

# A Probabilistic Generative Approach to Invariant Visual Inference and Learning

Zhenwen Dai and Jörg Lücke

*FIAS, Goethe-Universität Frankfurt am Main, Germany*

Inference and learning from visual data is a challenging task because of noise and the data's ambiguity. The most advanced vision systems to date are the sensory visual circuitries of higher vertebrates. Although artificial approaches make continuous progress, they are for the majority of applications an unequal match to such biological systems so far. To understand and to rebuild biological systems, they have been modeled using approaches from artificial intelligence, artificial neural networks, and probabilistic models. In terms of how the problem of invariant recognition is approached, these models can coarsely be grouped into two classes: models that passively treat invariances (e.g., [1,2]) and models that actively address the typical transformation invariances of object identities (e.g., [3-6]). The former approaches are often feed-forward while the latter approaches are usually recurrent.

In this work we study a probabilistic generative approach that explicitly addresses the translation invariance of objects in visual data. Object location is modeled using an explicit hidden variable while the object itself is encoded by a specific spatial combination of features. The hidden variable for object position corresponds in this respect to neural control units that were suggested as crucial computational elements in biological networks (e.g., [3,5,6]). The investigated generative model autonomously learns from unlabeled data with object identity and position. The data always contain the same object type at a random position and the back-ground of each data point is different. Note that such learning is a challenging task because it is initially difficult to infer the object position without knowledge about the object features, and without known object positions these features are difficult to infer. By using a probabilistic generative approach, we can nevertheless show that an object's feature combination can reliably be learned based on a maximum likelihood approach. We use expectation maximization (EM) to increase the data likelihood, which is similar, e.g., to approaches by Olshausen et. al. [3] or Williams and Titsias [7]. However, in contrast to these earlier studies we use vectorial features for image encoding and can robustly handle changing back-grounds. We demonstrate the algorithm using artificial and more realistic visual data.

- [1] Y. LeCun, F.J. Huang, and L. Bottou, "Learning methods for generic object recognition with invariance to pose and lighting," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 97–104, 2004.
- [2] M. Riesenhuber, and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature Neuroscience* 211: 1019–1025, 1999.
- [3] B.A. Olshausen, C.H. Anderson, and D.C. Van Essen, "A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information," *J Neurosci* 13: 4700–4719, 1993.
- [4] Lades, M., Vorbrüggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Würtz, R. P., and Konen, W, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans Comp* 42:300-311, 1993.
- [5] J. Lücke, C. Keck, and C. von der Malsburg, "Rapid convergence to feature layer correspondences," *Neural Computation* 20: 2441–2463, 2008.
- [6] G. Hinton, "A parallel computation that assigns canonical object-based frames of reference," *Proc. of the Seventh International Joint Conference on Artificial Intelligence*, 2, 683–685, 1981.
- [7] C.K. Williams and M.K. Titsias, "Greedy learning of multiple objects in images using robust statistics and factorial learning," *Neural Computation*, 16: 1039-1062, 2004.