Deep Learning Based Malicious Drone Detection Using Acoustic & Image Data

CONTENTS

MOTIVATION

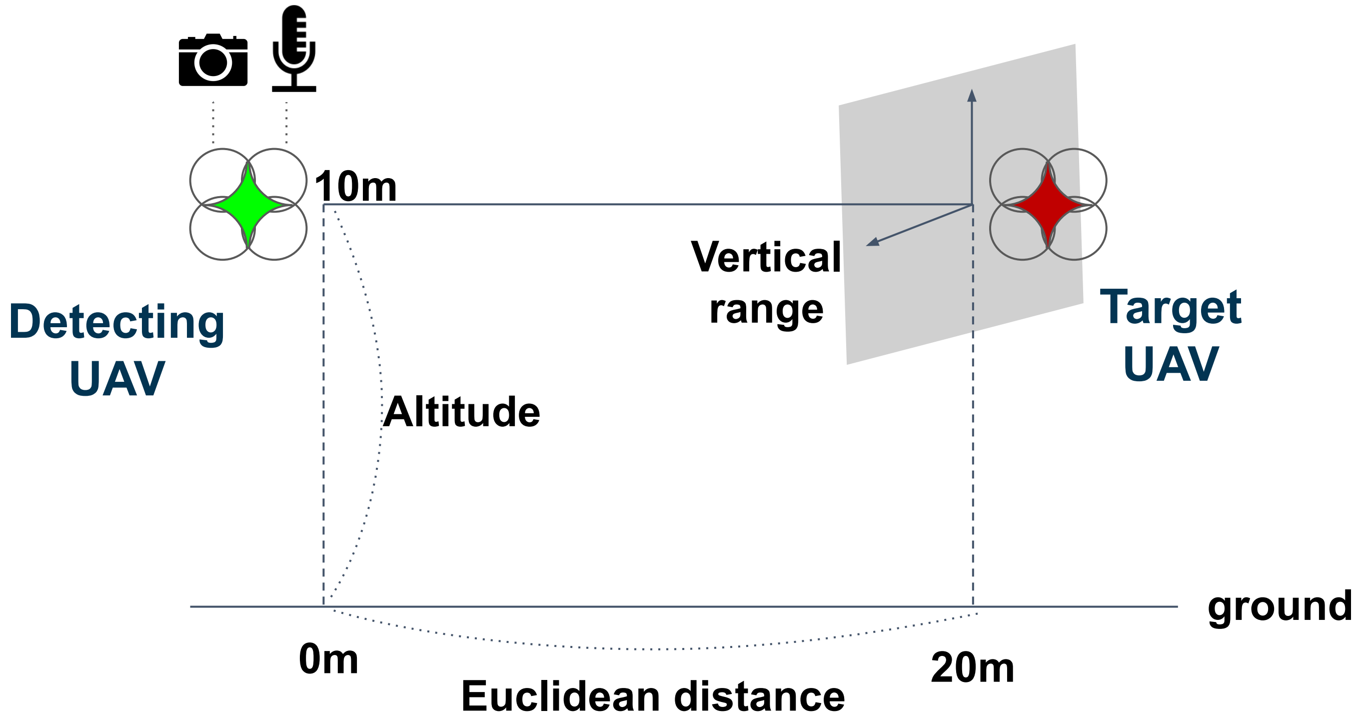
Drone Delivery Service

Counter UAV

RELATED WORK

	Advantages	Disadvantages
Heat (Thermal)	- Less affected by weather - Long range	- Low accuracy
RF Signal	- Obstacle-free	- Autonomous flight
Radar	- Less affected by weather - Long range	- High expense - Vulnerable to obstacle
Camera	- Low expense - Identification	- Highly affected by the weather - Vulnerable to obstacle
Microphone	- Miniaturized	- Low detection range

ENVIRONMENT SETTING 1



ENVIRONMENT SETTING 2

DJI Mavic 2 Pro
(detecting drone)

DJI Matrice 200 V2
(target drone)

ENVIRONMENT SETTING 3

Recording audio sound

Attached iPhone 6 using tapes

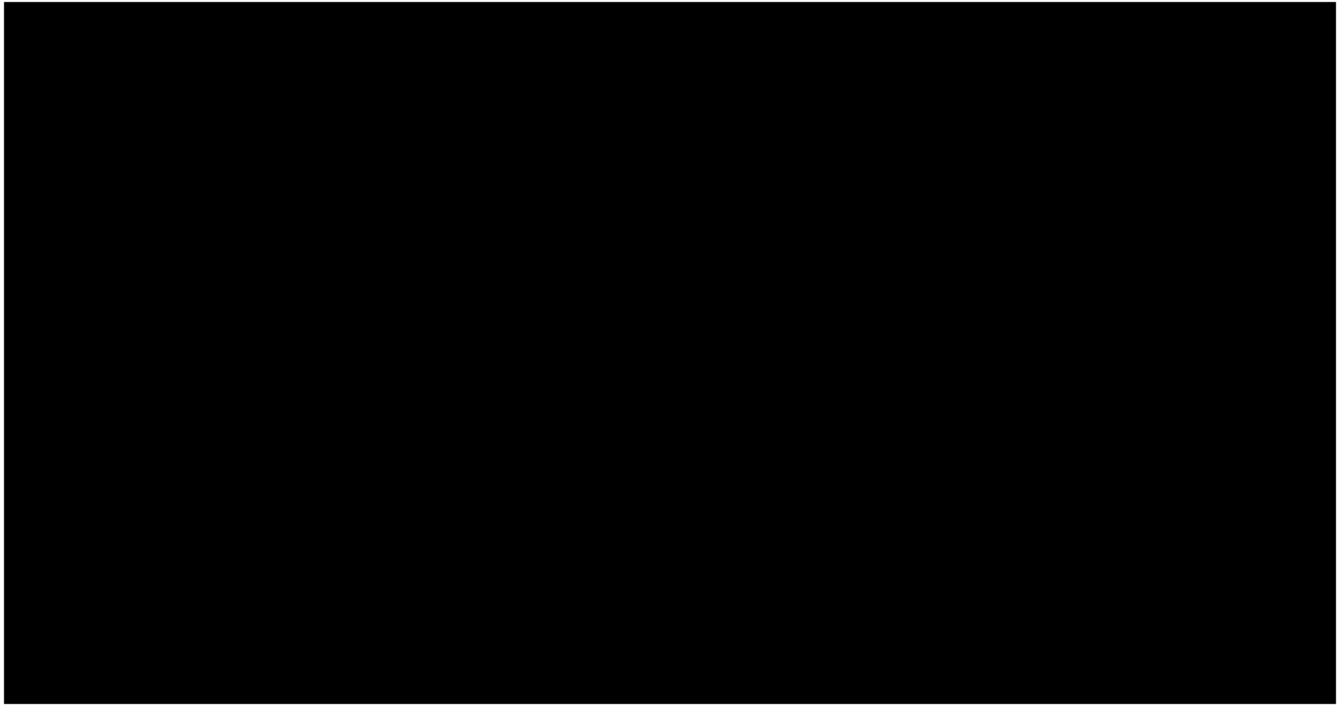
Filming video

WEATHER CONDITION

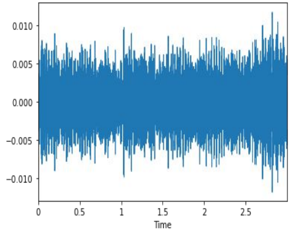
Foggy

Sunny

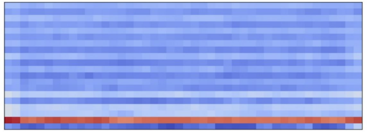
DATA COLLECTION



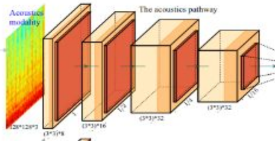
SYSTEM OVERVIEW



Acoustic Data



MFCC

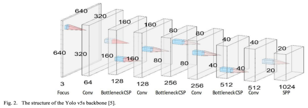


Classification (CNN)

Decision Fusion (OR)

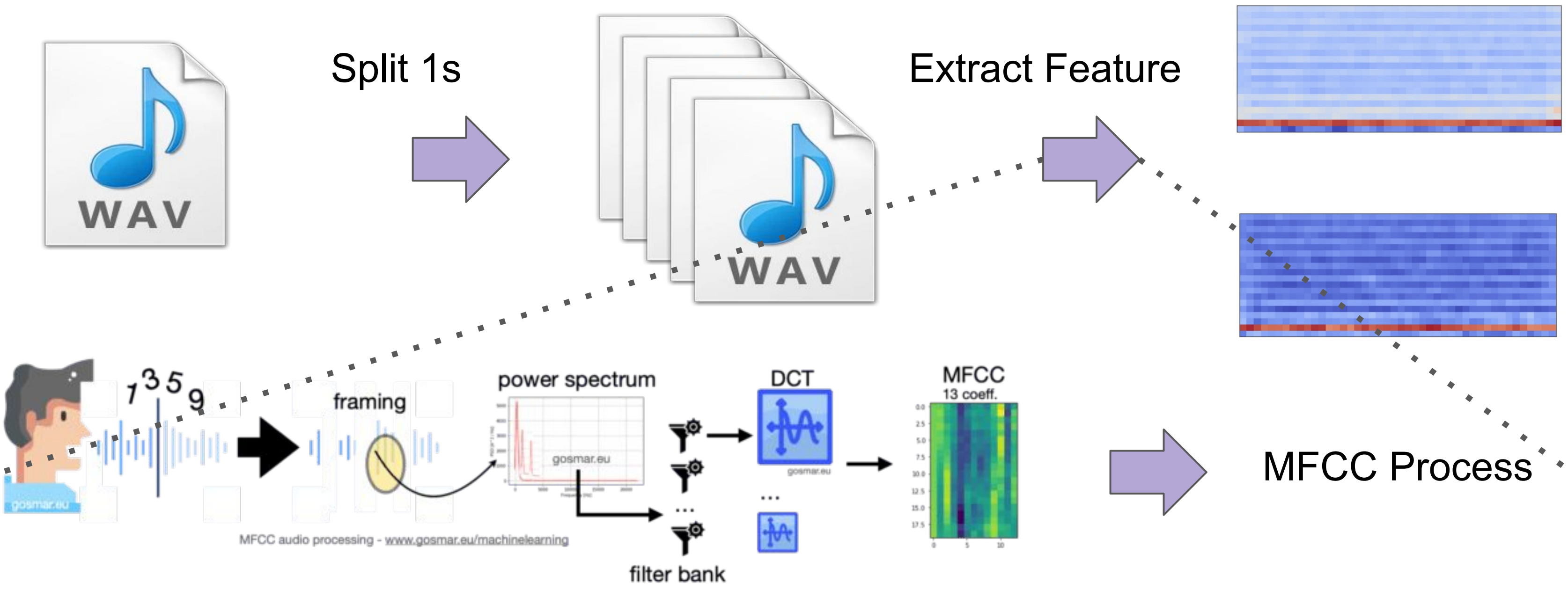
Image Data

Image Feature Map

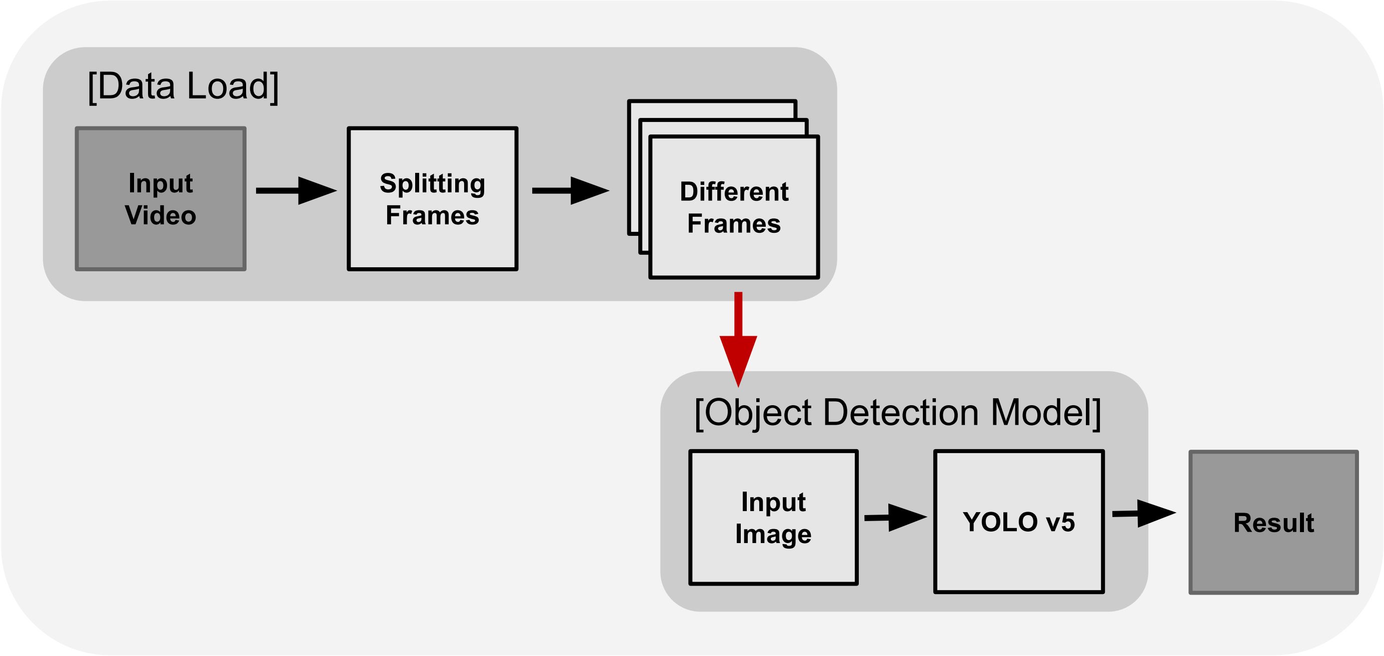


YOLOv5 (Classifier)

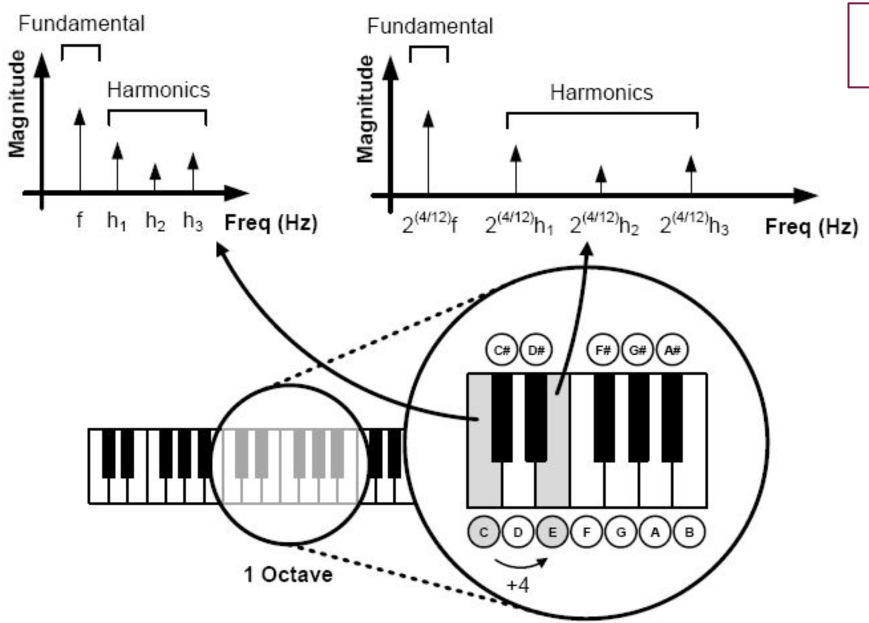
PREPROCESSING (A)



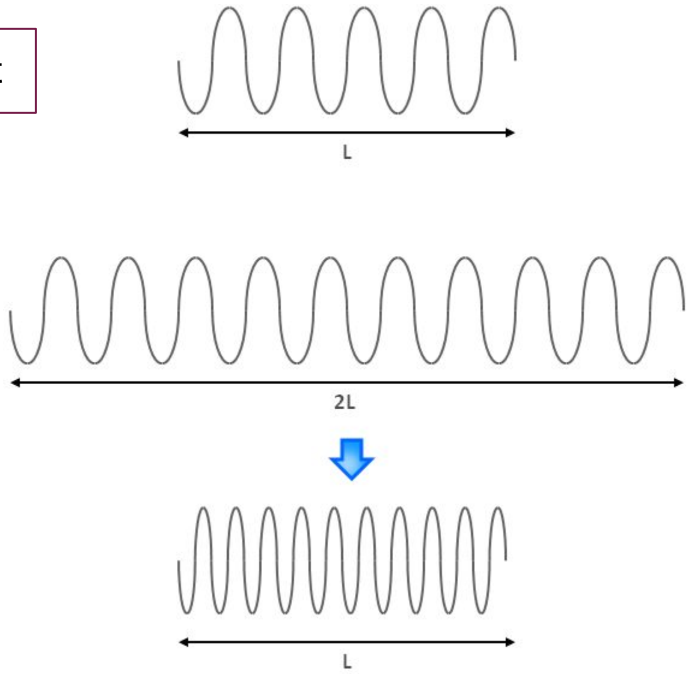
PREPROCESSING (V)



DATA AUGMENTATION (A)



Pitch Shift



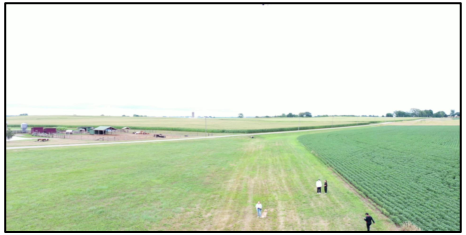
DATA AUGMENTATION (V)

blur | **horizontal flip** | **noising**

origin



modified



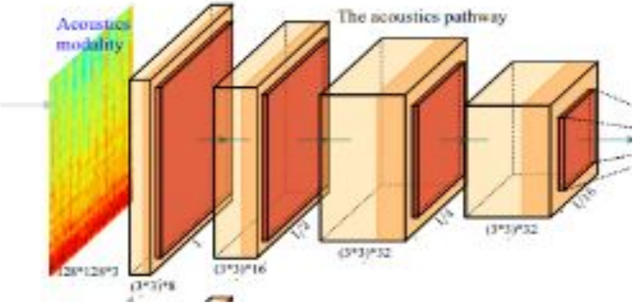
DATASET

Type of data	Class	Audio	Image	Augmented	Total times (s)
Train	drone	1055	1055	1055	4220
	no drone	1055	1055	1055	
Validation	drone	300	300	300	1200
	no drone	300	300	300	
Test	drone	154	154	-	308
	no drone	154	154	-	

Train	Val	Test
70%	20%	10%

TRAIN & TEST (A)

Layer Type	Output Shape	Parameters
Conv2D	(None, 44, 80, 32)	832
Conv2D	(None, 44, 80, 32)	25632
MaxPooling2D	(None, 22, 40, 32)	0
Dropout	(None, 22, 40, 32)	0
Conv2D	(None, 22, 40, 64)	18496
Conv2D	(None, 22, 40, 64)	36928
MaxPooling2D	(None, 11, 20, 64)	0
Dropout	(None, 11, 20, 64)	0
Flatten	(None, 14080)	0
Dense	(None, 256)	3604736
Dropout	(None, 256)	0
Dense	(None, 2)	514



light weight CNN

- iPhone6 -

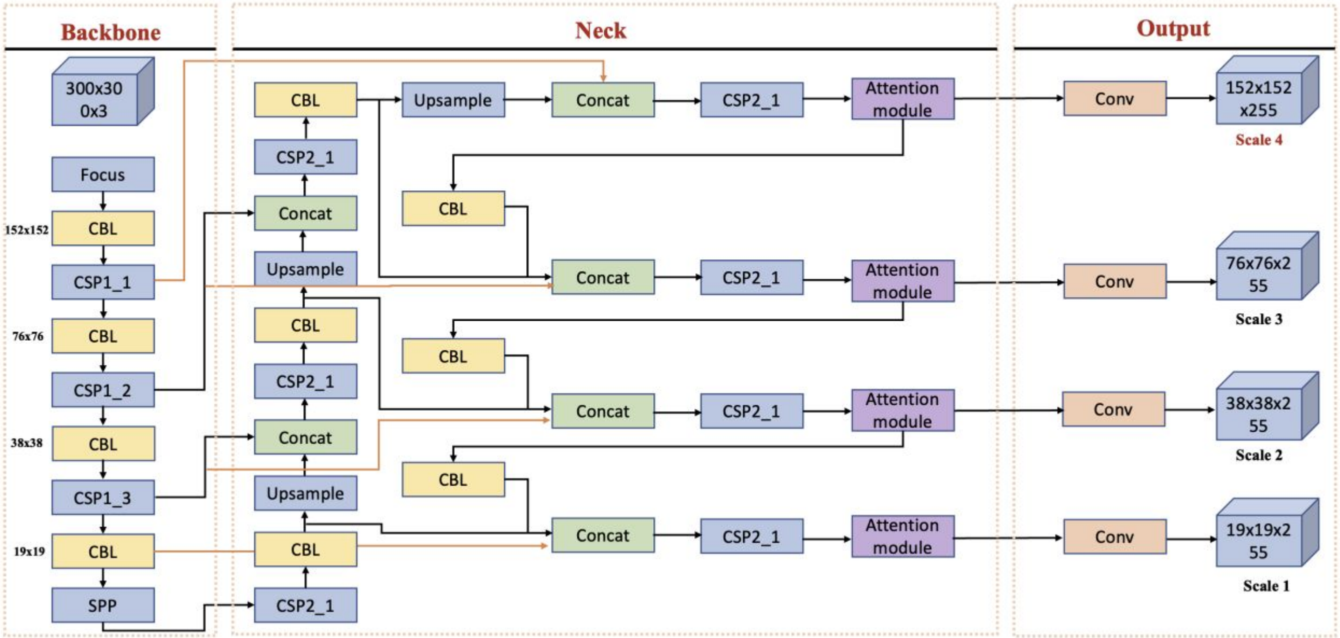
TRAIN & TEST (A)

n_mfcc

	20	40	80	120
sigmoid	*86.6%*	*88.6%*	*88.6%*	*88.6%*
softmax	*87.0%*	*87.3%*	*87.9%*	***88.9%***

activation func.

TRAIN & TEST (V)



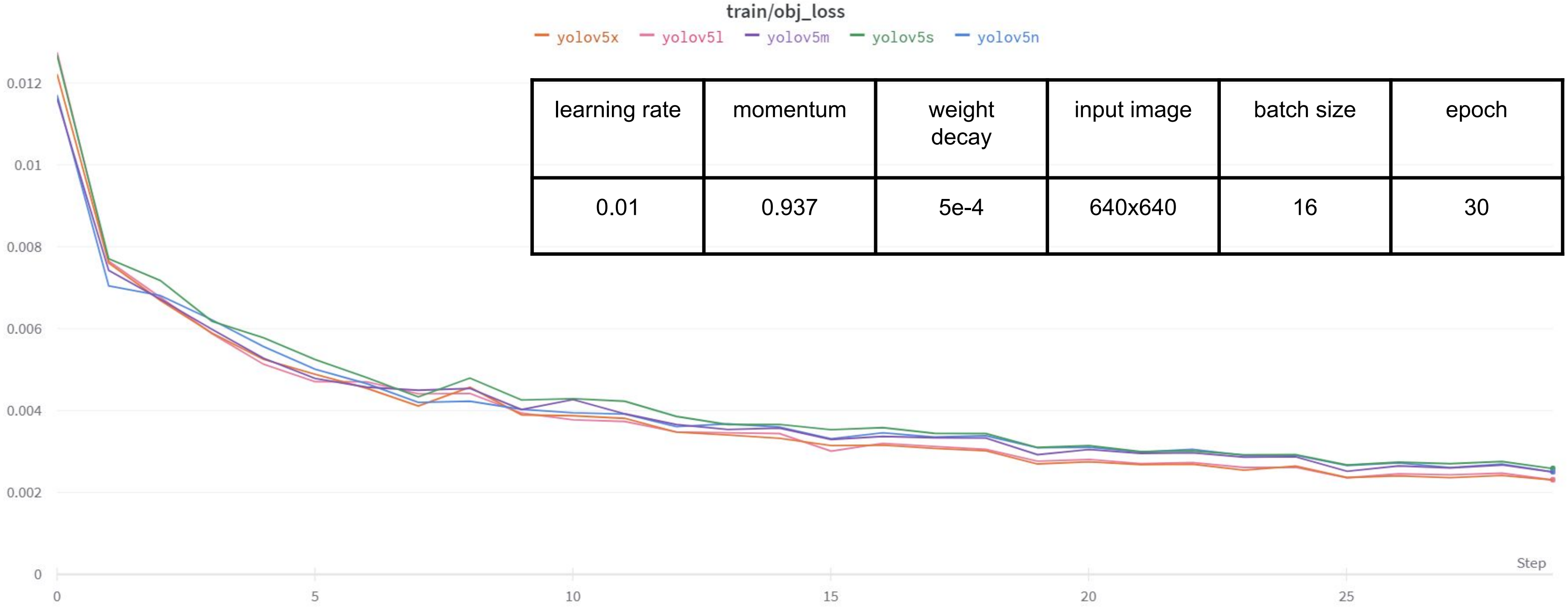
Backbone
Neck
Output
300x300x3
Focus
152x152
CBL
CSP1_1
76x76
CBL
CSP1_2
38x38
CBL
CSP1_3
19x19
CBL
SPP
CBL
CSP2_1
Concat
Upsample
CBL
CSP2_1
Concat
Upsample
CBL
CSP2_1
Upsample
Concat
CSP2_1
Attention module
CBL
Concat
CSP2_1
Attention module
CBL
Concat
CSP2_1
Attention module
CBL
Concat
CSP2_1
Attention module
Conv
152x152x255
Scale 4
Conv
76x76x255
Scale 3
Conv
38x38x255
Scale 2
Conv
19x19x255
Scale 1

TRAIN & TEST (V)


Nano
YOLOv5n
Small
YOLOv5s
Medium
YOLOv5m
Large
YOLOv5l
XLarge
YOLOv5x

TRAIN & TEST (V)

learning rate	momentum	weight decay	input image	batch size	epoch
0.01	0.937	5e-4	640x640	16	30



TRAIN & TEST (V)

Train result

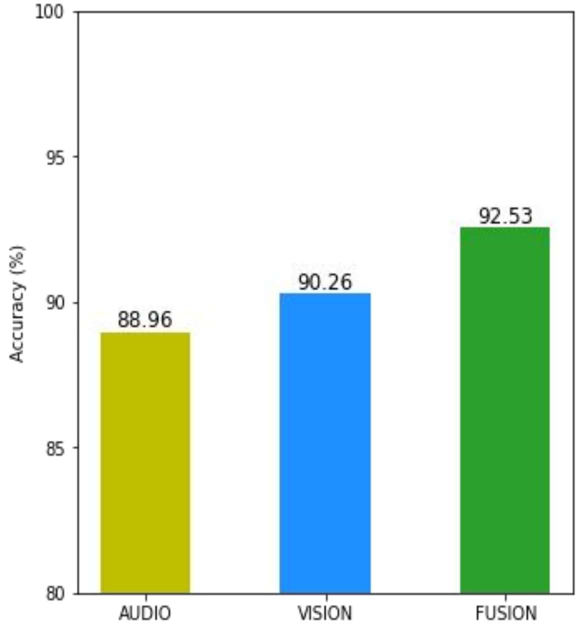
Models	mAP_0.5	mAP_0.5:0.95	Precision	Recall	F1-score
YOLOv5n	0.840	0.390	0.780	0.861	0.82
YOLOv5s	0.870	0.377	0.763	0.870	0.81
YOLOv5m	0.860	0.378	0.806	0.944	0.81
YOLOv5l	0.821	0.358	0.784	0.991	0.79
YOLOv5x	0.851	0.372	0.0.835	0.991	0.80

Test result

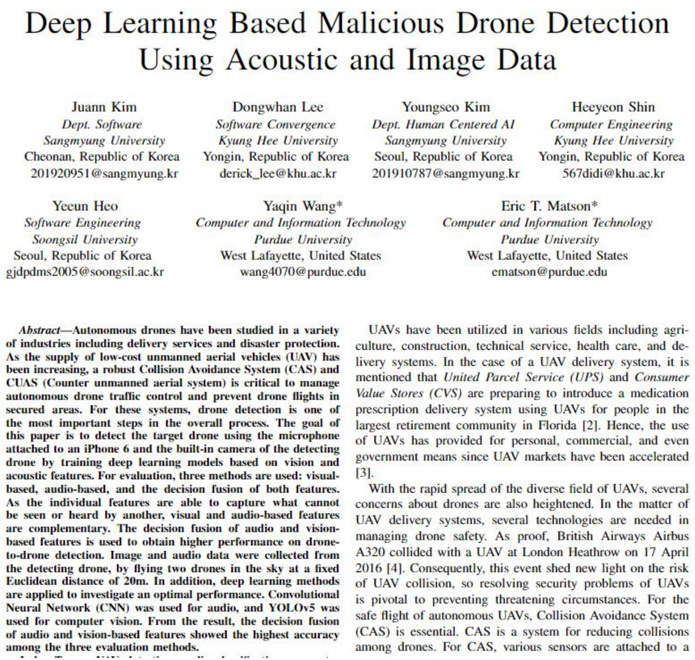
Models	mAP_0.5	mAP_0.5:0.95	Precision	Recall	F1-score
YOLOv5n	0.904	0.574	0.940	0.696	0.82
YOLOv5s	0.922	0.583	0.855	0.824	0.81
YOLOv5m	0.904	0.570	0.793	0.809	0.81
YOLOv5l	0.822	0.602	0.694	0.971	0.79
YOLOv5x	0.902	0.636	0.669	0.926	0.80

RESULT

	Acoustic	Vision	Audio + Vision
Drone Detection Accuracy	*88.96%*	*90.26%*	***92.53%***



PAPER



Deep Learning Based Malicious Drone Detection Using Acoustic and Image Data

Juann Kim
Dept. Software
Sangmyung University
Cheonan, Republic of Korea
201920951@sangmyung.kr

Dongwhan Lee
Software Convergence
Kyung Hee University
Yongin, Republic of Korea
derick_lee@khu.ac.kr

Youngseo Kim
Dept. Human Centered AI
Sangmyung University
Seoul, Republic of Korea
201910787@sangmyung.kr

Heeyeon Shin
Computer Engineering
Kyung Hee University
Yongin, Republic of Korea
567didi@khu.ac.kr

Yeeun Heo
Software Engineering
Soongsil University
Seoul, Republic of Korea
gjdpdms2005@soongsil.ac.kr

Yaqin Wang*
Computer and Information Technology
Purdue University
West Lafayette, United States
wang4070@purdue.edu

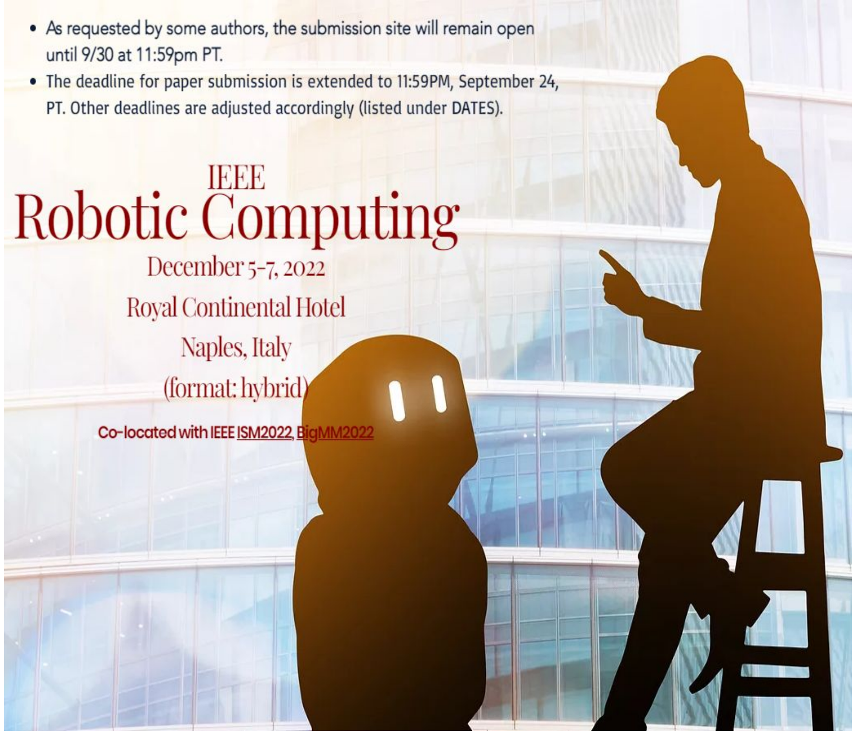
Eric T. Matson*
Computer and Information Technology
Purdue University
West Lafayette, United States
ematson@purdue.edu

***Abstract*—Autonomous drones have been studied in a variety of industries including delivery services and disaster protection. As the supply of low-cost unmanned aerial vehicles (UAV) has been increasing, a robust Collision Avoidance System (CAS) and CUAS (Counter unmanned aerial system) is critical to manage autonomous drone traffic control and prevent drone flights in secured areas. For these systems, drone detection is one of the most important steps in the overall process. The goal of this paper is to detect the target drone using the microphone attached to an iPhone 6 and the built-in camera of the detecting drone by training deep learning models based on vision and acoustic features. For evaluation, three methods are used: visual-based, audio-based, and the decision fusion of both features. As the individual features are able to capture what cannot be seen or heard by another, visual and audio-based features are complementary. The decision fusion of audio and vision-based features is used to obtain higher performance on drone-to-drone detection. Image and audio data were collected from the detecting drone, by flying two drones in the sky at a fixed Euclidean distance of 20m. In addition, deep learning methods are applied to investigate an optimal performance. Convolutional Neural Network (CNN) was used for audio, and YOLOv5 was used for computer vision. From the result, the decision fusion of audio and vision-based features showed the highest accuracy among the three evaluation methods.**

UAVs have been utilized in various fields including agriculture, construction, technical service, health care, and delivery systems. In the case of a UAV delivery system, it is mentioned that *United Parcel Service (UPS)* and *Consumer Value Stores (CVS)* are preparing to introduce a medication prescription delivery system using UAVs for people in the largest retirement community in Florida [2]. Hence, the use of UAVs has provided for personal, commercial, and even government means since UAV markets have been accelerated [3].

With the rapid spread of the diverse field of UAVs, several concerns about drones are also heightened. In the matter of UAV delivery systems, several technologies are needed in managing drone safety. As proof, British Airways Airbus A320 collided with a UAV at London Heathrow on 17 April 2016 [4]. Consequently, this event shed new light on the risk of UAV collision, so resolving security problems of UAVs is pivotal to preventing threatening circumstances. For the safe flight of autonomous UAVs, Collision Avoidance System (CAS) is essential. CAS is a system for reducing collisions among drones. For CAS, various sensors are attached to a

https://www.overleaf.com/project/62854acaf9acdaebd73333bf



https://www.ieee-irc.org/

PAPER 2

Drone Detect: How Far Can a Drone be Detected?

Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

Yaqin Wang*
Computer and Information Technology
Purdue University
West Lafayette, United States
wang4070@purdue.edu

Eric T. Matson*
Computer and Information Technology
Purdue University
West Lafayette, United States
ematson@purdue.edu

***Abstract*—The recent success of the drone detection model shows that leveraging the decision fusion of audio-based feature and vision-based feature can make high accuracy rather than using unitary features. However, they aim to detect the drone at the same location and fixed distance without any verification in the air. In this paper, we propose to figure out how does it appearance the drone detection performance ability depending on the distance. To deal with this problem, we implement two main strategy. First, we collect the high quality of drone-to-drone audio and video data manually. It should be able to great assistance about drone detection task from now on. Second, we make a various experiment attempt using deep learning techniques in order to gain adequate result.* Especially, Convolutional Neural Network (CNN) was utilized in the modeling of audio feature extraction, and YOLOv5 was adopted for visual feature extraction and drone detection in image. Finally, our model allows to comes out the infallible result that drone detection accuracy according to different distance.**

***Index Terms*—UAV detection, audio classification, object detection, collision avoidance system, deep learning, decision fusion**

I. INTRODUCTION

Overall, the main contributions of this work can be summarized as follows:

- Establishment of foundation for building CUAS system.
- We gather the high quality of drone-to-drone audio and video data that was collected by distance of 20 to 100 meters manually.
- In order to gain rational result of drone detection performance by distance, a number of experiments were attempted with deep learning technology.
- We propose a novel drone detection scheme that reduces the error rate, which using decision fusion.

II. RELATED WORK

A. RF & Radar

Currently, various methods have been used for drone detection and drone localization, including radar, Lidar, Computer vision, and Acoustic sensors. More specifically, radar is widely used for binary drone detection, classification of drones and birds, and multi-drone detection. In [3], distance estimation or drone localization was done in two ways. Firstly, the implemented FMCW radar system result with only one drone showed the maximum distance of the drone from the radar system that could be detected was greater than about 1005 to 1010 m. Meanwhile, when the two drones were flying at the same time, one frame of the detection results at the range of around 339 m. Thus, when comparing two results, the distance of drone-to-drone detection using radar is shorter than using only one drone flying. However, radar-based detection is not optimized for plastic material drone detection and small drone detection at widely varying ranges. Also, radar is high cost sensor and harmful to the environment. In this paper, low-cost products are used: a built-in camera for iPhone6 for audio-based features.

B. Deep learning - Audio

A radar system has a small cross-section, and RF-based systems do not operate well when GPS communication signals are small; therefore their performances are limited. However, the microphone array overcomes the shortcomings of the sensors and shows excellent performance in drone localization and drone tracking. In[17], four microphone sensors were used to predict the direction of drone arrival (DOA), and localization is performed by obtaining azimuth and elevation angles by a multi-signal classification algorithm(MUSIC). In fact, it showed a very low performance. Meanwhile, in [18-19], Acoustic Aircraft Detection (AAD) systems were developed and built. This system can detect and track small airplanes and helicopters, whereas it does not consider a situation with multiple noises.

C. Deep learning - Vision

Research on drone detection systems using computer vision is one of the traditional methods widely used in the past. Furthermore, research on drone detection systems using computer

*Corresponding authors.

https://www.overleaf.com/project/62854acaf9acdaebd73333bf

https://i2mtc2022.ieee-ims.org/

GANTT

	April	May	June	July	August	~ing
Research Topic						Submission completed
Indoor Test						
Data Collection						
Test & Paper						

THANK YOU