



2024 6th IKCEST Belt and Road International Big Data Competition

And the 10th Baidu & Xi'an Jiaotong University Big Data Competition

Team Name: Tang's Team

Presenter: Wang Bin

Presentation Date: November 2024



C O N T E N T



01

**Team and Competition
Retrospective**

02

**Data Analysis and
Preprocessing**

03

Solution Introduction

04

summary and Reflection



团队 & 赛题回顾

My Team & Review of the Competition Topic.

Team Introduction



Wang Bin



Supervisor

Tianjin Light Industry Vocational and Technical College - Instructor in Artificial Intelligence Technology and Applications

Previously served as a Senior AI Algorithm Researcher at Tencent, specializing in OCR (Optical Character Recognition).

Tang Guopeng



Team Captain

Student at Tianjin Light Industry Vocational and Technical College - Major in Artificial Intelligence Technology and Applications

Achievements:
Third Prize in the 2024 Baidu Business AI Technology Innovation Competition (Image Caption Generation Track for Advertisements)
Recipient of the 2024 National Perseverance Scholarship

Wu Xiaofei



Team Member"

Student at Tianjin Light Industry Vocational and Technical College - Major in Artificial Intelligence Technology and Applications

He Yiming



Team Member"

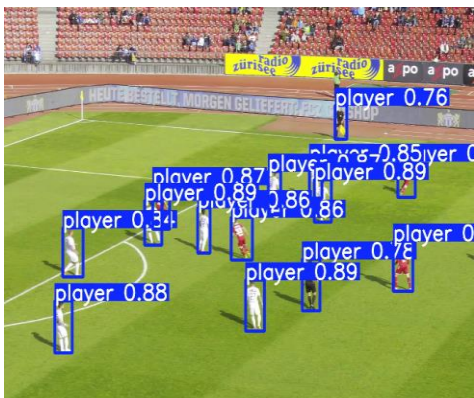
Student at Tianjin Light Industry Vocational and Technical College - Major in Artificial Intelligence Technology and Applications

赛题回顾

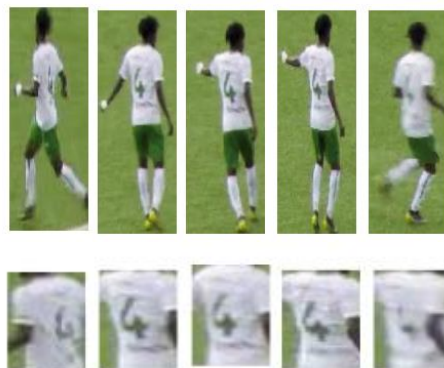
Review of the Competition Topic



1. Object Detection and Tracking



2. Jersey Number Recognition



3. Agent Commentary



1.Target detection and tracking

Check the players in the sports videos and track them

2.Jersey number identification

The shirt number of the player selected in the box is identified to identify the player

3.The agent explanation

The Baidu Wenxin agent platform develops an agent that can receive a video link to a football game as input and output the corresponding commentary



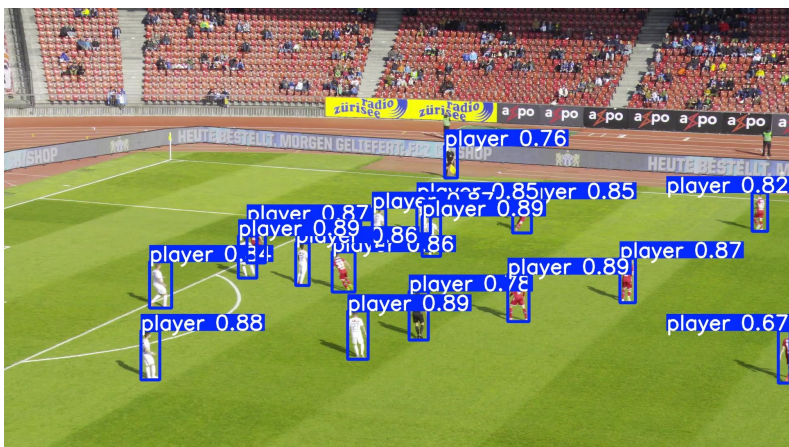
2

数 据 分 析 与 处 理

Data Analysis and Processing

数据处理—初赛阶段

Data Processing



Training Set and Validation Set:

A total of 57 videos were provided, and no division between a training set and a validation set had been performed.

- 1.Division: Use 42 for the training set and 15 for the validation set.
- 2.There are annotation errors in the data; we conducted a sample cleaning of 5%.
- 3.We used frame extraction from the videos for training purposes instead of using all available data. Considering the potential for error amplification, this approach could have a significant impact on the final results.
- 4.During training, we experimented with different resolutions to find a balance between efficiency and performance.

Ref: GLIP (Grounded Language – Image Pretraining)

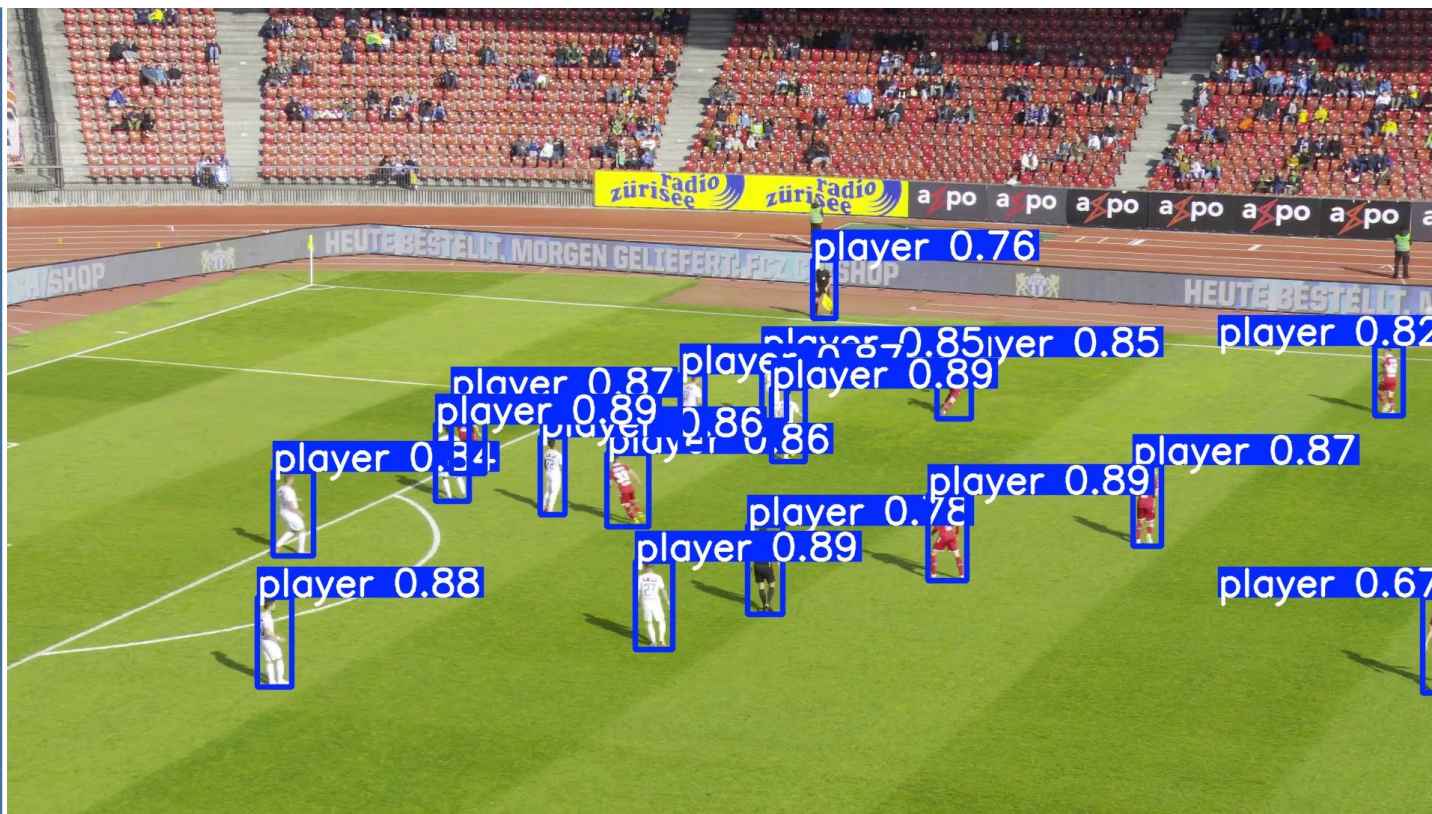
数据处理

Data Processing



Teacher-Student Training

1. Train a Teacher model on the manually annotated data.
2. Download additional soccer match videos and use the trained Teacher model to make predictions, generating pseudo labels.
3. Continue training using the combined dataset (original annotated data plus the new pseudo-labeled data) to obtain an updated model.
4. Repeat steps 1-3 in cycles to improve the metrics.



✶ Even if the teacher provides low-confidence predictions, they will be treated as ground truth in the generated dataset used by the student.

Ref: GLIP (Grounded Language - Image Pretraining)



方 案 介 绍

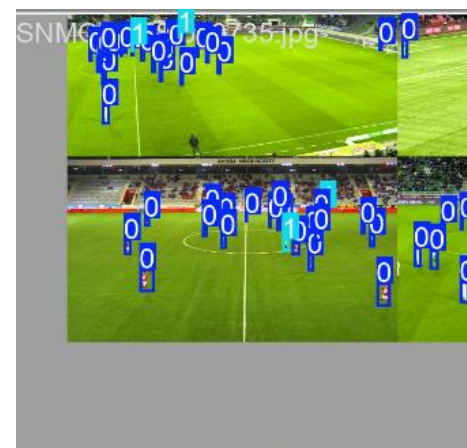
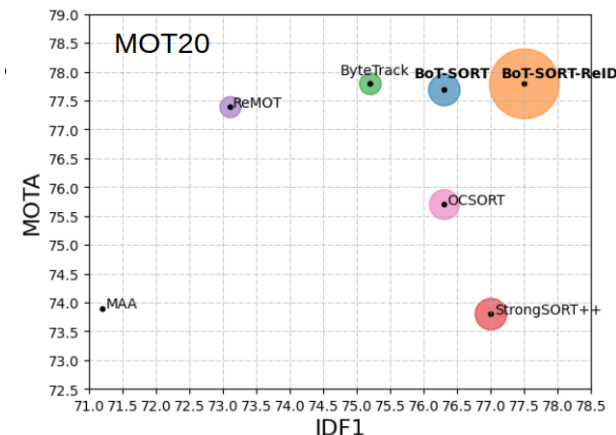
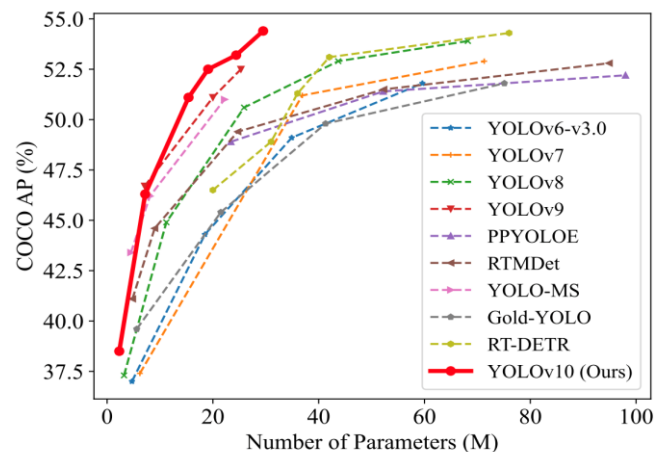
Detailed Introduction of the Solution

Detection and Tracking

jersey number detection

Model Selection and Configuration

- 1.YOLOv10-X: The super-large version achieves the highest accuracy and performance.
- 2.Input image size: 1024x1024
- 3.Degrees: Image rotation angle (± 10 degrees), while turning off flip, perspective, and other parameters to enhance effectiveness.



Ref: 检测 1. YOLOv10: Real-Time End-to-End Object Detection
追踪 BoT-SORT: Robust Associations Multi-Pedestrian Tracking

检测训练策略

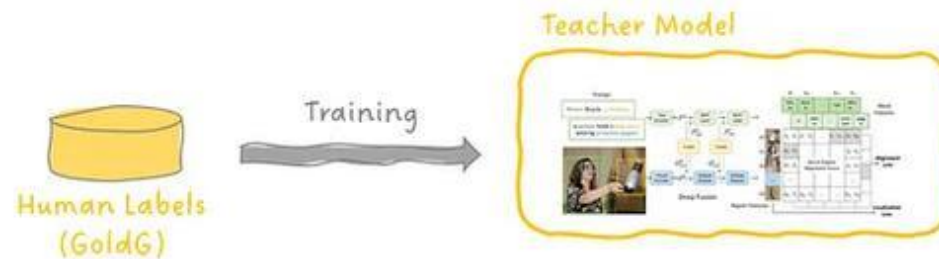
Training Strategy



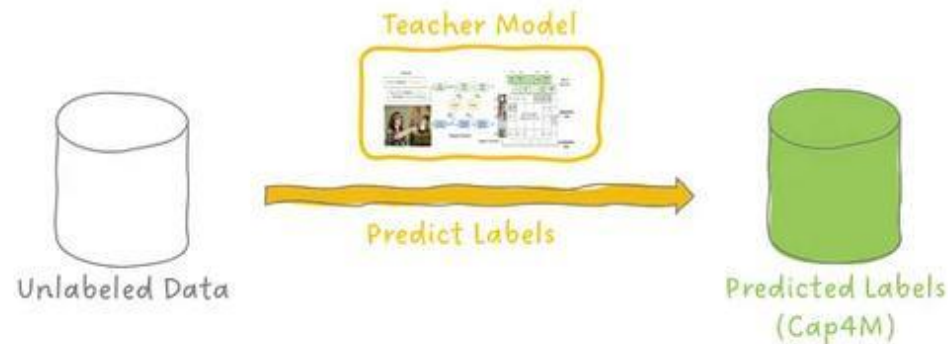
Teacher-Student Training

1. Train a Teacher model on manually annotated data.
2. Download additional soccer match videos and use the trained Teacher model to generate predictions, creating pseudo labels.
3. Continue training using the combined dataset (the original annotated data plus the newly generated pseudo-labeled data) to obtain an updated model.
4. Repeat steps 1-3 in cycles to improve the metrics.

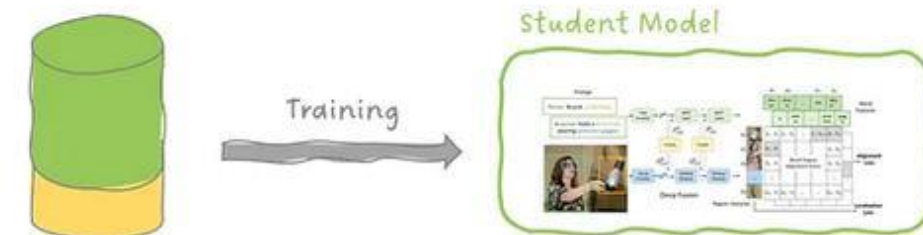
Phase 1:



Phase 2:



Phase 3:

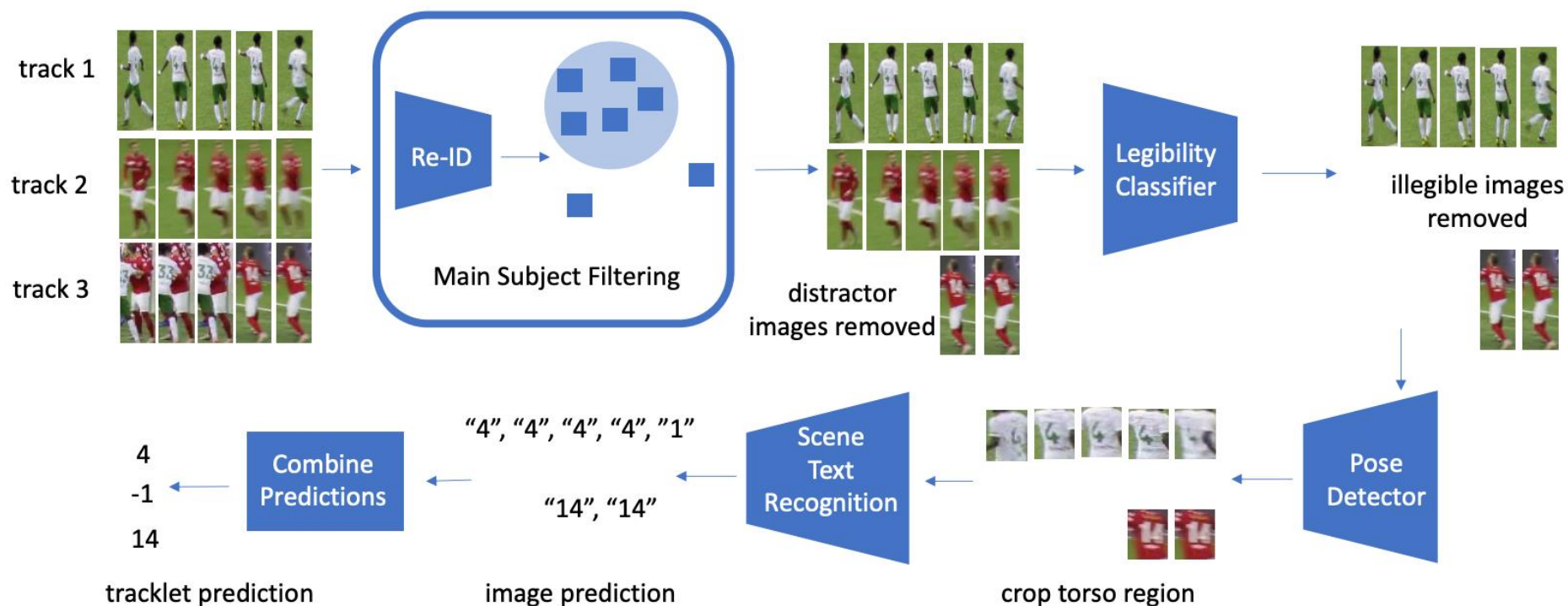


*Even if the teacher provides low-confidence predictions, they will be treated as ground truth in the generated dataset used by the student.

Ref: GLIP (Grounded Language - Image Pretraining)

球衣号码识别

jersey number detection and recognition

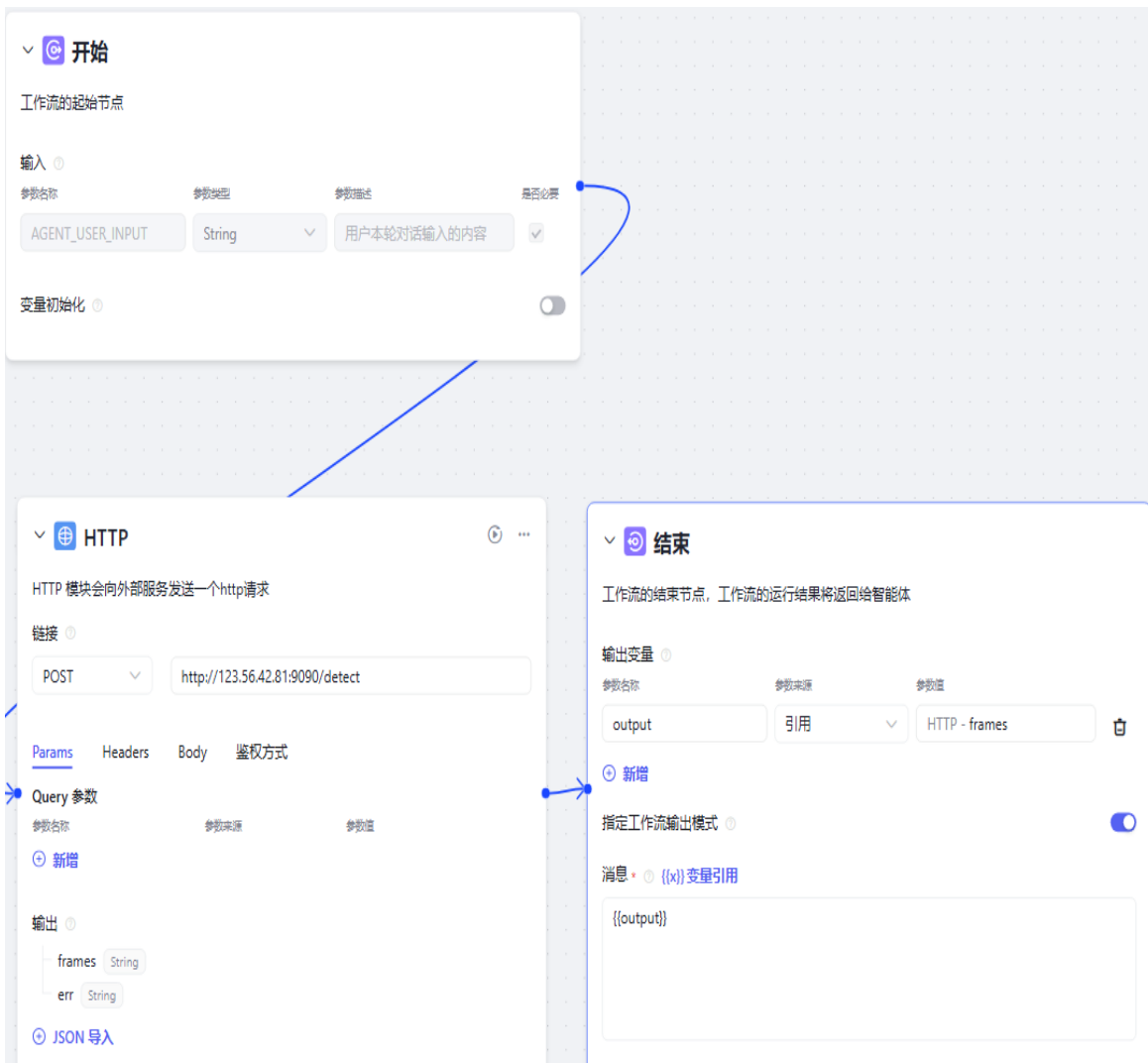
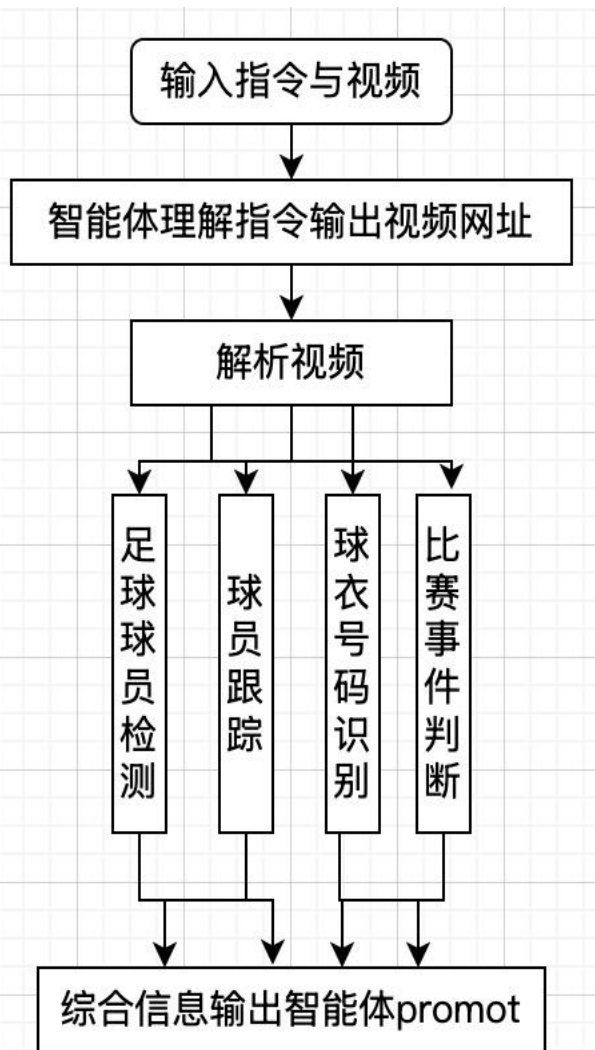


*Pipeline for Jersey Number Recognition

Ref: [A General Framework for Jersey Number Recognition in Sports Video](#)

智能体部分

Attempts in the Agent Section



Tang足球赛事解... 智能体

昨天 15:48

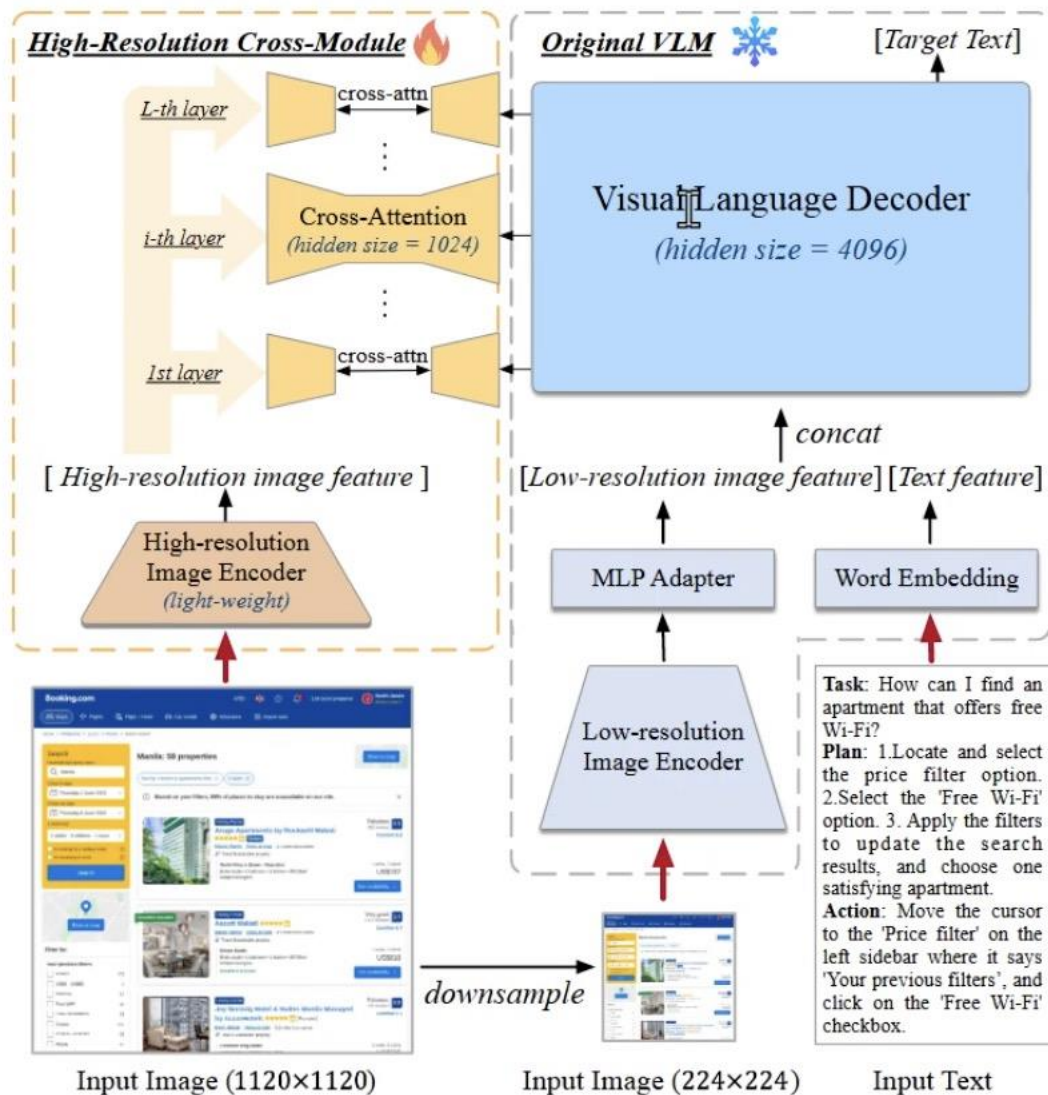
<https://www.ikcest.org/bigdata2024/spzs/content/240919990358354644.html>

这是一场紧张刺激的足球比赛, 让我们跟随镜头一起捕捉场上的精彩瞬间。0时刻, 场上球员们正在激烈拼抢。我们可以看到, 3057,1194位置的球员正在中场附近与对方争夺球权, 而1762,1523位置的球员则在前场寻找射门机会。同时, 后场的2698,1073位置和1998,1093位置的球员也保持高度警惕, 随时准备防守对方的反击。球场上快速传递, 双方球员都拼尽全力, 试图掌控比赛节奏。1时刻, 比赛进入高潮。1708,1581位置的球员成功拦截了对方的传球, 并迅速将球传给了前场的队友。此时, 2695,1220位置的球员接到球后, 带球突破对方的防线, 试图寻找射门机会。而对方球员也不甘示弱, 纷纷上前逼抢, 试图夺回球权。场上气氛紧张到了极点。2时刻, 场上形势瞬息万变。372,1460位置的球员在边路成功突破了对方的防守, 将球传给了中路的队友。中路球员接球后, 迅速调整姿势, 准备射门。然而, 对方门将及时出击, 成功地将球挡出。尽管这次进攻未能得分, 但球员们的士气并未受到影响, 他们继续积极进攻, 寻找下一个机会。3时刻, 比赛进入白热化阶段。394,1229位置的球员在前场接到了队友的传球, 面对对方门将, 他冷静地一脚射门, 球应声入网! 场上观众欢呼雀跃, 球队取得了领先。然而, 对方球队并未放弃, 他们加强了进攻, 试图扳平比分。4

可以问我任何问题...

智能体方案

Agent Solution



"Ladies and gentlemen, the match has begun! The midfield is buzzing with activity, with players from both teams fiercely jockeying for position. A quick pass moves towards the sideline but is intercepted by the opposition! The defenders are on high alert, clearing the ball away in time.

Now we have a race along the touchline, with players giving it their all. A cross enters the penalty area, but the goalkeeper reacts swiftly, making an excellent save! The crowd erupts into cheers, feeling the intensity of the match. The outcome remains uncertain!"

Due to the low output resolution, the model suffers from severe hallucinations after training.

智能体部署

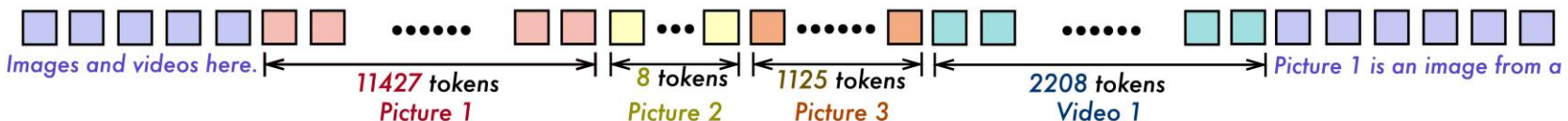
E2E Agent Solution

Dynamic Resolution

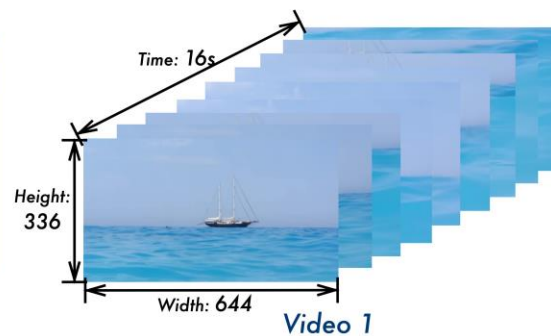
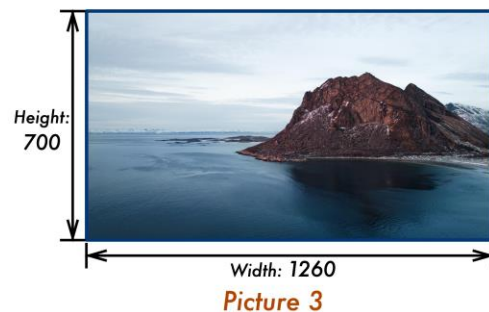
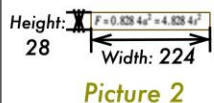
and videos here.



Picture 1 is an image from a blog



Native Resolution Input



Input Resolution: 1280x720

Sampling Frequency: 2 frames per second

Fine-tuned using LoRA (Low-Rank Adaptation)

Total video data: 5000 seconds

Trained for 10 hours on 8x NVIDIA A6000 GPUs

Final model quantized to INT8

Deployed with VLLM,
integrated into WenXin AI agent

72B model

Ultimately reduced hallucinations, resulting in
outputs that are more consistent with
the video content.

智能体部署

E2E Agent Solution



Tang足球赛事解... 智能体

昨天 15:48

<https://www.ikcest.org/bigdata2024/spzs/content/240919990358354644.html>

在一场激烈的足球比赛中，法国队和摩洛哥队正在场上展开激烈的争夺。比赛进行到78分钟时，法国队的球员们在中场附近积极组织进攻，试图突破摩洛哥队的防线。摩洛哥队的防守非常坚固，他们不断地拦截和破坏法国队的传球路线。

法国队的前锋在前场不断寻找机会，但摩洛哥队的后卫们紧紧盯防，不给对方任何射门的机会。比赛进入到飞奔的阶段，双方球员都更加频繁地尝试远射，希望能够打破僵局。

在一次快速反击中，法国队的中场球员成功将球传给了前场的队友，后者迅速带球突入禁区。面对摩洛哥队的门将，法国队的前锋冷静地起脚射门，皮球划过一道完美的弧线，飞进了球门的死角。

摩洛哥队的门将虽然做出了全力扑救，但无奈皮球已经越过了他的指尖，飞进了球网。法国队的球员们欢呼雀跃，庆祝这粒宝贵的进球。摩洛哥队的球员们则显得有些沮丧，但他们并没有放弃，继续努力寻找扳平比分的机会。

比赛还在继续，双方球员都在全力以赴，争取最后的胜利。

分享

👍 🗨





4

思考与总结

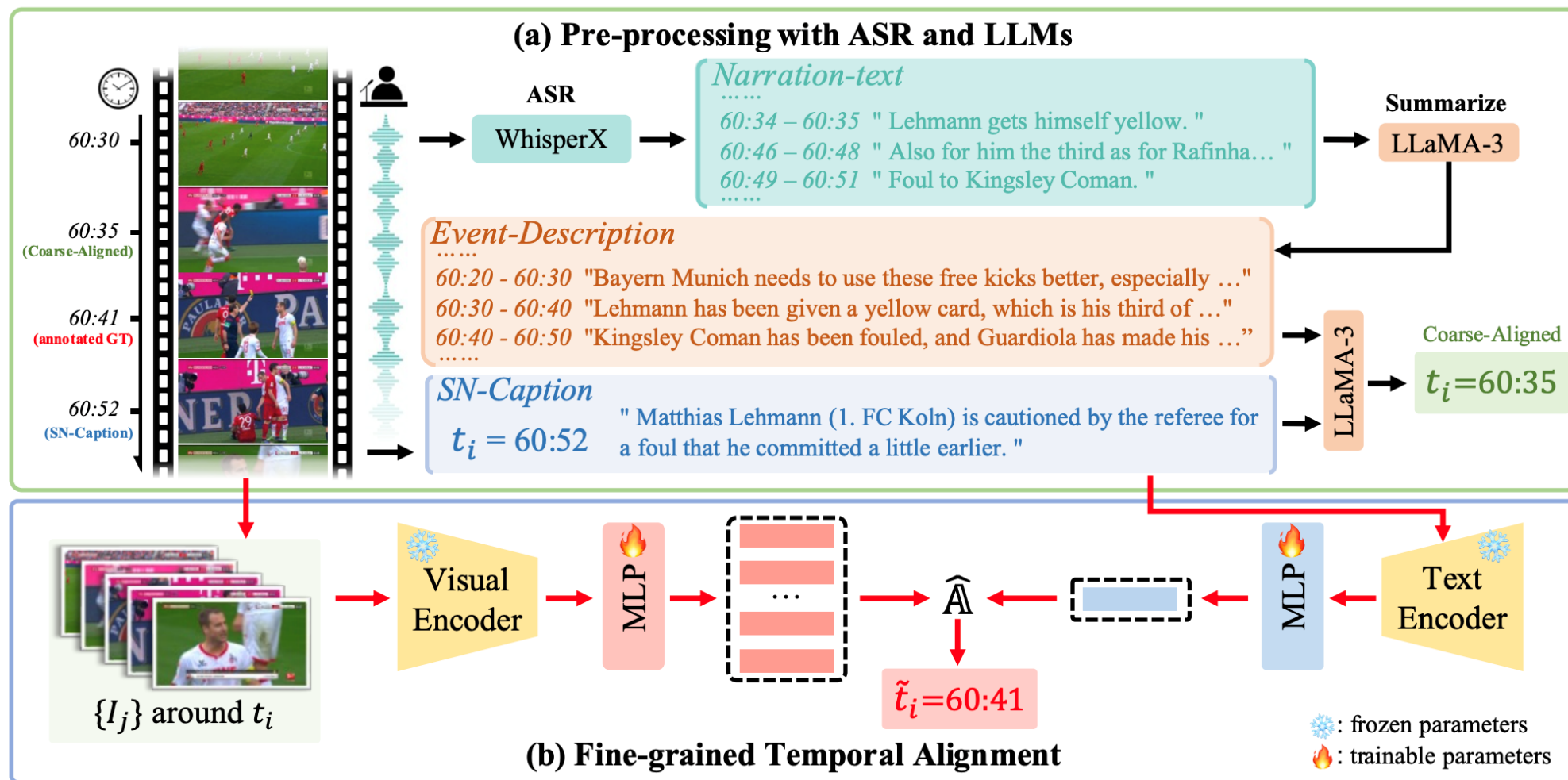
Reflection and Summary

后续思考

Further Reflections and Attempts



The football commentary dataset commonly suffers from the issue of temporal misalignment between video and text.



后续思考

Further Reflections and Attempts



The football commentary dataset commonly suffers from the issue of temporal misalignment between video and text.

	<p>SN-Caption: [PLAYER] ([TEAM]) is caught offside !</p> <p>MatchVoice: [PLAYER] ([TEAM]) is forced to stop his attacking move after the linesman signals for offside.</p> <p>GT: [PLAYER] ([TEAM]) is offside and the linesman raises his flag.</p>
	<p>SN-Caption: [PLAYER] ([TEAM]) takes the ball and sets it for the free kick .</p> <p>MatchVoice: [PLAYER] ([TEAM]) picks up the ball on the edge of the box and produces a brilliant low drive into the bottom right corner.</p> <p>GT: The ball is whipped in from the free kick and finds the head of [PLAYER] ([TEAM]), who rises and produces an amazing header inside the right post.</p>
	<p>SN-Caption: [PLAYER] ([TEAM]) takes the corner but fails to find any of his teammates .</p> <p>MatchVoice: [PLAYER] ([TEAM]) takes the corner with a short pass.</p> <p>GT: [PLAYER] ([TEAM]) quickly takes the corner kick with a short pass.</p>
	<p>SN-Caption: [PLAYER] ([TEAM]) is clearly asking for some medical attention with his painful gestures . The extent of his injury is yet to be discovered .</p> <p>MatchVoice: [PLAYER] ([TEAM]) is being forced to leave the pitch in order to receive medical treatment and his team will play with a man short for a few minutes.</p> <p>GT: The game is interrupted now, [PLAYER] ([TEAM]) picks up a knock and the physio has to come on.</p>

- a) Richer semantic descriptions, (b) complete commentary on multiple events within a single video, c) accuracy of the descriptions, and (d) prediction of incoming events.

Ref: MatchTime: Towards Automatic Soccer Game Commentary Generation



感谢评委的指导

Thank you for the judges' guidance.
