Gaussian Mixture Models Improve fMRI-based Image Reconstruction

Sanne Schoenmakers[1]*, Marcel van Gerven[1], Tom Heskes[2]

[1] *Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands*

[2] *Radboud University Nijmegen, Institute for Computing and Information Sciences, Nijmegen, The Netherlands*

* *Corresponding author at: Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour,*
Donders Centre for Cognition, P. O. Box 9104, 6500 HE Nijmegen, The Netherlands.
E-mail address: s.schoenmakers@donders.ru.nl. (S. Schoenmakers)

***Abstract*—New computational models have made it possible to reconstruct perceived images from BOLD responses in visual cortex. We expand a linear Gaussian framework for percept decoding with Gaussian mixture models to better represent the prior distribution of images. In our setup, different mixture components correspond to different letter categories. Our framework not only leads to more accurate reconstructions, but also automatically infers semantic categories from low-level visual areas of the human brain.**

I. Introduction

Machine learning techniques have made it possible to accurately decode mental states from neuroimaging data. Especially visual perception is a highly investigated modality since the visual system is relatively well understood and covers a large portion of the brain. Low-level visual features have been shown to allow the reconstruction of perceived stimuli. For instance, center-surround receptive fields have been used to reconstruct natural scenes from cat LGN with invasive recordings [1]. More recently, computational models operating on low-level visual features have been demonstrated to allow reconstruction of perceived images from V1 [2], [3].

A big challenge in fMRI-based image reconstruction is the relatively poor signal-to-noise ratio. In a probabilistic setting, reconstructions can be improved by combining the likelihood function with an image prior using Bayesian inversion. While the likelihood function models fMRI responses to the presented images, the image prior models the statistics of the input data [4].

A feasible image prior is one that encodes the covariance structure between pixels. Unfortunately, such a unimodal prior fails to capture higher-order statistical properties, for example when images belong to different semantic categories. To overcome this problem, we can try using multiple priors, one prior for each category. Previous studies have shown that it is possible to get an accurate read out of the category of a perceived image from fMRI data [5], [6], [7], [8]. In [9] it was shown that the use of semantic information greatly improved image reconstruction.

Here, we present a framework for image reconstruction using Gaussian mixture models in which semantic information can be integrated. That is, the image prior is taken to be multimodal, as captured by a mixture model whose mixture components reflect semantic categories for which the mixture weights are estimated from the fMRI data. We show that this formulation gives an analytical procedure to create image reconstructions from fMRI data that improves on previous work that makes use of a unimodal prior [4]. We evaluate the Gaussian mixture model by applying it to an fMRI dataset of people viewing handwritten characters. Figure 1 outlines the difference between the conventional approach and the mixture model approach.

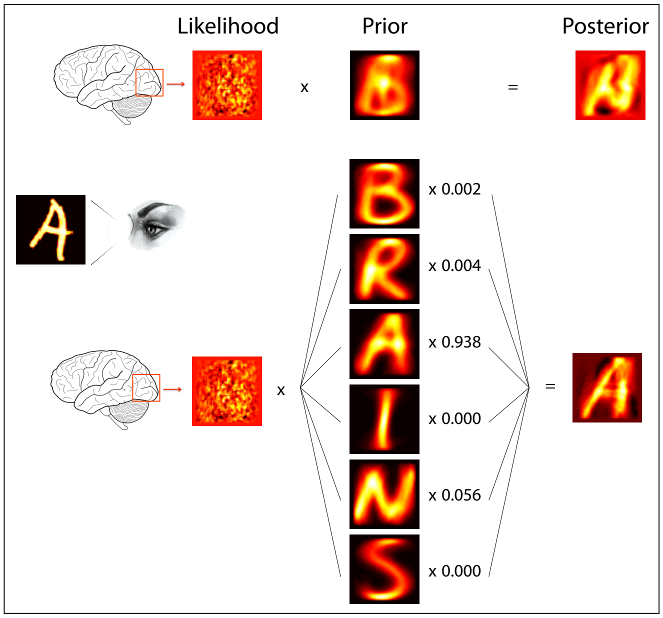

Fig. 1. Schematic diagram of the reconstruction steps. A subject looks at an image. The corresponding brain response is measured with fMRI and would, without any further knowledge, result into a noisy maximum likelihood reconstruction. Combining the likelihood of the brain response with a prior on images leads to a much more accurate reconstruction. A multimodal prior, as modeled by a mixture model, significantly improves over a unimodal prior.

978-1-4799-4149-0/14/$31.00 2014 IEEE

II. METHODS

A. Gaussian mixture models

As in [4], we make use of a linear Gaussian encoding model with image $\mathbf{x} = (x_1, \ldots, x_p)' \in \mathbb{R}^p$ and the associated measured response vector $\mathbf{y} = (y_1, \ldots, y_q)' \in \mathbb{R}^q$:

$$\mathbf{y} = \mathbf{B}'\mathbf{x} + \boldsymbol{\epsilon}$$

with $\boldsymbol{\epsilon}$ zero mean normally distributed noise. Regression coefficients $\mathbf{B}$ are estimated using regularized linear regression. The likelihood function is then given by

$$P(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{y}; \mathbf{B}'\mathbf{x}, \boldsymbol{\Sigma}) \,.$$

with diagonal covariance matrix $\boldsymbol{\Sigma}$. We assume that this mapping is independent of the context, e.g. the category i, in what follows.

For the prior distribution over images $\mathbf{x}$, we consider a Gaussian mixture model, where each mixture component corresponds to a different (letter) category:

$$P(\mathbf{x}) = \sum_i \pi_i \mathcal{N}(\mathbf{x}; \mathbf{m}_i, \mathbf{R}_i) \,,$$

with π_i the prior probability of category i, and $\mathbf{m}_i$ and $\mathbf{R}_i$ the mean and covariance matrix, respectively, of the corresponding Gaussian. The means and covariances are estimated from a separate image data set.

In this probabilistic framework, decoding boils down to computing the probability of a reconstruction $\mathbf{x}$ given an fMRI image $\mathbf{y}$. Following standard probabilistic inference, see e.g., [10], we obtain

$$P(\mathbf{x}|\mathbf{y}) = \sum_i P(i|\mathbf{y}) P(\mathbf{x}|\mathbf{y}, i) \,,$$

where both $P(\mathbf{x}|\mathbf{y}, i)$ and $P(i|\mathbf{y})$ follow from the application of Bayes' rule. That is,

$$P(\mathbf{x}|\mathbf{y}, i) = \frac{P(\mathbf{y}|\mathbf{x}) P(\mathbf{x}|i)}{P(\mathbf{y}|i)} \qquad (1)$$

with

$$P(\mathbf{y}|i) = \int d\mathbf{x} \, P(\mathbf{y}|\mathbf{x}) P(\mathbf{x}|i) \,. \qquad (2)$$

and

$$P(i|\mathbf{y}) = \frac{\pi_i P(\mathbf{y}|i)}{\sum_j \pi_j P(\mathbf{y}|j)} \,, \qquad (3)$$

Since both the likelihood $P(\mathbf{y}|\mathbf{x})$ and the prior $P(\mathbf{x}|i)$ have the form of a Gaussian in $\mathbf{x}$, so does their product. Therefore, deriving Equations (1), (2) and (3) is straightforward and we merely state the result here. The posterior $P(\mathbf{x}|\mathbf{y}, i)$ of a reconstruction $\mathbf{x}$ given an fMRI image $\mathbf{y}$ under the assumption that the corresponding category equals i is a Gaussian distribution with mean $\mathbf{n}_i(\mathbf{y})$ and variance $\mathbf{Q}_i$, which can be computed through

$$\mathbf{n}_i(\mathbf{y}) = \mathbf{Q}_i \bar{\mathbf{z}}(\mathbf{y}) + \mathbf{U}_i \mathbf{m}_i$$

where

$$\begin{aligned} \mathbf{U}_i &\equiv (\mathbf{I} + \mathbf{R}_i \mathbf{D})^{-1} \\ \mathbf{D} &\equiv \mathbf{B}\boldsymbol{\Sigma}^{-1}\mathbf{B}' \\ \mathbf{Q}_i &\equiv \mathbf{U}_i \mathbf{R}_i \\ \bar{\mathbf{z}}(\mathbf{y}) &\equiv \mathbf{B}\boldsymbol{\Sigma}^{-1}\mathbf{y} \,, \end{aligned}$$

and with $\mathbf{I}$ the identity matrix. The posterior probability $P(i|\mathbf{y})$ gives the probability that the category is indeed i given the fMRI image $\mathbf{y}$. It can be shown to obey

$$\begin{aligned} \log P(i|\mathbf{y}) = \log \pi_i + \frac{1}{2} \log \det \mathbf{U}_i + \frac{1}{2} \bar{\mathbf{z}}(\mathbf{y})' \mathbf{Q}_i \bar{\mathbf{z}}(\mathbf{y}) \\ - \frac{1}{2} \mathbf{m}_i' \mathbf{D} \mathbf{U}_i \mathbf{m}_i + \bar{\mathbf{z}}(\mathbf{y})' \mathbf{U}_i \mathbf{m}_i + C \,, \end{aligned}$$

where constants C can be ignored when normalizing $P(i|\mathbf{y})$ to sum to one since they are independent of i.

For the final reconstruction we then obtain

$$\mathbf{x}^*(\mathbf{y}) = \sum_i w_i(\mathbf{y}) \mathbf{n}_i(\mathbf{y}) \qquad (4)$$

with weights $w_i(\mathbf{y}) \propto P(i|\mathbf{y})^{1/T}$. Varying the temperature T introduces a natural way of interpolating between the most probable category and equal mixing of categories.

For temperature $T = 1$, we have $w_i(\mathbf{y}) = P(i|\mathbf{y})$ and the reconstruction is a standard weighted average of the reconstructions for each of the categories. In the limit $T \downarrow 0$, we zoom in on the reconstruction $\mathbf{n}_{i^*}(\mathbf{y})$ corresponding to the most probable category $i^* = \text{argmax}_i \, P(i|\mathbf{y})$. When no temperature is specified the model with $T = 1$ is implied.

B. Data Acquisition

To investigate the performance of the Gaussian mixture model we tested it on an fMRI dataset. The participant perceived instances of handwritten characters which were evenly distributed over six letter categories (B, R, A, I, N, S). The total set of images viewed by the participant contained 360 instances. BOLD estimates were acquired from the primary visual area V1 as in [4]. A regularized linear regression model was estimated to form a pixel-to-BOLD mapping. Graphnet was used for the regularisation of the linear model [11], which introduces sparseness and smoothing to the model. Image reconstructions were obtained via application of Eq. (4) for different settings of the temperature parameter. Class-specific means and covariances were estimated from a separate dataset containing 700 handwritten instances per letter category.

In order to quantify how much the reconstructions were alike to the originals the structural similarity metric (SSIM) was used. SSIM was specifically developed to match the properties of the human visual system when determining to what extent two images are alike. The measure is similar to taking the correlation between two images except that it takes into account noise and distortion of images and indexes images based on their structural similarity [12].

III. Results

In order to assess the performance of the Gaussian mixture model we compare the reconstructions that we obtain by using the multimodal prior for separate character classes with the reconstructions that are obtained when using a unimodal prior that contains all classes in one unimodal prior.

Figure 2 depicts the mixture weights w_i at $T = 1$ for the 72 instances of the test set. On the diagonal, blocks with high values are visible, demonstrating that many of the instances are correctly identified with the highest probability. In 63% of the instances the maximum of w_i provides the correct class. At chance level we would expect to get 17% of the instances correct. Furthermore, the figure reveals that often one or just a few categories actually contribute to the mixture. This ensures that some of the categories that are deemed very unlikely will not contribute to the reconstruction, which reduces the chance of a distorted result.

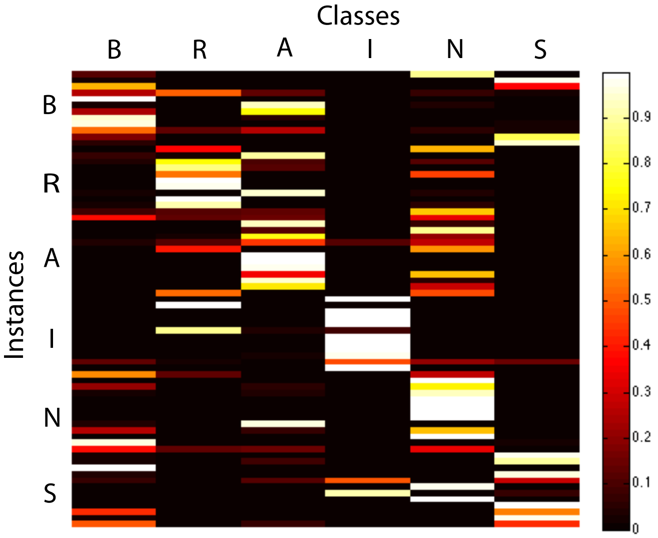

Fig. 2. Mixture weights per class for 72 test characters (twelve exemplars for all six categories).

Figure 3 displays the structural similarity metric per reconstruction for the multimodal prior relative to the unimodal prior in separate plots for each category. The plus sign represents correctly classified reconstructions and the circles show wrongly classified reconstructions. It can be seen that the multimodal prior improves upon the unimodal prior in all correctly classified cases and even in some of the wrongly classified cases. A paired samples t-test shows that the multimodal prior gives rise to a significant improvement (p $< 10^{-11}$) over the reconstructions relative to the reconstructions for the unimodal prior.

Figure 4 depicts in blue the average structural similarity score for different values of the temperature T. In red, the mean SSIM for the unimodal prior is shown. The average for the multimodal prior reaches a maximum when the temperature approaches zero. This shows that, according to the similarity metric, the best reconstructions are not formed when the mixture of categories is used, but rather when the category is chosen that is most likely according to the Gaussian mixture model. A paired samples t-test shows the similarity metric to be significantly lower ($p < 0.05$) for $T = 1$ compared with $T = 0$.

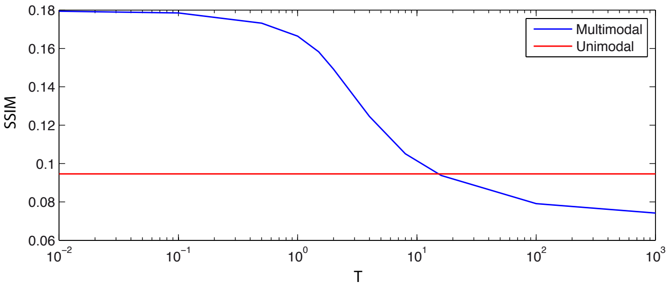

Fig. 4. Mean structural similarity metric for the multimodal prior at different temperatures in blue and in red the mean structural similarity metric for the unimodal prior.

Figure 5 demonstrates three sets of examples of the reconstructions. In the first column the original images are shown. The next column visualizes the reconstructions based on the unimodal prior, followed by the reconstructions with the multimodal prior at $T = 1$ and $T = 0$. Panel A illustrates some examples of reconstructions that lead to similar reconstructions for the different priors. Panel B contains examples that have greatly improved because of the mixture over categories and even more by taking the most likely category as according to the mixture weights. Finally, panel C shows examples of reconstructions that are incorrect under all reconstruction approaches. Notice that panel C presents reconstructions that look correct, but are actually showing reconstructions of the wrong character. This is particularly prominent for the case where $T = 0$.

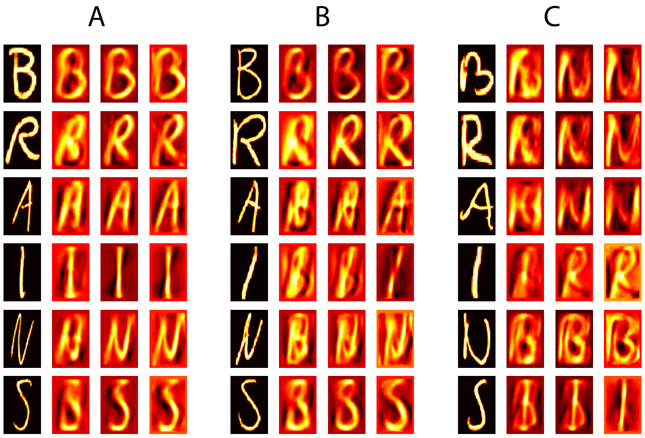

Fig. 5. Examples of reconstructions. The first column in a panel shows the original image, the second column shows the reconstructions based on the unimodal prior, the third column shows the reconstructions that follow from the multimodal prior and in the fourth column the reconstructions are shown for the most likely category. Panel A demonstrates examples of reconstructions that are good for all types of prior, Panel B shows examples that improve under the different priors and Panel C represents reconstructions that fail.

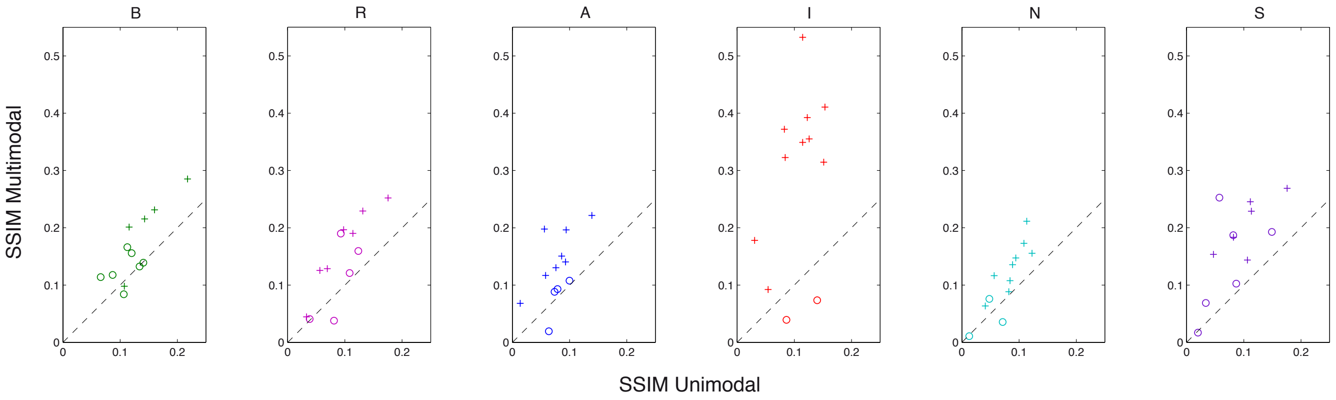

Fig. 3. Structural similarity metric per reconstruction for the multimodal prior relative to the unimodal prior in separate plots for each category. The plus signs correspond to image reconstructions for which the most likely category does happen to coincide with the actual category, the circles correspond to incorrect classifications.

IV. Conclusion

We not only showed that Gaussian mixture models lead to more accurate reconstructions, but also that by using such models one can automatically infer higher-order semantic categories from a low-level visual area in the brain. Furthermore, it appears that zooming in on the most probable category leads to better reconstructions (in terms of SSIM) than taking a standard weighted average over all categories. The drawback of choosing the most probable category is that reconstructions may converge towards the incorrect stimulus category.

The performance of correctly classifying 63% over a chance level of 17% is impressive for a multi-class classifier, but from a decoding perspective the model requires further improvement. Based on our experiences we suggest several potentially beneficial approaches to boost classification rate.

For one, it will be worthwhile to improve neuroimaging data quality using better acquisition protocols, more sophisticated analysis methods and longer recordings.

A second point of interest is the prior data. It might be the case that some of the character features are underrepresented in the data set used to estimate the prior.

Another improvement might be to include more brain data from extrastriate visual areas. The brain's ventral stream has a hierarchical organization leading all the way down to the anterior temporal lobe. These higher level brain regions provide a more explicit representation of semantic information which is expected to improve classification performance like in [9]. In contrast, we now use voxels in visual area V1.

Finally, the framework may be enhanced by learning categories directly from the data using expectation maximization. The inferred category structure is then driven by both image data and by the neural recording and this may improve estimation of mixture weights.

Summarizing, our results show that an analytical framework with a mixture model for the prior is effective in reconstructing images with an underlying class structure.

References

[1] G. B. Stanley, F. F. Li, and Y. Dan, "Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus," *The Journal of Neuroscience*, vol. 19, no. 18, pp. 8036–8042, 1999.

[2] B. Thirion, E. Duchesnay, E. Hubbard, J. Dubois, J.-B. Poline, D. Lebihan, and S. Dehaene, "Inverse retinotopy: inferring the visual content of images from brain activation patterns," *Neuroimage*, vol. 33, no. 4, pp. 1104–1116, 2006.

[3] Y. Miyawaki, H. Uchida, O. Yamashita, M. Sato, Y. Morito, H. C. Tanabe, N. Sadato, and Y. Kamitani, "Visual image reconstruction from human brain activity using a combination of multiscale local image decoders," *Neuron*, vol. 60, no. 5, pp. 915–929, 2008.

[4] S. Schoenmakers, M. Barth, T. Heskes, and M. van Gerven, "Linear reconstruction of perceived images from human brain activity," *NeuroImage*, vol. 83, p. 951, 2013.

[5] J. Haxby, M. Gobbini, M. Furey, A. Ishai, J. Schouten, and P. Pietrini, "Distributed and overlapping representations of faces and objects in ventral temporal cortex," *Science*, vol. 293, pp. 2425–2430, 2001.

[6] D. D. Cox and R. L. Savoy, "Functional magnetic resonance imaging (fMRI) brain reading: detecting and classifying distributed patterns of fMRI activity in human visual cortex," *Neuroimage*, vol. 19, no. 2, pp. 261–270, 2003.

[7] N. Kriegeskorte, M. Mur, D. A. Ruff, R. Kiani, J. Bodurka, H. Esteky, K. Tanaka, and P. A. Bandettini, "Matching categorical object representations in inferior temporal cortex of man and monkey," *Neuron*, vol. 60, no. 6, pp. 1126–1141, 2008.

[8] I. Simanova, P. Hagoort, R. Oostenveld, and M. van Gerven, "Modality-independent decoding of semantic information from the human brain," *Cerebral Cortex*, vol. 24, pp. 426–434, 2014.

[9] T. Naselaris, R. J. Prenger, K. N. Kay, M. Oliver, and J. L. Gallant, "Bayesian reconstruction of natural images from human brain activity," *Neuron*, vol. 63, no. 6, pp. 902–915, 2009.

[10] C. Bishop, *Pattern Recognition and Machine Learning*. Springer Verlag, 2006.

[11] L. Grosenick, B. Klingenberg, K. Katovich, B. Knutson, and J. E. Taylor, "Interpretable whole-brain prediction analysis with GraphNet.," *NeuroImage*, vol. 72, no. 2, pp. 304–321, 2013.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.