Deep Multimodal Autoencoders for Identifying Latent Representations of Spike Counts and Local Field Potentials

Arno Onken[1], Josue G. Yague[2], and Shuzo Sakata[3]

[1]School of Informatics, University of Edinburgh; [2]School of Medicine, Universidad San Pablo CEU

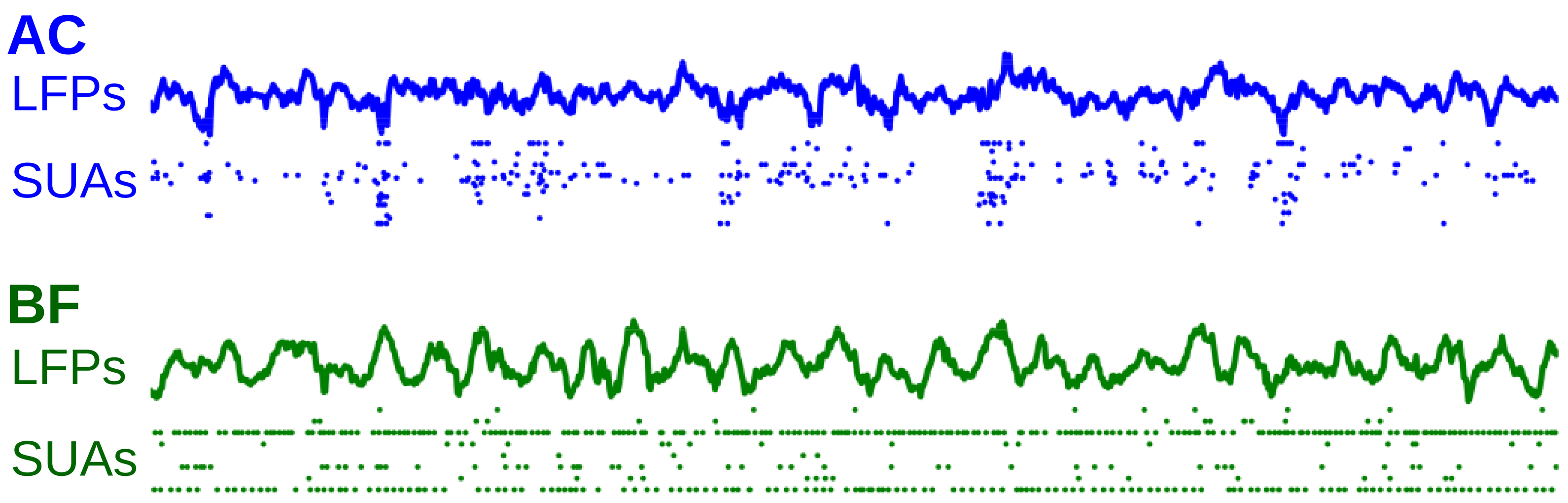
[3]Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde

Introduction

Advances in recording techniques lead to datasets of neural activity with ever increasing dimensionality and complexity. This trend calls for better analysis techniques to identify low-dimensional latent structure that can provide better insights into neural processing. Recently, various forms of deep autoencoders achieved remarkable successes in the identification of latent representations from complex high-dimensional data [e.g. 1]. Here, we apply deep autoencoders to model simultaneously recorded local field potentials and spike counts, exploring the benefits of multimodality for reconstruction.

Local Field Potentials and Spike Counts

- Electrophysiological recordings of spontaneous activity from auditory cortex (AC) and basal forebrain (BF) of urethane anesthetized mice [2]
- NeuroNexus Poly2 32 channels probe
- Local field potentials (LFPs): Lowpass filtered (100 Hz), downsampled to 1 kHz
- Single unit activities (SUAs): Spike sorted using Klusta package
- Between 8 and 31 well-separated single units



- Integration of activity in non-overlapping consecutive 100 ms bins
- Alternatively, LFP power in e.g. delta, theta and gamma bands
- Bins treated as independent samples
- 36000 samples split into training (60%), validation (20%) and test (20%) sets

Modality Noise

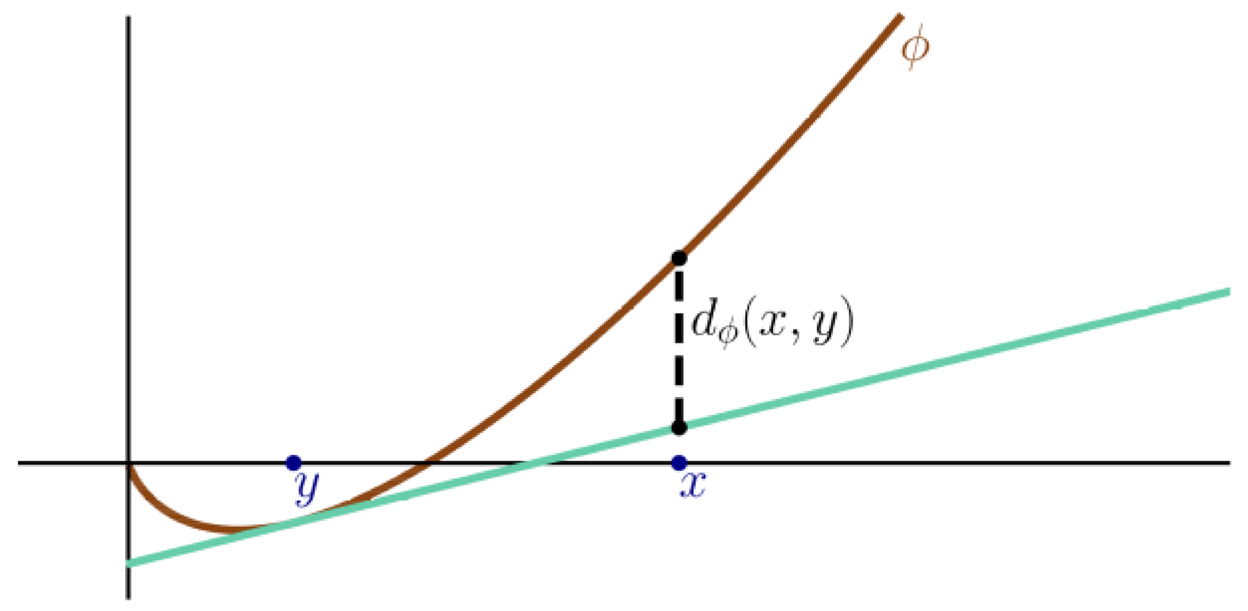
- Noise distributions of modalities can be radically different
- Regular exponential families include wide range of noise distributions (e.g. Gaussian, Poisson, gamma, binomial, negative binomial)
- Every regular exponential family p corresponds to a unique and distinct Bregman divergence d_ϕ [3]:

$$-\log(p(x|\theta)) = -\log(p_0(x)) - \phi(x) + d_\phi(x, g(\theta))$$

where

$$d_\phi(x, y) = \phi(x) - \phi(y) - \langle x - y, \nabla\phi(y)\rangle$$

and ϕ a differentiable strictly convex function



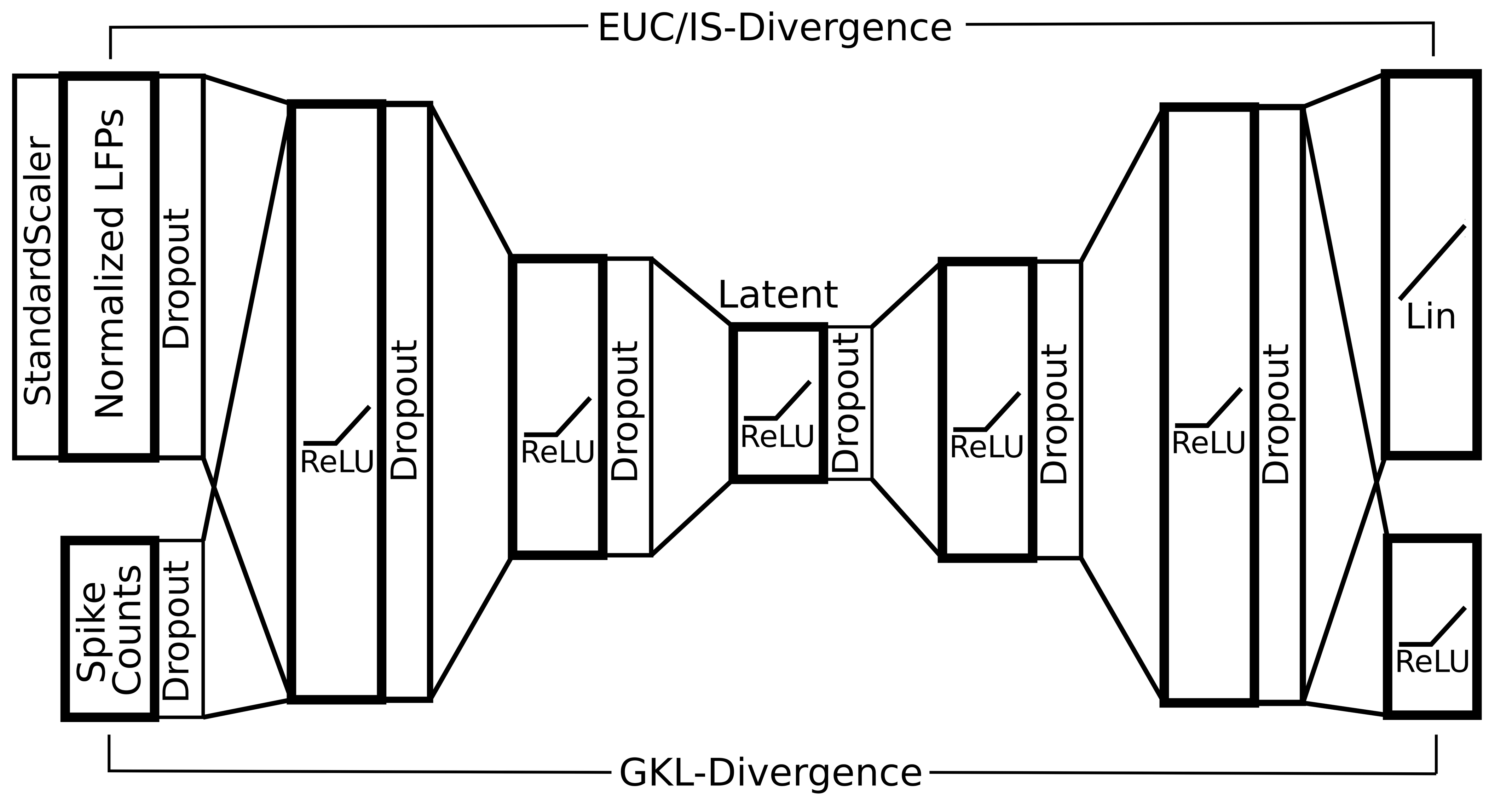
- g denotes link function transforming the natural parameters into variable lying in the domain of the data space
- Terms $-\log(p_0(x))$ and $-\phi(x)$ do not depend on θ
- Maximizing regular exponential family log likelihood is equivalent to minimizing corresponding Bregman divergence

Family	Divergence	$\phi(z)$
Gaussian	Euclidean	$\frac{1}{2}z^2$
Poisson	Generalized Kullback-Leibler	$z\log(z) - z$
Gamma	Itakura-Saito	$-\log(z)$

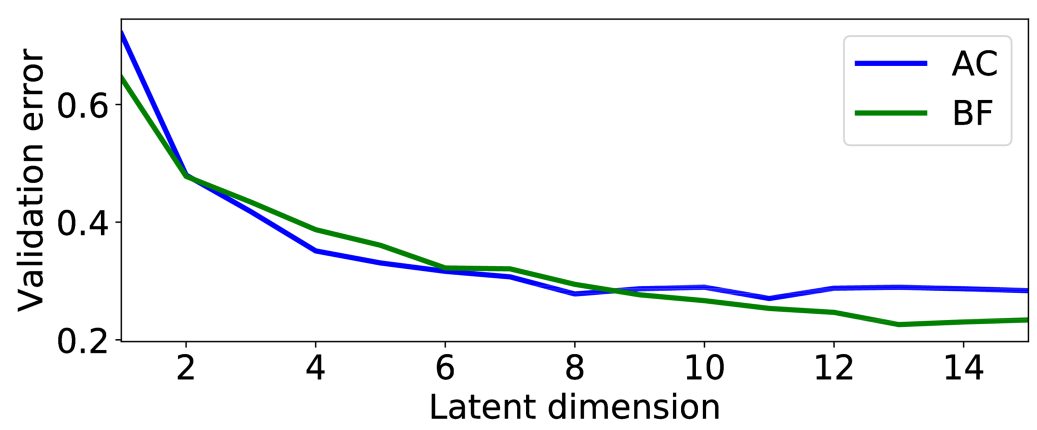
- Useful for LFPs: Gaussian, gamma
- Useful for spike counts: Poisson, binomial, negative binomial

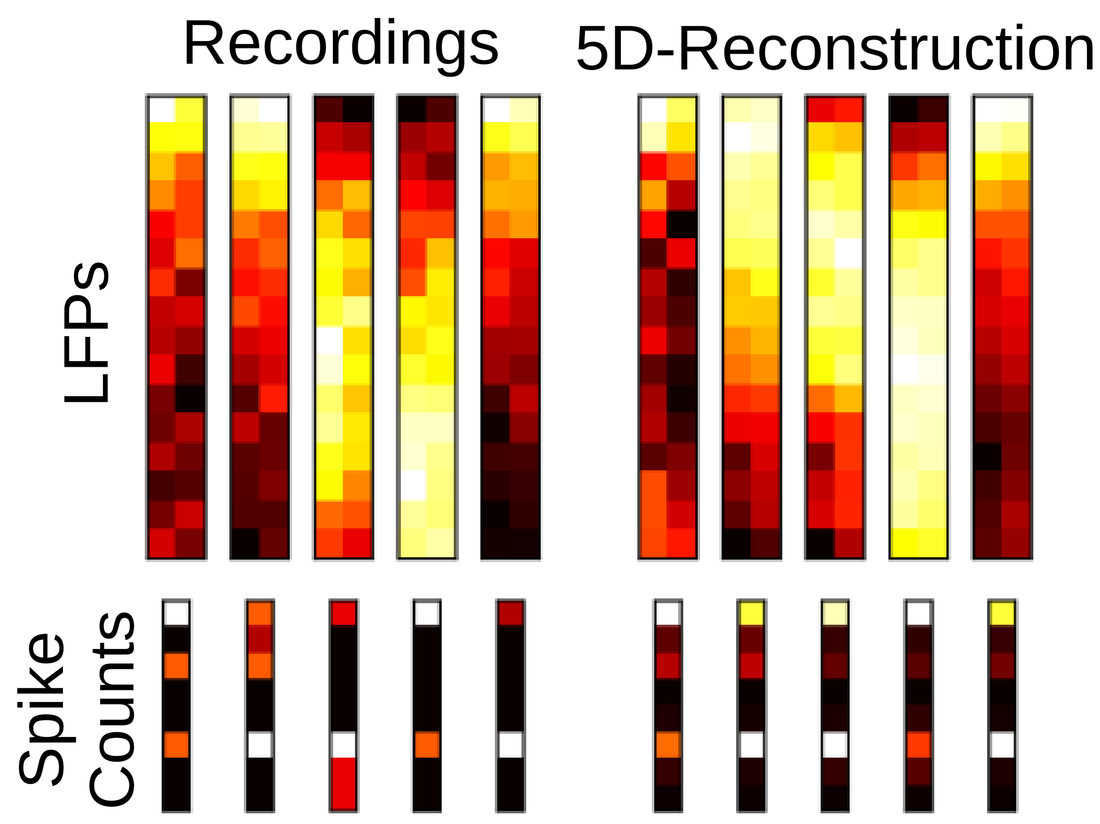
Multimodal Autoencoder

We train deep multimodal autoencoders with early modality fusion to find low dimensional representations of spike counts and local field potentials. For the different modalities, we use distinct loss functions implicitly modeling particular statistical distributions.

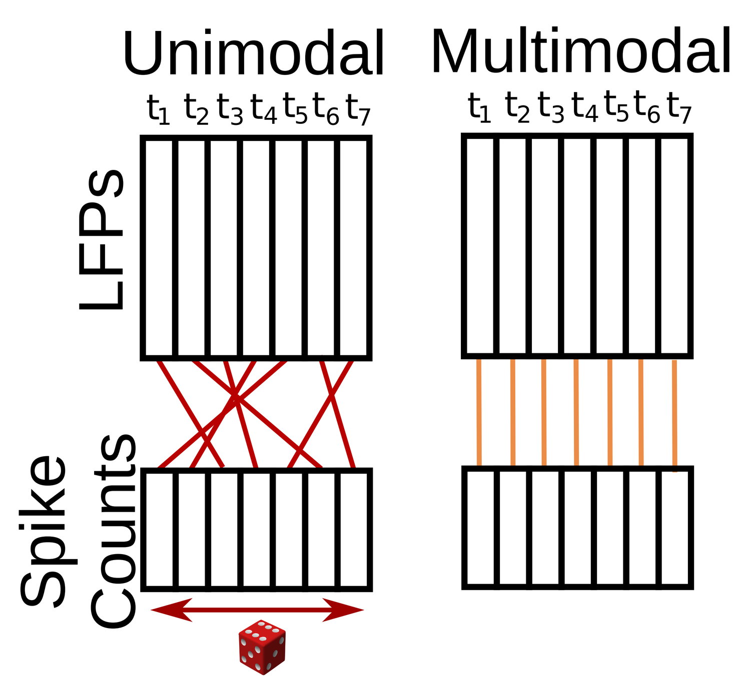


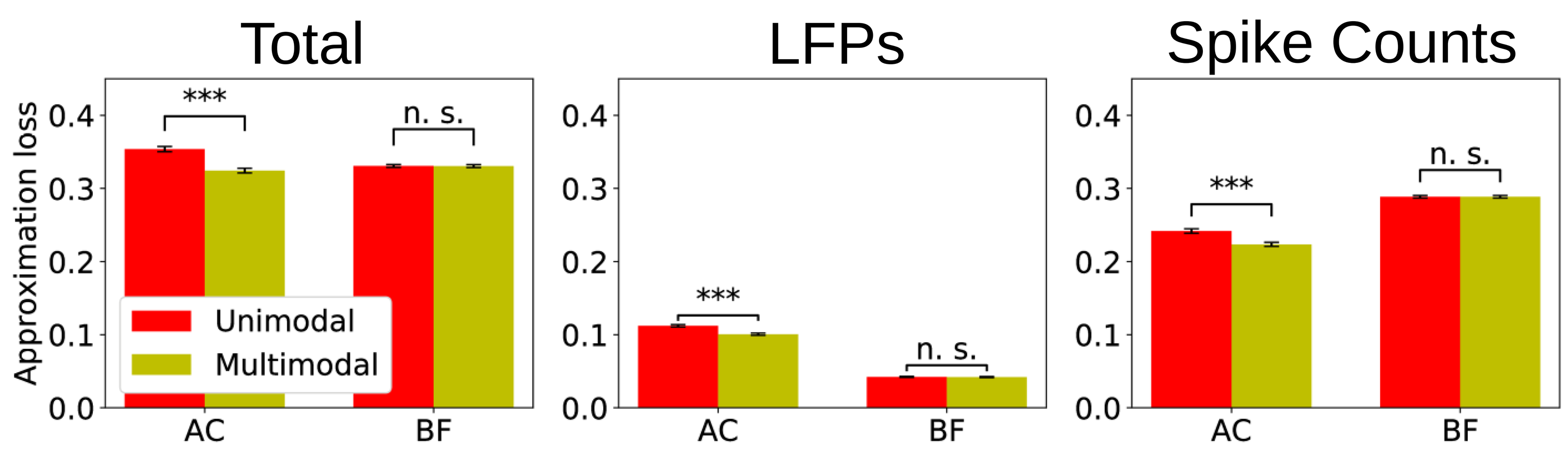
- Optimized number of layers and number of neurons per layer on validation set
- Adam optimizer and dropout with rate 0.2
- Max norm constraint on weights
- Batch size 32
- Keras implementation available [4]





Modality Shuffling





We compare the held-out approximation loss on the original data with the one on data where we shuffled the samples of one modality but not the other, thereby destroying information that one modality carries about the other. We refer to the latter as unimodal approximation loss.

Conclusions

We find that for the AC, multimodal reconstruction performance is significantly greater than the unimodal one whereas for the BF, the reconstruction performance does not benefit from a joint multimodal representation. Our results suggest differences in the AC and BF latent spaces that give rise to the observed spike counts and local field potentials. They further demonstrate that deep autoencoders are useful and versatile tools for identifying low-dimensional multimodal neural representations.

Acknowledgements

This work was supported by the Biotechnology and Biological Sciences Research Council under grant agreement numbers BB/K016830/1 and BB/M00905X/1.

References

1. Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 504-507.
2. Yague, J. G., Tsunematsu, T., & Sakata, S. (2017). Distinct temporal coordination of spontaneous population activity between basal forebrain and auditory cortex. *Frontiers in Neural Circuits*, 11, 64.
3. Banerjee, A., Merugu, S., Dhillon, I. S., & Ghosh, J. (2005). Clustering with Bregman divergences. *Journal of Machine Learning Research*, 6, 1705–1749.
4. Onken, A. (2018). mmae package for Keras. *GitHub Repository*, https://github.com/asnelt/mmae/