Live Repetition Counting

Ofir Levy, Lior Wolf
Tel Aviv University

Presented by: Aayush Jung Rana
Outline

• Motivation
• Proposed methods
• Experiments
• Results and conclusion
Motivation
Motivation

• Use CNN to detect length of an action
Motivation

• Use CNN to detect length of an action
• Count instances of repetitive motion
Motivation

- Use CNN to detect length of an action
- Count instances of repetitive motion
- Automatic segmentation of such activity
Motivation

• Train on synthetic data
• Real-time system
Overall System

\[ \hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n \]

Outer System
CNN architecture

Input Frames: 1@50x50x20

C1: 40@46x46
M2: 40@23x23
C3: 60@21x21
M4: 60@10x10
C5: 90@8x8
M6: 90@4x4

H7: 500
O8: 8

Max Pooling
Max Pooling
Max Pooling
Max Pooling

Fully connected
Softmax

[3, 4, 5, 6, 7, 8, 9, 10]
Integrating counter

\[ \hat{\gamma} \cdot \gamma \]

\[ p \rightarrow H(p) = -\sum_i p_i \log p_i \]

- \( \hat{\gamma} \): predicted cycle length
- \( p \): vector of 8 probability score
Outer system

- $\hat{y}$: predicted cycle length
- $p$: vector of 8 prediction scores
- $R$: no. of repetition till that frame
- $C$: no. of frames since last $R$ update
Start counting

• Average entropy of last 7 blocks

$$\frac{\sum_{i=n-6}^{n} H(p_i)}{7}$$
Start counting

- Average entropy of last 7 blocks
- If average < threshold
  - $R = \text{floor} \left( \frac{20}{\hat{y}} \right)$
  - $C = 20 \mod \hat{y}$
Update counting

• For each valid block
  • $C = C + 1$
• If $C \geq \text{median(last 4 } \hat{y} \text{)}$
  • $R = R + 1$
  • $C = 0$
End counting

- Average entropy of last 7 blocks
- End if average > threshold
Multiple detectors

• Action can be rapid, moderate, or slow
• Three detectors
  • Sampling interval [2, 5, 8]
Offline mode counting

• For all detectors
  • Run on whole video
  • Find average entropy
• Pick detector with lowest average entropy
• Use count from that detector
Online mode counting

• Use sliding window of all detectors
• Every 40 frame
  • Find average entropy of all detectors
  • If lowest average < threshold
    • Update global counter
ROI detection

- For each pixel
- Std. dev. in time
- Threshold at mean

20 frame sequence
ROI detection

Binary map (E)

10x10 Convolution kernel
### ROI detection

<table>
<thead>
<tr>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Binary map (E_2)**

[Image of the binary map (E_2)]
ROI detection

• Sort indices of positive pixels from $(E_2)$
• If no boundary detected, use whole frame
  • Also for moving camera
ROI detection
Dataset

• Synthetic (Training)
• Patterns on white noise background
• Cycle length 3-10
• 4 different pattern types (based on motion)
Synthetic data
Training CNN

• 30,000 training sequences (synthetic)
• 5,000 validation sequences (synthetic)
• 102 epochs on NVIDIA GTX580 (8.5% validation error)
Benchmark

• 100 YouTube videos selected
  • Exercising, cooking, building, etc.
• Main metric: absolute difference in counts
  • $100 \times \frac{|G-R|}{G}$
Benchmark – Online mode

• 25 YouTube videos with non-repetitive segments
  • Before and after action

• 1 hour long videos
  • 44 different actions
  • Uninterrupted
## Results

<table>
<thead>
<tr>
<th></th>
<th>Youtube benchmark in reverse</th>
<th></th>
<th>Mov. rev</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full</td>
<td>no ROI</td>
<td>regression</td>
</tr>
<tr>
<td>3.52±1.1</td>
<td>3.89±1.2</td>
<td>11.47±1.7</td>
<td>3.94±0.7</td>
</tr>
<tr>
<td>6.97±1.3</td>
<td>8.16±1.9</td>
<td>N/A</td>
<td>5.83±1.0</td>
</tr>
<tr>
<td>7.36±1.8</td>
<td>8.77±2.3</td>
<td>N/A</td>
<td>7.04±1.0</td>
</tr>
<tr>
<td>22.4±3.9</td>
<td>23.2±4.1</td>
<td>31.36±3.4</td>
<td>23.9±4.0</td>
</tr>
</tbody>
</table>
Results
Results
Results
Results
Thank You