**TinyML을 통한 소리 인식 및 위치 탐지**

Sound Recognition and

Localization with TinyML

요약

본 논문은 TinyML 기술을 활용하여 재난 생존자 위치를 빠르게 감지하고 구조하는 시스템을 개발하는 것을 목표로 한다. 시스템은 Edge Impulse 플랫폼을 사용하여 TinyML 모델을 학습하며, 드론에 장착하여 사고 피해자의 소리를 감지하고 위치를 파악한다. 소리 분류 및 소리 위치 파악 기술은 각각 2.1절과 2.2절에서 논의되었으며, TinyML 모델 개발 및 배포는 3절에서 제시됩니다. 이 논문은 특히 드론이 긴급 상황에 신속하게 대응하고 구조 작전의 성공률을 높이는 데 기여함으로써 재난 구호 작업에 기여할 것을 기대한다. 4절에서는 본 논문의 결론과 추후 연구 방향이 제안한다.

##### 1. 서론

최근 발생한 대지진과 같은 재난 상황에서 생존자를 빠르게 구조하는 것이 중요한 문제로 대두되고 있다. 기존의 드론 기술은 이미 재난 구조 작업에 활용되고 있지만, 생존자의 위치를 정확하게 파악하는 것이 여전히 어려운 문제가 있다. 이러한 문제를 해결하기 위해 논문은 TinyML 기술을 활용하여 드론이 생존자의 음성이나 소리를 감지하여 위치를 파악하고, 이를 기반으로 생존자를 신속하게 구조할 수 있는 시스템을 개발하는 것을 목표로 한다.

이번 논문에서는 Edge Impulse 플랫폼을 이용하여 TinyML 모델을 학습시켜, 드론에 장착하여 사고 피해자들의 소리를 감지하고 위치를 탐지하는 시스템을 구현할 것이다. 이를 통해 재난 구조 작업에서 드론이 보다 정확하고 신속하게 생존자를 찾을 수 있으며, 재난 구조 작업의 성공 확률을 높일 수 있도록 하는 것이 이번 논문의 핵심 목표이다.

이러한 논문은 재난 상황에서의 생명 구조 작업에 큰 도움을 줄 것으로 기대된다. 특히, 이번 논문에서 개발되는 시스템은 재난 구조 작업에서 드론이 빠르게 응급 상황에 대처할 수 있도록 도와주며, 이에 따라 많은 인명 구조 작업에서 적극적으로 활용될 것으로 예상된다.

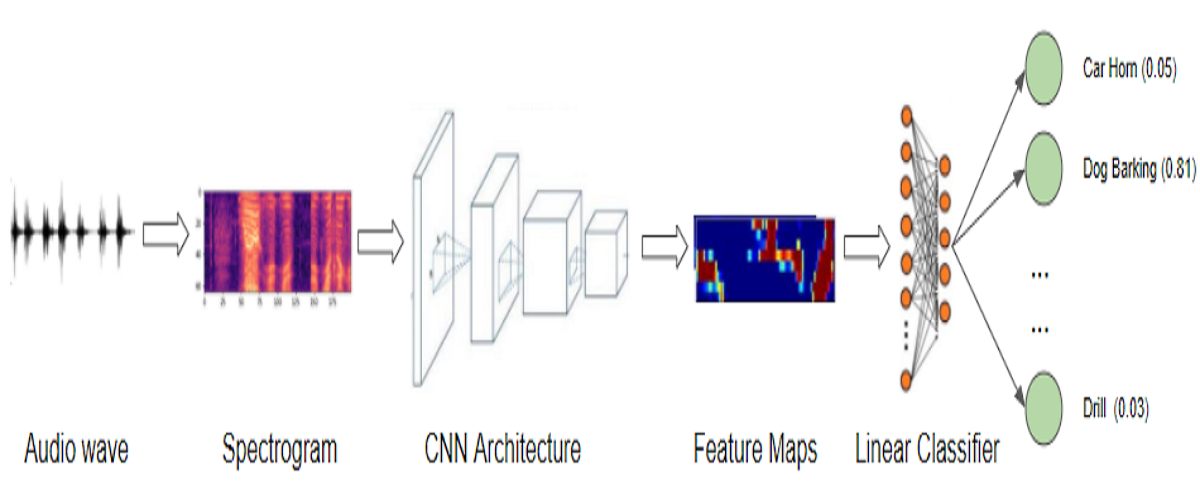
본 논문은 2절에서 Sound Classification과 Sound Localization을 조사한다. 2절에서 조사한 내용을 바탕으로 3절에서 Edge Impulse를 통해 TinyML을 설계하고, 4절에서는 결론 및 향후 연구 방향을 제시한다.

##### 2. 배경 이론 및 관련 연구

##### 2.1 Sound Classification

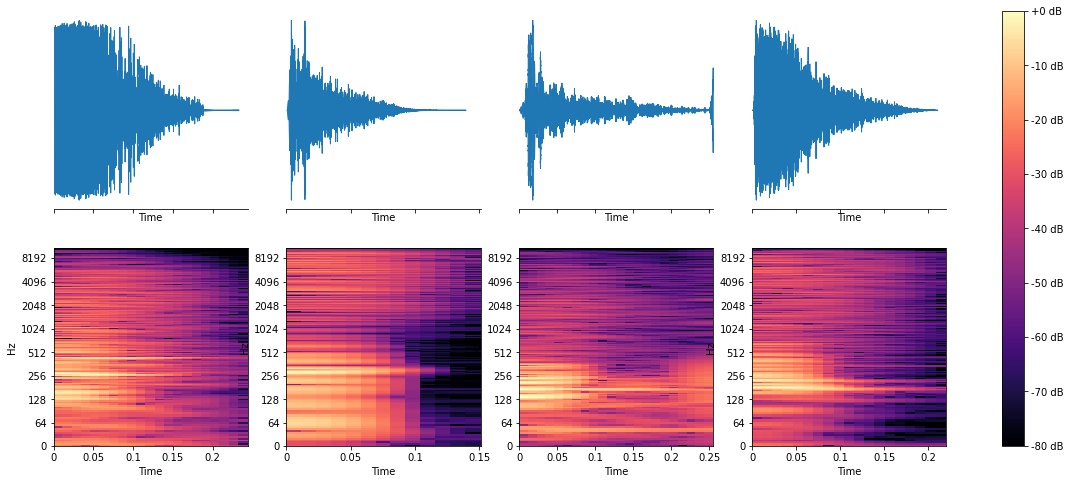
Sound Classification은 소리를 특정 범주로 분류하는 기술로, 소리를 녹음하여 이를 컴퓨터가 분석하고, 어떤 종류의 소리인지를 식별하는 기술이다.

Sound Classification은 딥 러닝 알고리즘을 사용하여 구현할 수 있다. 대표적으로는 Convolutional Neural Networks(CNN), Recurrent Neural Networks(RNN) 및 Transformer 등이 있다.[1] 이러한 알고리즘은 입력 데이터를 처리하고 분류하기 위한 모델을 학습한다. 학습된 모델은 입력 데이터의 패턴을 학습하여, 미지의 데이터를 분류하는데 사용된다.



[그림 1] Sound Classification 구조

Sound Classification을 구현할 때, 입력 데이터의 전처리가 중요하다. 일반적으로는 입력 데이터를 스펙트로그램으로 변환하여, 주파수 및 시간 정보를 추출한다. 이후, 데이터를 정규화하고, 데이터 어그멘테이션 및 데이터 증강 기술을 활용하여 데이터셋을 다양하게 만들어 모델 학습 효과를 향상시킨다.

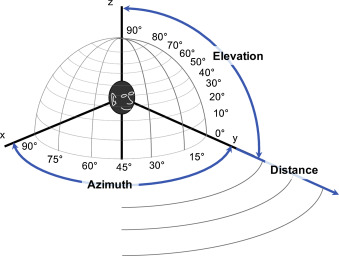


[그림 2] 음성 데이터 Spectogram

##### 2.2 Sound Localization

Sound Localization은 음향 신호가 발생하는 위치를 파악하는 기술로, 일반적으로 방향 산출(direction-of-arrival, DOA)이라고도 한다. 이 기술은 음향 신호가 발생하는 방향을 파악하는 것으로, 일반적으로 음향 신호의 도달 시간 차이를 측정하거나 다양한 마이크로 배열을 사용하여 파악한다.

이를 위해 일반적으로 음향 신호를 여러 개의 마이크로폰으로 수집하고, 마이크로폰 간의 도달 시간 차이(time delay)를 측정하여 해당 음향 신호가 발생한 방향을 파악한다.[2] 이러한 방법을 이용하여 구한 방향 정보는 보통 azimuth(방위각)와 elevation(고도각)으로 표현된다.



[그림 3] Sound Localization을 위한 Azimuth, Elevation

##### 2.3 Tiny ML

TinyML은 매우 작은 디바이스에서 작동하는 Machine Learning 모델을 의미한다. 예를 들어, 스마트폰, 웨어러블 기기, 드론, 로봇 등에서 작동할 수 있는 모델을 의미한다. 이러한 모델은 일반적으로 작은 크기와 낮은 전력 소비를 가지고 있으며, 이러한 특성은 모바일, IoT 및 에지 컴퓨팅 분야에서 중요하다.

##### Edge Impulse는 TinyML 모델 개발에 필요한 모든 도구를 제공하는 통합 개발 환경이다. 이 플랫폼을 사용하면 사용자는 디바이스에서 데이터를 수집하고, 모델을 훈련시키고, 배포할 수 있다. Edge Impulse는 데이터 수집, 데이터 전처리, 모델 선택, 모델 훈련 및 모델 배포와 같은 다양한 작업을 자동화한다. 이러한 기능을 통해 사용자는 TinyML 모델을 더 쉽고 빠르게 개발할 수 있다.

##### 

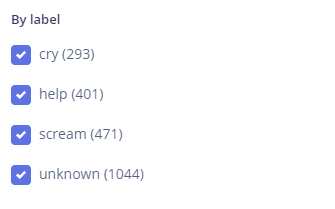
##### 본 논문에서는 Edge Impulse를 사용하여 TinyML 모델을 훈련시키고, 드론에서 사고 피해자의 소리를 감지하고 위치를 파악하는 시스템을 구현하고자 한다. 따라서 TinyML 모델을 구축하고 훈련시키는 방법에 대한 자세한 설명이 논문 3절에서 제공한다.

##### 

##### 

##### 3.1 데이터 준비

Sound Classification을 위한 데이터셋의 경우, 분류를 위한 다양한 음성 데이터를 사용한다. 우는 소리, 도와달라는 소리, 비명 소리를 포함하며 그 이외에는 다양한 소음 정보를 사용한다.



[그림 4] Sound Classification Dataset

Sound Localization을 위한 데이터셋의 경우, 드론 비행 음성 데이터를 사용한다. 음성 데이터 마다 azimuth, elevation 값이 정리되어 있으며 본 논문에서는 azimuth 값만을 활용하여 모델링을 한다. 고정 비행 정보에 대해선 고정된 azimuth값을 사용하며 동적인 비행 정보는 초 단위로 음성 정보를 분리하여 그에 맞는 azimuth 값을 라벨링한다.

[그림 5] Sound Data Split



[그림 6] Sound Localization Dataset

##### 3.2 모델 학습

Edge Impulse는 TinyML 모델을 학습, 테스트 및 배포할 수 있는 온라인 플랫폼으로, 간단한 인터페이스를 통해 사용자가 모델을 생성할 수 있다. 본 논문에서는 Edge Impulse를 활용하여 소리 인식 및 위치 탐지 모델을 학습시킨다. 학습 데이터는 재난 상황에서 생존자가 발생시키는 소리로 구성하여 Sound Classification 모델을 만들었고, 소리에 대한 azimuth값을 라벨링한 데이터를 통해 Sound Localization 모델을 학습시켰다.

Sound Classification 같은 경우 인간의 음성 언어 인식에 효과적인 MFCC를 통해 데이터를 전처리하고 Classification을 진행한다.

| Input Layers(650 features) |
| --- |
| Reshape Layer(13 columns) |
| 1D conv / pool layer  (8 neurons, 3 kernel size, 1 layer) |
| Dropout (rate 0.25) |
| 1D conv / pool layer  (16 neurons, 3 kernel size, 1 layer) |
| Dropout (rate 0.25) |
| Flatten layer |
| Output layer (4 classes) |

[표 1] Sound Classification layer 구조

| On-device performance | |
| --- | --- |
| Inferencing time | 2 ms |
| Peak Ram Usage | 5.1 k |
| Flash Usage | 26.8 k |

[표 2] Sound Classification 모델 자원 사용량

Sound Localization 같은 경우 Spectogram을 통해 음성 정보를 전처리한 뒤 Regression을 통해 Localization을 모델링한다.

| Input Layers(2,470 features) |
| --- |
| Reshape Layer(65 columns) |
| 1D conv / pool layer  (8 neurons, 3 kernel size, 1 layer) |
| Dropout (rate 0.25) |
| 1D conv / pool layer  (16 neurons, 3 kernel size, 1 layer) |
| Dropout (rate 0.25) |
| Flatten layer |
| Output layer (4 classes) |

[표 3] Sound Localization layer 구조

| On-device performance | |
| --- | --- |
| Inferencing time | 5 ms |
| Peak Ram Usage | 7.7 k |
| Flash Usage | 23.5 k |

[표 4] Sound Localization 모델 자원 사용량

##### 3.3 모델 배포

학습된 모델은 Edge Impulse 내에서 바로 배포한다. 또한, Edge Impulse는 다양한 디바이스에서 모델을 실행시킬 수 있는 SDK를 제공하며, 본 논문에서는 스마트폰을 사용하여 작동하는지 테스트한 뒤 Sony Spresense 디바이스를 사용하여 배포할 예정이다.

##### 4. 결론 및 향후 연구

이번 논문에서는 TinyML을 통해 학습된 모델을 스마트폰으로 배포하였다. 이후 Sony Spresense 디바이스로 모델을 배포할 예정이다.

Sound Classification과 Sound Localization을 통합한 모델을 드론에 탑재하면, 드론이 자동으로 구조 요청하는 소리를 인식하고 해당 위치를 탐지할 수 있다. 이러한 기술은 재난 구조 작업에서 매우 유용하게 활용될 수 있으며, 실제로 적용될 경우 구조 시간을 단축시키고 인명 구조에 큰 도움이 될 것이다.

향후 연구에서는 다양한 국적의 언어 데이터를 추가하여 모델의 범용성을 높일 예정이다. 또한, 이번 연구에서는 드론에 모델을 탑재하는 방법에 대해서만 다루었지만, 다른 디바이스에도 적용 가능한 방법을 탐구할 예정이다. 디바이스의 종류와 상관없이 모델을 쉽게 실행시키고 배포할 수 있는 기술을 개발하여 보다 효율적인 구조 작업에 활용하고자 한다.

5. 참고 문헌

[1] A. Khamparia, D. Gupta, N. G. Nguyen, A. Khanna, B. Pandey and P. Tiwari, "Sound Classification Using Convolutional Neural Network and Tensor Deep Stacking Network," in IEEE Access, vol. 7, pp. 7717-7727, 2019, doi: 10.1109/ACCESS.2018.2888882.

[2] A. Brendel and W. Kellermann, "Distributed Source Localization in Acoustic Sensor Networks Using the Coherent-to-Diffuse Power Ratio," in IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 1, pp. 61-75, March 2019, doi: 10.1109/JSTSP.2019.2900911.