UNIVERSITY OF
OXFORD

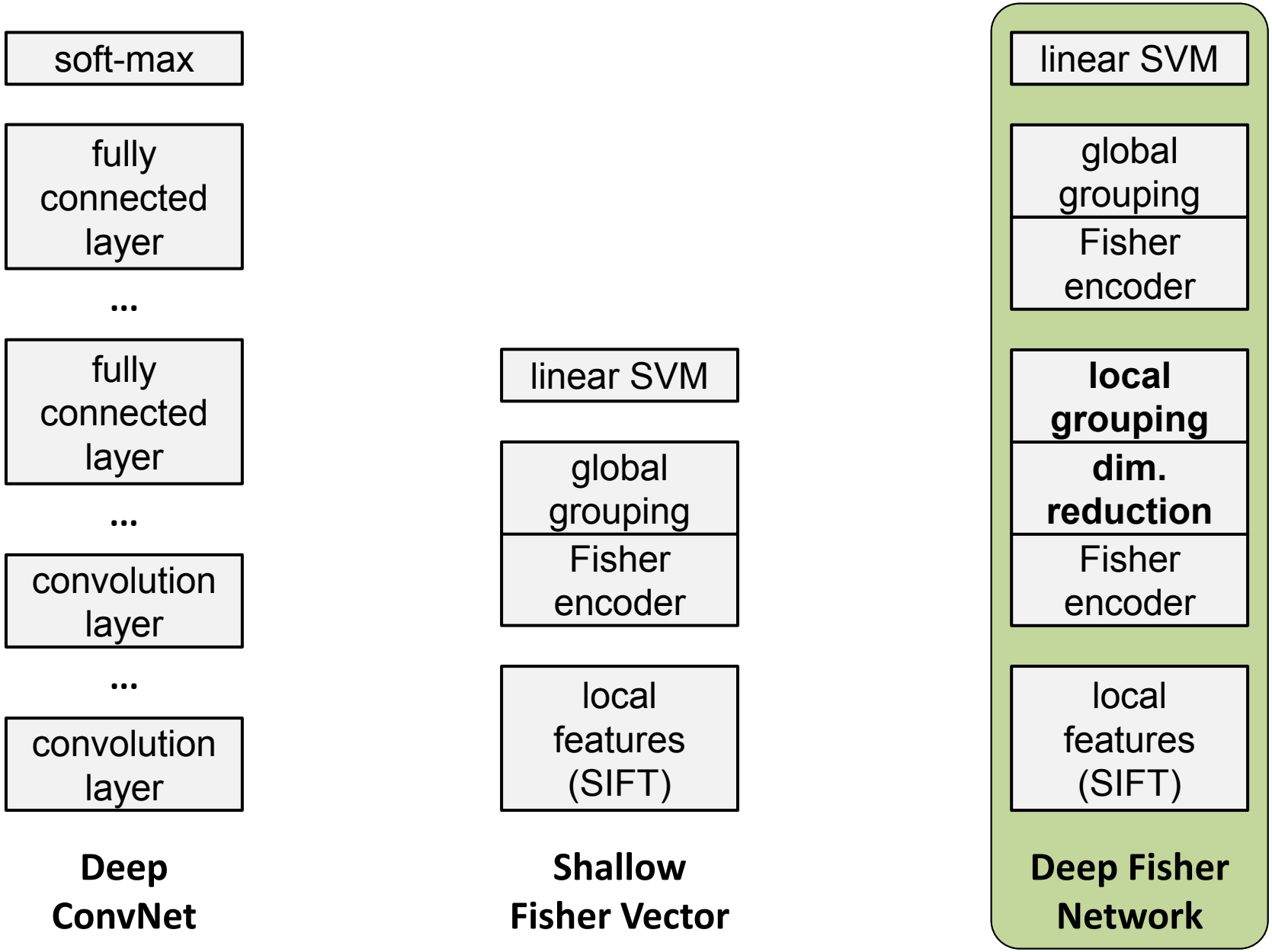
Deep Fisher Networks for Large-Scale Image Classification

Karén Simonyan, Andrea Vedaldi, Andrew Zisserman
Visual Geometry Group, University of Oxford

Deep learning achieves excellent performance in image classification.

Do hand-crafted image classification pipelines benefit from the increased depth too?

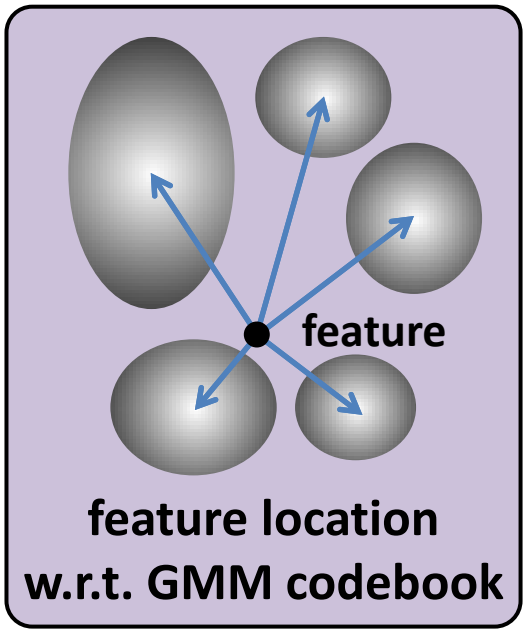
Image Classification Architectures



Deep Fisher Network

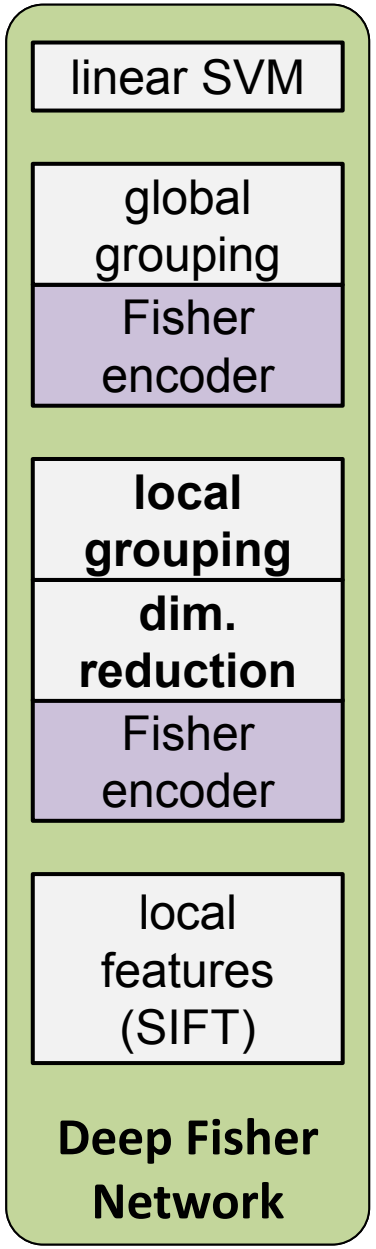
Why Fisher encoding?

- High-dimensional non-linear representation with small codebooks
- Outperforms other encodings (bag-of-words, sparse coding)



FisherNet

- Multiple Fisher layers made feasible by discriminative dimensionality reduction
- SIFT & colour features + 2 Fisher layers
- Learning: 2-3 days on 200 CPU cores (MATLAB + MEX implementation)



Large-Scale Image Classification

ImageNet challenge dataset:

- 1.2M images, 1K classes
- top-5 classification accuracy

Method	2010 challenge	2012 challenge
FV encoding	76.4%	72.7%
Deep FishNet	**79.2%**	**76.6%**
Deep ConvNet [Krizhevsky et al., 2012]	83.0%	81.8% 83.6% (5 ConvNets)
Deep ConvNet (our implement.)	83.2%	82.3%
Deep FishNet & Deep ConvNet	**85.6%**	**84.7%**