Università degli Studi di Milano - Bicocca
Università della Svizzera Italiana

DOUBLE DEGREE
MASTER OF SCIENCE IN COMPUTER SCIENCE

Interpretation of Neural Networks and Advanced Image Augmentation for Visual Control of Drones in Human Proximity

Supervisor: Dr. Alessandro Giusti
Co-Supervisor: PhD Stud. Dario Mantegazza

Master Thesis by
MARCO FERRI
ID 807130

PROBLEM INTRODUCTION

[Mantegazza et al. 2019]

HUMAN POSE ESTIMATION FOR FLYING A DRONE USING IMAGES FROM ITS CAMERA

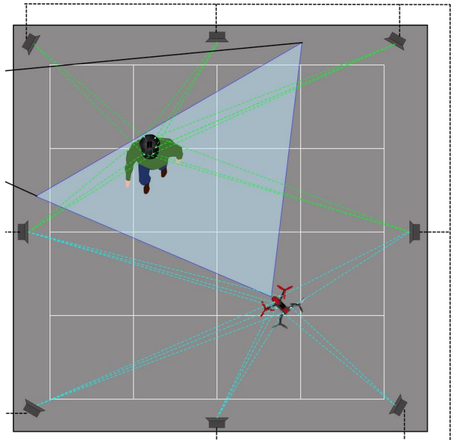
😥 ONLY WORKS IN THE DRONE ARENA

🎯 ENHANCE MODEL'S GENERALIZATION

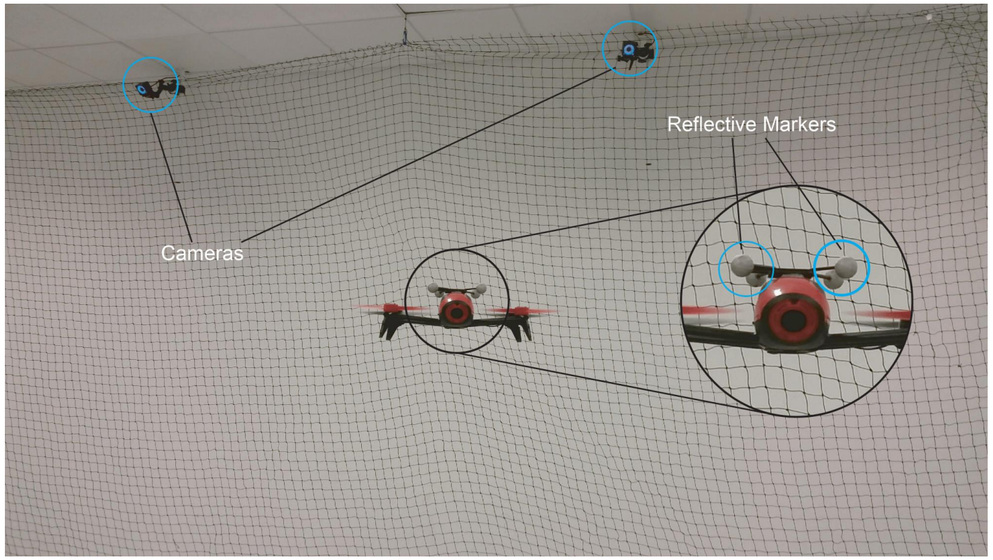
GIF link

HOW THE SYSTEM WORKS

- Supervised Learning
- Data collected in a dedicated drone arena
- Motion Capture to acquire the GT



USER'S 3D POSE WITH RESPECT TO THE DRONE



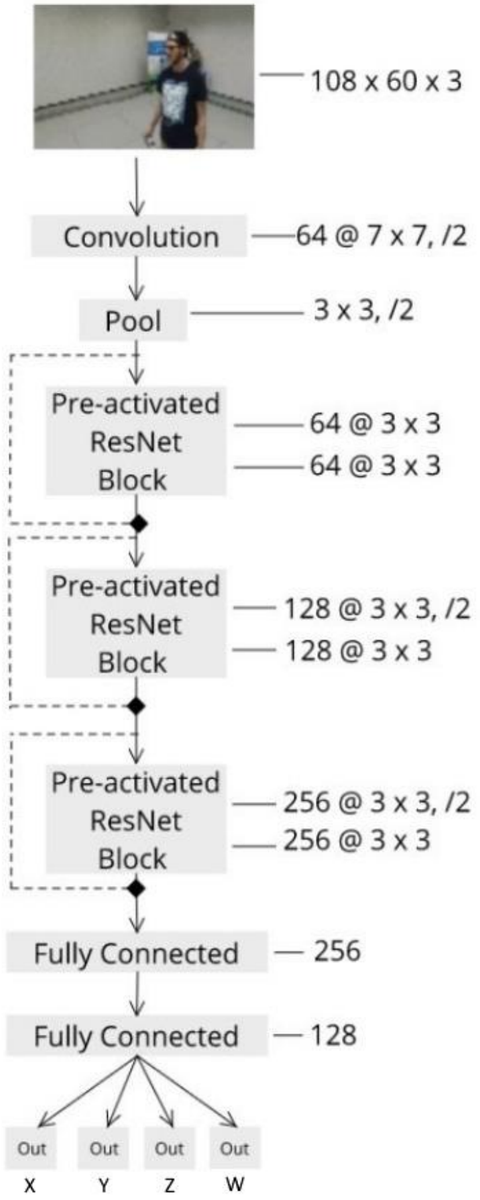
HOW THE SYSTEM WORKS

REGRESSION ON THE USER'S 3D COORDINATES

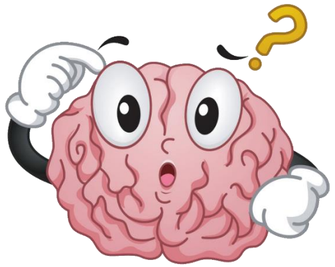
X	Distance from the camera
Y	Horizontal alignment
Z	Vertical alignment
W	Head's orientation (yaw)

RESNET | RESIDUAL NEURAL NETWORK

- **Input** camera images (108 x 60 RGB)
- **Output** 4 regression variables



OUR CONTRIBUTION



MODEL INTERPRETATION

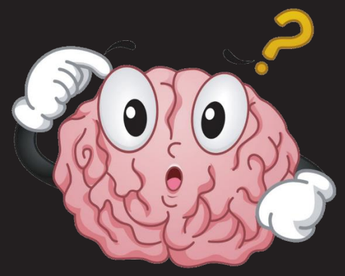
TO FIND THE PROBLEM

AUGMENTATION STRATEGY

TO GENERALIZE THE TASK

EXPERIMENTAL EVALUATIONS

TO VALIDATE THE SOLUTION



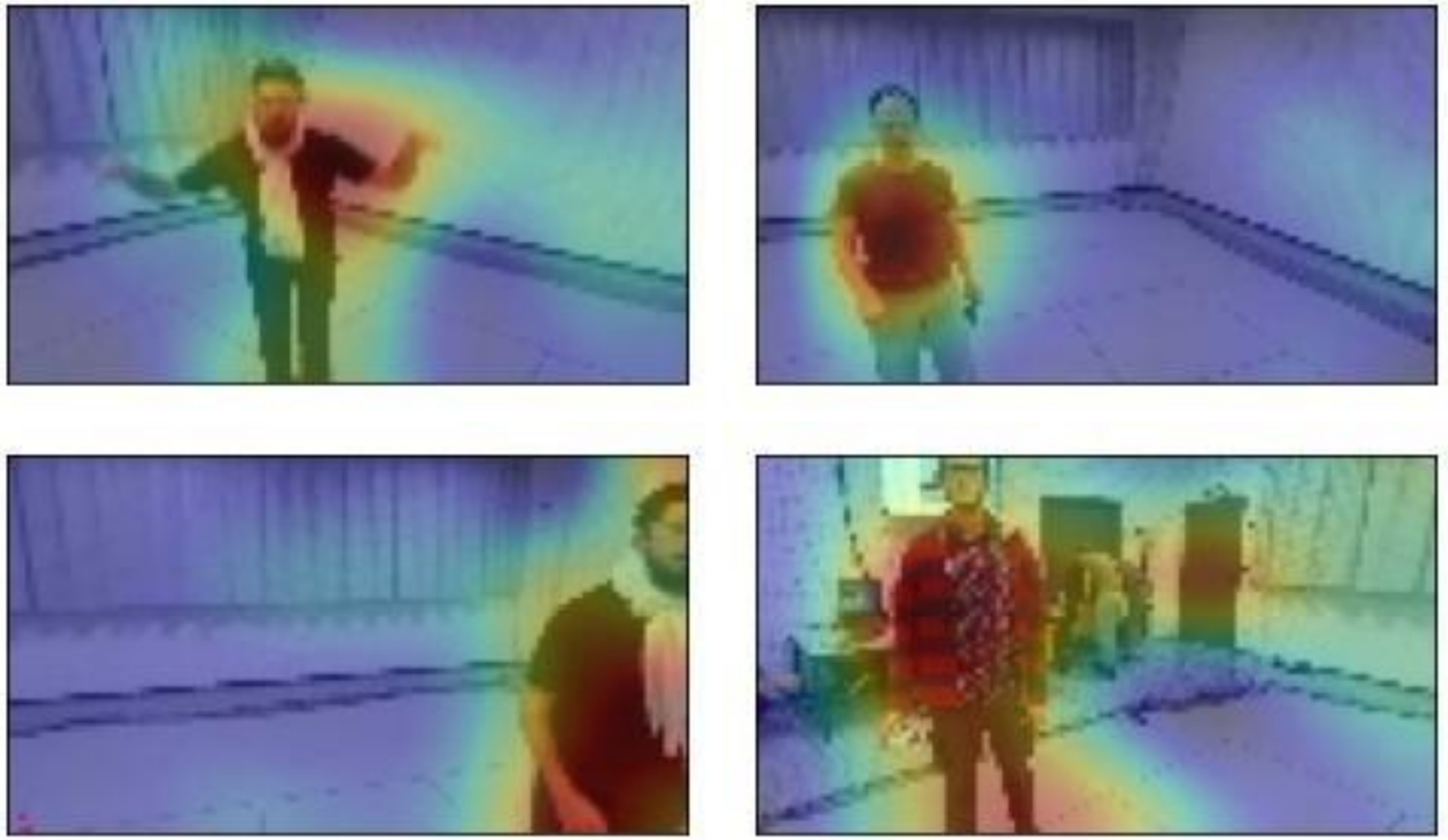
MODEL INTERPRETATION

TO FIND THE PROBLEM

GRAD-CAM

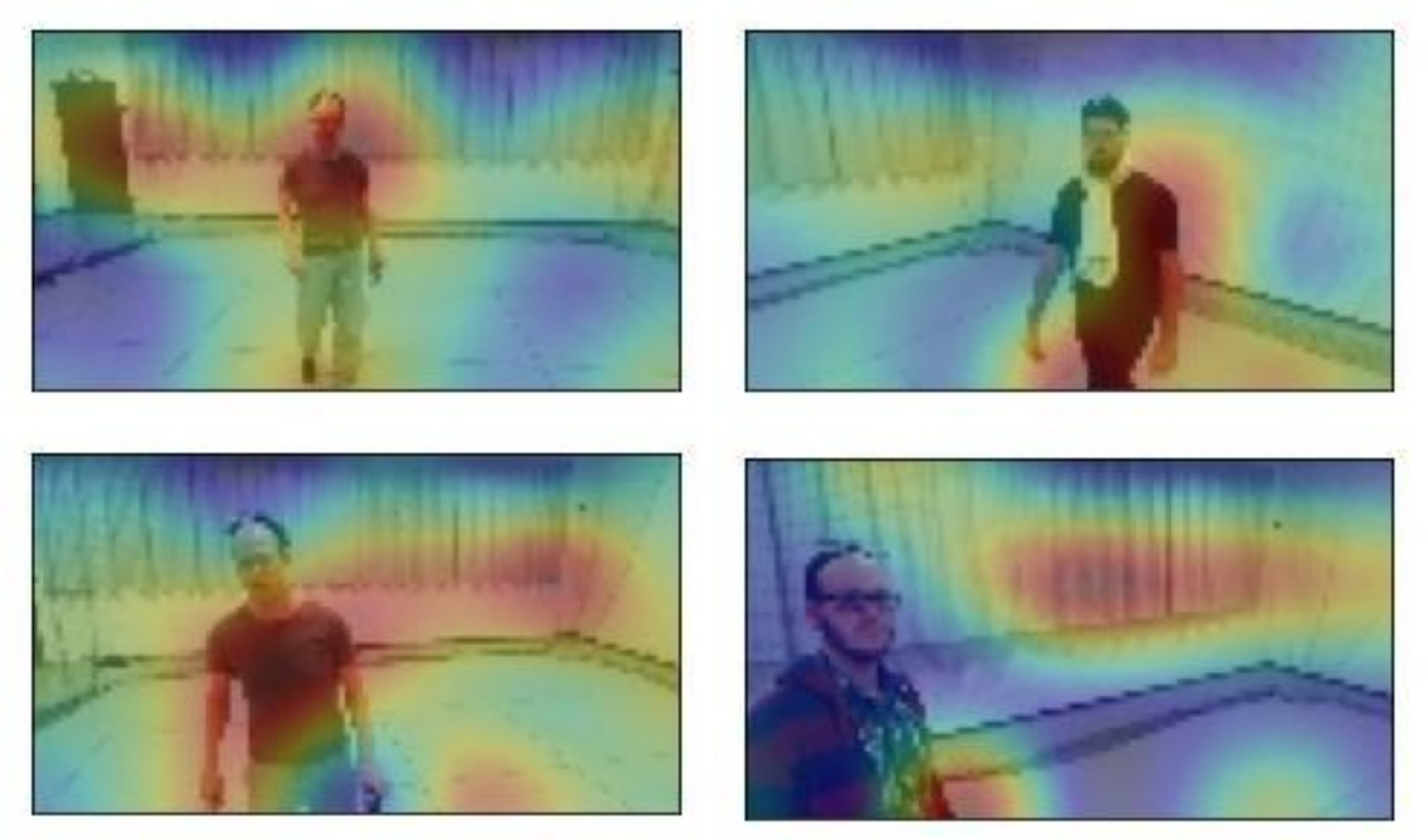
HIGHLIGHTS THE REGIONS OF THE INPUT IMAGES
MOST RESPONSIBLE FOR THE NEURAL NETWORK'S OUTPUT

GRAD-CAM **EXPECTED OBSERVATIONS**

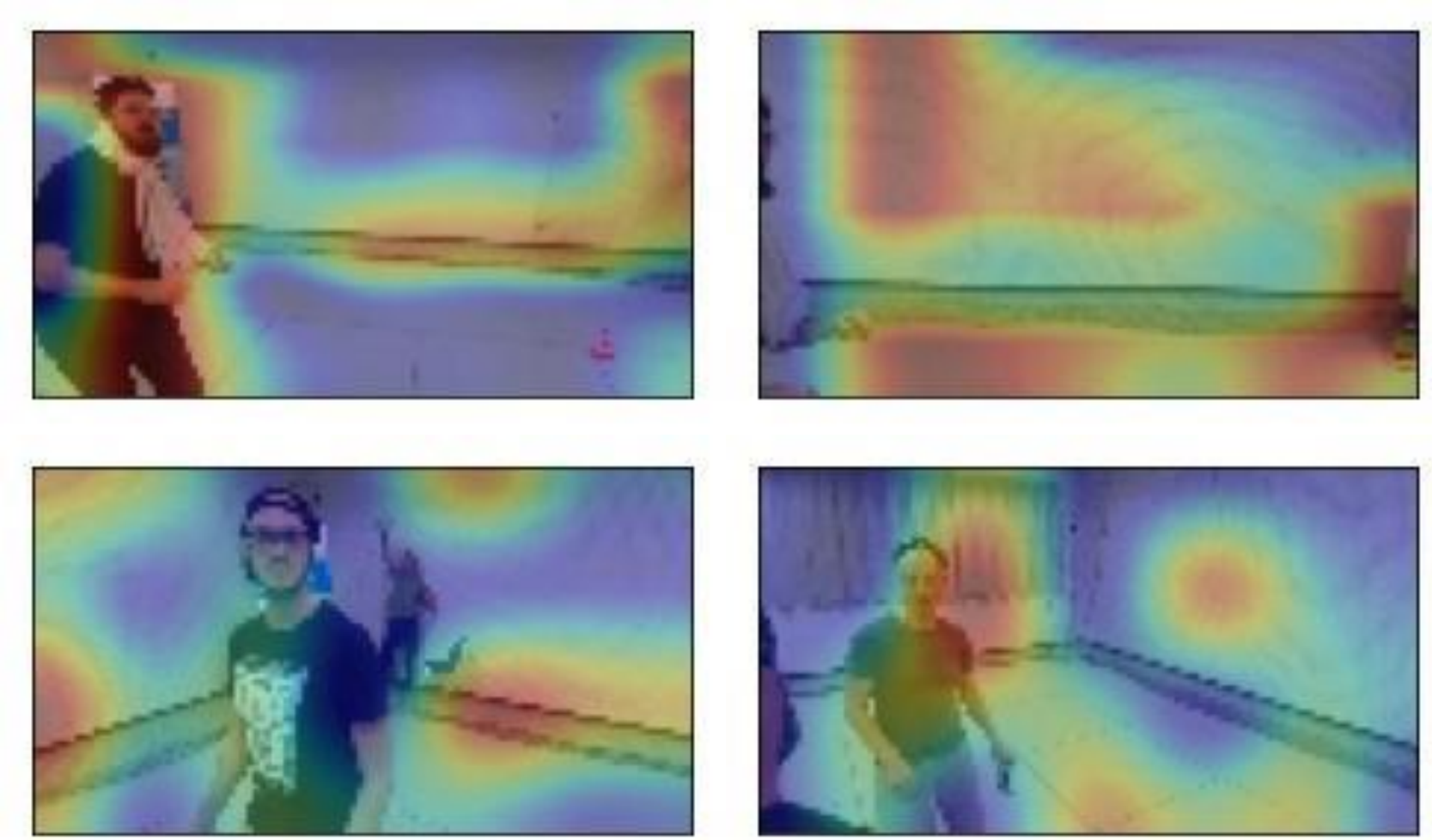


GRAD-CAM REVEALS: **BACKGROUND OBJECTS**

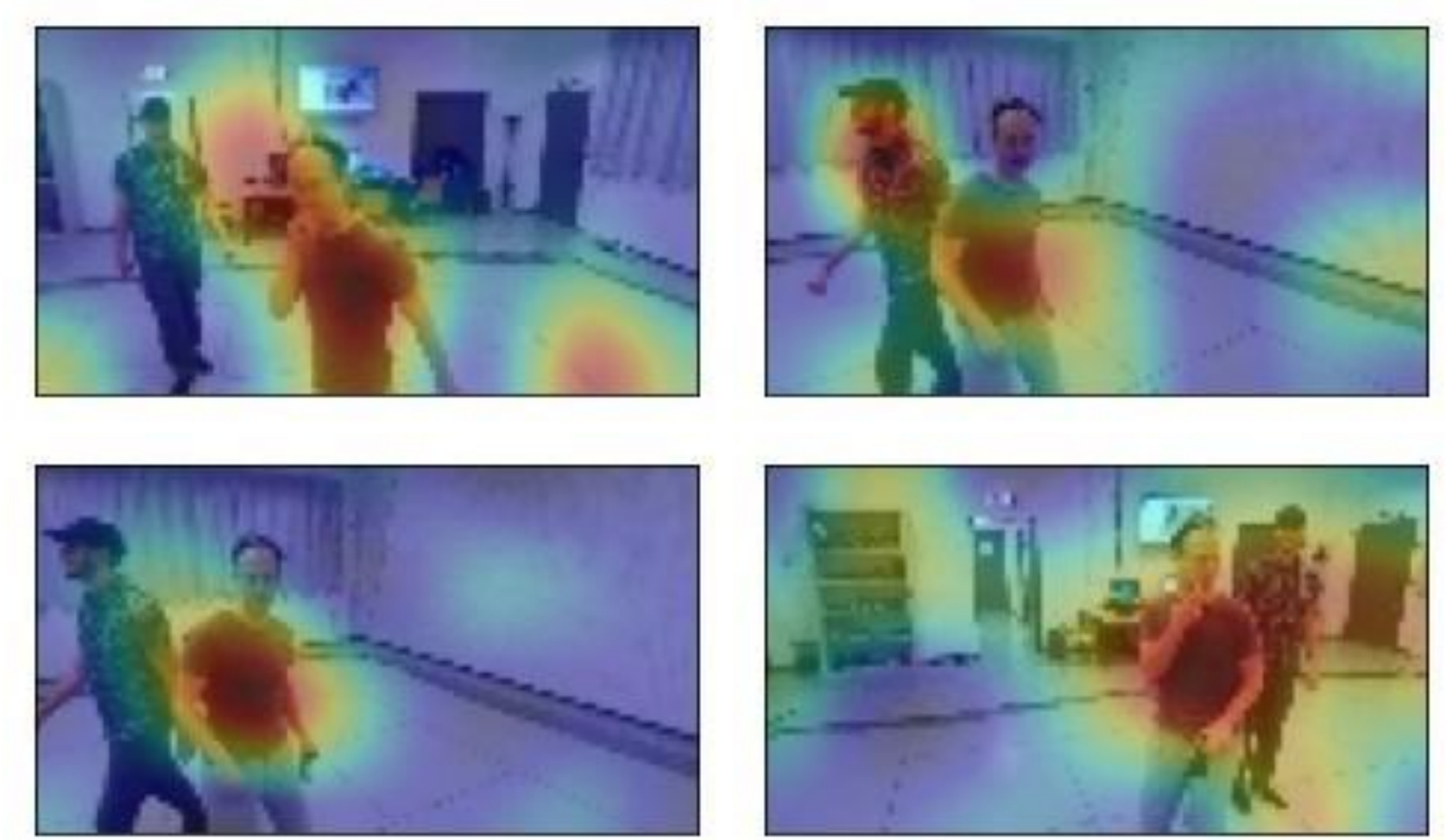
GRAD-CAM REVEALS: **CURTAINS**



GRAD-CAM REVEALS: **WALLS**



GRAD-CAM REVEALS: **MULTIPLE PEOPLE**



GRAD-CAM SUMMARY

THE MODEL IS DEPENDANT ON THE DRONE ARENA IN WHICH THE DATASET HAS BEEN COLLECTED

WE WANT TO ELIMINATE THE OVERFITTING ON THE TRAINING SET

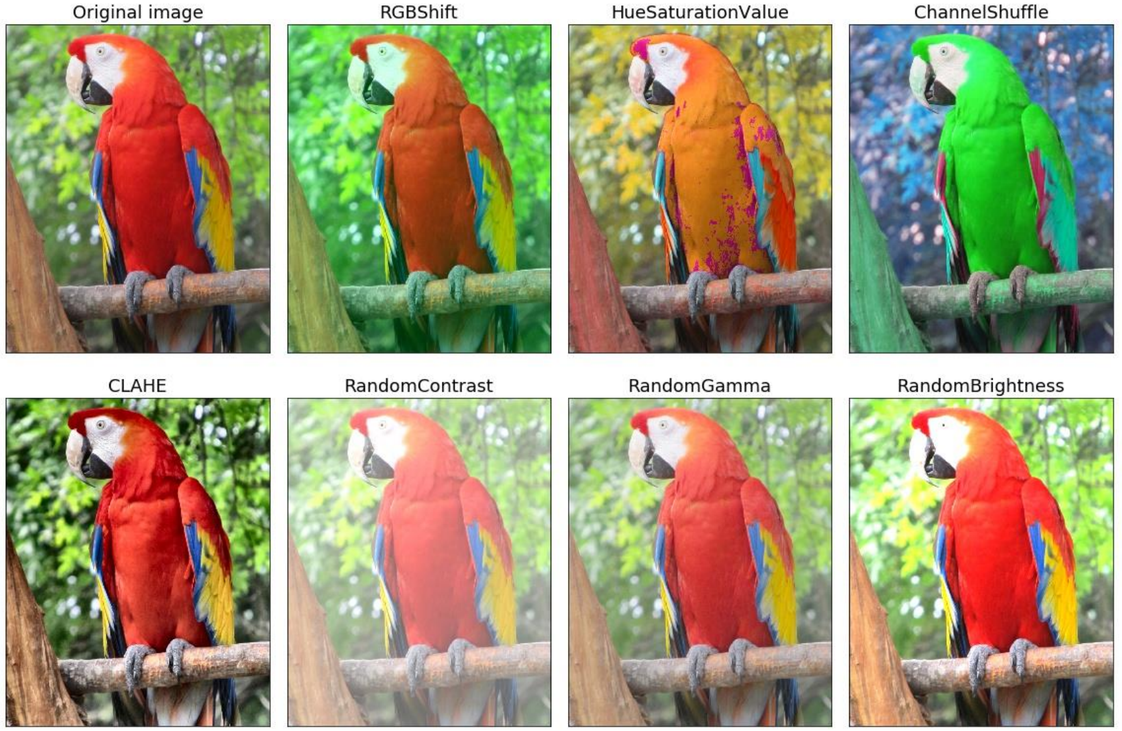
AUGMENTATION STRATEGY

TO GENERALIZE THE TASK

WHAT IS DATA AUGMENTATION

WIDELY USED IN MACHINE LEARNING TO GENERATE NEW TRAINING SAMPLES FOR REDUCING OVERFITTING

CLASSIC IMAGE AUGMENTATION

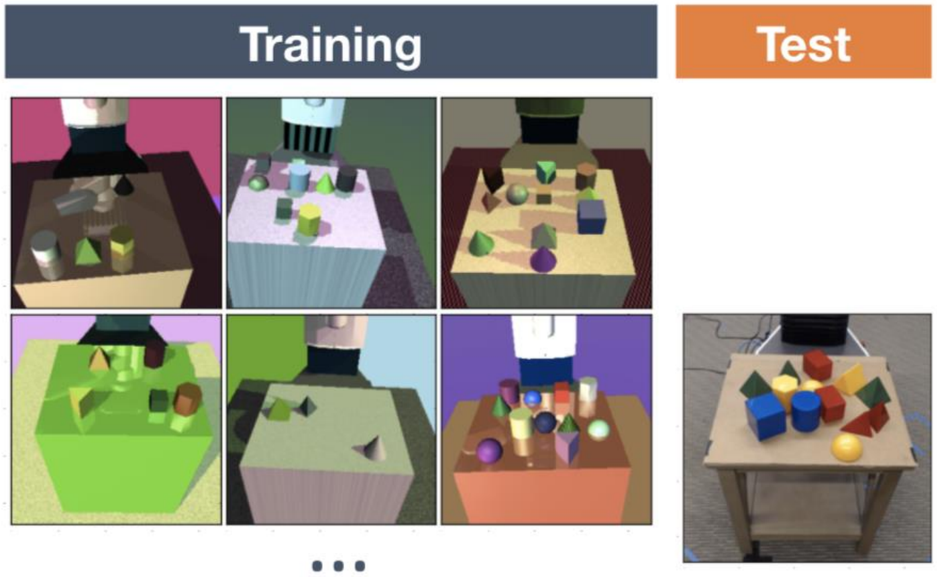


https://github.com/albumentations-team/albumentations

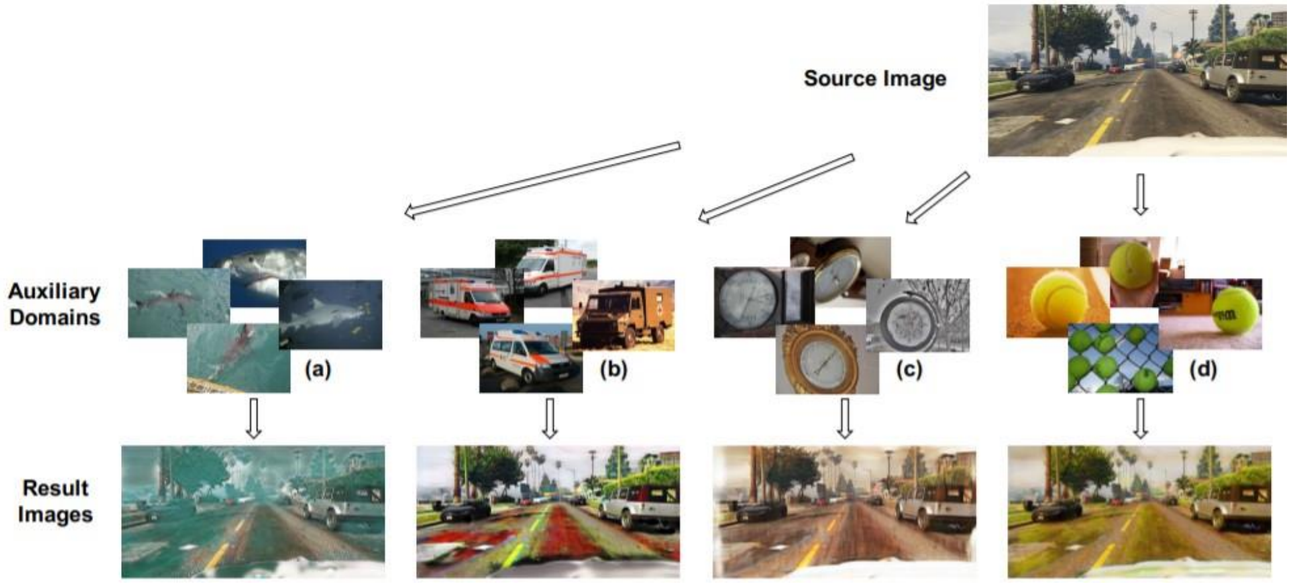
WHAT IS DATA AUGMENTATION

WIDELY USED IN MACHINE LEARNING TO GENERATE NEW TRAINING SAMPLES FOR REDUCING OVERFITTING

DOMAIN RANDOMIZATION

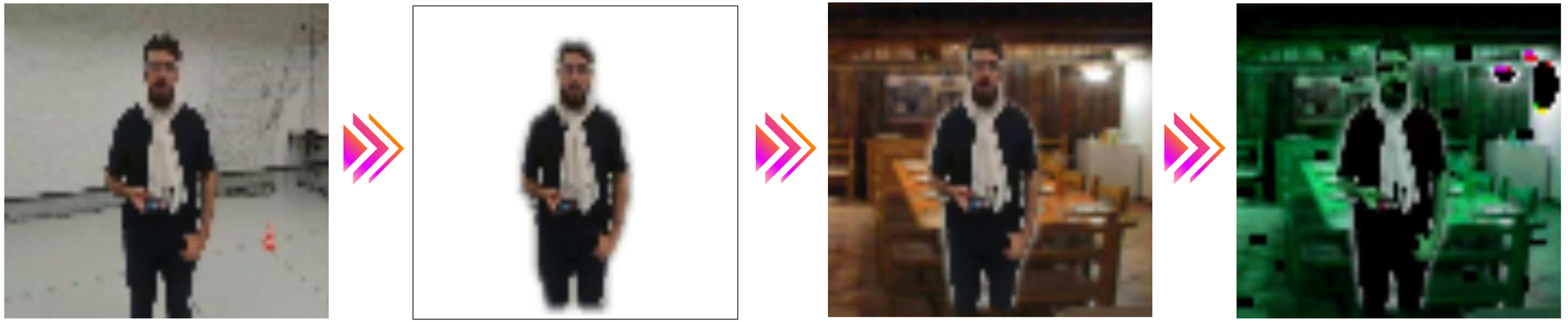


[Tobin et al. 2017]



[Yue et al. 2019]

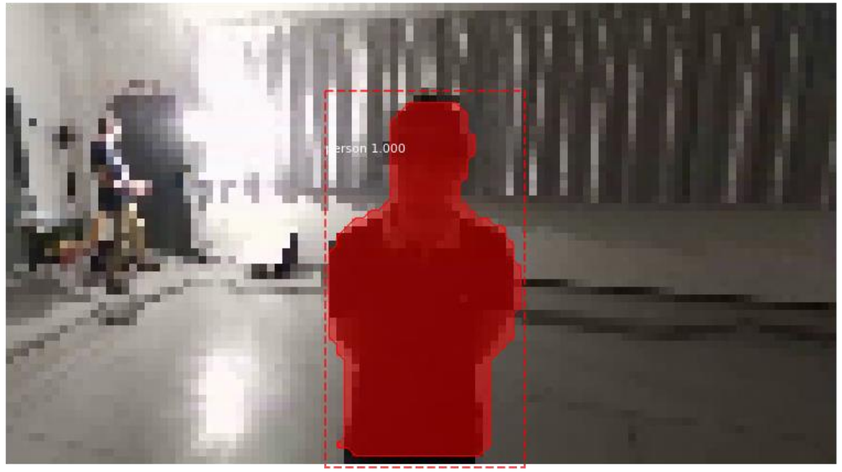
OUR AUGMENTATION STRATEGY

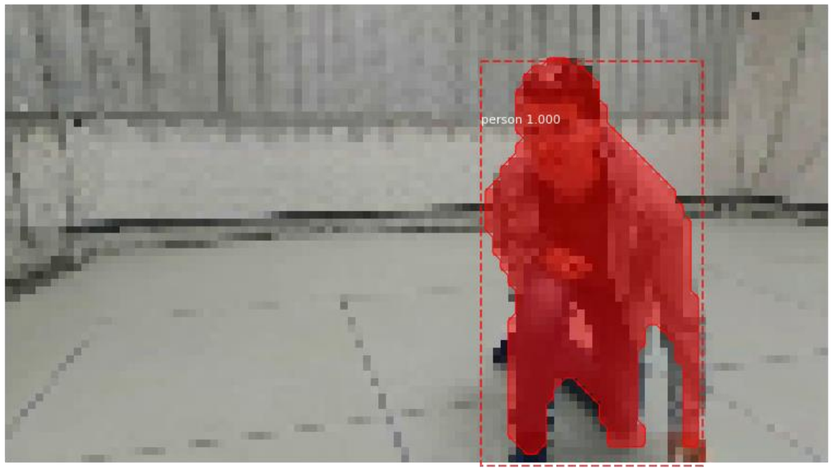


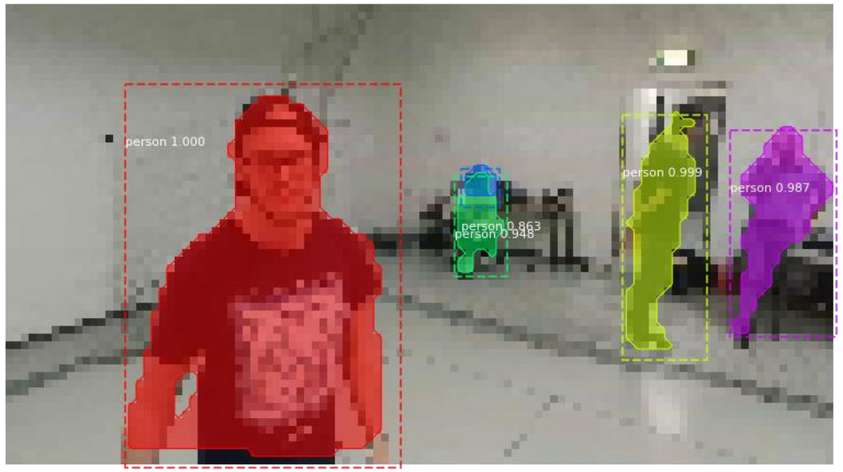
1. MASKING THE USERS IN THE IMAGES
2. REPLACING WITH NEW BACKGROUNDS
3. APPLYING CLASSIC IMAGE AUGMENTATIONS
4. TRAINING THE MODEL ON AUGMENTED DATASET

PERSON MASKING WITH MASK R-CNN

1







BACKGROUNDS DATASET

2





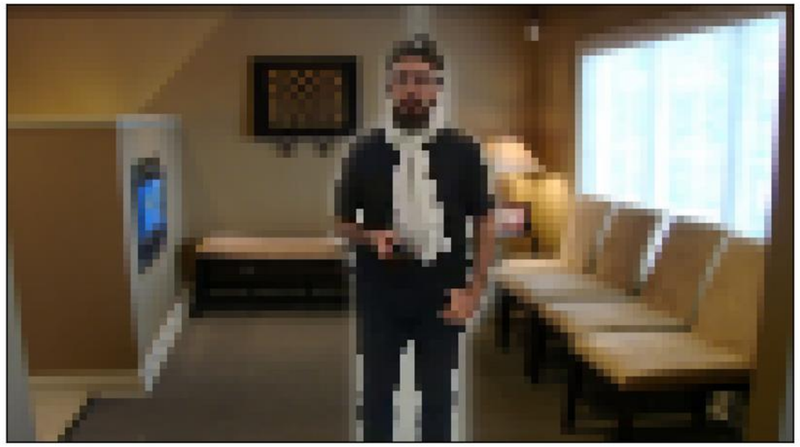
INDOOR SCENE RECOGNITION - CVPR 2009

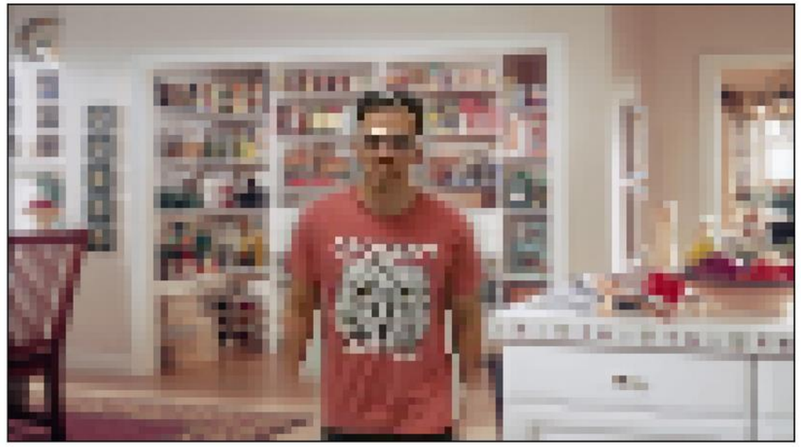
[Quattoni et al. 2009]

15’620 IMAGES
67 CATEGORIES

BACKGROUND REPLACEMENT

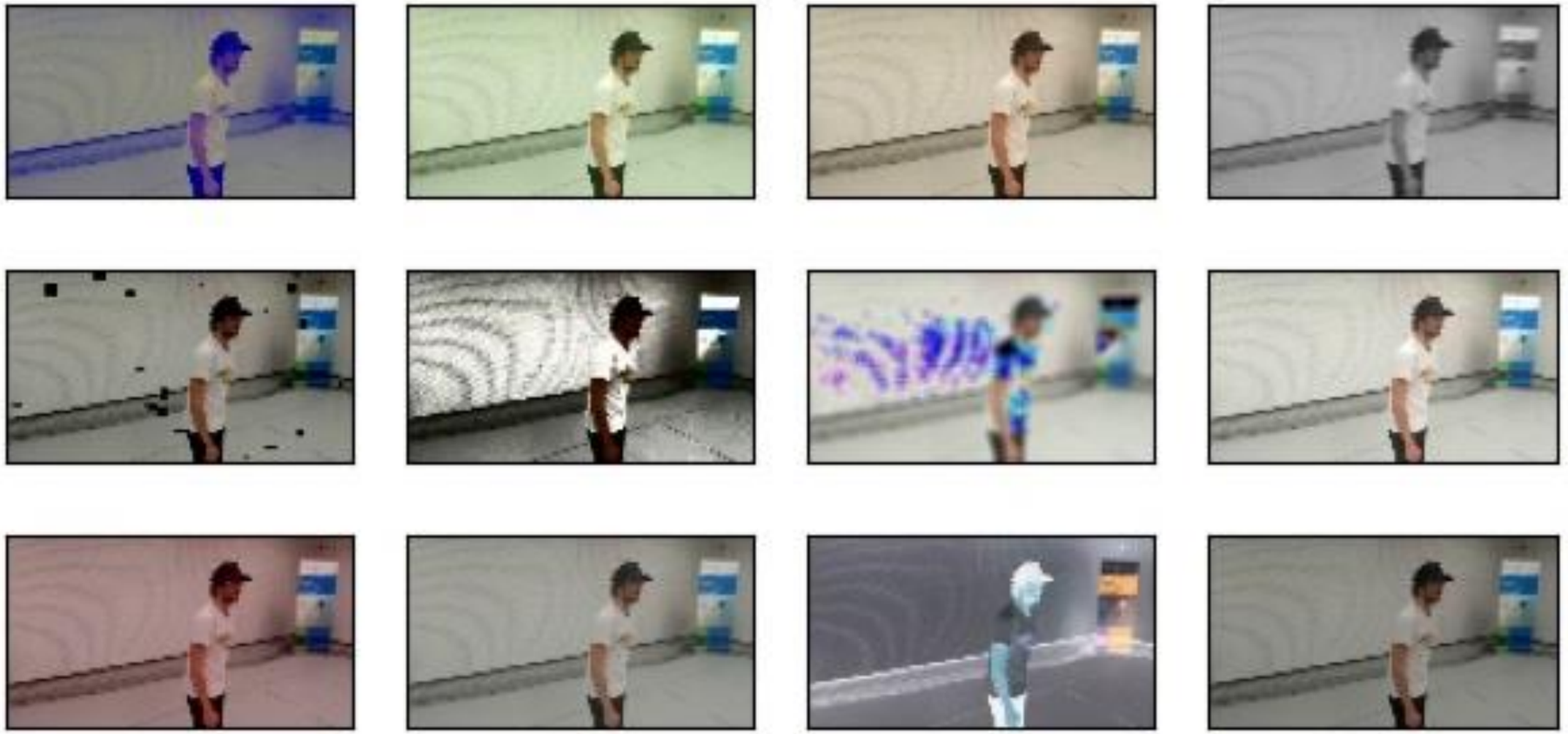
2





CLASSIC IMAGE AUGMENTATION

3



MODELS DEFINITION

4

TRAINED ON
ORIGINAL
DATASET

ARENA MODEL

TRAINED ON
BACKGROUND REPLACED
DATASET

CVPR MODEL

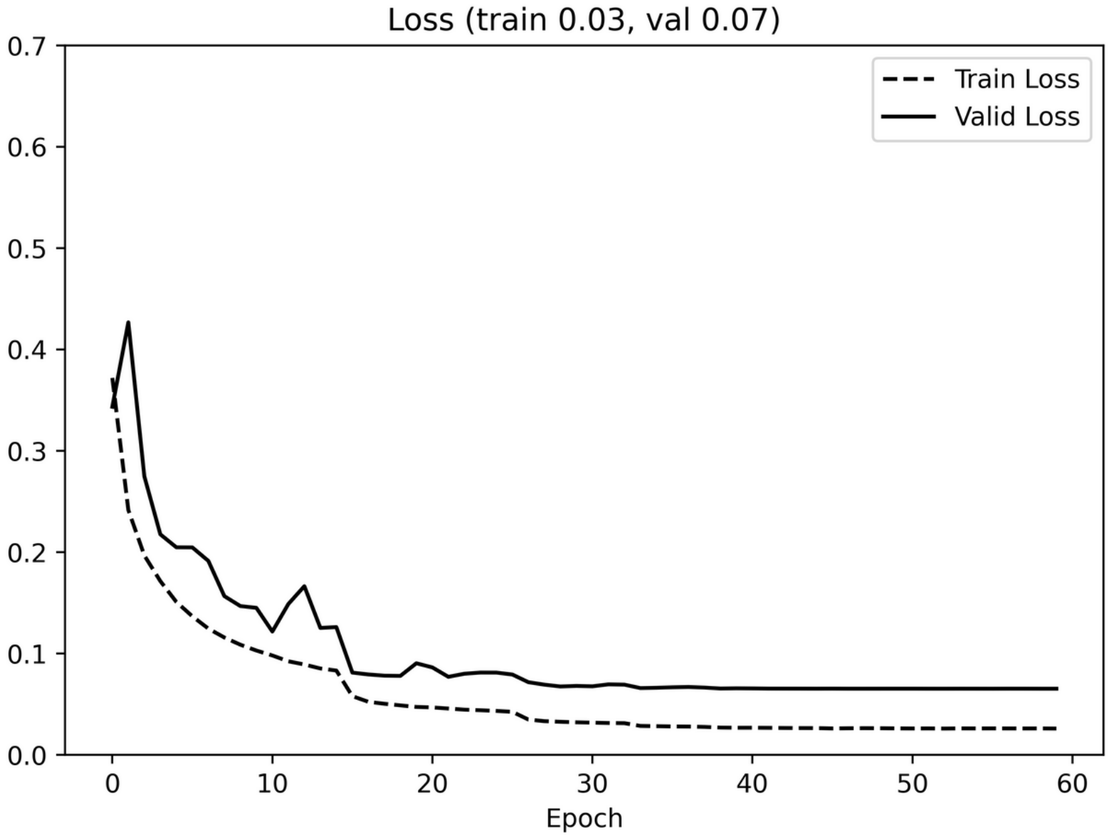
TRAINED ON
BG REPLACED & IMG
AUGMENTED DATASET

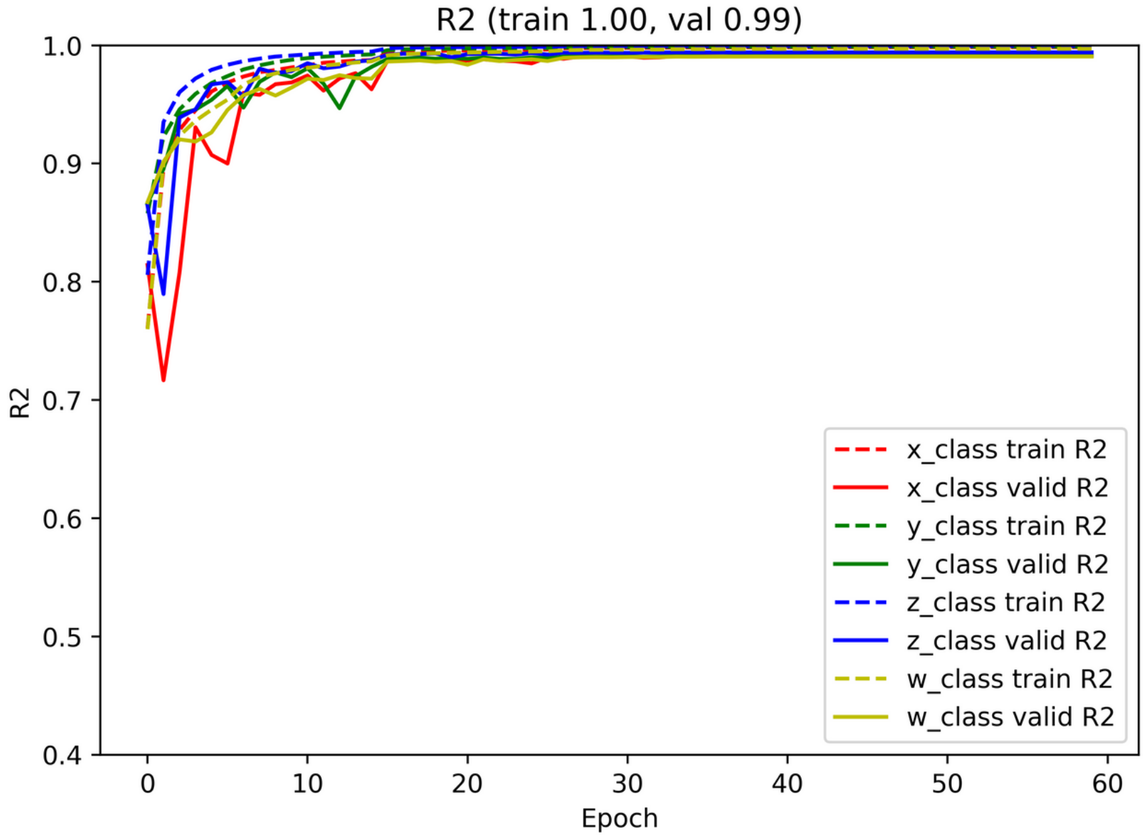
CVPR AUG MODEL

TRAINING

4

ARENA MODEL



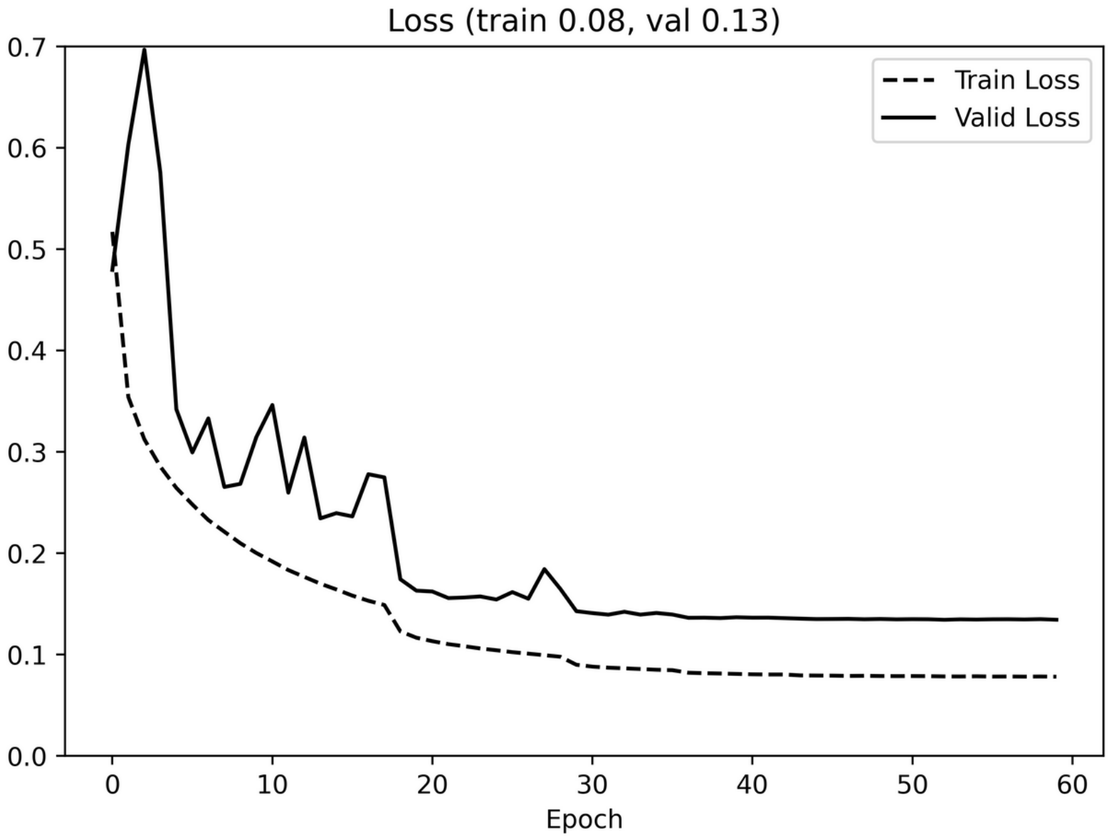


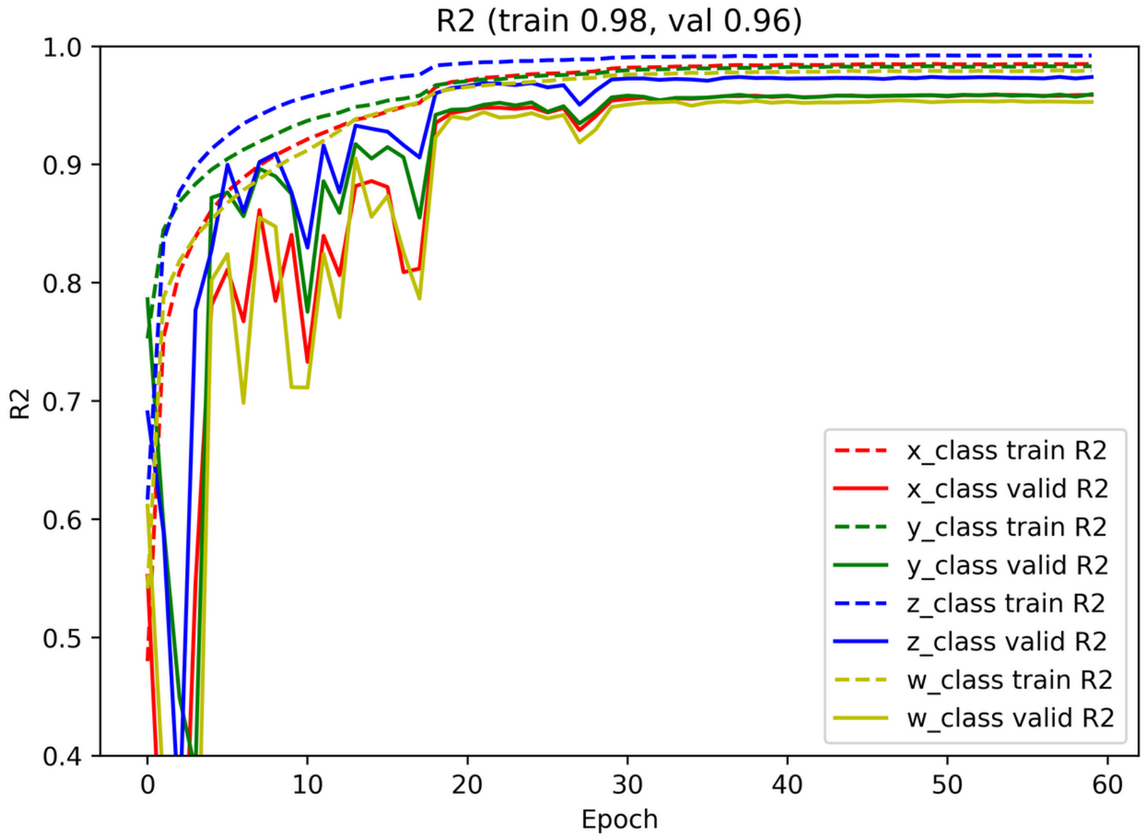
TRAINING

B

4

CVPR MODEL



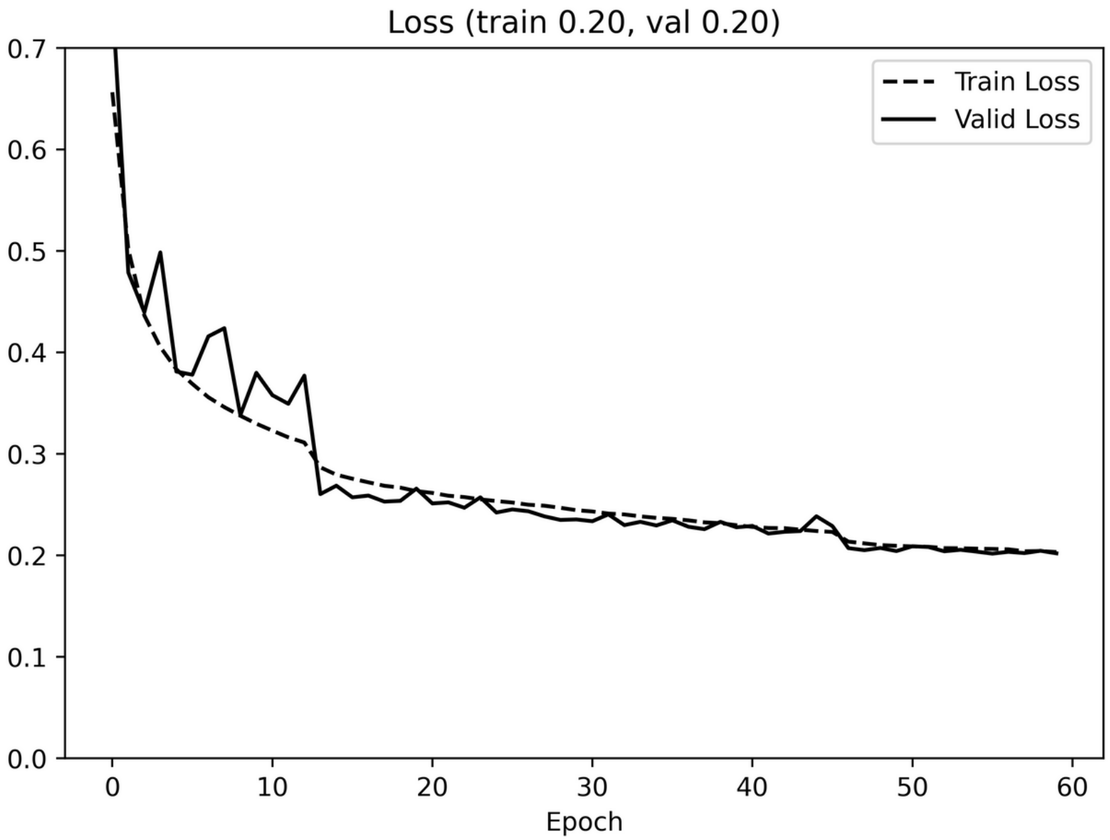


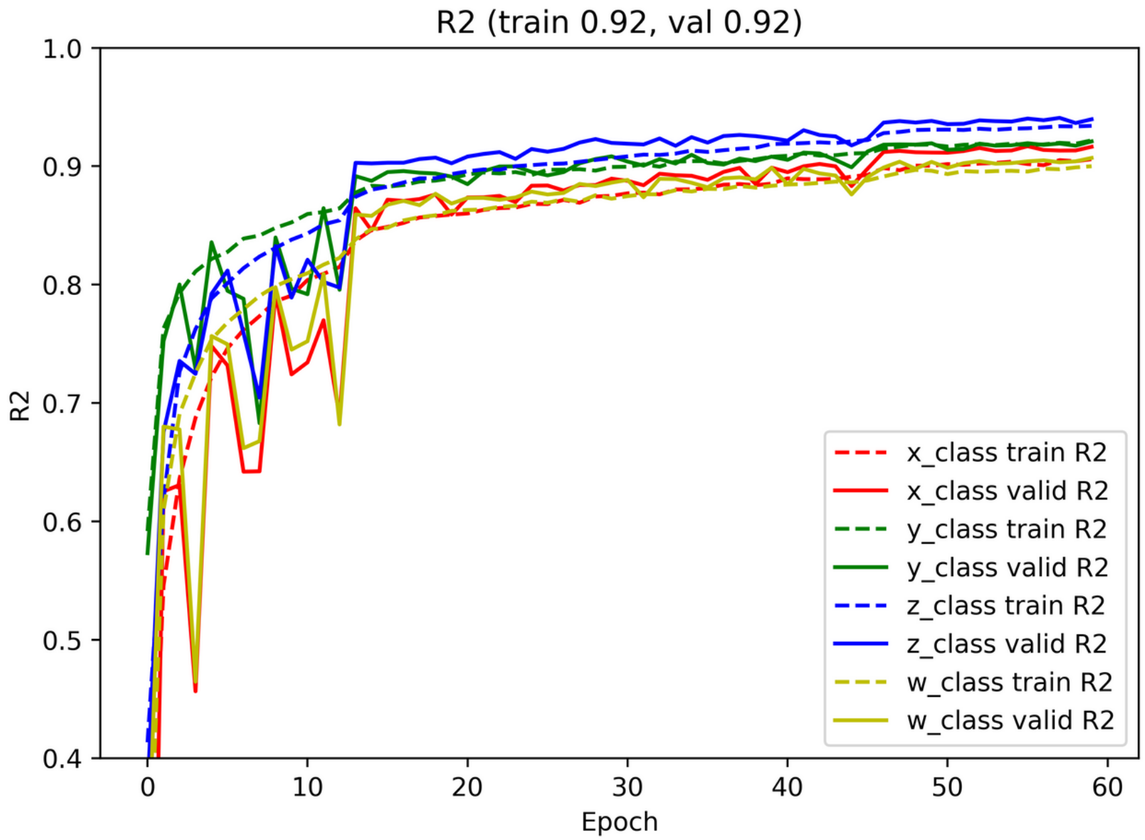
TRAINING

C

4

CVPR AUG MODEL





EXPERIMENTAL EVALUATIONS

TO VALIDATE THE SOLUTION

TEST SETS DEFINITION

Original

ARENA

TEST SET

through background replacement

INDOOR 1

TEST SET

through background replacement

INDOOR 2

TEST SET

QUANTITATIVE EVALUATION: ARENA TEST SET

		Arena	CVPR	CVPR Aug
TEST LOSS (MAE)		0,41	0,42	0,36
R2	**x**	0,81	0,69	0,83
	y	0,86	0,82	0,87
	z	0,79	0,79	0,86
	w	0,74	0,74	0,78

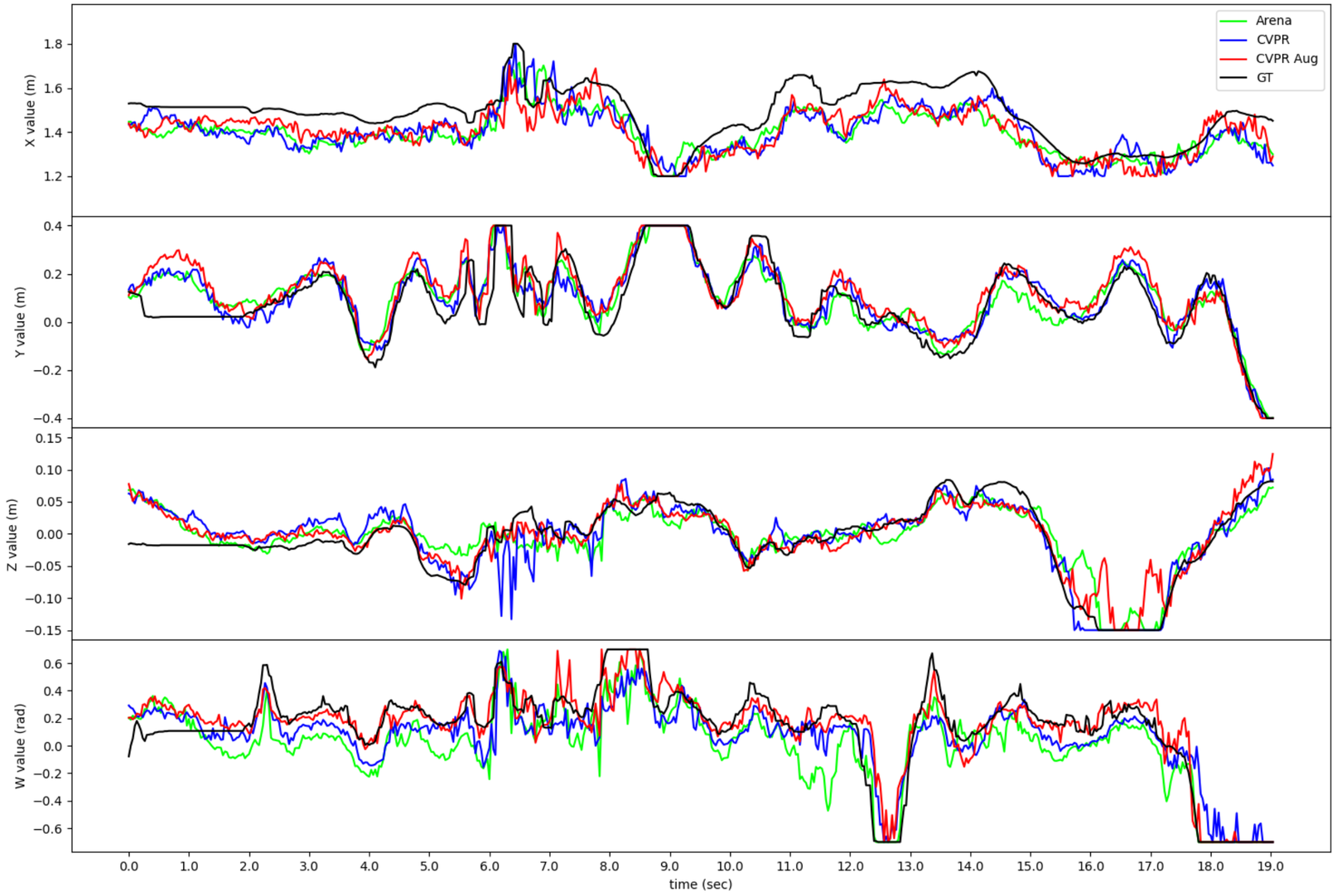
QUANTITATIVE EVALUATION: INDOOR 1 TEST SET

		A	B	C
		Arena	**CVPR**	**CVPR Aug**
TEST LOSS (MAE)		0,86	0,44	0,37
R2	**x**	0,16	0,58	0,83
	y	0,30	0,81	0,83
	z	0,21	0,78	0,85
	w	0,08	0,70	0,76

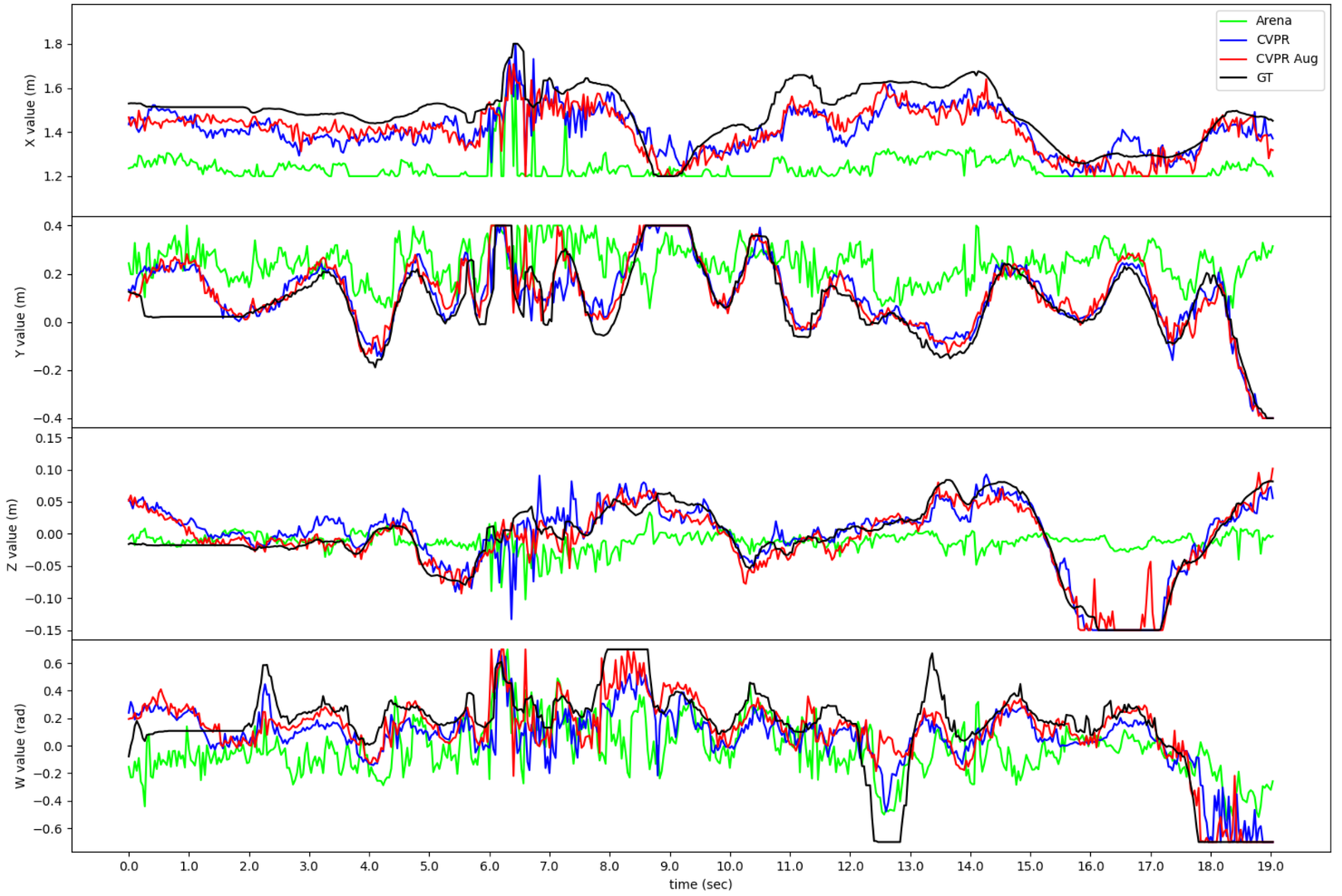
QUANTITATIVE EVALUATION: INDOOR 2 TEST SET

		A	B	C
		Arena	**CVPR**	**CVPR Aug**
TEST LOSS (MAE)		1,00	0,45	0,37
R2	**x**	-0,86	0,64	0,81
	y	0,51	0,82	0,84
	z	-2,78	0,77	0,86
	w	0,24	0,67	0,75

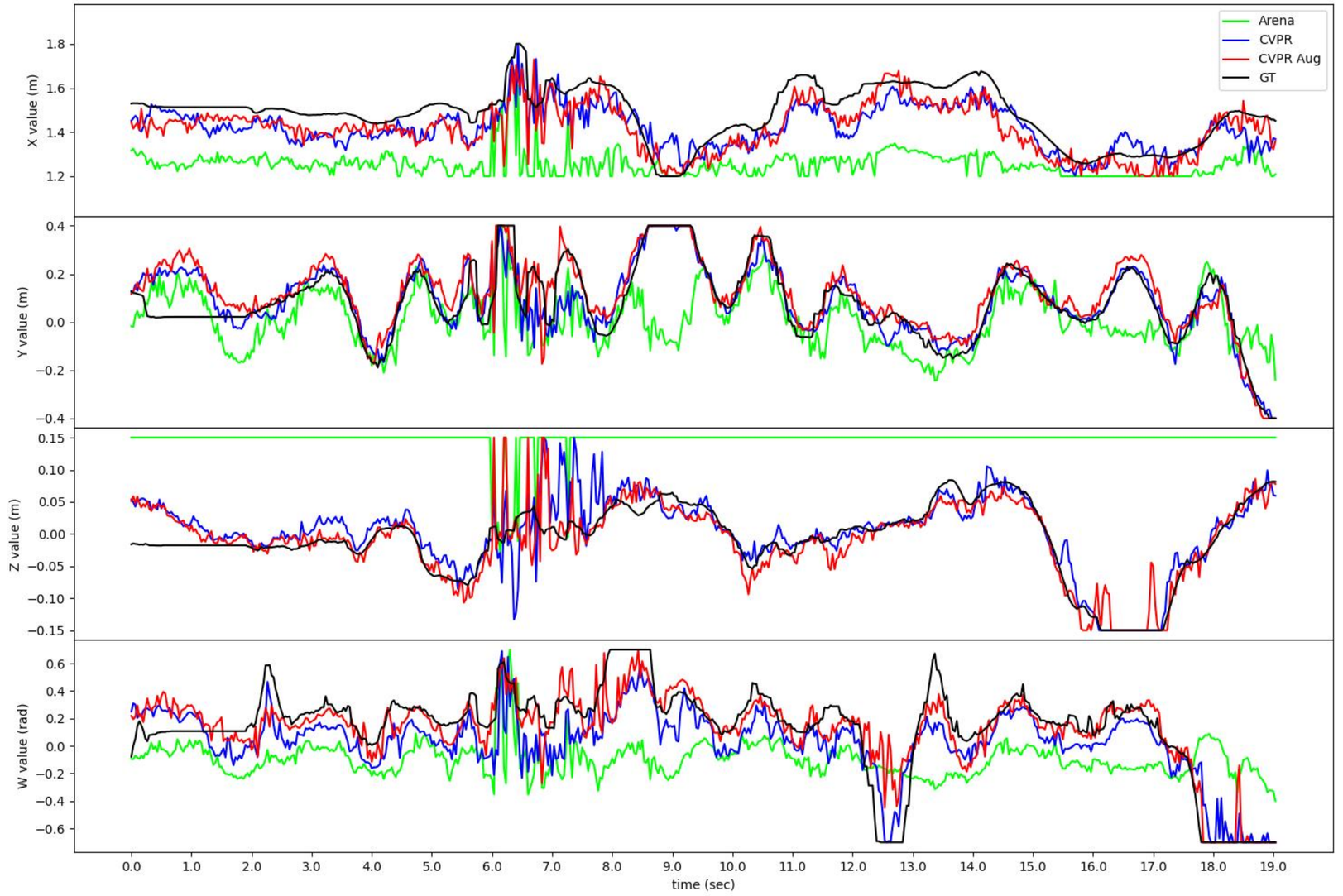
TIMELINE EVALUATION: ARENA TEST SET



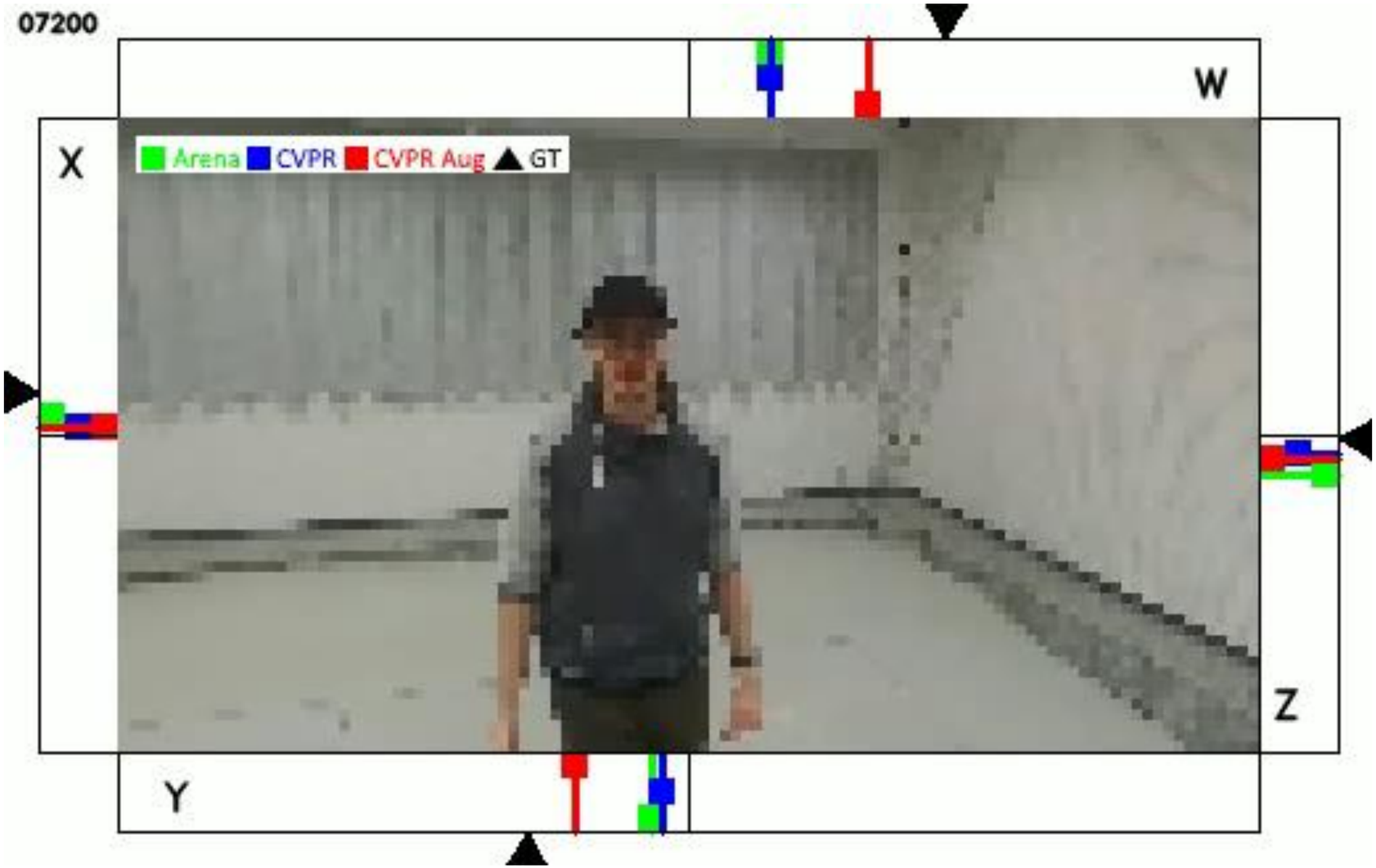
TIMELINE EVALUATION: INDOOR 1 TEST SET



TIMELINE EVALUATION: INDOOR 2 TEST SET



PER-FRAME EVALUATION: ARENA TEST SET

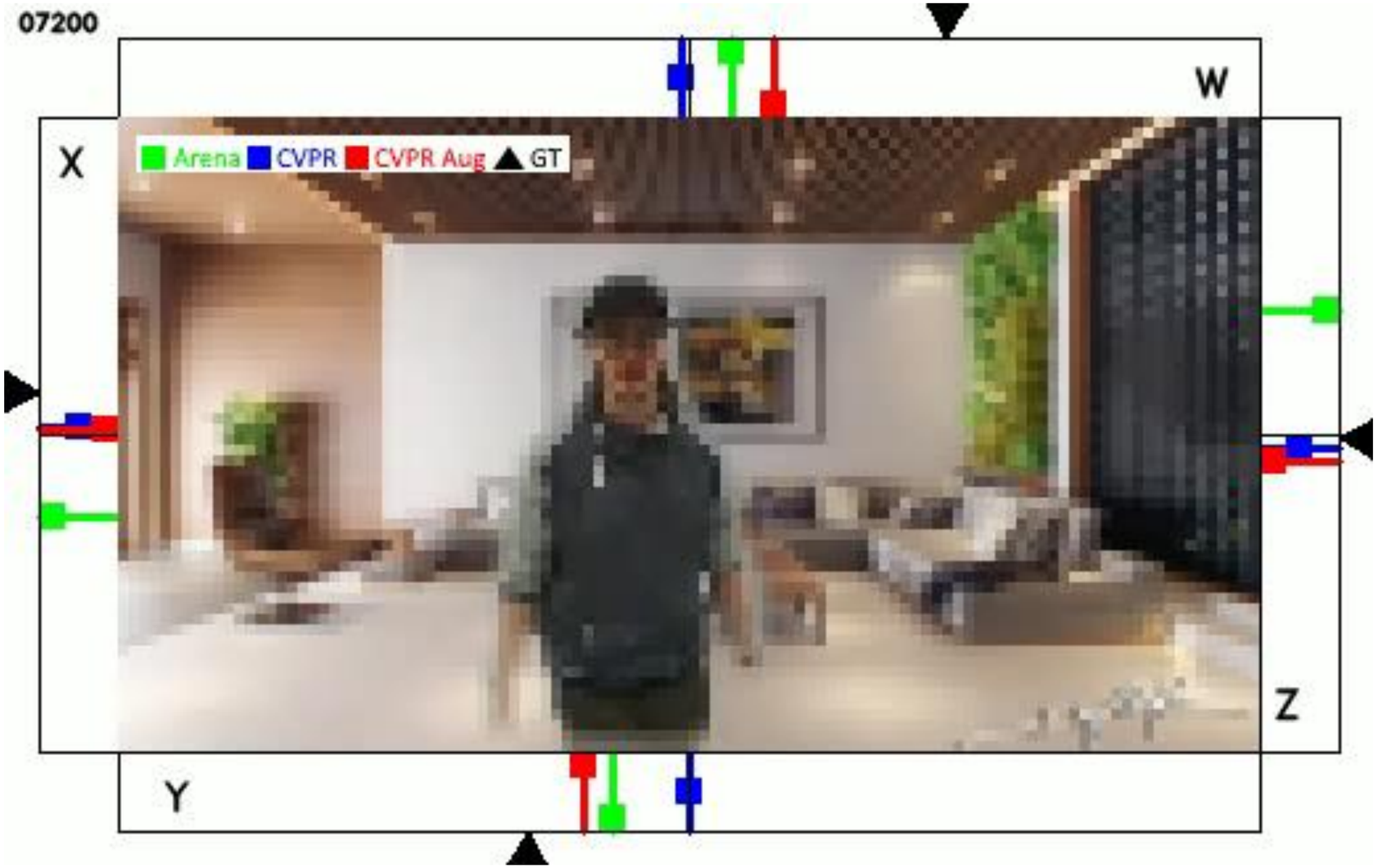


https://www.youtube.com/watch?v=VvESz9tJqE8

PER-FRAME EVALUATION: INDOOR 1 TEST SET

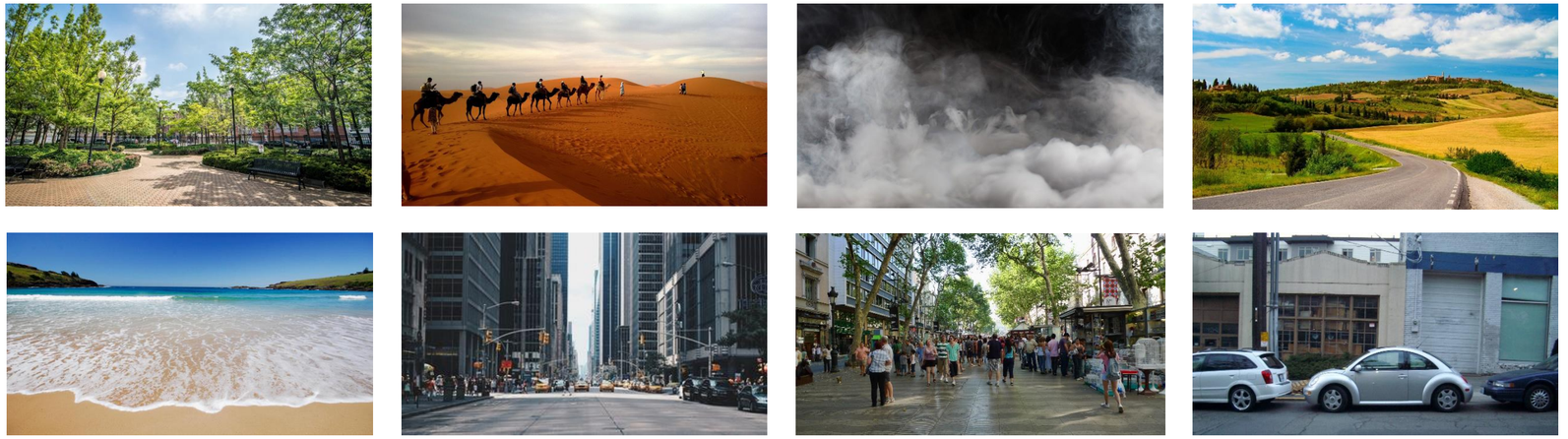
https://www.youtube.com/watch?v=-pII8ujJEng

PER-FRAME EVALUATION: INDOOR 2 TEST SET



https://www.youtube.com/watch?v=iFvtjjyx6Qs

ADDITIONAL TEST SET

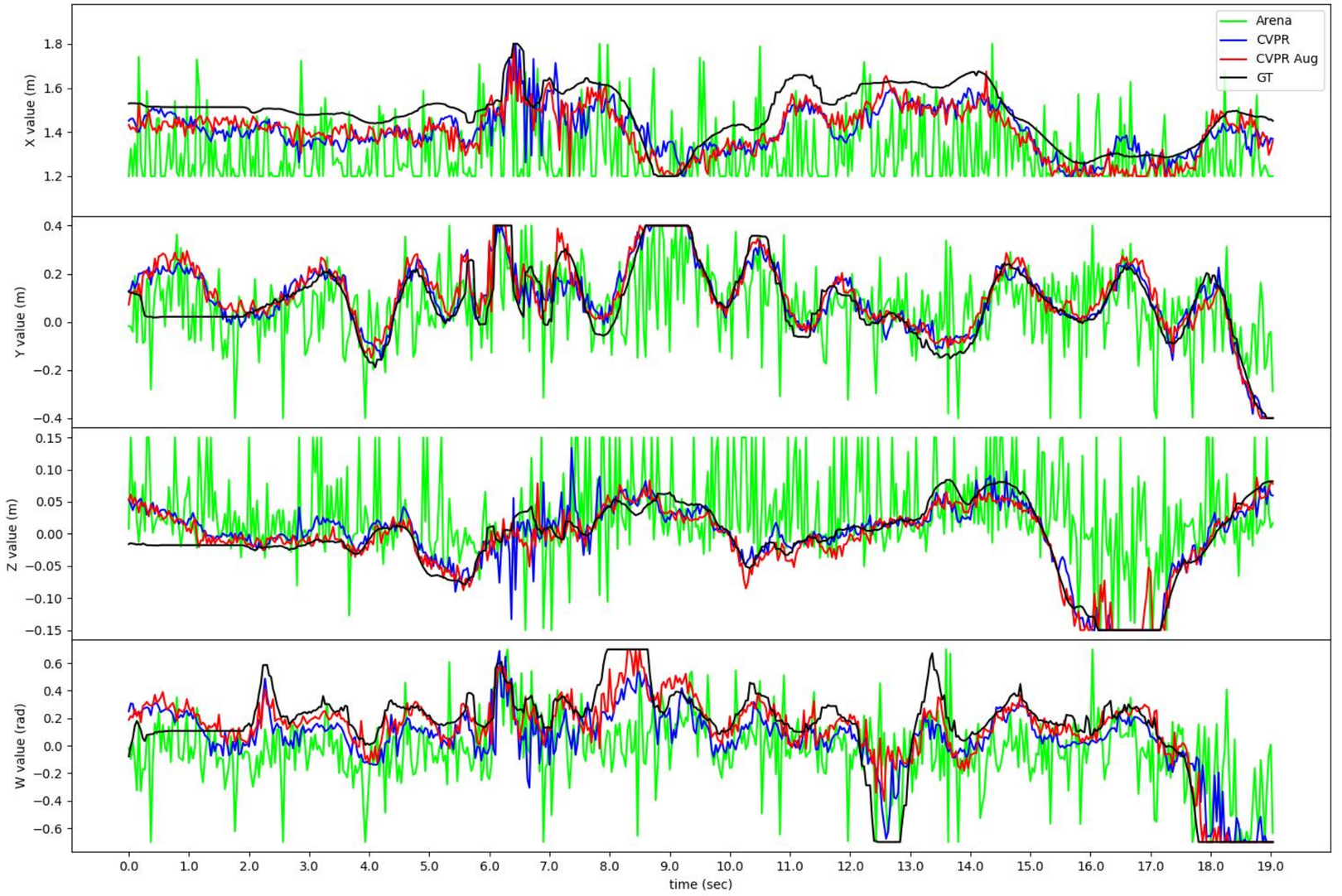


many more ...

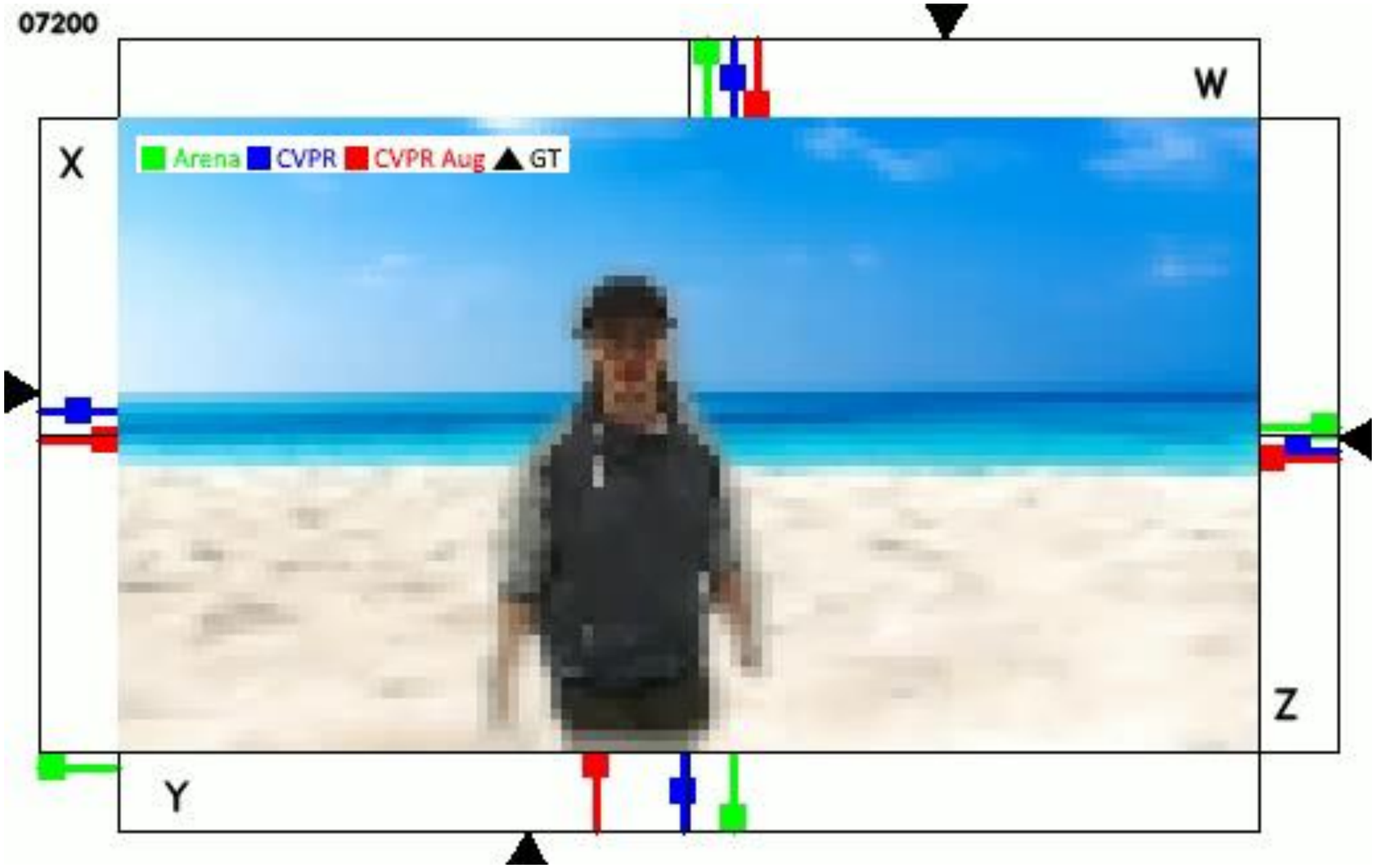
MIXED

TEST SET

TIMELINE EVALUATION: MIXED TEST SET

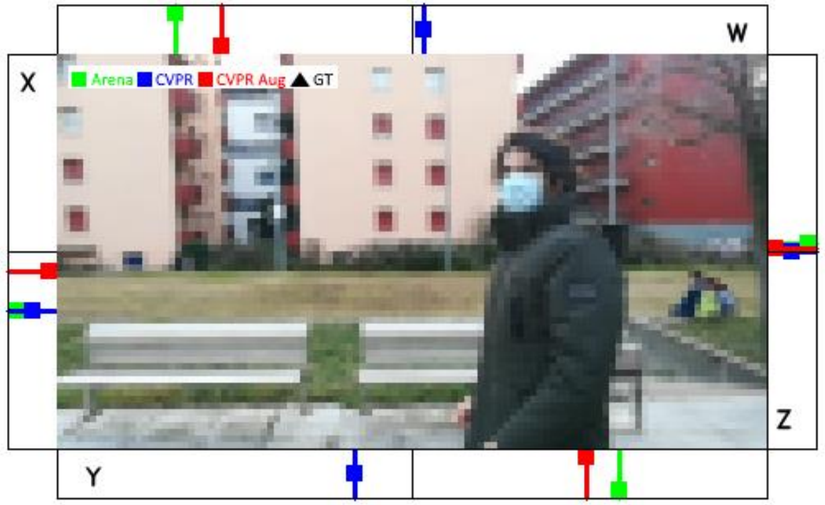


PER-FRAME EVALUATION: MIXED TEST SET



https://www.youtube.com/watch?v=MyYROdi2FR4

SMARTPHONE EVALUATION: OVERVIEW



SMARTPHONE EVALUATION: VERTICAL ALIGNMENT

https://www.youtube.com/watch?v=C37qapV6CbQ

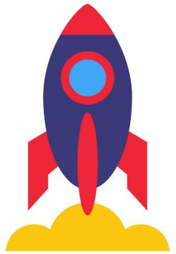
SMARTPHONE EVALUATION: USER'S DISTANCE

https://www.youtube.com/watch?v=1w1G246QR6M

CONCLUSIONS

QUANTITATIVE AND QUALITATIVE EVALUATION
PROVED THAT THE AUGMENTATION STRATEGY
IS ABLE TO ENHANCE MODEL'S GENERALIZATION

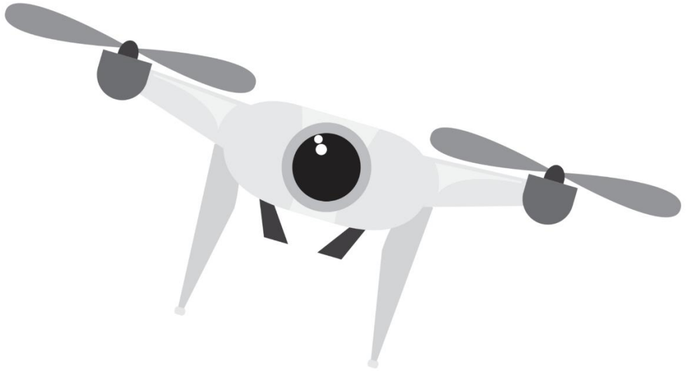
BOTH BACKGROUND REPLACEMENT AND
CLASSIC IMAGE AUGMENTATION ARE
REQUIRED TO OBTAIN ROBUST RESULTS



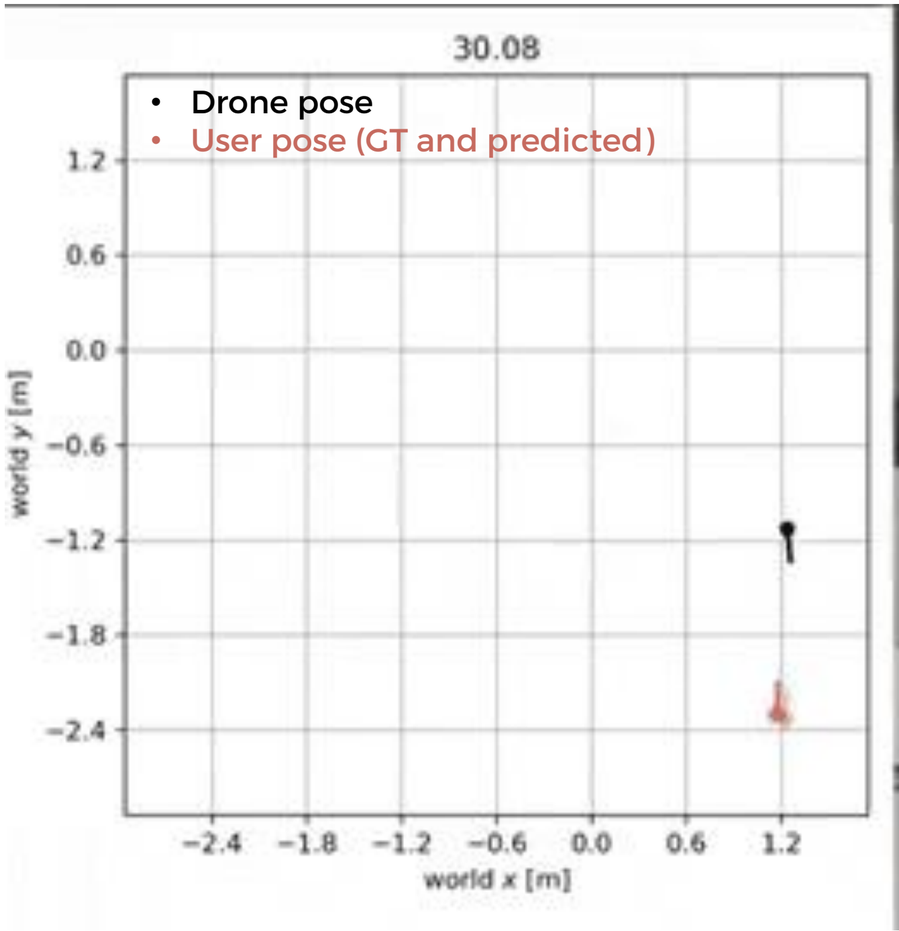
OUR APPROACH IS EXTENDABLE TO OTHER TASKS
WHICH REQUIRES OBJECT OR HUMAN RECOGNITION

FUTURE WORKS

- TEST THE MODEL ON THE **REAL DRONE (Parrot Bebop 2)**
- FURTHER **INTERPRETATION** ON THE NEW MODEL
- TEST THE AUGMENTATION STRATEGY ON **DIFFERENT DRONES**



FUTURE WORKS: CRAZYFLIE



THANKS FOR YOUR KIND ATTENTION

Supervisor: Dr. Alessandro Giusti
Co-Supervisor: PhD Stud. Dario Mantegazza

Master Thesis by
MARCO FERRI
ID 80730

ADDITIONAL SLIDES

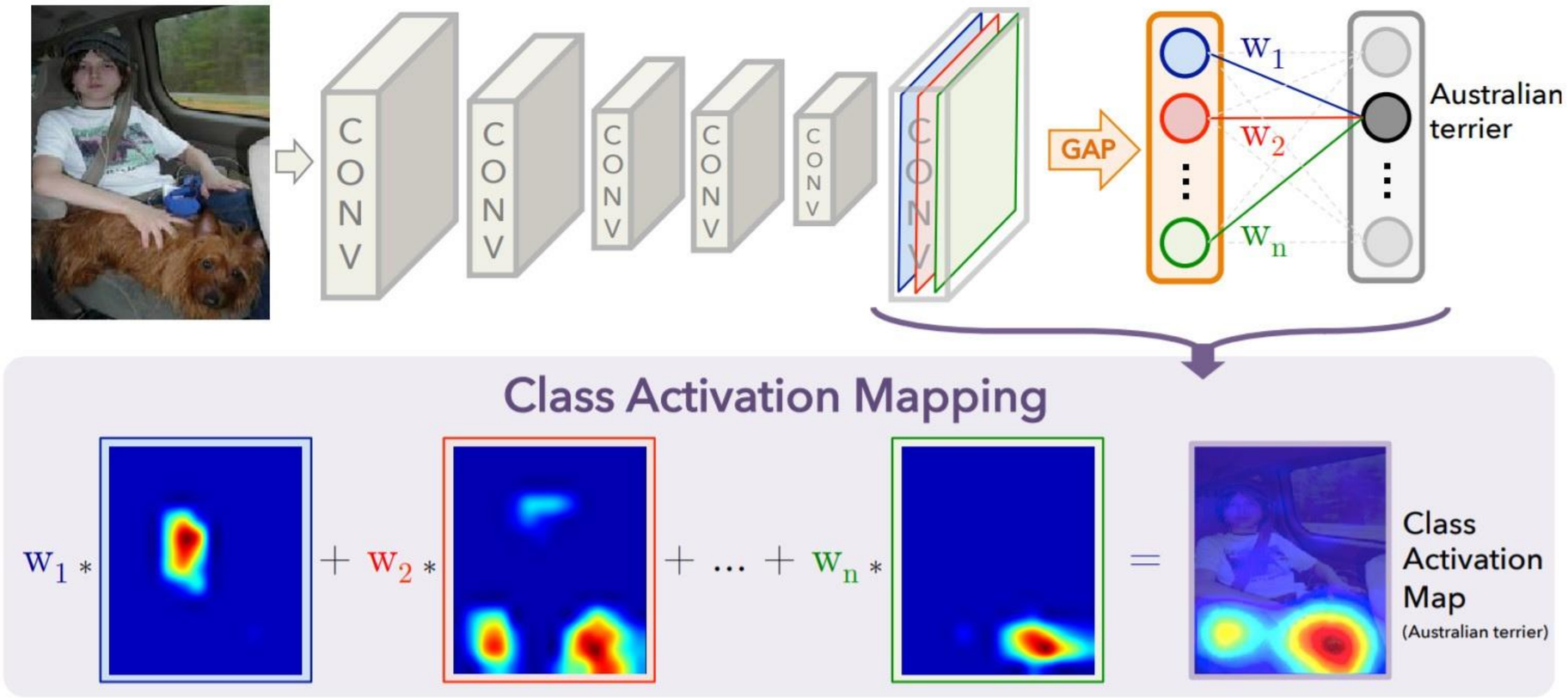
HOW GRAD-CAM WORKS

Grad-CAM for “Cat”

Grad-CAM for “Dog”

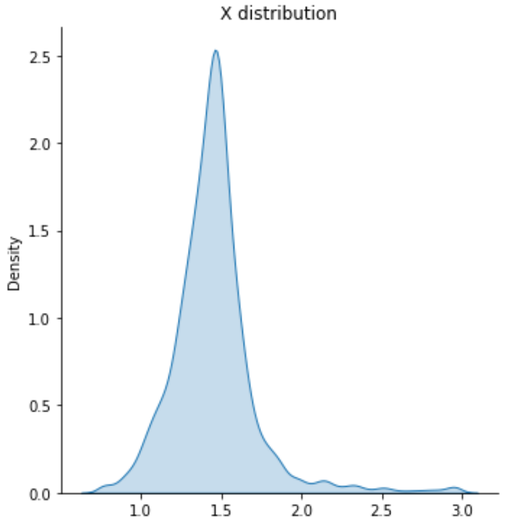
[Selvaraju et al. 2019]

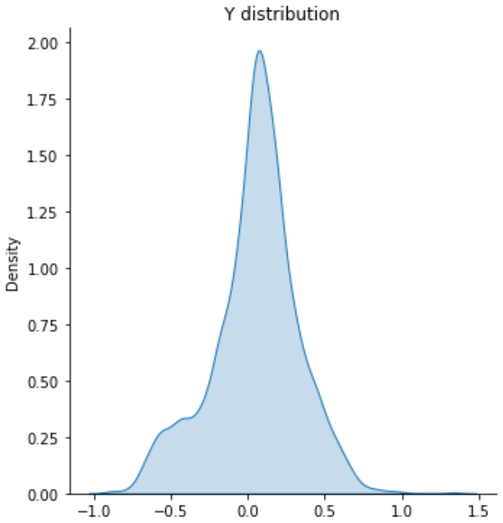
HOW GRAD-CAM WORKS

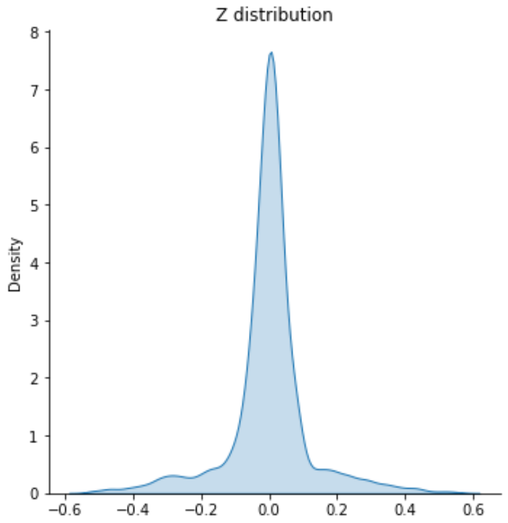


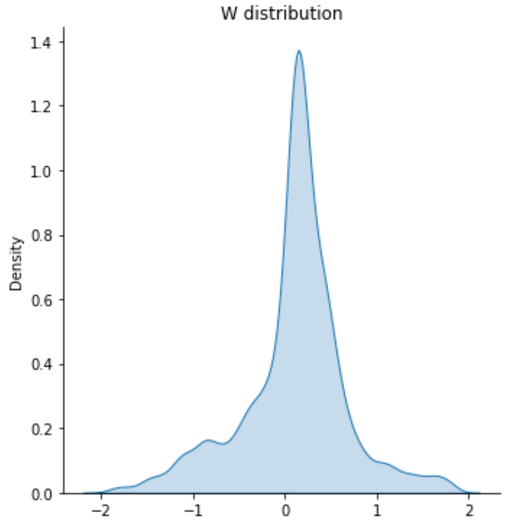
[Selvaraju et al. 2019]

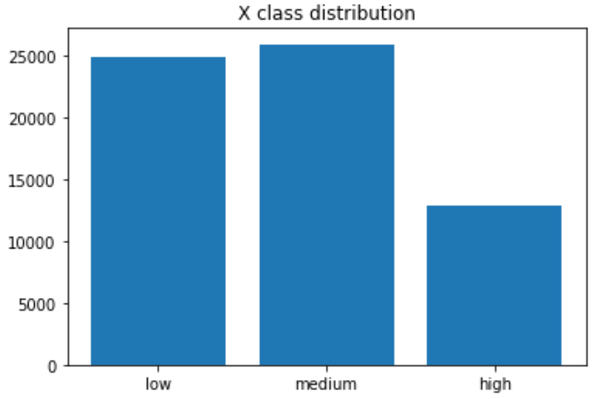
GRAD-CAM: FROM REGRESSION TO CLASSIFICATION

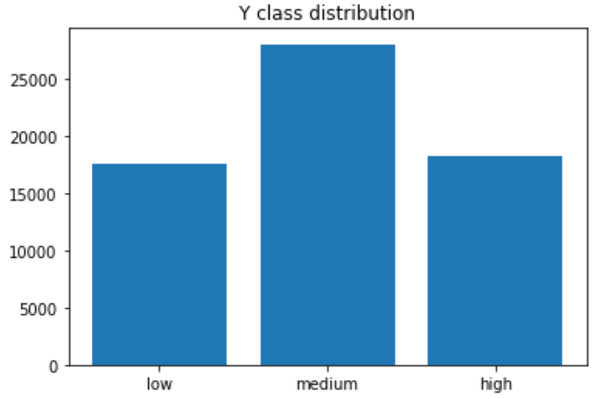


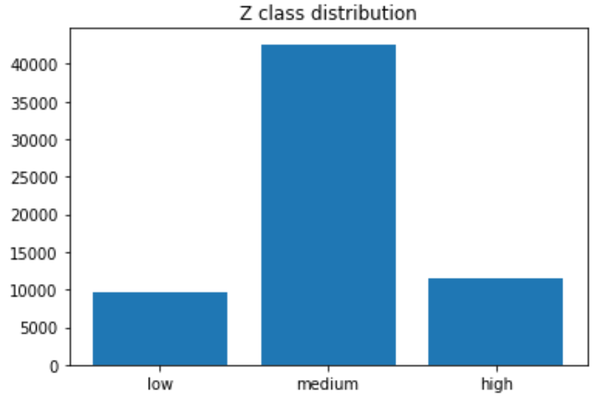




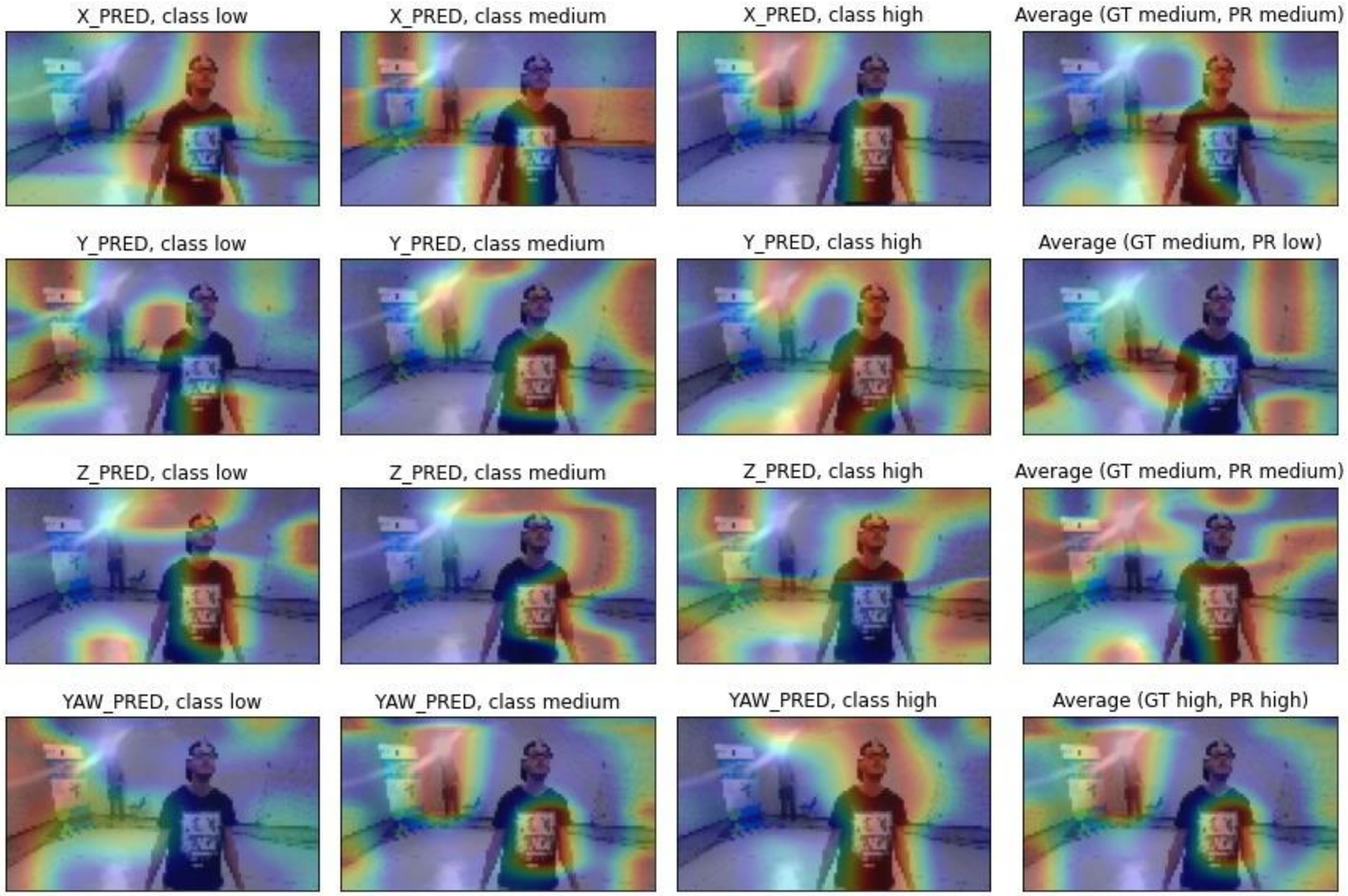




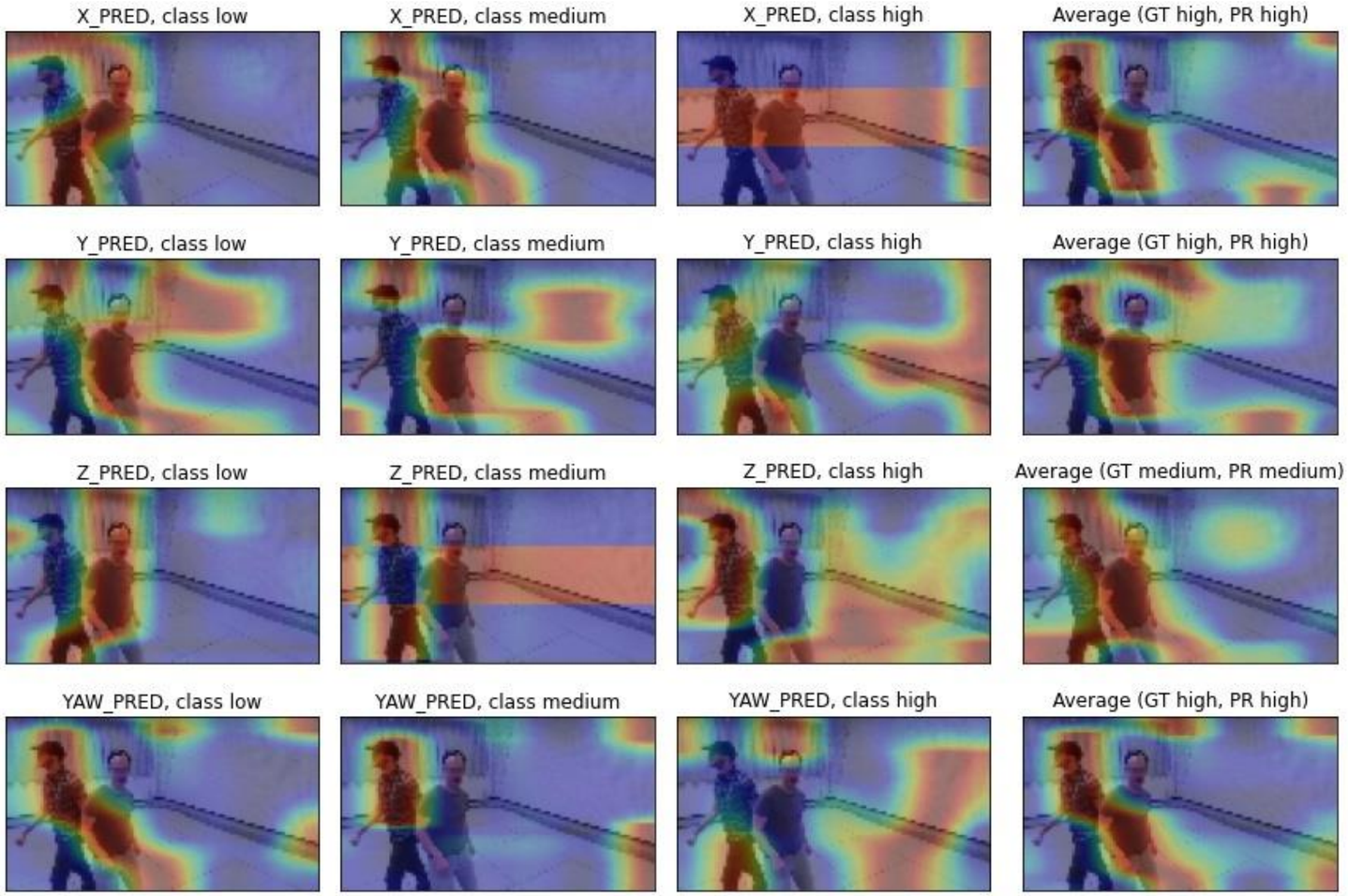




GRAD-CAM PER CLASS



GRAD-CAM PER CLASS



SMARTPHONE EVALUATION: MULTIPLE PEOPLE

https://www.youtube.com/watch?v=KWb3RNbr7Lk

SMARTPHONE EVALUATION: FAST MOVEMENTS

https://www.youtube.com/watch?v=zrp12f4aYCI

SMARTPHONE EVALUATION: FOREGROUND PERSON

https://www.youtube.com/watch?v=KWb3RNbr7Lk