Deep Learning & Small Training Samples

Dr. Simon Streltsov

Prof. Ilya Muchnik

ADSA

May 2018

info@longshortway.com

LongShortWay Inc.

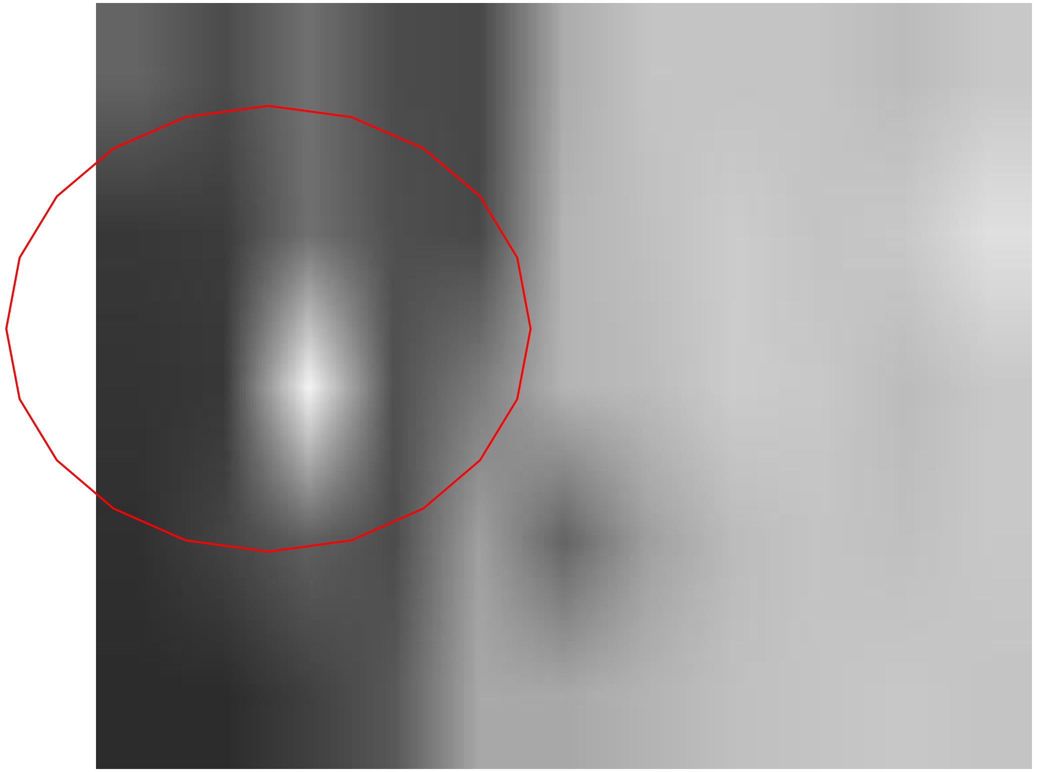
Boston MA

So what?

How to apply Deep Learning to practical problems with small number of training samples

- Deep Learning methods assume large amount of training data
- Large training set is not always available in practice
    - new objects, behaviors, environment
    - WORK ON REAL DATA even if small
    - prohibitive cost and time of collecting ground truth
    - assist a user earlier in data collection process
- Training image classifiers with small number of samples
    1. use off-the-shelf CNN, compute large number of features
    2. use good small sample methods - SVM, Random Forest, Logistic Regression
    3. use methods that accelerate convergence and increase robustness:
       privileged information, feature clustering, domain adaptation
- Challenge - changing data. Large training samples does not help!
    - Machine learning breaks engineering process
      See Leon Bottou https://icml.cc/2015/invited/LeonBottouICML2015.pdf

Example:
Small object detection in images and video
Classify real objects v. false alarms and clutter



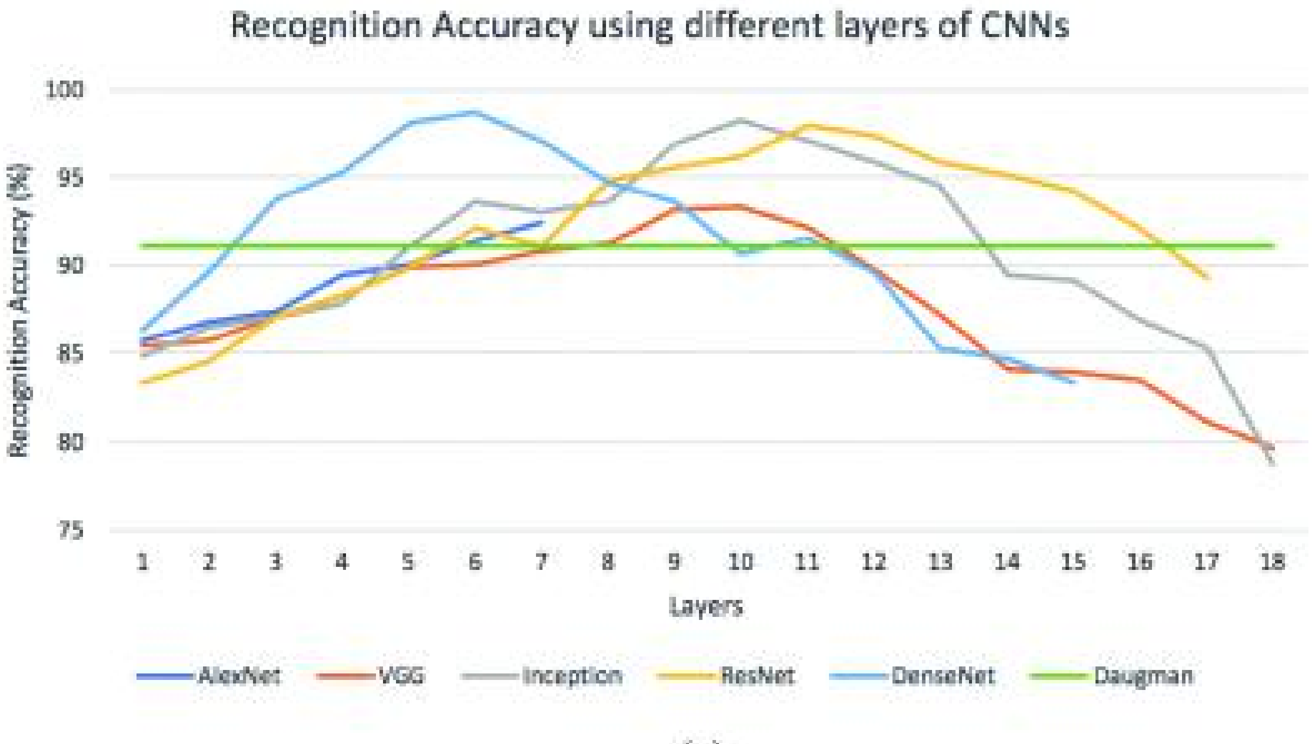
Detecting movers in low-res video

Detecting movers in low-res satellite imagery

How to learn

- Classical Machine Learning: Support Vector Machines (SVM)
  - Limited complexity of classification rules
  - Controls generalization error for a small training samples
- Deep Learning (DL)
  - Complex rules
  - Controls generalization numerically with large number of training samples
- Find appropriate DL architecture for small samples
  - Baldi, Pierre, Peter Sadowski, and Zhiqin Lu. "Learning in the machine: Random backpropagation and the deep learning channel." Artificial intelligence 260 (2018): 1-35. http://www.igb.uci.edu/~pfbaldi/
- Combine DL with classical methods for small sample learning
  - Use information accumulated by Deep Learning
  - Control generalization

CNN off the shelf - use CNN layers as features for new classification problems



(a)

figure from Nguyen, Kien & Fookes, Clinton & Ross, Arun & Sridharan, Sridha. (2017). Iris Recognition with Off-the-Shelf CNN Features: A Deep Learning Perspective. IEEE Access. PP. 1-1. 10.1109/ACCESS.2017.2784352. https://ieeexplore.ieee.org/document/8219390/

Deming, Ross, Roman Ilin, and Scott MacIntosh.) (2017) "Deep learning for fusing multi-sensor person-borne IED data." IEEE International Symposium on Technologies for Homeland Security (HST), Also at ASDA!

How to improve generalization

- Limit model complexity, maximize expected error - SVM
- Cross-validation: train, test
- Ensemble methods - Random Forest, Boosting

Advanced methods:

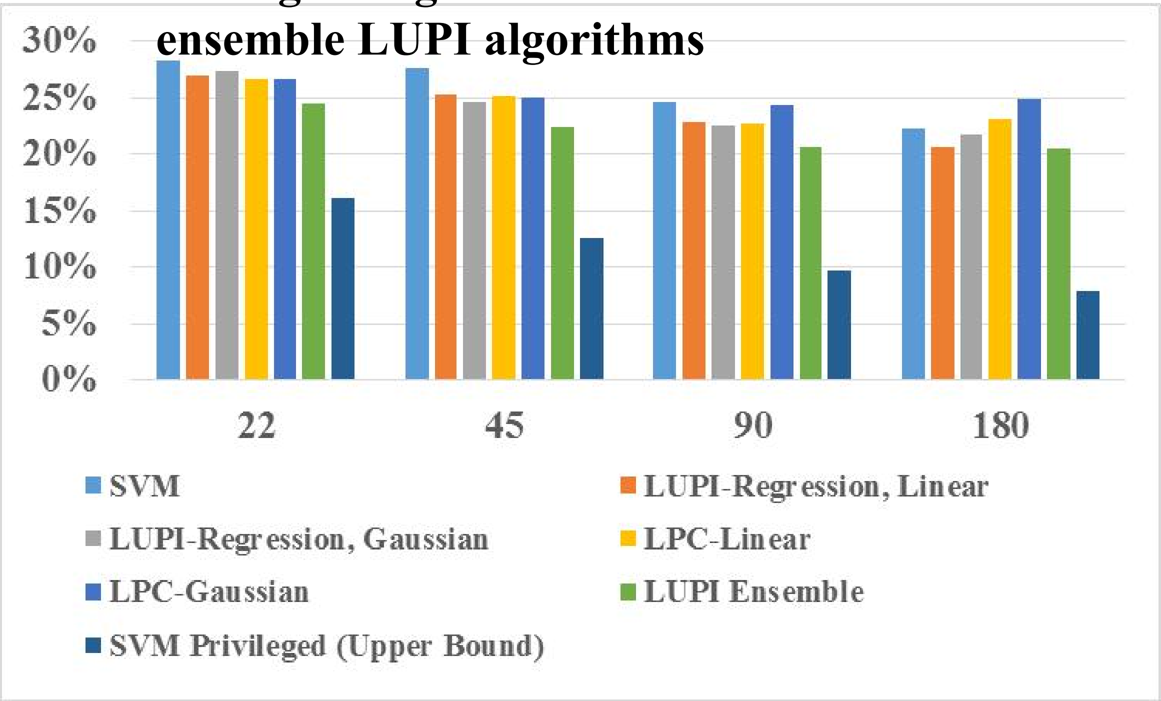
- Privileged Information - use expert information during training
- Feature Clustering - partition the feature space
- Domain Adaptation (new targets, new equipment) - use new unlabeled data to correct classifiers

Learning Under Privileged Information (LUPI)

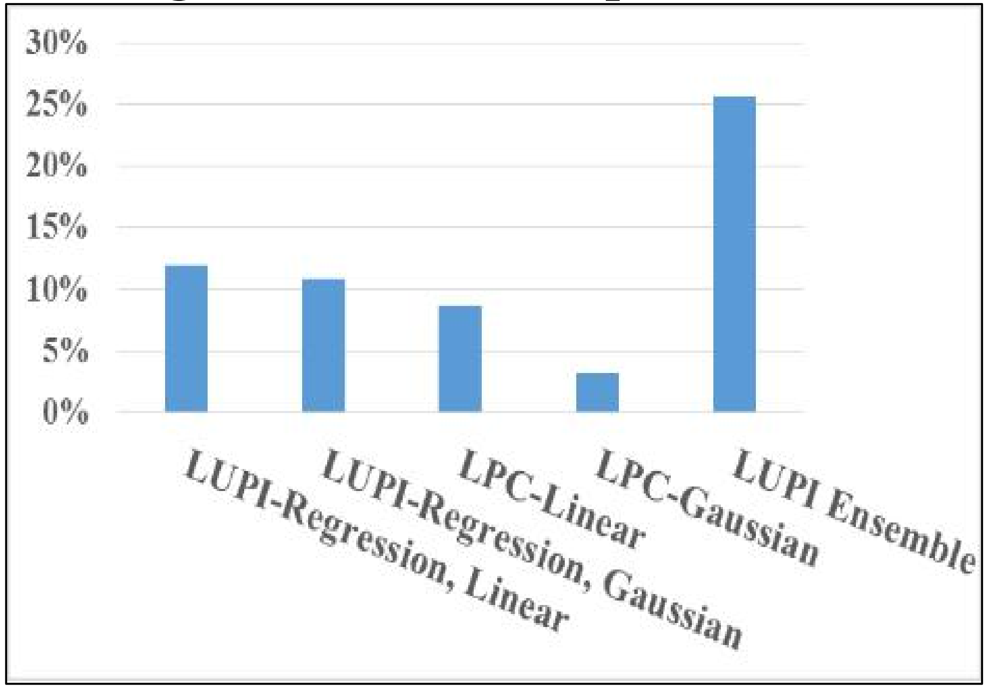
Extra data available during training, but not during detection.

The concept of a Teacher.

Average Weighted Error for individual and ensemble LUPI algorithms



Average Performance Improvement



LPC = Learning with Privileged Clusters

Pechyony, D., Izmailov, R., Vashist, A., Vapnik, V. (2010). SMO-Style Algorithms for Learning Using Privileged Information. In DMIN (pp. 235-241). https://link.springer.com/article/10.1007/s10472-017-9538-x

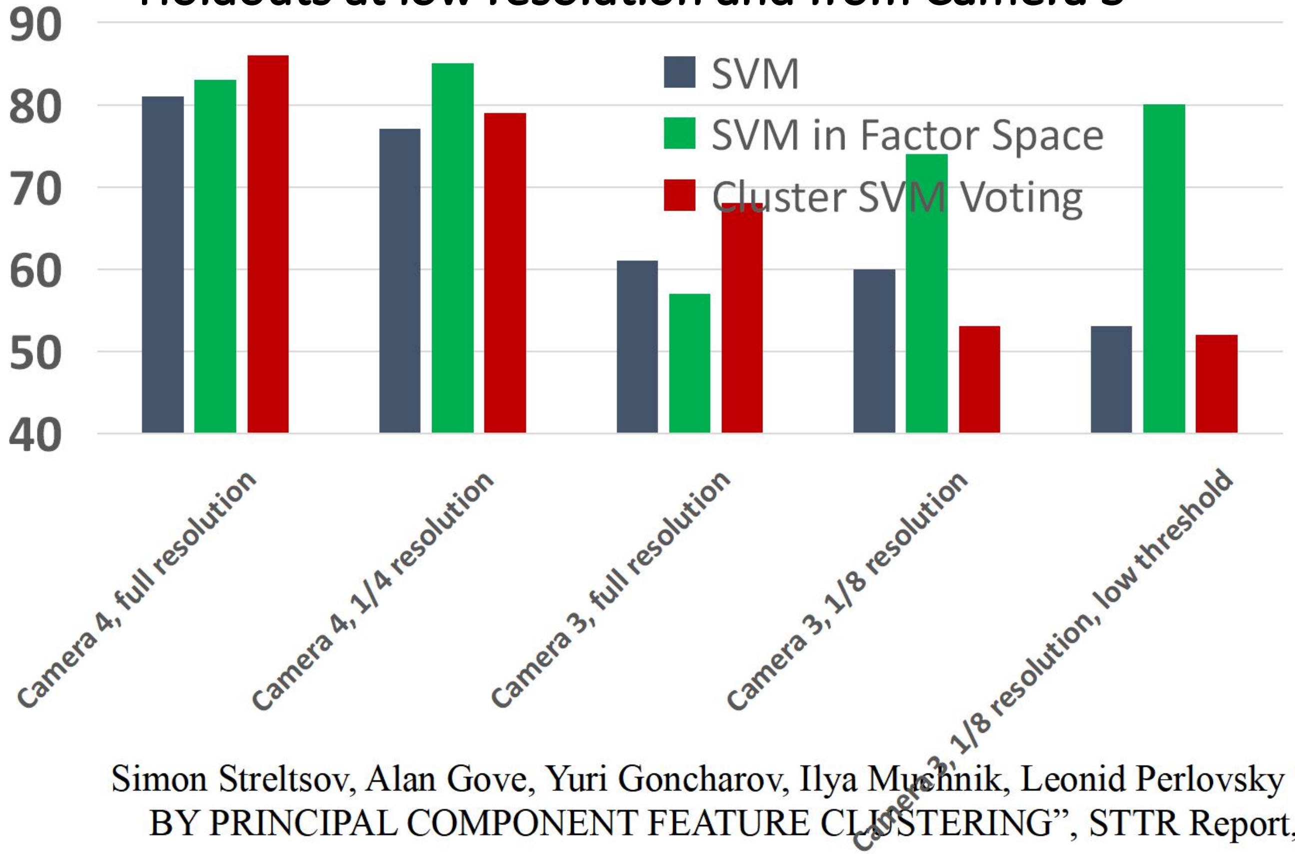
Lopez-Paz, D., Bottou, L., Schölkopf, B., Vapnik, V. (2015). Unifying distillation and privileged information. https://arxiv.org/abs/1511.03643

Ilin R., Izmailov R., Goncharov Y., Streltsov S. (2016) “Fusion of privileged features for efficient classifier training Information”, 19th International Conference on Fusion, Heidelberg, Germany. https://ieeexplore.ieee.org/abstract/document/7528045/

Feature Clustering

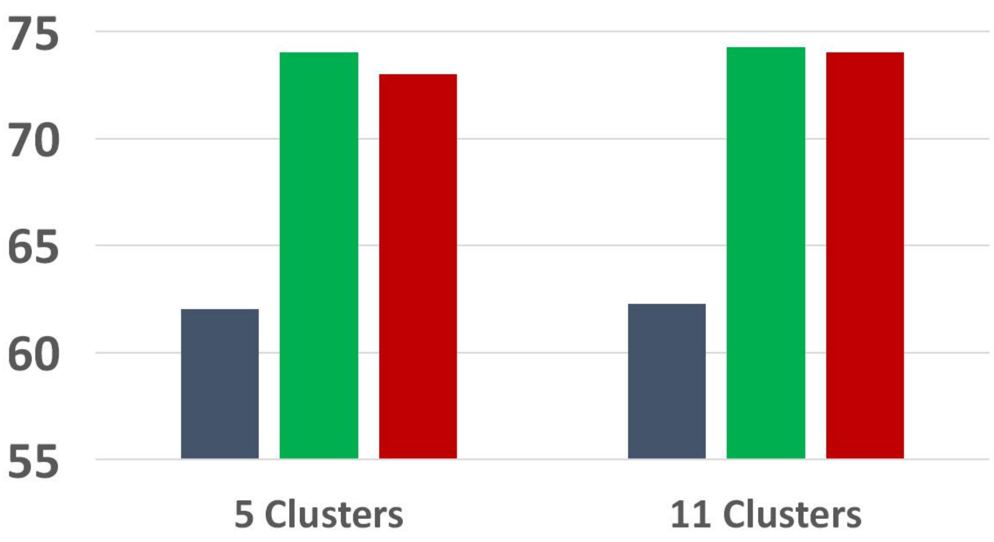
- Divides large feature sets into clusters of highly correlated features
- Increases robustness to changes in the data

Moving person detection in low-res video
Train, test using Camera 4, full resolution
Holdouts at low resolution and from Camera 3



Very small training set

4 Positives
4 Negatives



Simon Streltsov, Alan Gove, Yuri Goncharov, Ilya Muchnik, Leonid Perlovsky "ENHANCING MOTION IMAGERY CLASSIFIERS BY PRINCIPAL COMPONENT FEATURE CLUSTERING", STTR Report, AFRL-RY-WP-TR-2016-0023

Domain Adaptation (DA)

Machine Learning (ML) decisions are risky when new data can unexpectedly change.

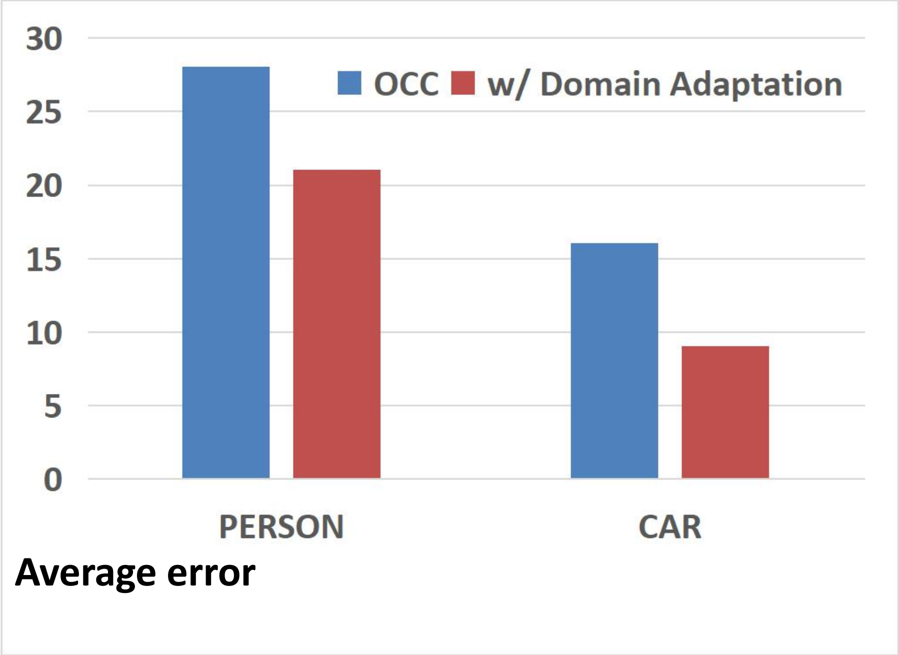
ML does not provide explicit performance guarantees when things change!

ML breaks software engineering paradigm! Weak contract.

DA evaluates distribution differences between new and training data without using labels and adjust decision rules to the new distribution.

- Can be used to train on different data
- Use data from different sensors
- A way to utilize synthetic data given to academia

Domain Adaptation with One Class Classifier (OCC)
900 CIFAR CNN features
Unbalanced multi-class data set



Leon Bottou, invited talk, ICML, 2015. https://icml.cc/2015/invited/LeonBottouICML2015.pdf)
Tzeng, Eric, Judy Hoffman, Kate Saenko, and Trevor Darrell. "Adversarial discriminative domain adaptation." In Computer Vision and Pattern Recognition (CVPR), vol. 1, no. 2, p. 4. 2017.