98 %"

5. 데모 영상

