Predicting Dynamical Evolution of Human Activities from a Single Image

Suhas Lohit\textsuperscript{1}, Ankan Bansal\textsuperscript{2}, Nitesh Shroff\textsuperscript{3}, Jaishanker Pillai\textsuperscript{4}, Pavan Turaga\textsuperscript{1} and Rama Chellappa\textsuperscript{2}

\textsuperscript{1} Arizona State University, \\
\textsuperscript{2} University of Maryland, College Park, \\
\textsuperscript{3} Zoox Inc., \textsuperscript{4} Google Research
Computational Framework for Predicting Activity from a Single Image

- **Modeling action segments**
  \[
  z_\phi(t + 1) = A_\phi z_\phi(t) + v_\phi(t), v_\phi(t) \sim N(0, \Xi) \\
  y_\phi(t) = C_\phi z_\phi(t) + w_\phi(t), w_\phi(t) \sim N(0, \Theta) \\
  \hat{\Omega}_\phi^T = \begin{bmatrix} C_\phi^T, (C_\phi A_\phi)^T, \ldots, (C_\phi A_\phi^{m-1})^T \end{bmatrix} \\
  \zeta^2(\Omega_i, \Omega_j) = p - tr(\Omega_j^T \Omega_i \Omega_i^T \Omega_j)
  \]

- **Density estimation on the Grassmannian**
  \[
  \hat{P}(\phi | \pi_s) = c_1 \sum_{\phi_i \in N_\phi(\pi_s)} \Psi(M^{-\frac{1}{2}} (I_d - \Omega_i^T \Omega \Omega^T \Omega_i) M^{-\frac{1}{2}})
  \]

- **Statistical inference: mode estimation**
  \[
  \hat{\phi}(\pi_s) = \arg \max_{\phi_i \in N_\phi(\pi_s)} \hat{P}(\phi_i | \pi_s)
  \]

- **Degree of Dynamic Information**
  \[
  \hat{H}(\phi | \pi_s) = -\frac{1}{|N_\phi(\pi_s)|} \sum_{\phi_i \in N_\phi(\pi_s)} \log \hat{P}(\phi_i | \pi_s) \\
  \text{DDI}(\pi) = \exp[-\hat{H}(\phi | \pi)]
  \]
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
Thank you