Deep Learning Method of Drone Detection Using Acoustic and Visual Features

K2S3
Dongwhan Lee, Juann Kim, Youngseo Kim
Heeyeon Shin, Yeeun Heo

Contents

Deep Learning Method of
Drone Detection using
Acoustic and Visual Features

1. Team Member

1. Team Member

Leader

Dongwhan Lee

Kyunghee Univ.

Youngseo Kim

Sangmyung Univ.

Juann Kim

Sangmyung Univ.

Heeyeon Shin

Kyunghee Univ.

Yeeun Heo

Soongsil Univ.

2. Introduction

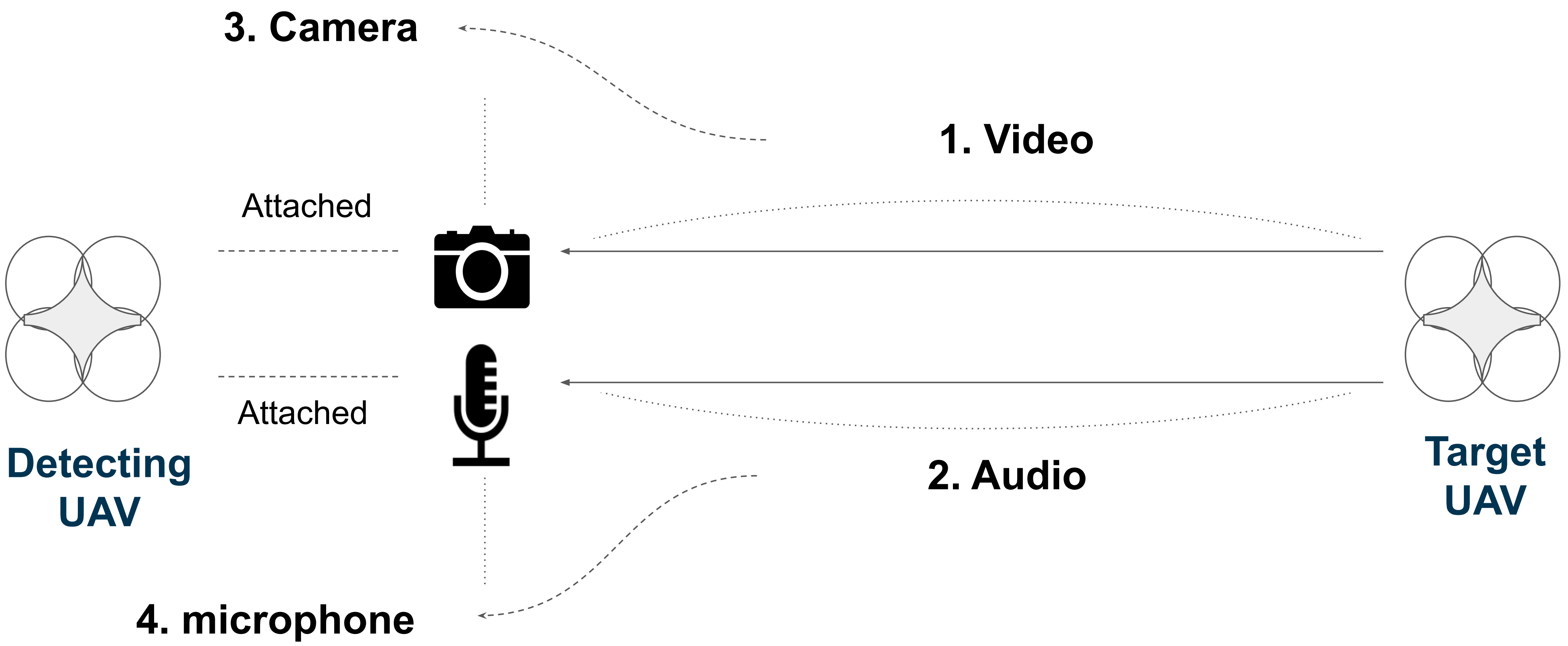
2.1. Aim of the project: (1) CAS

CAS (collision avoidance system)

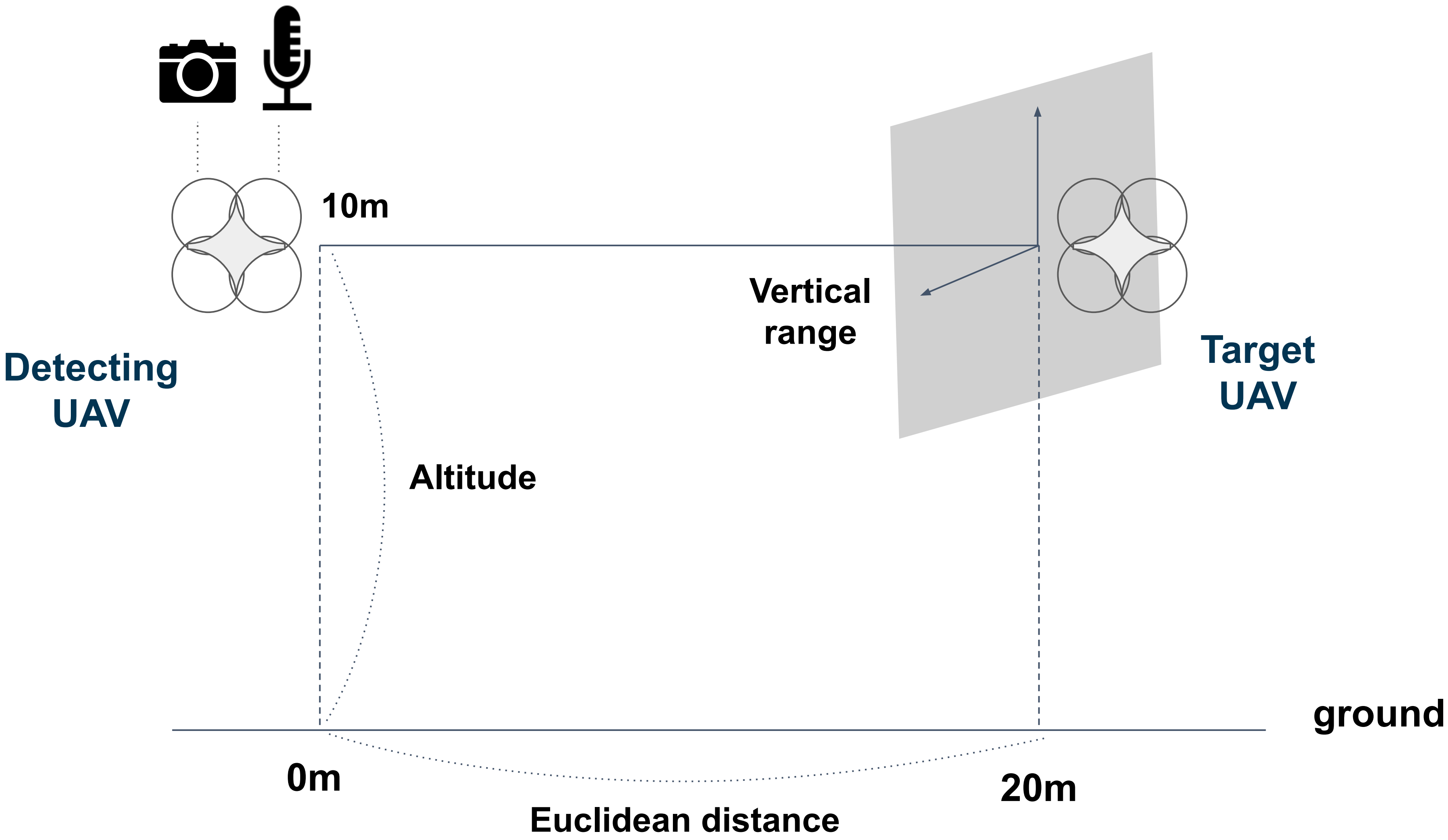
2.2. Aim of the project: (2) CUAS

CUAS (Counter Unmanned Aerial System)

2.2. Process of collecting data



2.3. Visualization of the experiment



3. Outdoor Experiment Setup

3.1. Drone Type
3.2. Method of Collecting Data
3.3. Weather Condition

3.1. Drone Types

DJI Mavic 2 Pro

Target Drone

DJI Matrice 200 V2

Detecting Drone

3.2. Method of collecting data - Audio & Video Files

Recording audio sound

Attached iPhone 6 using tapes

Filming video

3.3. Weather Condition

Foggy Days

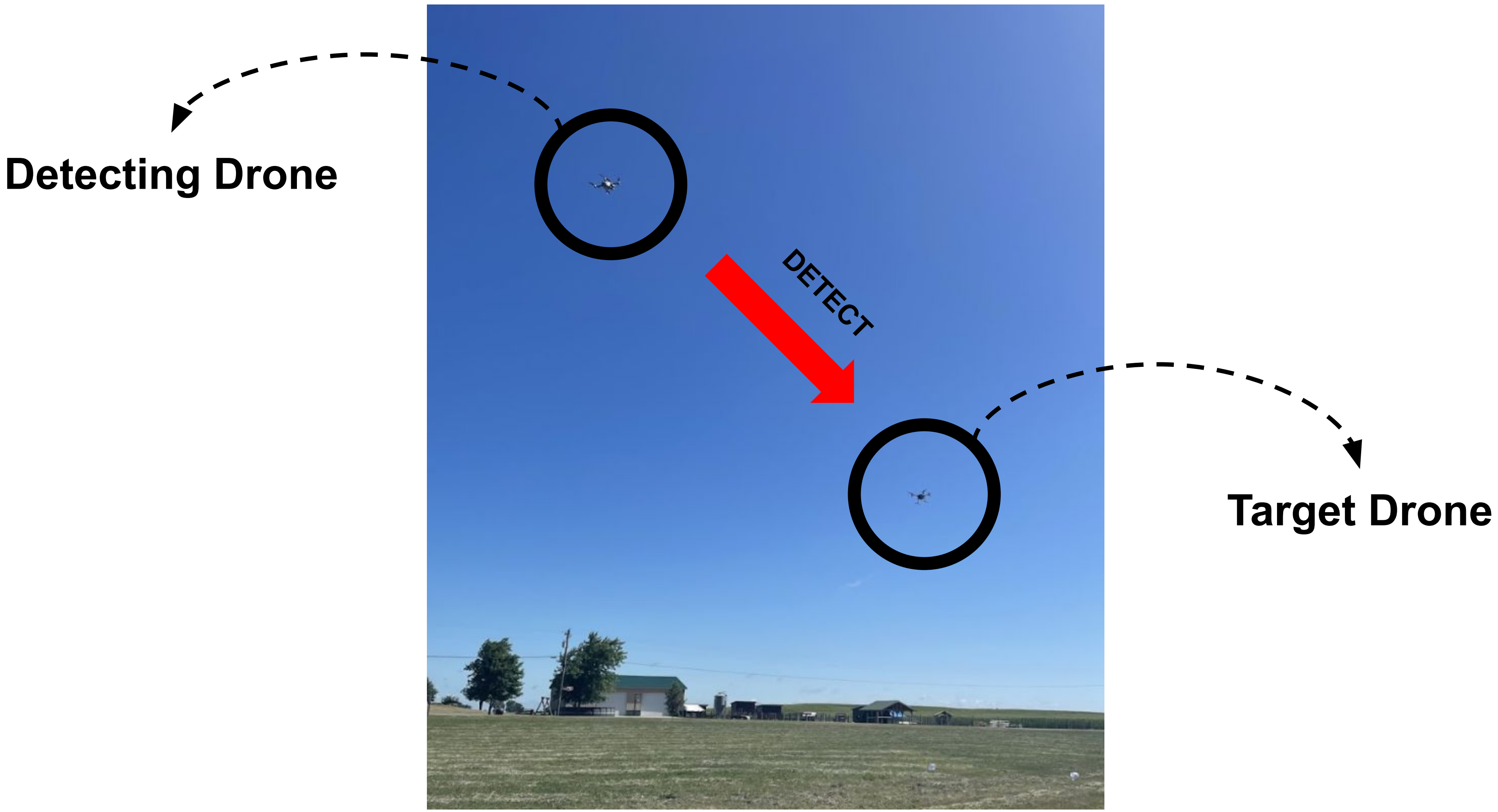
As a dense fog came down,
a target drone could be obscured

Sunny Days

As a view of farm opened out,
a single camera could capture a target drone clearly

3.3. Weather Condition

	Range	Median	Mean	Maximum	Minimum
Temp (°C)	17.2 ~ 29	27	25.65	29	17.2
Wind (mi/h)	2 ~ 12	7	7	12	2
Humidity (°C)	35 ~ 87	64	61.9	87	35
Precipitation (" in last 24h)	0 ~ 0.15	0	0.02	0.15	0


Detecting Drone
DETECT
Target Drone

4.

Experiment Process and Results

Overall Process

4.1.1 Three methods of evaluation
4.1.2 System overview
4.1.3 Total number of dataset

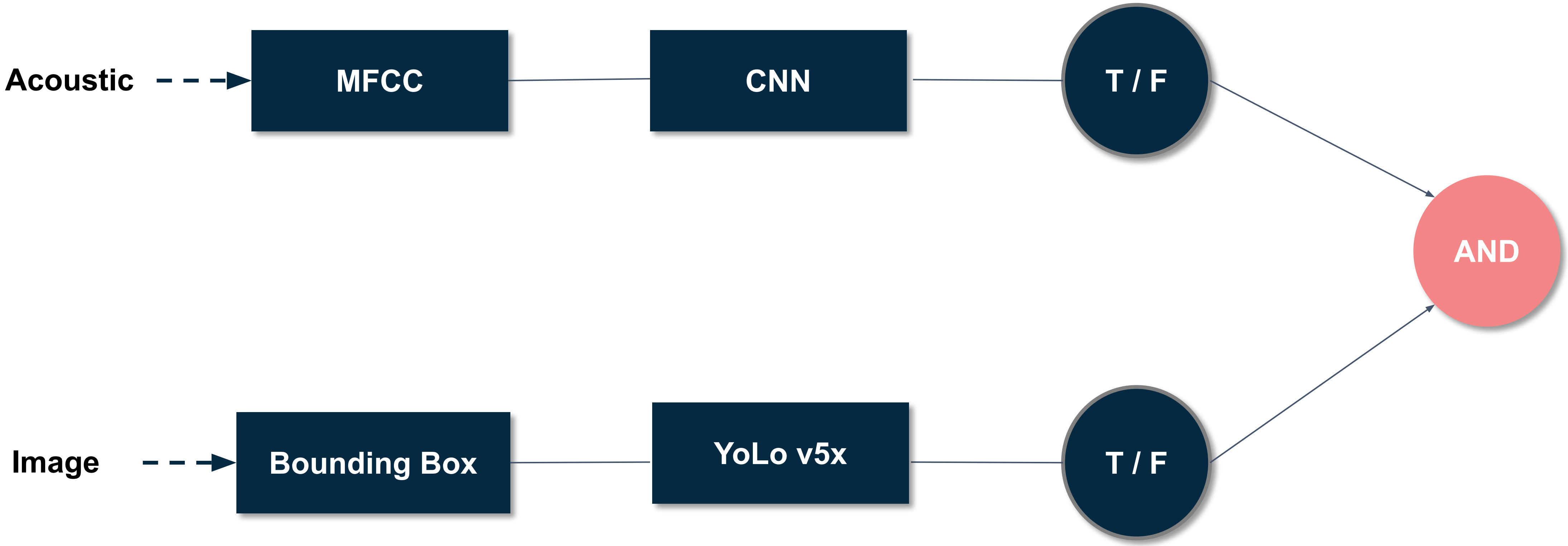
4.1.1. Three Methods of Evaluation

Acoustic-Based
Feature
Binary Classification

Multimodal
of
Acoustic and Visual
Based
Feature
Binary Classification

Visual-Based
Feature
Binary Classification

4.1.2. System Overview



4.1.3. Total Number of Dataset

Types of data	Class	# of Audio files (1sec)	# of Image files	Total Time (s)	Total Time (s)
Training	drone	1055	1055	2110	3026
	no drone	1055	1055		
Validation	drone	300	300	600	
	no drone	300	300		
Test	drone	158	158	316	
	no drone	158	158		

Training	Validation	Test
70%	**20%**	**10%**

Audio Process and Results

4.2.1. Data Processing

4.2.2. MFCC Feature Map

4.2.3. Acoustic Overview

4.2.4. Cross Evaluation - CNN&SVM, MFCC & Sigmoid & 5-CV

4.2.5. Highest Performance

4.3.3. Acoustic Overview

4.2.1 Data Processing

Split into
1 second
wav file

moving_dp20m_Matrice200V2_0726_1_1**.wav**
moving_dp20m_Matrice200V2_0726_1_2.wav
moving_dp20m_Matrice200V2_0726_1_3.wav

.
.
.

4.2.1 Data Processing

Sample of **1S**

Mel Frequency Cepstral Coefficients

MFCC

Mel Frequency
Cepstral Coefficients

Time
(=3018)

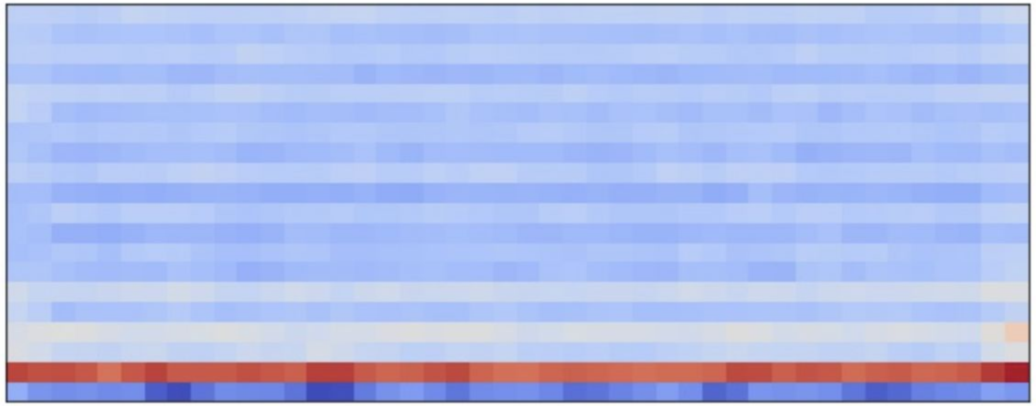
Feature
(n_mfcc = 20, 40)

2D	Y	
X		

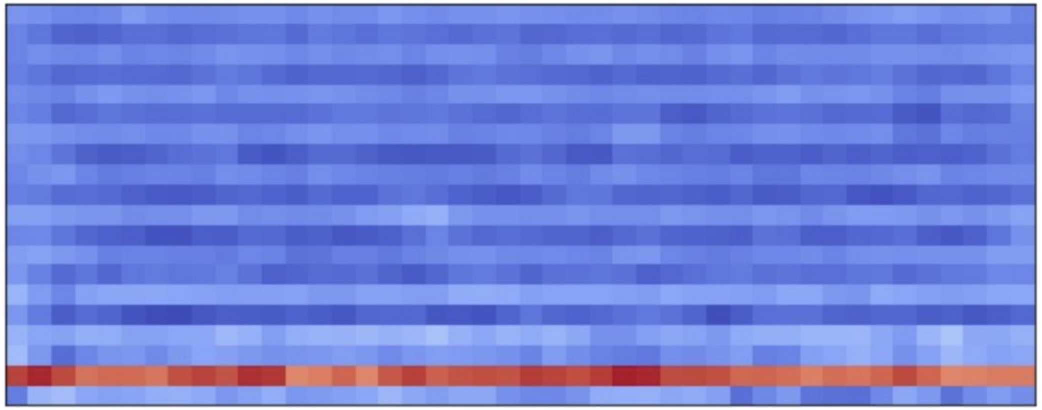
MFCC (Data Frame)

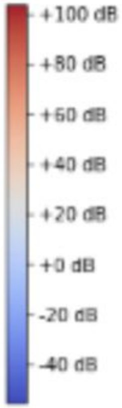
4.2.2. MFCC Feature Map

No Drone



Drone

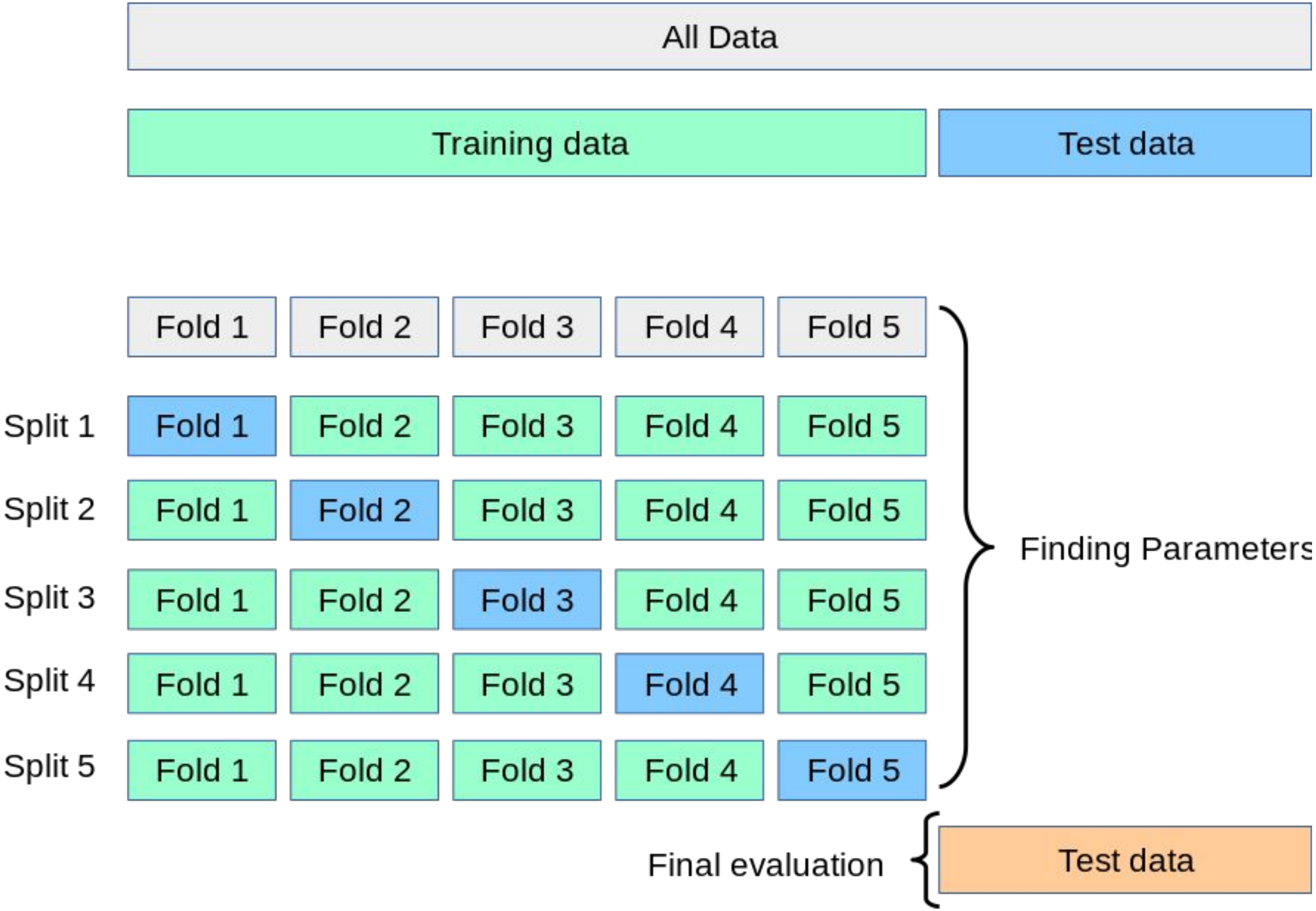




4.3.3. Model Summary

```
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_160 (Conv2D)          (None, 40, 16, 8)         208

max_pooling2d_160 (MaxPooli  (None, 20, 8, 8)          0
ng2D)

batch_normalization_160 (Ba  (None, 20, 8, 8)          32
tchNormalization)

conv2d_161 (Conv2D)          (None, 16, 4, 32)         6432

max_pooling2d_161 (MaxPooli  (None, 8, 2, 32)          0
ng2D)

batch_normalization_161 (Ba  (None, 8, 2, 32)          128
tchNormalization)

flatten_80 (Flatten)         (None, 512)               0

dense_160 (Dense)            (None, 32)                16416

dropout_80 (Dropout)         (None, 32)                0

dense_161 (Dense)            (None, 2)                 66

=================================================================
Total params: 23,282
Trainable params: 23,202
Non-trainable params: 80
```

4.3.3. 5-fold validation



4.3.4 Cross Evaluation – Accuracy :MFCC 20 & 40, Softmax & Sigmoid, 5-CV

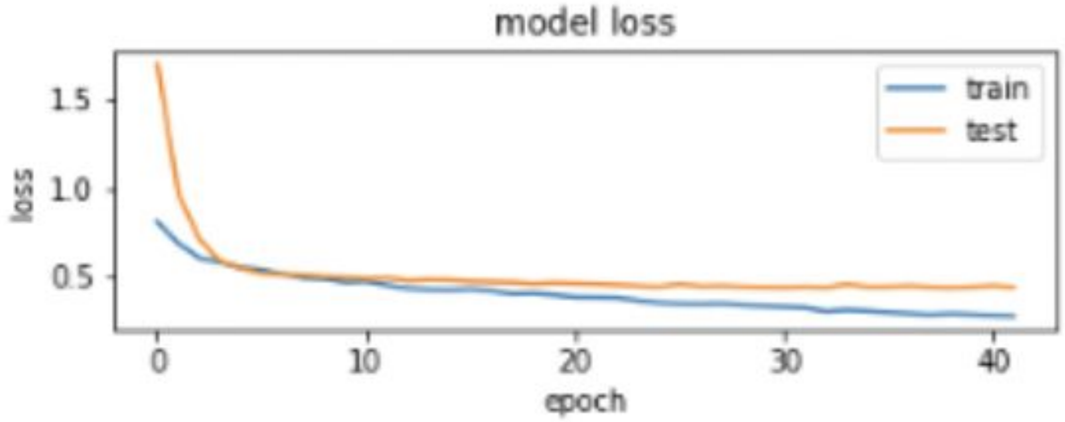
Monitor	Activation	n_mfcc	Batch size 16	Batch size 32	Batch size 64	Batch size 128
val_loss	Sigmoid	20	*77.21%*	*77.34%*	*78.18%*	*76.95%*
		40	*78.70%*	*79.74%*	*77.72%*	*77.07%*
	Softmax	20	*77.60%*	*77.69%*	*76.95%*	*76.36%*
		40	*78.83%*	*76.75%*	*78.31%*	*78.57%*
val_accuracy	Sigmoid	20	*77.01%*	*77.99%*	*75.78%*	*74.35%*
		40	*80.84%*	*79.74%*	*79.41%*	*80.06%*
	Softmax	20	*77.92%*	*77.92%*	*75.65%*	*76.23%*
		40	*81.81%*	*79.28%*	*78.96%*	*72.72%*

4.3.5. Highest Performance

Classes	Accuracy	precision	recall	f1-score
No drone	81.81%	*79%*	*82%*	*81%*
drone		*83.1%*	*79%*	*80%*

→ n_mfcc=40, activation=”softmax”, batch_size=16

4.3.6. Highest Performance



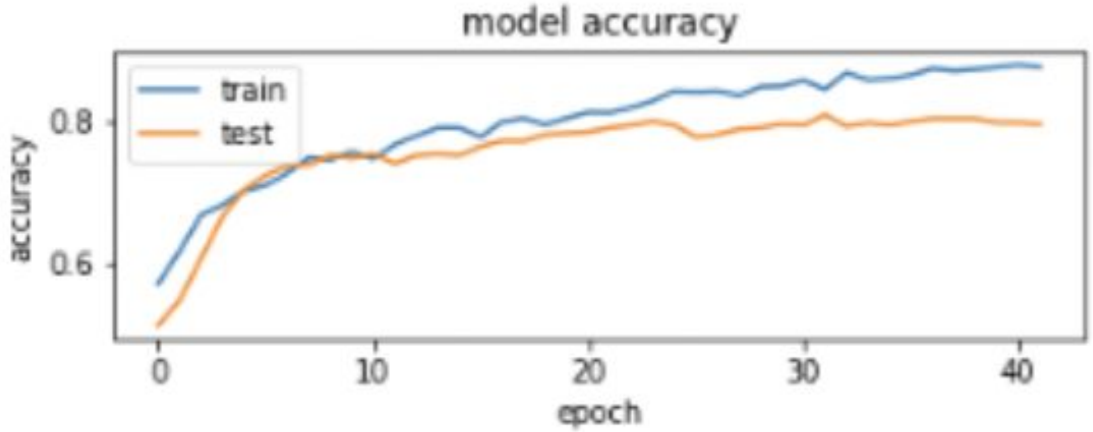


Image Process and Results

4.3.1. Vision Overview

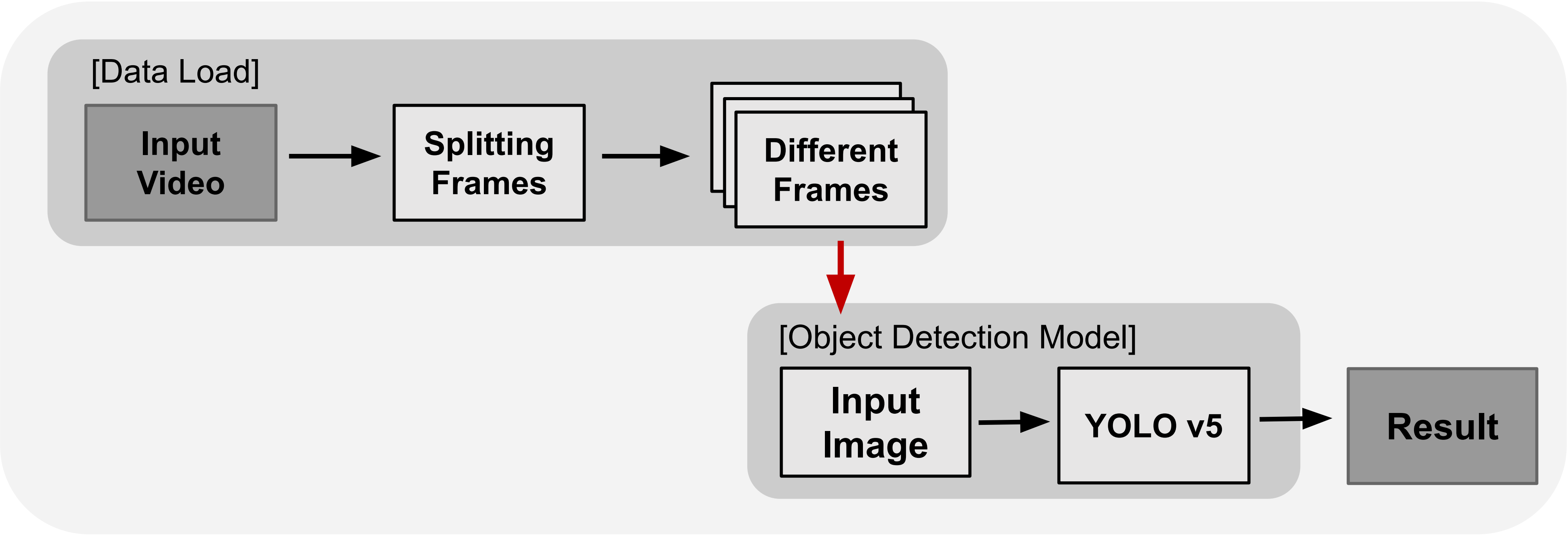
4.3.2. Collecting Data

4.3.3. YOLOv5 Model Architecture

4.3.4. Hyperparameters

4.3.5. Loss & Accuracy Curve about Train-Validation

4.3.1. Vision Overview



4.3.2. Data Processing

moving_dp20m_Matrice200V2_0726_1**.mp4**

moving_dp20m_Matrice200V2_0726_1_1**.jpg**
moving_dp20m_Matrice200V2_0726_1_2.jpg
moving_dp20m_Matrice200V2_0726_1_3.jpg
.
.
.

4.3.3. YOLOv5 Model Architecture

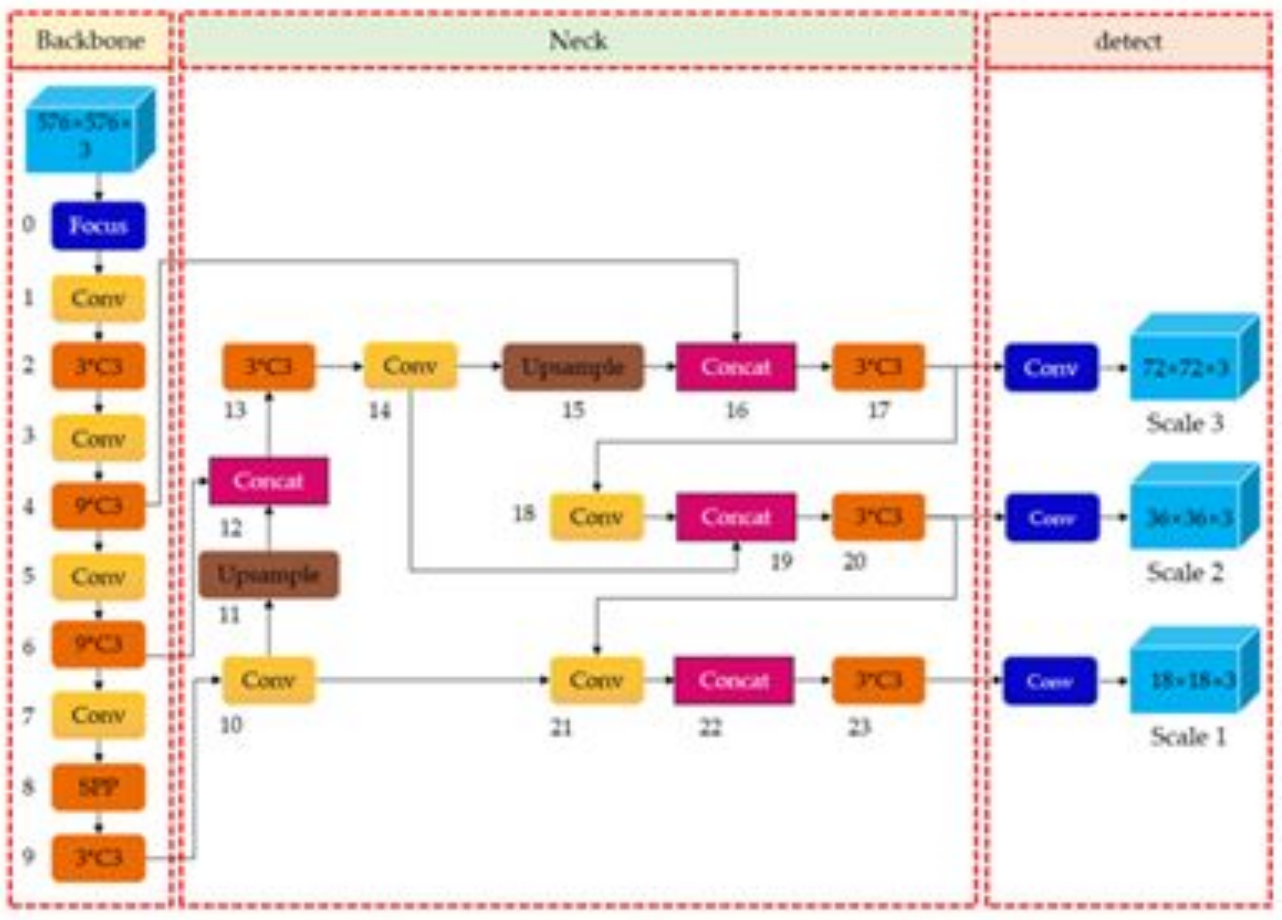
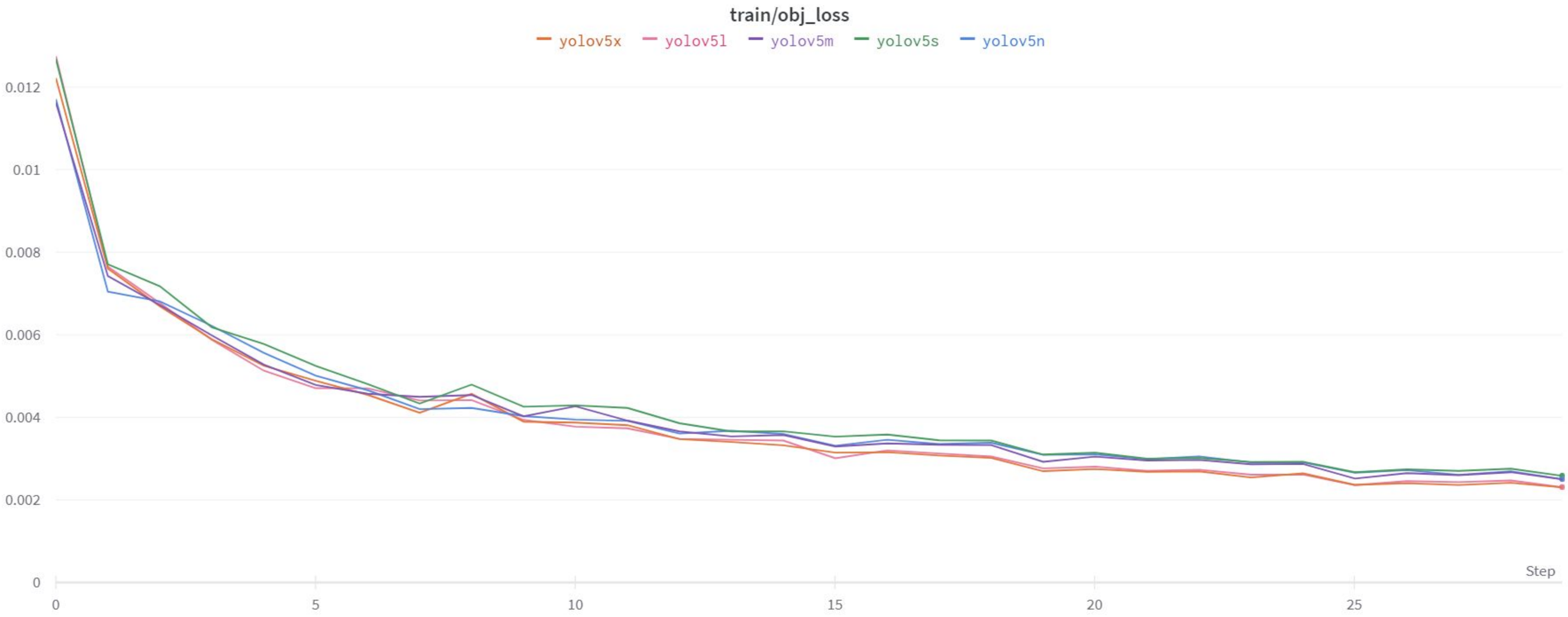


Figure 2. The architecture of the YOLOv5.

4.3.4. Hyperparameters

- Learning rate = 0.01
- Momentum = 0.937
- Weight_decay =5e-4
- Input Image = 640 x 640
- Batch size = 16
- Epoches = 30

4.3.5. Loss & Accuracy Curve about Train-Validation



4.3.6. Performance in Training Process

Models	mAP_0.5	Accuracy	Precision	Recall	F1-score
YoloV5n	0.744	0.867	0.589	0.914	0.716362
YoloV5s	0.824	0.876	0.646	0.914	0.756979
YoloV5m	0.801	0.879	0.639	0.857	0.732116
YoloV5l	0.76	0.883	0.619	0.886	0.728816
YoloV5x	**0.846**	0.899	0.605	0.914	0.728071

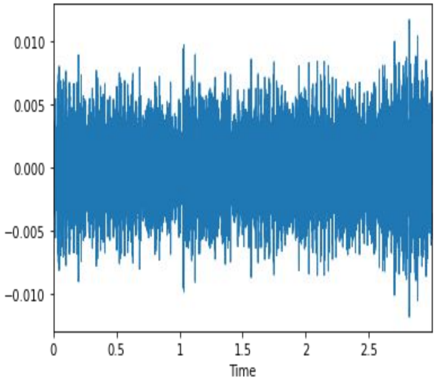
4.3.6. Performance in Training Process

Decision Fusion: Acoustic + Vision

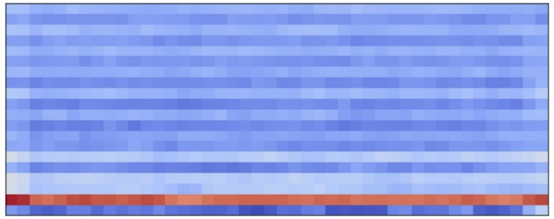
4.4.1. Decision Fusion

4.4.2. Result of Decision Fusion

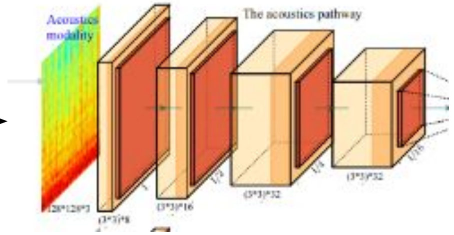
4.4.1. Decision Fusion



Audio Data



Audio Feature Map



Classification(CNN)

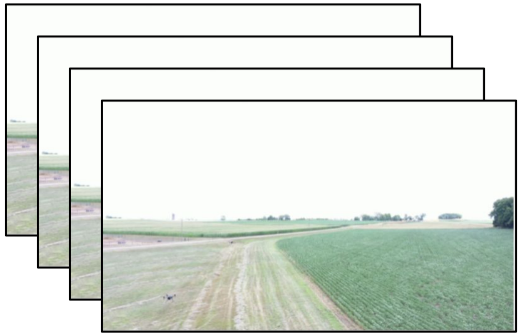
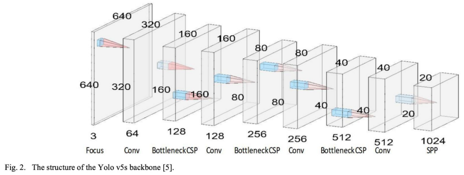


Image Data

Image Feature Map



YOLOv5 (Classifier)

Decision Fusion

4.4.2. Result of Decision Fusion

Classes	Vision based	Acoustic based	Vision + Audio
Drone Detection Accuracy	86.7%	81.81%	87.013%

→ YOLOv5n model with Audio processed with n_mfcc =40

5.

Conclusion

5.1.Limitation
5.2.Future work
5.3.Reference

5.1. Limitation

- Detecting Drone is just hovering in air space not moving
- The distance between detecting drone and target drone is only 20m.
- Only one target drone(DJI Matrice 200 V2) is used for experiments.

Future work

- Experiment with Detecting Drone in hovering and moving status is more reasonable
- The distance between detecting drone and target drone will be 20m~100m.
- Various type of drones will be sued as a target drone.

Reference

[1]Grand View Research, 2021. Commercial Drone Market Size, Share Trends Analysis Report by Product, by Application, by End-use, by Region, and Segment Forecasts, 2021-2028, April., 26, 2021.

[2] Meisenzahl, M., 2020. UPS will start delivering CVS prescriptions by drone in Florida —here's how it works. INSIDER, https://www.businessinsider.com/ups-deliver-cvs-prescription-medications-by-drone-the-villages-2020-4.

[3]Motlagh, Naser Hossein, Miloud Bagaa, and Tarik Taleb. "UAV-based IoT platform: A crowd surveillance use case." IEEE Communications Magazine 55.2 (2017): 128-134.

[4]Stevenson, B., 2016. A320 collision heightens UAV safety concerns.FlightGlobal,https://www.flightglobal.com/civil-uavs/a320-collision-heightens-uav-safety-concerns/120348.article.

[5]Science and Technology. 2022. Counter-Unmanned Aircraft Systems (C-UAS), https://www.dhs.gov/science-and-technology/counter-unmanned-aircraft-systems-c-uas.

[6]GUAN, X., LYU, R., SHI, H. and CHEN, J., 2020. A survey of safety separation management and collision avoidance approaches of civil UAS operating in integration national airspace system. Chinese Journal of Aeronautics , 33(11), pp.2851-2863.

[7]Yang, Bowon, et al. "UAV detection system with multiple acoustic nodes using machine learning models." 2019 Third IEEE International Conference on Robotic Computing (IRC). IEEE, 2019.

[8]McCoy, James, et al. "Ensemble Deep Learning for Sustainable Mul-timodal UAV Classification."IEEE Transactions on Intelligent Trans-portation Systems(2022).

[9]Abdullah, H., 2015. Man Detained for Flying Drone Near White House.NEWS,https://www.nbcnews.com/news/us-20news/20man-20detained-20trying-20fly-20drone-20near-20white-20house-20n359011

[10]A. R. Wagoner, D. K. Schrader and E. T. Matson, "Towards a vision-based targeting system for counter unmanned aerial systems (CUAS)," 2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications(CIVEMSA) , 2017, pp. 237-242, doi: 10.1109/CIVEMSA.2017.7995333.

[11]Yasin, Jawad N., et al. "Unmanned aerial vehicles (uavs): Collision avoidance systems and approaches." IEEE access 8 (2020): 105139-105155.

[12]Sharma, Sachin Umesh, and Dharmesh J. Shah. "A practical animal de-tection and collision avoidance system using computer vision technique."IEEE Access 5 (2016): 347-358.

[13]Liu, Hao, et al. "Drone detection based on an audio-assisted cameraarray."2017 IEEE Third International Conference on Multimedia BigData (BigMM). IEEE, 2017.

Thank you ☺

QnA

1. Why did we set a distance(horizontal) 20m for target drone?

2. Why did we set a altitude(vertical) 10m for detecting drone?

3. Why did we split 9:1 for audio and 7:2:1 for vision?

   -> We wanted to d