Sum-Product Networks for Modeling Activities with Stochastic Structure

Oregon State UNIVERSITY OSU

Mohamed R. Amer

Sinisa Todorovic

CVPR '12
Providence, Rhode Island
June 16 - 21, 2012

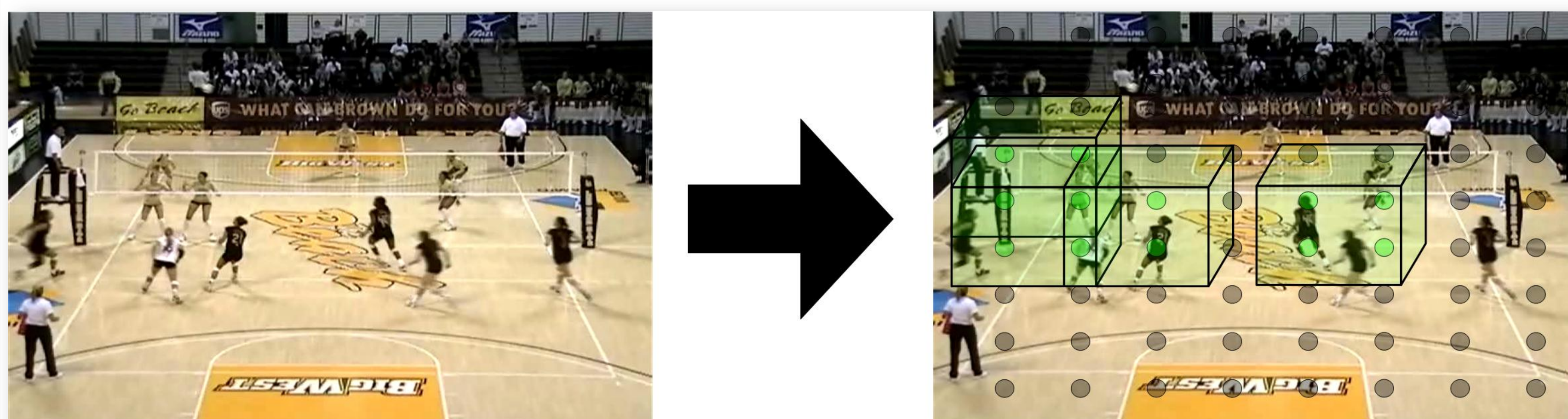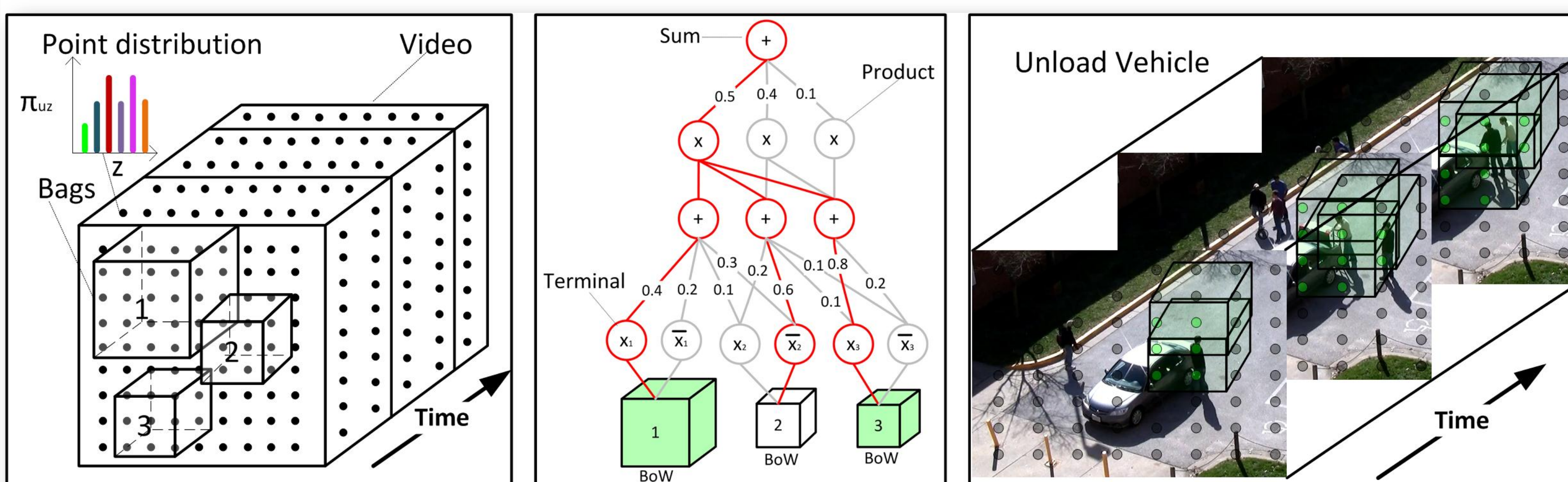## Problem Statement

Given Input Video → Classify & Localize

Stochastic activity – a random number of actors, parts, and configurations

Our Approach

Initial layout of bags — SPN — Classification & Localization

Probabilistic modeling of hierarchical and alternative configurations, grounded onto Bag of Words (BoW)

Contributions

- Activity representation using SPN [1]:
  - Product node = particular configuration
  - Sum node = alternative configuration
  - Terminal node = BoW

- New Volleyball Dataset

Low Level – Counting Grid

Posterior probability of BoW on a Counting Grid [2]:

$$P_{X_b|c_b} \propto \prod_z \left[\sum_{u \in H_b} \pi_{uz} \right]^{c_{bz} + \theta_z - 1}$$

Model – SPN

SPN:

$$S(C) = 0.5(0.4 x_1 P_1 + 0.2 \overline{x_1}(1 - P_1) + \cdots$$

Posterior:

$$P(X|C) = S(C)/S_{X=1}$$

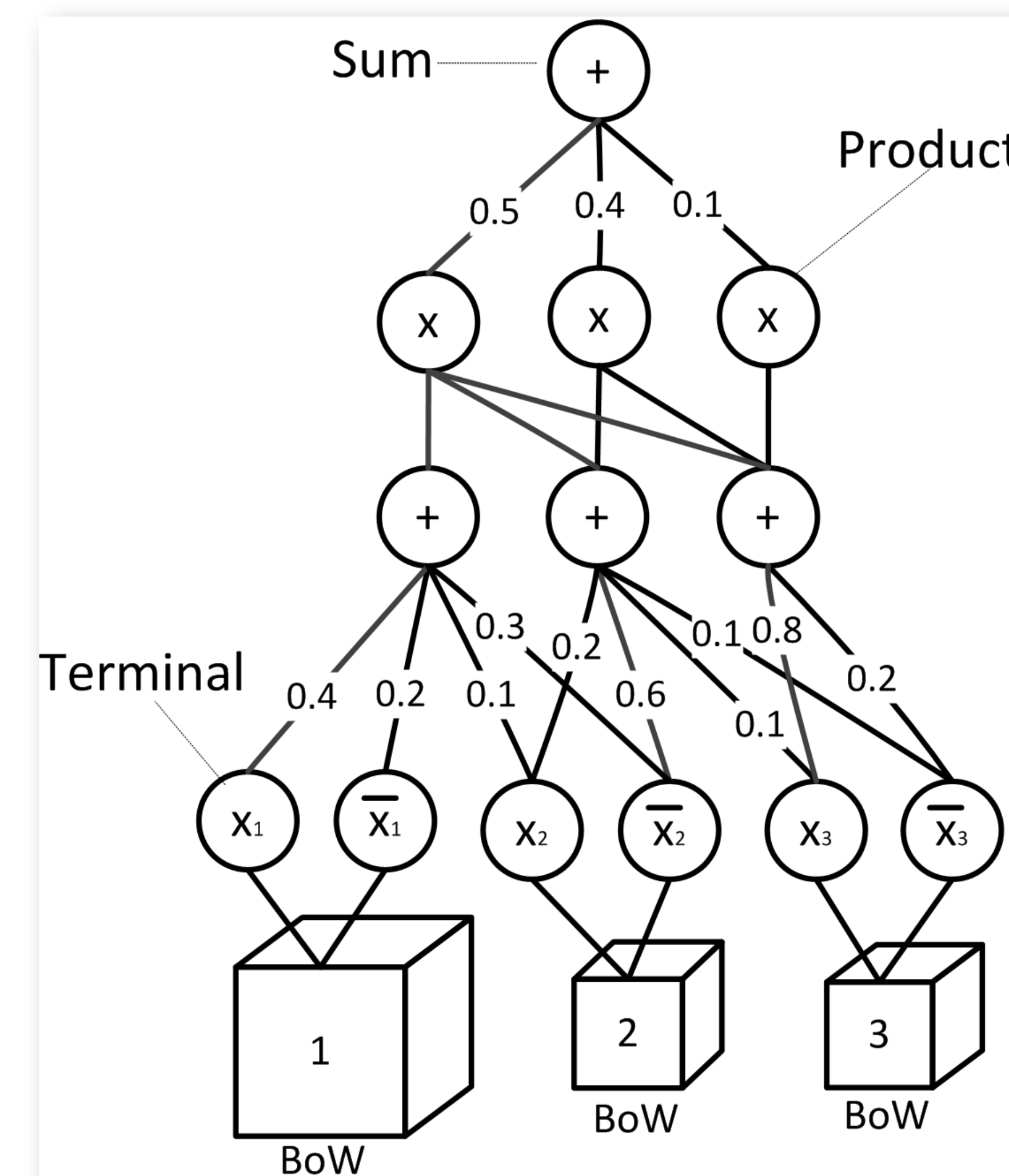**Sum Nodes:**

$$S_i(C) = \sum_{j \in i^+} w_{ij} S_j(C)$$

Product Nodes:
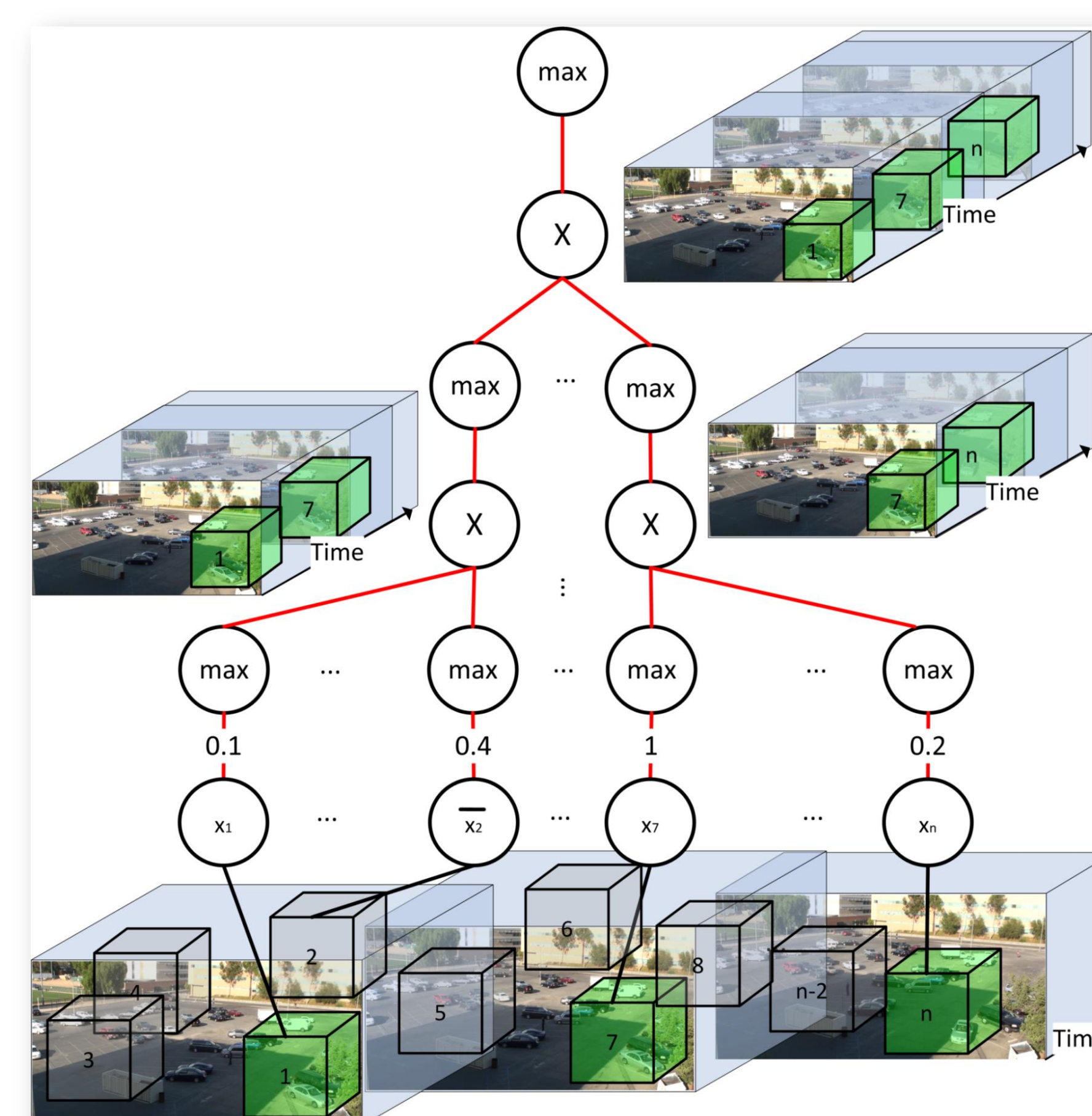
$$S_k(C) = \prod_{l \in k^+} S_l(C)$$

Terminal Nodes:

$$S_i(C) = \sum_{b \in i^+} \left[w_{ib1} x_b P_{X_b|c_b} + w_{ib2} \overline{x_b} \left(1 - P_{X_b|c_b} \right) \right]$$

$$x_b \in \{0,1\}, \overline{x_b} \in \{0,1\}, b = 1,2 \ldots$$

Inference – Bottom up/Top Down
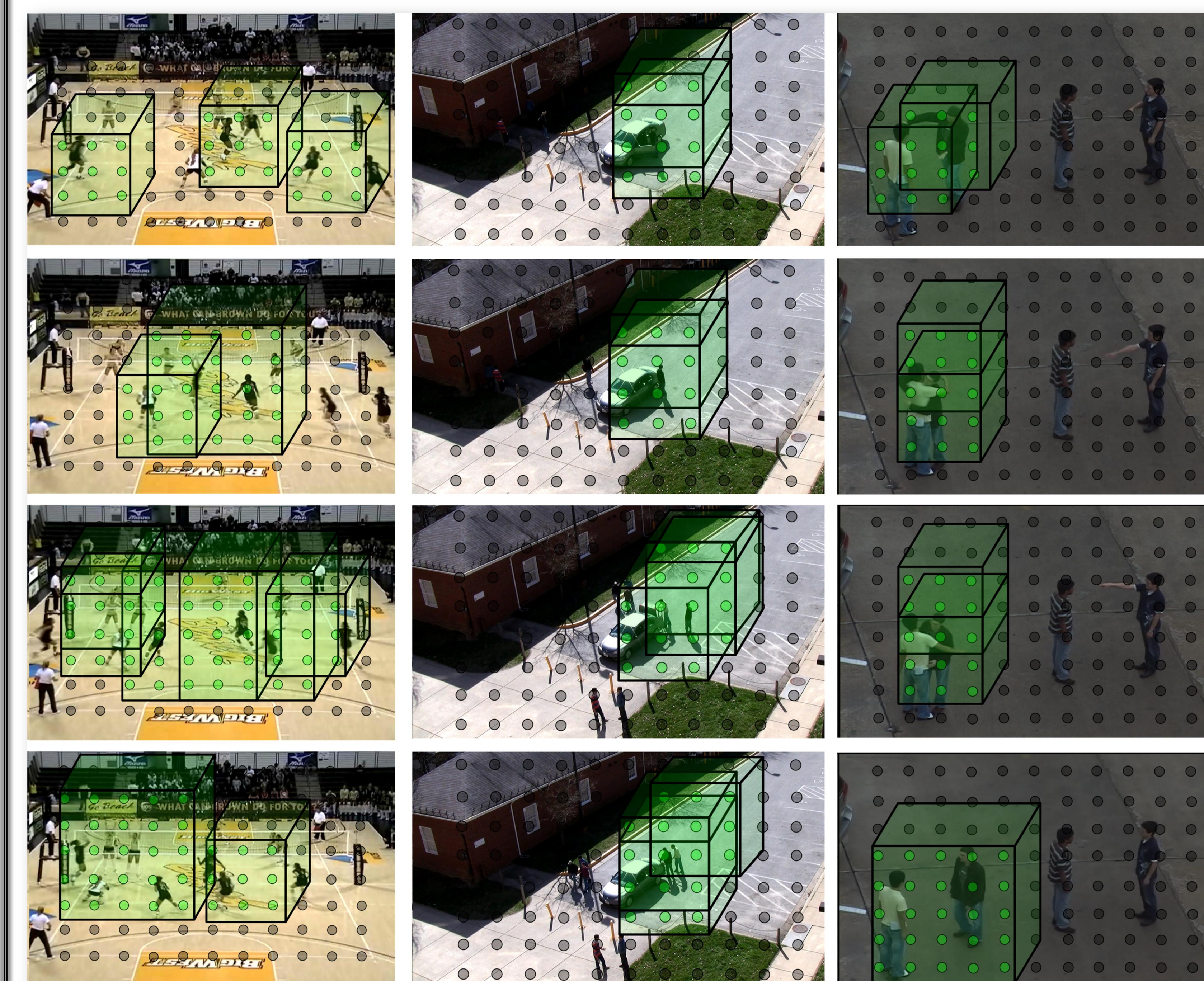
MPE: $\hat{a} = \text{argmax}_{a \in A} \hat{S}(C; a)$

Learning – Variational EM

1. Learn SPN structure parameters (w)

2. Learn Counting Grid parameters (π, θ)

Volleyball Dataset

- Classes: setting the ball to front/back left, right, or middle
- Dataset specifications:
  - **Videos per class**: 40 videos (20 train, 20 test)
  - **Resolution**: 848x480 pixels
  - **Length**: 4 seconds with 30 frames per second
- Ground-truth annotations are provided in terms of bounding boxes around the volleyball players engaged in the activity

Results

Results on the Volleyball, VIRAT, UT datasets. The green boxes are the foreground boxes highlighting the activity in the video.

Dataset	Var1	Var2	Var3	[3]
UT [4]	82.4 ± 2.6	80.3 ± 2.3	75.2 ± 5.3	76.0
KTH [5]	95.2 ± 2.5	94.1 ± 2.3	91.0 ± 4.2	93.9
VIRAT [6]	76.2 ± 3.1	72.5 ± 4.0	70.7 ± 4.2	68.1
Volleyball	69.8 ± 4.6	65.0 ± 4.6	59.0 ± 3.4	56.6

Average classification accuracy of our approach using different variants on KTH, UT , VIRAT, and Volleyball datasets.

References

[1] H. Poon, P. Domingos, UAI11
[2] N. Jojic, A. Perina, UAI11
[3] Q. V. Le. et al, CVPR11
[4] J. Aggarwal, M. Ryoo, ACM11
[5] C. Schueldt, et al, ICPR04
[6] S. Og et al., CVPR11

Acknowledgment