Live Repetition Counting

Ofir Levy, Lior Wolf

Tel Aviv University

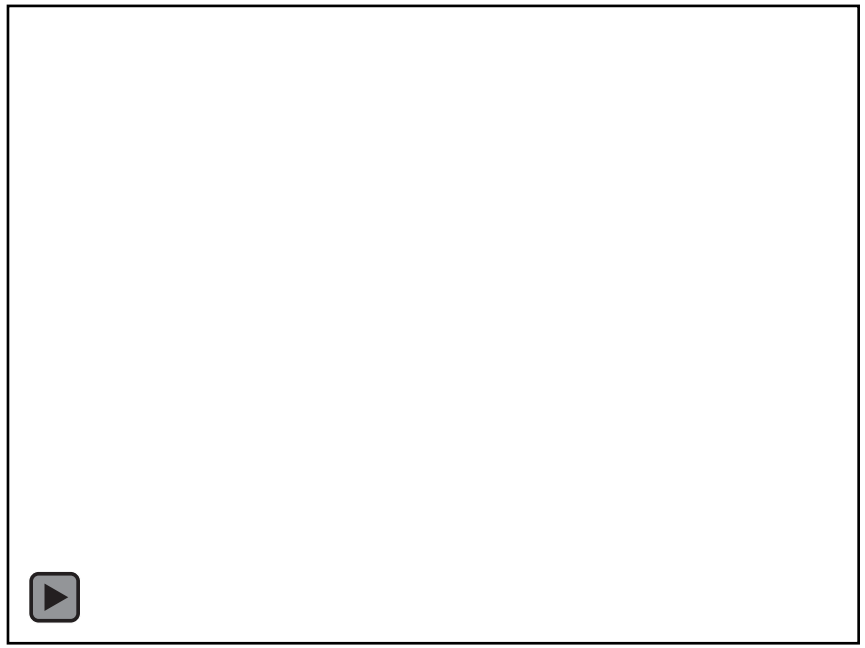
Presented by: Aayush Jung Rana

Outline

- Motivation
- Proposed methods
- Experiments
- Results and conclusion

Motivation



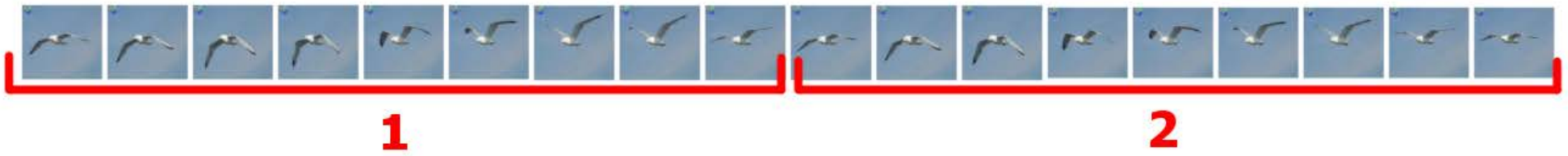


Motivation



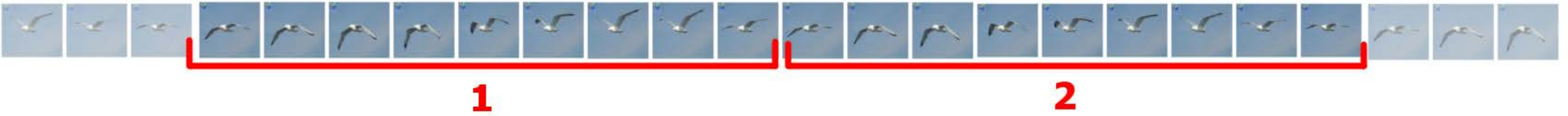
- Use CNN to detect length of an action

Motivation



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

Motivation

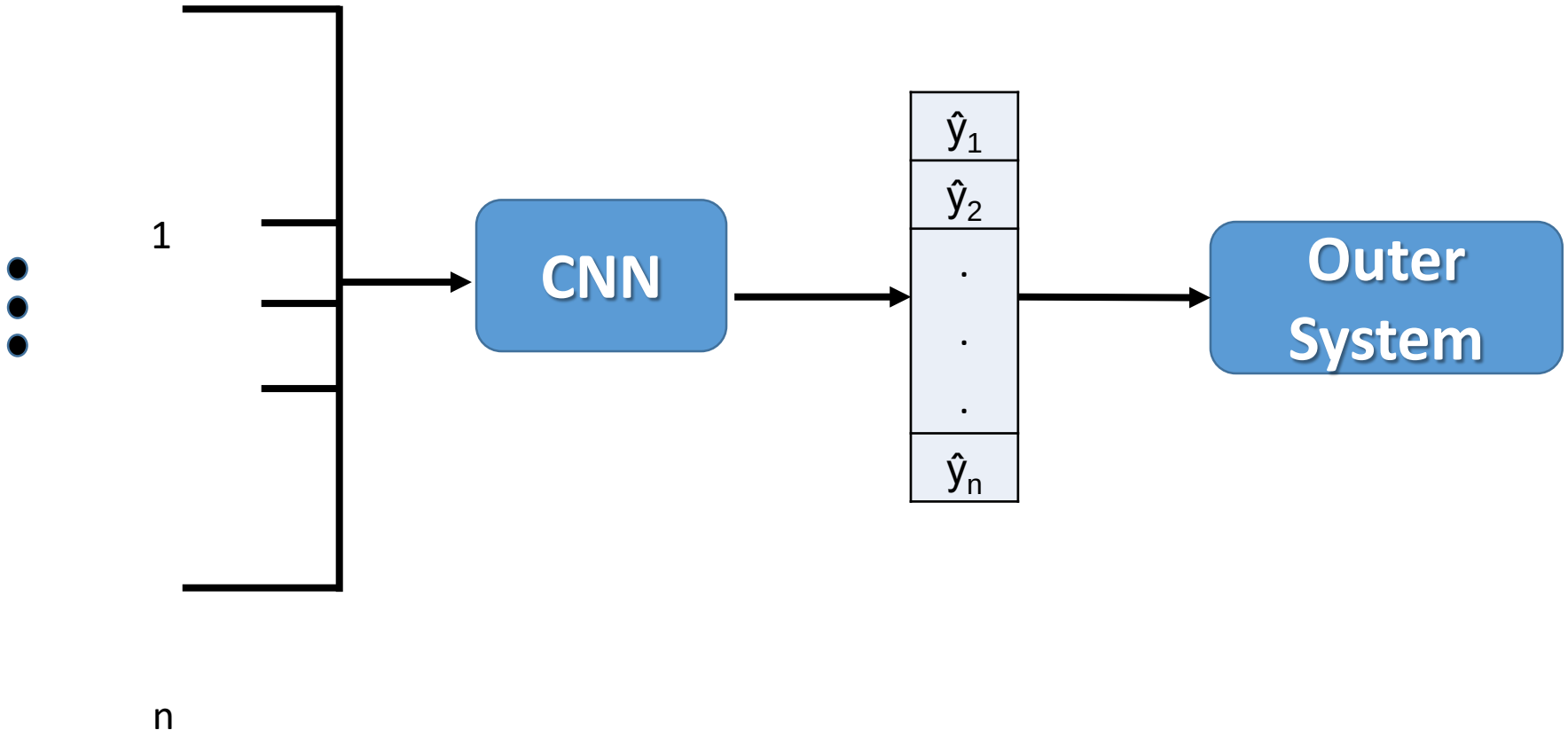


- Use CNN to detect length of an action
- Count instances of repetitive motion
- Automatic segmentation of such activity

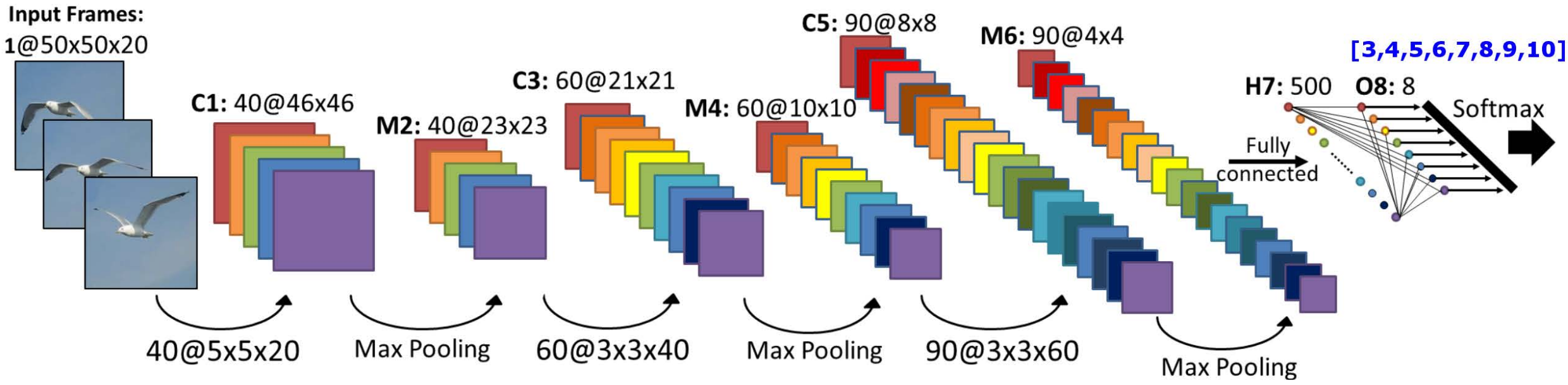
Motivation

- Train on synthetic data
- Real-time system

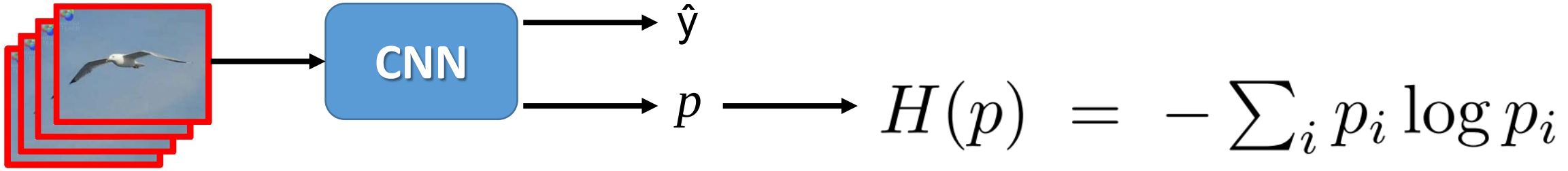
Overall System



CNN architecture



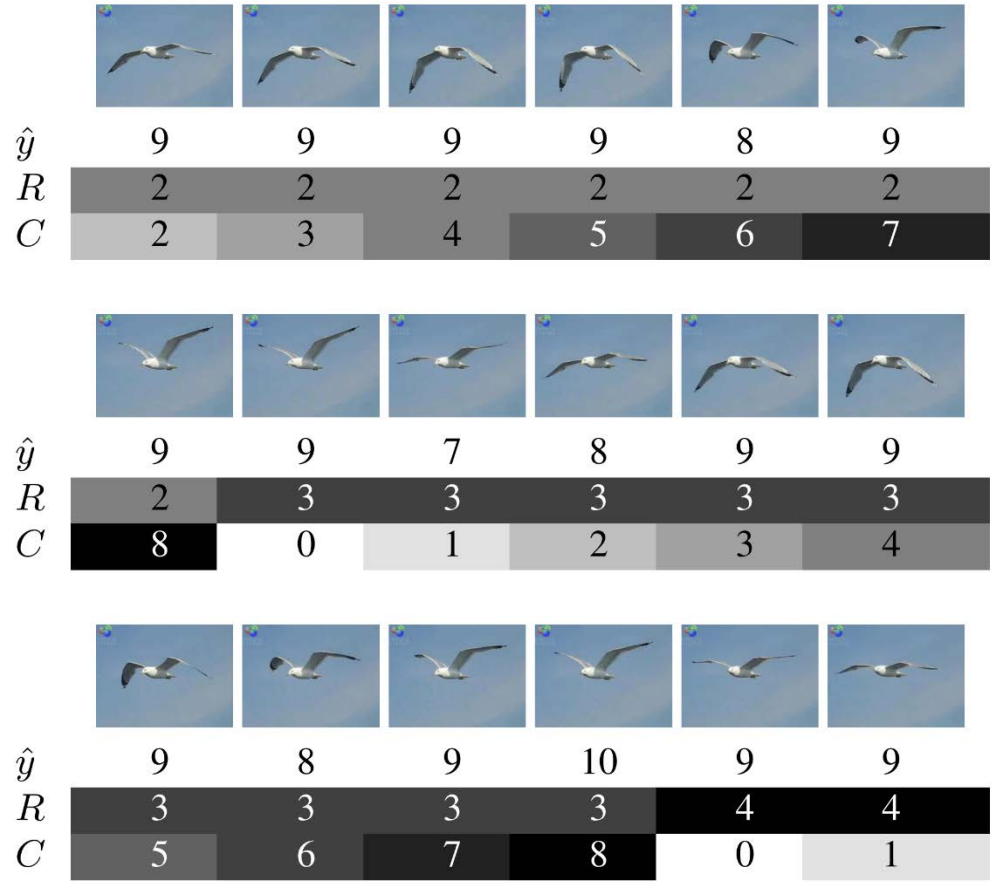
Integrating counter



- ŷ : 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



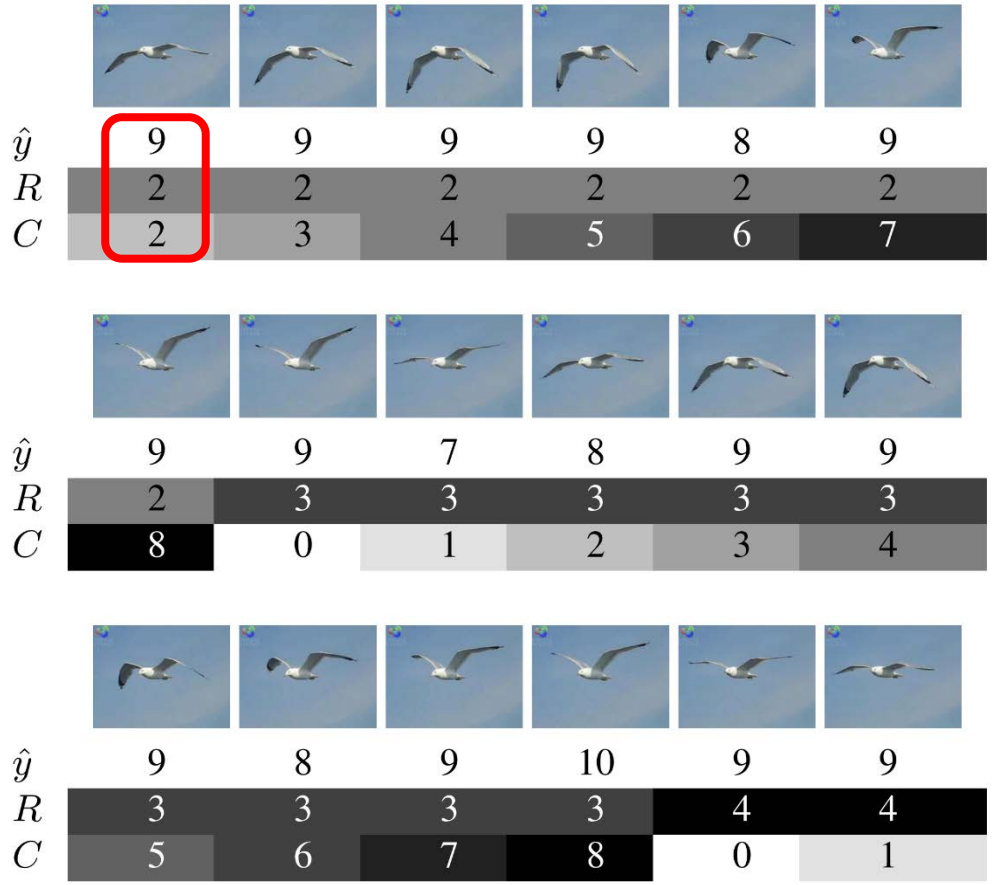
Start counting

- Average entropy of last 7 blocks

$$\frac{\sum_{i=n-6}^{n} \mathrm{H(pi)}}{7}$$

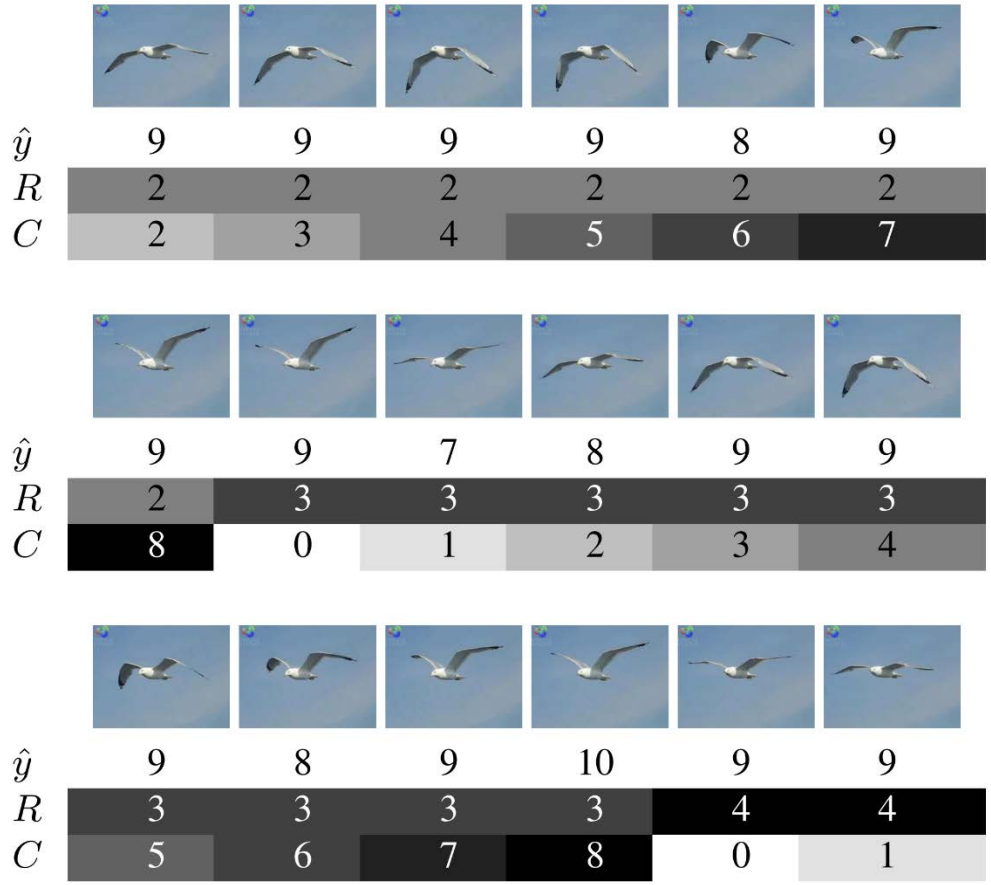
Start counting

- Average entropy of last 7 blocks
- If average < threshold
  - R = floor (20 / ŷ)
  - C = 20 mod ŷ



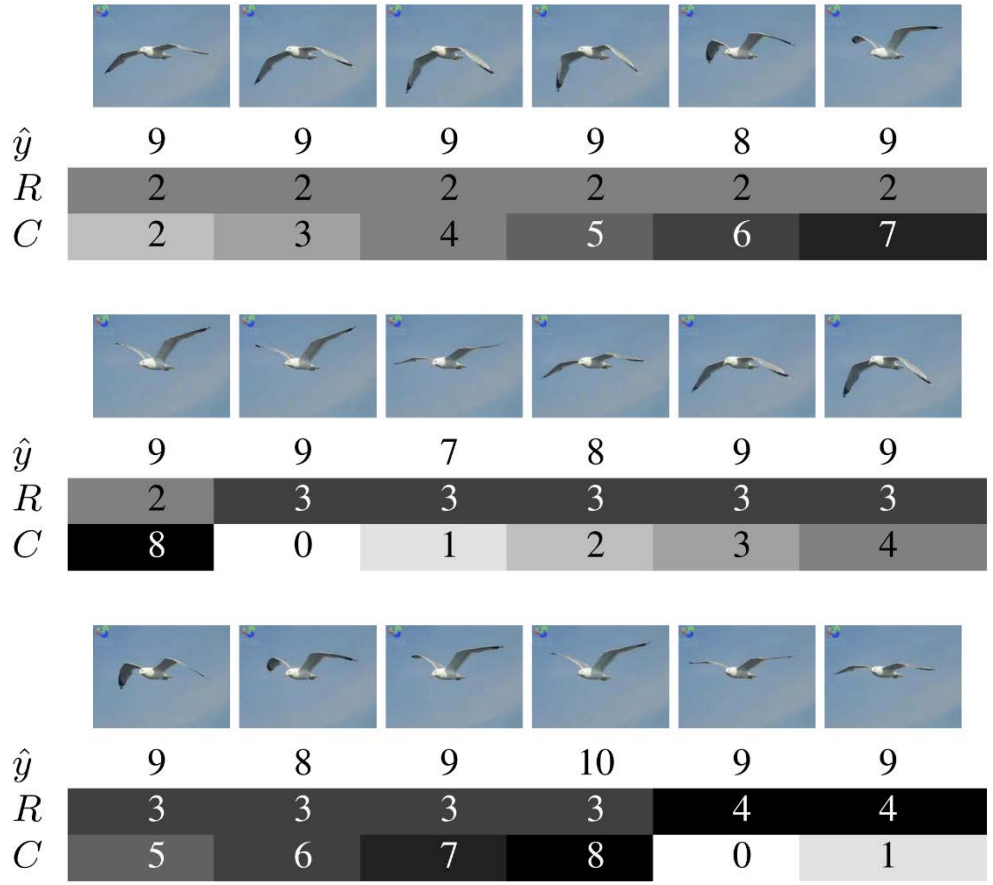
Update counting

- For each valid block
  - C = C + 1
- If C ≥ median(last 4 ŷ)
  - 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]



N = 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

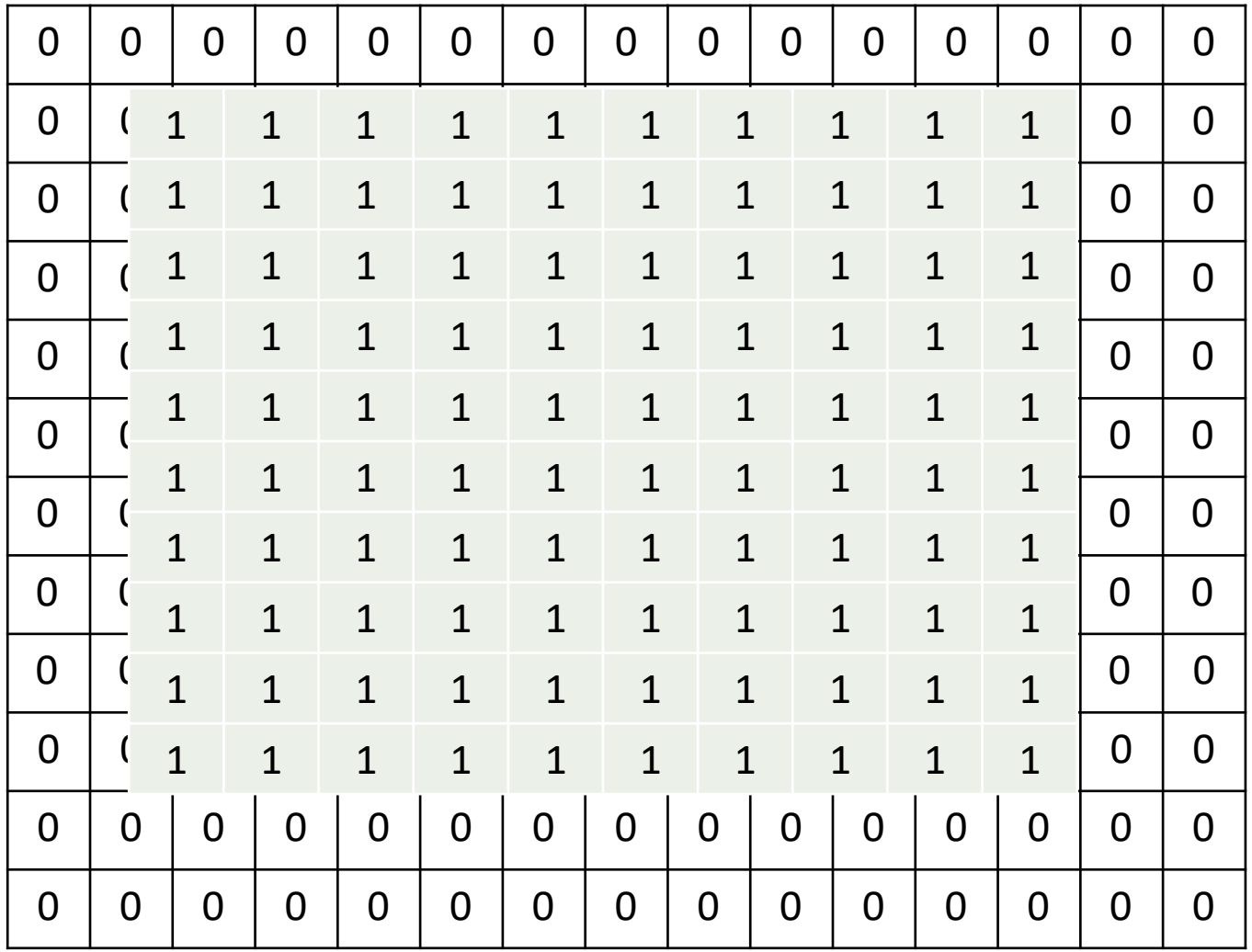
ROI detection

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



20 frame sequence

ROI detection



10x10 Convolution kernel

Binary map (E)

ROI detection

Binary map (E_2)

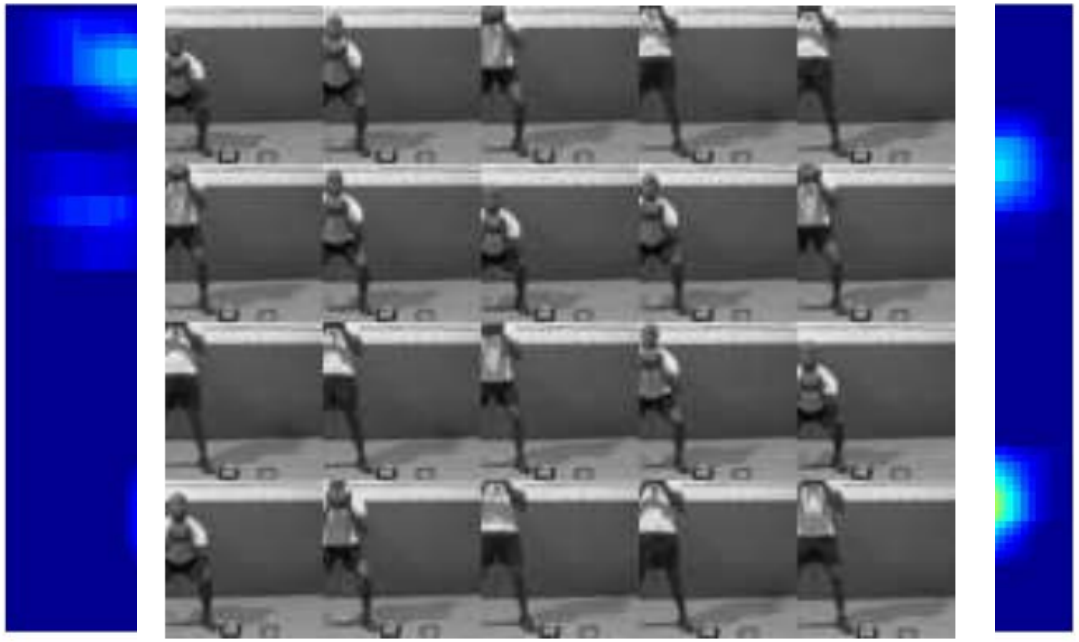
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0
0	0	**1**	**1**	0	0	0	0	0	0	**1**	0	0	0	0
0	0	0	**1**	**1**	**1**	0	0	0	**1**	0	0	0	0	0
0	0	0	0	**1**	**1**	**1**	**1**	**1**	0	0	0	0	0	0
0	0	0	0	0	**1**	**1**	**1**	0	0	0	0	0	0	0
0	0	0	0	0	0	**1**	**1**	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

ROI detection

- Sort indices of positive pixels from (E_2)
- If no boundary detected, use whole frame
  - Also for moving camera

ROI detection

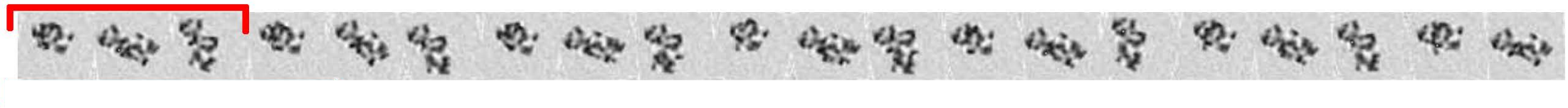




Dataset

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

Synthetic data

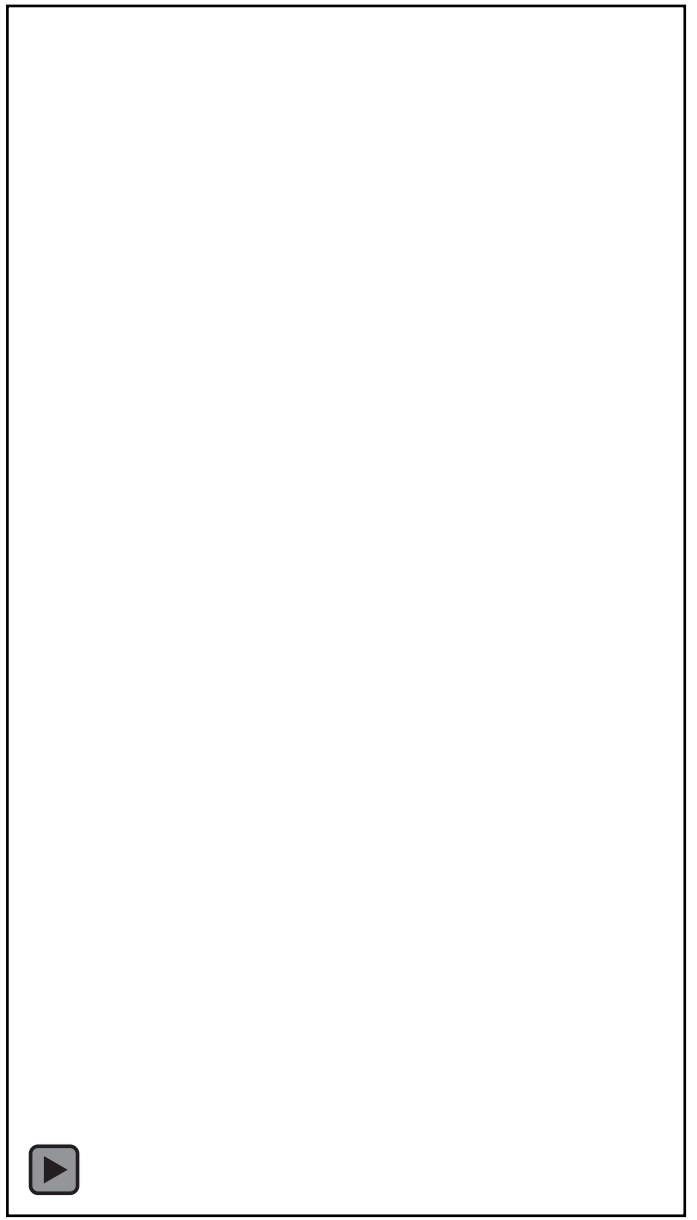


Training CNN

- 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
  - $100 * \frac{|G-R|}{G}$



Benchmark – Online mode

- 25 YouTube videos with non-repetitive segments
  - Before and after action
- 1 hour long videos
  - 44 different actions
  - Uninterrupted

Results

Youtube benchmark in reverse			Mov. rev
Full	no ROI	regression	Full
3.52±1.1	3.89±1.2	11.47±1.7	3.94±0.7
6.97±1.3	8.16±1.9	N/A	5.83±1.0
7.36±1.8	8.77±2.3	N/A	7.04±1.0
22.4±3.9	23.2±4.1	31.36±3.4	23.9±4.0

Results

Results



Results



Results



Thank You