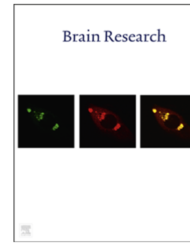
Available online at www.sciencedirect.com
ScienceDirect
www.elsevier.com/locate/brainres



Research Report

Natural image sequences constrain dynamic receptive fields and imply a sparse code

Chris Häusler[a,c,*,1], *Alex Susemihl*[b,c,1], *Martin P. Nawrot*[a,c]

[a]*Neuroinformatics and Theoretical Neuroscience Group, Freie Universität Berlin, Germany*
[b]*Methods of Artificial Intelligence Group, Berlin Institute of Technology, Germany*
[c]*Bernstein Center for Computational Neuroscience Berlin, Germany*

ARTICLE INFO

Article history:
Accepted 31 July 2013
Available online 8 August 2013

Keywords:
Autoencoding
Lifetime sparseness
Machine learning
Population sparseness
Restricted Boltzmann Machine
Visual cortex

ABSTRACT

In their natural environment, animals experience a complex and dynamic visual scenery. Under such natural stimulus conditions, neurons in the visual cortex employ a spatially and temporally sparse code. For the input scenario of natural still images, previous work demonstrated that unsupervised feature learning combined with the constraint of sparse coding can predict physiologically measured receptive fields of simple cells in the primary visual cortex. This convincingly indicated that the mammalian visual system is adapted to the natural spatial input statistics. Here, we extend this approach to the time domain in order to predict dynamic receptive fields that can account for both spatial and temporal sparse activation in biological neurons. We rely on temporal restricted Boltzmann machines and suggest a novel temporal autoencoding training procedure. When tested on a dynamic multi-variate benchmark dataset this method outperformed existing models of this class. Learning features on a large dataset of natural movies allowed us to model spatio-temporal receptive fields for single neurons. They resemble temporally smooth transformations of previously obtained static receptive fields and are thus consistent with existing theories. A neuronal spike response model demonstrates how the dynamic receptive field facilitates temporal and population sparseness. We discuss the potential mechanisms and benefits of a spatially and temporally sparse representation of natural visual input.

1. Introduction

Physiological and theoretical studies have argued that the sensory nervous systems of animals are evolutionarily adapted to their natural stimulus environment (for review see Reinagel, 2001). The question of how rich and dynamic natural stimulus conditions determine single neuron response properties and the functional network connectivity in

*Corresponding author at: Neuroinformatics and Theoretical Neuroscience Group, Freie Universität Berlin, Germany.
Fax: +49 3083857298.
E-mail addresses: chris.hausler@bccn-berlin.de (C. Häusler), alex.susemihl@bccn-berlin.de (A. Susemihl), martin.nawrot@fu-berlin.de (M.P. Nawrot).

[1]These authors contributed equally to the published work.

0006-8993
http://dx.doi.org/10.1016/j.brainres.2013.07.056

mammalian sensory pathways has thus become an important focus of interest for theories of sensory coding (for review see Simoncelli and Olshausen, 2001; Olshausen et al., 2004).

For a variety of animal species and for different modalities it has been demonstrated that single neurons respond in a temporally sparse manner (Reinagel, 2001; Jadhav et al., 2009; Olshausen et al., 2004; Hromádka et al., 2008) when stimulated with natural time-varying input. In the mammal this is intensely studied in the visual (Dan et al., 1996; Vinje and Gallant, 2000; Reinagel and Reid, 2002; Yen et al., 2007; Maldonado et al., 2008; Haider et al., 2010; Martin and Schröder, 2013) and the auditory (Hromádka et al., 2008; Chen et al., 2012; Carlson et al., 2012) pathway as well as in the rodent whisker system (Jadhav et al., 2009; Wolfe et al., 2010). Sparseness increases across sensory processing levels and is particularly high in the neocortex. Individual neurons emit only a few spikes positioned at specific instances during the presentation of a time-varying input. Repeated identical stimulations yield a high reliability and temporal precision of responses (Herikstad et al., 2011; Haider et al., 2010). Thus, single neurons focus only on a highly specific spatio-temporal feature from a complex input scenario.

Theoretical studies addressing the efficient coding of natural images in the mammalian visual system have been very successful. In a ground breaking study, Olshausen et al. (1996) learned a dictionary of features for reconstructing a large set of natural still images under the constraint of a sparse code to obtain receptive fields (RFs), which closely resembled the physiologically measured RFs of simple cells in the mammalian visual cortex. This approach was later extended to the temporal domain by van Hateren and Ruderman (1998), learning rich spatio-temporal receptive fields directly from movie patches. In recent years, it has been shown that a number of unsupervised learning algorithms, including the denoising Autoencoder (dAE) (Vincent et al., 2010) and the Restricted Boltzmann Machine (RBM) (Hinton and Salakhutdinov, 2006; Hinton et al., 2012; Mohamed et al., 2011), are able to learn structure from natural stimuli and that the types of structure learnt can again be related to cortical RFs as measured in the mammalian brain (Saxe et al., 2011; Lee et al., 2008, 2009).

Considering that sensory experience is per se dynamic and under the constraint of a temporally sparse stimulus representation at the level of single neurons, how could the static RF model, i.e. the learned spatial feature, extend into the time domain? Here we address this question with an unsupervised learning approach using RBMs as a model class. Building on an existing model, the Temporal Restricted Boltzmann Machine (TRBM) introduced by Sutskever and Hinton (2007), we introduce a novel learning algorithm with a temporal autoencoding approach to train RBMs on natural multi-dimensional input sequences. For validation of the method, we test the performance of our training approach on a reference dataset of kinematic variables of human walking motion and compare it against the existing TRBM model and the Conditional RBM (CRBM) as a benchmark (Taylor et al., 2007). As an application of our model, we train the TRBM using temporal autoencoding on natural movie sequences and find that the neural elements develop dynamic RFs that express smooth transitions, i.e. translations and rotations, of the static receptive field model. Our model neurons account for spatially and temporally sparse activities during stimulation with natural image sequences and we demonstrate this by simulation of neuronal spike train responses driven by the dynamic model responses. Our results propose how neural dynamic RFs may emerge naturally from smooth image sequences.

2. Results

We outline a novel method to learn temporal and spatial structure from dynamic stimuli – in our case smooth image sequences – with artificial neural networks. The hidden units (neurons) of these generative models develop dynamic RFs that represent smooth temporal evolutions of static RF models that have been described previously for natural still images. When stimulated with natural movie sequences the model units are activated sparsely, both in space and time. A point process model translates the model's unit activation into sparse neuronal spiking activity with few neurons being active at any given point in time and sparse single neuron firing patterns.

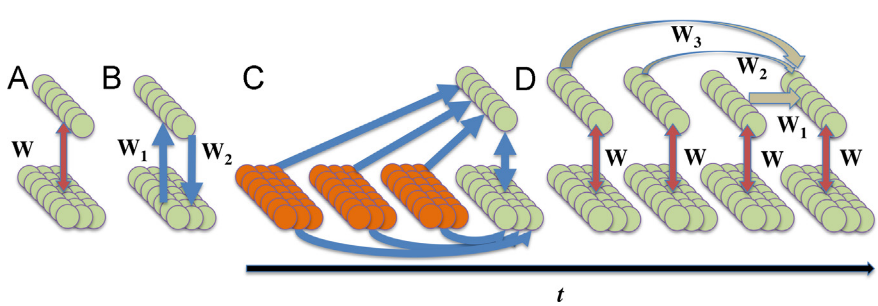


Fig. 1 – Described model architectures: (A) Autoencoder; (B) RBM; (C) Conditional RBM and (D) Temporal RBM. In the CRBM (subfigure C; see also Section 4), there is a *hidden* layer only at the current sample time whose activation is defined by weights connecting the current as well as previous activations of the *visible* layer. The TRBM (subfigure D) has a *hidden* layer instantiation for each sample time within the models delay dependency and the temporal evolution of the model is defined by lateral connections between the *hidden* units of consecutive time steps.