Invariants Based Architecture for Combining Small and Large Data Sets in Neural Networks.

Roelant Ossewaarde[1,2], Stefan Leijnen[1], Thijs van den Berg[1]

[1] Artificial Intelligence Research Group, HU University of Applied Science, Utrecht
[2] Neurolinguistics Research Group, Rijksuniversiteit Groningen, Groningen

UNIVERSITY OF APPLIED SCIENCES UTRECHT

Introduction In many practical applications, there is domain specific information available that could beneficially influence the training of deep learned data sets. Predictive models based on small data sets often have the advantage that white box AI techniques (interpretable) perform as well as black box AI techniques (less interpretable), such as Artificial Neural Networks (ANNs).
Our alternative neural network architecture is constructed so that partial representations (invariants) are learned in the intermediate layers, which can then be combined with a priori knowledge or with other predictive analyses of the same data. This leads to smaller training datasets due to more efficient learning. In addition, because this architecture allows inclusion of a priori knowledge and interpretable predictive models, the interpretability of the entire system increases while the data can still be used in a black box neural network.

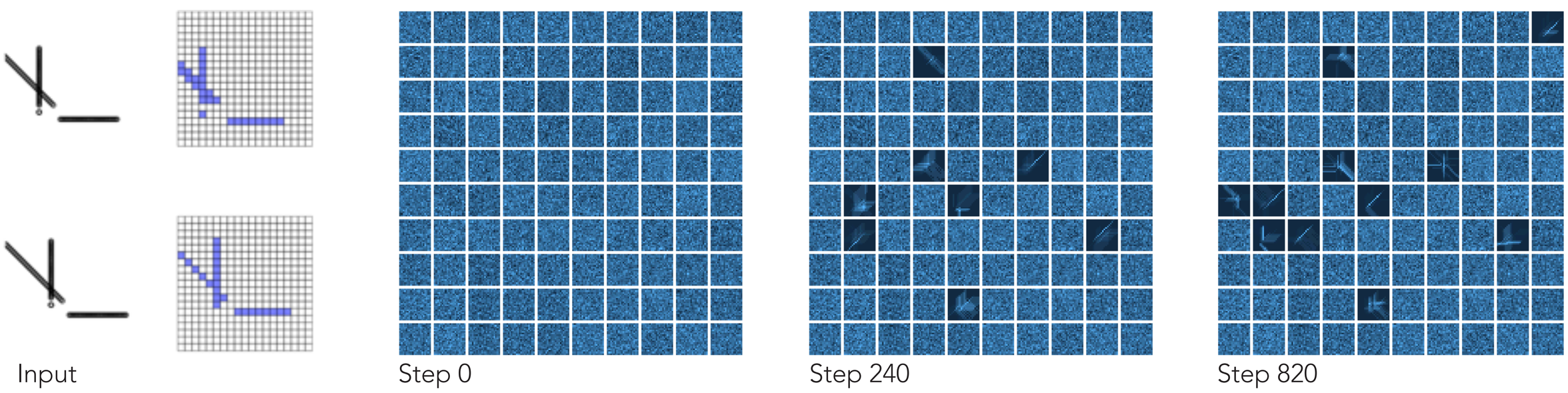
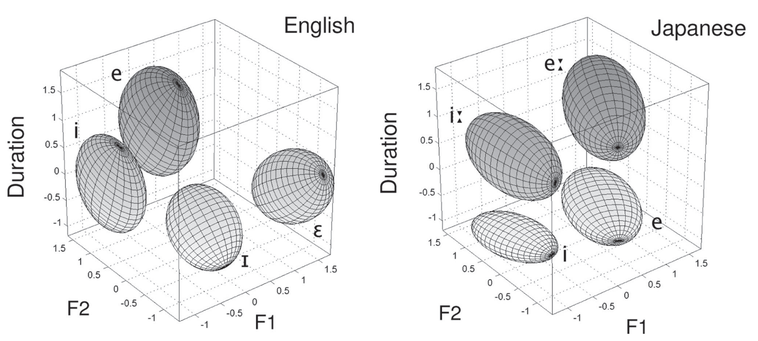


Fig.1 *Complex cells specialize in module hierarchies to recognize elements of the input*

Motivating problem: vowel learning

The properties of Japanese and English vowels form invariants based on their acoustic component waves. Humans quickly specialize to recognize these vowels, mixing episodic memory with associative memory (cf. affine transformations).



Representation of invariants

We distinguish between simple and complex neurons, where C-neurons pool S-neurons in a network. One C-neuron with its S-neurons is a module which features in a layered hierarchy.
An **invariant** representation is a particular activation pattern of different modules: a template.

https://bit.ly/3ERb8Hi (poster + data)

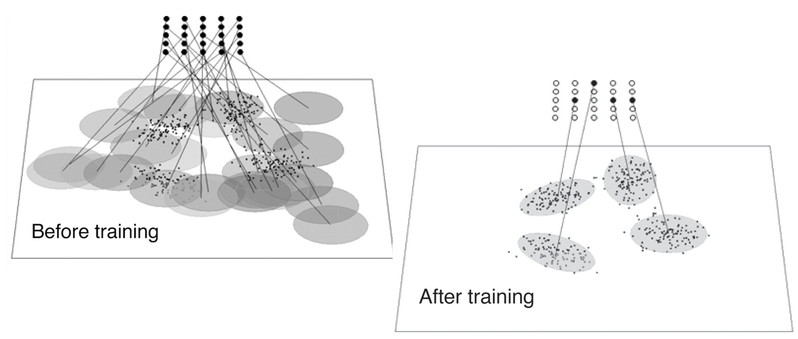
Two INSERT-statements

Invariance to auditory or visual translations (such as scale, duration, acoustic pitch height) can be built up by using two separate INSERT-operations for new data:

- INSERT-1: best rank-*r* approximation of a matrix is computed. A set of templates (the invariant properties) is first expressed as a matrix, which represents the specific activation patterns of HW-modules, then reduced using Singular Value Decomposition (SVD) and Principal Component Analysis (PCA).
- INSERT-2: random projections are used for dimensionality reduction. This is computationally less intensive but does not result in an outcome with correlated candidates.

INSERT-1 is slower but retains correlations between templates. It is like the human cortex (associative memory).
INSERT-2 is faster but the relations between templates necessary for invariance are lost. It is like the human hippocampus (episodic memory).

Output Our system correctly identifies the templates that represent invariants (Fig. 1). This implementation models two biologically plausible systems. After specialization (INSERT-1), invariants can be combined with analytical models as an alternative to memorization (INSERT-2).



Further work The current implementation is experimental and in R. Planned extensions include vowel recognition using a combination of invariant specialization (black box) and acoustic transformations (white box).

References

Poggio and Anselmi (2016). *Visual Cortex and Deep Networks: Learning Invariant Representations*. MIT Press.

Vallabha et al. (2007) *Unsupervised learning of vowel categories from infant-directed speech*, PNAS 104 (33)