ReHAR: Robust and Efficient Human Activity Recognition

Xin Li, Mooi Choo Chuah
WACV18’
OUTLINE

• Motivation
• The state-of-the-art scheme
• Our solution
• Evaluations
• Why does it work
• References
OUTLINE

• Motivation

• The state-of-the-art scheme

• Our solution

• Evaluations

• Why does it work

• References
Motivation

Large amount of Videos
Motivation

Public Safety
Motivation

Key events in sport videos
Motivation

An efficient scheme for identifying activities is critically important.
OUTLINE

• Motivation

• The state-of-the-art scheme

• Our solution

• Evaluations

• Why does it work

• References
Building Block

Long Short Term Memory Network (LSTM)

Figure from http://colah.github.io/posts/2015-08-Understanding-LSTMs/
Existing Work

Long-term recurrent convolutional networks for visual recognition and description
CVPR. 2015
Existing Work

A Hierarchical Deep Temporal Model for Group Activity Recognition
CVPR. 2016
Existing Work

[3] Xin Li, Mooi Choo Chuah
SBGAR: Semantics Based Group Activity Recognition
ICCV. 2017
OUTLINE

• Motivation

• The state-of-the-art scheme

• Our solution

• Evaluations

• Why does it work

• References
Activity Recognition

Our Solution

Video Frame RGB

Preprocess

Single Frame Representation Model

Activity Recognition Model

Video Frame RGB

Preprocess

Single Frame Representation Model

Activity Recognition Model

Video Frame RGB

Preprocess

Single Frame Representation Model

Activity Recognition Model

Video Frame RGB

Preprocess

Single Frame Representation Model

Activity Recognition Model
Activity Recognition

Our Solution

![Diagram of activity recognition model](image)

\[ \text{Loss} = \left( \sum_{t=1}^{\text{time_step}} \text{loss}_{1,t} \right) + \lambda \times \text{loss}_2 \]
OUTLINE

• Motivation

• The state-of-the-art scheme

• Our solution

• Evaluations

• Why does it work

• References
Dataset1: NCAA Basketball Dataset

NCAA Basketball dataset:
- **11436** training videos
- **856** validation videos
- **2256** testing videos

<table>
<thead>
<tr>
<th>Event</th>
<th>No. of videos Train (Test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-point succ.</td>
<td>895 (188)</td>
</tr>
<tr>
<td>3-point fail.</td>
<td>1934 (401)</td>
</tr>
<tr>
<td>free-throw succ.</td>
<td>552 (94)</td>
</tr>
<tr>
<td>free-throw fail.</td>
<td>344 (41)</td>
</tr>
<tr>
<td>layup succ.</td>
<td>1212 (233)</td>
</tr>
<tr>
<td>layup fail.</td>
<td>1286 (254)</td>
</tr>
<tr>
<td>2-point succ.</td>
<td>1039 (148)</td>
</tr>
<tr>
<td>2-point fail.</td>
<td>2014 (421)</td>
</tr>
<tr>
<td>slam dunk succ.</td>
<td>286 (54)</td>
</tr>
<tr>
<td>slam dunk fail.</td>
<td>47 (5)</td>
</tr>
<tr>
<td>steal</td>
<td>1827 (417)</td>
</tr>
</tbody>
</table>
### Test Result using NCAA Basketball Dataset

<table>
<thead>
<tr>
<th></th>
<th>3point S.</th>
<th>3point F.</th>
<th>throw S.</th>
<th>throw F.</th>
<th>layup S.</th>
<th>layup F.</th>
<th>2point S.</th>
<th>2point F.</th>
<th>dunk S.</th>
<th>dunk F.</th>
<th>steal</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDT[4]</td>
<td>0.370</td>
<td>0.501</td>
<td>0.778</td>
<td>0.365</td>
<td>0.283</td>
<td>0.278</td>
<td>0.136</td>
<td>0.303</td>
<td>0.197</td>
<td>0.004</td>
<td>0.555</td>
<td>0.343</td>
</tr>
<tr>
<td>IDT[4] player</td>
<td>0.428</td>
<td>0.481</td>
<td>0.703</td>
<td>0.623</td>
<td>0.300</td>
<td>0.311</td>
<td>0.233</td>
<td>0.285</td>
<td>0.171</td>
<td>0.010</td>
<td>0.473</td>
<td>0.365</td>
</tr>
<tr>
<td>C3D[5]</td>
<td>0.117</td>
<td>0.282</td>
<td>0.642</td>
<td>0.319</td>
<td>0.195</td>
<td>0.185</td>
<td>0.078</td>
<td>0.254</td>
<td>0.047</td>
<td>0.004</td>
<td>0.303</td>
<td>0.221</td>
</tr>
<tr>
<td>MIL[6]</td>
<td>0.237</td>
<td>0.335</td>
<td>0.597</td>
<td>0.318</td>
<td>0.257</td>
<td>0.247</td>
<td>0.224</td>
<td>0.299</td>
<td>0.112</td>
<td>0.005</td>
<td>0.843</td>
<td>0.316</td>
</tr>
<tr>
<td>LRCN[7]</td>
<td>0.462</td>
<td>0.564</td>
<td>0.876</td>
<td>0.584</td>
<td>0.463</td>
<td>0.386</td>
<td>0.257</td>
<td>0.378</td>
<td>0.285</td>
<td><strong>0.027</strong></td>
<td>0.876</td>
<td>0.469</td>
</tr>
<tr>
<td>Atten. no track[8]</td>
<td>0.583</td>
<td>0.668</td>
<td>0.892</td>
<td>0.671</td>
<td>0.489</td>
<td>0.426</td>
<td>0.281</td>
<td>0.442</td>
<td>0.210</td>
<td>0.006</td>
<td>0.886</td>
<td>0.505</td>
</tr>
<tr>
<td>Atten. track[8]</td>
<td>0.600</td>
<td>0.738</td>
<td>0.882</td>
<td>0.516</td>
<td>0.500</td>
<td><strong>0.445</strong></td>
<td>0.341</td>
<td><strong>0.291</strong></td>
<td>0.471</td>
<td>0.232</td>
<td>0.007</td>
<td>0.940</td>
</tr>
<tr>
<td>Ours</td>
<td><strong>0.753</strong></td>
<td><strong>0.766</strong></td>
<td><strong>0.933</strong></td>
<td><strong>0.857</strong></td>
<td><strong>0.613</strong></td>
<td>0.435</td>
<td><strong>0.405</strong></td>
<td><strong>0.542</strong></td>
<td>0.232</td>
<td>0.007</td>
<td>0.940</td>
<td>0.589</td>
</tr>
</tbody>
</table>

---


Test Result using NCAA Basketball Dataset

<table>
<thead>
<tr>
<th></th>
<th>3point S</th>
<th>3point F.</th>
<th>throw S</th>
<th>throw F.</th>
<th>layup S</th>
<th>laysup F.</th>
<th>2point S</th>
<th>2point F.</th>
<th>dunk S</th>
<th>dunk F.</th>
<th>steal</th>
</tr>
</thead>
<tbody>
<tr>
<td>61.17</td>
<td>11.17</td>
<td>0.53</td>
<td>0.00</td>
<td>2.13</td>
<td>1.60</td>
<td>18.09</td>
<td>4.79</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.53</td>
</tr>
<tr>
<td>1.75</td>
<td>72.07</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>23.19</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1.06</td>
<td>0.00</td>
<td>87.23</td>
<td>6.38</td>
<td>3.19</td>
<td>0.00</td>
<td>1.06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>4.88</td>
<td>17.07</td>
<td>75.61</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>2.44</td>
<td></td>
</tr>
<tr>
<td>2.58</td>
<td>1.29</td>
<td>0.43</td>
<td>0.00</td>
<td>59.66</td>
<td>12.02</td>
<td>15.88</td>
<td>6.01</td>
<td>1.72</td>
<td>0.00</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>3.94</td>
<td>0.00</td>
<td>0.00</td>
<td>8.66</td>
<td>47.64</td>
<td>1.57</td>
<td>35.83</td>
<td>0.00</td>
<td>0.00</td>
<td>2.36</td>
<td></td>
</tr>
<tr>
<td>12.16</td>
<td>4.05</td>
<td>0.68</td>
<td>0.00</td>
<td>31.08</td>
<td>6.08</td>
<td>36.49</td>
<td>7.43</td>
<td>0.00</td>
<td>0.00</td>
<td>2.03</td>
<td></td>
</tr>
<tr>
<td>1.66</td>
<td>16.86</td>
<td>0.00</td>
<td>0.24</td>
<td>1.19</td>
<td>17.10</td>
<td>0.95</td>
<td>58.43</td>
<td>0.00</td>
<td>0.00</td>
<td>3.56</td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.85</td>
<td>53.70</td>
<td>24.07</td>
<td>9.26</td>
<td>3.70</td>
<td>5.56</td>
<td>0.00</td>
<td>1.85</td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>60.00</td>
<td>0.00</td>
<td>0.00</td>
<td>20.00</td>
<td>0.00</td>
<td>0.00</td>
<td>20.00</td>
</tr>
<tr>
<td>0.00</td>
<td>4.32</td>
<td>0.00</td>
<td>0.24</td>
<td>1.92</td>
<td>5.04</td>
<td>0.00</td>
<td>6.24</td>
<td>0.00</td>
<td>0.00</td>
<td>82.25</td>
<td></td>
</tr>
</tbody>
</table>
Test Result using NCAA Basketball Dataset

Group 10 shooting-related actions (except “steal”) into 2 categories (success or failure)

<table>
<thead>
<tr>
<th>Number of testing videos</th>
<th>3-point</th>
<th>free-throw</th>
<th>layup</th>
<th>2-point</th>
<th>slam dunk</th>
<th>In total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success</td>
<td>188</td>
<td>94</td>
<td>233</td>
<td>148</td>
<td>54</td>
<td>717</td>
</tr>
<tr>
<td>Failure</td>
<td>401</td>
<td>41</td>
<td>254</td>
<td>421</td>
<td>5</td>
<td>1122</td>
</tr>
</tbody>
</table>

88% testing samples are correctly labeled into “Success” or “Failure” categories.
Dataset 2: UCF Sports Action Dataset

UCF Sports dataset:

- 103 training videos
- 47 testing videos

10 different sports categories
- Diving
- Golf
- Kicking
- Lifting
- Riding
- Run
- SkateBoarding
- Swing-Bench
- Swing-Side
- Walk
## Test Result UCF Sports Action Dataset

<table>
<thead>
<tr>
<th></th>
<th>Diving</th>
<th>Golf</th>
<th>Kicking</th>
<th>Lifting</th>
<th>Riding</th>
<th>Run</th>
<th>SkateB</th>
<th>Swing</th>
<th>SwingB</th>
<th>Walk</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gkioxari et al. [9]</td>
<td>0.758</td>
<td>0.693</td>
<td>0.546</td>
<td>0.991</td>
<td>0.896</td>
<td>0.549</td>
<td>0.298</td>
<td>0.887</td>
<td>0.745</td>
<td>0.447</td>
<td>0.681</td>
</tr>
<tr>
<td>Weinzaepfel et al. [10]</td>
<td>0.607</td>
<td>0.776</td>
<td>0.653</td>
<td>1.000</td>
<td>0.995</td>
<td>0.526</td>
<td>0.471</td>
<td>0.889</td>
<td>0.629</td>
<td>0.644</td>
<td>0.719</td>
</tr>
<tr>
<td>Peng et al. [11]</td>
<td>0.961</td>
<td>0.805</td>
<td>0.735</td>
<td>0.992</td>
<td>0.976</td>
<td>0.824</td>
<td>0.574</td>
<td>0.836</td>
<td>0.985</td>
<td>0.760</td>
<td>0.845</td>
</tr>
<tr>
<td>Hou et al. [12]</td>
<td>0.844</td>
<td>0.908</td>
<td>0.865</td>
<td>0.998</td>
<td>1.000</td>
<td>0.837</td>
<td>0.687</td>
<td>0.658</td>
<td>0.996</td>
<td>0.878</td>
<td>0.867</td>
</tr>
<tr>
<td><strong>Ours</strong></td>
<td><strong>1.000</strong></td>
<td><strong>0.955</strong></td>
<td><strong>1.000</strong></td>
<td><strong>1.000</strong></td>
<td><strong>1.000</strong></td>
<td><strong>0.806</strong></td>
<td><strong>0.626</strong></td>
<td><strong>1.000</strong></td>
<td><strong>1.000</strong></td>
<td><strong>0.888</strong></td>
<td><strong>0.928</strong></td>
</tr>
</tbody>
</table>


Test Result using UCF Sports Action Dataset

<table>
<thead>
<tr>
<th></th>
<th>Driving</th>
<th>Golf</th>
<th>Kicking</th>
<th>Lifting</th>
<th>Riding</th>
<th>Run</th>
<th>SkateB.</th>
<th>Swing</th>
<th>SwingB.</th>
<th>Walk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Driving</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Golf</td>
<td>0.00</td>
<td>83.33</td>
<td>16.67</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Kicking</td>
<td>0.00</td>
<td>0.00</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Lifting</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Riding</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>75.00</td>
<td>0.00</td>
<td>25.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Run</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>SkateB.</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>100.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Swing</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>57.14</td>
</tr>
<tr>
<td>SwingB.</td>
<td>0.00</td>
<td>14.29</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>28.57</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>CNN base net</td>
<td>Time on 10 Frames (ms)</td>
<td>Time on 24 Frames (ms)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>------------------------</td>
<td>------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VGG16</td>
<td>103.65</td>
<td>239.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>InceptionV3</td>
<td>78.40</td>
<td>192.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SBGAR [3]** model using

**InceptionV3** as feature extractor and

**10 input frames** was

**108.53 ms.**
OUTLINE

• Motivation

• The state-of-the-art scheme

• Our solution

• Evaluations

• Why does it work

• References
Why does our model work?

On UCF Sports Dataset

| Our Model | 0.928 |
Why does our model work?

(a) Correctly predict an “other 2-pointer success” event on Basketball Dataset.

(b) Correctly predict a “Steal Success” event on Basketball Dataset.
Why does our model work?

(c) Correctly predict a “Kicking” event on UCF Sports Action Dataset.

(d) Incorrectly predict a “Walking” event as “Golf” on UCF Sports Action Dataset.
Reference


