UCR Bourns College of Engineering Video Computing Group

Continuous Learning of Human Activity Models using Deep Nets

Mahmudul Hasan and Amit K. Roy-Chowdhury
University of California Riverside, CA-92521, USA.

ECCV '14 Zürich, September 6-12th, 2014 European Conference on Computer Vision

Motivation

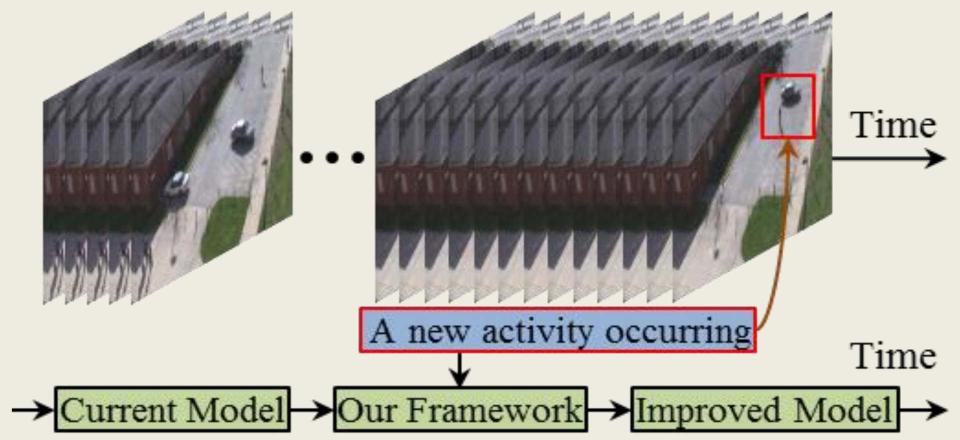
- Activity recognition strategies assume large amounts of labeled training data which require tedious human labor to label.
- They also use hand engineered features, which are not best for all applications, hence required to be done separately for each application.
- Several recognition strategies have benefited from deep learning for unsupervised feature selection, which has two important property – fine tuning and incremental update.

Question!

Can deep learning be leveraged upon for continuous learning of activity models from streaming videos?

Contributions

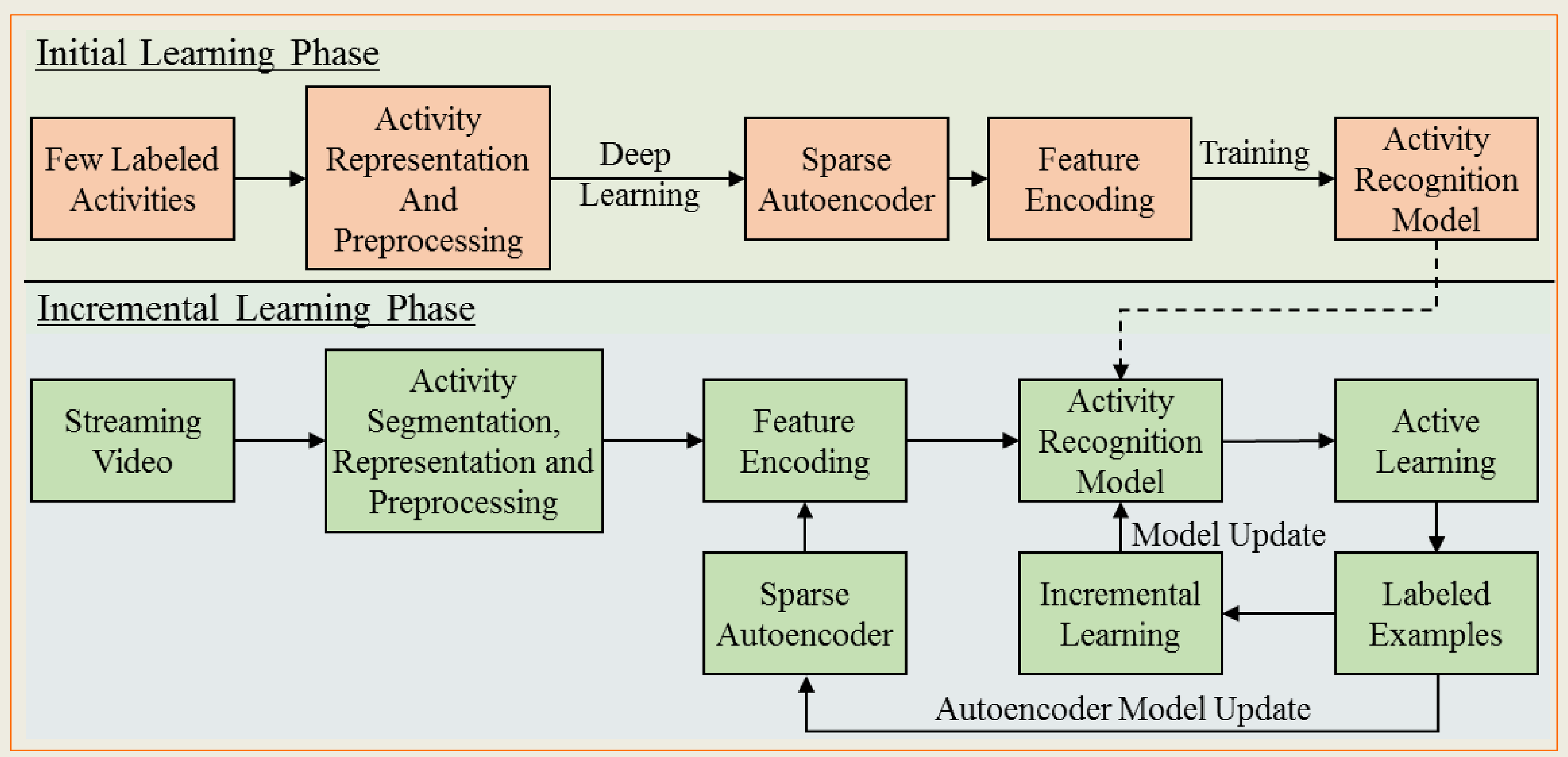
We propose a novel framework for continuous learning of activity models from streaming videos by intricately tying together deep learning and active learning.



Goals

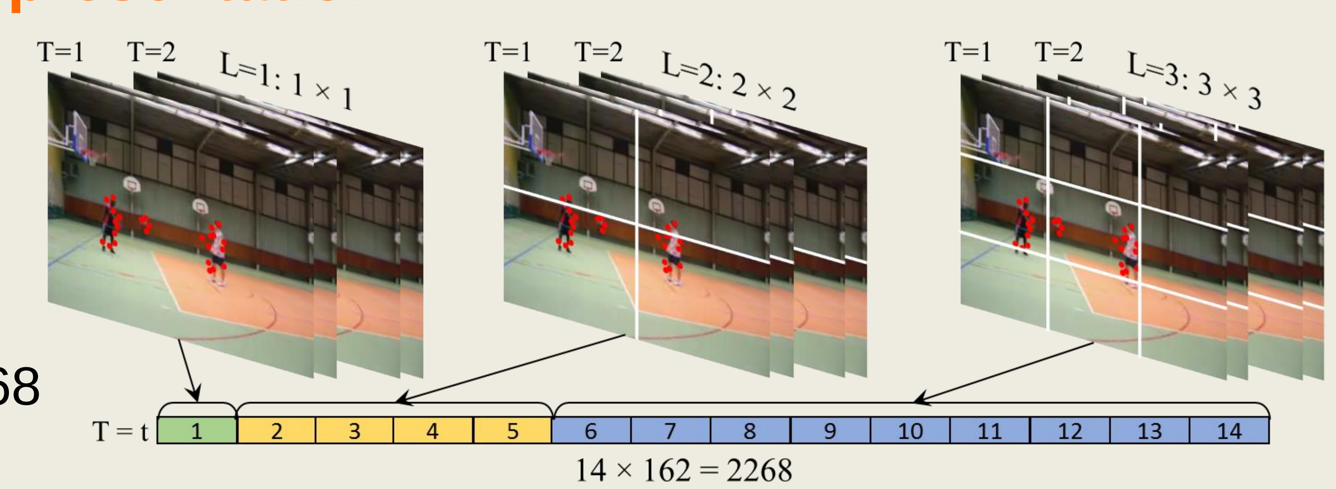
- Automatically learning the best set of features in unsupervised manner.
- Reducing the amount of manual labeling of the unlabeled instances.
- Retaining already learned information without storing all the previously seen data and continuously improve the existing activity models.

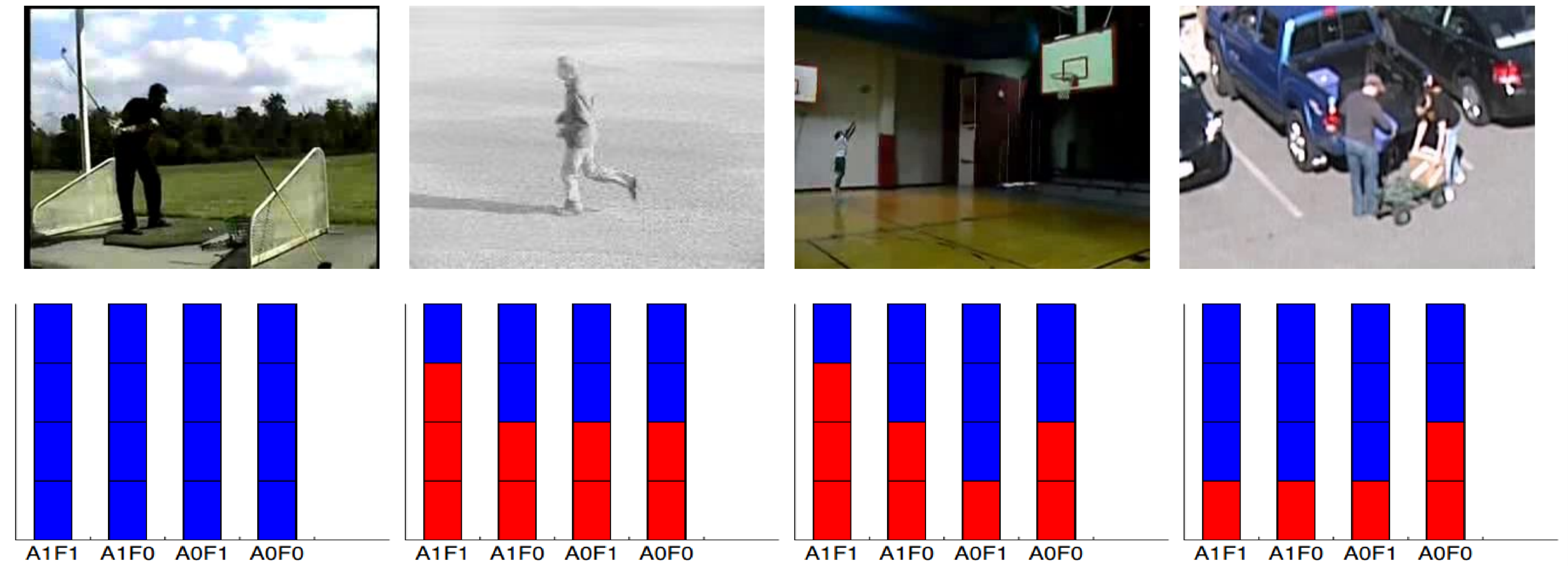
Framework



Initial Activity Representation

- STIP Feature
- Spatial-temporal pyramid
- Average Pooling
- Vector Size: T*2268





Effect of continuous learning on individual activity instances.

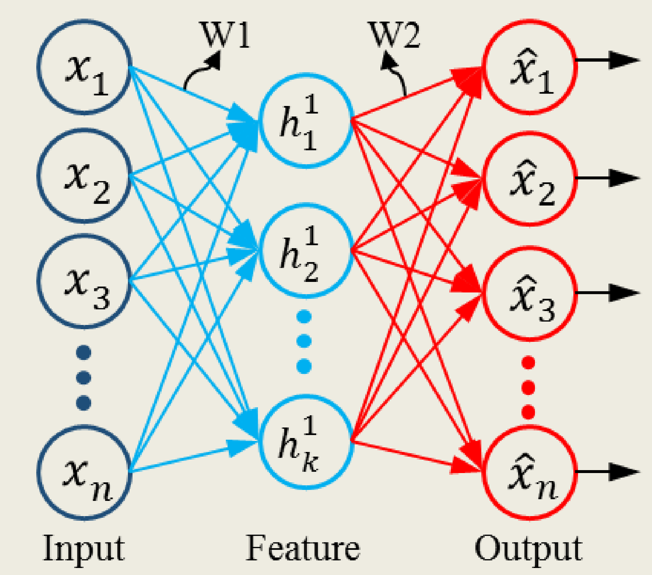
Sparse Autoencoder

$$\arg\min_{W} J_a(W) = \frac{1}{2m}\sum_{i=1}^{m}\left\|x^i - \hat{x}^i\right\|^2 + \lambda\|W\|^2 + \beta\sum_{j=1}^{k}\Psi(\rho\|\hat{\rho}_j)$$

$$\Psi(\rho\|\hat{\rho}_j) = \rho\ \log(\rho/\hat{\rho}_j) + (1-\rho)\ \log((1-\rho)/(1-\hat{\rho}_j))$$

Activity Model

$$\arg\min_{\theta} J_s(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{c} 1\{y^i = j\}\log P(y^i = j|x^i;\theta)$$



Autoencoder training

```
Training:
Initialize W = [W¹, W², b¹, b²]
Repeat for i = 1 : m
  Perform feedforward pass:
    Compute: x̂ⁱ.
  Perform backpropagation:
    Compute gradients: ∇_W J_a(W).
    Compute weight change: ΔW.
    Update weight W.
Feature Encoding:
Compute: x̄ⁱ = f(W¹xⁱ + b¹).
```

Active Learning

Two types of teacher –

1. Strong teacher – Human
2. Weak teacher – Classifier

We select one of them based on the $\Phi(x^i)$.

Expected Gradient Length –

$$\Phi(x^i) = \sum_{j=1}^{c} P(y^i = j|x^i)\left\|\nabla_{\theta_j} J_s(\theta)\right\|$$

$$U^* = \arg\min_{X\subseteq U \cap \left(\frac{|X|}{|U|}\right)=\alpha} \sum_{x\in X}\Phi(x)$$

Incremental Learning

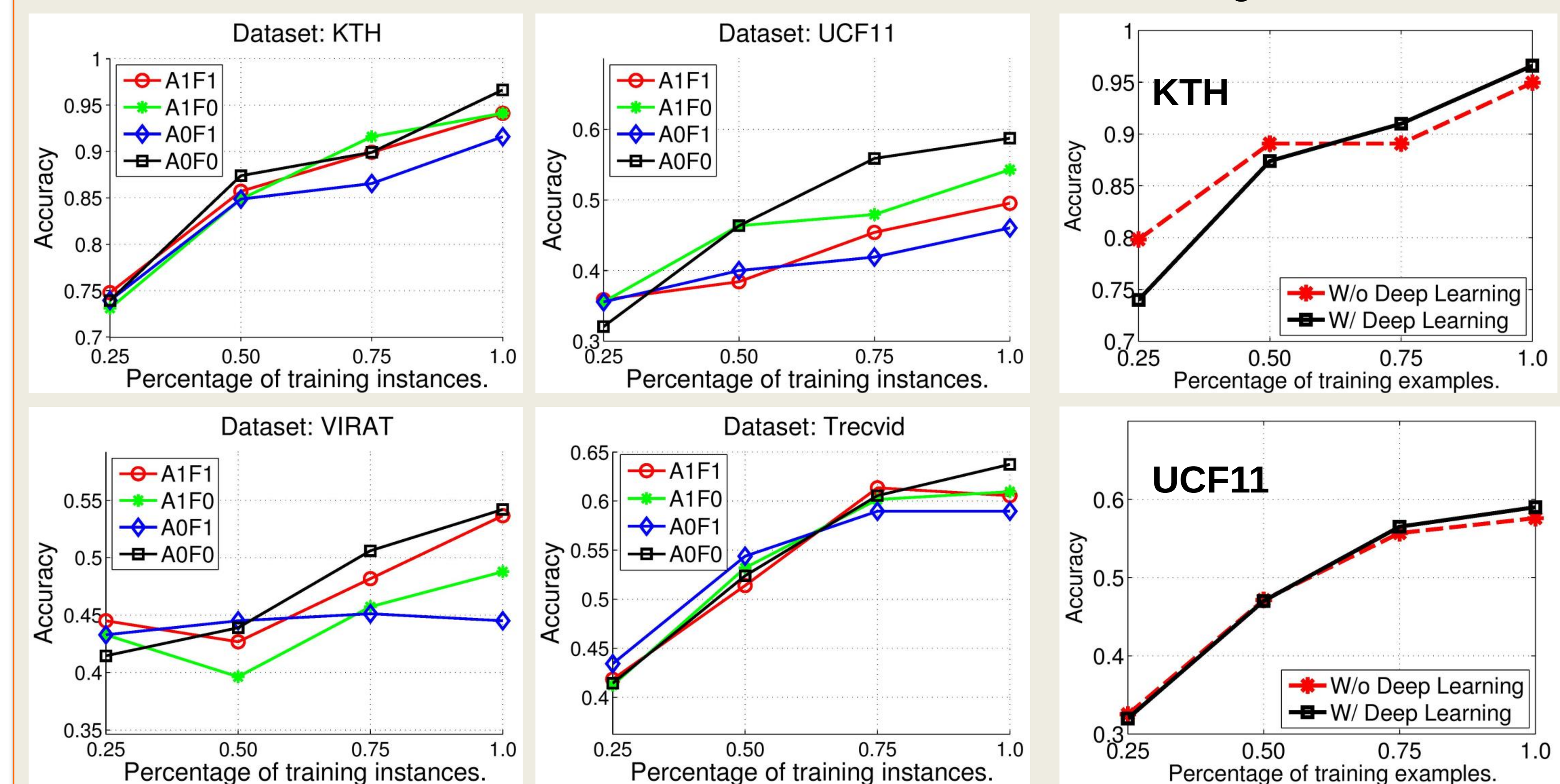
Mini-batch incremental learning	Most diverse instance selection

Initialize the weights.
Repeat the following steps:
If u training instances available:
  Process u training instances.
  Compute gradients.
  Update the weights.
Else
  Wait for stream data to arrive
	Repeat for each class c.
	  Available instances: N_c.
	  Available memory spaces: K_c
	  If $K_c < N_c$:
	    Use kmean clustering algo. to compute K_c clusters from N_c.
	    Assign N_c inst. to K_c clusters.
	    Store one instance per cluster.
	  Else
	    Store all of the N_c instances.

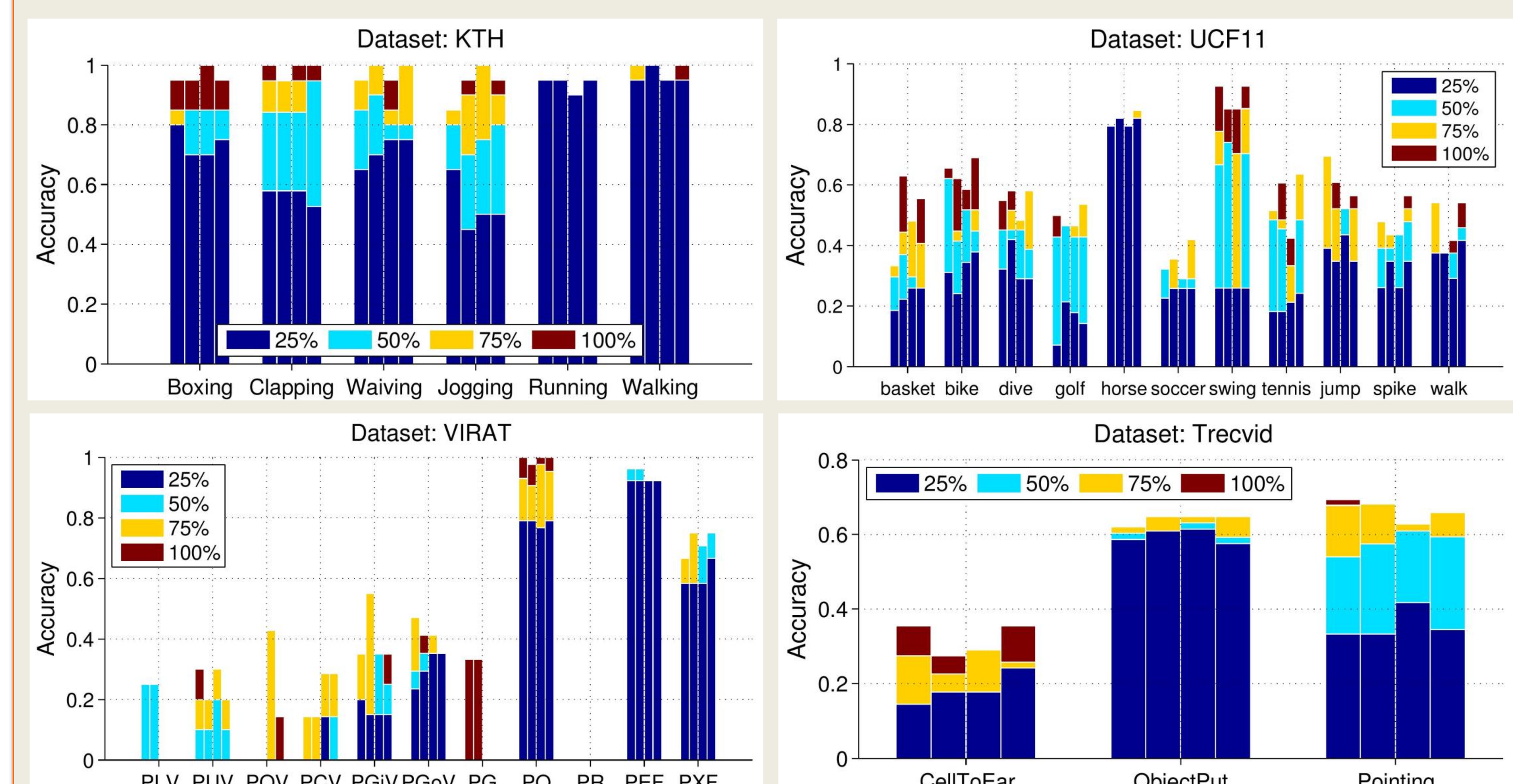
Experiments

- Four datasets –
- KTH – 6 classes, 600 Ins.
- UCF11 – 11 classes, 1600 ins.
- VIRAT – 11 classes, ~12hrs.
- TRECVID – 3 classes, 40hrs.

- Four test Scenarios
- A1F1 – Active Learning + Fixed buffer.
- A1F0 – Active Learning + Infinite buffer.
- A0F1 – No active learning + Fixed buffer.
- A0F0 – No active learning + infinite buffer.



Averaged performance over all activity classes increases as the new instances become available. Effect of Deep Learning



Performance of continuous learning over each activity class.

Summary

- Deep learning has significant impact on learning activity models continuously.
- Most realistic method A1F1 which is comprised of deep learning, active learning, and fixed buffer can achieve performance close to A0F0 which approximates the batch methods in the existing literature.
- When all the instances are seen, final accuracies of our methods in A1F1 are very competitive with state-of-the-art works.

Acknowledgement: This work was supported in part by ONR grant N00014-12-1-1026 and NSF grant IIS-1316934.