Sparse coding with stel dictionaries

Jeroen Chua
University of Toronto
10 King's College Rd.
Toronto, Ontario M5S 3G4
Canada
jeroen@psi.utoronto.ca

Brendan Frey
University of Toronto
10 King's College Rd.
Toronto, Ontario M5S 3G4
Canada
frey@psi.utoronto.ca

In the computer vision community, sparse coding [3] has become popular in object recognition frameworks as a means to learn and encode image features. A potential drawback of sparse coding, however, is that many codewords may represent similar local structures and differ only up to a change in color (e.g., uniform patches of red, green, blue). This kind of redundancy leads to an inefficient image feature representation since the number of dictionary elements must increase exponentially with the number of color channels. This can make feature learning and inference slow, but more importantly, can also cause patches with similar local structure (e.g., horizontal lines) but different coloring to appear very different in the feature representation, which is often undesirable. Figure 1 shows a dictionary learned using sparse coding on color images showing dictionary elements with redundant local structure.

We propose a framework in which sparse coding over dictionaries with explicit color models can be performed. Structure element (stel) models [2, 1] have previously been shown to be effective at constructing dictionaries with explicit color models. In this work, we apply the stel model framework to patches for dictionary learning, where each element represents local structure, and propose a modification to the standard sparse coding framework to allow for inference of the sparse coefficients over this dictionary. In our approach, we allow a structured transformation, T_p, to be applied to the dictionary for each image patch to explain away the effects of coloring in a patch. This structured deformation allows us to take a dictionary with a color model, and "color it in" in order to best explain the patch under evaluation. We use the following modified sparse coding formulation for inference:

$$[T_p^*, \vec{\alpha_p}^*] = \arg\min_{T_p \in \tau, \vec{\alpha_p}} \|\vec{x_p} - T_p(\mathbf{D}, x_p)\vec{\alpha_p}\|^2 + \lambda \|\vec{\alpha_p}\|_1 \tag{1}$$

where $\mathbf{D}$ is the dictionary, τ is a set of permissible structured transformations, p indexes patches, $\vec{x_p}$ are the observations for patch p, $\vec{\alpha_p}$ is the vector of sparse coefficients for patch p, and λ is the sparse penalty. Note that the transformation, T_p, is for each patch p, and that setting $T_p(\mathbf{D}, x_p) = \mathbf{D}$ leads to the standard sparse coding formulation. In this work, we show how to jointly infer the structured transformation, T_p, and the coefficients $\vec{\alpha_p}$ so as to perform sparse coding over a dictionary with an explicit color model. We demonstrate competitive object recognition results on Caltech28, and a relative improvement of 5% classification rate over previous stel models on Caltech101.

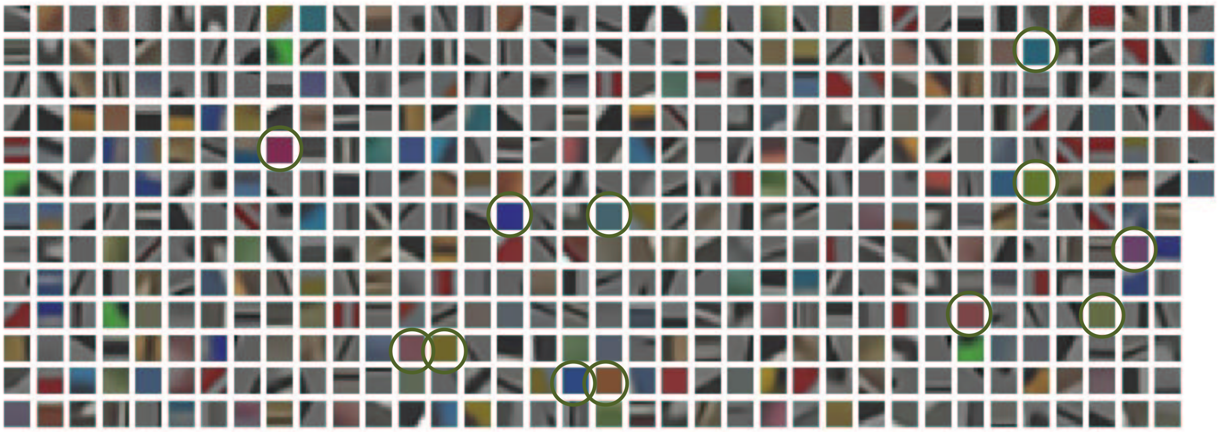

Figure 1: Dictionary elements learned using sparse coding on patches of 8x8 pixels on colored Caltech101 images. Circled are dictionary elements that, structurally, all represent the uniformly colored patch. These elements are highly redundant in that they only differ in their coloring, but not local structure.

Acknowledgements

We would like to thank Dr. Inmar Givoni (University of Toronto) and Professor Ryan Adams (Harvard University) for their helpful collaborations on this work.

References

[1] N. Jojic and Y. Caspi. Capturing image structure with probabilistic index maps. In *CVPR*, 2004.

[2] N. Jojic, A. Perina, M. Cristani, V. Murino, and B. J. Frey. Stel component analysis: Modeling spatial correlations in image class structure. In *Proceedings of CVPR*, pages 2044–2051, 2009.

[3] B. A. Olshausen and D. J. Fieldt. Sparse coding with an overcomplete basis set: a strategy employed by v1. *Vision Research*, 37:3311–3325, 1997.

Topic: vision
Preference:oral/poster