

# 드론 소리 분석을 위한 Convolutional Neural Networks

김다은<sup>1</sup>, 김현중<sup>2</sup>, 박성욱<sup>1</sup>, 이석훈<sup>1</sup>, Shulin Li<sup>3</sup>, Kyle Moffitt<sup>3</sup>, John C. Gallagher<sup>4</sup>

<sup>1</sup>충남대학교 컴퓨터공학과

<sup>2</sup>충남대학교 메카트로닉스공학과

<sup>3</sup>퍼듀대학교 컴퓨터정보기술학과

<sup>4</sup>라이트주립대학교 컴퓨터공학과

[kde0820@gmail.com](mailto:kde0820@gmail.com), [hyujong96@gmail.com](mailto:hyujong96@gmail.com), [parksu0199@gmail.com](mailto:parksu0199@gmail.com),

[sukhoon0975@gmail.com](mailto:sukhoon0975@gmail.com), [li2602@purdue.edu](mailto:li2602@purdue.edu), [kmoffittx@gmail.com](mailto:kmoffittx@gmail.com), [john.gallagher@wright.edu](mailto:john.gallagher@wright.edu)

## Convolutional Neural Networks for Analyzing Unmanned Aerial Vehicles Sound

Daeun Kim<sup>1</sup>, HyunJong Kim<sup>2</sup>, SungWook Park<sup>1</sup>, Sukhoon Lee<sup>1</sup>, Nancy Li<sup>3</sup>, Kyle Moffitt<sup>3</sup>, John C.

Gallagher<sup>4</sup>

<sup>1</sup>Department of Computer Science and Engineering, Chungnam National University

<sup>2</sup>Department of mechatronics Engineering, Chungnam National University

<sup>3</sup>Department of Computer information and Technology, Purdue University

<sup>4</sup>Department of Computer Science and Engineering, Wright State University

### 요약

컨볼루셔널 뉴럴넷(Convolutional Neural Network)은 딥러닝의 여러 모델 중 하나로 이미지 인식, 자연어 처리 등 다양한 분야에서 사용되고 있다. 본 연구에서는 컨볼루셔널 뉴럴넷(Convolutional neural network)을 이용하여 드론에서 생성되는 소리를 기반으로 해당 드론의 위험 물질 탑재 여부를 판별하는 시스템을 설계한다. 본 논문에서는 녹음된 드론의 소리를 고속 푸리에 변환(FFT) 과 Mel 주파수 캡스트럼(MFCC)을 통해 스펙트럼 데이터로 전처리한 후 CNN 모델의 입력값으로 주어 학습시키고, 테스트를 통해 결과를 도출하였다.

### 1. 서론

기술의 발전에 따라 드론의 보급도 점차 활성화 되고, 물건을 운반하는 영역에도 드론이 사용되기 시작했다.<sup>[1]</sup> 이에 따라 드론을 통한 택배 서비스까지 구상할 수 있게 되었고, 편리성이 증대될 것이라 예상 가능하게 되었다. 그러나 이러한 드론의 보급과 발전은 위험성 또한 증대시켰다. 폭발물 등의 위험물을 실은 드론을 통하여 테러리스트, 범죄자 및 악의적인 사용자가 쉽게 공격을 가할 수 있게 되었다. 그렇기에 이러한 위협으로부터 사람들을 보호하기 위한 수단의 필요성이 증가하고 있다.

드론의 소리를 통해 위험성을 판단하는 연구는 이미 진행된바 있다.<sup>[2],[3],[4]</sup> 지난 연구 또한 FFT를 통해 드론의 소리를 스펙트럼으로 변환 후 CNN 모델에 입력 값으로 주어 위험성 여부를 판단한다.<sup>[5]</sup> 다만 드론의 소리를 스펙트럼으로 변환하여 입력 값으로 줄 때 그 범위가 적다는 것이다. 입력값의 범위가 적다는 것은 소음에 큰 영향을 받는다는 것을 의미한다. 이번 연구는 이러한 점을 해결하기 위해서 드론의 소리를 약 5~10초 정도의 크기를 기준으로 데이터를 샘플링하여 일시적인 소음에는 영향을 크게 받지 않도록 하였다.

5~10초를 기준으로 샘플링한 데이터의 크기가 무수히 많아 머신러닝의 트레이닝에 많은 시간을 소비하게 된다. 상대적으로 많은 시간을 소비하는 것을 방지하기 위해 Mel 주파수 캡스트럼(MFCC)를 사용하였다. MFCC를 통해 데이터의 특징을 추출할 수 있기 때문에 데이터의 크기를 1/90로 줄일 수 있고, 그러므로 트레이닝의 시간 또한 줄일 수 있다.

위의 가공시킨 데이터를 CNN 모델에 학습시켜 “적재된 드론”, “적재되지 않은 드론”, “소음”으로 구분 할 수 있다. 이 중 “적재되었지만 적재 되지 않았다”, “적재되었지만 소음”으로 도출되는 결과가 나오는데, 이러한 것을 부정오류라고 할 수 있다. 부정오류의 경우 위험을 발생시킬 가능성이 크기 때문에 이러한 점을 최소화 하기 위해 가중치를 사용하였다.

2. 연구방식

2.1 데이터 수집

수집한 데이터의 표준 드론 모델은 DJI Phantom2이다. 무게 1000g, 이륙 중량 1300g 이하, 최대 비행 속도 15m/s의 성능을 나타내고 있다.

데이터의 수집은 비행장에서 진행되었으며, 주로 적재되지 않았을 때의 Phantom2에 대한 여러 움직임에 대한 소리, 적재 되었을 때의 움직임에 대한 소리 및 그 외 드론 소리가 들어가지 않은 여러 배경 및 소음에 대하여 데이터 수집을 하였다.

2.2 데이터 전처리

우리는 수 많은 데이터를 효율적으로 모델에 적용하기 위해서 수집된 데이터에 몇가지 전처리를 수행하였다.

첫 번째는 정규화 처리를 하였다. 정규화된 입력 데이터를 사용하면, 경사하강법에 대해서 빠르게 최적화 지점을 찾을 수 있다는 장점을 가지게 된다. 또한 학습률을 적게 설정해야한다는 단점을 없애 한번에 많은 학습을 수행할 수 있다.

두 번째는 Pre-emphasis Filter를 적용하였다. 후에 푸리에 변환을 사용하면 저주파수 성분에서의 주파수 반응이 고주파수 성분에서의 주파수 반응보다 훨씬 크게 나오므로 Pre-emphasis Filter를 사용하여 모든 주파수 대역에서 거의 일정한 주파수 반응을 갖도록 하였다.

세 번째는 STFT(단시간 푸리에 변환)를 적용하였다. STFT 사용시 윈도우의 종류는 Hamming, 길이는 863로 사용했으며 이는 44.1kHz에서 0.0116초에 해당한다. 이를 이용하여 음성 데이터를 2차원 스펙트로그램으로 변환한다. 즉 STFT는 CNN 모델에 맞는 입력 데이터 형식으로 바꾸어주는 역할을 하며 또한 시간 도메인에서 불가능하던 소리 분석을 주파수 도메인을 추가해 가능하게 한다.

네 번째는 MFCC이다. STFT를 통해 얻어진 데이터는 너무 많은 데이터를 가지고 있다. 우리가 필요한 소리도 들어있지만 또한 Noise도 들어있으며 주파수 데이터가 1024개가 나오도록 했기때문에 데이터 양도 많고 쓸데없는 데이터 또한 섞여있는데 MFCC를 사용해 특징추출을 하여 데이터의 수를 약 1/90로 줄일 수 있다.

2.3 Convolutional Neural Network

우리가 구축한 모델은 녹음된 소리들을 적재된 드론, 적재되지 않은 드론, 소음 이렇게 세 종류로 분류하는 것으로 CNN을 사용하였다.

CNN 모델은 다음과 같은 Layer들로 구성되어져 있다.

1_Convolutional	window size: 2x4x1 number of filter : 32
1_Relu	activation function
1_MaxPool	kernel size : [1,2,4,1] stride : [1,2,4,1]
2_Convolutional	window size: 2x9x32 number of filter : 64
2_Relu	activation function

2_MaxPool	kernel size : [1,1,2,1] stride : [1,1,2,1]
3_Fully Connected Neural Network	input size: 24192 output size : 100
3_Relu	activation function
3_DropOut	keeping rate : 70%
4_Fully Connected Neural Network	input size : 100 output size : 3

실제로는 적재되었지만 적재 되지 않았다고 분류 하거나 소음으로 분류하는 경우를 부정오류라고 할 수 있다. 부정 오류가 발생하면 위험을 유발할 가능성이 크기 때문에 부정 오류를 최소화 해야한다. 따라서 CNN 모델로 학습시킬 때 부정 오류의 결과에 대한 가중치를 더 크게 주도록 하여 학습을 할 때 효과적으로 cost를 조정하게 되어 부정 오류가 발생할 확률을 낮추도록 하였다.

3. 실험 결과

학습 데이터들과 똑같은 방식으로 처리된 56000개의 데이터들로 테스트를 수행하였다. 99.5%의 정확도를 얻을 수 있었으며 그 중 0.005%의 부정오류가 발생하였다.

	소음	적재되지 않은 드론	적재된 드론
소음	46766	60	174
적재되지 않은 드론	0	5000	0
적재된 드론	3	0	3997

4. 결론

기존 CNN 모델과 데이터 전처리를 통하여 드론의 적재 상태를 구분 할 수 있음을 알 수 있었다. 더욱이 저번 연구에서 문제가 될 수 있던 부정 오류에 대한 도출이 가중치를 통해 해결 할 수 있음을 확인하였다. 이후 연구를 발전시켜 휴대폰이나 휴대용 기기를 통하여 음성을 녹음하여, 실시간으로 그 음성에 대하여 적재 여부를 판단할 수 있도록 할 것이다.

## 참 고 문 헌

- [1] Seung Jin Shin, 2016, "Possibility of logistics delivery through drone (drone) and overseas case," Monthly KOTI Magazine on Transport, pp. 45~53.
- [2] Dong-Hyun Lee, 2017, "Convolutional Neural Network-based Real-Time Drone Detection Algorithm," The Journal of Korea Robotics Society, Vol. 12, No. 4, pp. 425~431.
- [3] Kee-Woong Lee, Kyoung-Min Song, Jung-Hwan Song, Chul-Ho Jung, Woo-kyung Lee, Myeong-Jin Lee and Yong-Kyu Song, 2017, "Implementation of Radar Drone Detection Based on ISAR Technique," The Journal of Korean Institute of Electromagnetic Engineering and Science, Vol. 28, No. 2, pp. 159~162.
- [4] Lin Shi, Yujing He and KyungHi Chang, 2016, "Acoustic-based Classification for Drone Identification using Hidden Markov Model with MFCC Technique," Proceedings of Symposium of the Korean Institute of communications and Information Sciences, , pp. 204~205.
- [5] SangGi Hong, GaYoung Kim, HeonGyeom Kim, SangHee Lee, DongHyun Lim, Austin Snair, Lucy Gotwals and Jone C. Gallagher, 2017, "Practically Classifying Unmanned Aerial Vehicles Sound using Convolutional Neural Networks," Proceedings of the Korean Information Science Society, pp. 1856~1858.