CONVOLUTIONAL SPARSE CODING IS IMPROVED BY HETEROGENEOUS UNCERTAINTY MODELING

Hugo J. Ladret
Institut de Neurosciences de la Timone
UMR 7289, CNRS and Aix-Marseille Université
Marseille, 13005, France
& School of Optometry
Université de Montréal
Montréal, QC H3C 3J7, Canada
hugo.ladret@univ-amu.fr

Laurent U.Perrinet
Institut de Neurosciences de la Timone
UMR 7289, CNRS and Aix-Marseille Université
Marseille, 13005, France
laurent.perrinet@univ-amu.fr

ABSTRACT

Aleatoric uncertainty characterizes the variability of features found in natural images, and echoes the epistemic uncertainty ubiquitously found in computer vision models. We explore this "uncertainty in, uncertainty out" relationship by generating convolutional sparse coding dictionaries with parametric epistemic uncertainty. This improves sparseness, resilience and reconstruction of natural images by providing the model a way to explicitly represent the aleatoric uncertainty of its input. We demonstrate how hierarchical processing can make use of this scheme by training a deep convolutional neural network to classify a sparse-coded CIFAR-10 dataset, showing that encoding uncertainty in a sparse code is as efficient as using conventional images, with additional beneficial computational properties. Overall, this work empirically demonstrates the advantage of partitioning epistemic uncertainty in sparse coding algorithms.

1 INTRODUCTION

Sensory processing is constrained by both uncertainty outside and uncertainty within, respectfully designated as aleatoric and epistemic uncertainty (Hüllermeier & Waegeman, 2021). The former arises due to inherent stochasticity in the structure of the inputs to a sensory system, and is a characteristic property of many naturalistic inputs, such as sounds (Nakamura & Nakadai, 2015), haptics (Pettypiece et al., 2010) and images (Ruderman, 1994). This aleatoric uncertainty is challenging to predict, as it often relies on factors that are outside possible measurements or control, and has proven to be an arduous modeling challenge to computer vision (Gousseau & Morel, 2001). On the other hand, epistemic uncertainty arises due to the lack of knowledge about the global input and model system, and can stem from various sources, the foremost being due to inputs that have ambiguous visual parameters properties, such as lighting conditions, object pose, or orientation. This creates aleatoric uncertainty in the input, which then results in epistemic uncertainty in the model (Coppola et al., 1998).

An optimal policy to describe the relationship between the two types of uncertainty is crucial to maximize the performance of a model, and to try to minimize decision uncertainty. Certain frameworks, such as Bayesian modeling, offer an explicit rule to connect the squared inverse of both uncertainties (i.e., precisions) and guide the model's updates accordingly, which is well-accounted for by neuroscientific models of sensory processing (Helmholtz, 1924; Orbán et al., 2016; Hénaff et al., 2020; Ladret et al., 2022). Yet, even if a model does not explicitly take into account aleatoric uncertainty, it cannot escape being under its implicit constraint. For instance, fundamental machine learning models such as sparse coding (Olshausen & Field, 1996), that forms the basis of many computer vision algorithms (Zhang & Ghanem, 2018), display parameters (dictionaries) that are strikingly similar to their biological counterparts (Olshausen & Field, 1997). Notably, they display heterogeneous epistemic uncertainty, feature positioning and scaling (Figure 1). Position- and scale-free representation of features provides the basis for convolutional sparse algorithms, yet these models cannot represent a single feature at multiple uncertainty levels.

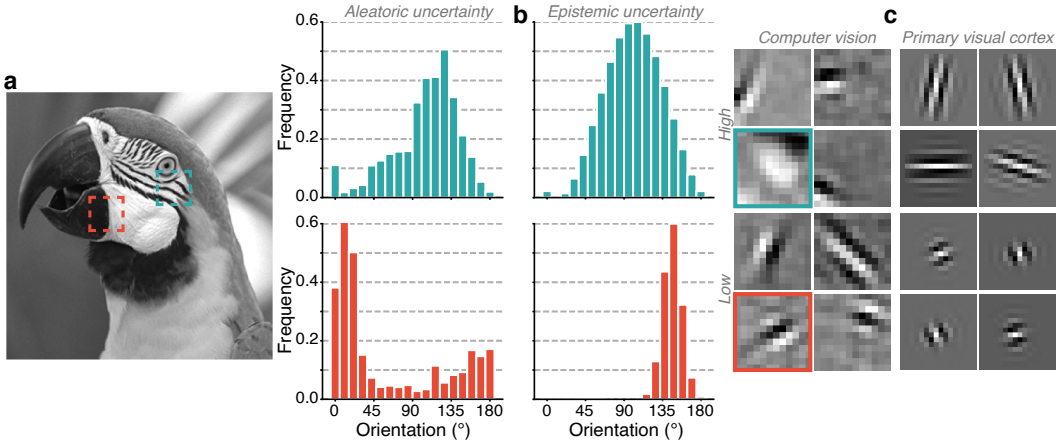


Figure 1: Aleatoric uncertainty reflects epistemic uncertainty. **(a)** High (blue) and low (red) aleatoric uncertainty in the feature (orientation WRT the vertical) space from two patches of a natural image (using a Histogram of Oriented Gradients on 32×32 pixel patches from an image of the Kodak dataset (Franzen, 2013)). **(b)** High and low epistemic uncertainty emerging by learning statistics of natural images (Olshausen & Field, 1996) using a Sparse Coding algorithm. **(c)** Comparison with receptive fields of cat primary visual cortex.

Here, we sought to alleviate this shortcoming, by showing how convolutional sparse coding might benefit from factoring-in uncertainty. We present a generative model of dictionary with parametrized feature uncertainty. This shows that heterogeneous epistemic uncertainty in features dictionaries improves sparsity, PSNR and resilience of convolutional sparse coding algorithms. We then provide evidence that this structure also naturally emerges from training on a dataset of natural images, supporting the idea that epistemic uncertainty is causal and advantageous when dealing with datasets that contain high levels of aleatoric uncertainty. Finally, we show that sparse-coded natural images can be re-used as inputs to deep neural networks, boosting performance and providing resilience to input degradation.

2 Background

2.1 Sparse coding

Sparse coding (SC) is a widely used model for learning the inverse representation of an input signal (Lee et al., 2006). Given the assumption that a signal can be represented as a linear mixture of basis functions, the optimization problem solved by sparse coding is one that tries to minimize the number of basis functions that are used to represent the input signal, yielding a compact and efficient representation of the original signal (Perrinet, 2015). Here, we focused on sparse coding as the problem of reconstructing an image $\boldsymbol{S}$ from sparse representations x while minimizing a ℓ_1 norm of the representation. This problem can be formulated as:

$$\underset{x}{\operatorname{argmin}} \frac{1}{2}||\boldsymbol{S} - \mathbb{D}x||^2 + \lambda||x||_1 \tag{1}$$

where $\mathbb{D}$ is a dictionary (i.e. a set of basis functions used to represent $\boldsymbol{S}$) and λ a regularization parameter that controls the trade-off between fidelity and sparsity. This problem can be approached with a Basis Pursuit DeNoising (BPDN) algorithm (Chen et al., 2001).

Convolutional sparse coding (CSC) uses dictionary elements (kernels) that are spatially localized and replicated on the full input space, resulting in convolutional kernels. The number of basis functions in the dictionary defines the number of features, or *channels*, which is multiplied by the number of position, compared to standard SC. As a result, a convolution allows to explicitly represent the spatial structure of the signal to be reconstructed. This further allows reducing the number of basis functions required to achieve a good performance of the image whilst providing shift-invariant representations. As the convolution is a linear operator, CSC problems can be solved with convolutional

BPDN algorithms (Wohlberg, 2015) by extending equation (1):

$$\underset{\{x_k\}}{\operatorname{argmin}} \frac{1}{2}||\boldsymbol{S} - \sum_{k=1}^{K} d_k * x_k||^2 + \lambda \sum_{k=1}^{K} ||x_k||_1 \tag{2}$$

where x_k is an N^2 dimensional coefficient map (given an N^2 sized image), d_k is one kernel (among K channels) and $*$ is the convolution operator. Here, we used the Python SPORCO package (Wohlberg, 2017) to implement sparse coding methods, using an Alternating Direction Method of Multipliers (ADMM) algorithm (Wang et al., 2019) which splits Convolutional Sparse Coding problems into two sub-problems. In our setting, by introducing an auxiliary variable Y (Wohlberg, 2014), the ADMM problems can be decomposed into a standard form:

$$\underset{x,y}{\operatorname{argmin}} f(x) + g(y) \tag{3}$$

with the constraint $x = y$. ADMM can be readily applied to equation (2), such that the BPDN problem becomes:

$$\underset{\{x_k\},\{y_k\}}{\operatorname{argmin}} \frac{1}{2}||\sum_{k=1}^{K} \mathrm{d}_k * x_k - \boldsymbol{S}||_2^2 + \lambda \sum_{k=1}^{K} ||y_k||_1 \text{ s.t. } \mathrm{x}_k = \mathrm{y}_k \tag{4}$$

which is solved by alternating between the two equations:

$$\{x_k\}_{i+1} = \underset{\{x_k\}}{\operatorname{argmin}} \frac{1}{2}||\sum_{k=1}^{K} \mathrm{d}_k * x_k - \boldsymbol{S}||^2 + \frac{\rho}{2}||x_k - y_{k,i} + \mathrm{u}_{k,i}||^2 \tag{5}$$

$$\{\mathrm{y}_k\}_{i+1} = \underset{\{y_k\}}{\operatorname{argmin}} \lambda \sum_{k=1}^{K} ||y_k||_1 + \frac{\rho}{2}||x_{k,i+1} - y_k + \mathrm{u}_{k,i}||^2 \tag{6}$$

where ρ is a penalty parameter that controls the convergence rate of the iterations. x and y are residuals whose equality is enforced by the prediction error :

$$\mathrm{u}_{k,i+1} = \mathrm{u}_{k,i} + x_{k,i+1} - y_{k,i+1} \tag{7}$$

CSC was performed on the "targe" subset of the database from Serre et al. (2007), which consists of 600 high-quality color images (256×256 pixel size) that were grayscaled and high-passed filtered.

2.2 DICTIONARIES

We built dictionaries made of localized, oriented elements (Olshausen & Field, 1996) in a parametric fashion, using a set of log-Gabor filters that captures the log-frequency structure of the image and ensure its optimal reconstruction (Fischer et al., 2007a). Each filter is defined in the frequency plane by polar coordinates (f, θ) by a bi-dimensional log-Gabor filter (Fischer et al., 2007b), whose envelope is given by:

$$G(f, \theta) = \exp\left(-\frac{1}{2} \cdot \frac{\log(f/f_0)^2}{\log(\sigma_f/f_0)^2}\right) \cdot \exp\left(\frac{\cos(2 \cdot (\theta - \theta_0))}{4 \cdot \sigma_\theta^2}\right) \tag{8}$$

where f_0 is the center frequency, σ_f the bandwidth parameter for the frequency, θ_0 the center orientation and σ_θ the standard deviation for the orientation. We kept $f_0 = \sigma_f = 0.4$ cpd and varied only the orientation parameters to build the dictionaries. From σ_θ (in octaves), we defined the angular bandwidth (in degrees) of the log-Gabor filter as $B_\theta = \sigma_\theta \sqrt{2 \log 2}$ (Swindale, 1998).

Dictionaries could also undergo learning based on the dataset, in which case convolutional sparse coding was alternated with a dictionary update equation in a multi-image setting:

$$\underset{\{x_k,j\}}{\operatorname{argmin}} \frac{1}{2} \sum_{j=1}^{J} ||\sum_{k=1}^{K} \mathrm{d}_k * x_{k,j} - \boldsymbol{S}_j||^2 + \lambda \sum_{k}^{K} \sum_{j}^{J} ||x_{k,j}||_1 \text{ s.t. } \forall k, ||\mathrm{d}_k||_2 = 1 \tag{9}$$

where $\boldsymbol{S}_j$ is the j-th image in the dataset and $x_{k,j}$ is the coefficient map for the k-th filter and the j-th image. We evaluated the performance of the convolutional sparse coding algorithm with two

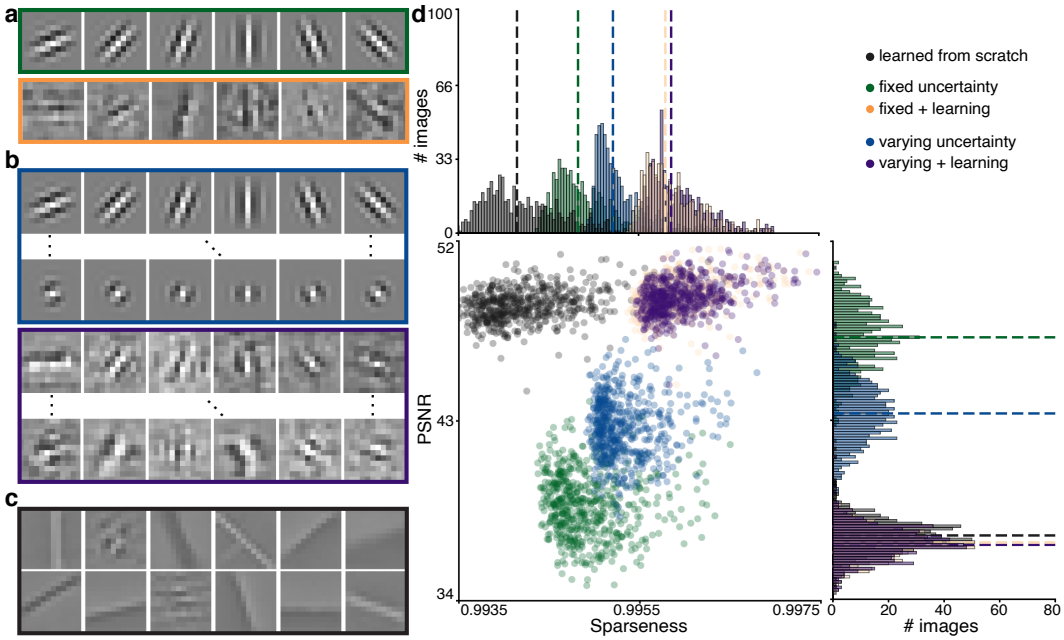


Figure 2: Epistemic uncertainty in a CSC dictionary improves both sparseness and reconstruction performance. **(a)** Elements from dictionaries with fixed epistemic uncertainty before (green) and after dictionary learning (orange). **(b)** Elements from a dictionary with heterogeneous epistemic uncertainty before (blue) and after dictionary learning (purple). **(c)** Elements from a dictionary learned from scratch. **(d)** Distribution of the sparseness (top) and Peak Signal-to-Noise Ratio (PSNR, right) of the five dictionaries, shown as a scatter plot for each of the 600 images of the dataset (center). Median values are shown as dashed line on the histograms.

metrics. The reconstruction quality was measured using the Peak Signal-to-Noise Ratio (PSNR), which for grayscale images used here is defined as:

$$\text{PSNR}(I_1, I_2) = 20\log_{10}(\max I_1) - 10\log_{10}\left(\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(I_1 - I_2)^2\right) \tag{10}$$

where $(\max \mathrm{I}_1)$ is the maximum pixel intensity of the source image. The right hand-side term of the PSNR is the $\log_{10}$ of the mean squared error, where I_1 and I_2 represent the pixel intensity in the source and reconstructed images, respectively. We also measured the sparseness of the algorithm, defined as the ratio of zero/non-zero coefficients used in reconstruction.

2.3 Image classification using deep learning

We sparse-coded the images from the CIFAR-10 dataset, which consists of 32 color images, making up 10 classes of 6000 images. Images were upscaled to a 128×128 resolution with bilinear interpolation, then grayscaled. After sparse coding, they were used as inputs to a convolutional neural network from the Visual Geometry Group (Simonyan & Zisserman, 2014) with 16-layers configuration (VGG-16), which was retrained from scratch using a standard PyTorch implementation. We split the data into a training set of 50000 images and a test set of 10000 images, and trained VGG-16 using stochastic gradient descent, using a learning rate of 0.001 and a momentum of 0.9. The cross-entropy loss function was minimized on batches of 64 sparse-coded images for 100 epochs while applying a weight decay of $5e-4$ for every iteration of the gradient descent to prevent overfitting.

3 RESULTS

3.1 EPISTEMIC UNCERTAINTY IMPROVES PSNR AND SPARSENESS

We used five types of dictionaries to probe the role of epistemic uncertainty in the encoding of natural images, all with similar size and with duplicated atoms over two phases $(0; \pi)$. Two dictionaries were built with Log-Gabor filters: either with one fixed epistemic uncertainty ($B_\theta = 12.0°$) and 72 orientations θ_0 ranging from $0°$ to $180°$ (as shown in Figure 2a, green), or with 12 values of θ_0 but heterogeneous epistemic uncertainty, i.e. 6 values of B_θ from $3°$ to $30°$ (Figure 2b, blue). These two dictionaries were compared with versions that learned features on a dataset of natural images (Figure 2a, orange; b, purple) and with another dictionary learned from scratch on the same dataset (Figure 2c, black). All dictionaries had consistent performance across the 600 images (Figure 2d). The reconstruction with heterogeneous epistemic uncertainty outperformed the reconstruction with fixed epistemic uncertainty in both PSNR ($U = 15503.0$, $p < 0.001$, Mann-Whitney U-test) and sparseness ($U = 66793.0$, $p < 0.001$). The learning procedure nonetheless yielded superior results in terms of PSNR ($U = 93.0$, $p < 0.001$) when comparing the scratch version to both non-learned. This same learning also improved the Log-Gabor dictionaries, resulting in an increase in PSNR ($U = 0.0$, $p < 0.001$; $U = 177128.0$, $p < 0.001$, fixed and heterogeneous epistemic dictionary, respectively) and sparseness ($U = 14520.0, p < 0.001$; $U = 205670.0, p < 0.001$) after learning. The post-learning heterogeneous epistemic dictionary had higher sparseness compared to the post-learned fixed uncertainty dictionary ($U = 205670.0$, $p < 0.001$), but similar PSNR ($U = 177128.0$, $p = 0.31$). As both have comparable reconstruction quality, we will now concentrate on comparing the two Log-Gabor dictionaries prior to learning and the heterogeneous epistemic uncertainty dictionary post-learning. Further information regarding the post-learning version of the dictionary with fixed epistemic uncertainty can be found in Appendix B.

Overall, without learning, the optimal dictionary for encoding natural images contains epistemic uncertainty. Even though learned dictionaries have a higher computational cost, they exhibit more gain in PSNR and sparseness than by simply adding heterogeneous epistemic uncertainty into a dictionary. The learning procedure led to changes in the coefficients of the dictionary, both in terms of features (orientations θ_0) and their epistemic uncertainty (B_θ). Learning from the dataset introduced a new bias towards the cardinal orientations (Figure 2a, upper), which reflects the bias present in natural images (Appelle, 1972), but which was absent from the initial dictionary. The coefficients also became distributed non-uniformly across multiple levels of epistemic uncertainty (Figure 2a, lower). Notably, learning from the dataset resulted in coefficients that were completely unused (i.e. sparseness $= 1$) for higher B_θ (Figure 2d, blue) becoming active (Figure 2d, purple). The resulting patterns of coefficient distribution became consistent over multiple levels of epistemic uncertainty (Figure 2b), aligning with the multiple levels of aleatoric uncertainty present in the dataset. Thus, the improvement in performance resulting from the learning procedure relies on a bias of features θ_0 and a redistribution of epistemic uncertainty B_θ, both of which reflect the structure of the dataset.

3.2 EPISTEMIC UNCERTAINTY BOOSTS RESILIENCE OF THE NEURAL CODE

The coefficients of all dictionaries followed a prototypical Dirac-Laplacian function, enforced by the l_1 norm of the algorithm (see Appendix A). The parameters of the function were highly stereotypical across images for a given dictionary, indicating a common structure, which opens up the possibility of manipulating the dictionary's activation. By utilizing the stereotypical pattern of activation, it is possible to remove the less activated coefficients (described by the Dirac function) to further increase sparseness and evaluate the resilience of the code to the removal of its less used elements. We zero-ed the activation of coefficients whose absolute value fell under a given threshold, iterating from 0.001 to 0.5 in 8 steps. The resulting pruning increased sparseness, which correlated non-linearly with a decrease in PSNR for all dictionaries (Figure 4a). The pre-learning heterogeneous epistemic dictionary's PSNR was significantly more resilient to coefficient degradation than the pre-learning fixed epistemic dictionary ($p < 0.05$ for all pruning levels). Post-learning, both the fixed and heterogeneous epistemic uncertainty dictionary showed similar PSNR for pruning < 0.1, as would be expected from the similar PSNR already found without pruning (Figure 2). However, the post-learning heterogeneous epistemic uncertainty dictionary emerged as the clear winner for all other pruning levels (> 0.1, $p < 0.001$), outperforming all the other dictionaries in the test.

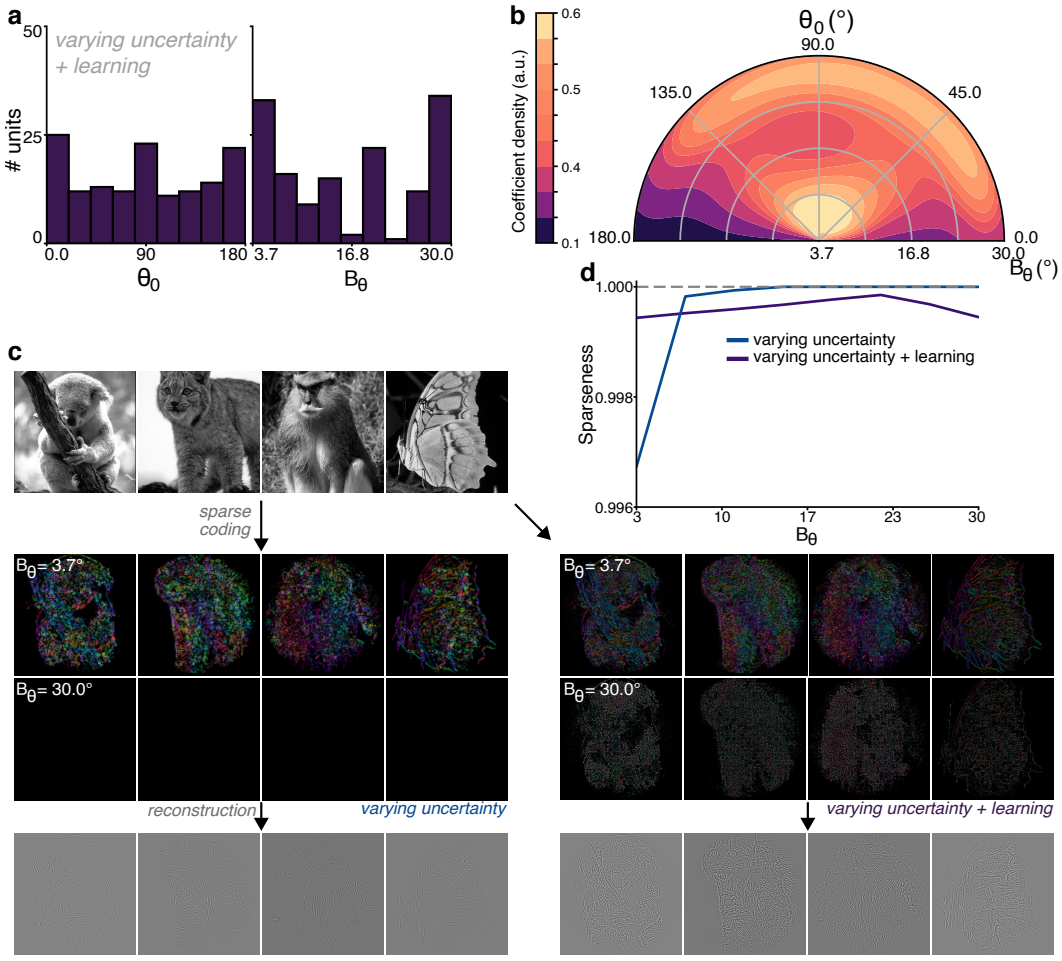


Figure 3: Learning balances coefficient distribution and reflects dataset structure. **(a)** Distribution of coefficients over θ_0 and B_θ after learning. **(b)** Kernel density estimation of the coefficients before and after learning. **(c)** Example images from the dataset (first row), sparse coded (second row, color coded by each coefficient's θ_0) and reconstructed (third row), for pre- and post-learning heterogeneous uncertainty dictionaries. **(d)** Sparseness of the dictionaries as a function of epistemic uncertainty B_θ. Sparseness $= 1$ (i.e. no activation) is represented as a gray dashed line.

This highlights the superiority of the heterogeneous epistemic uncertainty dictionary in terms of resilience and encoding efficiency for natural images.

Overall, these findings demonstrate that epistemic uncertainty in sparse codes exhibits desirable properties : improved reconstruction quality and sparseness (Figure 2d), more evenly distributed activation (Figure 3b) and greater resilience to code degradation (Figure 4a).

3.3 Epistemic uncertainty improves deep neural network performances

We used a deep convolutional neural network trained to classify sparse-coded images, to see whether the desirable properties of the sparse code could improve deep networks. Here, we used VGG-16 to classify with high accuracy up- and gray-scaled images from the CIFAR-10 dataset (see Methods), reaching a maximum top-1 accuracy of 82.27 in 60 epochs (Table 3.3). After sparse coding of the dataset, the dictionary with heterogeneous epistemic uncertainty post-learning had the highest classification accuracy (82.63 top-1 accuracy in 72 epochs). Similarly, at high degradation condition, this dictionary showed the highest resilience, maintaining 80.68 top-1 accuracy, compared to a 75.73 top-1 accuracy for the pre-learned dictionary and a 74.44 top-1 accuracy for the fixed epis-

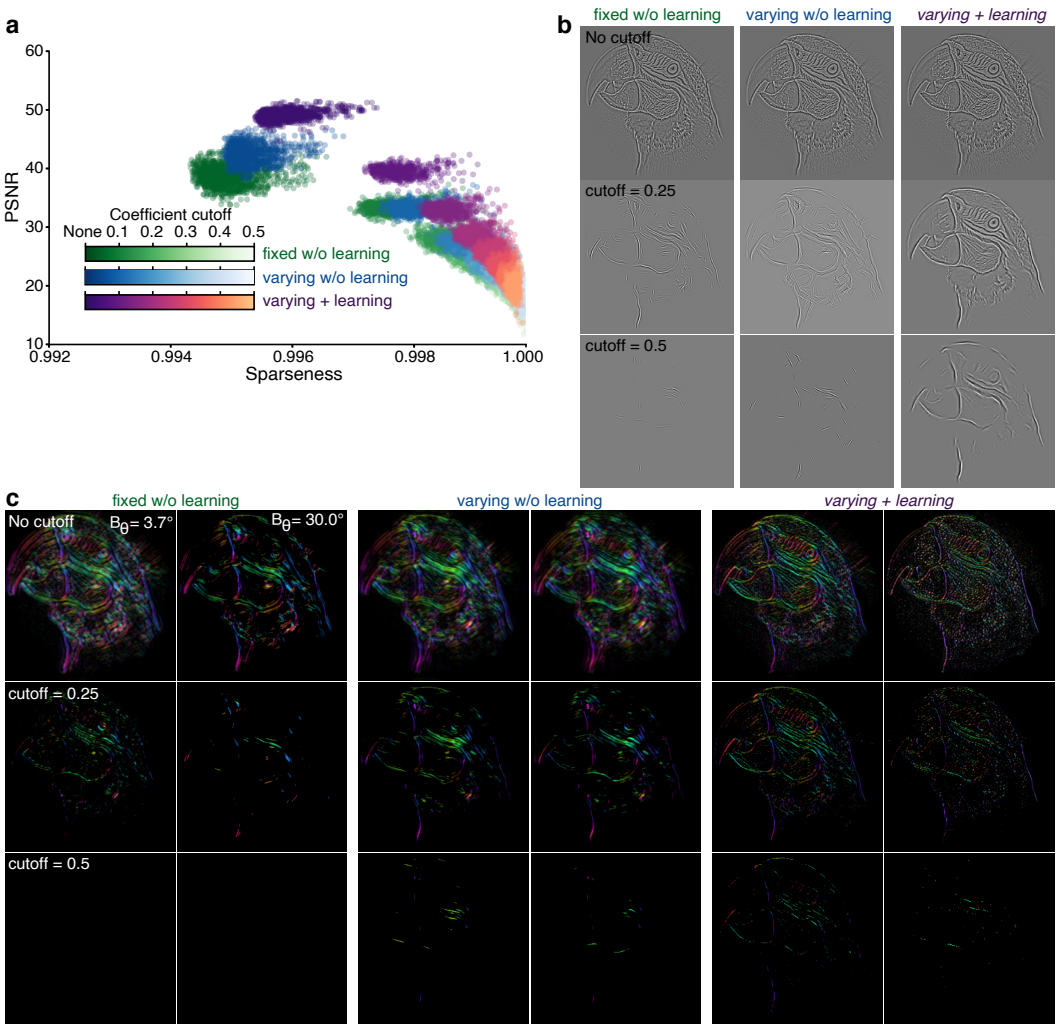


Figure 4: Sparse coefficients can be pruned to boost sparsity. **(a)** Pruning of the coefficients based on their values and resulting sparseness/PSNR for both dictionaries. **(b)** Reconstruction of the image shown in Figure 1 with different cutoff levels. **(c)** Coefficients for the same image, for the minimum and maximum epistemic uncertainty and for different cutoff levels.

Table 1: Top-1 accuracy of VGG-16 for varying CIFAR-10 encoding schemes. $c = 0.2$ and $c = 0.5$ indicate the cutoff of the coefficient's activations, as done in Figure 4.

Encoding scheme	**full accuracy**	**c=0.2 accuracy**	**c=0.5 accuracy**
Upscaled, no sparse coding	82.27		
Heterogeneous, pre-learning	81.69	75.73	67.22
Heterogeneous, post-learning	**82.63**	**80.68**	**74.28**
Fixed, pre-learning	79.71	74.44	67.82
Fixed, post-learning	82.45	80.2	69.14

temic uncertainty dictionary. Thus, learning with epistemic uncertainty provided the best accuracy and resilience, outperforming the classification on the raw dataset. Despite the removal of information from the images by the sparse coding process, the usage of heterogeneous levels of epistemic uncertainty allows outperforming by a small margin the dense classification performance.

4 CONCLUSION

In this study, we investigated the impact of incorporating heterogeneous epistemic uncertainty in a convolutional sparse coding dictionary. Our results demonstrate that this approach outperforms traditional dictionaries with fixed epistemic uncertainty in terms of reconstruction performance, sparseness, and resilience. Additionally, we have shown that these dictionaries can be effectively utilized in subsequent stages of visual processing, leading to improved classification performance in deep networks. These findings suggest that incorporating epistemic uncertainty in sparse coding dictionaries can significantly enhance the encoding and processing of natural images. While the performance of the deep network is lower than the current state-of-the-art (94.71% accuracy), we did not use color images nor specific performance tuning (Zhu et al., 2021), but rather sought to compare the performance of models with different levels of epistemic uncertainty in their first layer. Future directions could also include using the sparse coefficients directly, rather than the reconstructed images, as an input to a deep network. This would enable highly sparse deep learning, which might also prove to be more robust by eliminating low-level noise and making the network more resilient to adversarial attacks (Madry et al., 2017).

ACKNOWLEDGMENTS

Acknowledgements to colleagues and funding agencies will be updated after the double-blind review process.

REFERENCES

Stuart Appelle. Perception and discrimination as a function of stimulus orientation: the" oblique effect" in man and animals. *Psychological bulletin*, 78(4):266, 1972.

Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.

David M Coppola, Harriett R Purves, Allison N McCoy, and Dale Purves. The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences*, 95(7): 4002–4006, 1998.

Sylvain Fischer, Rafael Redondo, Laurent Perrinet, Gabriel Cristóbal, Gabriel Cristobal, Gabriel Cristóbal, and Gabriel Cristobal. Sparse Approximation of Images Inspired from the Functional Architecture of the Primary Visual Areas. *EURASIP Journal on Advances in Signal Processing*, 2007(1):1–17, 2007a. ISSN 16876172. doi: 10.1155/2007/90727. URL http://www.hindawi.com/journals/asp/2007/090727.abs.htmlhttp://asp.eurasipjournals.com/content/2007/1/090727.

Sylvain Fischer, Filip Šroubek, Laurent U Perrinet, Rafael Redondo, and Gabriel Cristóbal. Self-invertible 2d log-gabor wavelets. *International Journal of Computer Vision*, 75(2):231–246, 2007b.

Rich Franzen. The kodak color image dataset, available online at http://r0k.us/graphics/kodak/, 2013.

Yann Gousseau and Jean-Michel Morel. Are natural images of bounded variation? *SIAM Journal on Mathematical Analysis*, 33(3):634–648, 2001.

HLF von Helmholtz. *Treatise on physiological optics, 3 vols.* Optical Society of America, 1924.

Olivier J Hénaff, Zoe M Boundy-Singer, Kristof Meding, Corey M Ziemba, and Robbe LT Goris. Representation of visual uncertainty through neural gain variability. *Nature communications*, 11 (1):1–12, 2020.

Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 110(3):457–506, 2021.

Hugo J Ladret, Nelson Cortes, Lamyae Ikan, Frédéric Chavane, Christian Casanova, and Laurent U Perrinet. Dynamical processing of orientation precision in the primary visual cortex. *bioRxiv*, pp. 2021–03, 2022.

Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Ng. Efficient sparse coding algorithms. *Advances in neural information processing systems*, 19, 2006.

Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*, 2017.

Keisuke Nakamura and Kazuhiro Nakadai. Robot audition based acoustic event identification using a bayesian model considering spectral and temporal uncertainties. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4840–4845. IEEE, 2015.

Bruno A Olshausen and David J Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996.

Bruno A Olshausen and David J Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision research*, 37(23):3311–3325, 1997.

Gergő Orbán, Pietro Berkes, József Fiser, and Máté Lengyel. Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron*, 92(2):530–543, 2016.

Laurent U Perrinet. Sparse Models for Computer Vision. In Matthias Keil, Gabriel Cristóbal, and Laurent U Perrinet (eds.), *Biologically Inspired Computer Vision*, pp. 319–346. Wiley-VCH Verlag GmbH & Co. KGaA, 2015. doi: 10.1002/9783527680863.ch14. URL `http://onlinelibrary.wiley.com/doi/10.1002/9783527680863.ch14/summary`.

Charles E Pettypiece, Melvyn A Goodale, and Jody C Culham. Integration of haptic and visual size cues in perception and action revealed through cross-modal conflict. *Experimental brain research*, 201(4):863–873, 2010.

Daniel L Ruderman. The statistics of natural images. *Network: computation in neural systems*, 5 (4):517, 1994.

Thomas Serre, Aude Oliva, and Tomaso Poggio. A feedforward architecture accounts for rapid categorization. *Proceedings of the national academy of sciences*, 104(15):6424–6429, 2007.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Nicholas V Swindale. Orientation tuning curves: empirical description and estimation of parameters. *Biological cybernetics*, 78(1):45–56, 1998.

Yu Wang, Wotao Yin, and Jinshan Zeng. Global convergence of admm in nonconvex nonsmooth optimization. *Journal of Scientific Computing*, 78(1):29–63, 2019.

Brendt Wohlberg. Efficient convolutional sparse coding. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7173–7177. IEEE, 2014.

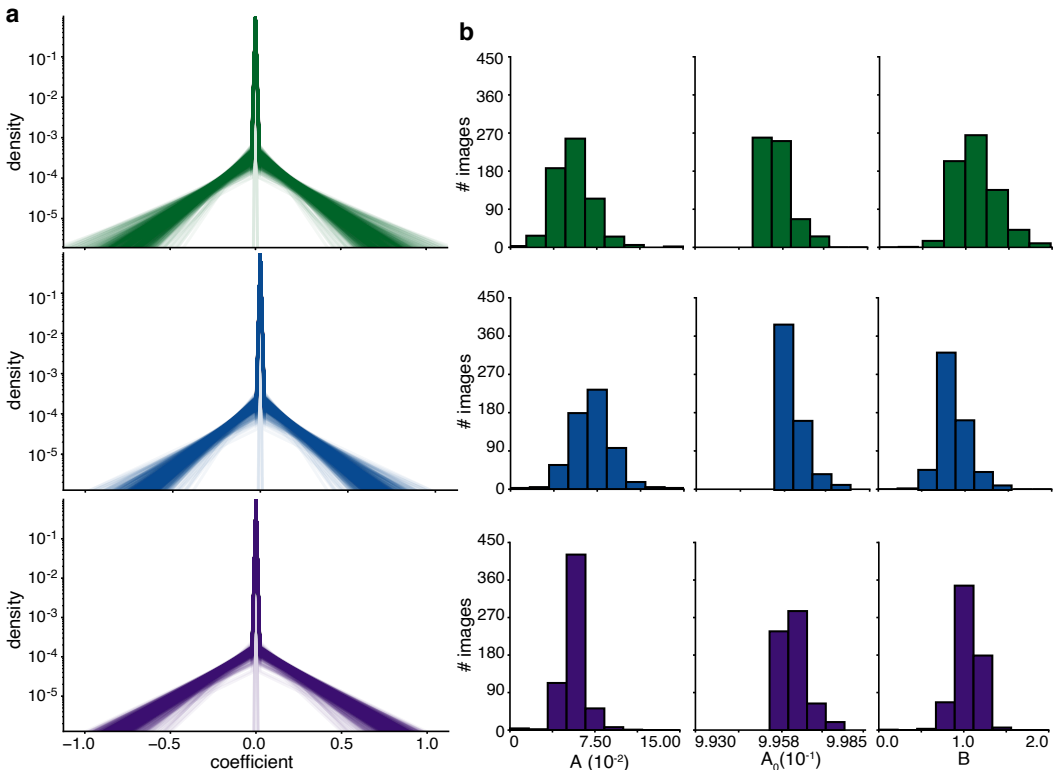
Brendt Wohlberg. Efficient algorithms for convolutional sparse representations. *IEEE Transactions on Image Processing*, 25(1):301–315, 2015.

Brendt Wohlberg. Sporco: A python package for standard and convolutional sparse representations. In *Proceedings of the 15th Python in Science Conference, Austin, TX, USA*, pp. 1–8, 2017.

Jian Zhang and Bernard Ghanem. Ista-net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1828–1837, 2018.

Chen Zhu, Renkun Ni, Zheng Xu, Kezhi Kong, W Ronny Huang, and Tom Goldstein. Gradinit: Learning to initialize neural networks for stable and efficient training. *Advances in Neural Information Processing Systems*, 34:16410–16422, 2021.

A APPENDIX A



Appendix A Figure 1: Sparse coefficients of natural images follow a Dirac-Laplacian distribution. **(a)** Distribution of the coefficients' activation from the fixed uncertainty (top, green); heterogeneous uncertainty pre-learning (middle, blue) and post-learning (bottom, purple) dictionaries. Each solid line represents a single Dirac-Laplacian fit of the activations over a single image. **(b)** Distribution of the parameters of the Dirac-Laplacian fit, showing that the best invariance is obtained with the heterogeneous uncertainty post-learning dictionary.

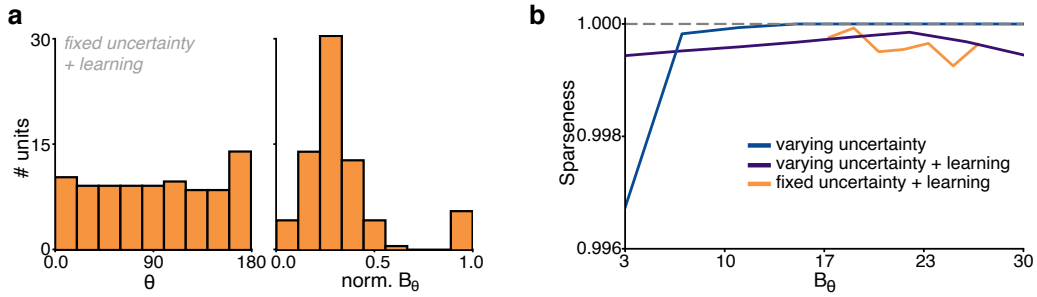
For each of the images in the dataset, the pattern of activations of the dictionary, i.e. the usage of coefficients for the reconstruction of a single an image, were highly stereotypical. The relationship between a coefficient's absolute activation and its density (the inverse of its sparseness) was well described by a mixture of Dirac and Laplacian functions:

$$y(x) = (1 - A_0) \exp\left(\frac{-|x|}{B}\right) A + A_0 \tag{11}$$

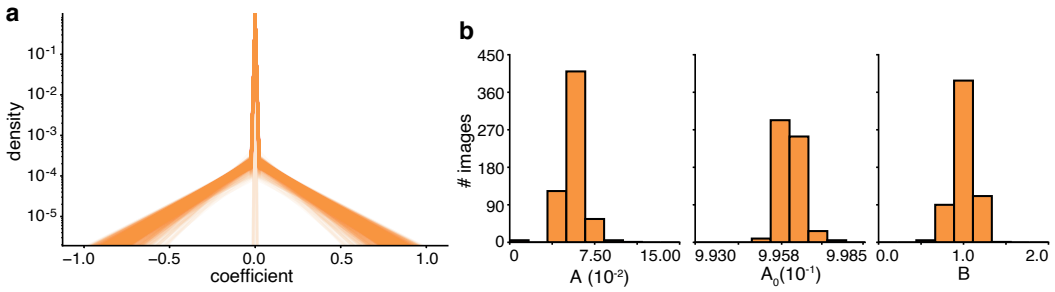
where B is the rate parameter of the exponential function, A_0 is a baseline parameter and A the gain of the function. This function captured the activation of all dictionaries, regardless of their structure (Figure 1a). However, the specific parametrization of the function depended on the dictionary, and revealed functional differences that echoed their structural differences. For instance, the heterogeneous epistemic uncertainty dictionary showed significantly higher exponential rate parameter B ($U = 255133.0$, $p < 0.001$) and baseline gain A_0 ($U = 90642.0$, $p < 0.001$) compared to the fixed epistemic dictionary (Figure 1a, green and blue). Fewer extreme coefficients were thus used (less spread function), but overall more coefficients were used, which resulted in a smaller spread function and a greater use of coefficients, leading to higher global sparseness for the heterogeneous epistemic dictionary (as seen in Figure 2d). This suggests that the encoding process was achieved with increased efficiency. Post-learning, the exponential rate parameter increased significantly ($U = 65922.0$, $p < 0.001$) while the baseline gained decreased ($U = 106921.0$, $p < 0.001$) compared to the pre-learning heterogeneous epistemic uncertainty dictionary. This implies less baseline activation, which can be attributed to better distribution of the coefficients with respect to the dataset (Figure 3b, d).

APPENDIX B

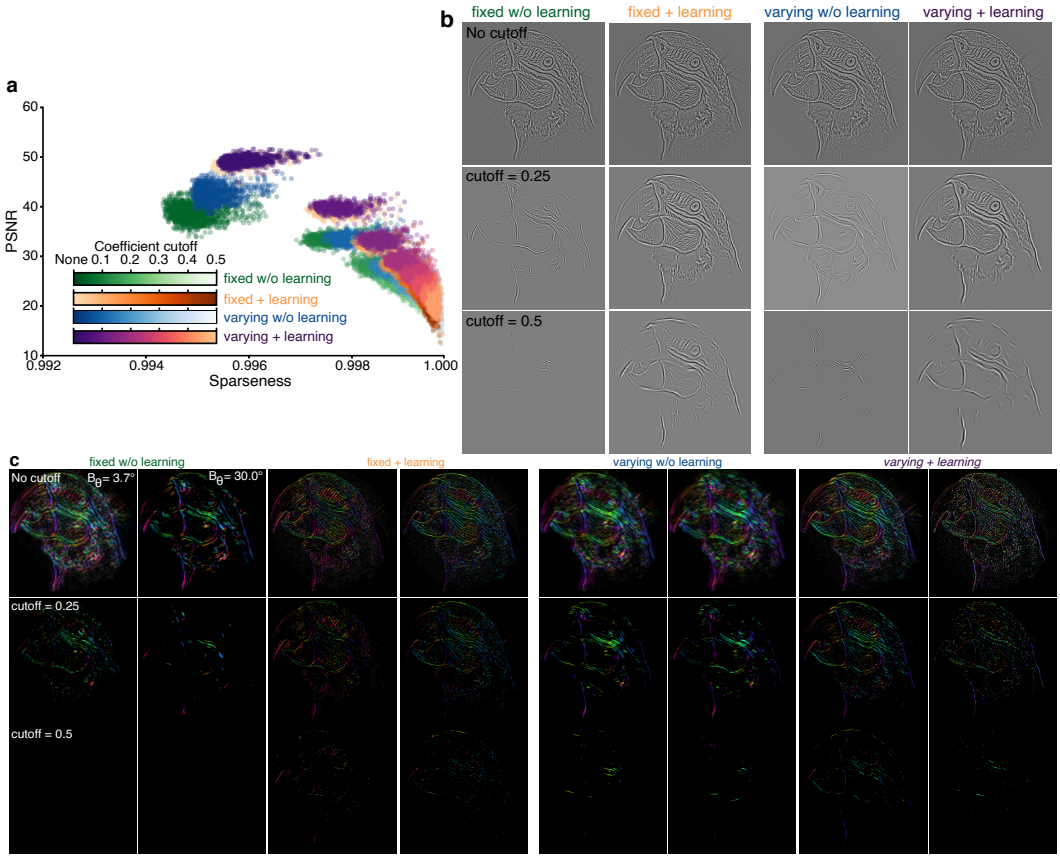
Results from the main text are shown here for the fixed epistemic uncertainty dictionary, post-learning (i.e. Figure 2a, orange).



Appendix B Figure 1: Learning balances coefficient distribution. **(a)** Distribution of coefficients over θ_0 and B_θ after learning. **(b)** Sparseness of coefficients for each B_θ. Sparseness $= 1$ is represented as a gray dashed line.



Appendix B Figure 2: Sparse coefficients of natural images follow a Dirac-Laplacian distribution. **(a)** Distribution of the coefficients' activation from the fixed uncertainty dictionary post-learning. Each solid line represents a single Dirac-Laplacian fit from the activation over a single image. **(b)** Distribution of the parameters of the Dirac-Laplacian fit.



Appendix B Figure 3: Sparse coefficients can be pruned to boost sparsity. **(a)** Pruning of the coefficients based on their values and resulting sparseness/PSNR for both dictionaries. **(b)** Reconstruction of the image shown in Figure 1 with different cutoff levels. **(c)** Coefficients for the same image, for the minimum and maximum epistemic uncertainty and for different cutoff levels.