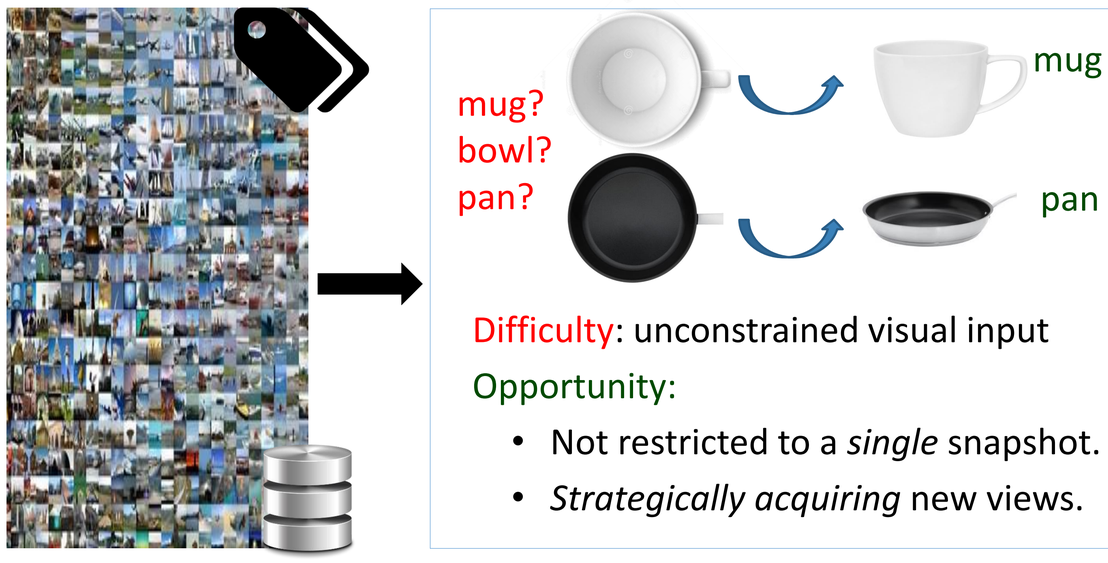
Look-ahead before you leap: End-to-end active recognition by forecasting the effect of motion

Dinesh Jayaraman and Kristen Grauman
UT Austin

Active recognition setting

passive 1-view setting

active setting

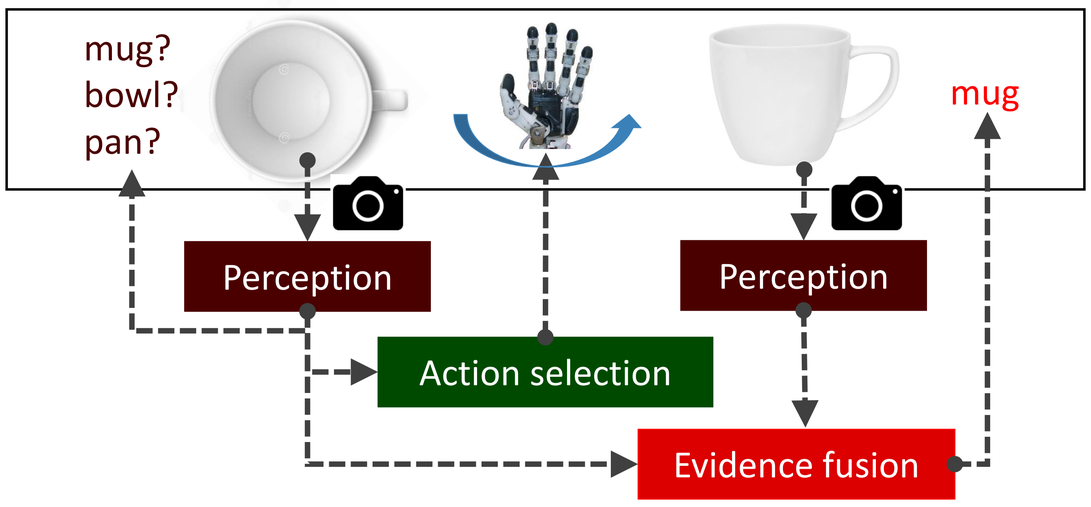


Difficulty: unconstrained visual input

Opportunity:

- Not restricted to a *single* snapshot.
- *Strategically acquiring* new views.

Components of the active recognition pipeline



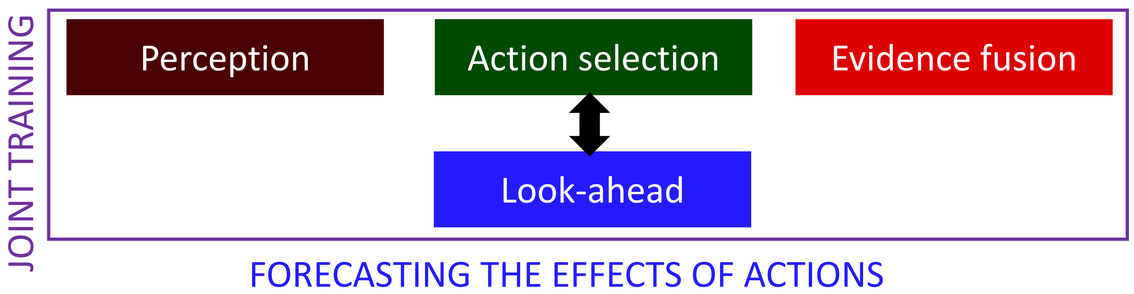
Closely intertwined perception, action and fusion components

Our idea

Prior art: independent, often heuristic components

[Wilkes 1992, Dickinson 1997, Borotschnig 1998, Schiele 1998, Denzler 2002, Soatto 2009, Ramanathan 2011, Aloimonos 2011, Borotschnig 2011, Wu 2015, Malmir 2015, Johns 2016, …]

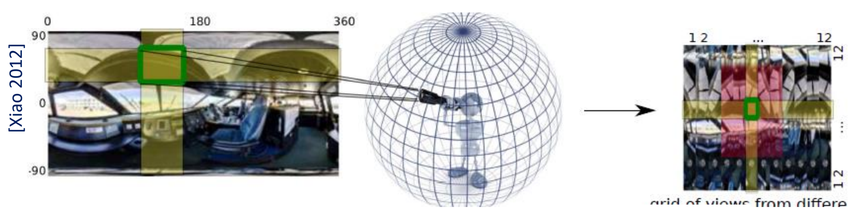
Our idea: Multi-task joint training of components for active recognition + auxiliary internally supervised "look-ahead" task.



FORECASTING THE EFFECTS OF ACTIONS

Towards real-world active recognition

SCENE CATEGORIES
SUN 360 panoramas

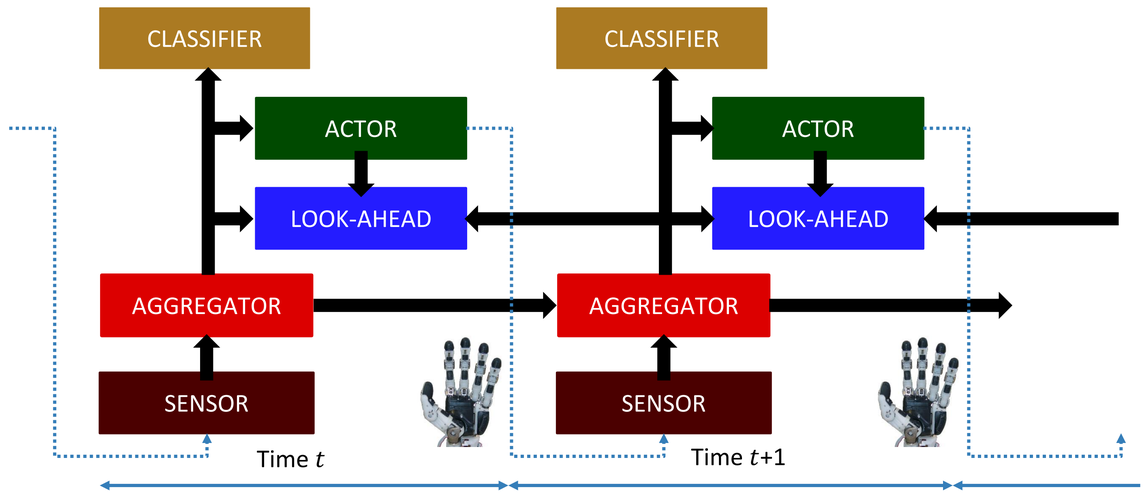


HAND-HELD OBJECTS
GERMS toy manipulation

3-D CAD MODELS
ModelNet-10 3-D models

Complex real-world categories + easily benchmarkable setups.

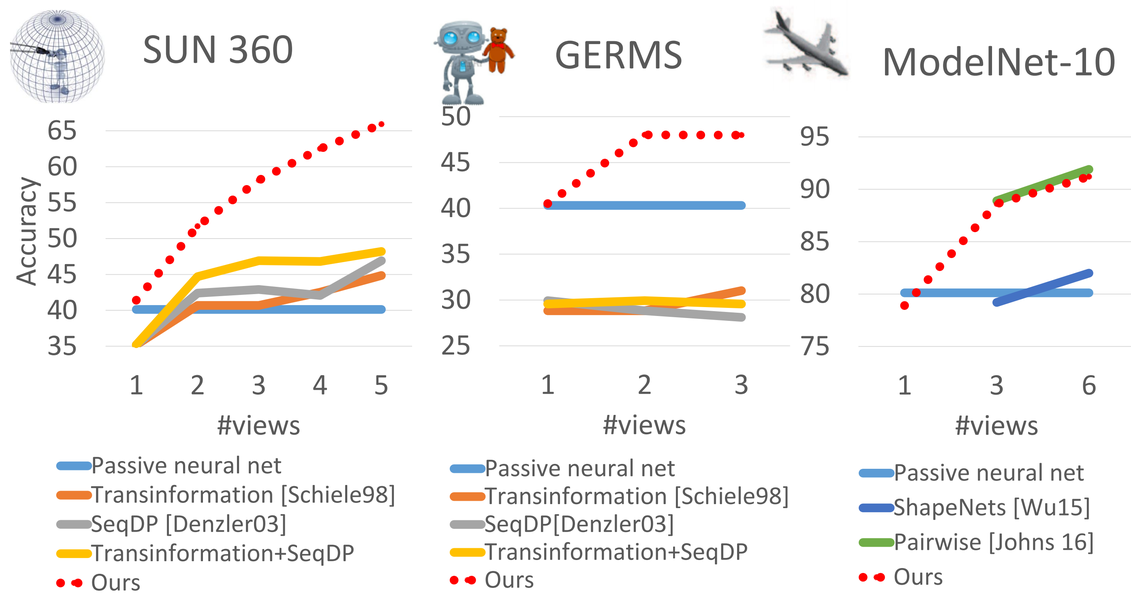
High-level active RNN system architecture



End-to-end joint training with gradient descent + REINFORCE

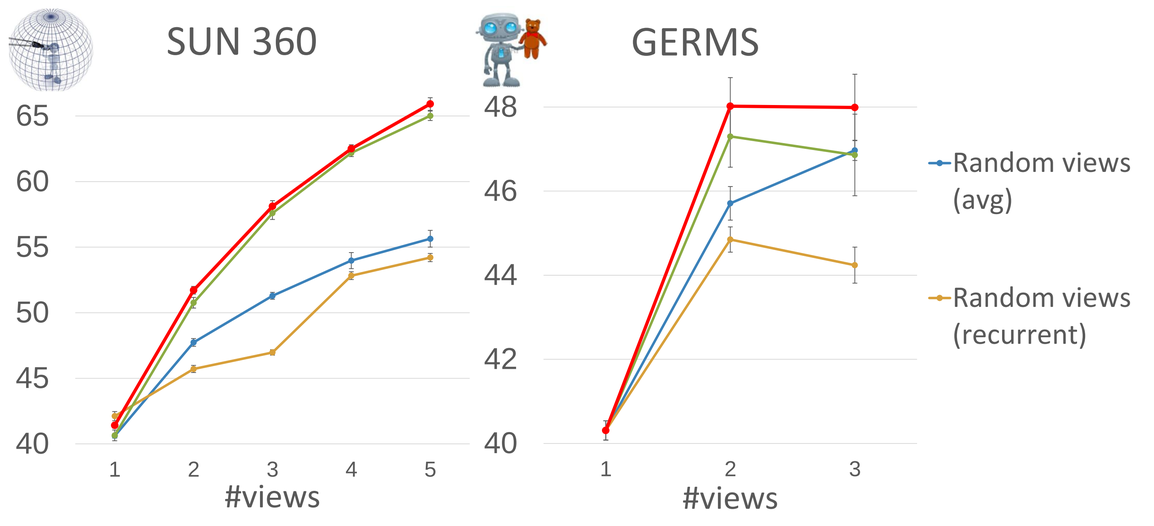
Selected view sequence examples

Quantitative evaluation results



Our method strongly outperforms representative traditional active recognition approaches on all tasks.

Ablation studies



Training all 3 components jointly is most critical to performance.

Component module architectures