Exploring the ability of CNNs to generalise to previously unseen scales over wide scale ranges

Ylva Jansson and Tony Lindeberg, KTH Royal Institute of Technology, Stockholm, Sweden.

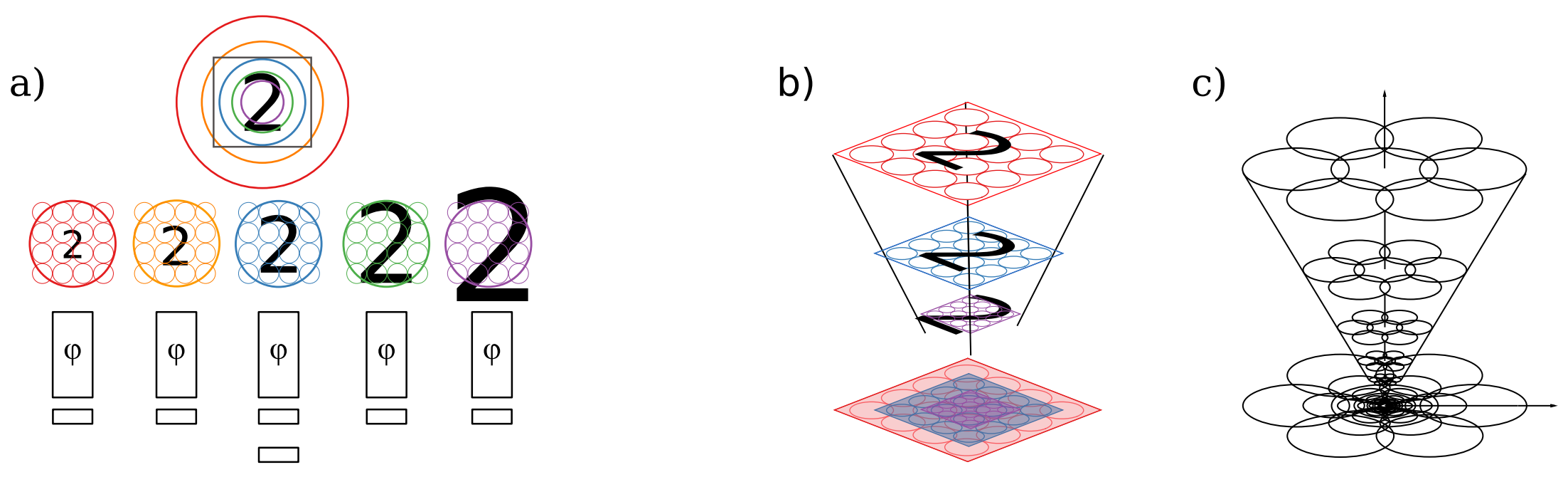
Contributions

- Detailed study of scale channel networks, obtained by applying the same CNN to multiple rescaled copies of the input image.
- Formalism for analysing such networks, showing provable scale covariance and scale invariance for continuous network model.
- Experimental investigation of the scale generalization properties of different types of networks, when performing testing at scales not spanned by the training data, showing: (i) very good scale generalization properties of our FovMax and FovAvg networks, (ii) poor generalization properties of vanilla CNNs and scale concatenation networks, and (iii) moderate scale generalization properties of a sliding windows networks.
- Experiments showing that our foveated scale channel networks perform better compared to the other approaches when training on small sample sets with large scale variability.

Scale channel networks

Apply same CNN to multiple rescaled copies of any image — *shared weights* between the scale channels.

- FovMax - with max pooling over the multiple scale channels.
- FovAvg - with average pooling over the multiple scale channels.



Provable scale covariance

With scaling operator S_s defined by

$$(\mathcal{S}_s f)(x) = f(S_s^{-1} x) = f_s(x) = f(\tfrac{x}{s}).$$

and feature maps $\Gamma^{(i)}$, the raw scale channels are *scale covariant*:

$$(\Gamma^{(i)} \mathcal{S}_t f)(x, c, s) = (\Gamma^{(i)} f)(x, c, st).$$

"Resizing of the input image corresponds to a mere shift in the scale channels of the scale channel network."

Proof in paper based on both (i) operator notation for the scaling group and (ii) an integral representation of the scale channel network.

Translational covariance: With shift operator $(\mathcal{D}_\delta f)(x) = f(x - \delta)$

$$(\Gamma^{(i)} \, \mathcal{D}_\delta \, f)(x, c, s) = (\Gamma^{(i)} f)(x - S_s \delta, c, s).$$

"Translational shift is rescaled depending upon the scale channel.'

Provable scale invariance

Given an *infinite* number of scale channels, either continuous or discrete γ^i for $\gamma > 1$, the supremum over the scale channels

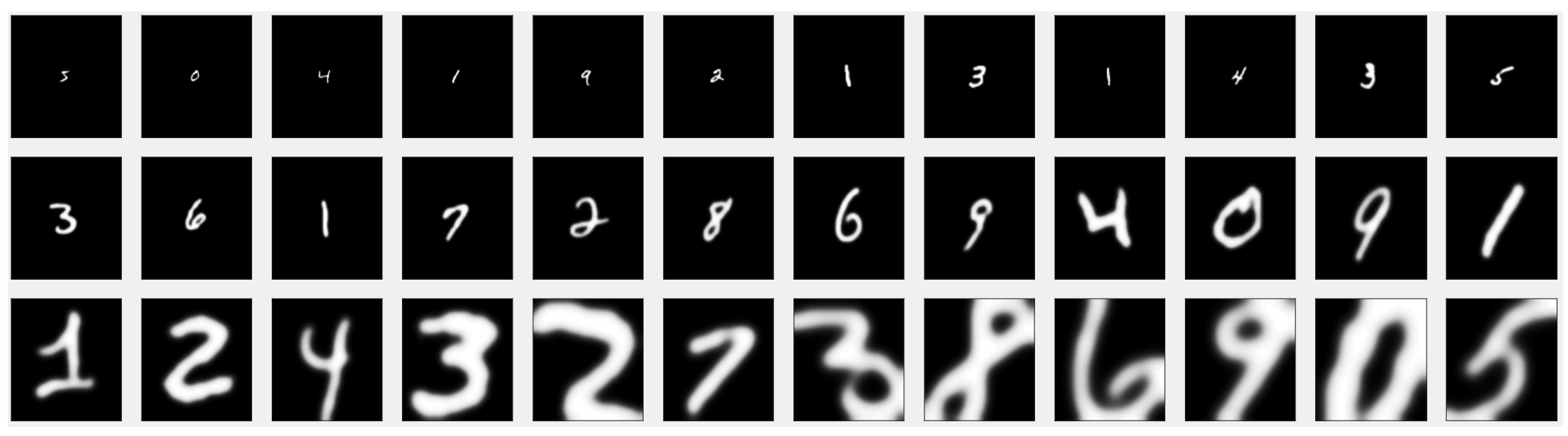
$$(\Lambda_{\sup} f)(x, c) = \sup_{s \in S} [(\phi_s f)(x, c, s)]$$

is *provably scale invariant* (proof in the paper).

Any other *permutation invariant pooling operation*, such as the average, is also provably scale invariant.

MNISTLargeScale dataset

Images from the original MNIST dataset 28 × 28 rescaled by factors between 1/2 and 8 and embedded in images of size 112 × 112.



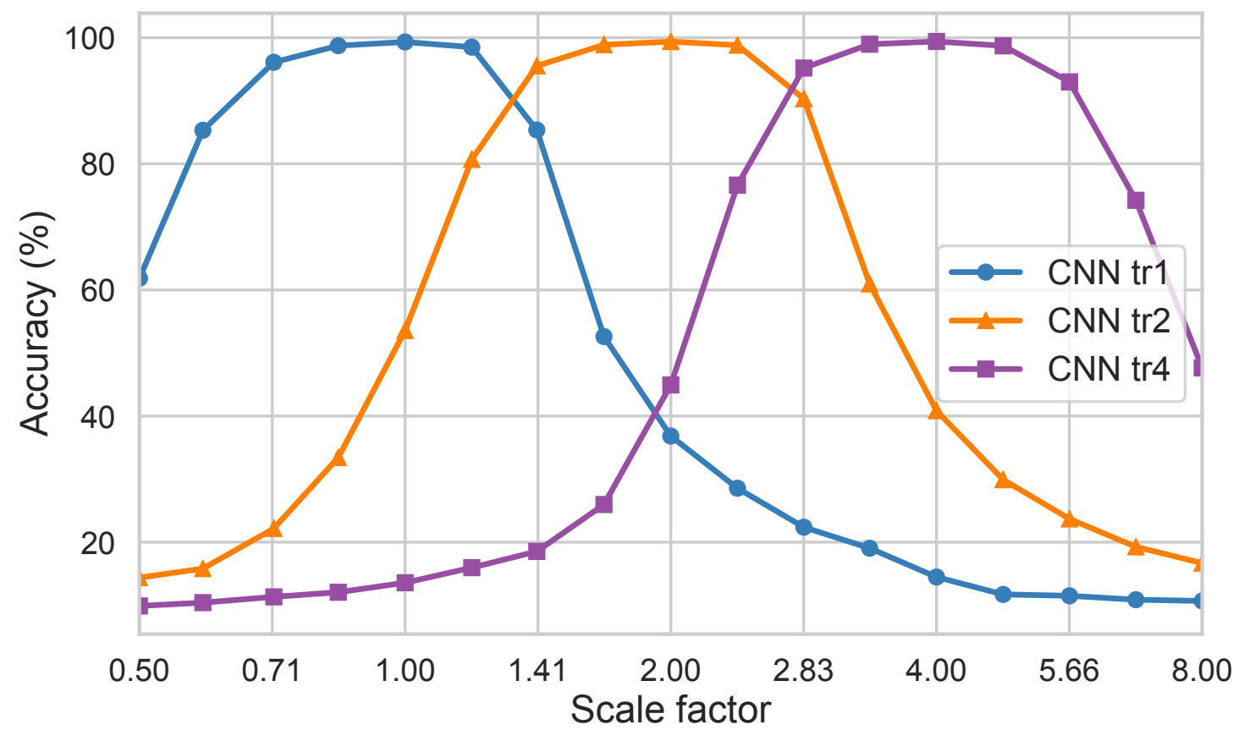
Training data for sizes 1, 2 and 4.

Testing data for sizes between 1/2 and 8 with relative size ratio $\sqrt[4]{2}$.

50 000 training images, 10 000 testing images, 10 000 validation images.

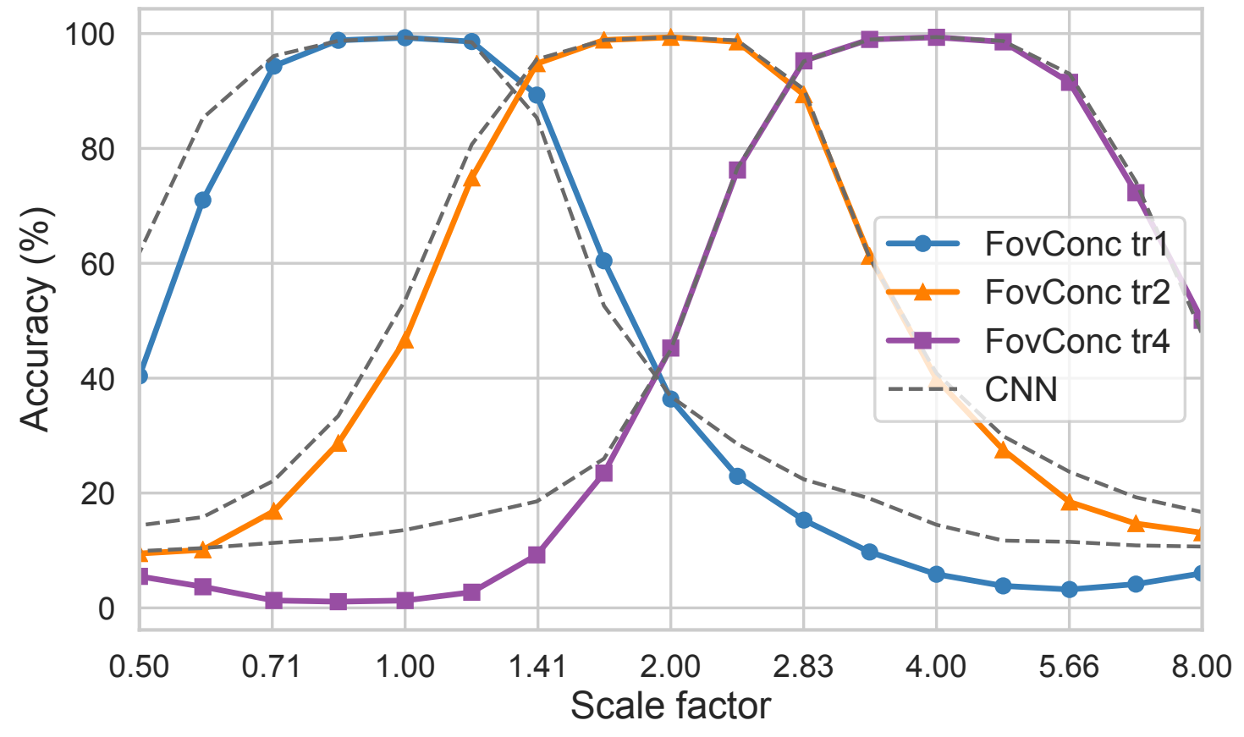
Scale generalization vanilla CNN

Trained for each one of sizes 1, 2 and 4.
Tested at all sizes in $[1/2, 8]$.



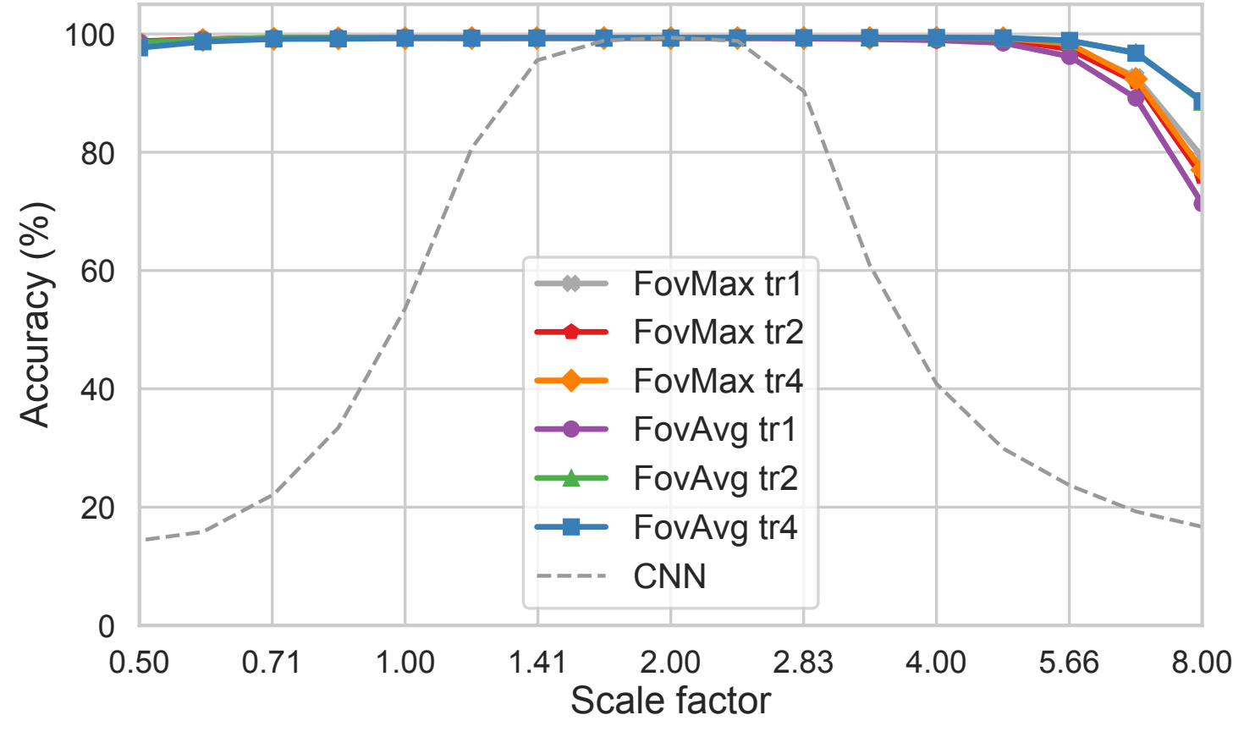
Poor generalization to previously unseen scales (no invariance mechanism)

Scale concatenation network



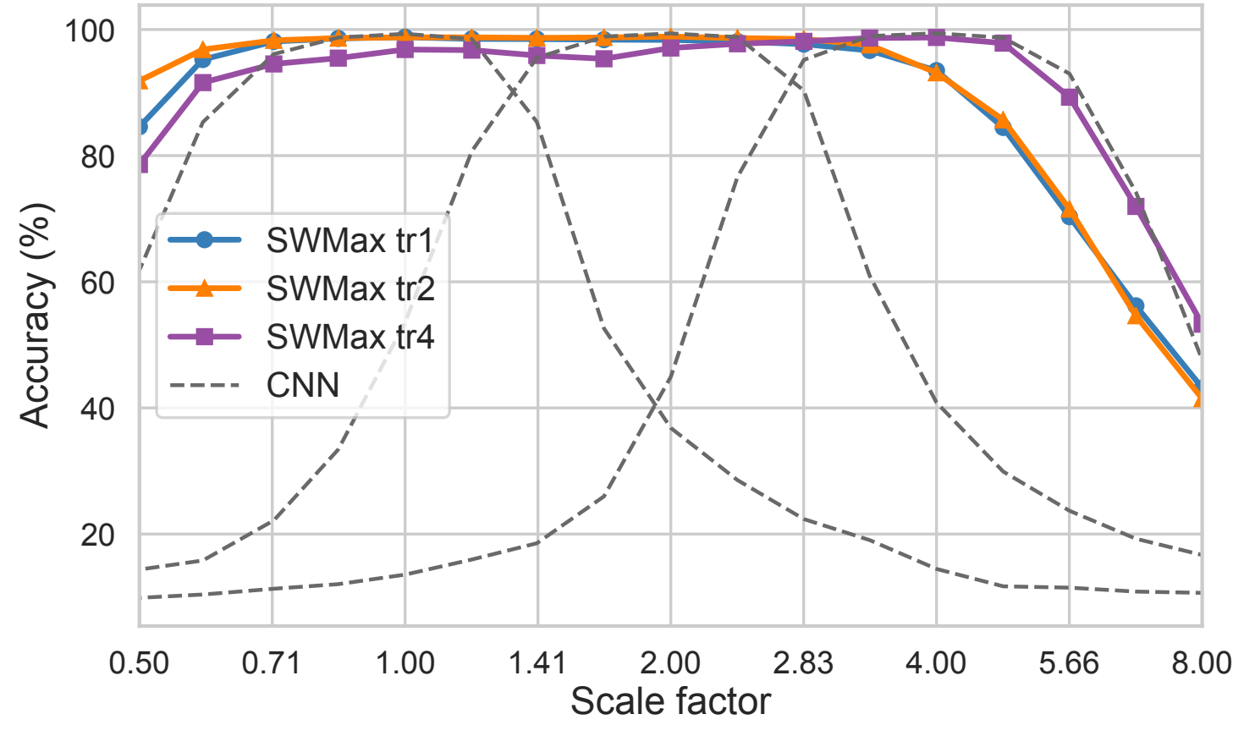
Poor generalization to previously unseen scales (no invariance mechanism)

FovMax and FovAvg networks



The invariance mechanism gives excellent generalization to previously unseen scales.

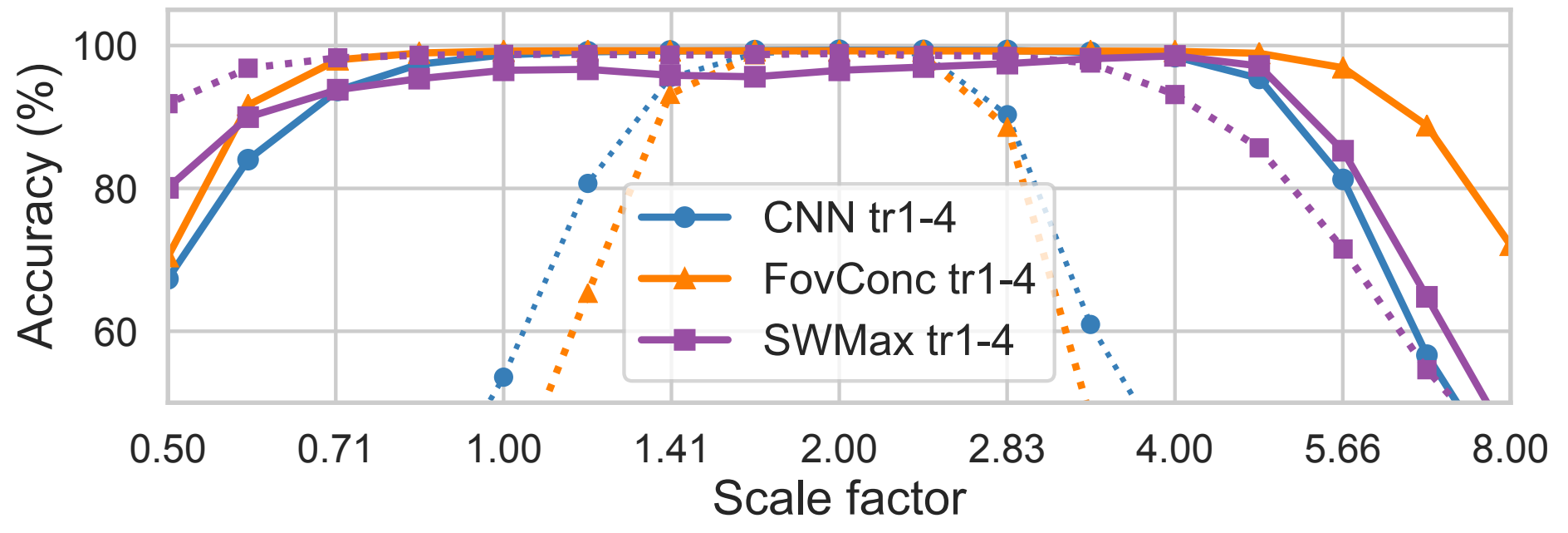
Sliding windows network



Some scale generalization but not as good as for FovMax and FovAvg.

Training over multiple scales

Both training data and testing data with uniform distribution over sizes between 1 and 4.

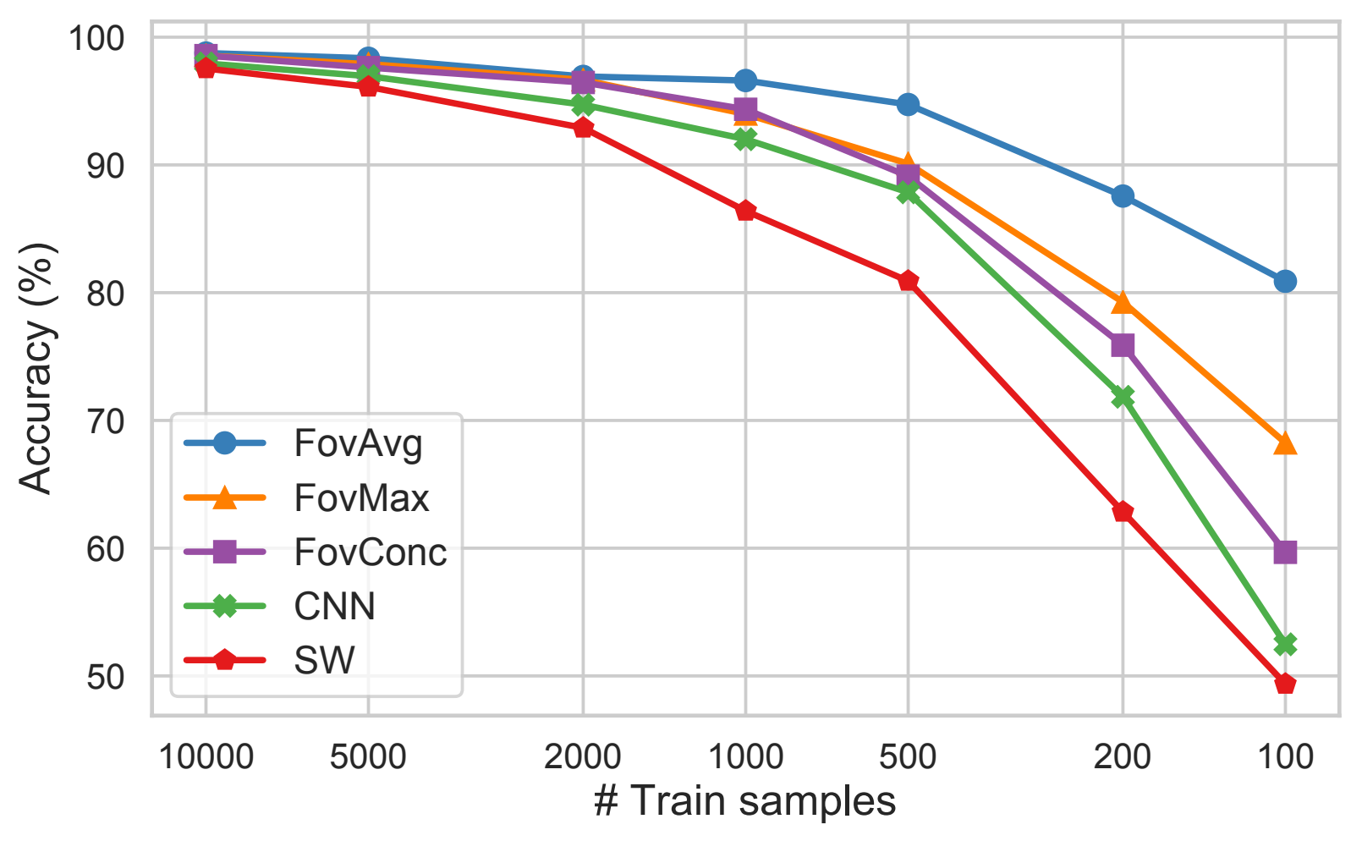


The vanilla CNN and the scale concatenation network are very much helped by training over multiple scales.

For the FovMax and FovAvg networks, training at a single scale is basically as good as training over multiple scales (see paper for details).

Small training sets, large scale varia

Both training data and testing data with uniform distribution over sizes between 1 and 4.



The FovMax and FovAvg networks make more efficient use of small amounts of training data that contain large scale variability.

Summary and conclusions

Presented methodology to *handle scaling transformations in deep networks* by scale channel networks.

Presented formalism to analyse scale channel networks and shown that they are *scale covariant* and translationally covariant.

Combined with max pooling or average pooling over the scale channels, the foveated scale channel networks are also *provably scale invariant*.

Shown that the FovMax and FovAvg are robust to scaling transformations and allow for *scale generalization*, with very good testing performance *at scales not spanned by the training data*.

Investigated limited scale generalization performance of vanilla CNNs, scale concatenation networks and sliding window networks.

Demonstrated that the FovMax and FovAvg networks lead to improvements for *multi-scale training data in the small sample regime*.

References

R. Ghosh and A. K. Gupta, "Scale steerable filters for locally scale-invariant convolutional neural networks", *arXiv preprint arXiv:1906.03861*.

Y. Jansson and T. Lindeberg, "MNISTLargeScale dataset" [Online]. Available at: https://www.zenodo.org/record/3820247. DOI:10.5281/zenodo.3820247, 2020.

Y. Jansson and T. Lindeberg, "Exploring the ability of CNNs to generalise to previously unseen scales over wide scale ranges", *arXiv preprint arXiv:2004.01536*, 2020. (contains more details)

A. Kanazawa, A. Sharma, and D. W. Jacobs, "Locally scale-invariant convolutional neural networks," *arXiv preprint arXiv:1412.5104*, 2014.

D. Laptev, N. Savinov, J. M. Buhmann, and M. Pollefeys, "TI-pooling: Transformation-invariant pooling for feature learning in convolutional neural networks", in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, pp. 289–297, 2016.

T. Lindeberg, "Scale-covariant and scale-invariant Gaussian derivative networks", *arXiv preprint arXiv:2011.14759*, 2020.

T. Lindeberg and L. Florack, "Foveal scale-space and linear increase of receptive field size as a function of eccentricity," Technical report ISRN KTH/NA/P--94/27--SE, Dept. of Numerical Analysis and Computer Science, KTH Royal Institute of Technology, 1994.

D. Marcos, B. Kellenberger, S. Lobry, and D. Tuia, "Scale equivariance in CNNs with vector fields," *arXiv preprint arXiv:1807.11783*, 2018.

P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated recognition, localization and detection using convolutional networks," *arXiv preprint arXiv:1312.6229*, 2013.

D. E. Worrall and M. Welling, "Deep scale-spaces: Equivariance over scale," *arXiv preprint arXiv:1905.11697*, 2019.

Y. Xu, T. Xiao, J. Zhang, K. Yang, and Z. Zhang, "Scale-invariant convolutional neural networks," *arXiv preprint arXiv:1411.6369*, 2014.