Computer Vision in Sports: Scalable Solutions for Downmarket Leagues

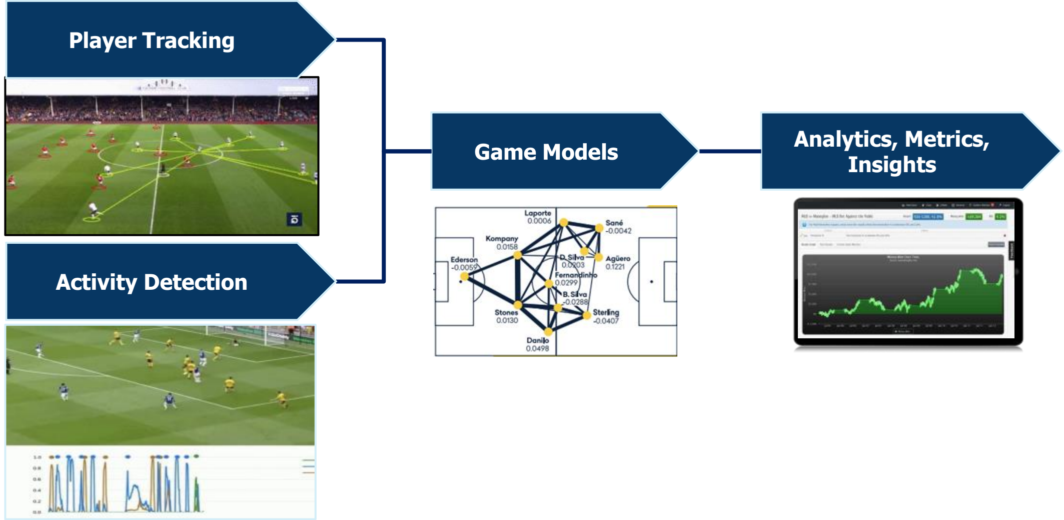
Mehrsan Javan

CTO

Sportlogiq

Sport is Data

- Analytics are everywhere
  - Performance evaluation, scouting, media content creation, prediction, …
- Types of data for analytics
  - Player location tracking and game events/activities
  - 2D/3D body pose data, biosignals from wearables



Data Acquisition in Sports

- [Semi-]Manual annotation on videos
- Wearables
  - Passive/active localization sensors
- Computer vision
  - Multi-camera systems (~4-24) static cameras in venues
  - Single-camera feed processing
    - One single panoramic feed
    - One single broadcast/tactical feed

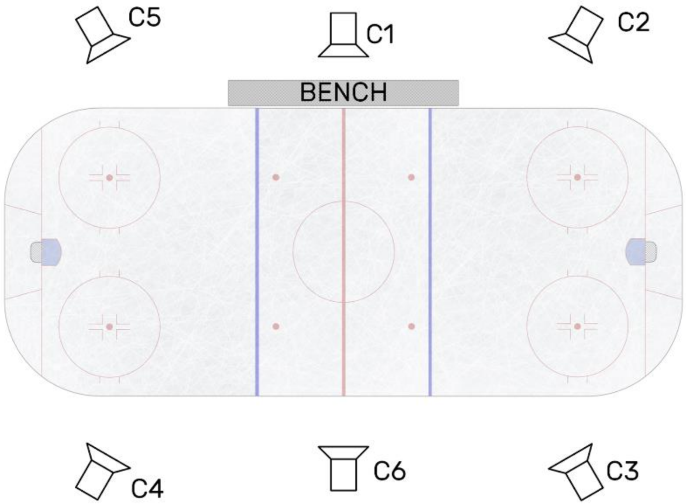
Data Acquisition in Sports

- Top tier pro leagues – Thousands of games
  - Player tracking: Wearables (NFL, NHL) or multi-camera systems (EPL, NBA, NHL)
  - Game events: Manual or semi-automated video tagging
- 2nd tier pro & draft eligible leagues - Tens of thousands of games
  - Player tracking: Single static camera or broadcast videos
- 3rd tier and amateur leagues - Millions of games
  - Video streams from inexpensive cameras in local sports facilities

Data Acquisition in Top Tier Pro Leagues

- On premise hardware for video capture and processing
  - 4k 30(60)fps feeds
- Real-time processing
  - 99.9% tracking accuracy with less than 10 cm localization error

An example six-camera setup in Hockey



Data Acquisition in Top Tier Pro Leagues

- Pros:
  - Full observation and accurate tracking data
  - Accurate event data
  - Manual QA processes
- Cons:
  - Cost – only affordable for top tier pro leagues

Example output data

Extending CV Beyond Pro Leagues

- Objectives
  - Generate comparable data in downmarket leagues
  - Acceptable quality and affordable cost
- Constraints
  - Use available video feeds
  - Ensure data comparability across millions of games

Democratizing Pro Tools – Example Player Metrics and Written Content in Downmarket

Smith Showing Improvement in Moving the Puck up the Ice

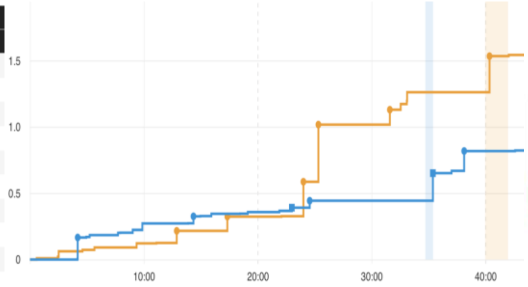
November 24, 2022

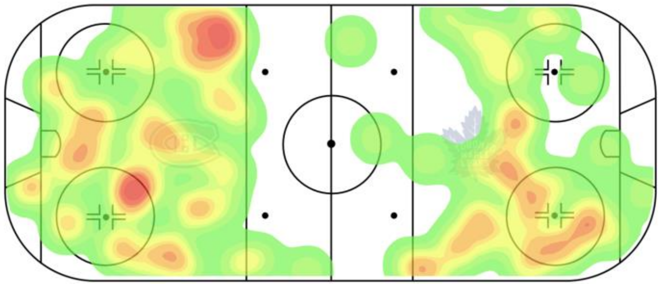
Last night Matt Smith faced off against one of the fastest offensive lines he has faced so far this season.

Though the Tigers ended up losing 3-1, Matt forced the majority of turnovers and made the Warriors earn every point. He had a decent passing game between Gorchek and Lansel, but the highlight of Matt's night was his ability to move the puck up ice.

Fast offensive players have been a menace to Matt and his defensive teammates all throughout the first half of the season. Earlier this month, however, they began to shut down stronger players, turning the puck over and creating more offensive zone time for the Tigers. That steady improvement was on display last night with 80% of the...

Players			Strength of Opposition vs Teammates		
	TOI	OZ start%	SOO	SOOΔ	SOT
#47 Engvall • #15 Kerfoot • #65 Mikheyev	10:25	50.0%	18.1	4.17	22.3
#97 Thornton • #19 Spezza • #71 Foligno	5:43	50.0%	18.0	0.18	18.2
#11 Hyman • #16 Marner • #34 Matthews	95:30	40.8%	17.9	12.0	29.9
#12 Galchenyuk • #88 Nylander • #15 Kerfoot	37:02	43.6%	20.8	1.37	22.2
#24 Simmonds • #15 Kerfoot • #65 Mikheyev	7:15	0%	18.2	1.15	19.3
#16 Marner • #88 Nylander • #34 Matthews	12:29	70.6%	18.4	12.9	31.3
#12 Galchenyuk • #71 Foligno • #88 Nylander	17:26	27.8%	17.8	1.94	19.7
#97 Thornton • #19 Spezza • #77 Brooks	15:26	33.3%	20.0	-6.46	13.5
#24 Simmonds • #47 Engvall • #65 Mikheyev	31:00	5.6%	18.6	-3.60	15.0
#97 Thornton • #19 Spezza • #24 Simmonds	13:47	50.0%	17.6	-0.33	17.2





Extending CV to Downmarket Hockey

- Data acquisition
  - Game segmentation - Faceoff/Whistles
  - Player/puck tracking - Detection, tracking, player & team ID, re-ID
  - Event detection, e.g., passes, shots, goals, penalties
- Insights and analytics
  - Manpower estimation and situational filters
  - Player impact ratings, expected goals and shots impact values
  - Game/player summary

Game Segmentation

- Broadcast feeds in downmarket leagues
  - Unreliable noisy game clock data
  - Partial observation of the game
  - Acceptable audio signal
- Panoramic feeds in lower-level leagues
  - No game clock
  - Poor audio signal
  - Full observation of the game

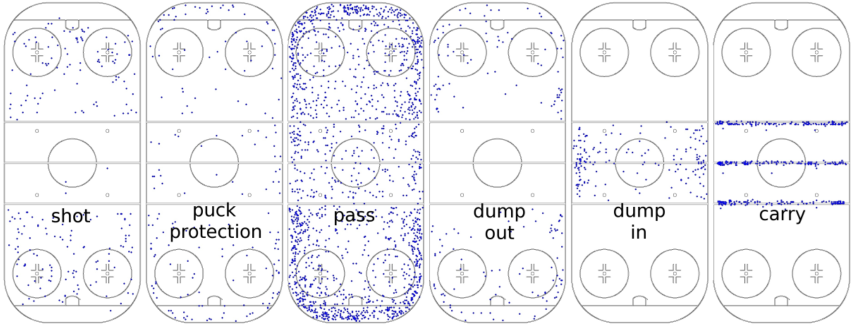
Game Segmentation - Multi-Stream Process

- Single-stream faceoff/whistle detection
  - Visual stream, 0.91 F-score
  - Trajectory stream, 0.92 F-score
  - Game clock reading, 0.75 F-score
  - Audio stream, 0.85 F-score
- Multi-stream with late fusion
  - 99.9 F-score on all games

F-score is the harmonic mean of the precision and recall

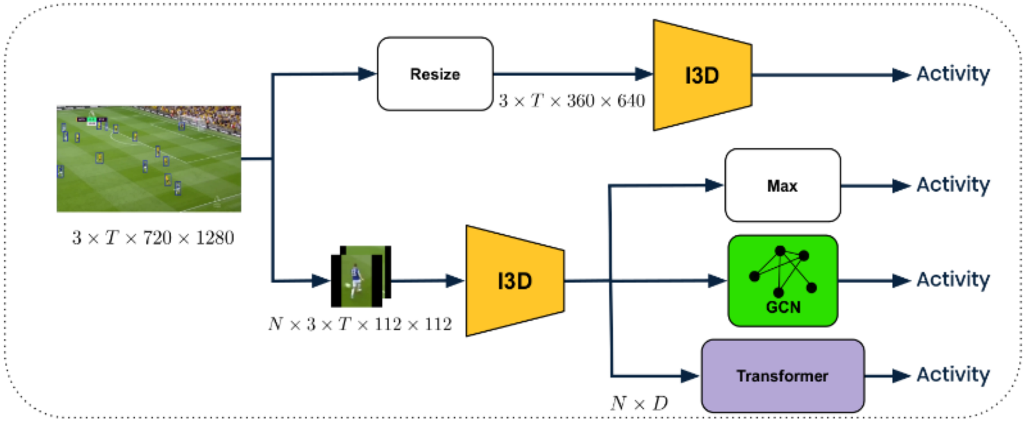
Event Detection

- Game events are firmly associated with spatial locations
- Some actions are visually similar, dump-in vs dump-out
- Some actions may need other sources of information, such as end of stick and body pose



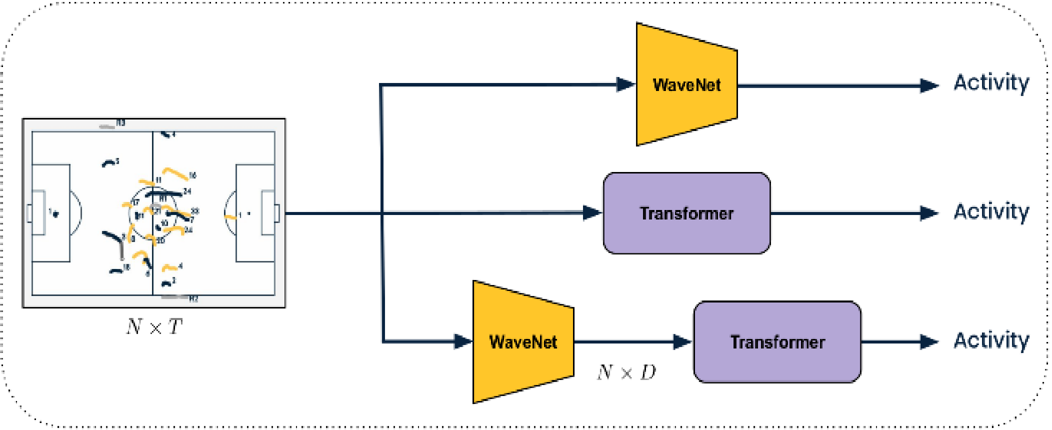
Event Detection - Multi-Stream Process

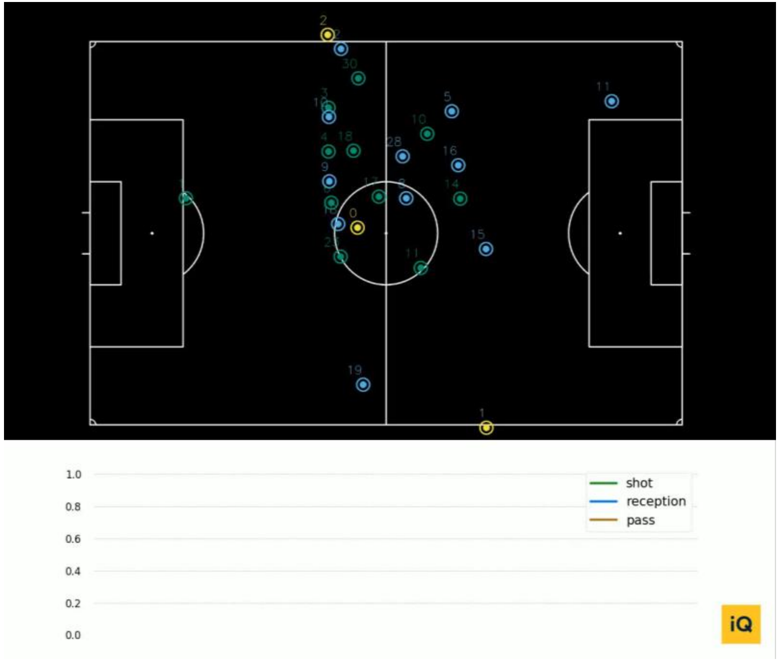
Visual Stream



Event Detection - Multi-Stream Process

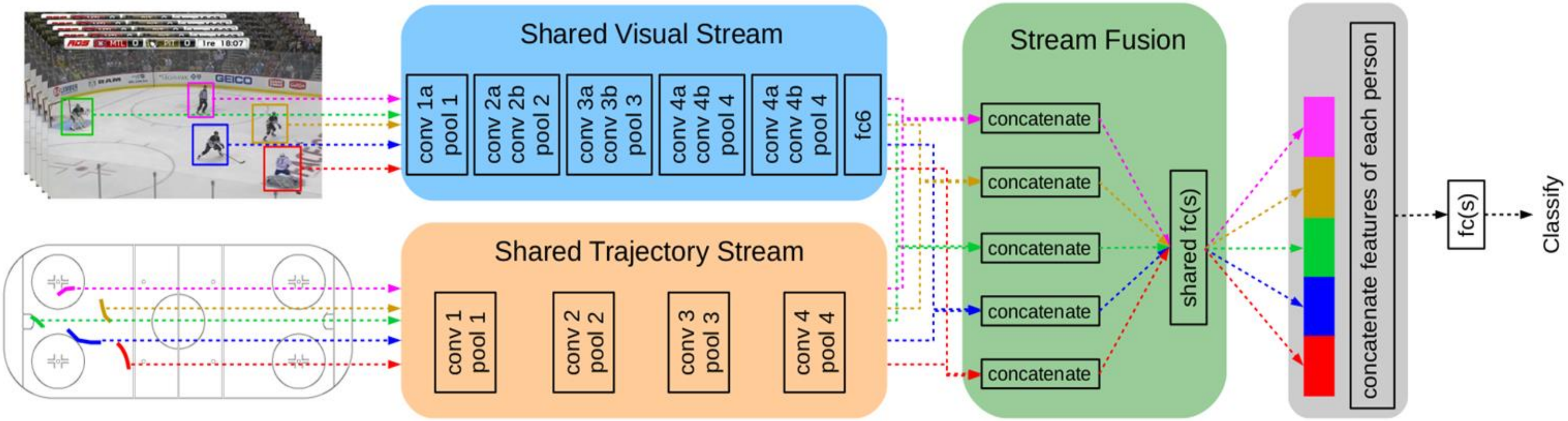
Trajectory Stream



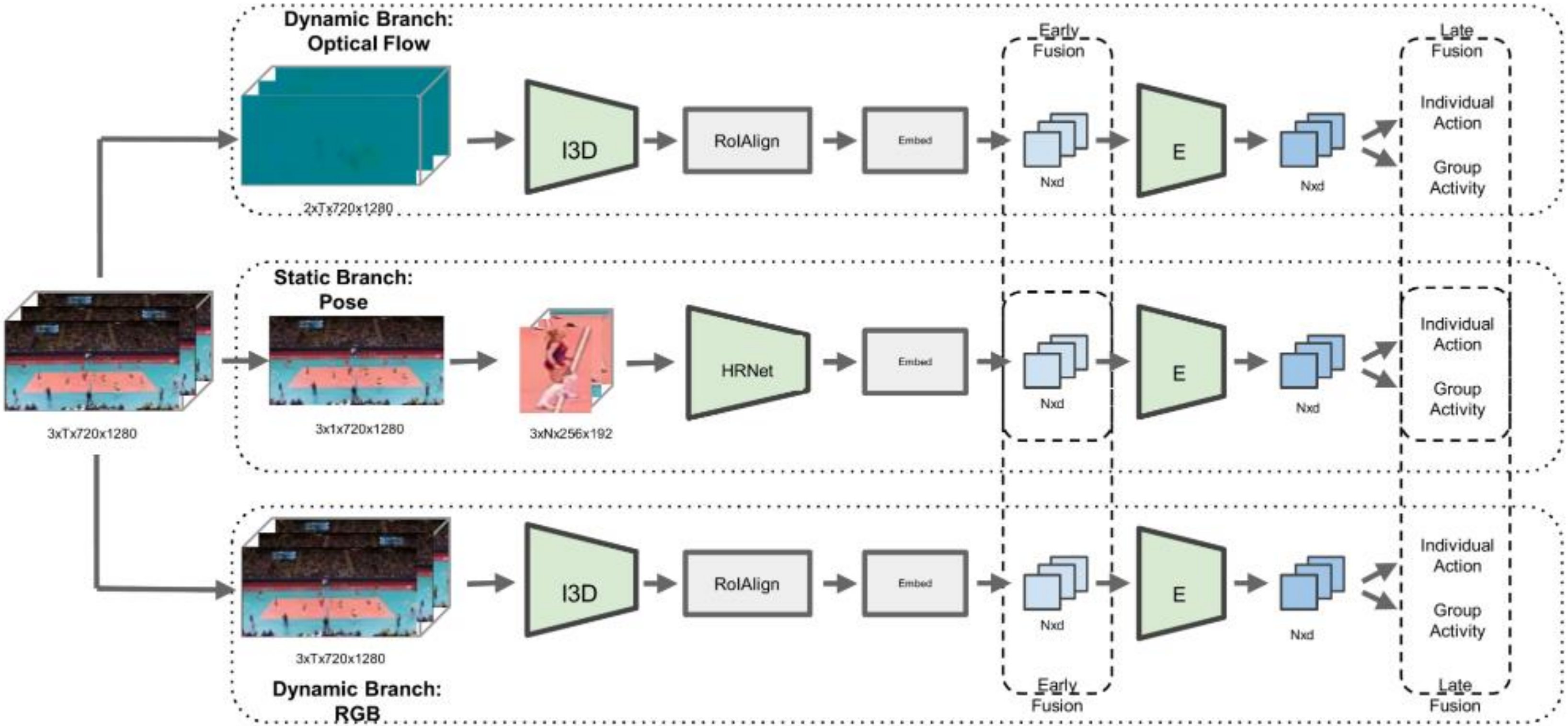


Event Detection - Multi-Stream Process

Fusion Between Trajectory and Visual Streams



Event Detection - Multi-Stream Process



Patent pending

Event Detection - Multi-Stream Process

- Explicitly including body pose is useful

Method	Backbone	Group Activity	Individual Action
I3D Baseline	I3D	93.0%	-
Transformer (RGB+Flow)	I3D	93.0%	83.7%
Transformer (Pose+RGB)	HRNet + I3D	93.5%	85.7%
Transformer (Pose+Flow)	HRNet + I3D	94.4%	85.9%

- Early fusions vs late fusion

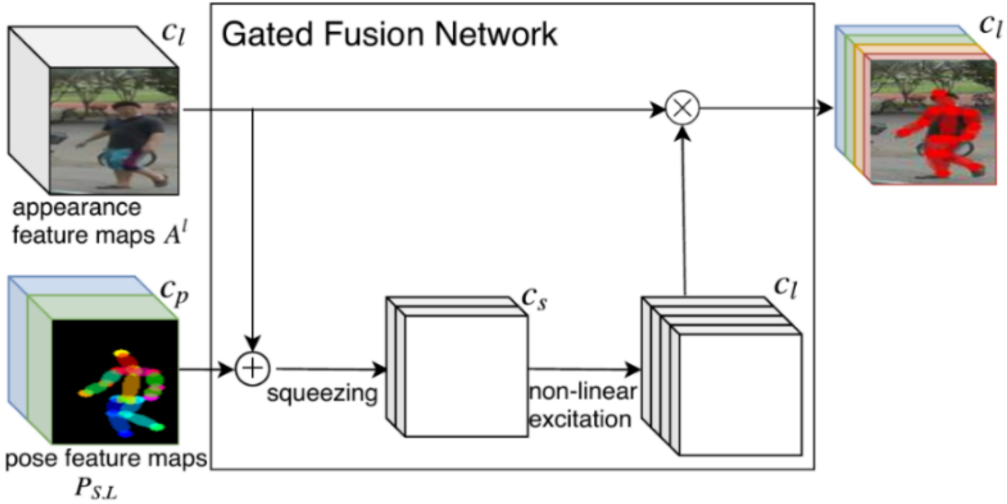
Method	Pose + RGB	Pose + Flow
Early - summation	91.2%	88.5%
Early - concatenation	91.8%	89.7%
Late	93.5%	94.4%

Event Detection Summary

- Multi-stream fusion improves event detection results
  - Trajectory + body pose + holistic visual models
  - Late fusion results in better performance
- Overall performance boost of at least 0.15 F-score

Multi-Stream Processing for Re-ID

- Mid-layer fusion
  - Joint-training of pose and re-ID models
- More than 8% improvement for sport players re-ID



Method and Dataset	Rank-1	mAP
Baseline (CUHK03-NP dataset)	56.93	54.64
Gated Fusion(CUHK03-NP dataset)	59.21	56.63
Baseline (Market-1501dataset)	91.81	77.85
Gated Fusion (Market-1501dataset)	93.47	91.02
Baseline (Duke dataset)	75.31	57.28
Gated Fusion (Duke Dataset)	76.39	60.38

Downmarket Data Acquisition – Conclusions

- Due to partial observations and incomplete information, single stream methods do not result in acceptable accuracy
- Multi-stream processing helps to extract maximum possible information from a video – Downside, increased processing cost
  - Tracking data, visual data, body pose data
- Late fusion combined with attention mechanism works the best for sport data collection
- Mid-Level fusion is optimal for some applications but difficult to train

Downmarket Data Acquisition – Challenges

- Significant varieties in the input distributions
  - Concept drift and continuous learning
- Unrecoverable edge cases
  - Poor video quality
  - Corrupted feed
  - Missing footage
  - Picture in picture
  - Bad camera placement

Other Challenges

- Move from offline processing to near real-time
- In-venue processing on edge devices
- Generalization guarantees, robustness, and scalability of the vision systems
- Full automation on content creation

For More Information

Videos

All videos and demos are available at https://sportlogiq.com

References

Shi F. et al. Self-Supervised Shape Alignment for Sports Field Registration, WACV2022

Gavrilyuk k. et al. Actor-Transformers for Group Activity Recognition, CVPR2020

Bhuiyan A. et al. Pose Guided Gated Fusion for Person Re-identification, WACV2020

Sanford R. et al. Group activity detection from trajectory and video data in soccer, CVPRW2020

Liu G. et al. Learning agent representations for ice hockey, NeurIPS2020

US Patents and pending applications 9,824,281; 11,130,040; 11,185,621; 11,176,706; 17/809,175; 17/817,454