8 Dec 2022
Tsinghua University

In partial fulfilment of the requirement for the degree of Master of Science in Computer Science and Technology

DETECTING SHUTTLECOCK HIT EVENTS IN PROFESSIONAL AND AMATEUR BADMINTON VIDEOS

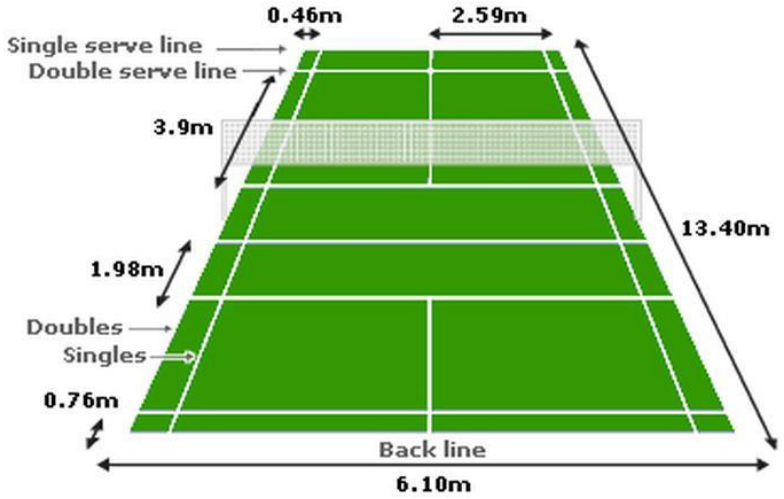
Student: Yoke Kai Wen, 2020280598
Thesis Supervisor: Professor Li Jianmin

CONTENTS

Motivation
Challenges
Usefulness of CV in badminton video annotation
Thesis Scope
Contributions

1. INTRODUCTION

1. INTRODUCTION - BADMINTON



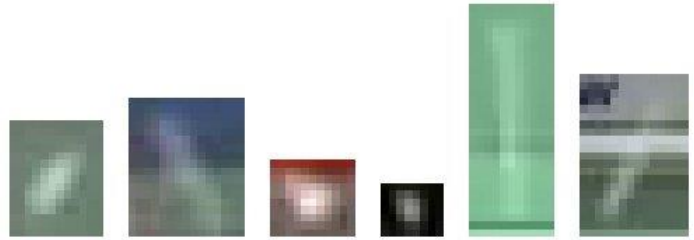
1. INTRODUCTION - MOTIVATION

1. Badminton is a popular sport
2. Badminton videos are very abundant, both professional and amateur
   - Many badminton hobbyists take match videos from arbitrary angles, and in a variety of court environments
   - Televised professional broadcast videos are also abundant
3. Badminton video annotation is useful
   - Downstream tasks: gameplay analysis, match judgement assistance
   - Useful features to annotate: objects (court, player, shuttlecock); events (stroke, shot, scoring), context (playing style, stroke quality, player strengths)

1. INTRODUCTION - MOTIVATION

1. INTRODUCTION – CHALLENGES IN BADMINTON VIDEO ANNOTATION

- Less attention compared to team sports like basketball, football
- Shuttlecock difficult to track
  - Small, fast, often occluded
- Automatic player detection and identification not trivial
  - Players switch sides after a match
  - Occlusion
  - Players in the same team for a doubles match wear the same kit
- Difficulty in generalizing to wide variety of camera angles and backgrounds
- Large variation in stroke appearance
  - Left and right handedness
  - variation in player skill level



1. INTRODUCTION – CV USEFUL FOR BADMINTON VIDEO ANNOTATION

1. Visible court lines provide free 3D calibration
2. Player stroke actions are clearly segmented around hit point
3. Less occlusion due to small number of players compared to team sports
4. Small motion variance among professional players (not for amateur)
5. Generic object detectors pretrained on massive datasets can be applied robustly to badminton videos

1. INTRODUCTION – THESIS SCOPE

1. Focus on representative event level task – detecting player hits temporally and spatially within a rally segment, after obtaining object-level annotations
2. Objectives:
    1. Improve hit detection accuracy
    2. Examine robustness of various techniques towards different camera angles and players of different skill levels
    3. Consider preliminarily how techniques that work on singles matches can extent to doubles matches
3. Assumptions:
    1. Camera: Monocular baseline-angled camera view, static and no shake
    2. Rally segment video as input
    3. Predominant focus on singles matches vs doubles matches
    4. Larger dataset of professional matches compared to amateur matches

1. INTRODUCTION - CONTRIBUTIONS

1. A small annotated dataset of amateur videos taken from a variety of camera angles
2. Video annotation tools and object-level annotation pipeline
3. A lightweight GRU-based hit detection algorithm that utilised player pose and shuttlecock features
4. An analysis of what techniques work and fail to generalize to amateur and doubles matches

Racket sports video analysis
Badminton video datasets
Gaps
CV and DL for badminton video analysis

2. RELATED WORK

2. RELATED WORK – RACKET SPORTS VIDEO ANALYSIS

Feature extraction

- Broadcast-related cues: interleaving between gameplay/closeups/replay, scoreboards[5,6,7,8]
- Court level: automatic vs manual[12, 7], traditional[9, 10, 11, 12, 13] vs DL[14]
- Frame level: HoG[11, 6], RGB deep visual embeddings[8, 14, 6]
- Player level: person detector + player id + feature extraction within bounding box[1, 7], training player detector from scratch[6], unsupervised temporal clustering for player identification[8, 14]
- Shuttlecock: traditional[27, 12, 28] vs DL (TrackNet)[29, 2]
- Audio: associating scene with sound[31, 5], simple [32, 7], audio-visual fusion[27, 28]

Feature processing for hit event detection

- Rule-based: Ball position coordinates[7, 1, 8]
- Learning-based: badminton domain specific features as input (MonoTrack)[1, 7, 30, 28] generic RGB features as input[6, 14]

2. RELATED WORK – BADMINTON VIDEO DATASETS

Most works prepare their own dataset

Only publicly available dataset is TrackNetv2 dataset[2]

- 50000 frames from 26 unique professional broadcast matches + 3 amateur matches
- Labeled with shuttlecock coordinates and hit timestamps

2. RELATED WORK - GAPS

Few works deal with multiple camera angles for badminton hit event detection

- Court detection in multiple camera angles[1, 13]
- Table-tennis hit event detection with player level features in unstructured environments[14]
- Shuttlecock tracking in various camera views[2]

Few works deal with doubles matches

- Only one work deals with table-tennis doubles matches, for rally scene detection[21]
- Another work deals with tennis ball trajectory estimation in both singles and doubles matches[27]

No previous works on amateur match videos

2. RELATED WORK – CV AND DL FOR BADMINTON VIDEO ANALYSIS

Pretrained CNNs for visual feature extraction – VGG, Inception, ResNet. *Easily accessible in tensorflow, pytorch.*

Pretrained object detectors for person extraction – R-CNN, YOLO. *Easily accessible in MMDetection, Detectron2.*

Pretrained 2D pose estimation – HRNet, Hourglass. *Easily accessible in MMPose, OpenPose.*

Video action recognition: 3D convolutions, multiple-stream networks, 2D convolution followed by temporal pooling

Badminton video datasets
Object level annotation pipeline
Hit detection algorithms
Extending to doubles matches

3. METHODOLOGY

3. METHODOLOGY – BADMINTON VIDEO DATASETS

Table 3.1 Dataset statistics for Professional and Amateur Datasets

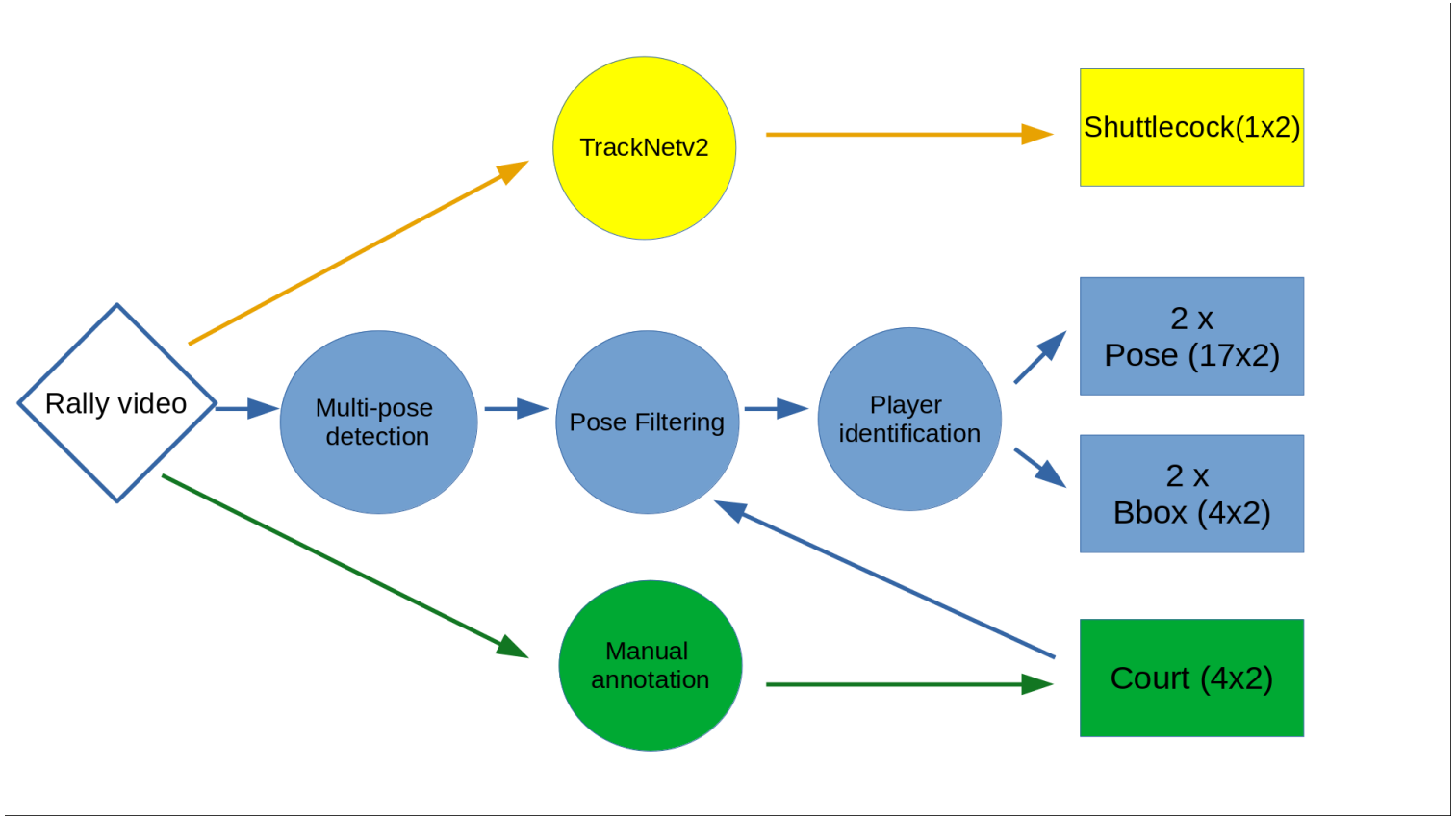
	Pro-train	Pro-test	Am-singles	Am-doubles
Number of unique matches	23	3	6	4
Number of rally videos	152	28	35	20
Number of frames	60379	10603	11393	7115
Number of hits	2222	365	331	264

TrackNetv2 train

TrackNetv2 test

Amateur

3. METHODOLOGY – OBJECT-LEVEL ANNOTATION PIPELINE



3. METHODOLOGY – COURT COORDINATES

Manual court annotation of four court corners

3. METHODOLOGY – PLAYER COORDINATES

Assume that four court corners are known

Multi-pose detection with pretrained R-CNN + HRNet models

Differentiate players and non-players by checking if they lie within court boundaries

- Check left or right foot joint

Differentiate between players

- Singles: near vs far player by checking y-coordinate
- Doubles: (not automatic) – match ReID features of bboxes with manually input gallery of player crops

3. METHODOLOGY – POSE DETECTION DEMO

3. METHODOLOGY – SHUTTLECOCK COORDINATES

TrackNetv2 model pretrained on TrackNetv2 dataset

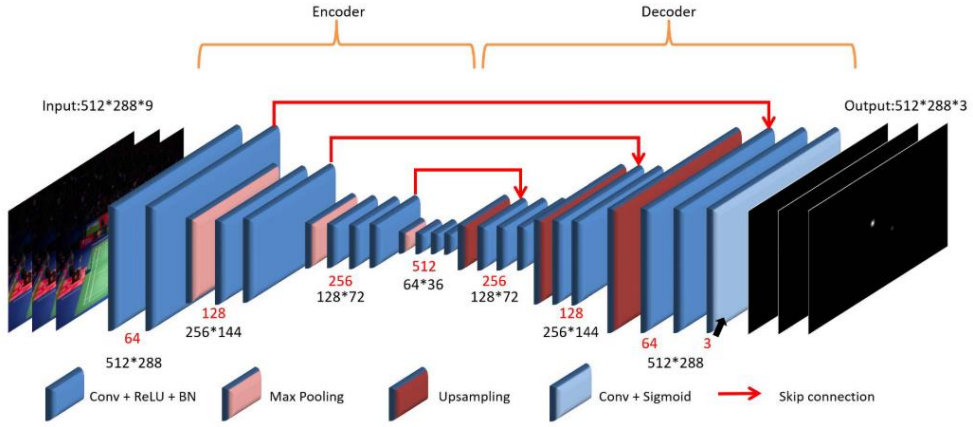
- SOTA CNN-based solution for shuttlecock tracking

Takes in 3 consecutive frames, outputs 3 consecutive ball detection heatmaps

- Exploit ball trajectory patterns
- Combat motion blur

U-Net based encoder-decoder architecture with skip connections between encoder and decoder layers

- Skip connections prevent information on tiny objects from getting lost



3. METHODOLOGY – SHUTTLE TRACKING DEMO

3. METHODOLOGY – HIT EVENT DETECTION

Aim: Localize hit event in space and time

Reformulate as multi-class classification problem

- Each frame can be assigned one out of three classes: (1) no hit, (2) hit by near player, (3) hit by far player
- For singles matches

3. METHODOLOGY – GRU-BASED HIT DETECTION ALGORITHM

Motivation

- Shallow and lightweight for fast performance
- Exploit temporal infomration

I/O:

- Input: k-dim features extracted from n consec frames
- Output: Hit class label

A sequence of n frames is considered a hit if a hit was present in the last n/2 frames

Table 3.2 Architecture of hit detection RNN, with $n = 16, k = 78$, GRU as the RNN layer

Layer	Output Shape	Param #
Input	(None, 16, 78)	0
Dense	(None, 16, 32)	2528
GRU	(None, 16, 32)	6336
GRU	(None, 32)	6336
Dropout	(None, 32)	0
Dense	(None, 3)	99
Total params		15299

3. METHODOLOGY – RNN INPUT FEATURES

Badminton domain features as model input

- Court (4x2), player pose (17x2), shuttlecock coordinates (1x2), following MonoTrack

Generic image features as model input (Poor)

- HoG, deep visual embeddings

Audio features (Failed)

3. METHODOLOGY – RESNET BASELINE

Standard image classification model, no temporal information

I/O:

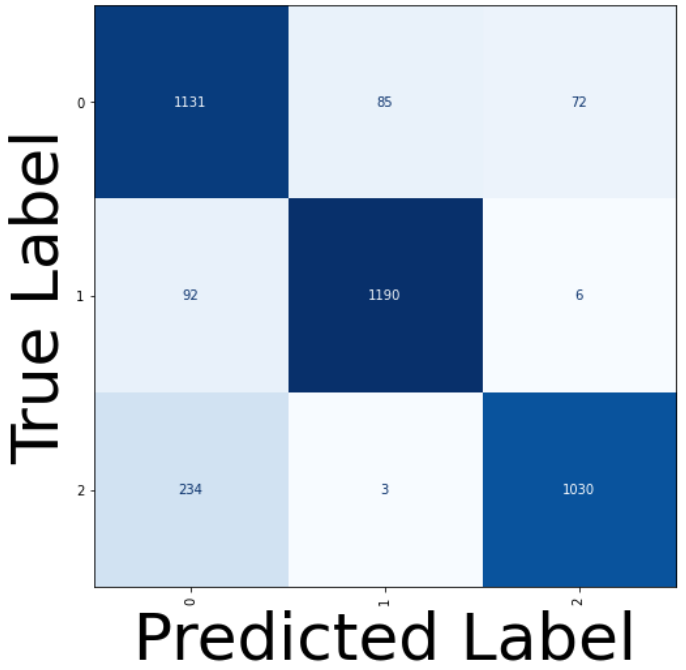
- Input: player crop
- Output: class label: no hit, near player hit, far player hit

Finetuned a pretrained ResNet-50 on player image crops

- Trained on Professional train dataset (~6k per class)
- Achieved good preliminary classification accuracy of 87.4% on static player crops from Professional test dataset (~1.2k per class)

Inference finetuned ResNet on badminton videos:

- ResNet runs twice per frame, once on each player crop
- Two labels for each frame are aggregated into one final label with simple rules

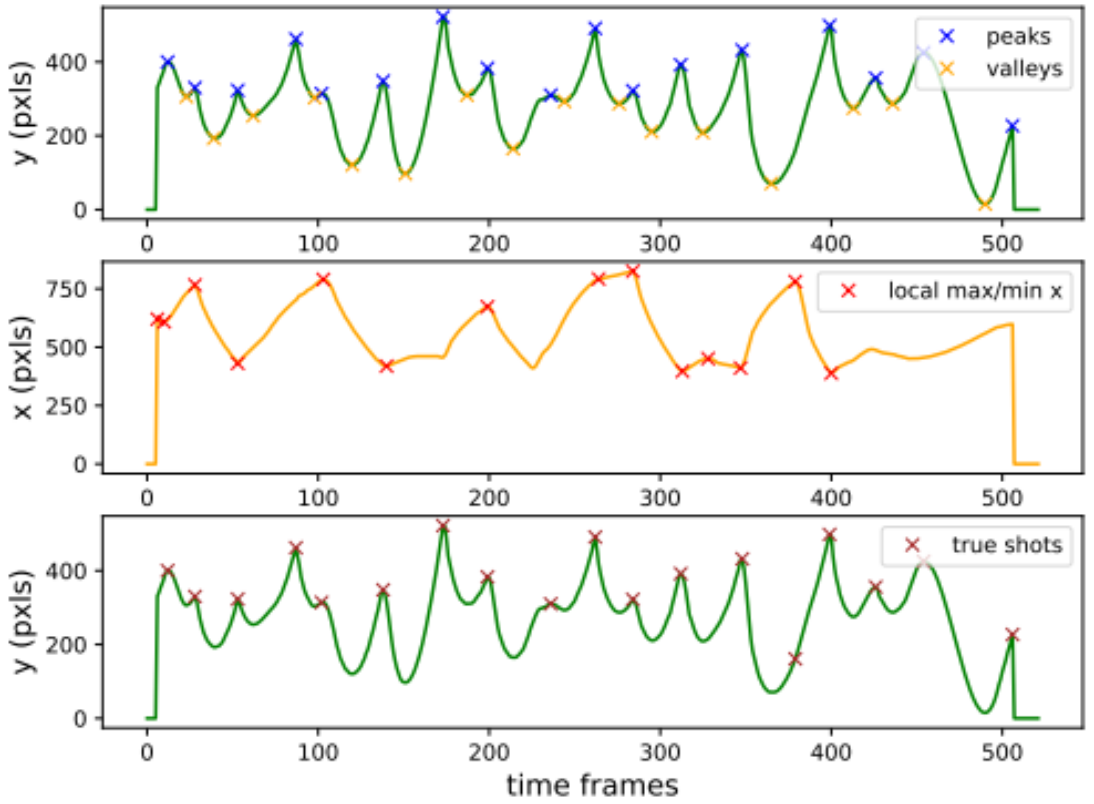


3. METHODOLOGY – RULE-BASED BASELINE

Hit is detected when second derivative of shuttlecock x or y-coordinates exceed threshold

Shuttlecock trajectory is first interpolated with a 3rd order polynomial, before obtaining 2nd derivative

2nd derivative threshold is set at 25, empirically determined to achieve balance between precision and recall



3. METHODOLOGY – EXTENDING HIT DETECTION TO DOUBLES MATCHES

Hit detection algorithms are trained only on singles matches

Propose to extend to doubles matches

- Four people instead of two
- Straightforward: divide four players into two groups of 1 near + 1 far player
- Run hit detector twice for each frame, once per group
- Aggregate the two output labels for each frame into one label

4. EXPERIMENTS

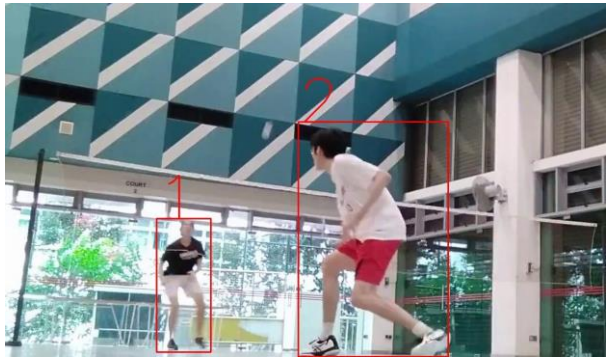
Object-level annotation pipeline performance
Evaluation implementation
Ablation study with different input features
Performance comparison with baselines
Qualitative analysis and demos

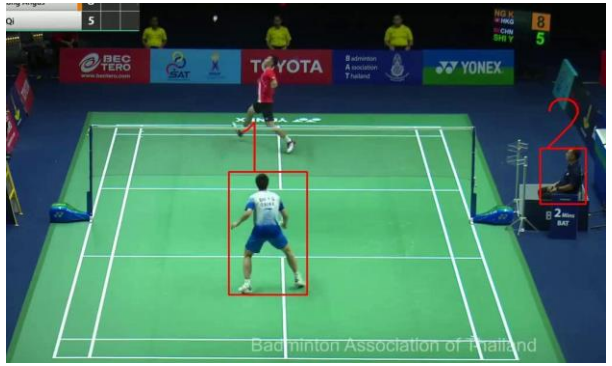
4. EXPERIMENTS – PERFORMANCE OF OBJECT-LEVEL ANNOTATION PIPELINE

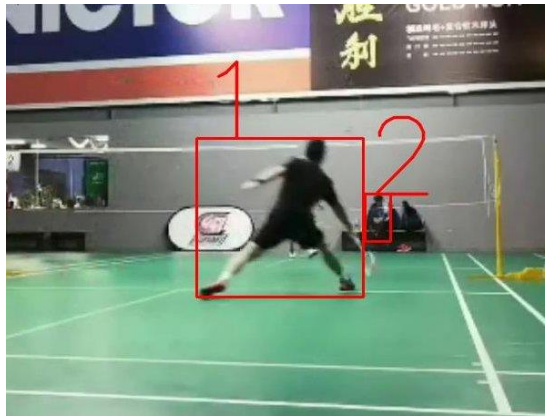
Table 4.1 Performance of annotation pipeline on Professional and Amateur Datasets

		Pro	Am-singles
Player ID (Near)	Visibility	0.99871	0.993652
	Accuracy	0.99929	0.92172
Player ID (Far)	Visibility	0.99871	0.93470
	Accuracy	0.98328	0.95494
Shuttlecock	Visibility	0.87309	0.83180
	Accuracy	-	0.91731
	Precision	-	0.92182
	Recall	-	0.91737
Court	Visibility	1	0.875

Player annotation error cases:







4. EXPERIMENTS – EVALUATION IMPLEMENTATION

Overall evaluation

- Ignores temporal structure of entire rally
- Input is a sequence of consecutive 14 frames, a ground-truth positive hit label is assigned as long as last 6 frames contain at least one positive hit label
- Window step-size of 3 frames for test set
- Metrics: Precision, Recall, F1, but taking only hits into account
  - G = {(frame, hit label != 0)}

$$P = \frac{|G \cap G'|}{|G'|}$$

$$R = \frac{|G \cap G'|}{|G|}$$

$$A = \frac{|G \cap G'|}{|G \cup G'|}$$

Video-level evaluation

- Considers temporal structure of entire rally
- Input is a sequence of 14 consecutive frames, with window step-size of 1 frame
- Hit action is treated as a 0.2s time interval → reformulate as temporal action localization problem, evaluate with mAP at different IoU thresholds

$$IoU = \frac{\text{predicted time window} \cap \text{ground truth time window}}{\text{predicted time window} \cup \text{ground truth time window}}$$

$$P = \frac{TP}{TP + FP} = \frac{\text{number of correct predicted windows}}{\text{total number of predicted windows}}$$

$$AP_k = \text{Average of } P \text{ for class } k \text{ across all videos}$$

$$mAP = \text{Average of } AP \text{ across all classes}$$

4. EXPERIMENTS – ABLATION STUDY WITH VARIOUS RNN INPUT FEATURES

Table 4.4 Hit classification performance with different input features

	Professional				Amateur			
	P	R	A	F1	P	R	A	F1
Player position	0.67137	0.84795	0.59923	0.74939	0.42354	0.54025	0.31133	0.47483
Pose	0.78137	0.82740	0.67186	0.80373	0.59234	0.67028	0.45869	0.62890
Shuttle	0.82576	0.89589	0.75346	0.85940	0.60823	0.61765	0.44186	0.61290
Shuttle + Pose	**0.92583**	**0.95753**	**0.88931**	**0.94141**	**0.77063**	**0.72291**	**0.59490**	**0.74601**
Court + Shuttle + Pose	0.88965	0.89452	0.80518	0.89208	0.54646	0.53715	0.37152	0.54176

4. EXPERIMENTS – PROPOSED VS BASELINES @ IOU=0.2

Proposed algorithm: optimal combination of shuttlecock + pose coordinates as input into RNN

Baselines:

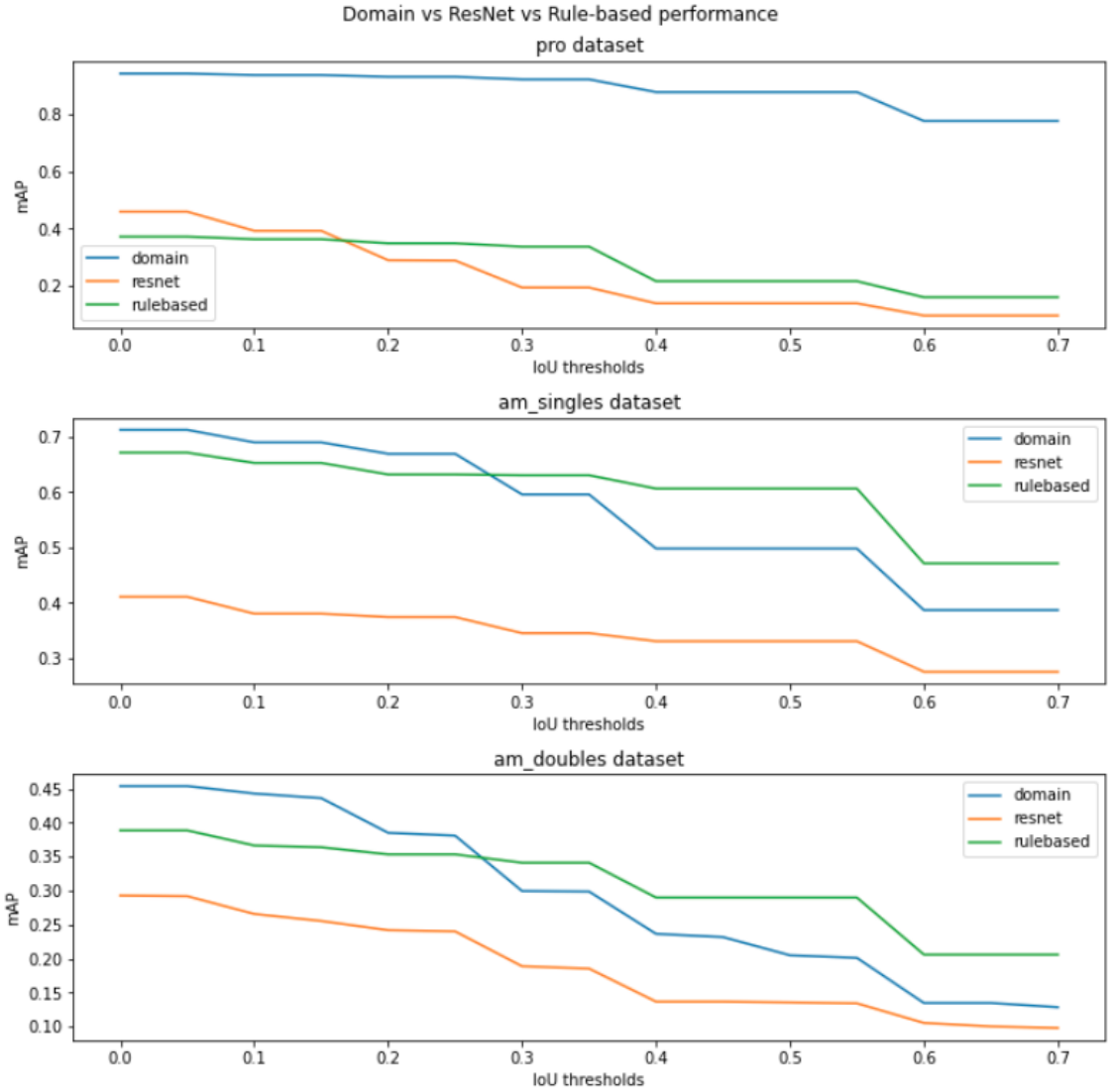
- ResNet
- rules-based

Table 4.5 Performance of proposed algorithm compared with two baselines @ IoU=0.2

		Proposed (domain)	ResNet	Rule-based
Pro	mAP	**0.93106**	0.29017	0.34918
	AP1	**0.90962**	0.24375	-
	AP2	**0.95250**	0.33660	-
Am-singles	mAP	**0.66961**	0.37464	0.63225
	AP1	**0.77700**	0.30106	-
	AP2	**0.56226**	0.44822	-
Am-doubles	mAP	**0.38513**	0.24190	0.35339
	AP1	**0.41382**	0.22411	-
	AP2	**0.35644**	0.25970	-

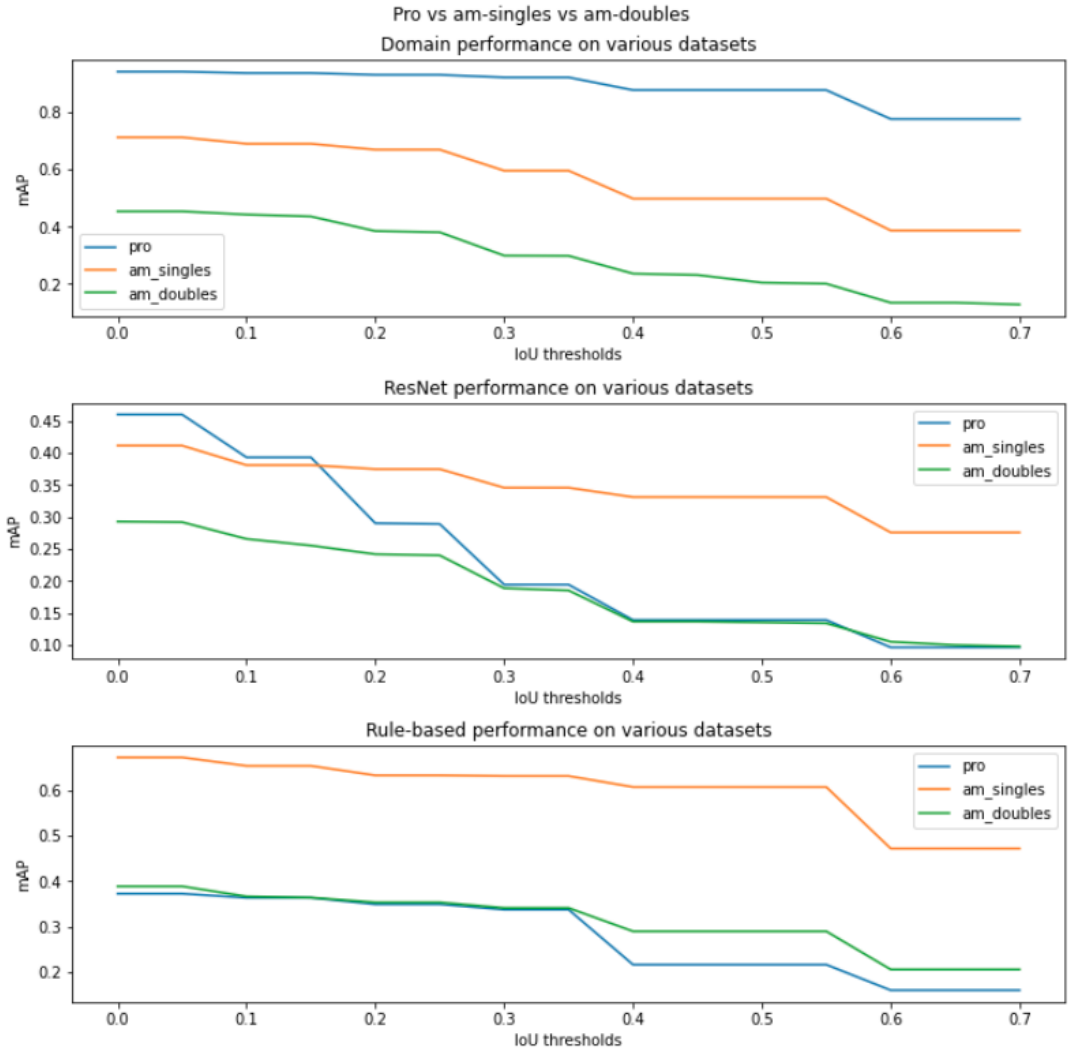
4. EXPERIMENTS – PERFORMANCE AT VARYING IOU THRESHOLDS

Proposed vs ResNet vs Rule-based



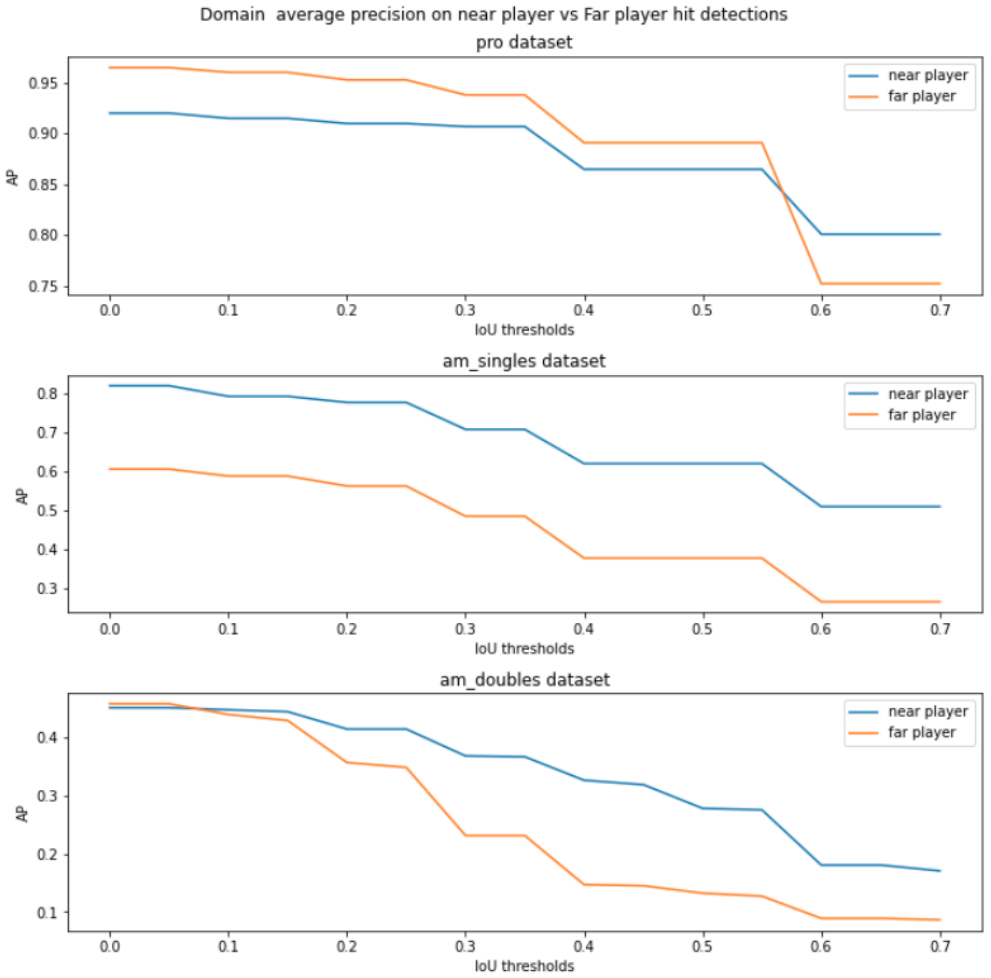
4. EXPERIMENTS – PERFORMANCE AT VARYING IOU THRESHOLDS

Pro vs am-singles vs am-doubles



4. EXPERIMENTS – PERFORMANCE AT VARYING IOU THRESHOLDS

Near player vs far player



4. EXPERIMENTS – DEMOS OF PROPOSED HIT DETECTION ALGORITHM

https://github.com/kwyoke/Badminton-hit-detection

- Flash blue – near player hit
- Flash red – far player hit

Professional demo: https://youtu.be/Sga5BMbK9Qk

Amateur singles demo: https://youtu.be/WpQMvr3_JuY

Amateur doubles demo: https://youtu.be/79Vh_RI03KY

4. EXPERIMENTS – QUALITATIVE ANALYSIS

Camera angle

- Proposed domain-based method suffers from large camera angle variations in amateur videos – hit from near player is often detected as hit by far player. E.g. https://www.youtube.com/watch?v=ieLmhx0r1PQ , https://www.youtube.com/watch?v=bM7ez-uBKwo

Robust to errors in shuttlecocks, poses

- Even when shuttlecocks or poses are not visible for a long time, proposed method is able to detect hit. E.g. https://www.youtube.com/watch?v=Gle4XFsr6t8 , https://www.youtube.com/watch?v=PrToOe11lbI , https://youtu.be/Sga5BMbK9Qk

Ability to extend from singles to doubles

- https://youtu.be/79Vh_RI03KY

5. CONCLUSION

5. CONCLUSION

Value of work done

- Proposed GRU-based model is small and lightweight, and yields excellent performance on hit detection for Pro dataset
- Proposed method has decent transferability to the Amateur Dataset
- Extending to doubles matches shows encouraging results
- Highlight the difficulty of hit detection on amateur and doubles videos

Limitations and future work

- Information in doubles matches not exploited
- Poor performance of baselines, especially ResNet
- Multimodal feature learning: domain objects + RGB + audio
- Amateur dataset is too small
- Expanding from hit detection to other badminton video analysis tasks

KEY REFERENCES

[1] Liu P, Wang J H. Monotrack: Shuttle trajectory reconstruction from monocular badminton video [M/OL]. arXiv, 2022. https://arxiv.org/abs/2204.01899. DOI: 10.48550/ARXIV.2204.01899.

[2] Sun N E, Lin Y C, Chuang S P, et al. Tracknetv2: Efficient shuttlecock tracking network[C/OL]// 2020 International Conference on Pervasive Artificial Intelligence (ICPAI). 2020: 86-91. DOI: 10.1109/ICPAI51961.2020.00023.

[3] BadmintonBites. How many badminton players are there in the world?[J/OL]. BadmintonBites, 2021. https://badmintonbites.com/how-many-badminton-players-are-there-in-the-world/.

[4] Brennan E. Badminton world federation tops sportonsocial ranking for 2022[J/OL]. Inside the Games, 2022. https://www.insidethegames.biz/articles/1121673/badminton-world-federation-tops-ranking.

[5] Liu C, Huang Q, Jiang S, et al. A framework for flexible summarization of racquet sports video using multiple modalitiesf[J/OL]. Computer Vision and Image Understanding, 2009, 113 (3): 415-424. https://www.sciencedirect.com/science/article/pii/S1077314208001355. DOI: https://doi.org/10.1016/j.cviu.2008.08.002.

[6] Ghosh A, Singh S, Jawahar C V. Towards structured analysis of broadcast badminton videos [C/OL]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV). 2018: 296-304. DOI: 10.1109/WACV.2018.00039.

[7] Bosi S. Audio-video techniques for the analysis of players behaviour in badminton matches[D]. 2021.

[8] Deng D, Wu J, Wang J, et al. Eventanchor: Reducing human interactions in event annotation of racket sports videos[J/OL]. CoRR, 2021, abs/2101.04954. https://arxiv.org/abs/2101.04954.

[9] Han J, Farin D, de With P H N. Broadcast court-net sports video analysis using fast 3-d camera modeling[J/OL]. IEEE Transactions on Circuits and Systems for Video Technology, 2008, 18 (11): 1628-1638. DOI: 10.1109/TCSVT.2008.2005611.

[10] Chu W T, Situmeang S. Badminton video analysis based on spatiotemporal and stroke features [C/OL]//ICMR '17: Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval. New York, NY, USA: Association for Computing Machinery, 2017: 448–451. https://doi.org/10.1145/3078971.3079032.

[11] Ghosh A, Jawahar C V. Smarttennistv: Automatic indexing of tennis videos[J/OL]. CoRR, 2018, abs/1801.01430. http://arxiv.org/abs/1801.01430.

[12] Dierickx T. Badminton game analysis from video sequences[D]. 2014.

[13] Ma H, Ding X. Robust automatic camera calibration in badminton court recognition[C/OL]// 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC). 2022: 893-898. DOI: 10.1109/IPEC54454.2022.9777532.

[14] Chaudhury S, Kimura D, Vinayavekhin P, et al. Unsupervised temporal feature aggregation for event detection in unstructured sports videos[J/OL]. CoRR, 2020, abs/2002.08097. https: //arxiv.org/abs/2002.08097.

[16] Yoshikawa F, Kobayashi T, Watanabe K, et al. Automated service scene detection for badminton game analysis using chlac and mra[J/OL]. International Journal of Computer and Information Engineering, 2010, 4(2): 331 - 334. https://publications.waset.org/vol/38.

[21] Yan C, Li X, Li G. A new action recognition framework for video highlights summarization in sporting events[J/OL]. CoRR, 2020, abs/2012.00253. https://arxiv.org/abs/2012.00253.

[23] Yoshikawa Y, Shishido H, Suita M, et al. Shot detection using skeleton position in badminton videos[C/OL]//Nakajima M, Kim J G, Lie W N, et al. International Workshop on Advanced Imaging Technology (IWAIT) 2021: volume 11766. SPIE, 2021: 312 - 317. https://doi.org/10

.1117/12.2590407.

[24] Cao Z, Hidalgo G, Simon T, et al. Openpose: Realtime multi-person 2d pose estimation using part affinity fields[J/OL]. CoRR, 2018, abs/1812.08008. http://arxiv.org/abs/1812.08008.

[25] Wang J, Sun K, Cheng T, et al. Deep high-resolution representation learning for visual recognition[J/OL]. CoRR, 2019, abs/1908.07919. http://arxiv.org/abs/1908.07919.

[26] Contributors M. Openmmlab pose estimation toolbox and benchmark[EB/OL]. 2020. https://github.com/open-mmlab/mmpose.

[27] Yan F, Kittler J, Windridge D, et al. Automatic annotation of tennis games: An integration of audio, vision, and learning[J/OL]. Image and Vision Computing, 2014, 32(11): 896-903. https://www.sciencedirect.com/science/article/pii/S0262885614001309. DOI: https://doi.org/ 10.1016/j.imavis.2014.08.004.

[28] Huang Q, Cox S, Zhou X, et al. Detection of ball hits in a tennis game using audio and visual information[C]//Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference. 2012: 1-10.

[29] Huang Y C, Liao I N, Chen C H, et al. Tracknet: A deep learning network for tracking highspeed and tiny objects in sports applications[C/OL]//2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). 2019: 1-8. DOI: 10.1109/AVSS.2019.8909871.

[30] Ju N P, Yu D R, İk T U, et al. Trajectory-based badminton shots detection[C/OL]//2020 International Conference on Pervasive Artificial Intelligence (ICPAI). 2020: 64-71. DOI:10.1109/ICPAI51961.2020.00020.

[31] Xing L, Ye Q, Zhang W, et al. A scheme for racquet sports video analysis with the combination of audio-visual information[C/OL]//Li S, Pereira F, Shum H Y, et al. Visual Communications and Image Processing 2005: volume 5960. SPIE, 2005: 259 - 267. https://doi.org/10.1117/12.631382.

[32] Guo X, Zhong B, Lin L, et al. Recognition of technical action types based on main frequency of hitting sound spectrum of elite badminton players[C/OL]//2021 International Conference on Information Technology and Contemporary Sports (TCS). 2021: 571-577. DOI: 10.1109/TC S52929.2021.00122.

THANK YOU FOR LISTENING!

Any questions?