Learning from Small Sample Sets by Combining Unsupervised Meta-Training with CNNs

Carnegie Mellon University — The Robotics Institute

Yu-Xiong Wang and Martial Hebert

Email: {yuxiongw, hebert}@cs.cmu.edu

Motivation

- **Transferability of Supervised CNNs**: Negatively affected by the specialization of top layer units to their original task → decouple these units from such ties
- **Unsupervised Meta-Training**: Original tiny sampling biased to a selection of categories → a massive set of unlabeled images as a much less biased sampling
- **More Generic, Richer Description**: Diverse sets of separations discriminating the data manifold from its surroundings in all non-manifold directions [Bengio]
- **Structure/Manifold Assumption**: Encourage multiple top layer units to generate low-density separators that do not cross high-density regions

Quasi-Classes Visualization

Conclusions and Future Work

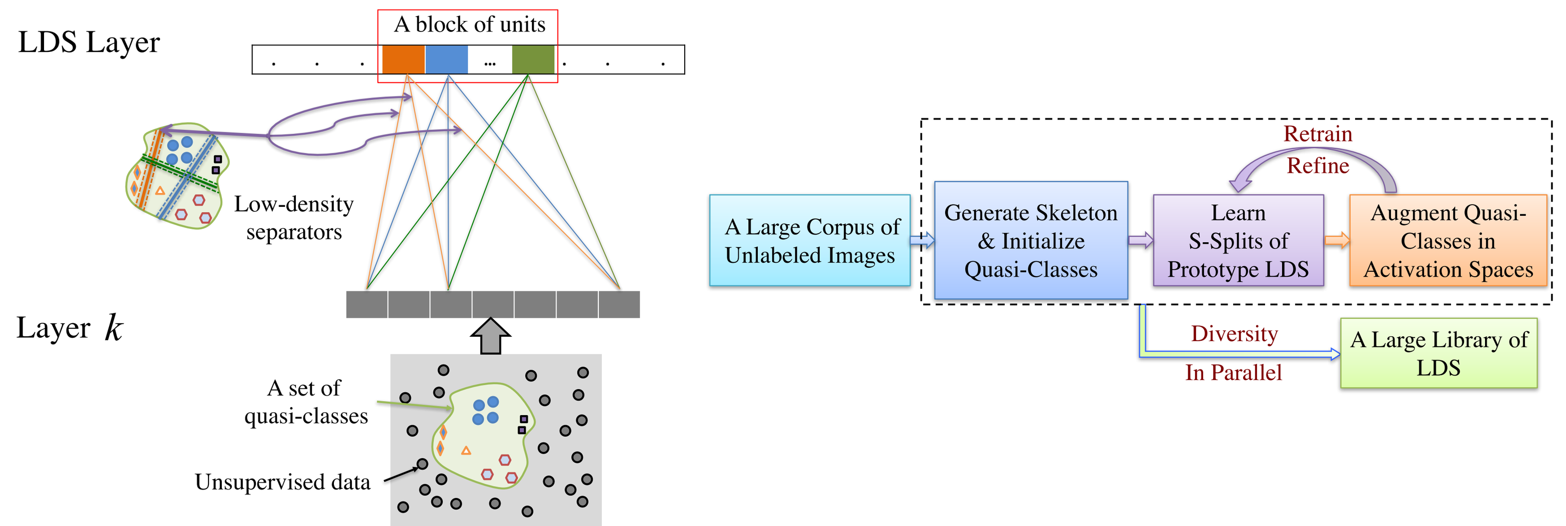
- Structure learning in a large set of unlabeled real-world images to improve the overall transferability of supervised CNNs
- Combination of supervised and unsupervised learning to facilitate the recognition of novel categories from few examples
- Integration into the current CNN backpropagation framework both learning low-density separators and gradually estimating high-density quasi-classes

Unsupervised Meta-Training of Low-Density Separators

- **Approach Overview**: Seeking low-density separators (LDS) while identifying high-density quasi-classes (HDQC)

$$\text{find} \quad \boldsymbol{W} \in \text{LDS}, \ \boldsymbol{T} \in \text{HDQC}$$
$$\text{subject to} \quad \boldsymbol{W} \text{ separate } \boldsymbol{T}$$

  - **Unlabeled Data Corpus**: Yahoo/Flickr 100-million
  - **Feature Space**: Activation space of layer k of a pre-trained ImageNet CNN
  - **Unsupervised Margin Maximization**: A vector of weights ↔ a separator or decision boundary in the activation space



- **Learning Low-Density Separators**: Generalization of supervised predictable discriminative binary codes [Rastegari et al.]

$$\min_{\boldsymbol{W},\boldsymbol{L},\boldsymbol{\Phi}} \sum_{s=1}^{S}\|\boldsymbol{w^s}\|^2+\eta\sum_{i=1}^{N}\sum_{s=1}^{S} I_i\left[1-l_i^s\left(\boldsymbol{w^s}^T\boldsymbol{x_i}\right)\right]_+ + \frac{\lambda_1}{2}\sum_{c=1}^{C}\sum_{\substack{u=1\\v=1}}^{N} T_{c,u}T_{c,v}d\left(\boldsymbol{\phi}_u,\boldsymbol{\phi}_v\right) - \frac{\lambda_2}{2}\sum_{c'=1}^{C}\sum_{\substack{c''=1\\c''\neq c'}}^{C}\sum_{\substack{p=1\\q=1}}^{N} T_{c',p}T_{c'',q}d\left(\boldsymbol{\phi}_p,\boldsymbol{\phi}_q\right)$$

- **Generating High-Density Quasi-Classes**: A coarse-to-fine procedure that combines max-min sampling [Dai and Van Gool] and bootstrap learning [Choi et al.]

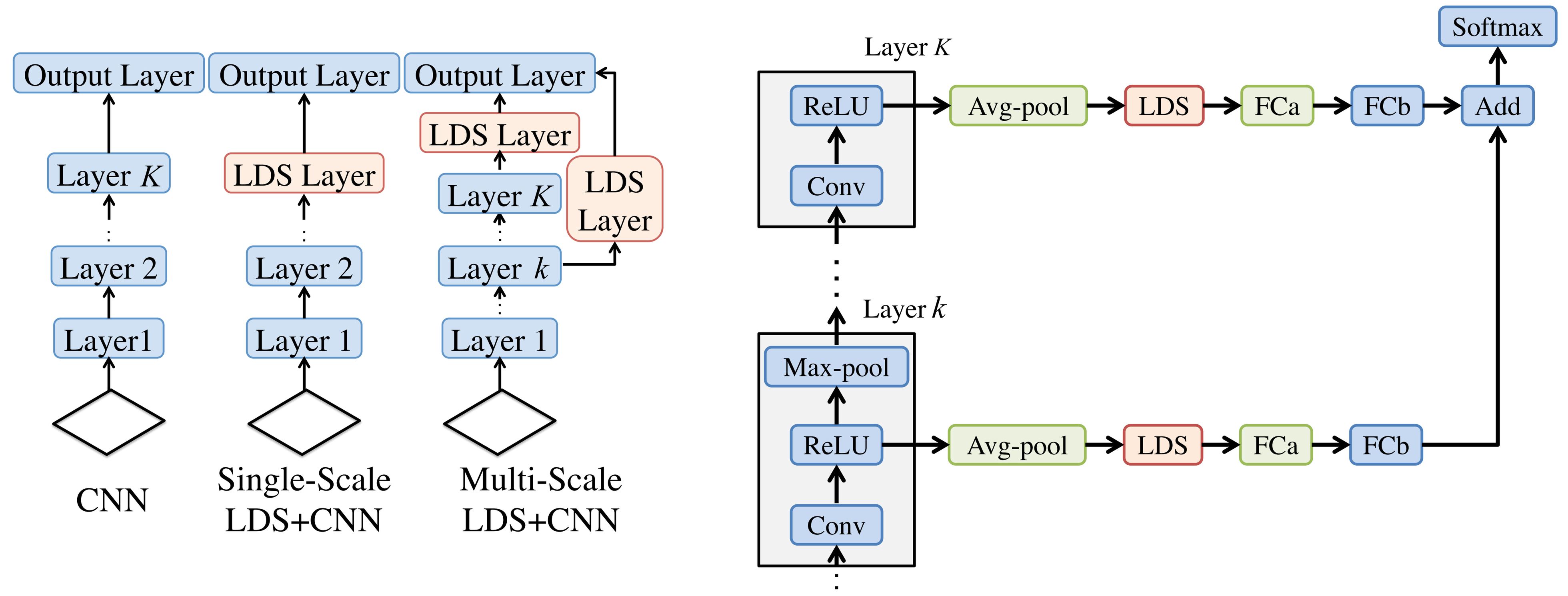
$$\min_{\boldsymbol{T},\boldsymbol{h}_c^{\mathcal{X}},\boldsymbol{h}_c^{\mathcal{F}}} \alpha\sum_{c=1}^{C}\left(\left\|\boldsymbol{h}_c^{\mathcal{X}}\right\|_2^2+\lambda_{\mathcal{X}}\sum_{i=1}^{N} I_i\left[1-y_{c,i}\left(\boldsymbol{h}_c^{\mathcal{X}^T}\boldsymbol{x}_i\right)\right]_+\right)+\sum_{c'=1}^{C}\sum_{\substack{c''=1\\c'\neq c''}}^{C}\sum_{j=1}^{N} T_{c',j}T_{c'',j} + \beta\sum_{c=1}^{C}\left(\left\|\boldsymbol{h}_c^{\mathcal{F}}\right\|_2^2+\lambda_{\mathcal{F}}\sum_{i=1}^{N} I_i\left[1-y_{c,i}\left(\boldsymbol{h}_c^{\mathcal{F}^T}\boldsymbol{\phi}_i\right)\right]_+ - \sum_{j=1}^{N} T_{c,j}\left(\boldsymbol{h}_c^{\mathcal{F}^T}\boldsymbol{\phi}_j\right)\right)$$

$$s.t. \quad \tau_0 \leq \sum_{i=1}^{N} T_{c,i} \leq \tau, \forall c \in \{1,\ldots,C\}$$

> $T_{c,i}=1$ if image $\mathcal{I}_i$ is selected for assignment to quasi-class c and zero otherwise
> $I_i=0$ if $\mathcal{I}_i$ is not selected for assignment to any quasi-class (i.e., $\sum_{c=1}^{C} T_{c,i}=0$) and one otherwise
> $\phi_i^s = f\left(\boldsymbol{w^s}^T\boldsymbol{x_i}\right)$, $f(\cdot)$ is a non-linear function (e.g., ReLU)

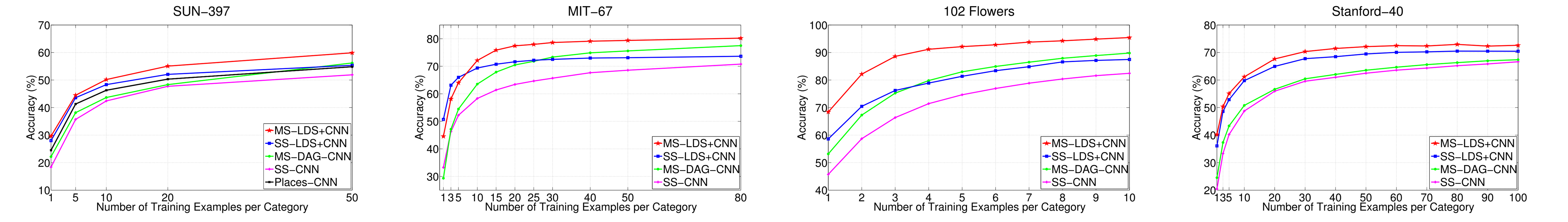
Low-Density Separator Networks

- **Single-Scale Layer-Wise Training**: Break the LDS units into blocks to prevent co-adaptation & enforce diversity
- **Multi-Scale Structure**: Modification of multi-scale DAG-CNN architecture [Yang and Ramanan]
- **SS-LDS+CNN**: LDS with 2,000 blocks of 10 units in activation space of $fc7$ for AlexNet & VGG19
- **MS-LDS+CNN**: LDS in $Conv3$, $Conv4$, $Conv5$, $fc6$, $fc7$ for AlexNet & in $Conv43$, $Conv44$, $Conv51$, $Conv52$, $fc6$ for VGG19



Learning from Few Examples

- **Target Tasks**: Novel category recognition for scene classification | fine-grained recognition | action recognition
- **Evaluation**: VGG19 LDS+CNN & CNN as off-the-shelf features | influence of number of training examples per category



Learning in the Moderate Number of Examples Regime

- Comparison to weakly-supervised CNNs [Joulin et al.]

Type	Approach	SUN-397	MIT-67	102 Flowers	Stanford-40
Weakly-supervised CNNs	Flickr-AlexNet	42.7	55.8	74.2	53.0
	Flickr-GoogLeNet	44.4	55.6	65.8	52.8
	Combined-AlexNet	47.3	58.8	83.3	56.4
	Combined-GoogLeNet	55.0	67.9	83.7	69.2
Ours	SS-LDS+CNN	55.4	73.6	87.5	70.5
	MS-LDS+CNN	**59.9**	**80.2**	**95.4**	**72.6**

- Fine-tuning (AlexNet)