Tiled Convolutional Neural Networks

Quoc V. Le, Jiquan Ngiam, Zhenghao Chen, Daniel Chia, Pang Wei Koh, and Andrew Y. Ng

1 Abstract

Convolutional neural networks (CNNs) have been successfully applied to many tasks such as digit and object recognition. Using convolutional (tied) weights significantly reduces the number of parameters that have to be learned, and also allows translational invariance to be hard-coded into the architecture. In this paper, we consider the problem of learning invariances, rather than relying on hard-coding. We propose *tiled* convolutional neural networks (Tiled CNNs), which use a regular "tiled" pattern of tied weights that does not require that adjacent hidden units share identical weights, but instead requires only that hidden units k steps away from each other to have tied weights. By pooling over neighboring units, this architecture is able to learn complex invariances (such as scale and rotational invariance) beyond translational invariance. Further, it also enjoys much of CNNs' advantage of having a relatively small number of learned parameters (such as ease of learning and greater scalability). We provide an efficient learning algorithm for Tiled CNNs based on Topographic ICA, and show that learning complex invariant features allows us to achieve highly competitive results for both the NORB and CIFAR-10 datasets.

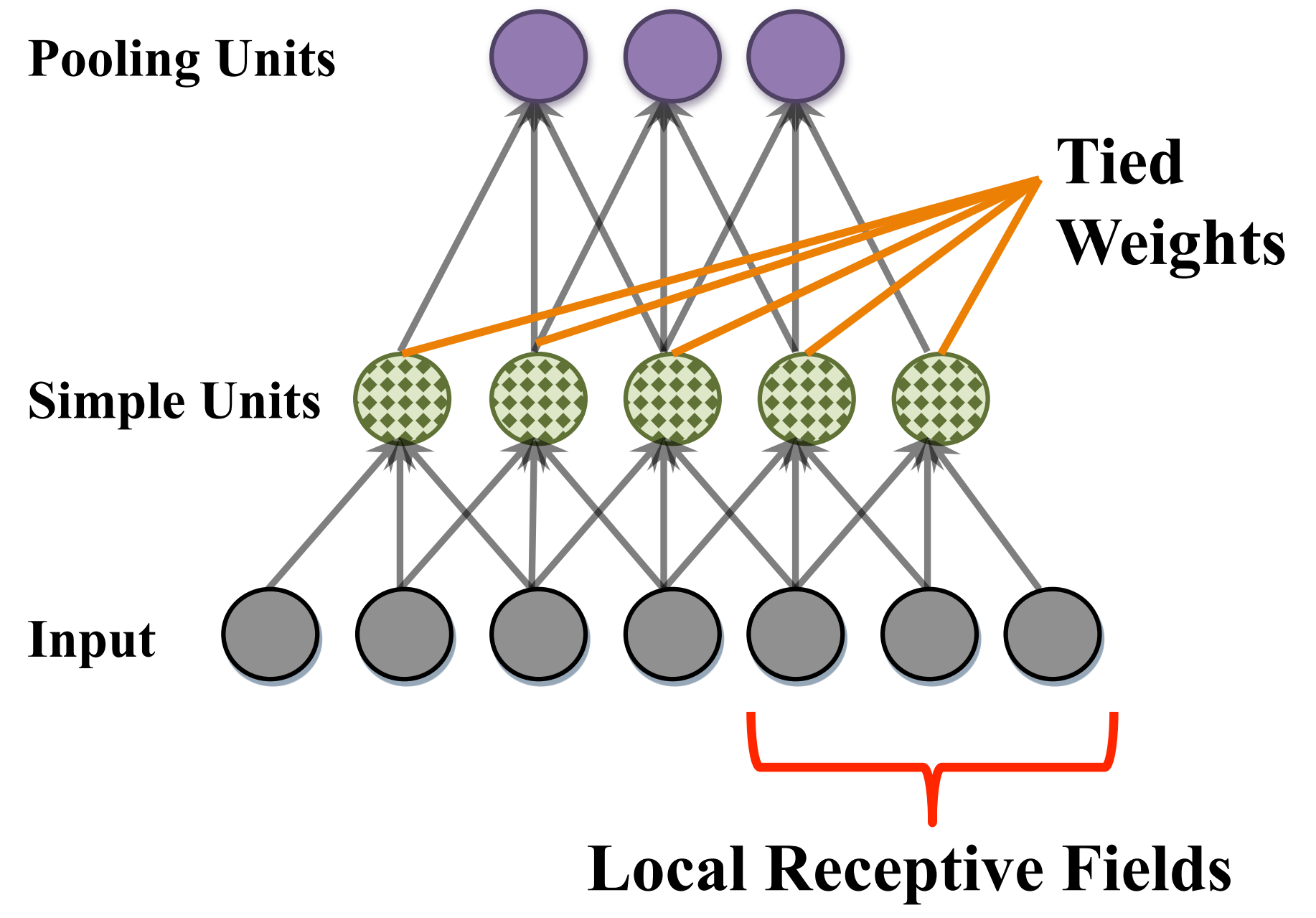
2 Motivation

Convolutional neural networks [1] work well for many recognition tasks:

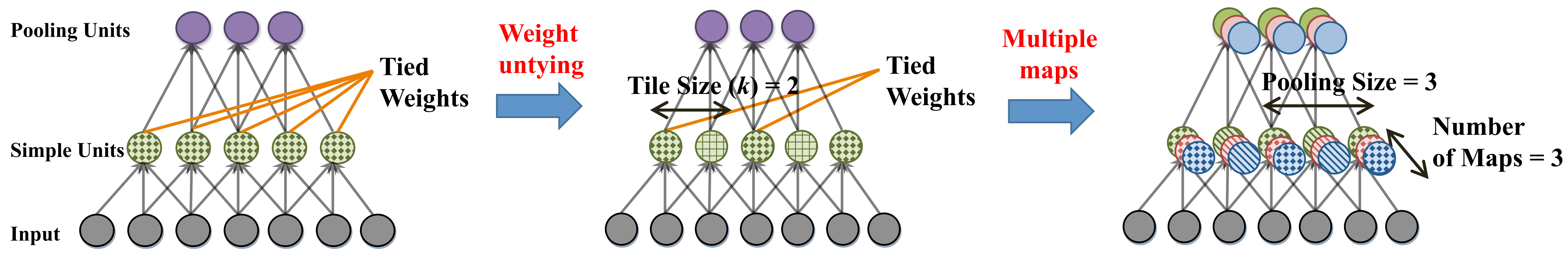
- Local receptive fields for computational reasons
- Weight sharing gives translational invariance

However, weight sharing can be restrictive because it prevents us from learning other kinds of invariances.

Convolutional Neural Networks



3 Tiled Convolutional Neural Networks

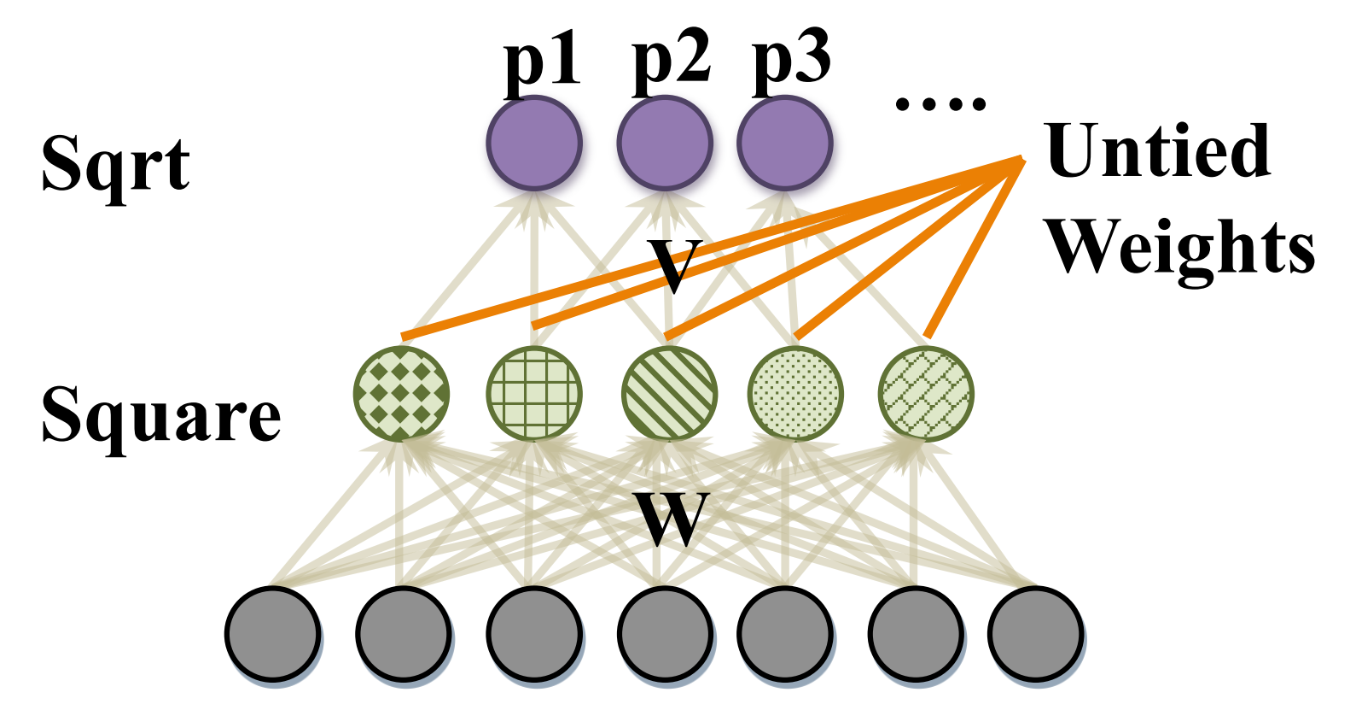


CNN

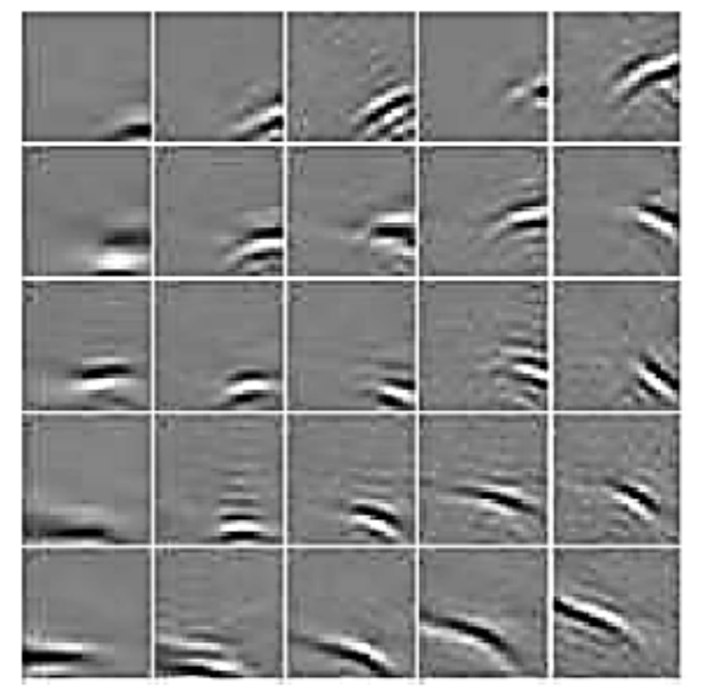
Tiled CNN

Tiled CNN with multiple feature maps
(Our model)

4 Pretraining with Topographic ICA



TICA network architecture



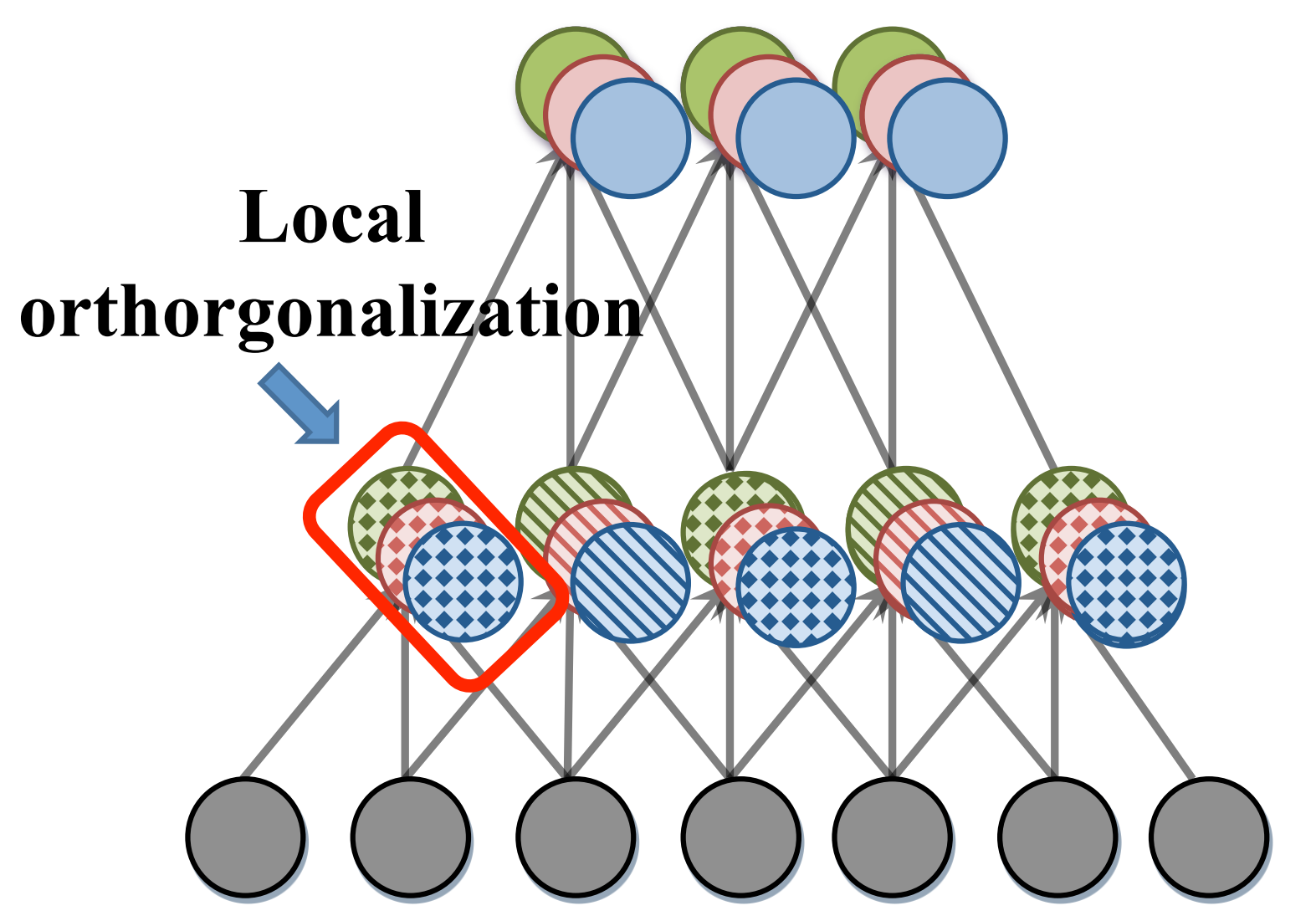
TICA first layer filters
(2D topography, 25 rows of W).

$$\underset{W}{\text{minimize}} \quad \sum_{t=1}^{T}\sum_{i=1}^{m} p_i(x^{(t)}; W, V), \ \text{subject to} \quad WW^T = \mathbf{I}$$

$$p_i(x^{(t)}; W, V) = \sqrt{\sum_{k=1}^{m} V_{ik}(\sum_{j=1}^{n} W_{kj}x_j^{(t)})^2}$$

- Algorithms for pretraining convolutional neural networks [2,3] do not use untied weights to learn invariances.
- TICA can be used to pretrain Tiled CNNs because it can learn invariances even when trained only on unlabeled data [4, 5].

Local Orthorgonalization



Overcompleteness (multiple maps) can be achieved by local orthogonalization. We localize neurons that have identical receptive fields

5 Algorithm

Algorithm 1 Unsupervised pretraining of Tiled CNNs with TICA (line search)

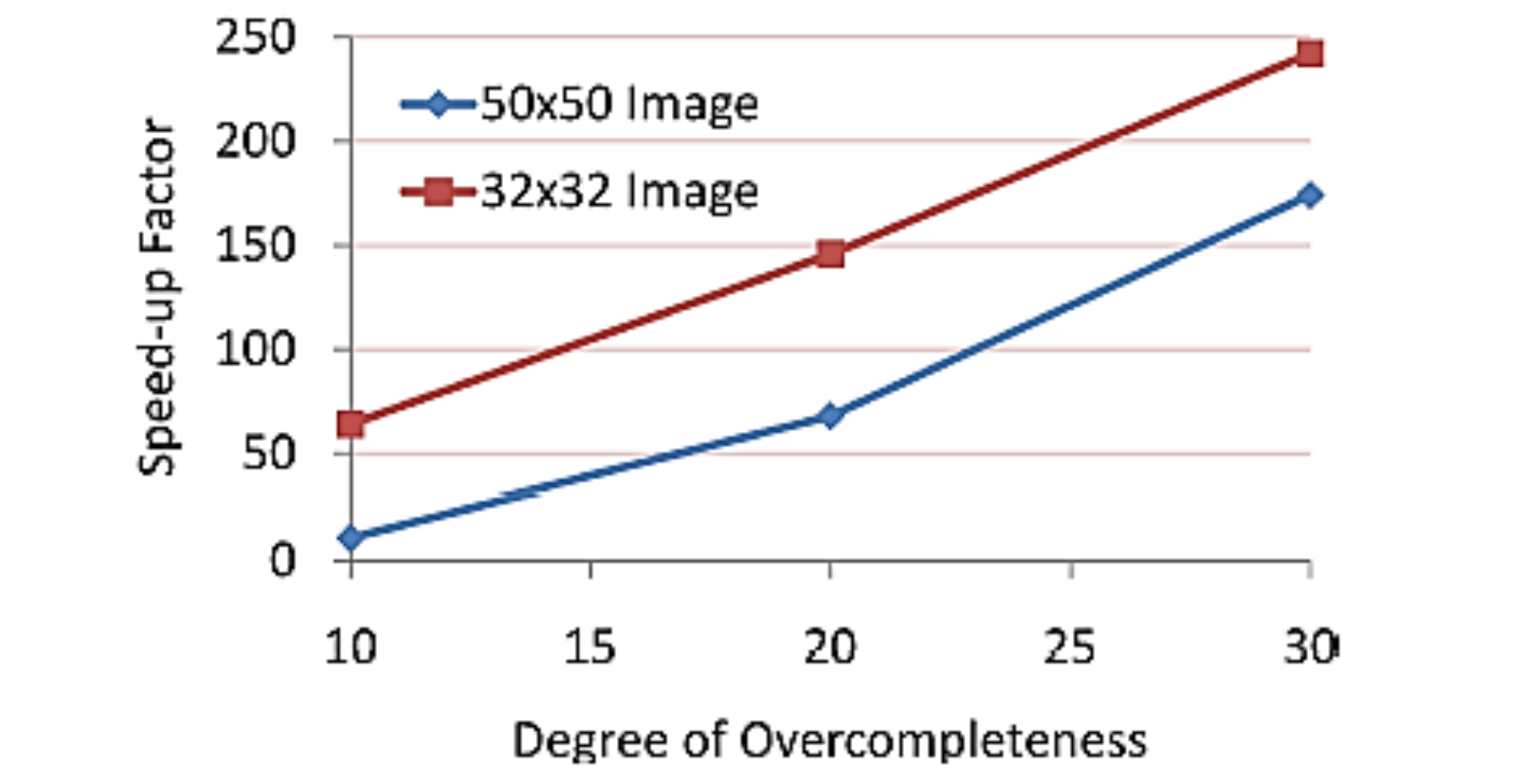
```
Input: $\{x^{(t)}\}_{t=1}^{T}, W, V, k, s$        // $k$ is the tile size, $s$ is the receptive field size
Output: $W$
  repeat
    $f^{old} \leftarrow \sum_{t=1}^{T}\sum_{i=1}^{m}\sqrt{\sum_{k=1}^{m} V_{ik}\left(\sum_{j=1}^{n} W_{kj}x_j^{(t)}\right)^2}$, $g \leftarrow \frac{\partial\left[\sum_{t=1}^{T}\sum_{i=1}^{m}\sqrt{\sum_{k=1}^{m} V_{ik}\left(\sum_{j=1}^{n} W_{kj}x_j^{(t)}\right)^2}\right]}{\partial W}$
    $f^{new} \leftarrow +\infty$, $\alpha \leftarrow 1$
    while $f^{new} \geq f^{old}$ do
      $W^{new} \leftarrow W - \alpha g$
      $W^{new} \leftarrow localize(W^{new}, s)$
      $W^{new} \leftarrow tie\_weights(W^{new}, k)$
      $W^{new} \leftarrow orthogonalize\_local\_RF(W^{new})$
      $f^{new} \leftarrow \sum_{t=1}^{T}\sum_{i=1}^{m}\sqrt{\sum_{k=1}^{m} V_{ik}\left(\sum_{j=1}^{n} W_{kj}^{new}x_j^{(t)}\right)^2}$
      $\alpha \leftarrow 0.5\alpha$
    end while
    $W \leftarrow W^{new}$
  until convergence
```

Optimization for Sparsity at Pooling units

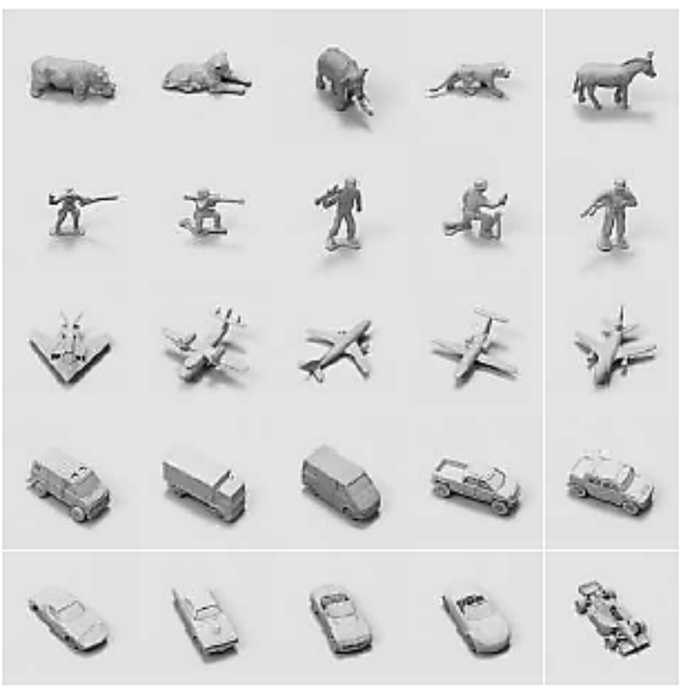
Projection step
Locality, Tie weight and Orthogonality contraints

6 TICA Speedup

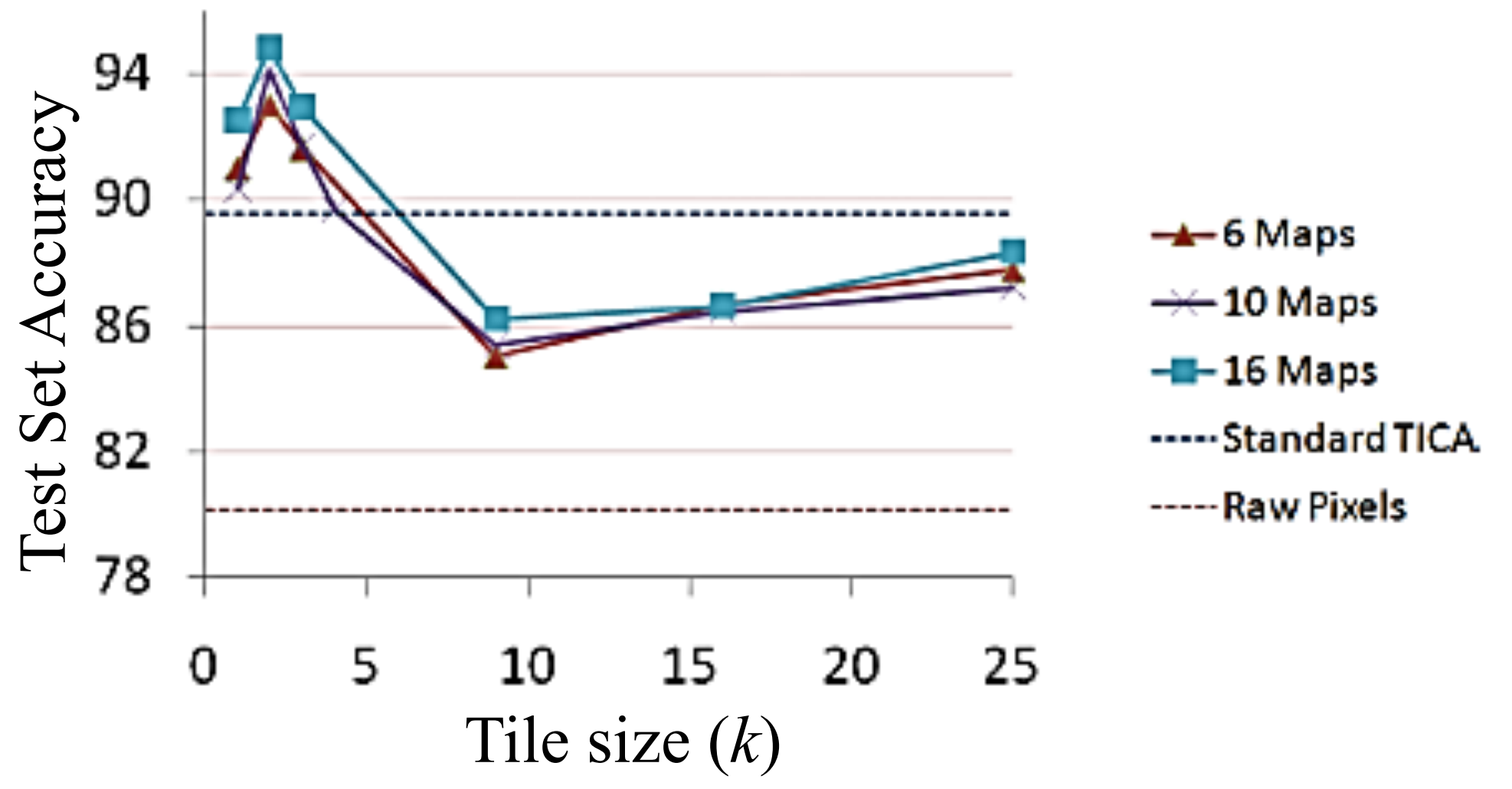
Speedup over non-local (fully-connected) TICA



7 Results on the NORB dataset



Algorithms	Accuracy
Deep Tiled CNNs [this work]	**96.1%**
CNNs [LeCun et al]	94.1%
3D Deep Belief Networks [Nair & Hinton]	93.5%
Deep Boltzmann Machines [Salakhutdinov et al]	92.8%
TICA	89.6%
SVMs	88.4%

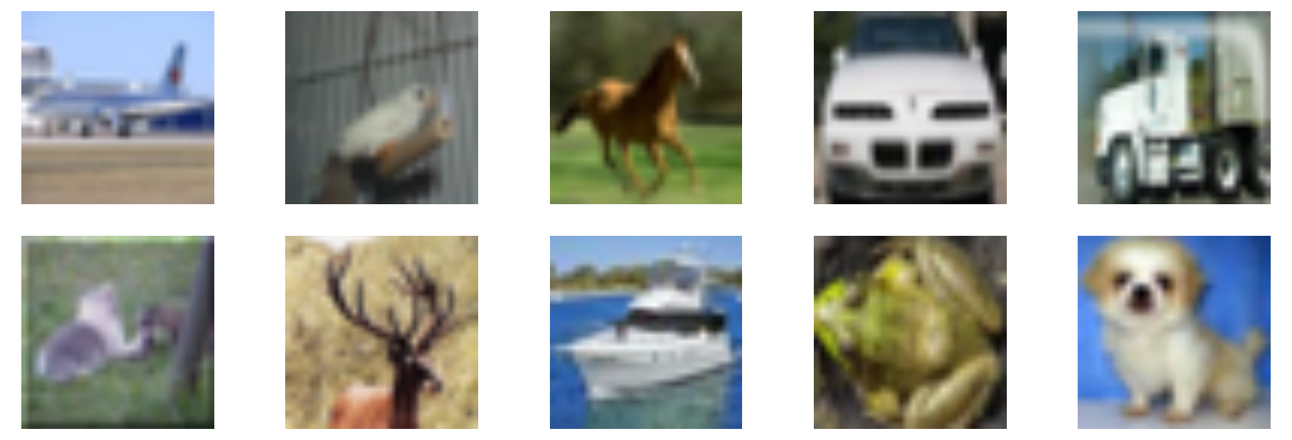


Tiled CNNs are more flexible and usually better than fully convolutional neural networks.

Pretraining with TICA finds invariant and discriminative features and works well with finetuning.

State-of-the-art results on NORB

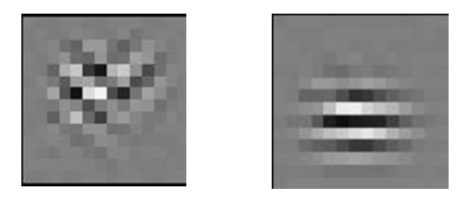
8 Results on the CIFAR-10 dataset



Algorithms	Accuracy
LCC++ [Yu et al]	74.5%
Deep Tiled CNNs [this work]	**73.1%**
LCC [Yu et al]	72.3%
mcRBMs [Ranzato & Hinton]	71.0%
Best of all RBMs [Krizhevsky et al]	64.8%
TICA	56.1%
Raw pixels	41.1%

Evaluating benefits of convolutional training
Training on 8x8 samples and using these weights in a Tiled CNN obtains only 51.54% on the test set compared to 58.66% using our proposed method.

Visualization:



Networks learn concepts like edge detectors, corner detectors
Invariant to translation, rotation and scaling

9 References

[1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient based learning applied to document recognition. *Proceeding of the IEEE*, 1998.
[2] H. Lee, R. Grosse, R. Ranganath, and A.Y. Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In ICML, 2009.
[3] M.A. Ranzato, K. Jarrett, K. Kavukcuoglu and Y. LeCun. What is the best multi-stage architecture for object recognition? In *ICCV*, 2009.
[4] A. Hyvarinen and P. Hoyer. Topographic independent component analysis as a model of V1 organization and receptive fields. *Neural Computation*, 2001.
[5] A. Hyvarinen, J. Hurri, and P. Hoyer. *Natural Image Statistics*. Springer, 2009.
[6] K. Kavukcuoglu, M. Ranzato, R. Fergus, Y. LeCun. Learning invariant features through topographic filter maps . In *CVPR*, 2009.
[7] K. Gregor, Y. LeCun. Emergence of Complex-Like Cells in a Temporal Product Network with Local Receptive Fields. *ARXIV*, 2010.

http://ai.stanford.edu/~quocle/