Predicting Dynamical Evolution of Human Activities from a Single Image

Suhas Lohit[1], Ankan Bansal[2], Nitesh Shroff[3], Jaishanker Pillai[4], Pavan Turaga[1] and Rama Chellappa[2]

[1] Arizona State University,
[2] University of Maryland, College Park,
[3] Zoox Inc., [4] Google Research

Computational Framework for Predicting Activity from a Single Image

- Modeling action segments

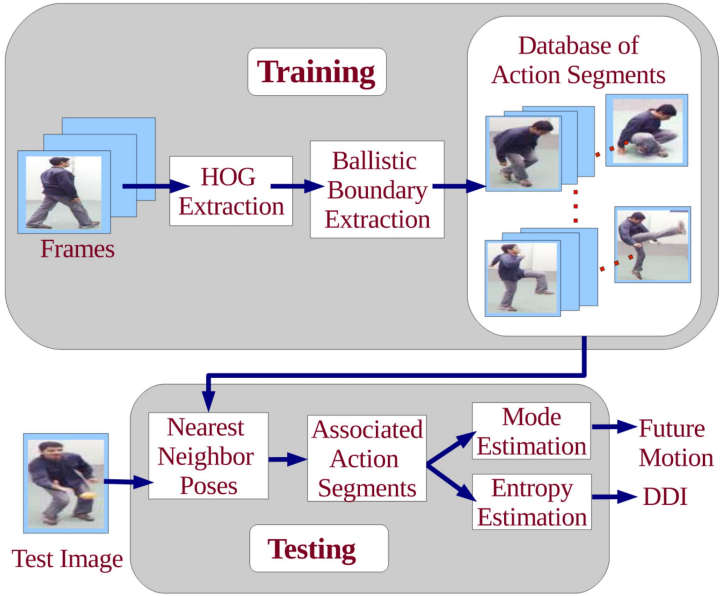
$$z_\phi(t+1) = A_\phi z_\phi(t) + v_\phi(t), v_\phi(t) \sim N(0, \Xi)$$
$$y_\phi(t) = C_\phi z_\phi(t) + w_\phi(t), w_\phi(t) \sim N(0, \Theta)$$
$$\hat{\Omega}_\phi^\top = \left[C_\phi^\top, (C_\phi A_\phi)^\top, \ldots, (C_\phi A_\phi^{m-1})^\top\right]$$
$$\zeta^2(\Omega_i, \Omega_j) = p - tr(\Omega_j^T \Omega_i \Omega_i^T \Omega_j)$$

- Density estimation on the Grassmannian

$$\hat{\mathcal{P}}(\phi|\pi_s) = c_1 \sum_{\phi_i \in \mathcal{N}_{\phi(\pi_s)}} \Psi(M^{-\frac{1}{2}}(I_d - \Omega_i^\top \Omega \Omega^\top \Omega_i) M^{-\frac{1}{2}})$$

- Statistical inference : mode estimation

$$\hat{\phi}(\pi_s) = \underset{\phi_i \in \mathcal{N}_{\phi(\pi_s)}}{\arg\max}\ \hat{\mathcal{P}}(\phi_i|\pi_s)$$



- Degree of Dynamic Information

$$\hat{\mathcal{H}}(\phi|\pi_s) = -\frac{1}{|\mathcal{N}_{\phi(\pi_s)}|} \sum_{\phi_i \in \mathcal{N}_{\phi(\pi_s)}} \log \hat{\mathcal{P}}(\phi_i|\pi_s)$$

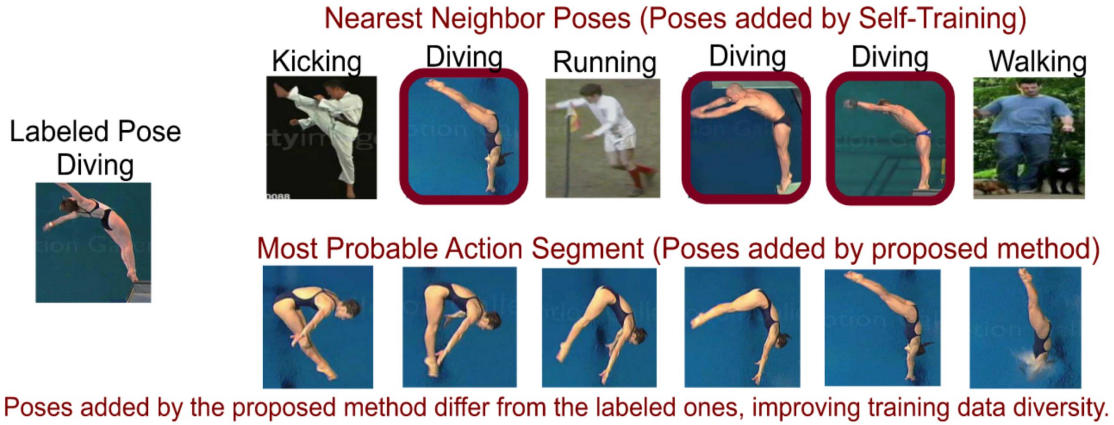
$$\mathrm{DDI}(\pi) = \exp[-\mathcal{H}(\phi|\pi)]$$

Application: Motion Prediction from a Single Image

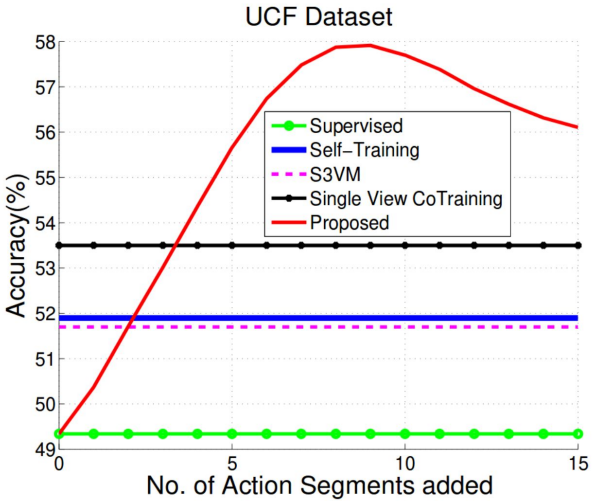
- Given a single frame, we can predict the most possible action segment

Application: Single-Image Semi-supervised Action Recognition

- We evaluate the label-propagation technique for semi-supervised action recognition
- Our method fares better than competitive approaches



Poses added by the proposed method differ from the labeled ones, improving training data diversity.



Thank you