



TECHNICKÁ UNIVERZITA
V KOŠICIACH



KKUI
Katedra kybernetiky
a umelej inteligencie

Dokumentácia

SEMESTRÁLNE ZADANIE

Predmet: Pokročilé počítačové videnie

Študenti: Bc. Mark Chernomorchenko
Bc. Illia Postylha
Bc. Bohdan Orydoroha
Bc. Eduard Matovka
Bc. Heorhii Tolstoukhov
Bc. Laura Pituková

Vyučujúci: Ing. Ján Magyar, PhD.
doc. Ing. Marek Bundzel, PhD.

Obsah

1	Popis úlohy	2
2	Popis riešenia - Detekcia a sledovanie objektov	2
2.1	Centroidové trackovanie	2
3	Popis riešenia – Rozpoznávanie čísel na dresoch	3
4	Popis riešenia – Titulkovanie videa	3
4.1	Konfigurácia modelu	4
4.2	Generovanie komentárov	5
4.3	Experimenty a optimalizácia	5
5	Záver	5

1 Popis úlohy

Cieľom tohto projektu je riešiť tri úlohy definované platformou *SoccerNet*, ktoré sa týkajú spracovania futbalových videí:

- **Sledovanie hráčov (Player Tracking)** – sledovanie pohybu futbalistov medzi jednotlivými snímkami videa.
- **Rozpoznávanie čísiel na dresoch (Jersey Number Recognition)** – identifikácia čísiel na hráčskych dresoch.
- **Titulkovanie videa (Dense Video Captioning)** – generovanie opisov akcií počas zápasu na časovej osi.

Úlohy sú zamerané na analýzu videí futbalových zápasov pomocou metód počítačového videnia a hlbokého učenia.

2 Popis riešenia - Detekcia a sledovanie objektov

Videozáznamy futbalových zápasov vo formáte MP4, ideálne v rozlíšení minimálne 720p.

Na detekciu hráčov bol použitý predtrénovaný model **YOLOv8**, ktorý poskytuje bounding boxy v tvare (x, y, w, h) pre každého hráča. Model bol spustený pomocou knižnice **Ultralytics** v prostredí Python.

Kalibrácia kamery bola realizovaná nástrojom **TVCalib**. Z vybraných snímkov s viditeľnými čiarami na ihrisku sa vygenerovala homografická matica, ktorou sa transformovali stredy bounding boxov do reálnych súradníc ihriska (105×68 m).

Transformované pozície boli vykreslené na 2D minimape, kde boli hráči reprezentovaní farebnými bodmi podľa tímu a rozhodcovia inou farbou.

Na sledovanie pohybu objektov vo videosekvencii bol použitý YOLOv5 a algoritmus centroidového trackovania - CentroidTracker

2.1 Centroidové trackovanie

Centroid tracker sleduje stred spodnej hrany každého bounding boxu ako reprezentatívny bod objektu. Ak sa daný objekt v nasledujúcej snímke nenájde, algoritmus dočasne eviduje jeho „zmiznutie“. Objekt sa z trackovania odstráni až po prekročení prahovej hodnoty **maxDisappeared** (napr. 50 snímkov), čo znižuje riziko prerušenia trajektórie pri krátkodobom výpadku detekcie.

Aby sa odstránil šum spôsobený drobnými odchýlkami v detekcii polohy, bola implementovaná funkcia **smooth_trajectory()**, ktorá aplikuje kľzavý priemer na historické pozície objektov. Výsledkom sú plynulejšie trajektórie, ktoré lepšie odzrkadľujú skutočný pohyb.

Z každého bounding boxu bola extrahovaná dolná tretina (časť dresu s číslom), pričom sa vyberali snímky, kde hráč stál chrbtom ku kamere. Na rozpoznávanie čísla bol použitý modul **MMOCR**. OCR model bol aplikovaný na extrahované výrezy hráčov. Výstupom bolo rozpoznané číslo hráča ako reťazec. Vizualita hráčov bola sledovaná pomocou nástroja **PRReID**, ktorý spájal výskyty rovnakého hráča naprieč snímkami. Výsledné ID vzniklo kombináciou informácií z OCR a ReID.

Na vyhodnotenie kvality trackingu bola použitá metrika **GS-HOTA**, ktorá berie do úvahy presnosť pozícií a správnosť identity hráča. Hodnotenie vyžaduje CSV súbor vo formáte:

```
timestamp, team, player_id, x, y
```

3 Popis riešenia – Rozpoznávanie čísel na dresoch

Pre rozpoznávanie čísiel na dresoch hráčov je implementovaný kód v piatich krokoch:

1. **Generovanie sledov hráčov (trackletov):** Pre každého hráča vo videu sú vygenerované sledy pomocou modelu Centroid-ReID. Tento model priradzuje jedinečný identifikátor hráčovi a odstraňuje snímky, ktoré nepatria hlavnému subjektu.
2. **Filtrácia čitateľnosti čísla:** Používa sa *Legibility Classifier*, ktorý eliminuje snímky, na ktorých číslo nie je dostatočne čitateľné.
3. **Detekcia kľúčových bodov tela:** Na zvyšné snímky sa aplikuje model ViTPose, ktorý detekuje kľúčové body tela hráča a identifikuje oblasť trupu, kde sa očakáva číslo.
4. **Extrahovanie a OCR rozpoznávanie:** Detekovaná oblasť sa vyreže a odošle do OCR modelu PARSeq, ktorý rozpozná číslo.
5. **Agregácia a výber finálneho čísla:** Predikcie z OCR modelu sa agregujú a na základe všetkých predikcií sa vyberie finálne číslo pre celý tracklet.

4 Popis riešenia – Titulkovanie videa

Pre generovanie komentárov na futbalové zápasy sme použili model ‘unsloth/Meta-Llama-3.1-8B-Instruct’. Tento model bol upravený pre špecifickú úlohu generovania komentárov zo štruktúrovaných údajov o zápasoch.

Dataset, uložený v CSV formáte, obsahuje štruktúrované údaje o zápasoch, ktoré zahŕňajú pozície hráčov, lopty a ďalšie dôležité informácie, ktoré sú potrebné pre generovanie komentárov. Každý riadok obsahuje informácie o aktuálnom stave zápasu, ako sú pozície hráčov, lopty, rýchlosť lopty a ďalšie metadáta.

Na tréningovanie modelu sme použili nasledujúce knižnice:

- *Transformers*: pre prácu s modelmi typu Llama.
- *Accelerate*: na efektívne distribuované tréningovanie.
- *Axolotl*: na správu tréningového procesu a experimentov.

Tieto knižnice je možno nainštalovať pomocou pip:

```
pip install transformers accelerate axolotl
```

4.1 Konfigurácia modelu

Model je predtrénovaný na úlohy spracovania prirodzeného jazyka a bol upravený pomocou technológie LoRA (Low-Rank Adaptation), čo umožnilo jeho optimalizáciu pre generovanie komentárov v reálnom čase počas futbalového zápasu.

Pre tréning modelu je potrebné nastaviť konfiguráciu, ktorá určuje parametre modelu, ako aj tréningové nastavenia. Nižšie sú uvedené parametre konfigurácie, ktoré sme použili pre tréning:

```
# model params
base_model: unsloth/Meta-Llama-3.1-8B-Instruct

# dataset params
datasets:
  - path: dataset.csv
    type:
      system_prompt: "You are a professional live football (soccer) commentator.
      You are observing real-time match data: the positions of all players
      and the ball on the field. Your job is to generate vivid, exciting,
      and natural commentary based on what is currently happening in the
      match. You speak with enthusiasm and clarity, describing the actions
      as if broadcasting to a live audience on television. You understand
      the roles of different players, team strategies, typical movements
      (passes, shots, tackles, runs, positioning), and can infer the
      likely intentions behind them. You do not provide analysis or
      explanations | your role is to comment on what is happening as it
      unfolds, moment-by-moment. Your commentary should sound like that of
      a professional commentator such as Martin Tyler or Peter Drury.
      Input format: You receive structured match data including player
      positions, the ball position, and optionally the ball's velocity.
      Output format: Generate 1{3 sentences of vivid, flowing, real-time
      commentary for the current moment in the match. Use a natural,
      spoken tone. Be concise, energetic, and immersive. If a key moment
      is occurring | like a shot, goal, tackle, or counterattack | focus
      on building tension and excitement. Always pretend this is a real
      match happening right now."
      field_system: system
      field_instruction: field_state
      field_output: output
      format: "<|user|>\n {instruction} </s>\n<|assistant|>"
      no_input_format: "<|user|> {instruction} </s>\n<|assistant|>"

output_dir: ./models/Llama3-Commentator_v1

# model params
sequence_length: 1024
bf16: auto
tf32: false
```

```

# training params
micro_batch_size: 4
num_epochs: 4
optimizer: adamw_bnb_8bit
learning_rate: 0.0002

logging_steps: 1

# LoRA
adapter: lora
lora_r: 16
lora_alpha: 16
lora_dropout: 0.05
lora_target_linear: true

# Gradient Accumulation
gradient_accumulation_steps: 1

# Gradient Checkpointing
gradient_checkpointing: true

```

4.2 Generovanie komentárov

Na tréning modelu sme použili 4 epóky s mikro dávkami o veľkosti 4 a optimalizátorom *adamw_bnb_8bit*. Trénovanie modelu trvalo niekoľko hodín na vyžaduje minimálne GPU so 16 GB VRAM.

Po úspešnom natrénovaní modelu môžeme generovať komentáre na základe štruktúrovaných dát o zápase. Model prijíma ako vstup údaje o aktuálnych pozíciách hráčov, lopty a ďalších metadátach, a na ich základe generuje komentáre v reálnom čase.

4.3 Experimenty a optimalizácia

Pre optimalizáciu modelu sme použili technológiu LoRA, ktorá umožnila efektívnu adaptáciu modelu na špecifickú úlohu komentovania zápasov. Taktiež sme využili gradientné hromadenie a kontrolu gradientov, aby sme znížili pamäťové nároky počas tréningu.

Tento postup vám umožní replikovať našu implementáciu a generovať živé komentáre na futbalové zápasy pomocou modelu ‘unsloth/Meta-Llama-3.1-8B-Instruct’.

5 Záver

V tomto projekte sme úspešne vytvorili systém na analýzu futbalových videí, ktorý rieši tri základné úlohy: sledovanie hráčov, rozpoznávanie čísel na dresoch a generovanie komentárov k zápasu. Každá časť bola implementovaná s využitím moderných metód počítačového videnia a hlbokého učenia.

Na detekciu hráčov bol použitý model YOLOv8, na sledovanie centroidový tracker a pre rozpoznávanie čísel model MMOCR. Následne sme využili techniky kalibrácie a transformácie súradníc na zobrazenie hráčov na 2D mape ihriska. Pre zlepšenie vizuálnej kontinuity pohybu bol implementovaný systém vyhladzovania trajektórií.

Najpokročilejšou časťou riešenia bolo automatické generovanie komentárov pomocou veľkého jazykového modelu LLaMA-3, ktorý bol prispôsobený na danú úlohu pomocou techniky LoRA. Model bol trénovaný na štruktúrovaných dátach o zápase a dokázal vytvárať pútavé, dynamické a štýlovo vhodné komentáre v reálnom čase.

Projekt ukazuje, že prepojenie počítačového videnia s generatívnymi modelmi umožňuje vytvárať inteligentné systémy, ktoré môžu nájsť uplatnenie pri športovej analýze, automatickom spracovaní prenosov alebo obohatení diváckeho zážitku.

Kód je možné nájsť na GitHub repozitári:

<https://github.com/TeraKrono/SoccerNetTrackingYOLO.git>