Unsupervised Learning of Invariant Visual Representations

Marian Stewart Bartlett[1] and **Terrence J. Sejnowski**[2]
University of California San Diego
[1] Departments of Cognitive Science and Psychology
[2] Department of Biology and Howard Hughes Medical Institute
Computational Neurobiology Lab, The Salk Institute
10010 N. Torrey Pines Road, La Jolla, CA 92037
{marni, terry}@salk.edu

Abstract

The appearance of an object or a face changes continuously as the observer moves through the environment or as a face changes expression or pose. Recognizing an object or a face despite these image changes is a challenging problem for computer vision systems, yet we perform the task quickly and easily. In natural visual experience, different views of an object tend to appear in close temporal proximity. Capturing the temporal relationships among patterns is a way to automatically associate different views of an object without requiring complex geometrical transformations or three dimensional structural descriptions [1].

Temporal association may be an important aspect of invariance learning in the ventral visual stream [2]. There is evidence for temporal association of complex visual patterns by neurons in the inferior temporal lobe [3]. A temporal window for Hebbian learning could be provided by the long open-time of the NMDA channel [4], or by reciprocal connections between cortical regions [5]. Hebbian learning of temporal associations has been explored with idealized input representations [6, 7, 5].

In our first model we used a one layer feedforward network to demonstrate the capability of such mechanisms to acquire invariant representations of complex representations such as gray level images of faces [8; See also 9]. A temporal association (TA) learning rule based on Competitive Learning [10] clusters input patterns by a combination of spatial similarity and temporal proximity. A two layer network presented with sequences of images of faces as subjects gradually change pose learned representations that are invariant to pose (Figure 1)[11].

We next explored the the development of invariances with an attractor state representation. In this framework, representations are sustained pattern of activity across an interconnected assembly of units. Attractor networks with Hebbian learning mechanisms are capable of acquiring temporal associations between randomly generated patterns [12]. We investigate the ability of attractor network dynamics to acquire temporal associations between different views of a face or object. Input to this system consisted of images of faces convolved with Gabor filters of three spatial scales and 4 orientations at 64 locations in the image. Faces are presented to the system in sequence as they rotate from the frontal view. The learning algorithm [12] caused input patterns that are spatially similar and temporally proximal to be drawn to the same basin of attraction. Corrections in attractor states corresponding to the temporal proximity of the associated input patterns during training revealed the degree to which the network acquired invariant representations.

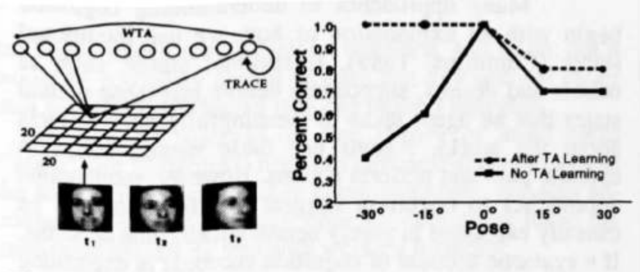

Figure 1: Left: Network architecture. Right: Mean "tuning curve" for pose following competitive learning of frontal views only, and after temporal association (TA) learning. Outputs for non-frontal views were classed as correct if they matched the output for the frontal view of the same face. Data are means for 10 different faces.

References

1. Stryker, M. 1991. Temporal Associations. Nature: 354:108-109.
2. Rolls, E. 1995. Learning Mechanisms in the temporal lobe visual cortex. Behav. Brain Res. 66:177-185.
3. Miyashita, Y. 1988. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. Nature: 335(27):817-820.
4. Rhodes, P. 1992. The long open time of the NMDA channel facilitates the self-organization of invariant object responses in cortex. Soc. Neurosci. Abst. 18:740.
5. O'Reilly, R. & Johnson, M. 1994. Object recognition and sensitive periods: A computational analysis of visual imprinting. Neural Comp. 6:357-389.
6. Foldiak, P. 1991. Learning invariance from transformation sequences. Neural Comp. 3:194-200.
7. Weinshall, D.& Edelman, S. 1991. A self-organizing multiple view representation of 3D objects. Bio. Cyber. 64(3):209-219.
8. Bartlett, M. Stewart, and Sejnowski, T., in press. Unsupervised learning of invariant representations of faces through temporal association. In *The Neurobiology of Computation* J.M. Bower, ed. Kluwer Academic Publishers, Boston.
9. Wallis, G. & Rolls, E. 1996. A model of invariant object recognition in the visual system. Technical Report, Oxford University Department of Experimental Psychology.
10. Rumelhart, D. & Zipser, D. 1985. Feature discovery by competitive learning. Cog. Sci. 9:75-112.
11. We thank David Beymer of the MIT Media Lab for the images.
12. Griniasty, M., Tsodyks, M., & Amit, D. (1993). Conversion of temporal correlations between stimuli to spatial correlations between attractors. Neural Comp. 5:1-17.