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

Bavesh Balaji, Jerrin Bright, Harish Prakash, Yuhao Chen, David Clausi and John Zelek
University of Waterloo, Waterloo, Ontario, Canada

Key Contributions

- We delve into the **inherent bias in real-world video data**(absence of jersey numbers) and formulate a solution that aims at alleviating this bias.
- We introduce **a keyframe identification module** that is robust to blur and occlusions using **RoI** and **Spatial Context Aware filtering** to facilitate effective jersey number recognition.
- We conduct an extensive study to determine the **best training strategy** for our model by experimenting with **different heads for the loss function**.

Dataset

Table 1: 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

Table 2: Comparison of datasets in literature. (†) **- Uses temporal data**

Dataset	Number of Images
Gerke et al	8,281
Liu et al [1]	3,567
Kanav et al	54,251
Li et al	215, 036
Kanav et al (†)	670,410
SoccerNet (†)	2,052,306

Loss Function

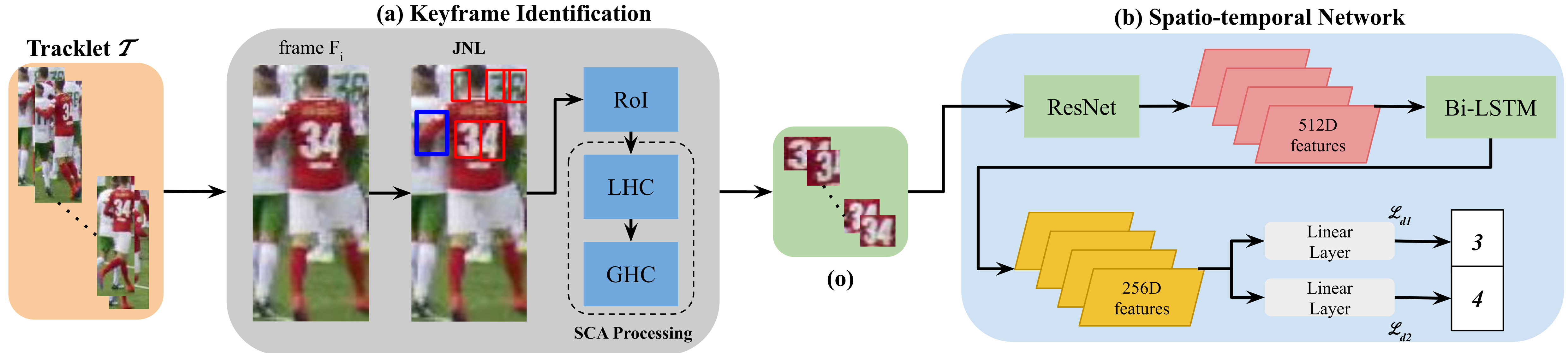
We employ a multi-task loss function for each digit $d_0 \in \mathbb{R}^{11}$ and $d_1 \in \mathbb{R}^{11}$.

$$L_{tot} = 0.5 * L_{d1} + 0.5 * L_{d2} \quad (1)$$

where, L_{d1} and L_{d2} are cross-entropy losses with ground-truth digits p_1 and p_2 respectively.

$$L_{d_1} = -\Sigma_{i=0}^{10} d_1^i log p_1^i \qquad L_{d_2} = -\Sigma_{j=0}^{10} d_2^j log p_2^j \quad (2)$$

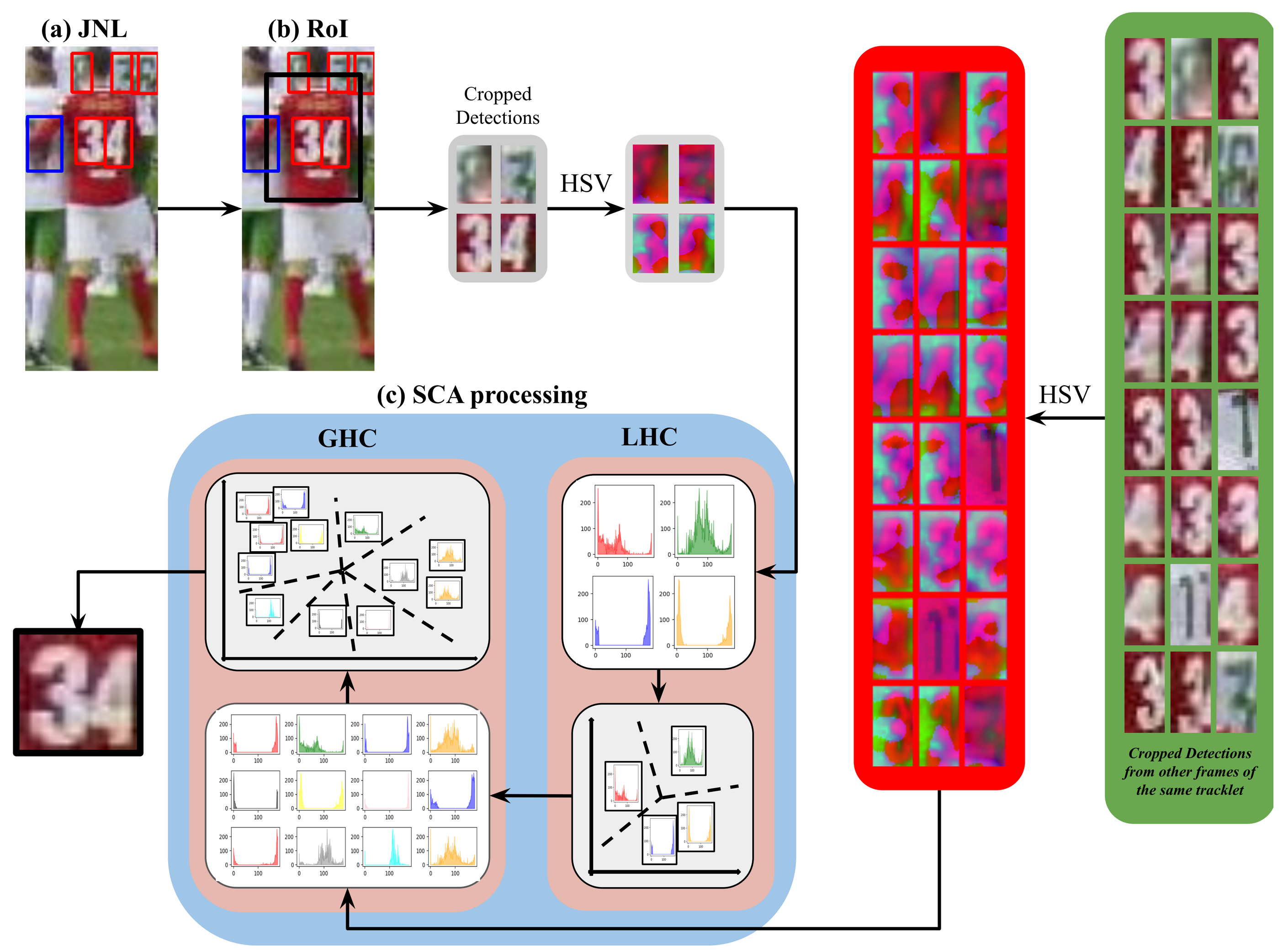
Overall Architecture



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 $\mathcal{F}_t$.

Keyframe Identification



Algorithm 1: KfID Module

Input: Tracklet $\mathcal{T} = \{F_i : F_i \in \mathbb{R}^{H \times W \times 3}\}_{i=1}^{t}$
Output: $KfID(\mathcal{T}) = \mathcal{T} \setminus \{F_{n_1}, F_{n_2}, ..., F_{n_k}\}$ where $\{F_{n_1}, F_{n_2}, ..., F_{n_k}\}$ are noisy and redundant frames.

```
Initialize f_local;
for F_i in T do
    Initialize det_i ;
    h_i ← F_i.height, w_i ← F_i.width ;
    det_i ← JNL(F_i) ;
    Initialize roi_i ← [x_min, y_min, x_max, y_max], f^i_filtered ;
    for d_i in det_i do
        s ← IoU(d_i, roi_i) ;
        if IoU(d_i, roi_i) ≥ a threshold t then
            f^i_filtered.append(d_i)
    f^i_local ← LHC(f^i_filtered) ;
    f_local.append(f^i_local) ;
KfID(T) ← GHC(f_local)
```

Conclusion

1. **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.
2. **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.
3. **Significant Improvement on SOTA:** We consistently outperform the existing state-of-the-art by 15%, underscoring the impact of bias in existing networks.

Results

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

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: Quantitative comparison with the state-of-the-art methods

Method	Test Acc	Challenge Acc
Gerke et al	32.57	35.79
Vats et al [2]	46.73	49.88
Li et al	47.85	50.60
Vats et al [3]	52.91	58.45
Ours	**68.53**	**73.77**

Ablation Study

Table 5: Different heads for the loss function

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

Table 6: Different training sequence length

Seq. Length	Test Acc
20	65.45
30	66.52
40	**68.53**
50	67.03
60	65.80

References

[1] Hengyue Liu and Bir Bhanu. Pose-guided r-cnn for jersey number recognition in sports. *2019 IEEE/CVF CVPRW*, pages 2457–2466, 2019.

[2] Kanav Vats, Mehrnaz Fani, David A Clausi, and John Zelek. Multi-task learning for jersey number recognition in ice hockey. In *Proceedings of the 4th International Workshop on MMSports*, pages 11–15, 2021.

[3] Kanav Vats, William J. McNally, Pascale Walters, David A Clausi, and John S. Zelek. Ice hockey player identification via transformers and weakly supervised learning. *2022 IEEE/CVF CVPRW*, pages 3450–3459, 2021.

Acknowledgement

Our work was supported by Stathletes through the Mitacs Accelerate Program. We also acknowledge the Natural Sciences and Engineering Research Council of Canada for their support.