Boxing Punch Detection and Analysis Using YOLOv8 and MediaPipe

1. Introduction

The objective of this project was to develop a deep learning-based system capable of detecting and classifying boxing punches using YOLOv8 while integrating 2D pose estimation via MediaPipe to determine punch landing.

2. Dataset Creation

2.1 Data Collection

- Videos were sourced from YouTube and converted into individual frames.
- Specific frames where a boxer was executing a punch were manually selected for annotation.

2.2 Annotation

- Used **LabelImg** to annotate frames.
- Defined four classes: **"Jab", "Hook", "Uppercut", "Punching bag"**.
- The **"Cross"** class was omitted due to its similarity to "Jab" in static images.
- **Training Set:** 198 images (after augmentation)
- **Test Set:** 72 images
- Applied **data augmentation** techniques (rotations, flips, color adjustments) to enhance dataset diversity.

3. Model Implementation

3.1 YOLOv8 Training

- Due to computational constraints and usability challenges with YOLOv7, **YOLOv8** was used instead.
- Training was conducted on **Google Colab**.
- Results:
    - **Epochs:** 100
    - **Training Time:** 0.042 hours
    - **Best Model Size:** 6.2MB
    - **Validation Metrics:**
        - **Precision:** 0.986
        - **Recall:** 1.0
        - **mAP50:** 0.995

        - **mAP50-95:** 0.984

3.2 Punch Detection & Pose Estimation

- YOLOv8 was used to detect punches and the punching bag.
- MediaPipe was used for **2D pose estimation** to track the boxer's key joints.

4. Punch Landing Detection

4.1 Methodology

- Extracted **wrist key points** from MediaPipe.
- Defined a **bounding box region** around each wrist.
- If a wrist bounding box overlapped with the **punching bag's bounding box**, the punch was classified as landing successfully.

4.2 Evaluation

- **Punch Contact Detection:** Implemented a bounding box intersection function to determine contact.
- **Frame Output Format:**

```
{
   "frame": {
      "Boxer": {
         "Punch": "Jab",
         "Landing": "Punching bag"
      },
      "Punching bag": {
         "Action": "Standing",
         "Landing": "NAN"
      }
   }
}
```

5. Evaluation Metrics

- **Model Performance:** Evaluated using **precision, recall, and F1-score**.
- **Pose Estimation Accuracy:** Ensured correct keypoint tracking using MediaPipe.
- **Punch Landing Accuracy:** Estimated accuracy of punch landing detection.

6. Challenges & Future Work

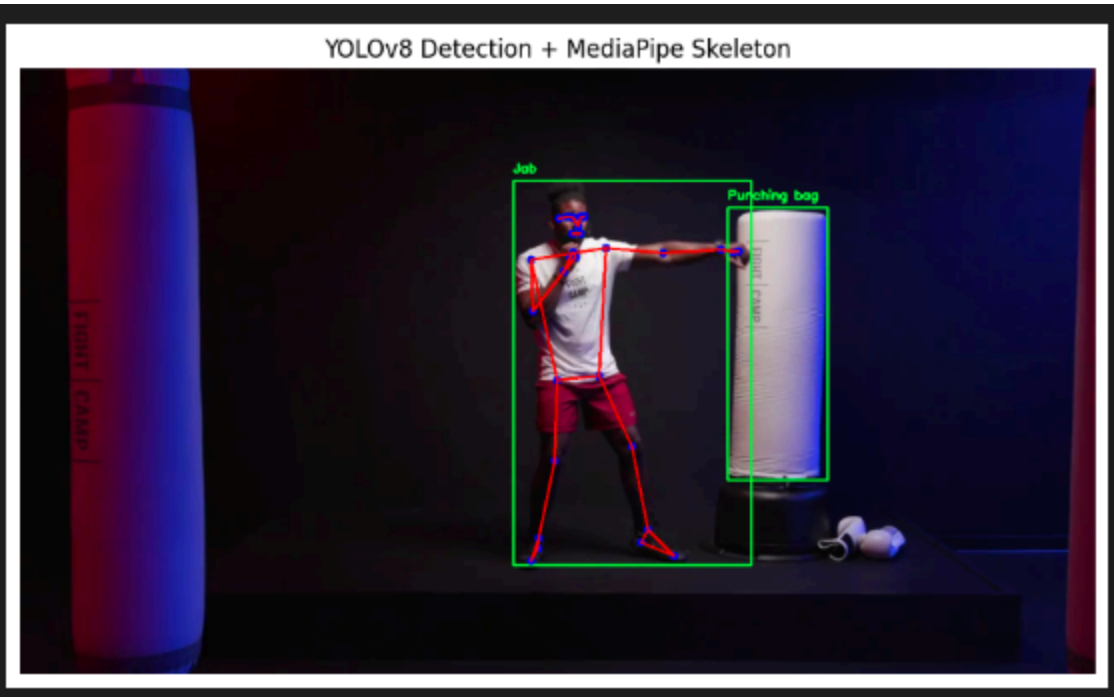
6.1 Challenges

- Computational constraints limited YOLOv7 training, leading to a switch to YOLOv8.
- Limited annotated dataset resulted in potential overfitting.
- Lack of **tracking** led to missing motion-based features.

6.2 Future Enhancements

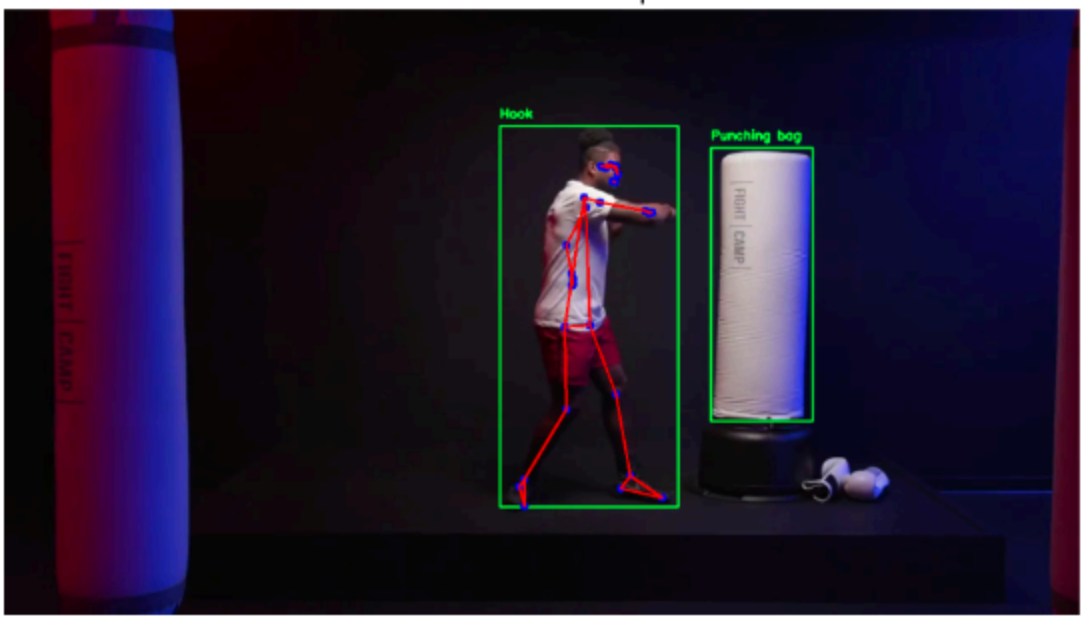
- **Expand the dataset** with more diverse angles and scenarios.
- Implement **object tracking** for smoother punch detection over video frames.
- Extend to **3D pose estimation** to better differentiate similar punch types.
- Improve **punch landing detection** using additional features (e.g., velocity tracking).

7. Results & Visuals

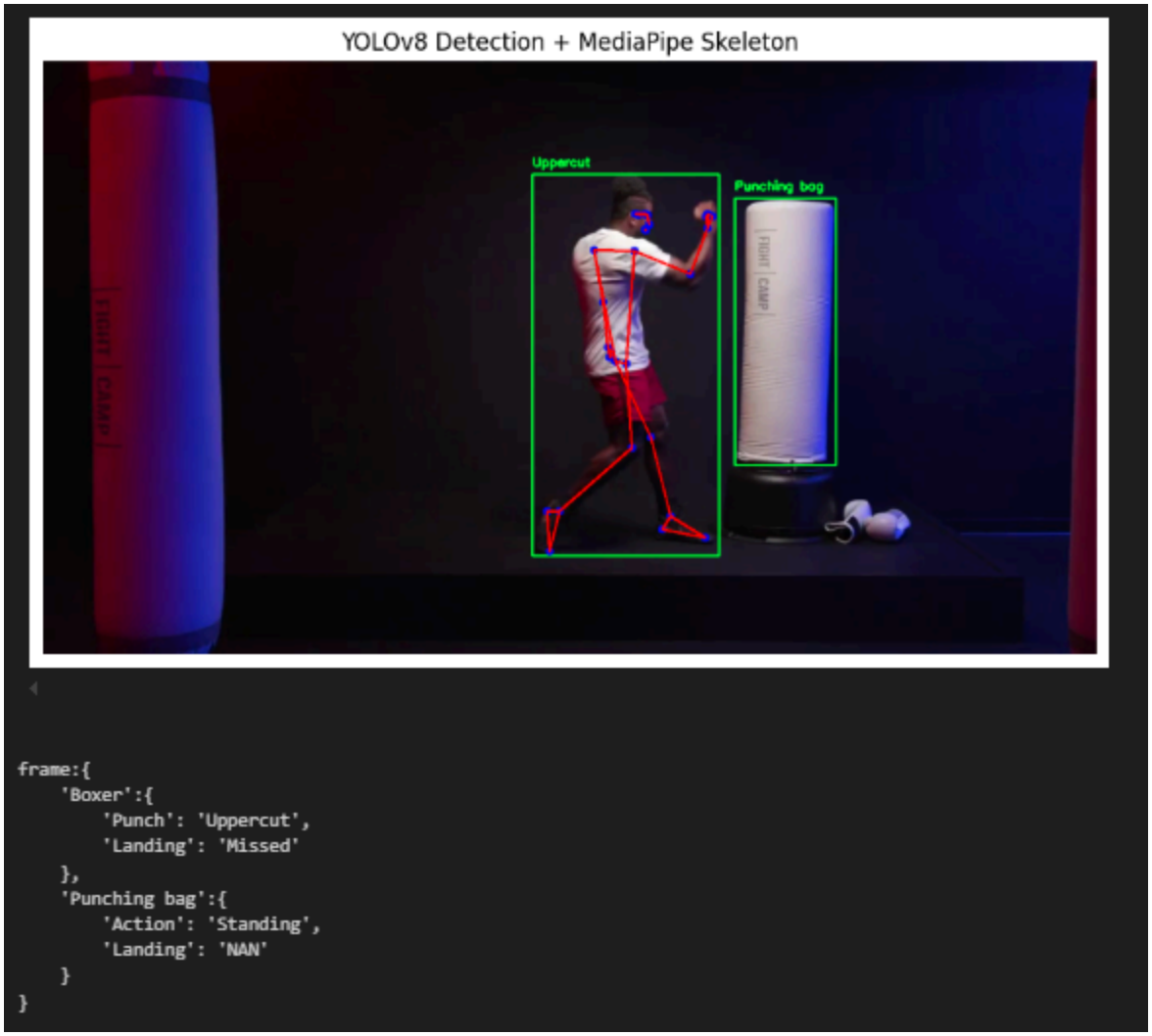



```
{
    "frame": {
        "Boxer": {
            "Punch": "Jab",
            "Landing": "Head"
        },
        "Punching bag": {
            "Action": "Standing",
            "Landing": "NAN"
        }
    }
}
```

YOLOv8 Detection + MediaPipe Skeleton




```
{
    "frame": {
        "Boxer": {
            "Punch": "Hook",
            "Landing": "Missed"
        },
        "Punching bag": {
            "Action": "Standing",
            "Landing": "NAN"
        }
    }
}
```




```
frame:{
    'Boxer':{
        'Punch': 'Uppercut',
        'Landing': 'Missed'
    },
    'Punching bag':{
        'Action': 'Standing',
        'Landing': 'NAN'
    }
}
```

8. Conclusion

This project successfully implemented YOLOv8 for punch classification and MediaPipe for 2D pose estimation, achieving high detection accuracy. Future improvements will focus on dataset expansion, tracking integration, and 3D pose estimation to enhance robustness.
