**Discriminatively Activated Sparselets**

Ross Girshick*, Hyun Oh Song*, Trevor Darrell

UC Berkeley, *indicates equal contribution

---

**Goal**
- Shared predictive model with sparse activation vectors
- Efficient inference for linear structured output predictors
- Example application: realtime object recognition in CV/faster retrieval in IR, etc.

**Object recognition**

- Figure showing a general approach for speeding up inference in any coding and dictionary learning.
- Example application: realtime object detection with mixtures of deformable part models.

---

**Sparselet review**

- Set of model filters: $W = \{w_1, ..., w_K\}$
- Set of sparselet filters: $S = \{s_1, ..., s_d\}$

Set factorization point of view

$S = \sum_{i=1}^{d} \sum_{j=1}^{K} a_{ij} s_j w_i \rightarrow \min_{a_{ij}} \sum_{i=1}^{d} \sum_{j=1}^{K} \|a_{ij}\|_1$ subject to $\|a_{ij}\|_0 \leq \epsilon$ for $i = 1, ..., K$ and $\|a_{ij}\|_0 \leq 1$ for $i = 1, ..., d$

Visualized sparselet blocks on HOG

- Left: Sparselet dictionary of size 128
- Right: Top 16 activated sparselets for PASCAL motorcycle class

---

**Joint feature map: multiclass classification**

- Model parameter: sparse activation vector
- Feature: sparselet response

**Joint feature map: object detection**

- Length $n$ window at position $y$ in slot $k$
- Projected feature window from position $y$ installed in slot $k$

---

**Experiments**

- Table showing accuracy comparison:
  - PASCAL VOC 2007: object detection
  - Caltech-256: Image/scene detection (Planar)

- Graphs illustrating performance metrics:
  - Sparsity vs. accuracy
  - Precision/Recall curves for different models

---

**Conclusion**

- Training activation vectors discriminatively significantly outperforms reconstructive training
- Generalized framework for training activation vectors discriminatively using SVM with sparsity constraints

---

**Discriminative sparselet activation**

Original $w_j$ = Sparselet approximation $w_{kj}$

- Intuition: model weights might be composed of smaller building blocks/sets
- For a matrix $X$ with $k$ number of elements, reshape the matrix such that the ratio between the full rank and actual rank of the reshaped matrix is maximized:

$\arg\max_{m, k/m} \min \{m, k/m\}$

- Rank (reshaped $[X, m, k/m]$)

---

**Structural SVM for DAS**

- Parameter vector: $\beta = (\alpha^T, \omega^T)^T$
- Transformed features: $\Phi(x, y) = (\mathbf{e} \cdot S, \ldots, \mathbf{e} \cdot S)^T$
- Aggregate feature vector:

$\Phi(x, y) = (\Phi_1(x, y), ..., \Phi_K(x, y))$

- Training

$\hat{\beta} = \arg\min_{\beta} R(\beta) + \frac{1}{2} \|\omega\|^2 + \frac{1}{2} \sum_{i=1}^{K} \max \{\beta^T \Phi_i(x, y) \Delta(y, y) - \beta^T \Phi_i(x, y)\}$

- Sparsity inducing norm

---

**Sparsity enforcing norms**

- I. Lasso penalty: $R_{lasso}(\alpha) = \lambda_1 \|\alpha\|_1$
- II. Elastic net penalty: $R_{en}(\alpha) = \lambda_1 \|\alpha\|_1 + \lambda_2 \|\alpha\|_2$
- III. Combined $t_0$ and $t_2$ penalty: $R_{ct}(\alpha) = \lambda_1 \|\alpha\|_1 + \gamma \|\alpha\|_2$ subject to $\|\alpha\|_0 \leq t_0$
- A. Overshoot, rank, and threshold (ORT)
- B. Orthogonal matching pursuit (OMP)
- C. Reconstructive

---

**Joint feature map: object detection**

- Inference: $f_{w_k}(x) = \arg\max_{k \in \{1, ..., K\}} \beta^T \Phi(x, y)$
- Feature: $\Phi(x, y) = (\mathbf{e} \cdot S, \ldots, \mathbf{e} \cdot S)^T$

---

**Joint feature map: multiclass classification**

- Inference: $f_{w_k}(x) = \arg\max_{k \in \{1, ..., K\}} \beta^T \Phi(x, k)$
- Feature: $\Phi(x, k) = (0, ..., 0, x^T, 0, ..., 0)^T$