Jersey Number Recognition using Keyframe Identification from Low-Resolution Broadcast Videos

Bavesh Balaji, Jerrin Bright, Harish Prakash, Yuhao Chen, David Clausi and John Zelek

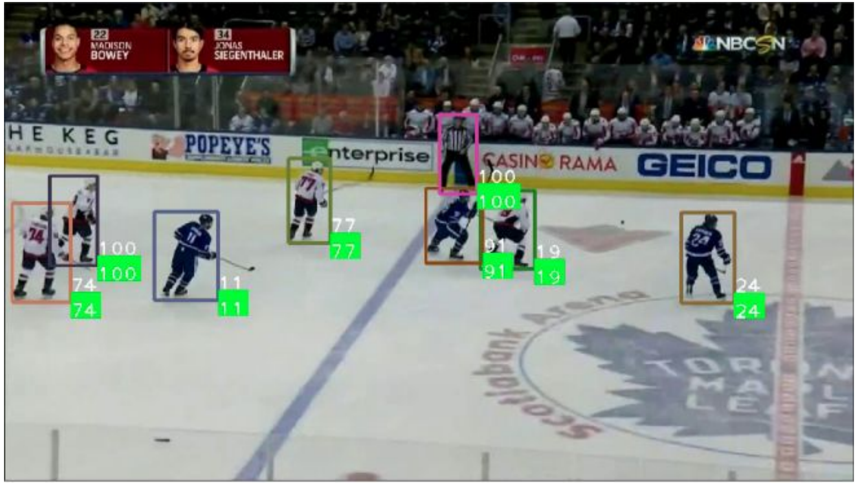
Vision and Image Processing Lab,

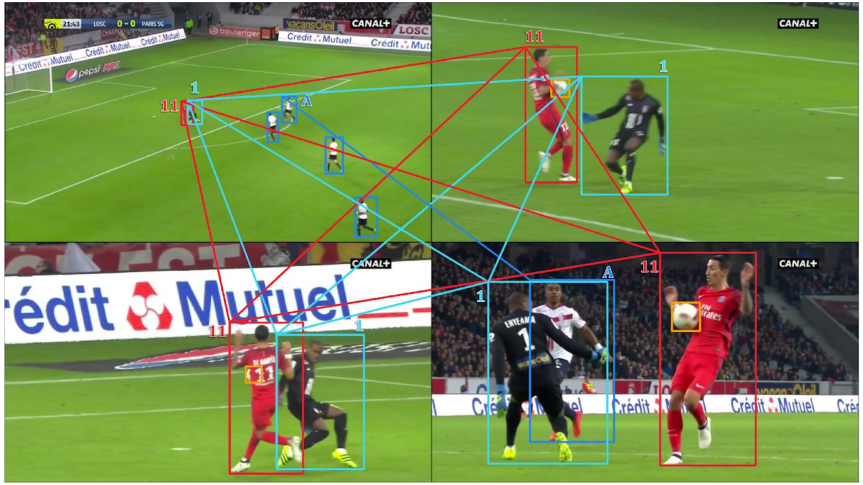
University of Waterloo,

Waterloo, Ontario, Canada

MOTIVATION

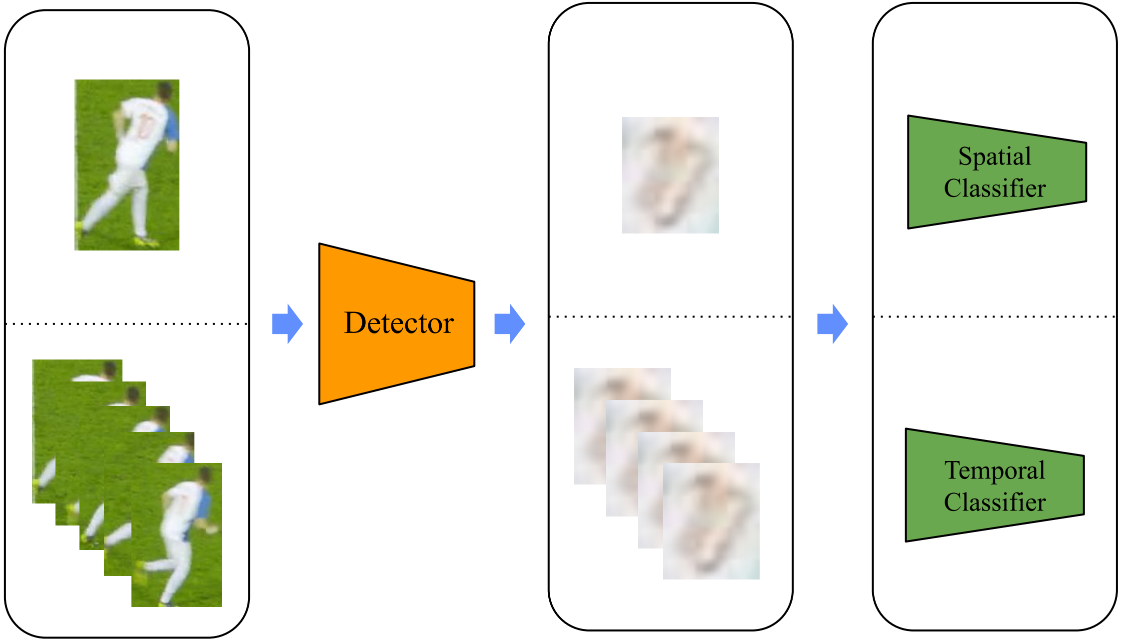
- Common approach for player identification.
- In-game analytics, enhanced broadcast experience.





EXISTING WORKS

- Formulate as a classification problem.
  - Most methods operate on static images[1, 2].
    - Do not consider temporal aspect.
    - Datasets created in isolated environments.
  - Few works use tracklets[3, 4].
    - Consider temporal aspect.

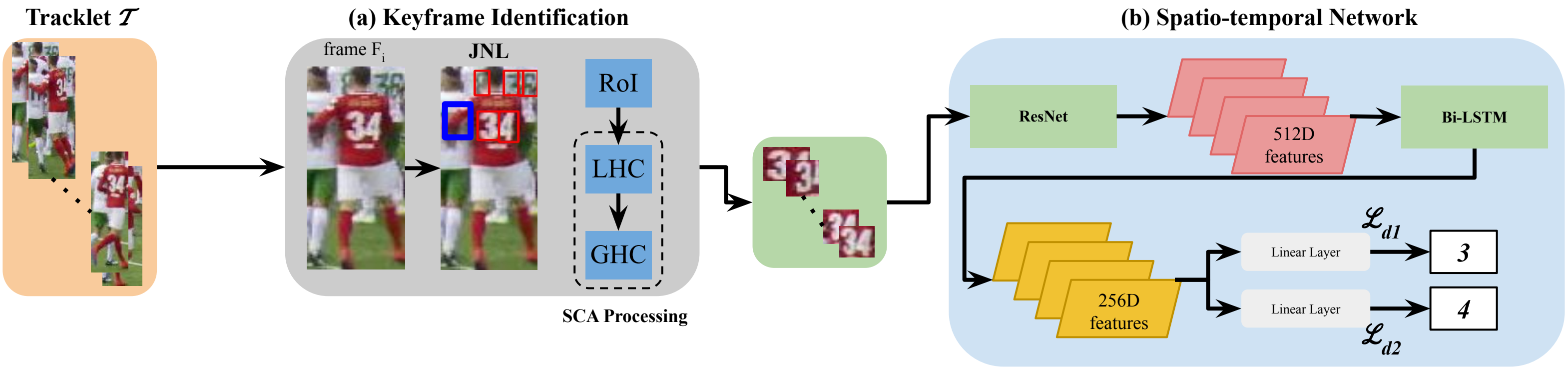


[1] D. Bhargavi, E. P. Coyotl, and S. Gholami, “Knock, knock. who’s there? – identifying football player jersey numbers with synthetic data,” 2022.
[2] G. Li, S. Xu, X. Liu, L. Li, and C. Wang, “Jersey number recognition with semi-supervised spatial transformer network,” in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1864–18647, 2018.
[3] K. Vats, W. J. McNally, P. Walters, D. A. Clausi, and J. S. Zelek, “Ice hockey player identification via transformers and weakly supervised learning,” 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).
[4] A. Chan, M. D. Levine, and M. Javan, “Player identification in hockey broadcast videos,” Expert Syst. Appl., vol. 165, p. 113891, 2020.

LIMITATIONS

- **Bias** - JN not visible in most frames.
- Error gets accumulated over frames in real-world data.
- Prone to motion blur & occlusions.

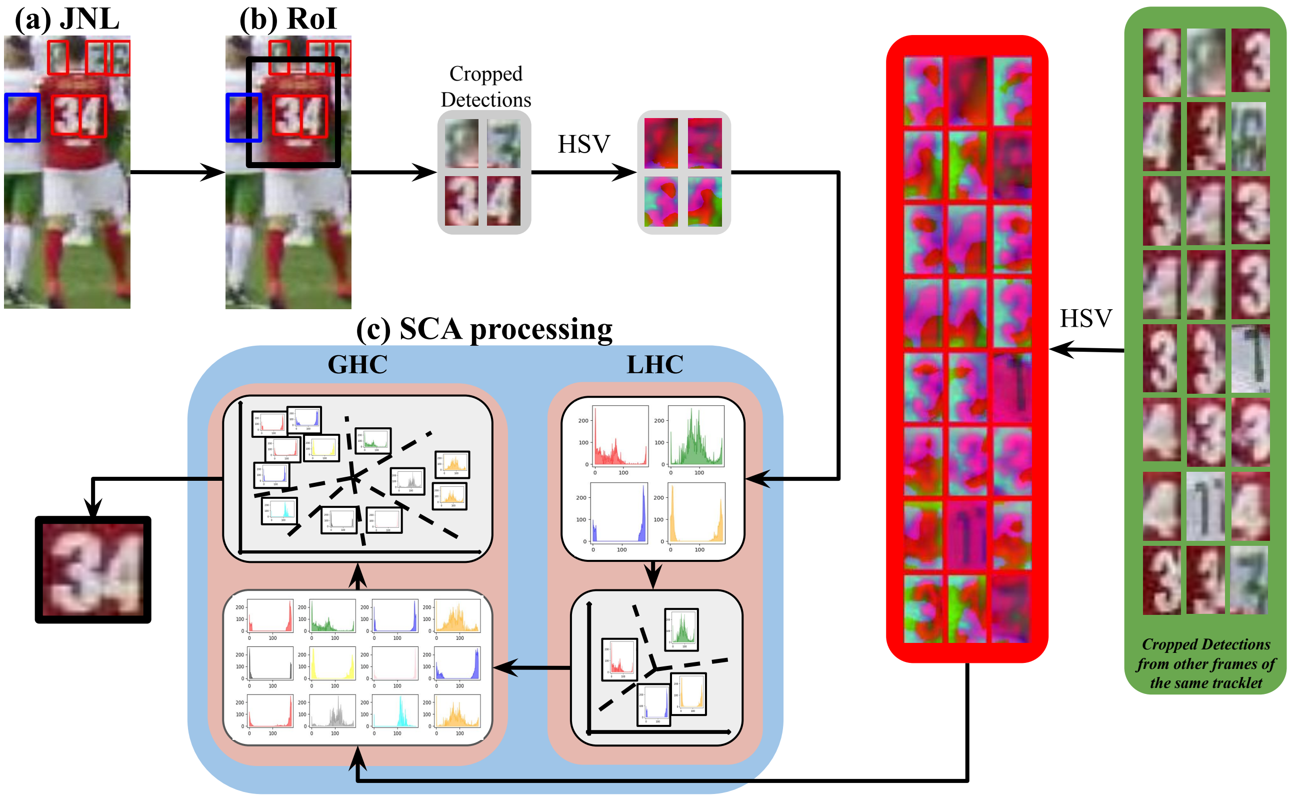
OUR METHOD



The proposed approach comprises several key steps:

1. Keyframe Identification: Each input tracklet is passed through the KfID module which identifies keyframes that contain high-level context of the jersey number, and localizes it.
2. Spatio-temporal Feature Extractor: The extracted frames are then passed through a spatio-temporal neural network that extracts the spatial features $\mathcal{F}_s$ and temporal features $\mathcal{F}_t$ necessary to identify the jersey number reliably.
3. Multi-task Classifier We leverage 2 classification heads to classify each digit $d_{i\in\{1,2\}} \in \mathbb{R}^{11}$ separately from $\mathcal{F}_t$

KfID MODULE



(a) JNL
(b) RoI
Cropped Detections
HSV
HSV
(c) SCA processing
GHC
LHC
Cropped Detections from other frames of the same tracklet

SOCCERNET DATA

- Largest open-source dataset.
- Real-world broadcast videos.
- 2853 tracklets from 550 soccer broadcast videos.
- Low-res with motion blur & occlusions.

Table 1. Comparison of datasets in literature. (†) - Uses temporal data

Dataset	Number of Images
Gerke et al	8,281
Liu et al	3,567
Kanav et al	54,251
Li et al	215, 036
Kanav et al (†)	670,410
SoccerNet (†)	2,052,306

RESULTS

Table 2. Quantitative comparison with the state-of-the-art methods

Method	Test Acc	Challenge Acc
Gerke et al	32.57	35.79
Kanav et al	46.73	49.88
Li et al	47.85	50.60
Kanav et al	52.91	58.45
Ours	**68.53**	**73.77**

IMPACT OF KfID

VIP

Table 3. Results with and without KfId Module. (†) - with the KfId module

Method	Test Acc	Challenge Acc
TCN	27.08	30.17
ViT	19.90	23.78
LSTM	**30.89**	**36.07**
TCN (†)	67.54 (+40.46)	63.81 (+33.64)
ViT (†)	58.62 (+38.72)	65.37 (+41.59)
LSTM (†)	**68.53 (+37.81)**	**73.77 (+37.70)**

Table 4. Dataset split for training, validation, and testing

Dataset	Tracklets	Number of Images	Keyframes
Train	1,141	587,543	68,881
Validation	286	146,886	17,220
Test	1,211	565,758	68,745
Challenge	1,426	750,092	98,504
Total	4,064	2,052,306	253,350

ABLATION STUDY

Table 5. Ablation study on different heads for the loss function

HO	DW	LC	Test Acc
✓			55.71
✓	✓		62.39
✓	✓	✓	65.14
	✓	✓	63.77
	✓		**68.53**

Table 6. Ablation study on different training sequence length

Sequence Length	Test Acc
10	62.82
20	65.45
30	66.52
40	**68.53**
50	67.03
60	65.80

CONCLUSION

- **Efficacy of our KfID Module**: We demonstrate that incorporating our novel keyframe identification module results in a significant 37.81% and 37.70% increase in the accuracies of 2 different test sets with domain gaps.
- **Digit-wise Classification**: We carefully compare the impact of auxiliary tasks such as length prediction, and empirically showcase that digit-wise classification is the best training strategy for unique player identification.
- **Significant Improvement on SOTA**: We consistently outperform the existing state-of-the-art by ~15%, underscoring the impact of bias in existing networks.

ACKNOWLEDGEMENT

Thank You!

Open to any questions