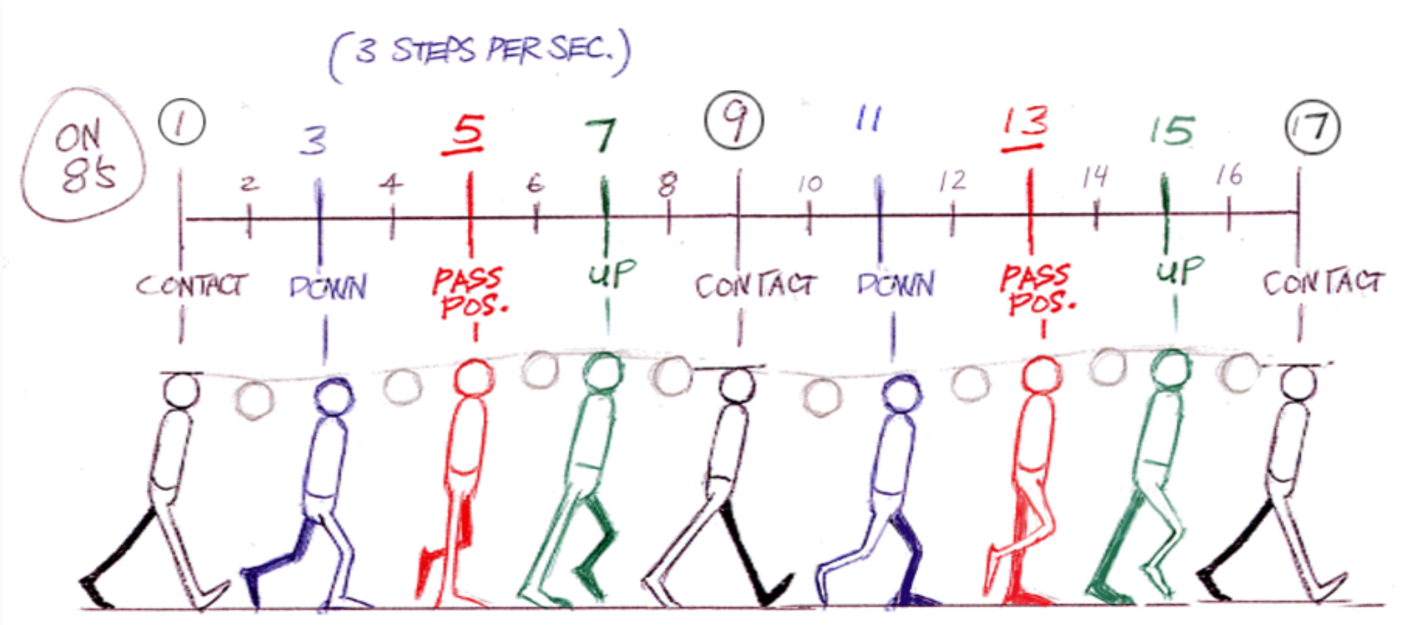
KeyIn: Discovering Subgoal Structure with Keyframe-based Video Prediction

Penn Engineering | GRASP Laboratory, General Robotics, Automation, Sensing & Perception Lab
SAMSUNG AI CENTRE
USC University of Southern California
Ryerson University

Karl Pertsch*, Oleh Rybkin*, Jingyun Yang,
Konstantinos G. Derpanis, Joseph Lim, Kostas Daniilidis, Andrew Jaegle
(*equal contribution)

Idea



Inspiration: Real-world videos can often be summarized with just few key snapshots (keyframes).

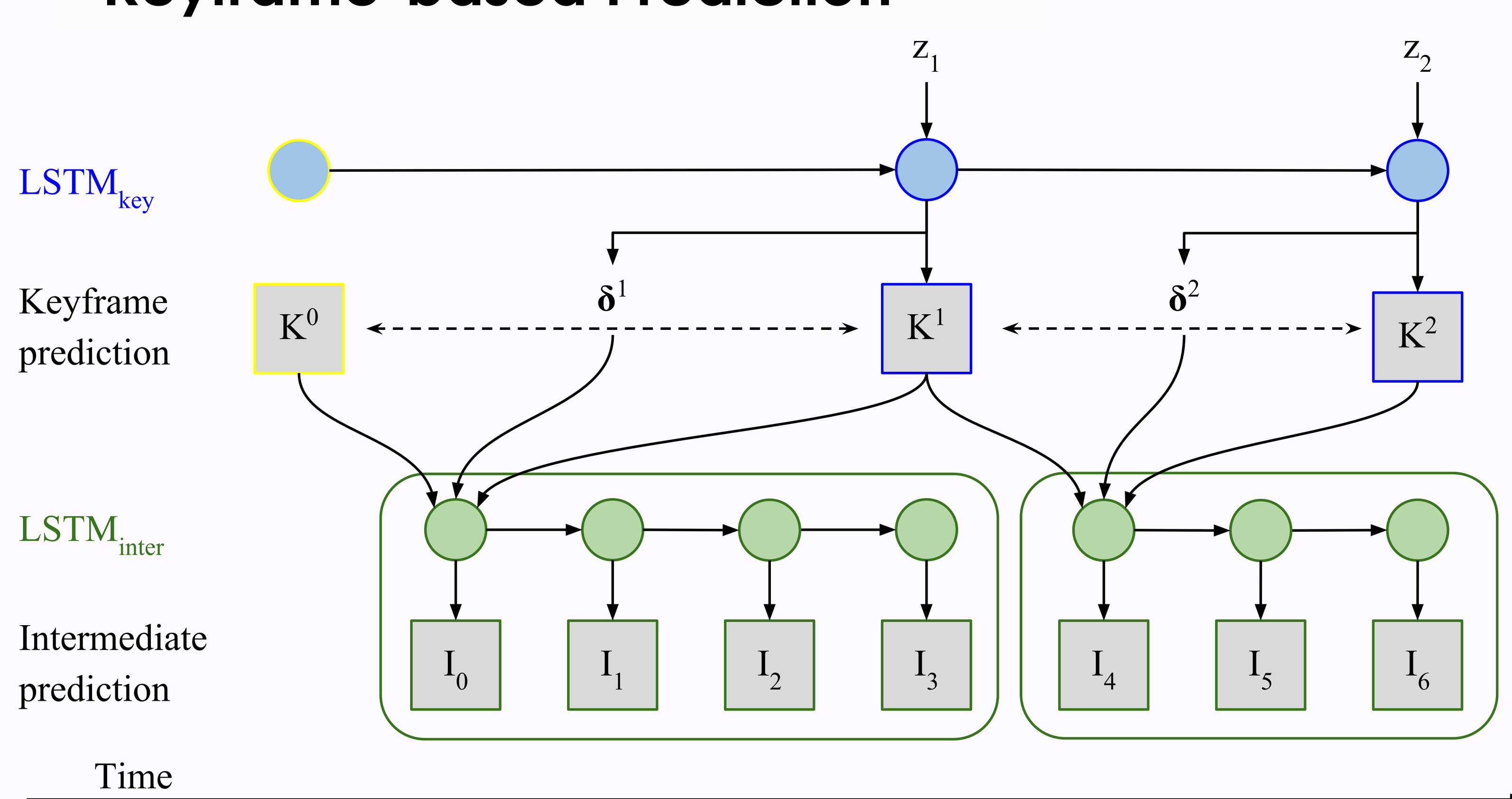
Task: We want to discover keyframes in videos by finding the subset of frames that best describes the sequence.

Idea: Train a variational model to select frames from which it can reconstruct the rest of the video.

Application: We focus on hierarchical planning: we first plan the subgoals (keyframes) and then plan how to reach them.

Approach

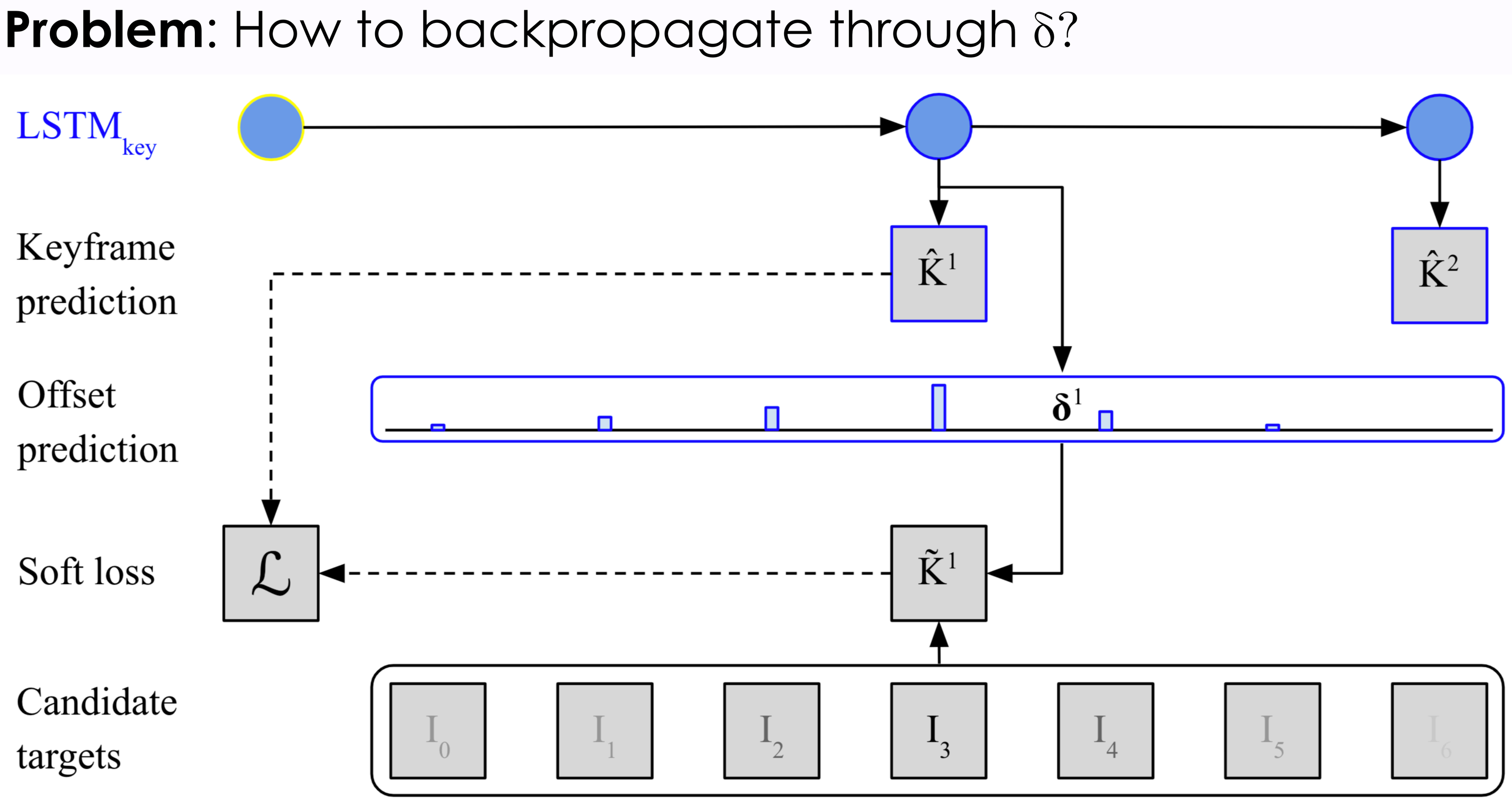
Keyframe-based Prediction



The high-level network predicts a sequence of keyframes K and distributions over time offsets δ, the low level network interpolates between each pair of keyframes.

Soft Reconstruction Loss

Problem: How to backpropagate through δ?



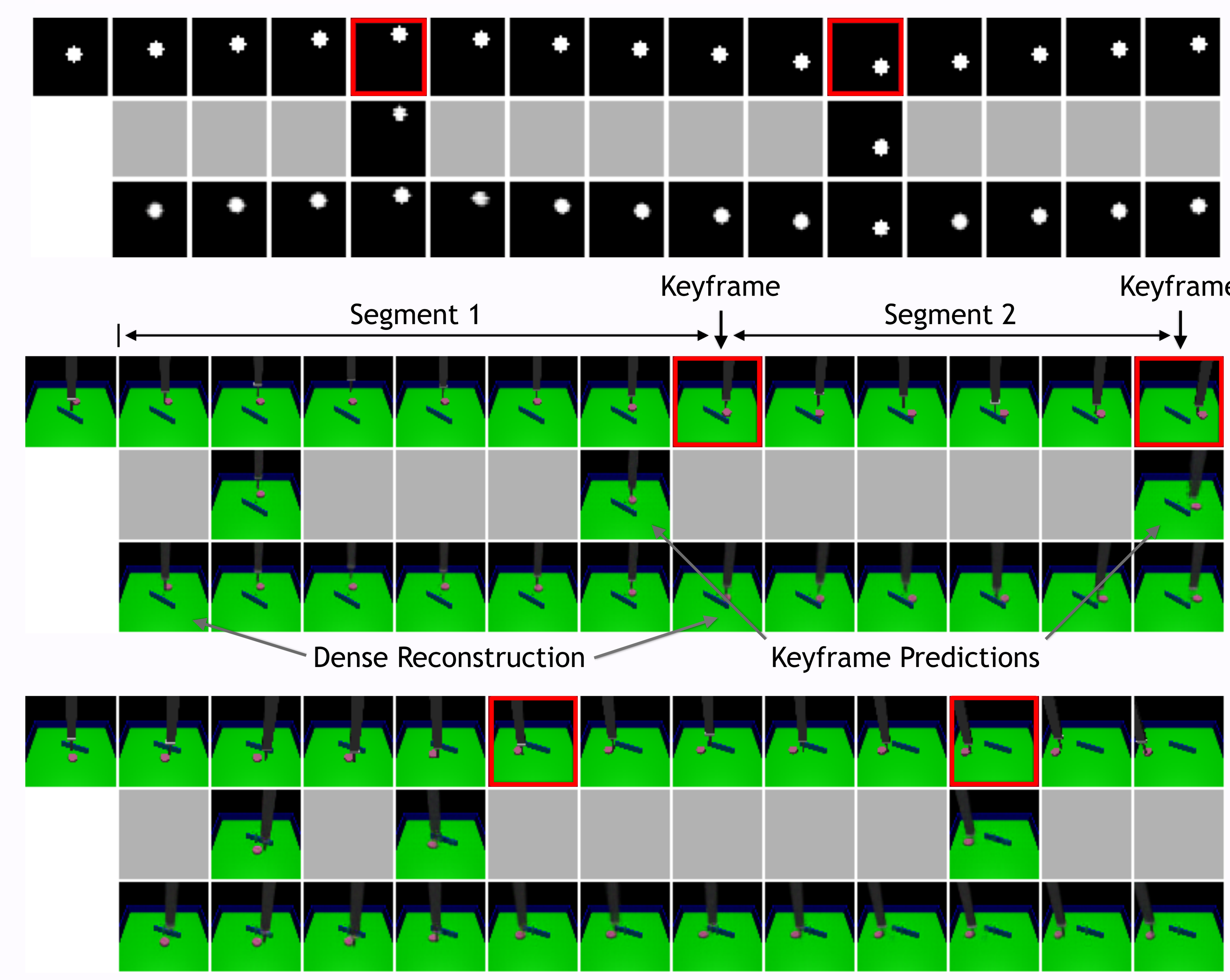
Proposed solution: The reconstruction loss is computed as an expectation over δ. There is no sampling.

$$\mathcal{L}_{rec} = \sum_t c^t \beta_{ki} ||\hat{K}^t - \tilde{K}^t||^2 + \sum_{t,i} ||I_i^t - \tilde{I}_i^t||^2$$

$$\tilde{K}^t = \textstyle\sum_j \tilde{\delta}_j^t I_j \qquad \tilde{I}_j = (\sum_{t,i} \tilde{\delta}_{i,j}^t \hat{I}_i^t) / \sum_{t,i} \tilde{\delta}_{i,j}^t$$

Experiments

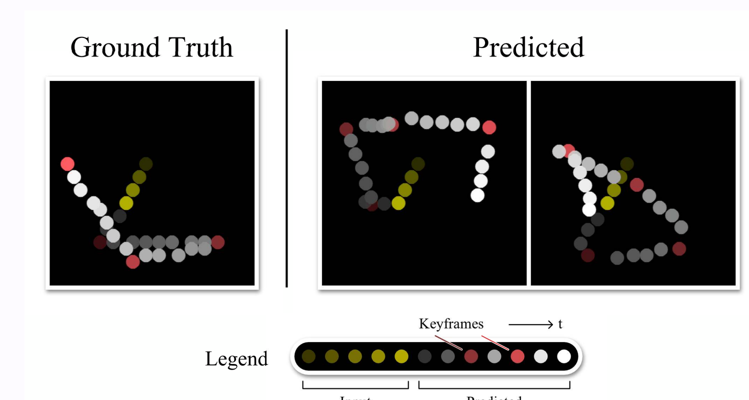
Keyframe Discovery



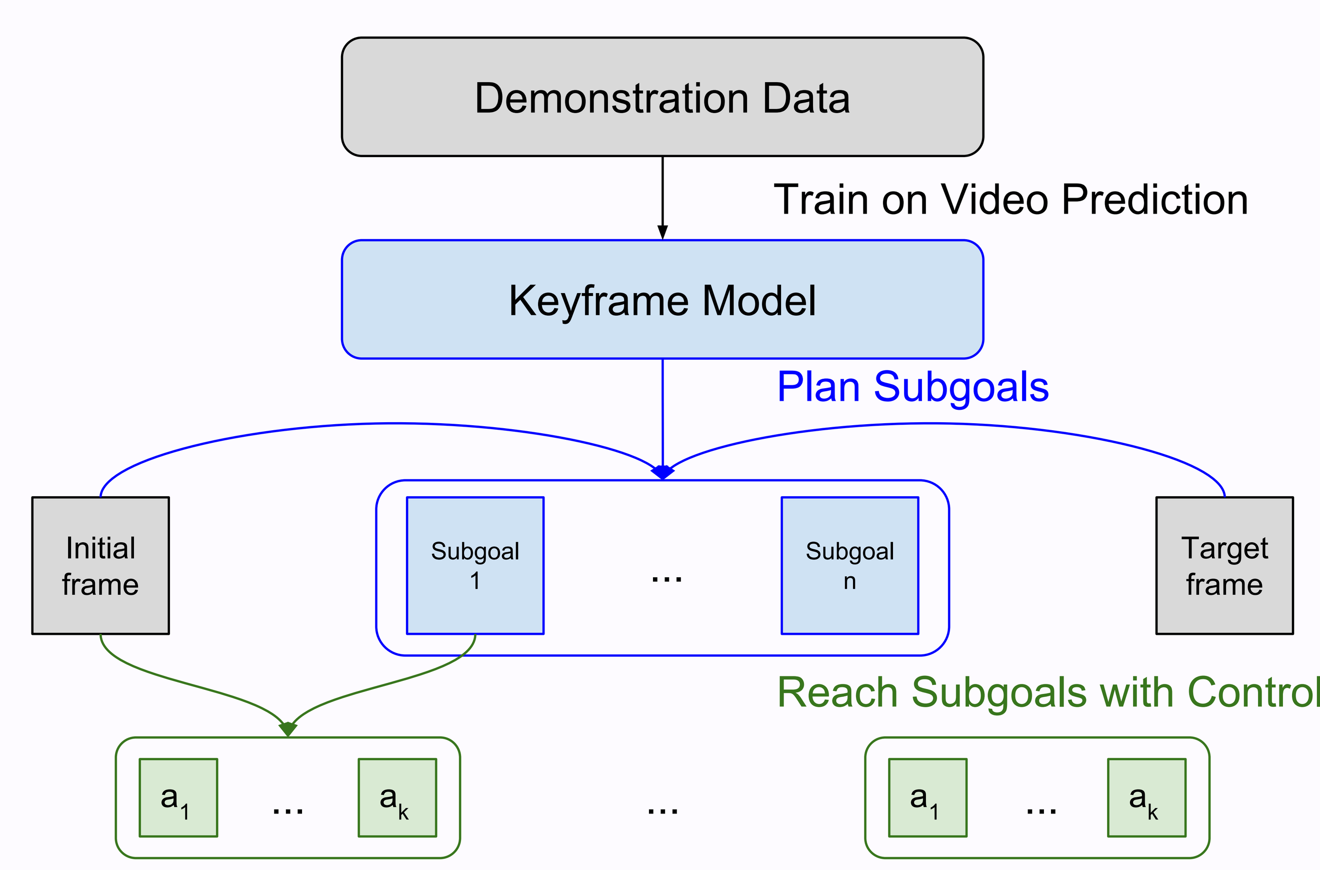
Our model can correctly select descriptive keyframes from a given sequence (above) and predict a distribution of possible next keyframes (below, right).

F1 score on keyframe discovery

Method	SBM	Push
Random	0.15	0.18
Constant	0.17	0.23
Static	0.21	0.18
Surprise	0.73	0.10
KeyIn (Ours)	**0.84**	**0.30**



Hierarchical Planning



We use a model predictive control method (CEM) to (a) generate a plan of high-level subgoals (the keyframes predicted by KeyIn), and (b) plan trajectories between these subgoals.

Planning performance

Method	Position error	Success Rate
Intitial	1.32 ± 0.06	-
Random	1.32 ± 0.07	-
No subgoals	0.90 ± 0.14	15.0 %
TAP	0.80 ± 0.16	23.3 %
Surprise	0.64 ± 0.28	50.8 %
Jumpy	0.62 ± 0.33	58.8 %
KeyIn (Ours)	$\mathbf{0.50 \pm 0.26}$	**64.2 %**