

Fast and Flexible Convolutional Sparse Coding

Felix Heide^{1,3}, Wolfgang Heidrich^{2,3}, Gordon Wetzstein¹

¹Stanford University ²KAUST ³UBC

Convolutional sparse coding (CSC) has become an increasingly important tool in machine learning and computer vision. Image features can be learned and subsequently used for classification and reconstruction tasks. As opposed to patch-based methods, convolutional sparse coding operates on whole images, thereby seamlessly capturing the correlation between local neighborhoods. Grosse et al. [3] were the first to propose a frequency domain method for 1D audio signals, while [1, 2] later demonstrate efficient frequency domain approaches for 2D image data. While this is the first step towards making CSC practical, these frequency methods can introduce boundary artifacts for both learning and reconstruction [4] and, as inherently global approaches, make it difficult to work with incomplete data.

In this paper, we propose a new approach to solving CSC problems. We propose a new splitting-based approach to convolutional sparse coding and show that our method converges significantly faster and also finds better solutions than the state of the art. In addition, the proposed method is the first efficient approach to allow for proper boundary conditions to be imposed and it also supports feature learning from incomplete data as well as general reconstruction problems.

We propose the following, general formulation for convolutional sparse coding:

$$\begin{aligned} \underset{\mathbf{d}, \mathbf{z}}{\operatorname{argmin}} \quad & \frac{1}{2} \|\mathbf{x} - \mathbf{M} \sum_{k=1}^K \mathbf{d}_k * \mathbf{z}_k\|_2^2 + \beta \sum_{k=1}^K \|\mathbf{z}_k\|_1 \\ \text{subject to} \quad & \|\mathbf{d}_k\|_2 \leq 1 \quad \forall k \in \{1, \dots, K\}. \end{aligned} \quad (1)$$

where \mathbf{z}_k are sparse feature maps that approximate the data term \mathbf{x} when convolved with the corresponding filters \mathbf{d}_k of fixed spatial support. \mathbf{M} is a diagonal or block-diagonal matrix, such that it decouples linear systems of the form $(\mathbf{M}^T \mathbf{M} + \mathbb{I})\mathbf{x} = \mathbf{b}$ into many small and independent systems that are efficiently solved. This allows us to use unmodified filters in boundary regions, thus preserving the convolutional nature of the problem without requiring circular boundaries or other conditions. Furthermore, we show that \mathbf{M} allows for efficient learning and reconstruction from incomplete data.

To efficiently solve (1), we reformulate it as the following sum of functions in (2).

$$\begin{aligned} \underset{\mathbf{d}, \mathbf{z}}{\operatorname{argmin}} \quad & f_1(\mathbf{D}\mathbf{z}) + \sum_{k=1}^K (f_2(\mathbf{z}_k) + f_3(\mathbf{d}_k)), \quad \text{with} \\ f_1(\mathbf{v}) = \quad & \frac{1}{2} \|\mathbf{x} - \mathbf{M}\mathbf{v}\|_2^2, f_2(\mathbf{v}) = \beta \|\mathbf{v}\|_1, f_3(\mathbf{v}) = \operatorname{ind}_C(\mathbf{v}), \end{aligned} \quad (2)$$

where \mathbf{D} is the matrix representing the sum of convolutions with all filters. The splitting into different functions, which might be unintuitive on the first sight, leads to an efficient optimization method that separates the filtering via \mathbf{D} from the masking operator \mathbf{M} in f_1 . The subproblem involving filtering can then be solved efficiently in the spectral domain, while the the subproblem involving \mathbf{M} can be solved in the spatial domain. To achieve this, we derive an optimization method for the general sum-of-function objectives from (3).

$$\underset{\mathbf{z}}{\operatorname{argmin}} \quad \sum_{i=1}^I f_i(\mathbf{K}_i \mathbf{z}), \quad (3)$$

where $\mathbf{K}_i : \mathbb{R}^{b_i \times a_i}$ are arbitrary matrices, $f_i : \mathbb{R}^{b_i} \rightarrow \mathbb{R}$ are closed, proper, convex functions, and $i \in \{1, \dots, I\}$, such that $f_i(\mathbf{K}_i \cdot) : \mathbb{R}^{a_i} \rightarrow \mathbb{R}$. For the objective from (2) we derive a fast and flexible ADMM-based method which then ultimately is specialized to solve (2).

For the popular datasets from [5], we plot the empirical convergence of the proposed algorithm and compare it to the state of the art in Fig. 1. In both cases we learn $K = 100$ filters. Our method outperforms recent methods [1, 2] by a large margin (even when the others are augmented with our factorization strategy).

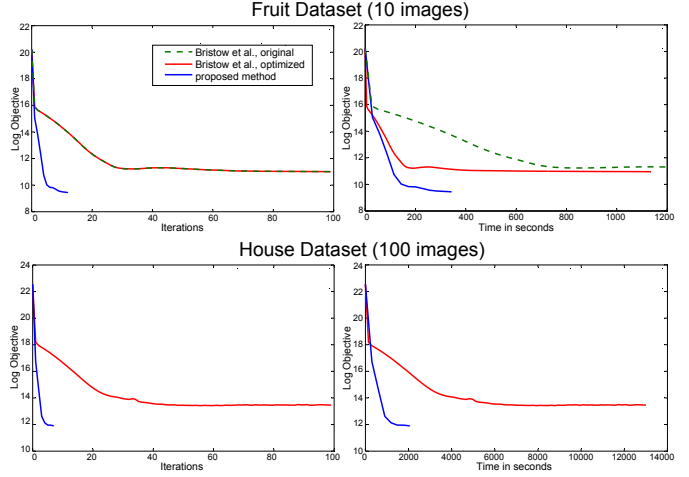


Figure 1: Convergence for two datasets (left $N = 10$ images, right $N = 100$). The proposed algorithm converges to a better solution in less time than competing methods.

Fig. 1 also shows that our method does in fact find *lower objective* to the non-convex CSC problem. Fig. 2 shows the resulting filters after convergence (ours after 13 iterations, Bristow after 300 iterations).



Figure 2: Filters learned on city dataset [5]. Filters learned with our method (left) and with that described in [1, 2] (right). Our method finds a local optimum with objective 3–4× lower than comparable methods.

In summary, we propose a new method for learning and reconstruction problems using convolutional sparse coding. Our formulation is flexible in allowing for proper boundary conditions, it allows for feature learning from incomplete observations, or any type of linear operator applied to the estimation. We demonstrate that our framework is faster than the state of the art and converges to better solutions.

- [1] Hilton Bristow and Simon Lucey. Optimization methods for convolutional sparse coding. *arXiv:1406.2407*, 2014.
- [2] Hilton Bristow, Anders Eriksson, and Simon Lucey. Fast convolutional sparse coding. In *Proc. CVPR*, pages 391–398, 2013.
- [3] Roger B. Grosse, Rajat Raina, Helen Kwong, and Andrew Y. Ng. Shift-invariance sparse coding for audio classification. In *Proc. UAI*, pages 149–158, 2007.
- [4] Koray Kavukcuoglu, Pierre Sermanet, Y-Lan Boureau, Karol Gregor, Michaël Mathieu, and Yann LeCun. Learning convolutional feature hierarchies for visual recognition. In *Proc. NIPS*, 2010.
- [5] Matthew D Zeiler, Dilip Krishnan, Graham W Taylor, and Robert Fergus. Deconvolutional networks. In *Proc. CVPR*, pages 2528–2535, 2010.