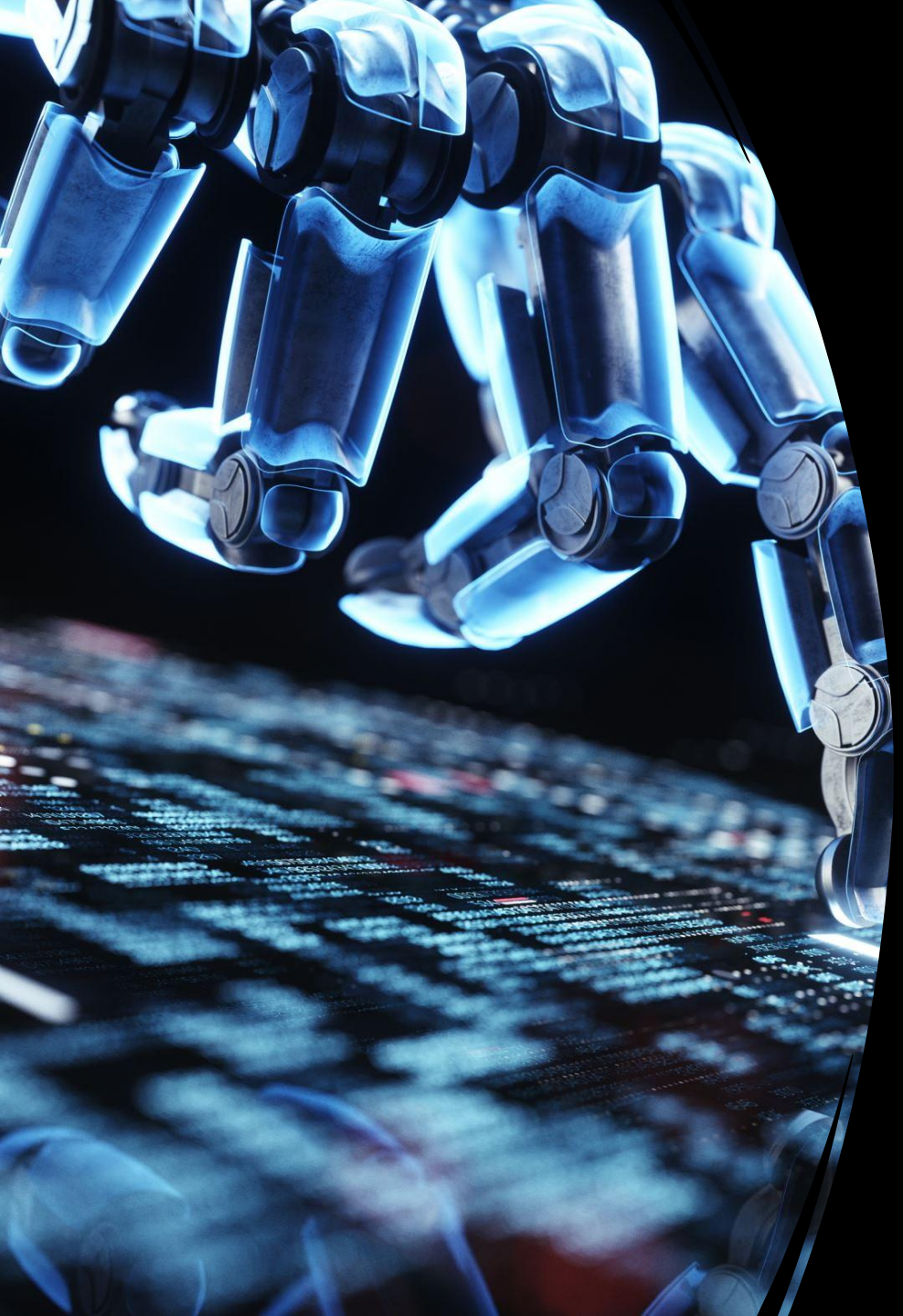
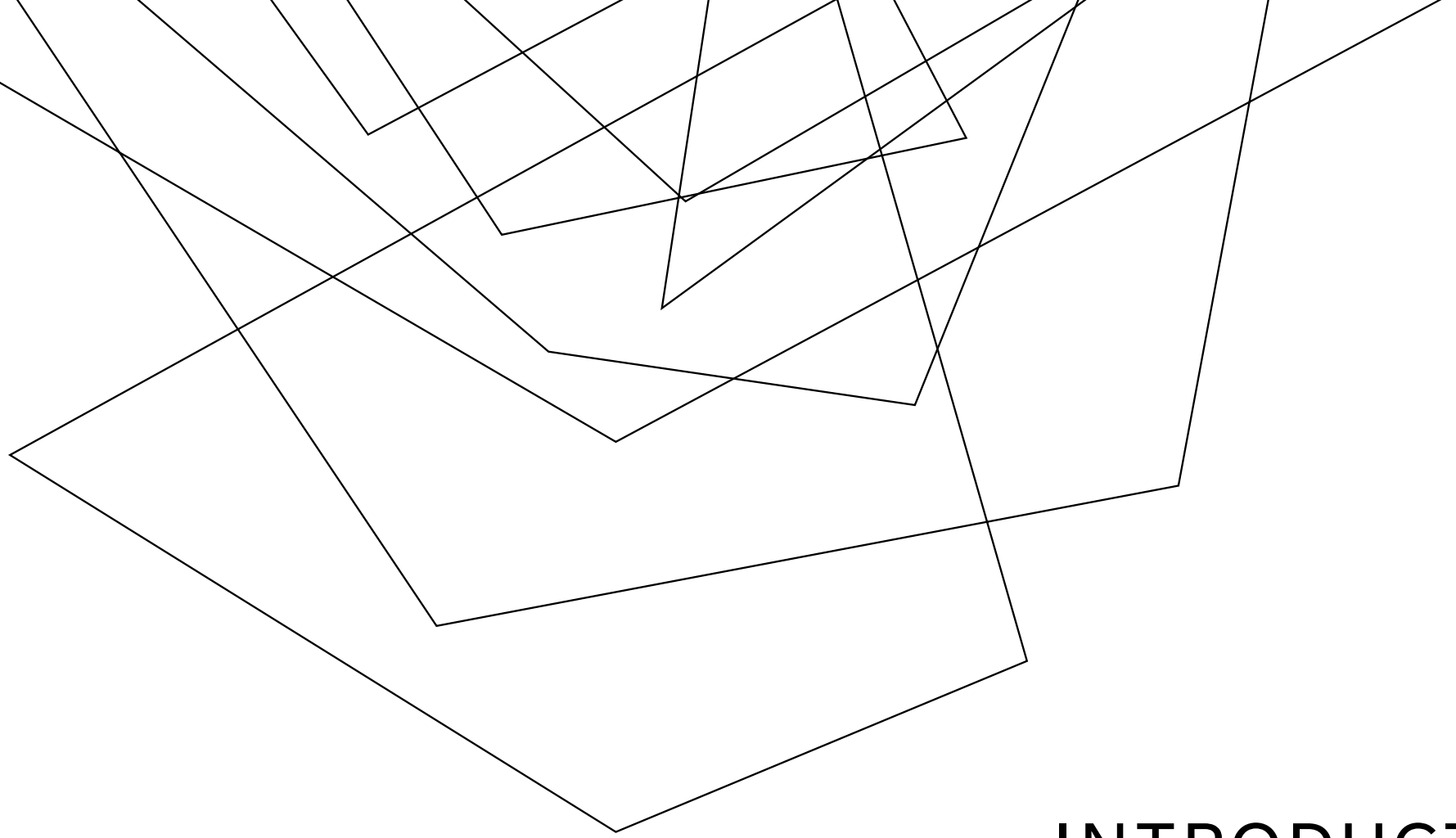


FROM VISION TO TYPING: AI AUTOMATION FOR MASTERING TYPING GAMES

Naris Chanpaiboonkit

DADS 4

6520422025



INTRODUCTION

INTRODUCTION

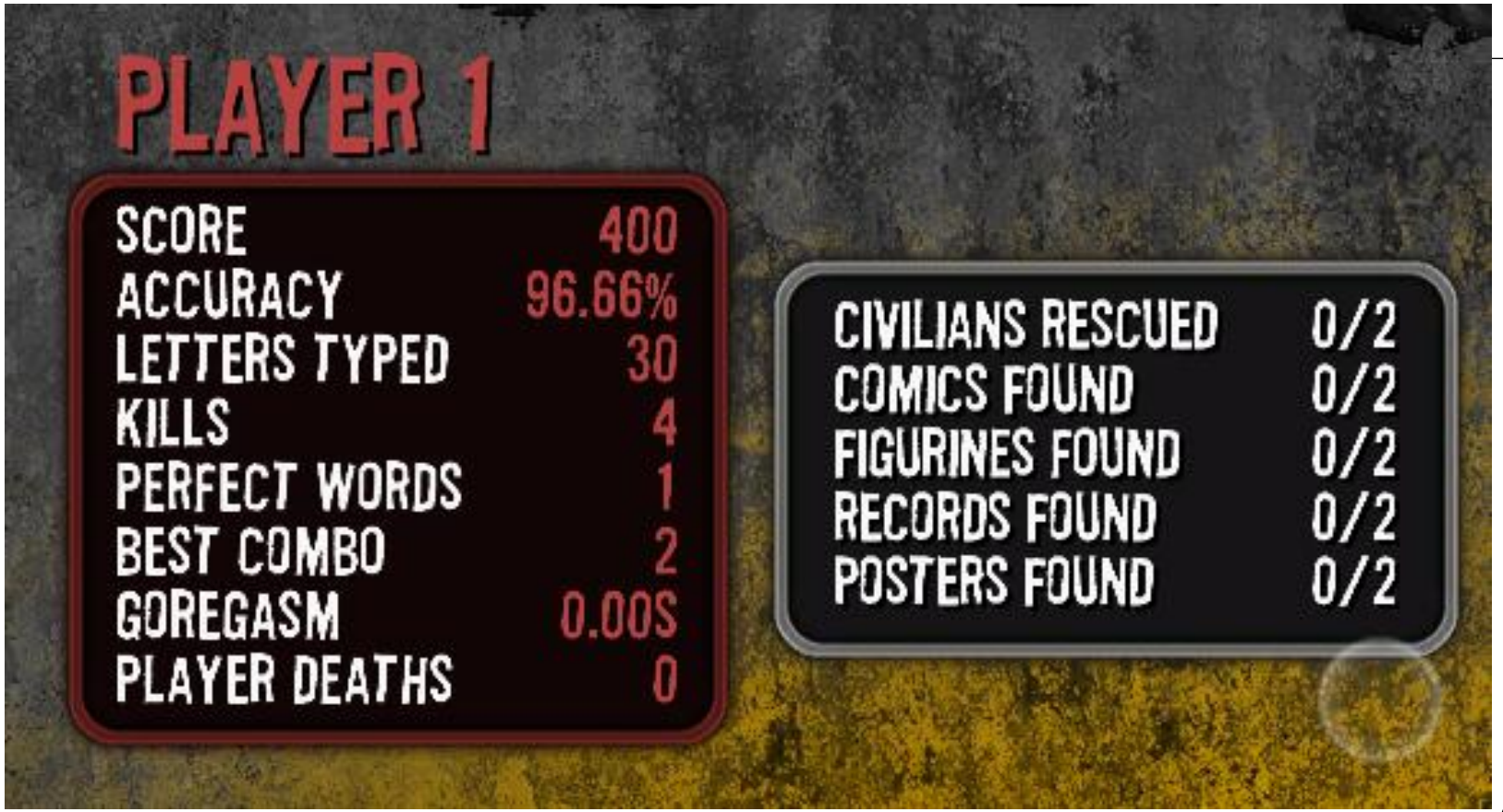
- Develop an automated typing system to achieve high scores efficiently.
- Optimize real-time text recognition to enhance input speed

MAIN OBJECTIVE

MAIN OBJECTIVE

At the highest difficulty, typing accurately and completing certain stages is extremely challenging, emphasizing the need for automation.

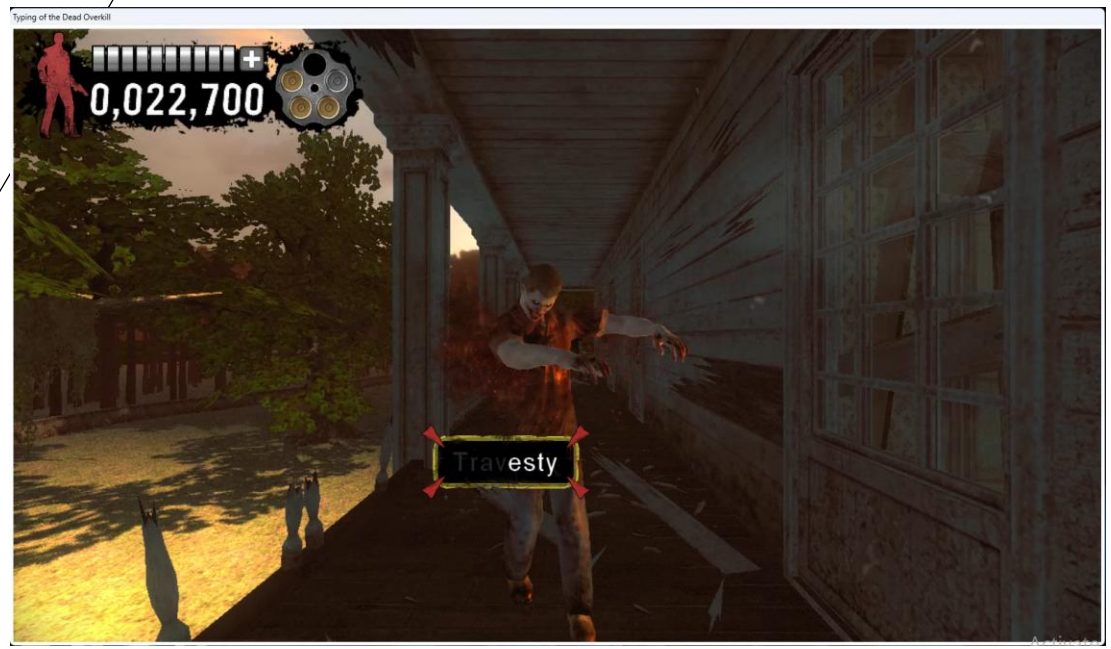
MAIN OBJECTIVE

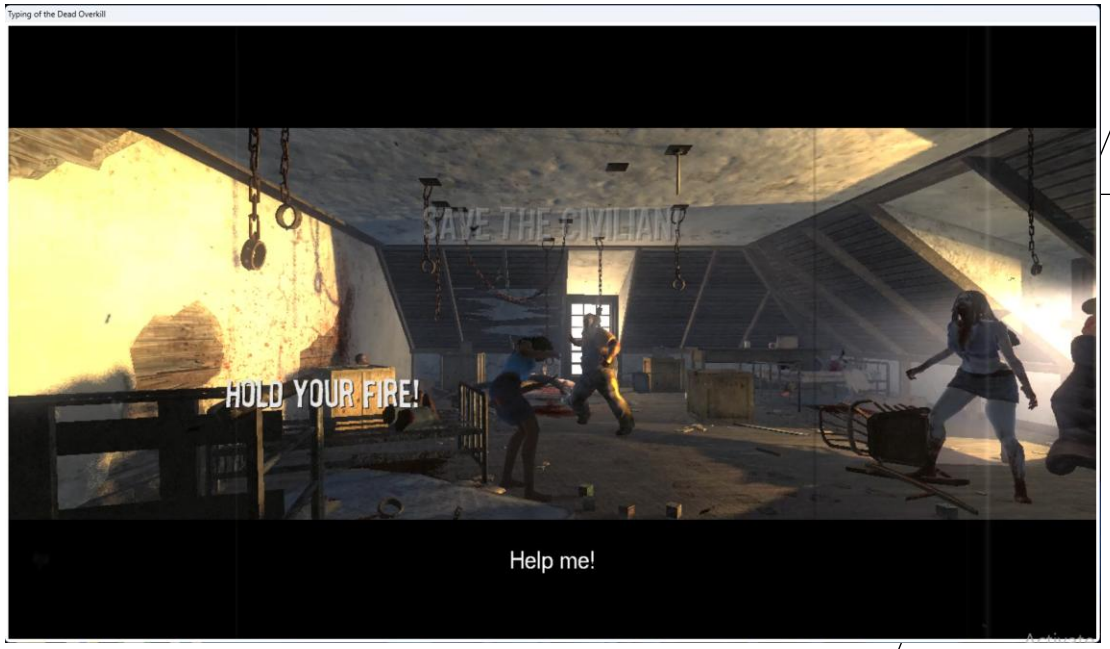


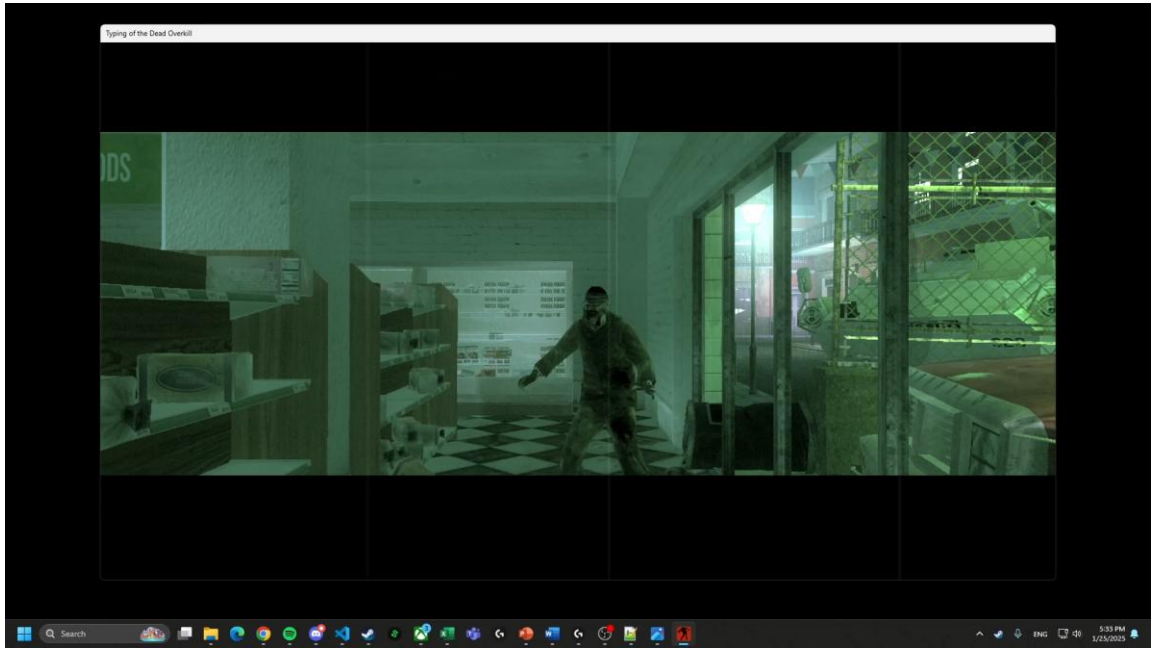
MAIN OBJECTIVE

MAIN OBJECTIVE

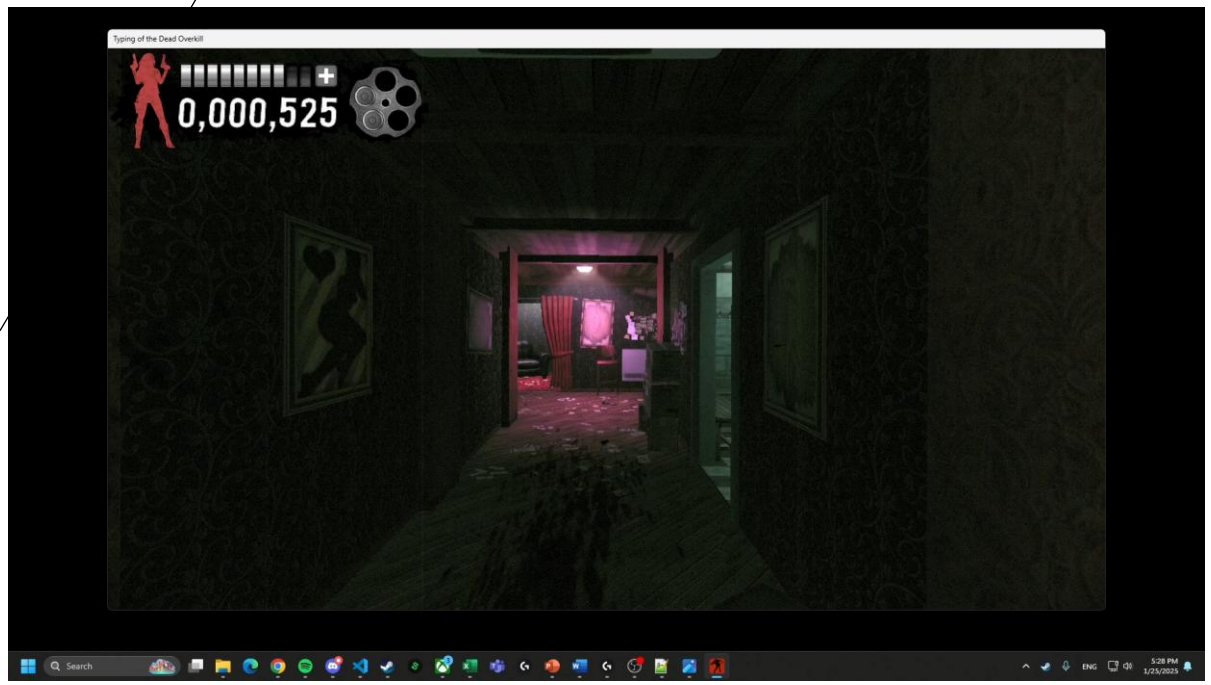
MAIN OBJECTIVE

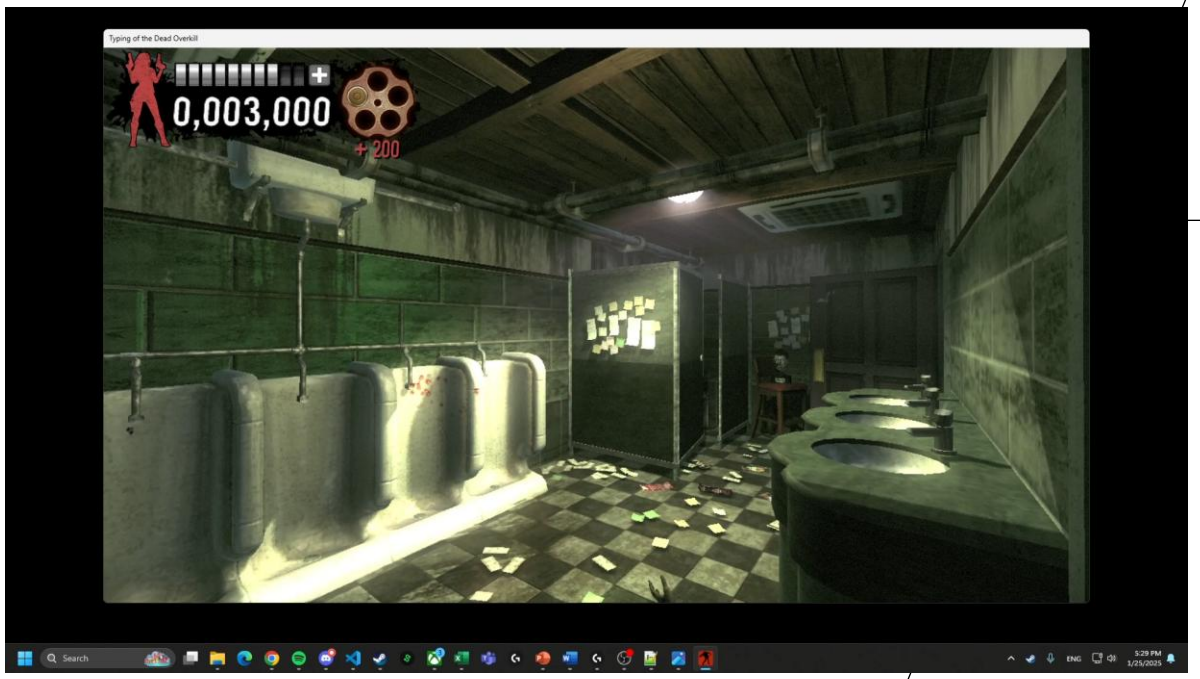


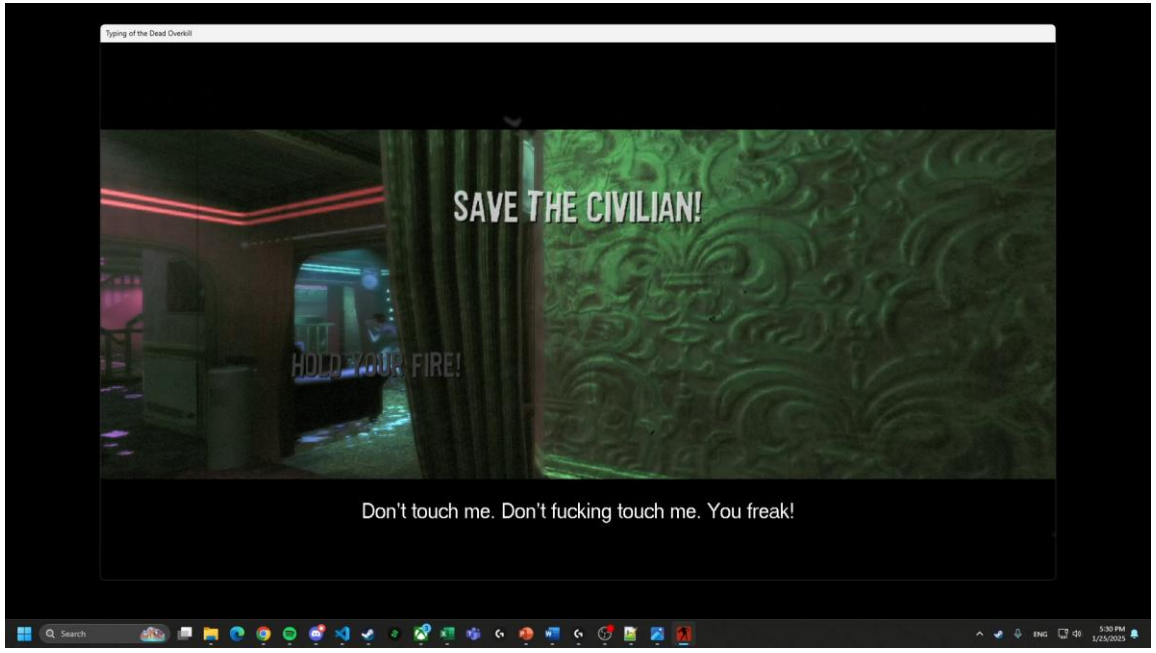


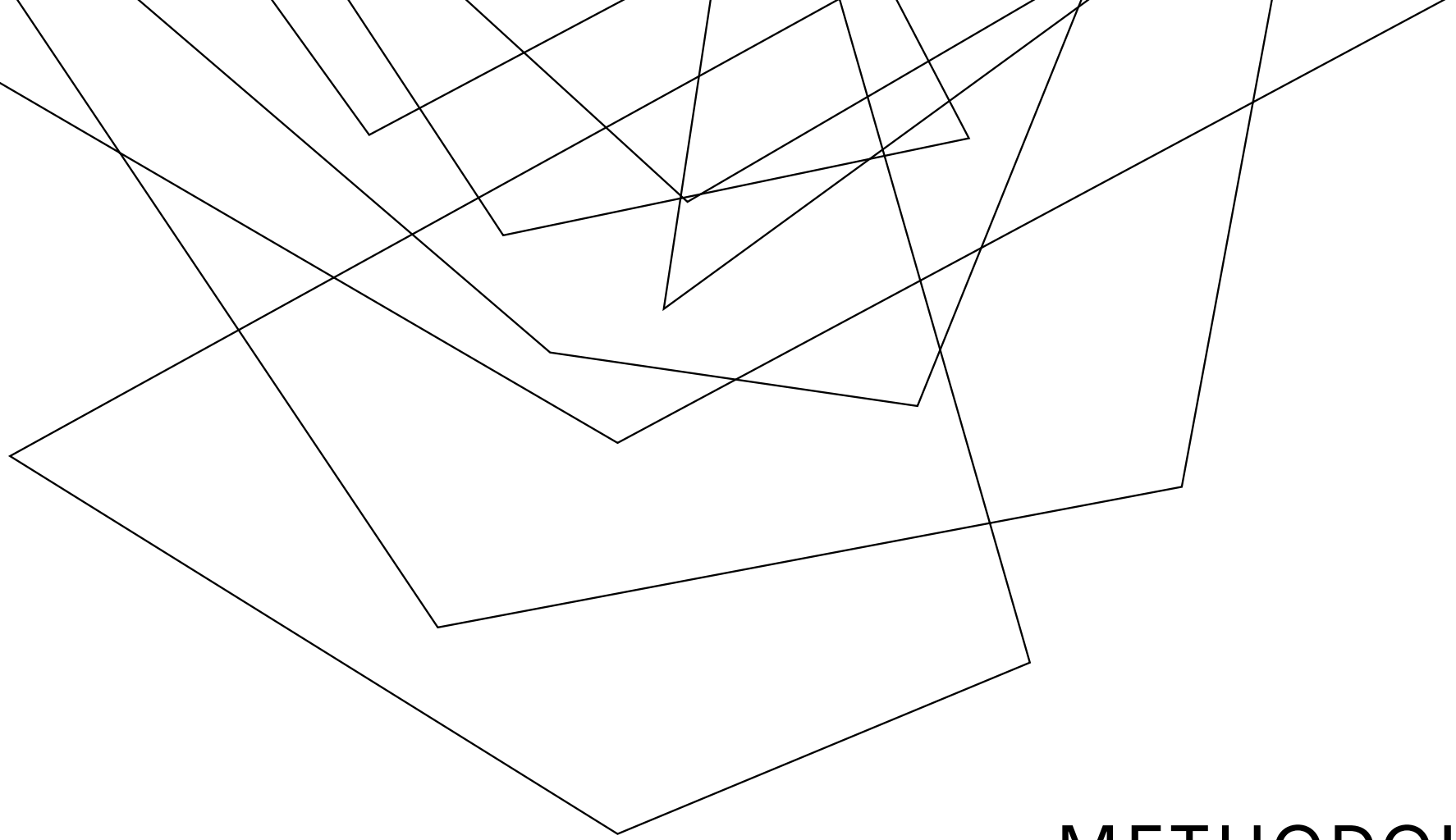


MAIN OBJECTIVE









METHODOLOGY

METHODOLOGY

Knowledge

• **Computer Vision:**
Detect text and objects using **YOLOv11**; preprocess and recognize text via OCR frameworks **EasyOCR or Tesseract**.

• **NLP:**
Tokenize, translate, and correct recognized text for accurate in-game commands using tools like **SymSpell**.

• **Automation:**
Automate typing and interactions with **PyAutoGUI**.

METHODOLOGY – YOLO V11

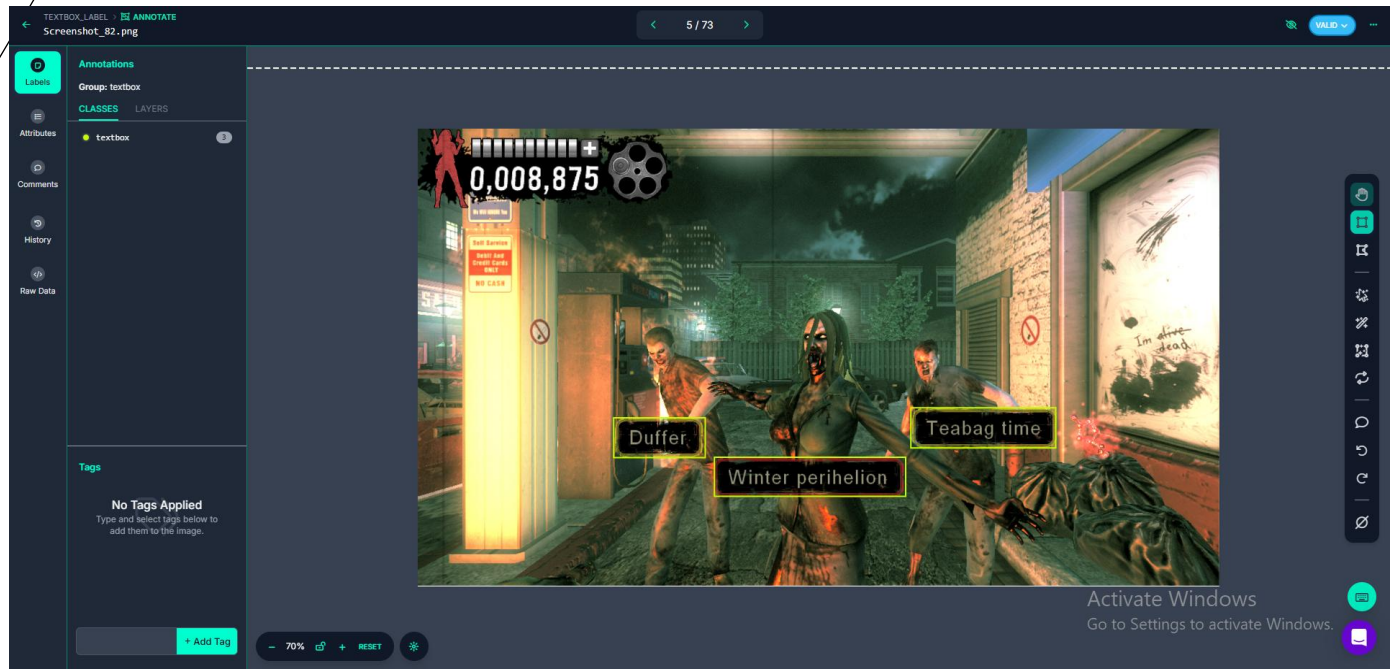
Data Collection

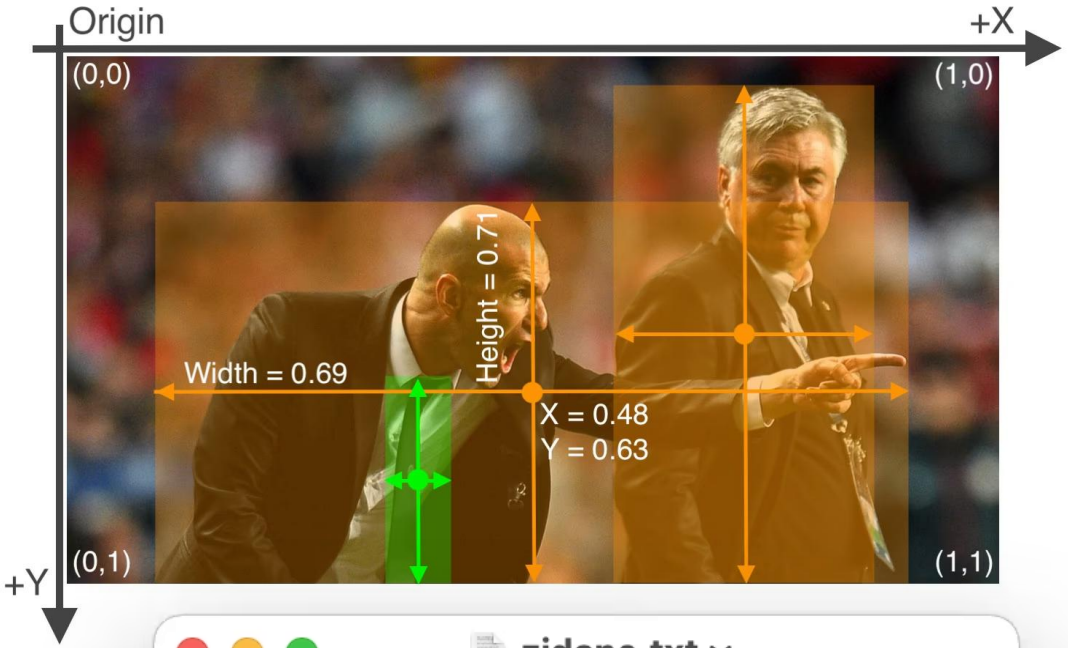
METHODOLOGY – YOLO V11

Data Labeling

METHODOLOGY – YOLO V11

Data Labeling





zidane.txt

```
0  0.481719 0.634028 0.690625 0.713278
0  0.741094 0.524306 0.314750 0.933389
27 0.364844 0.795833 0.078125 0.400000
```

"Class X Y Width Height"

METHODOLOGY – YOLO V11

Data Preprocesing

Labeled Images

Number of Images

360

1 missing annotations

1 null examples

Number of Annotations

687

1.9 per image (average)

Across 8 classes

8 Classes

- textbox 456
- quick_text 58
- slow 57
- health 41
- figure 26
- record 23
- comic 17
- poster 9

Preprocessing Resize: Stretch to 640x640

Augmentations

Outputs per training example: 3

Crop: 0% Minimum Zoom, 21% Maximum Zoom

Rotation: Between -12° and +12°

Grayscale: Apply to 15% of images

Labeled and Augmented Images

Dataset Split

TRAIN SET 94%

912 Images

VALID SET 6%

56 Images

METHODOLOGY – YOLO V11

Our brain

YOLO V11

Model	size (pixels)	mAP^{val} 50-95	Speed CPU ONNX (ms)	Speed T4 TensorRT10 (ms)	params (M)	FLOPs (B)
YOLO11n	640	39.5	56.1 ± 0.8	1.5 ± 0.0	2.6	6.5
YOLO11s	640	47.0	90.0 ± 1.2	2.5 ± 0.0	9.4	21.5
YOLO11m	640	51.5	183.2 ± 2.0	4.7 ± 0.1	20.1	68.0
YOLO11l	640	53.4	238.6 ± 1.4	6.2 ± 0.1	25.3	86.9
YOLO11x	640	54.7	462.8 ± 6.7	11.3 ± 0.2	56.9	194.9

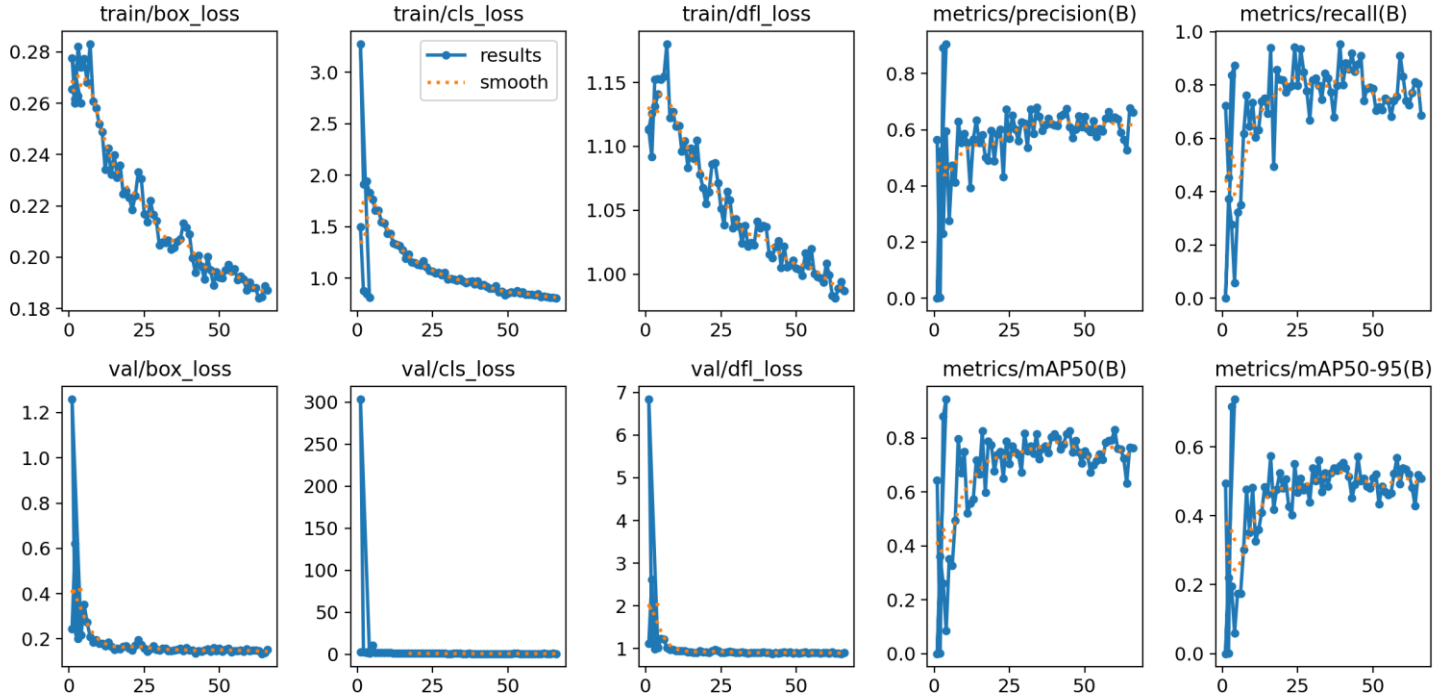
Training configuration

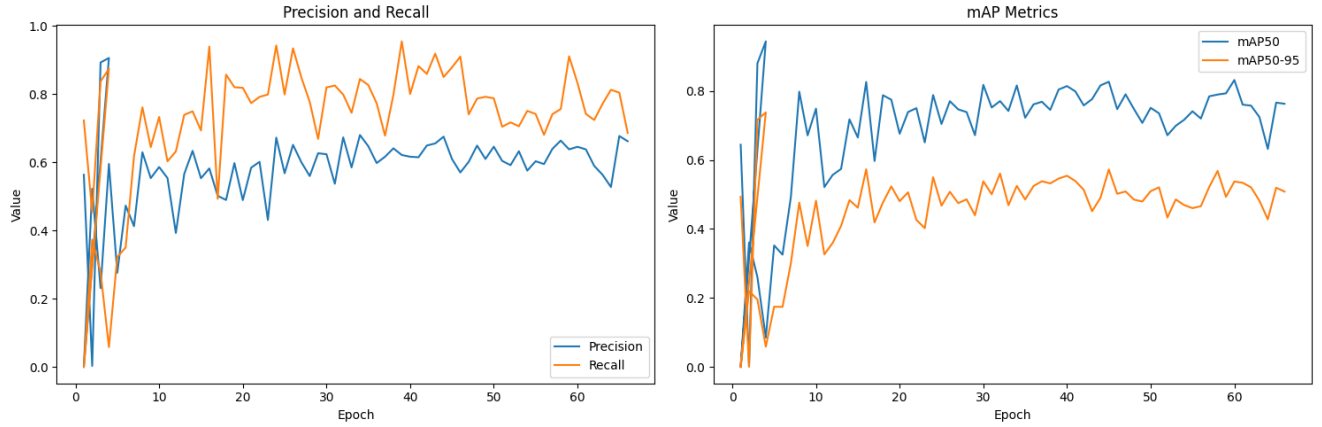
Parameter	Value	Description
epochs	300	Number of training rounds
imgsz	640	Input image size
patience	50	Early stopping patience
lr0	0.01	Initial learning rate
optimizer	AdamW	Optimizer type
warmup_epochs	5	Number of warmup epochs
cls	0.5	Classification Loss weight
dfl	1.5	Distribution Focal Loss weight
box	7.5	Box Loss weight

METHODOLOGY – YOLO V11

YOLOv11
Training result

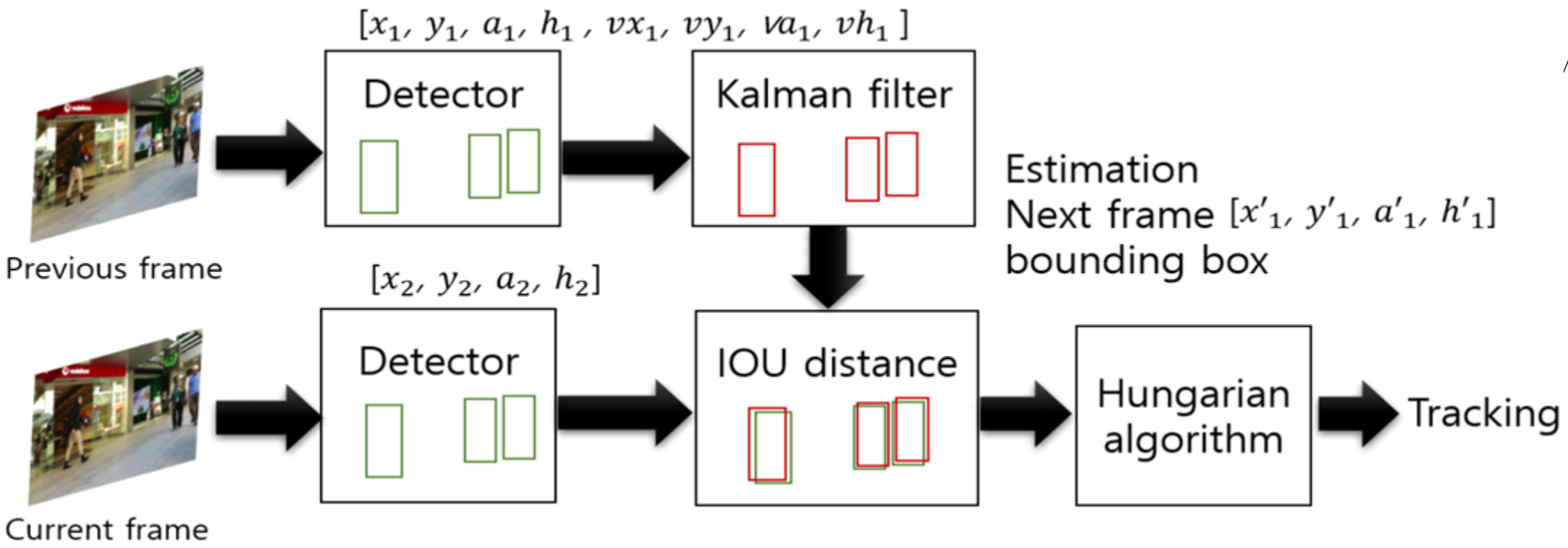
```
 Epoch    GPU_mem   box_loss   cls_loss   dfl_loss  Instances       Size
66/300      16.6G     0.1871     0.8045     0.9869         47        640: 100%|██████████| 57/57 [03:33<00:00,  3.74s/it]
               Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 100%|██████████| 2/2 [00:10<00:00,  5.25s/it]
                 all         56        119      0.662      0.686      0.763      0.508
```





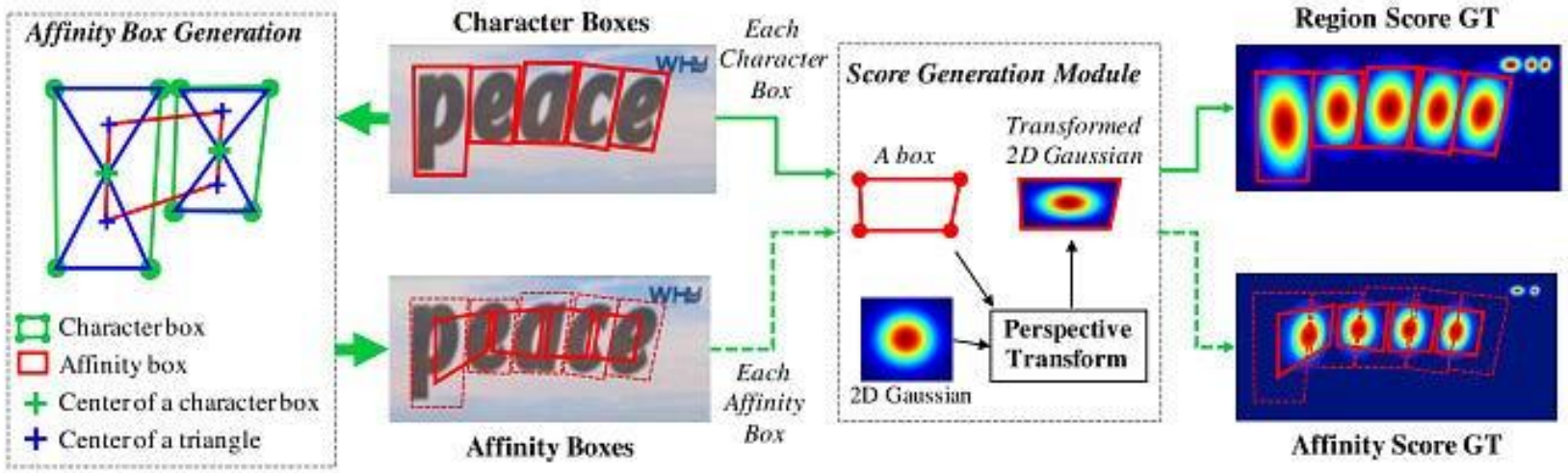
METHODOLOGY – SORT (SIMPLE ONLINE AND REALTIME TRACKING)

YOLOv11
Training result

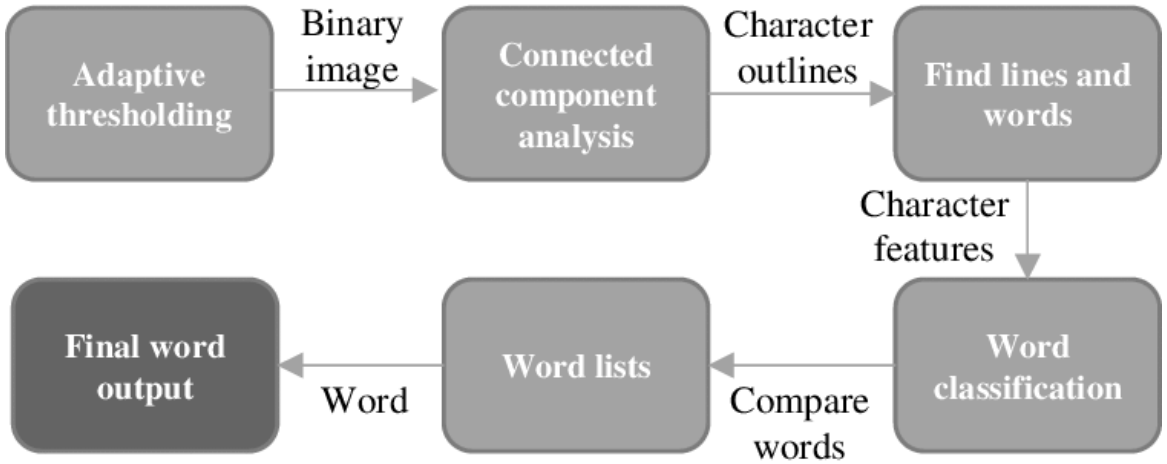


METHODOLOGY – OCR

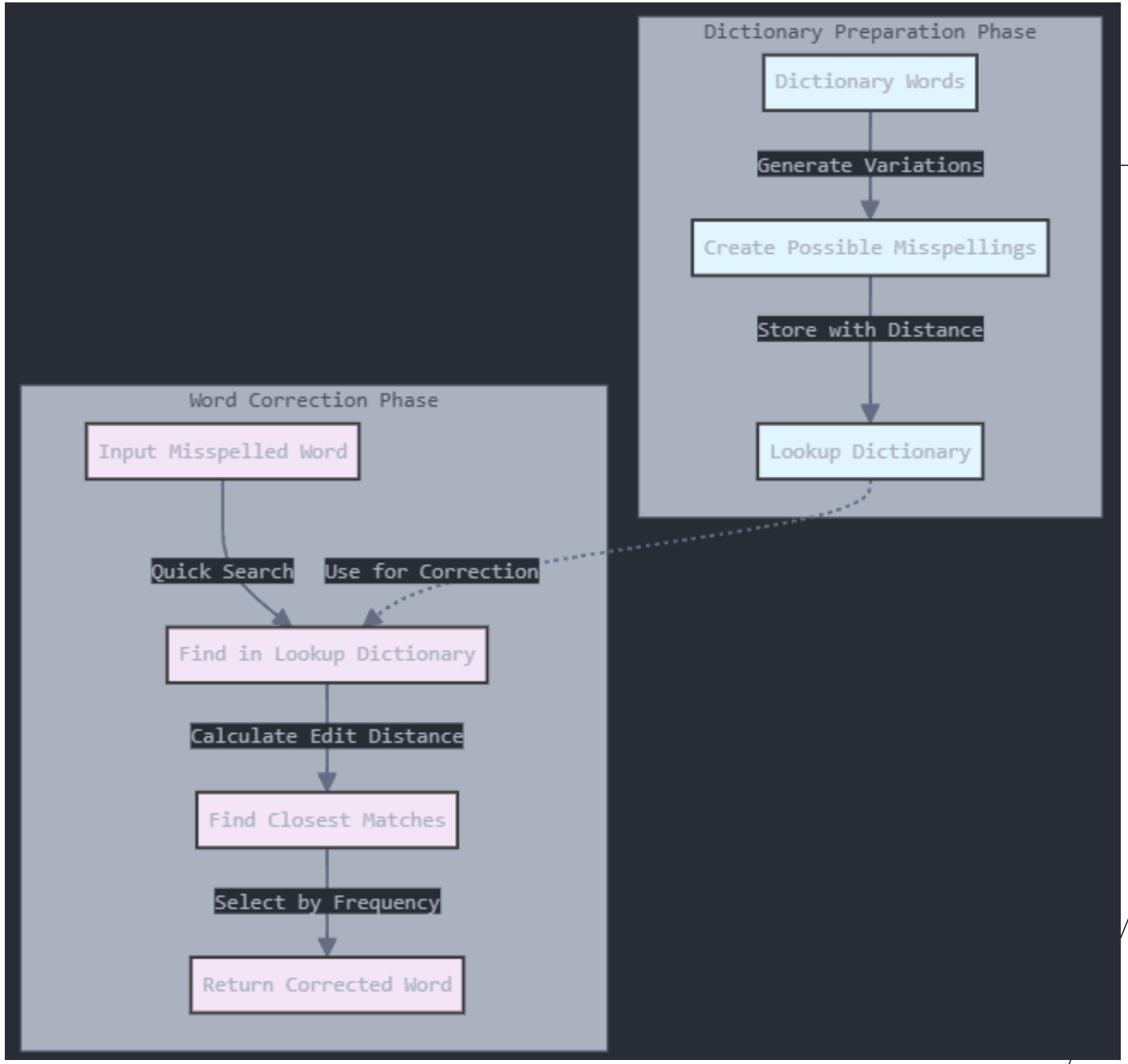
EasyOCR(CRAFT)



Tesseract



METHODOLOGY – SYMSPELL

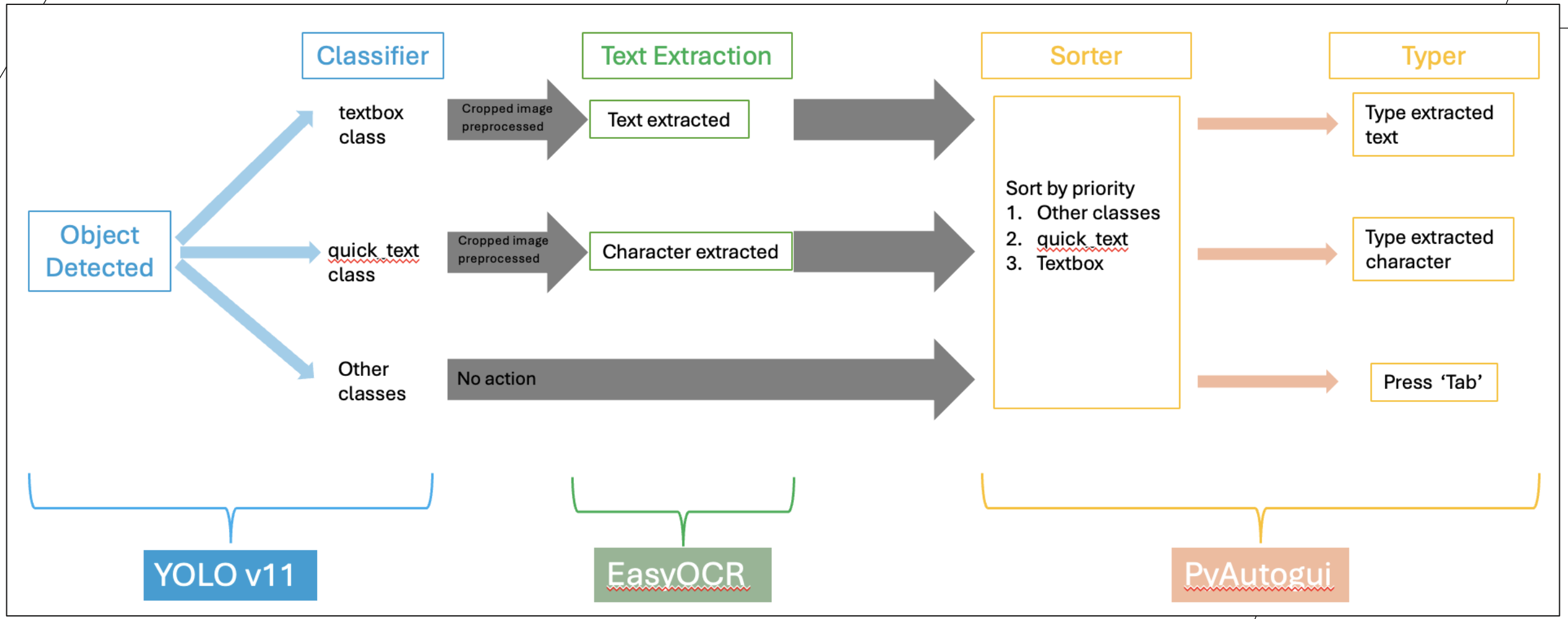


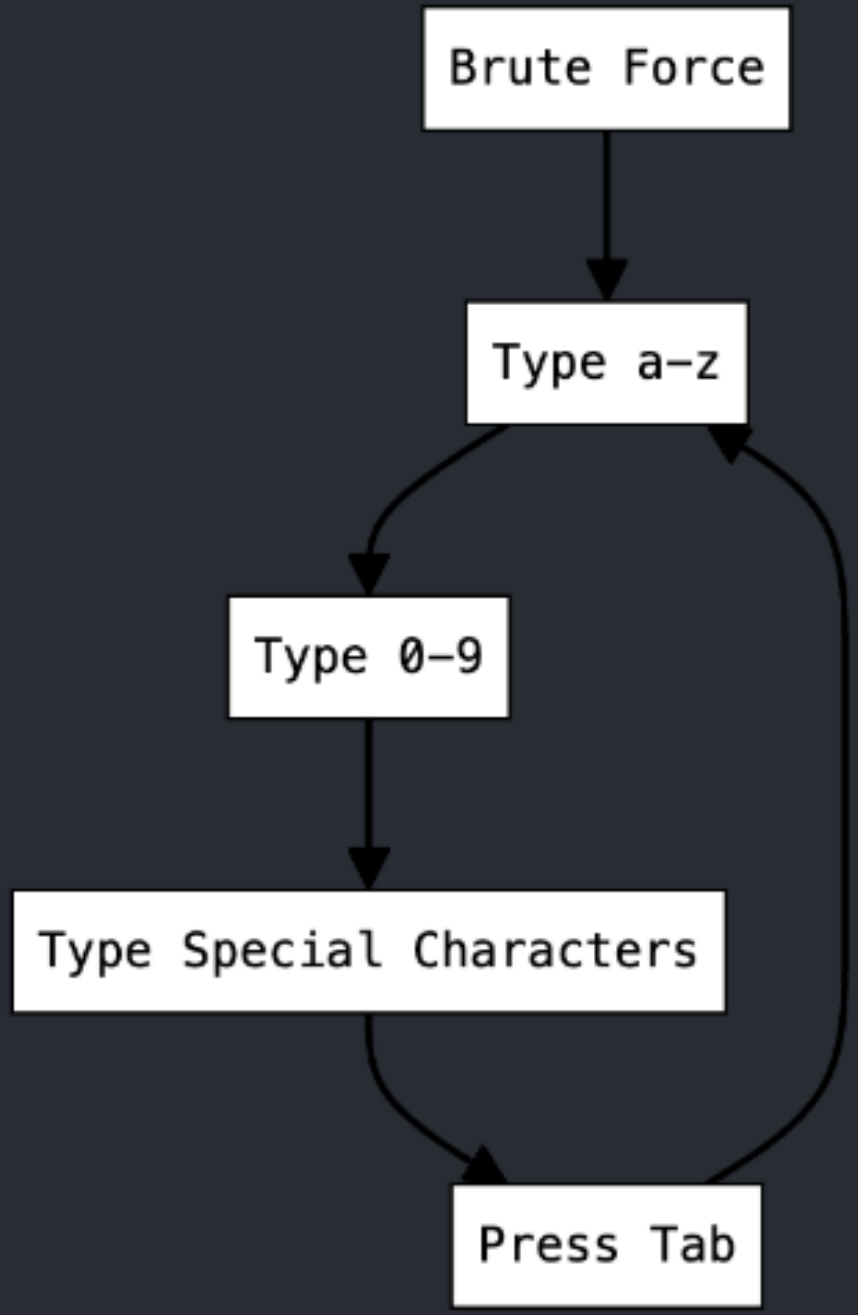
METHODOLOGY

5 **Methods**

1) Brute Force
2) YOLOv11 + EasyOCR
3) YOLOv11 + Tesseract OCR
4) YOLOv11 + EasyOCR + SymSpell
5) YOLOv11 + EasyOCR + SymSpell + SORT + Brute Force

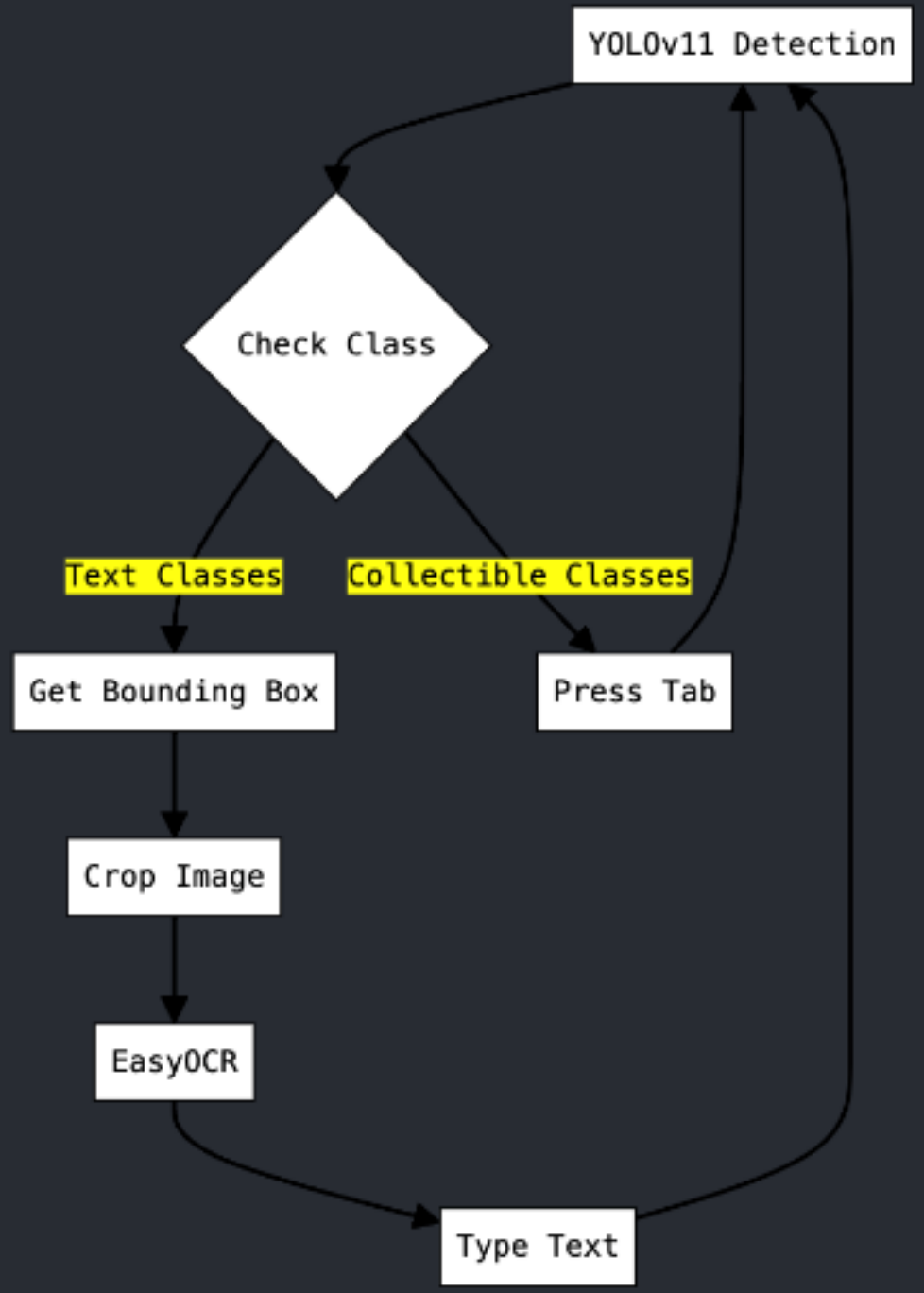
METHODOLOGY





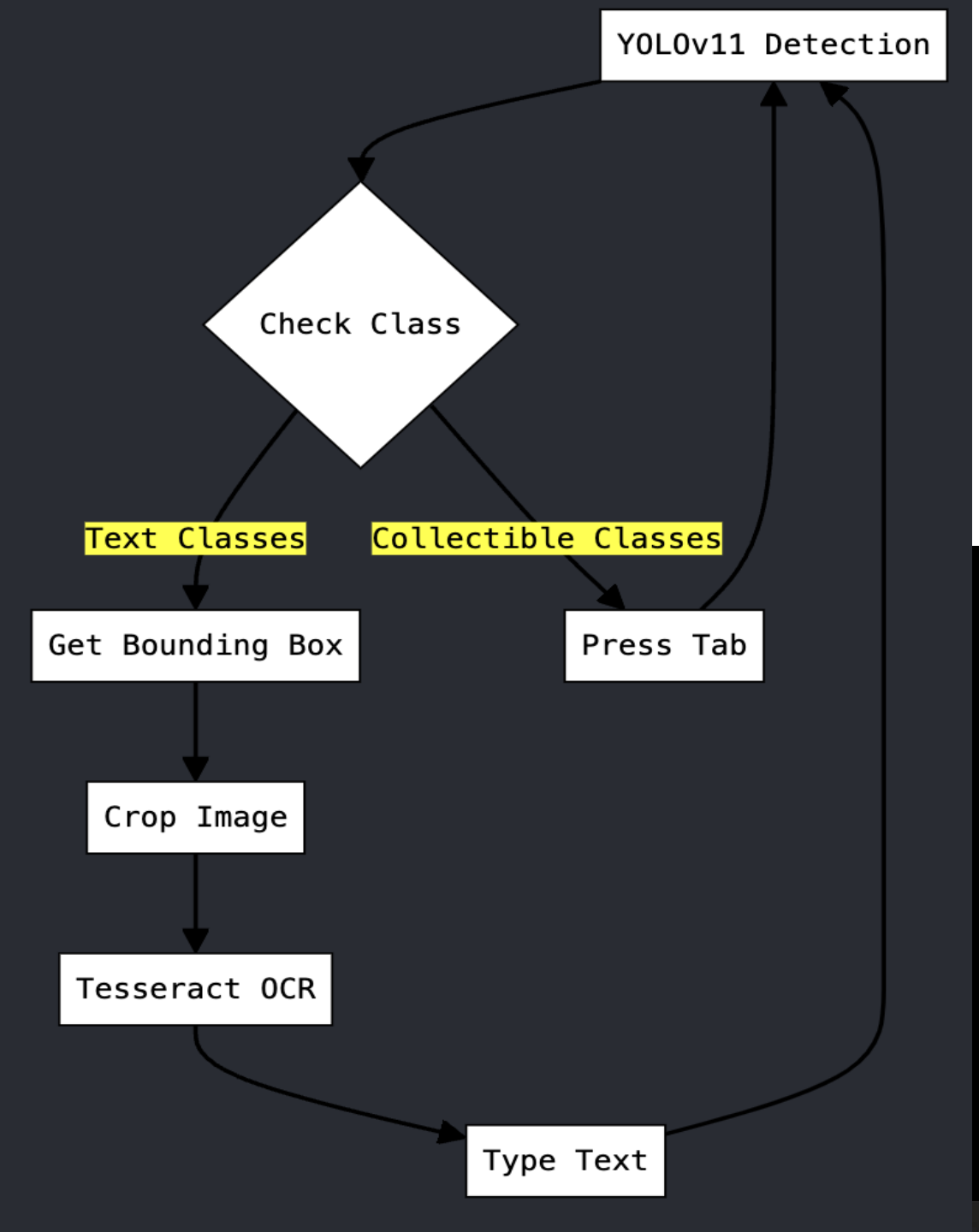
5 Methods

1) Brute Force
2) YOLOv11 + EasyOCR
3) YOLOv11 + Tesseract OCR
4) YOLOv11 + EasyOCR + SymSpell
5) YOLOv11 + EasyOCR + SymSpell + SORT + Brute Force



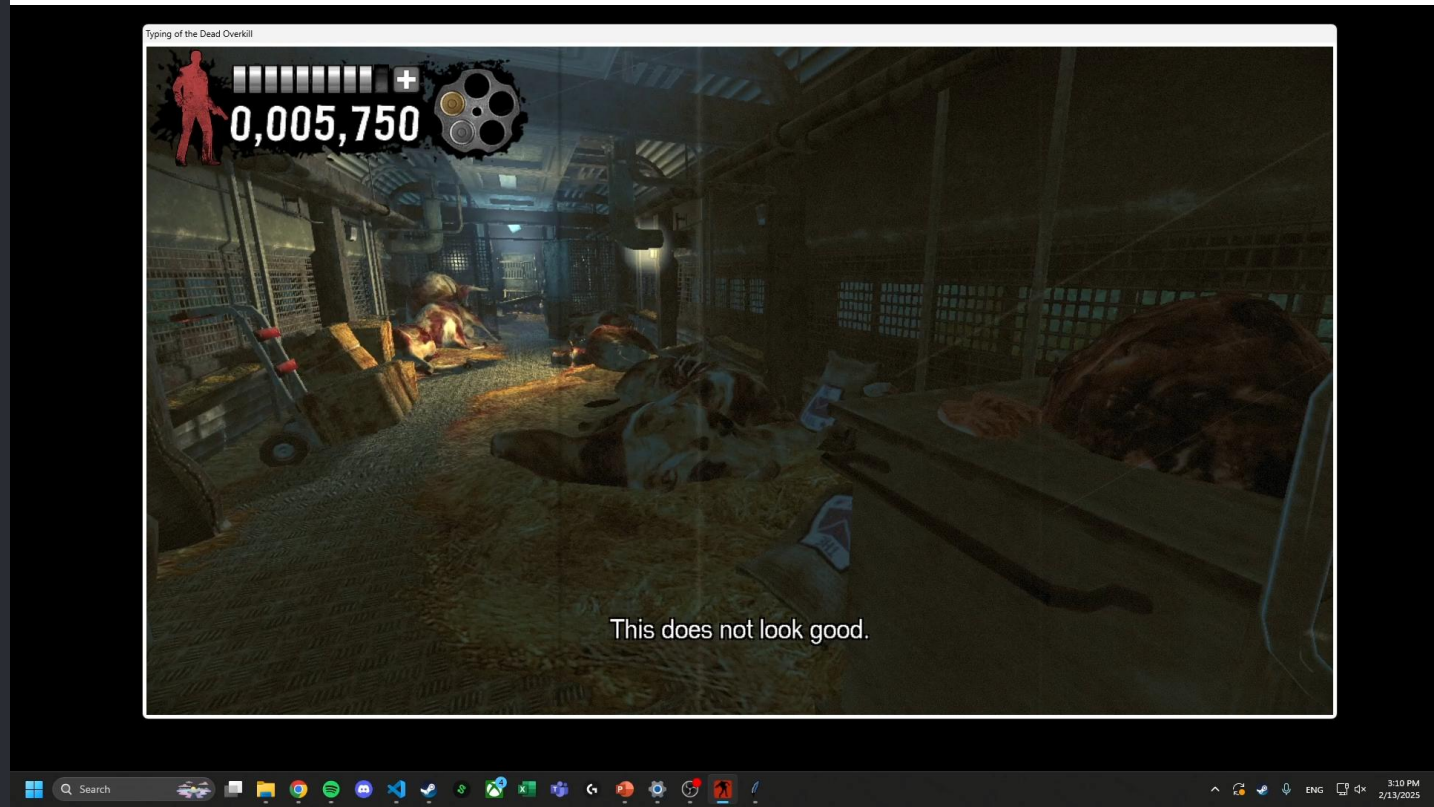
5 Methods

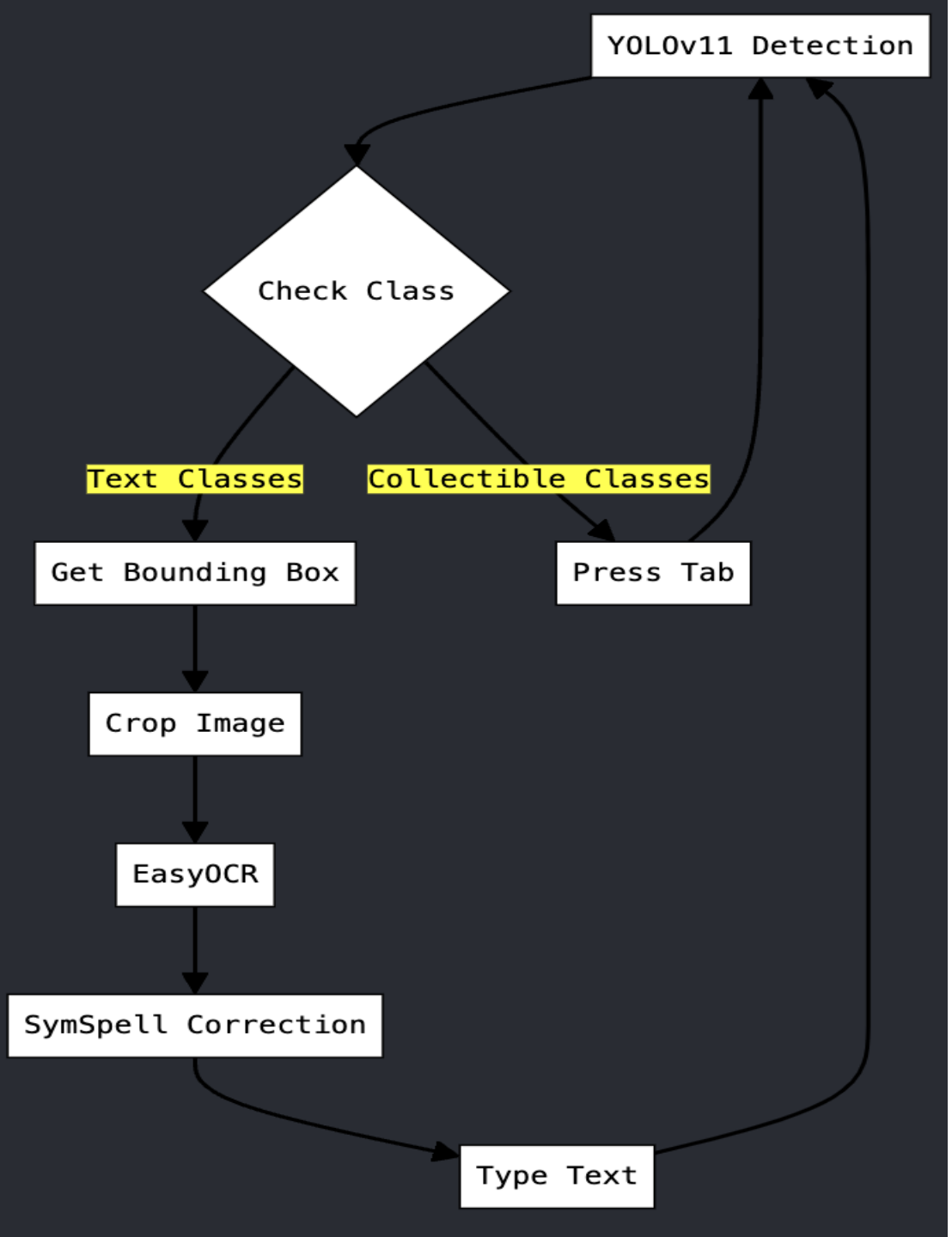
1) Brute Force
2) YOLOv11 + EasyOCR
3) YOLOv11 + Tesseract OCR
4) YOLOv11 + EasyOCR + SymSpell
5) YOLOv11 + EasyOCR + SymSpell + SORT + Brute Force



5 Methods

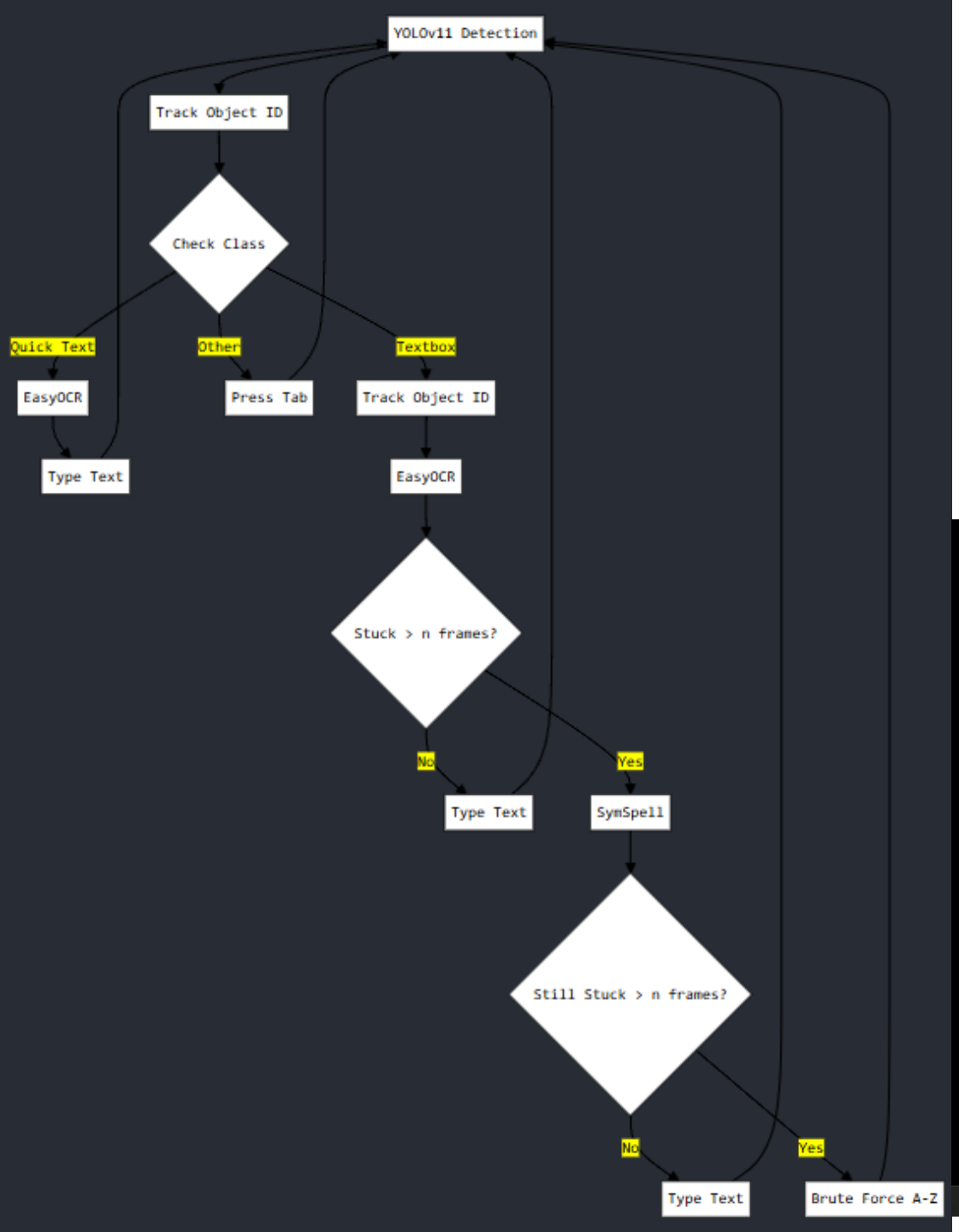
1) Brute Force
2) YOLOv11 + EasyOCR
3) YOLOv11 + Tesseract OCR
4) YOLOv11 + EasyOCR + SymSpell
5) YOLOv11 + EasyOCR + SymSpell + SORT + Brute Force





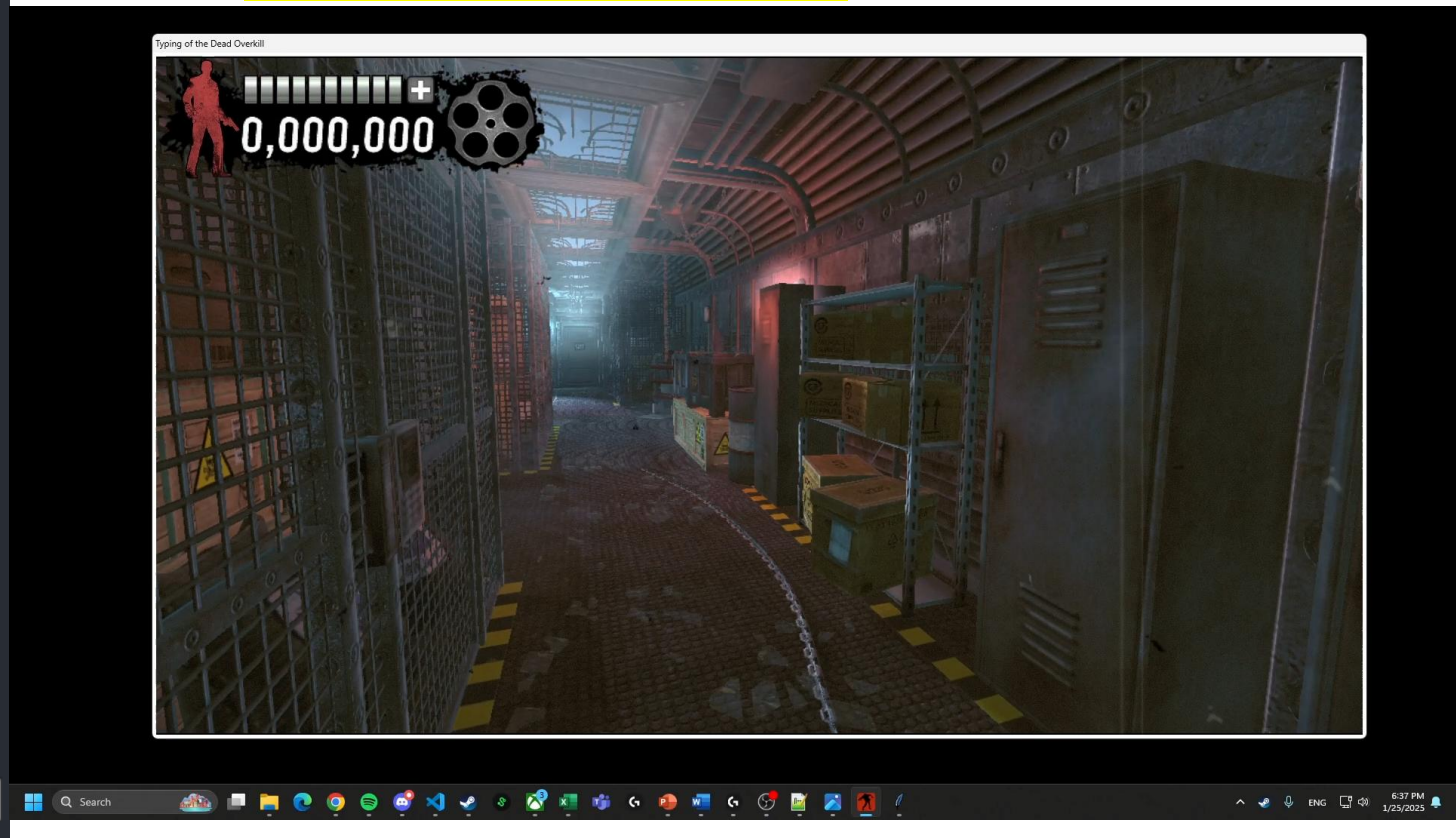
5 Methods

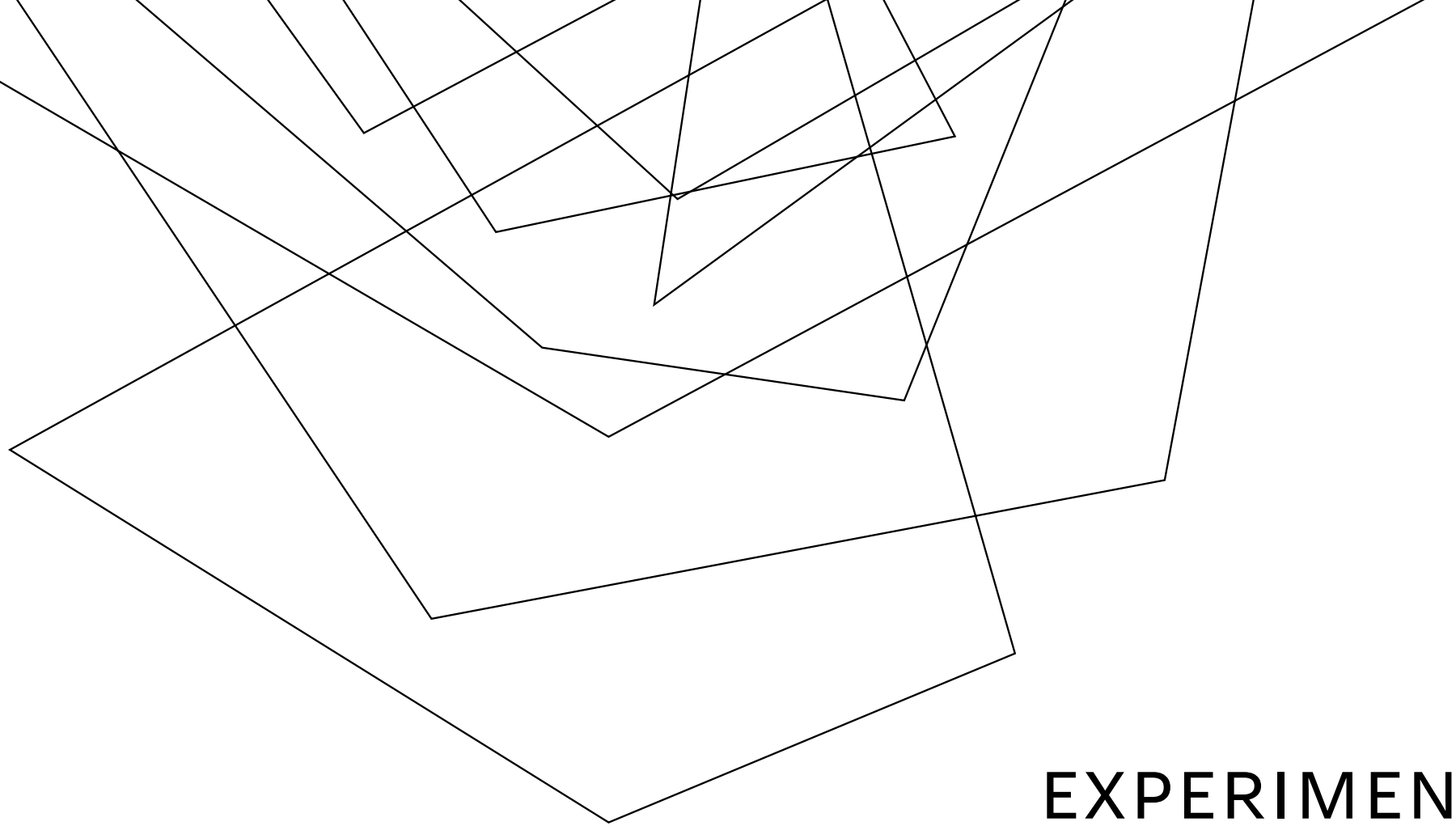
1) Brute Force
2) YOLOv11 + EasyOCR
3) YOLOv11 + Tesseract OCR
4) YOLOv11 + EasyOCR + SymSpell
5) YOLOv11 + EasyOCR + SymSpell + SORT + Brute Force



5 Methods

1) Brute Force
2) YOLOv11 + EasyOCR
3) YOLOv11 + Tesseract OCR
4) YOLOv11 + EasyOCR + SymSpell
5) YOLOv11 + EasyOCR + SymSpell + SORT + Brute Force





EXPERIMENTS & RESULTS

EXPERIMENTS & RESULTS

Experiment Environment

Maps

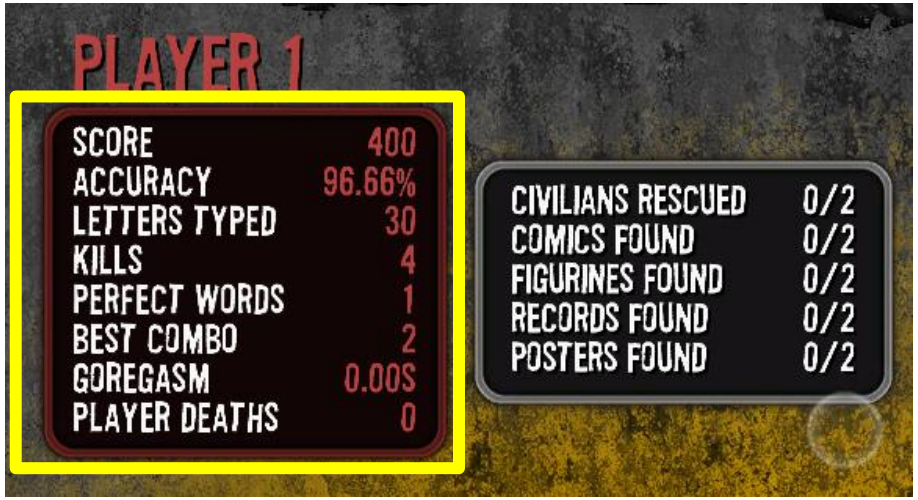
- Tested on all 5 stages at the hard difficulty level, with two attempts per map.

Hardware

GPU : NVIDIA RTX 4060TI
CPU : Intel(R) Core(TM) i7-14700F 2.10 GHz
RAM : 32 GB

EXPERIMENTS & RESULTS

In game Evaluation



Method	Brute Force	Yolo11+ EasyOCR	Yolo11+ EasyOCR+SymSpell +SortTrack+Brute Force	Yolo11+EasyOCR+SymSpell	Yolo11+Tesseract
Score Avg (points)	60,543.30	91,919.40	**102,806.90**	82,600.60	43,468.10
Acc Avg (%)	4.36	**84.53**	84.39	77.82	76.34
Letters Typed Avg (count)	**45,930.90**	2,640.20	2,676.30	2,526.80	2,943.10
Player Deaths Avg (count)	0.00	0.00	0.00	0.40	**2.00**
time Avg (seconds)	12.16	9.99	11.08	10.31	**12.28**

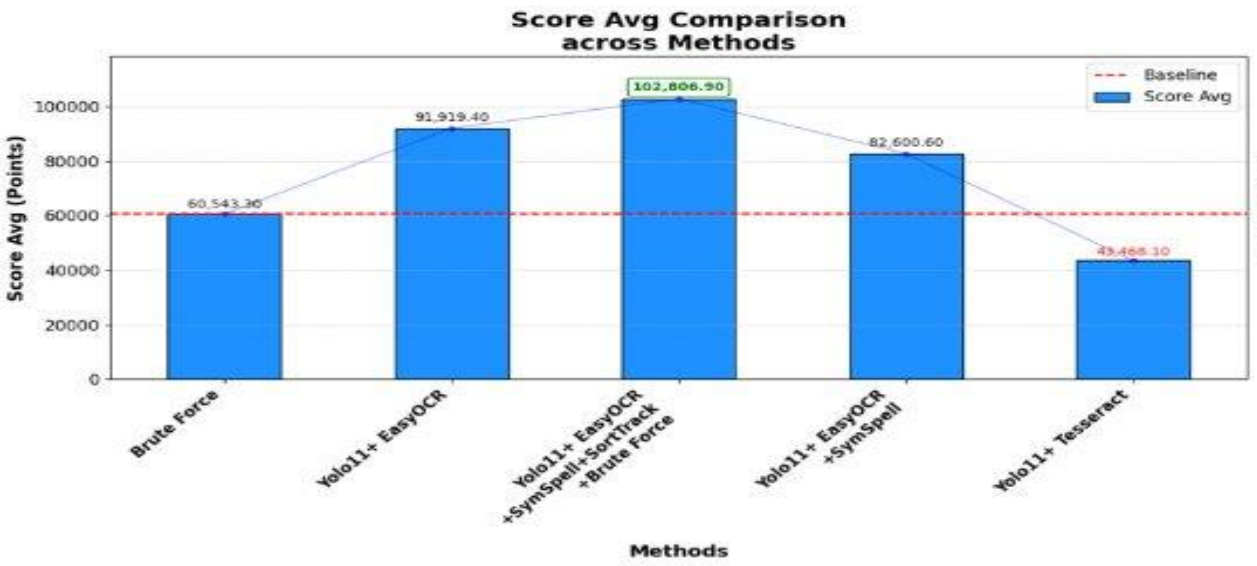
Good

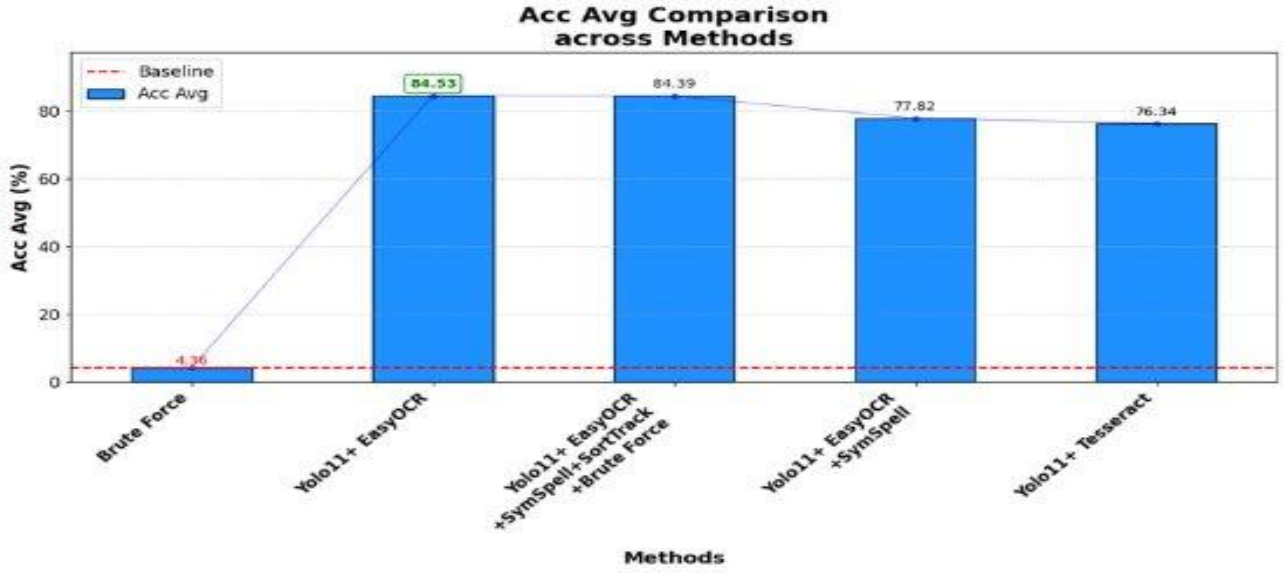
Bad

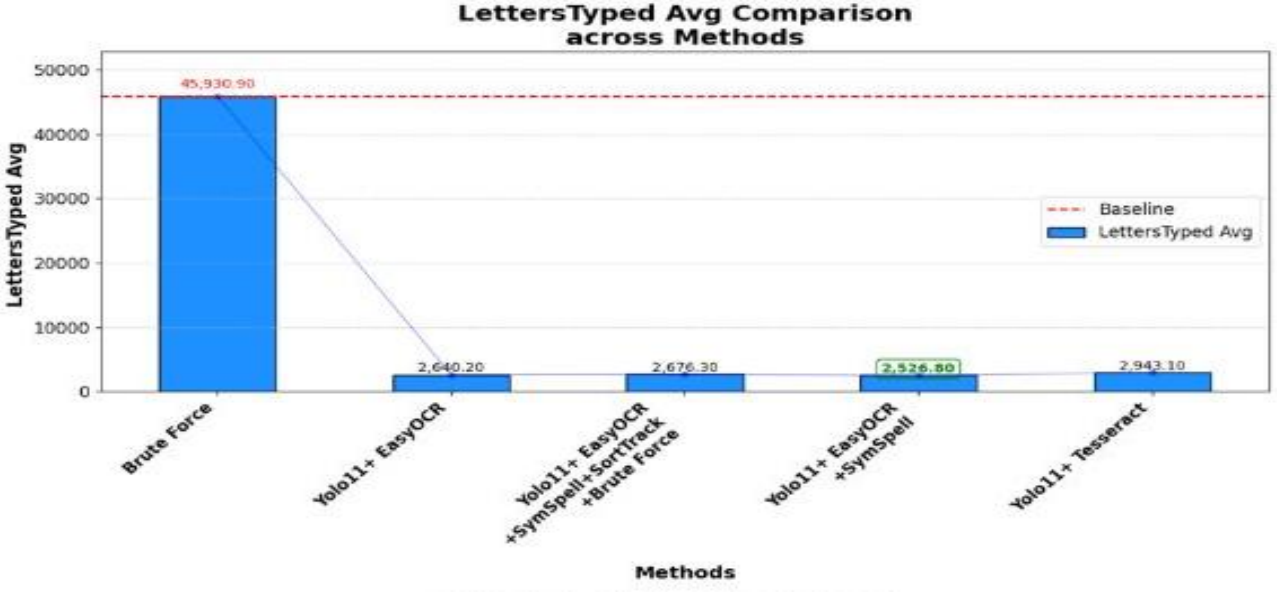
$$Accuracy\ (Acc\ Avg) = \frac{Number\ of\ Correctly\ Typed\ Characters}{Total\ Number\ of\ Characters\ Typed} \times 100$$

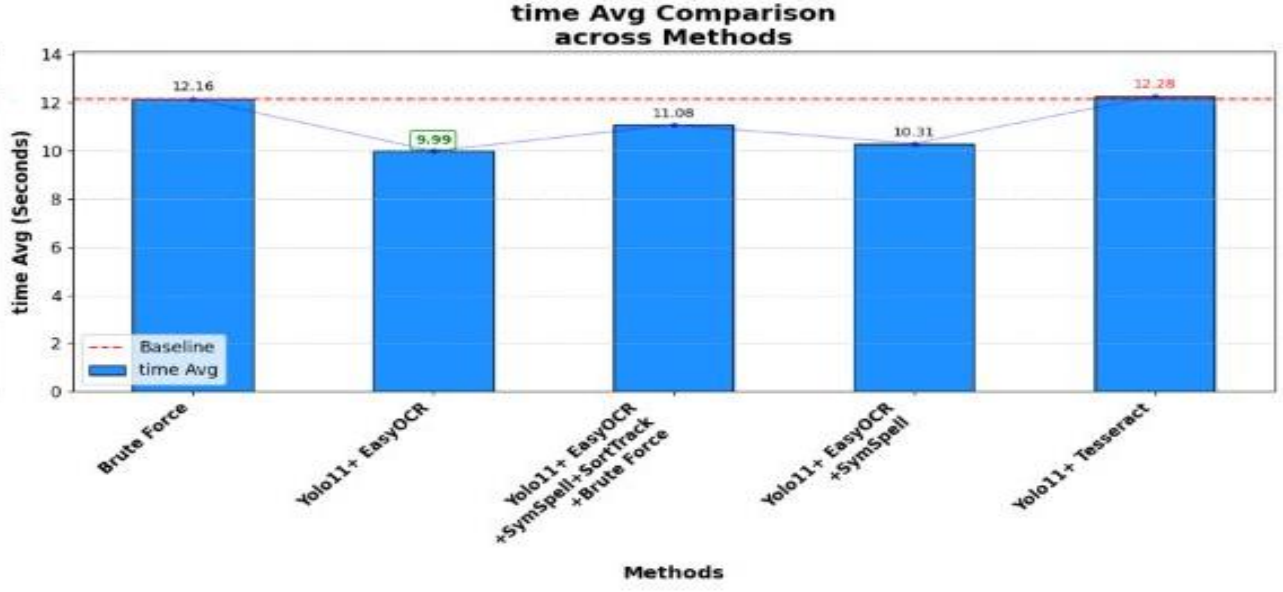
EXPERIMENTS & RESULTS

In game
Evaluation
for main objective



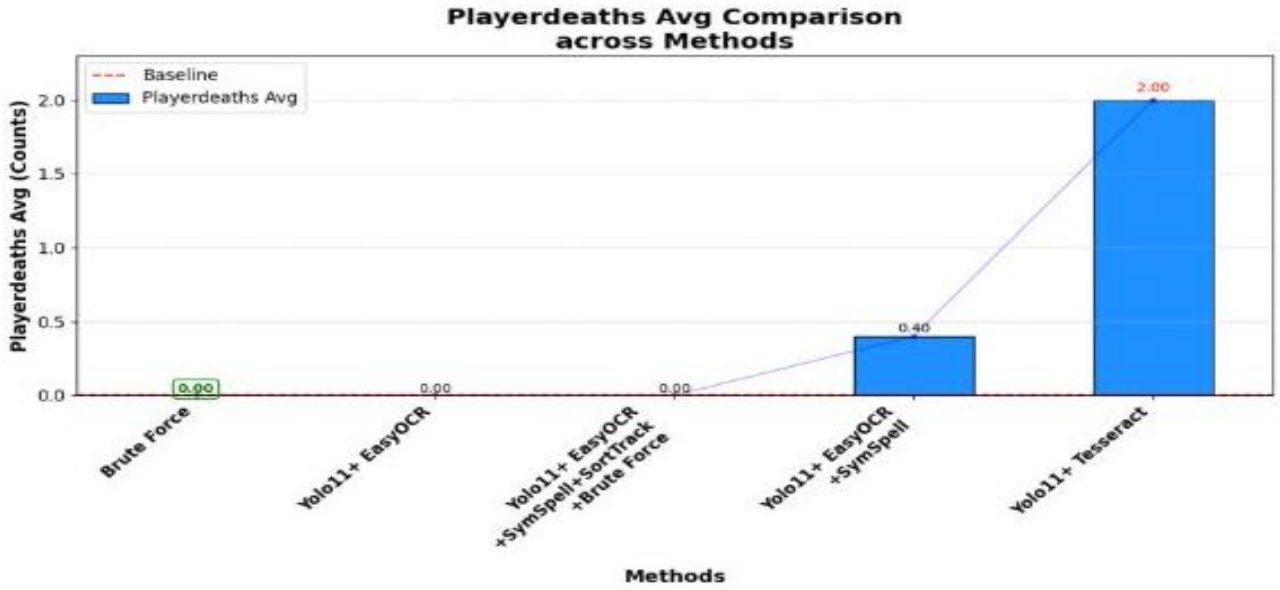






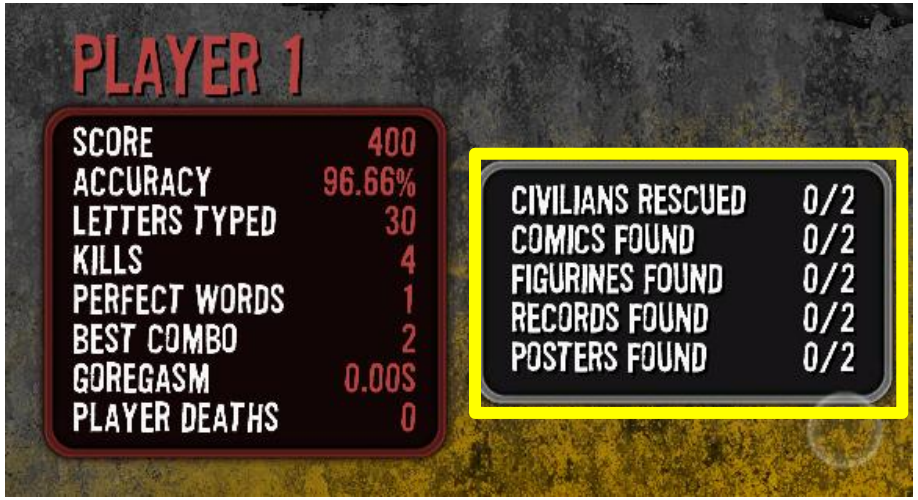
EXPERIMENTS & RESULTS

In game
Evaluation
for main objective



EXPERIMENTS & RESULTS

In game
Evaluation
for main objective



Map	Civilians Rescued	Comics Found	Figures Found	Records Found	Posters Found
4	5/5	2/2	2/2	1/1	4/4
5	2/2	2/2	2/2	2/2	2/2
6	2/2	3/3	3/3	1/1	2/2
7	4/4	2/2	2/2	1/1	3/3
8	3/3	0/0	2/2	1/1	2/2

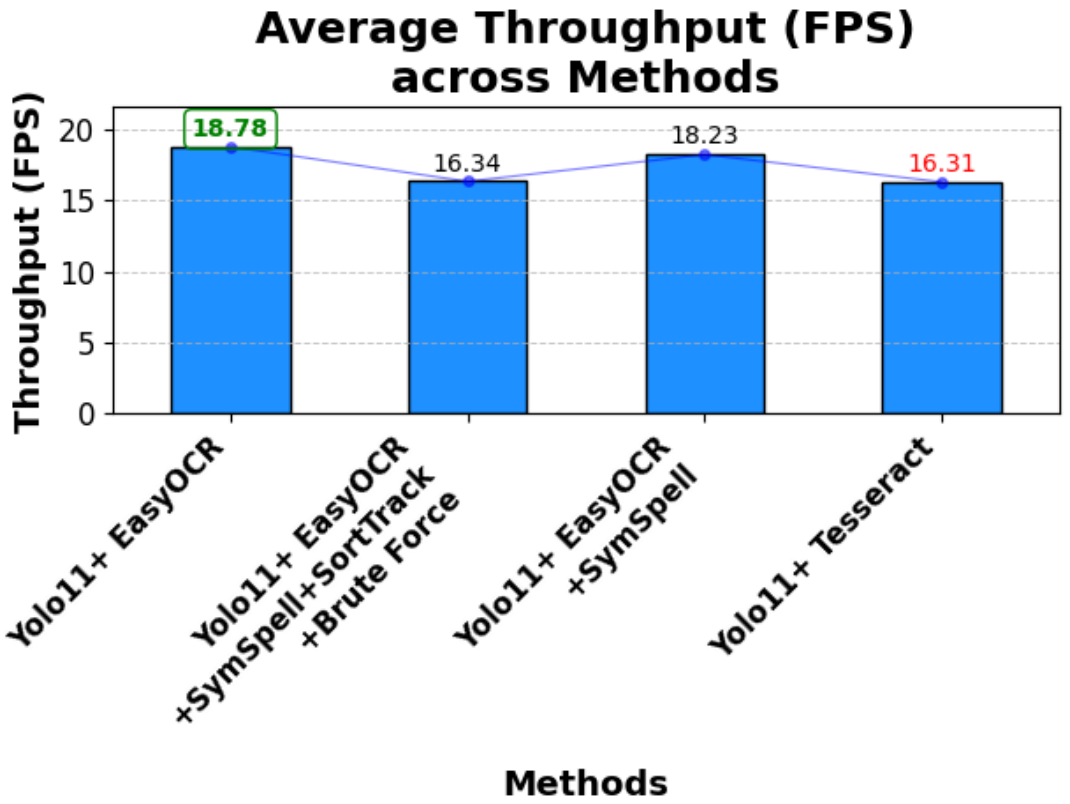
EXPERIMENTS & RESULTS

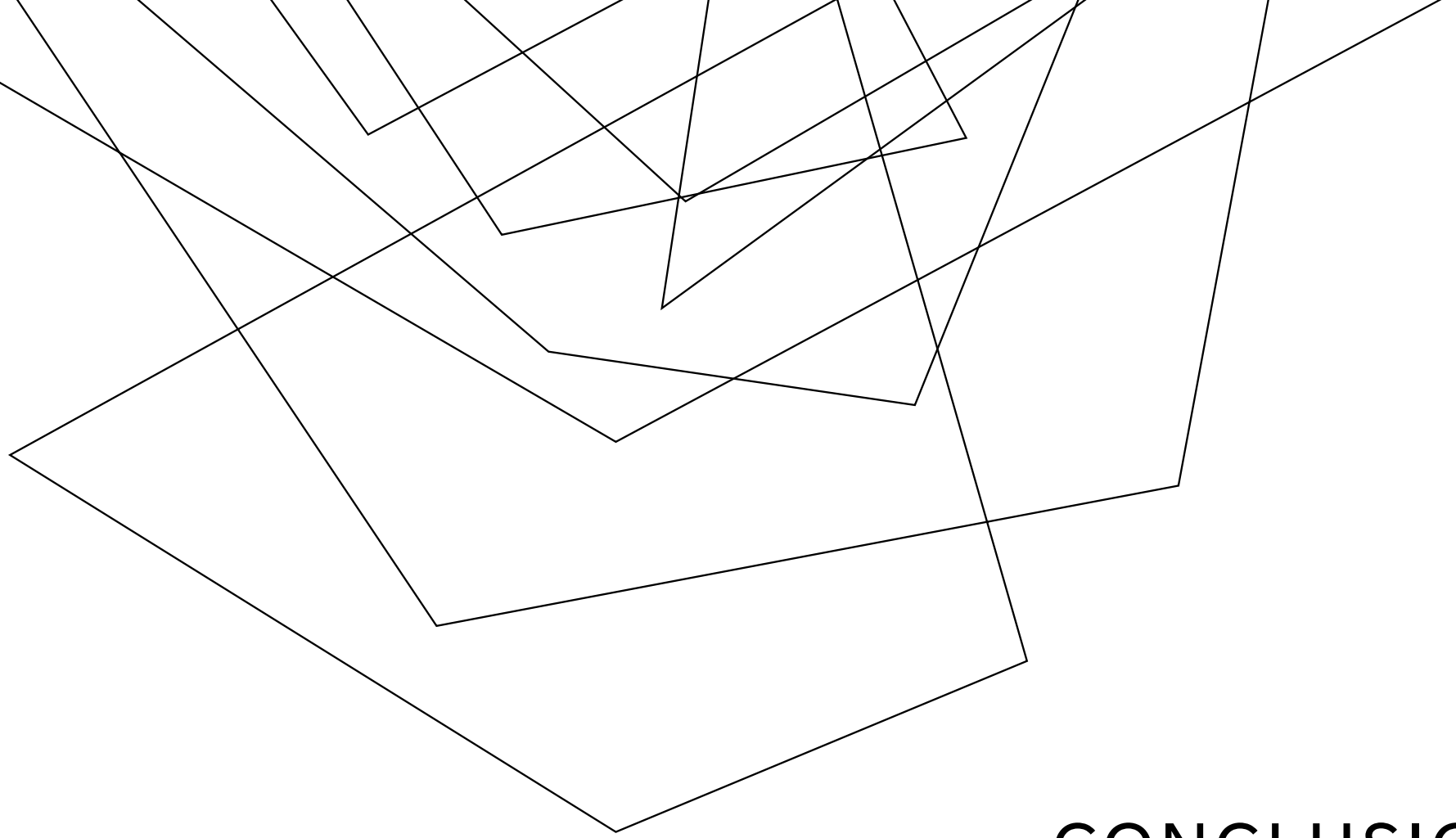
Algorithms Performance Analysis

Method	Level	Throughput (FPS)
Yolo11+ EasyOCR	Map4	19.1
Yolo11+Tesseract	Map4	13.68
Yolo11+EasyOCR+SymSpell	Map4	18.05
Yolo11+ EasyOCR+SymSpell+SortTrack+Brute Force	Map4	16.16
Yolo11+ EasyOCR	Map5	19.27
Yolo11+Tesseract	Map5	16.7
Yolo11+EasyOCR+SymSpell	Map5	18.83
Yolo11+ EasyOCR+SymSpell+SortTrack+Brute Force	Map5	18.38
Yolo11+ EasyOCR	Map6	18.6
Yolo11+Tesseract	Map6	16.97
Yolo11+EasyOCR+SymSpell	Map6	18.18
Yolo11+ EasyOCR+SymSpell+SortTrack+Brute Force	Map6	16.7
Yolo11+ EasyOCR	Map7	19.72
Yolo11+Tesseract	Map7	17.45
Yolo11+EasyOCR+SymSpell	Map7	19.17
Yolo11+ EasyOCR+SymSpell+SortTrack+Brute Force	Map7	15.26
Yolo11+ EasyOCR	Map8	17.2
Yolo11+Tesseract	Map8	16.73
Yolo11+EasyOCR+SymSpell	Map8	16.9
Yolo11+ EasyOCR+SymSpell+SortTrack+Brute Force	Map8	15.22

$$FPS = \frac{1}{Latency\ (second\ per\ frame)}$$

Method	Throughput (FPS)
Yolo11+ EasyOCR	18.778
Yolo11+EasyOCR+SymSpell	18.226
Yolo11+ EasyOCR+SymSpell+SortTrack+Brute Force	16.344
Yolo11+Tesseract	16.306





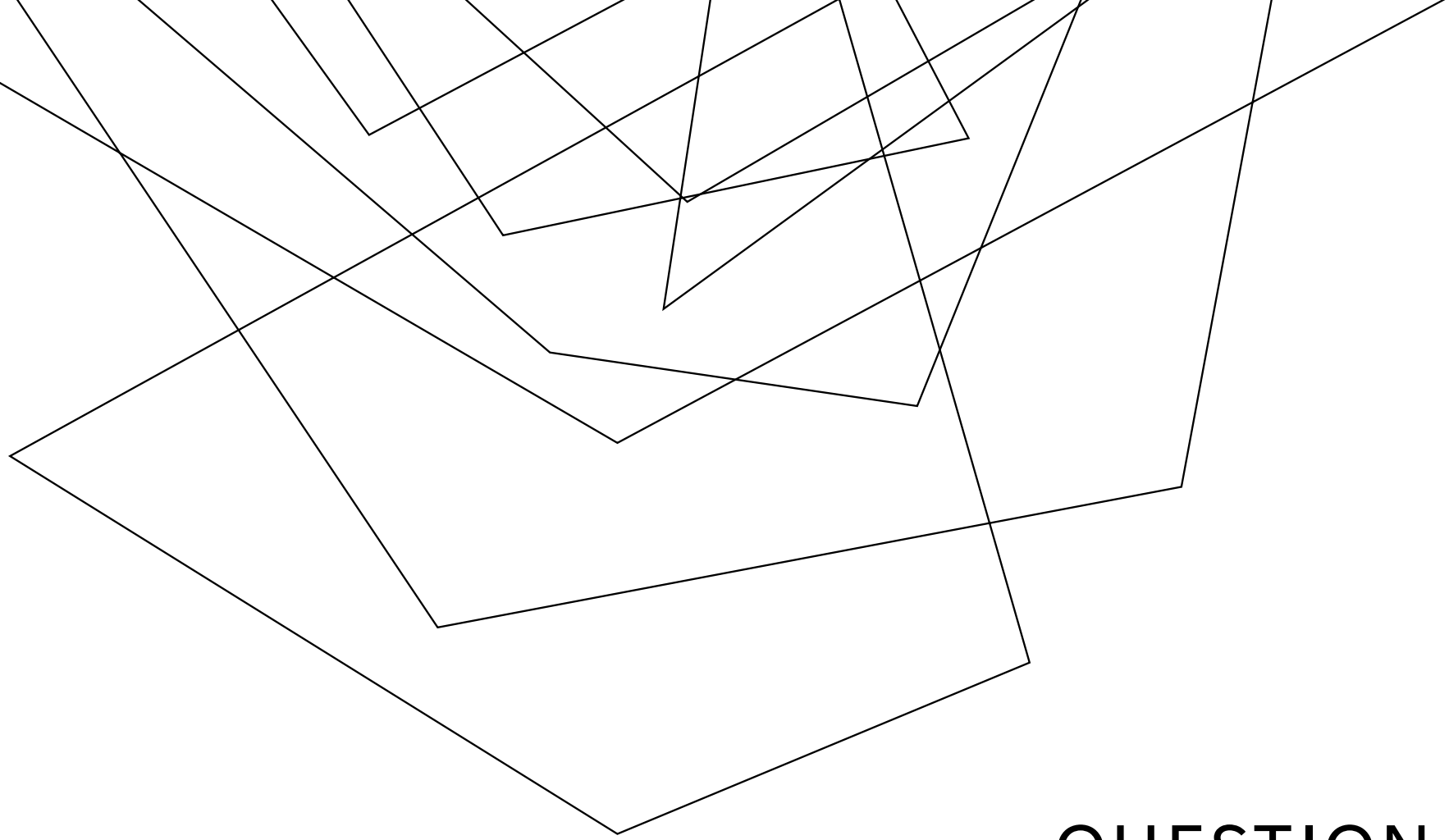
CONCLUSION

CONCLUSION

Method	Brute Force	Yolo11+ EasyOCR	Yolo11+ EasyOCR+SymSpell +SortTrack+Brute Force	Yolo11+EasyOCR+SymSpell	Yolo11+Tesseract
Score Avg (points)	60,543.30	91,919.40	**102,806.90**	82,600.60	43,468.10
Acc Avg (%)	4.36	**84.53**	84.39	77.82	76.34
Letters Typed Avg (count)	**45,930.90**	2,640.20	2,676.30	2,526.80	2,943.10
Player Deaths Avg (count)	0.00	0.00	0.00	0.40	**2.00**
time Avg (seconds)	12.16	9.99	11.08	10.31	**12.28**

Method	Throughput (FPS)
Yolo11+ EasyOCR	18.778
Yolo11+EasyOCR+SymSpell	18.226
Yolo11+ EasyOCR+SymSpell+SortTrack+Brute Force	16.344
Yolo11+Tesseract	16.306

Method	Pros ✅	Cons ❌
Brute Force	- Guaranteed to pass all levels without dying. - Completes all objectives.	- Lowest accuracy. - Types the highest number of characters. - Slow time spend per map.
YOLOv11 + EasyOCR	- Provides good overall evaluation results. - Flexible in recognizing uncommon words. - Most FPS	- Lower score compared to enhanced versions.
YOLOv11 + EasyOCR + SymSpell + SORT + Brute Force	- Best overall performance. - Achieves the highest game score (10k higher than YOLOv11 + EasyOCR).	- Player deaths due to non-dynamic words ocr. - Slow time spend per map. - Low FPS
YOLOv11 + EasyOCR + SymSpell	- Smallest number of letter typed.	- Does not provide significantly better overall results compared to previous methods. - Player deaths due to non-dynamic words ocr.
YOLOv11 + Tesseract OCR	- Higher accuracy than Brute Force.	- Results in the most player deaths, leading to lower scores. - Takes the longest time per level. - Types an excessive number of characters. - Low FPS



QUESTION

THANK YOU

Naris Chanpaboonkit

6520422025@stu.nida.ac.thwww