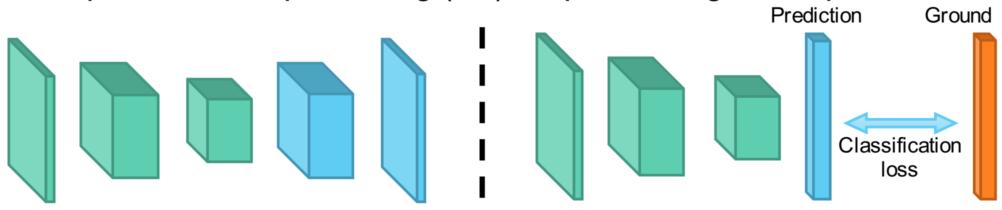
Augmenting Supervised Neural Networks with Unsupervised Objectives for Large-scale Image Classification

University of Michigan

Yuting Zhang, Kibok Lee, Honglak Lee

1 Supervised and unsupervised deep learning

- Deep representations can be obtained with
  - Unsupervised learning: informative preservation
    E.g. Stacked autoencoders (SAE), DBN, DBM
    Generating data from feature representations (related to invertibility)
  - Supervised learning: task-specific, not necessarily invertible
- Unsupervised deep learning (DL) as pretraining for supervised DL

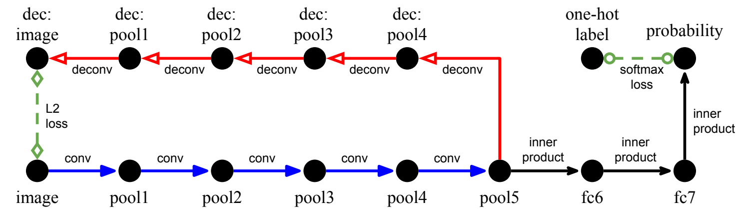


- Such pretraining became unnecessary given proper initialization and and large amount of labeled data.
- To revisit the importance of unsupervised deep learning, people incorporate unsupervised objectives into supervised training.

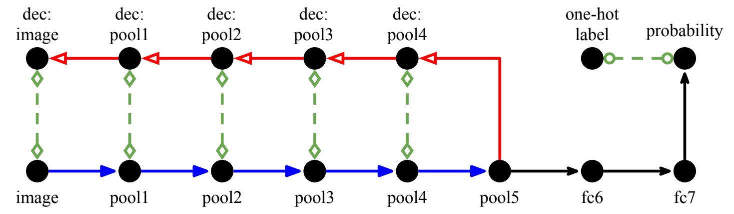
  Autoencoders + Classifiers
  - Ladder network: layer-wise skip links & pathway combinators
  - Stacked "what-where" AE (SWWAE): unpooling switches
- Promising results exist, but no evidence on both
  - Large amount of labeled data
  - Very deep networks

2 Augmenting classification networks with decoding pathways

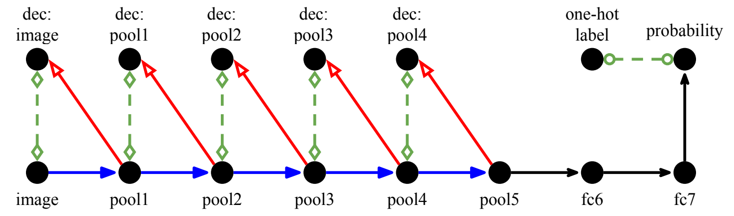
- SAE/ SWWAE -first



- SAE/ SWWAE -all



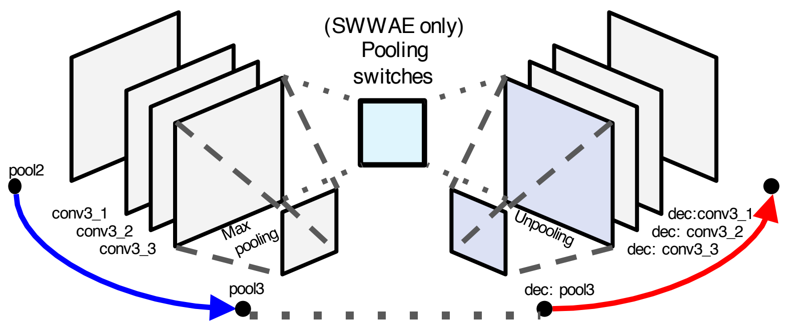
- SAE/ SWWAE -layerwise

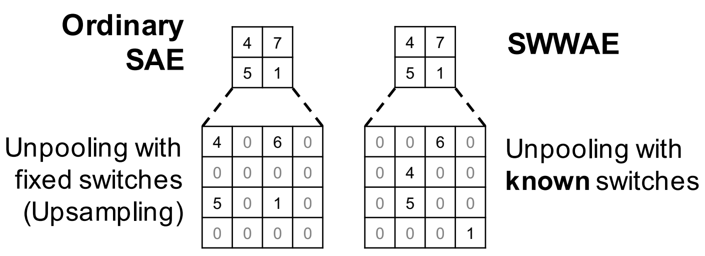


- ●: node; →: encoding pathway; ←: decoding pathway; →: classification pathway; ○- -○: reconstruction loss; ○- -○: classification loss.
- **Training procedure:**
  Step 1: Initialize the classification network with pretrained weights.
  Step 2: Train (randomly initialized) "layerwise" decoding pathways.
  Step 3: Train the top-down decoding pathways. (Inverting a network)
  Step 4: Finetune the entire augmented network. (Improving a network)

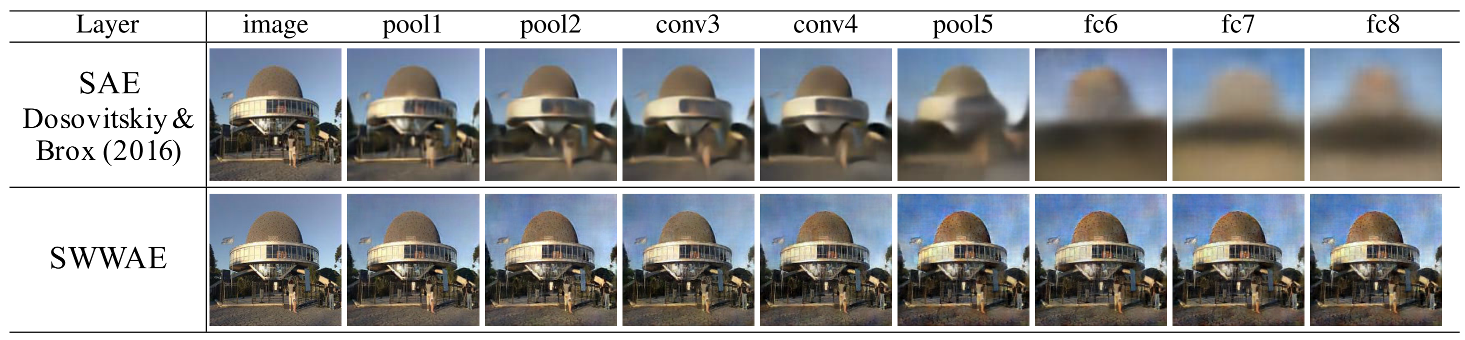
3 Invertibility of large-scale classification networks

a. Micro-architectures for SAE & SWWAE

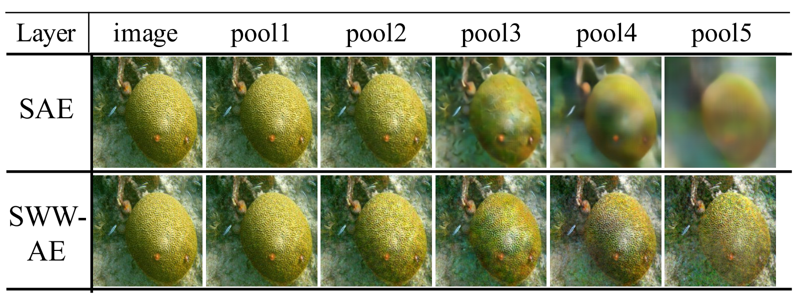




b. Inverting AlexNet



c. Inverting 16-layer VGGNet



d. Observations & Hypotheses

- Max-pooling is the main source of information loss (SWWAE sufficiently recovers it)
- Convolutional filters and non-linearity cause minor information loss

Take it as a helpful property and enhance it.

4 Improving large-scale classification networks with decoding pathways

a. Experiments

- 16-layer VGGNet on ILSVRC2012
- Rescale the shorter edge to 256px
- Singe crop: 224px patch at center
- Convolution: dense sampling

Sampling	Single crop				Convolution	
Errors	Top-1		Top-5		Top-1	Top-5
Model	Train	Val.	Train	Val.	Validation	
VGGNet (baseline)	17.43	29.05	4.02	10.07	26.97	8.94
+ SAE-first	15.36	27.70	3.13	9.28	26.09	8.30
+ SAE-all	15.64	27.54	3.23	9.17	26.10	8.21
+ SAE-layerwise	16.20	27.60	3.42	9.19	26.06	8.17
+ SWWAE-first	**15.10**	27.60	**3.08**	9.23	25.87	8.14
+ SWWAE-all	15.67	**27.39**	3.24	**9.06**	**25.79**	**8.13**
+ SWWAE-layerwise	15.42	27.53	3.32	9.10	25.97	8.20

b. Conclusions

- A simple and effective way to incorporate unsupervised objectives into large-scale classification network learning.
- We improved the image classification performance of the 16-layer VGGNet, a strong baseline model, by a noticeable margin.
- Comparison among the variants of our models
  - Pooling switch connections in SWWAE slightly benefit classification performance.
  - The decoding pathways mainly help the supervised objective reach a better optimum.
  - The layer-wise reconstruction loss can regularize the solution to the joint objective.

Main references:

- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks", in *NIPS*, 2012.
- Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition", in *ICLR*, 2015.
- Alexey Dosovitskiy and Thomas Brox, "Inverting visual representations with convolutional networks", in *CVPR*, 2016.
- Junbo Zhao, Michael Mathieu, Ross Goroshin, and Yann LeCun, "Stacked what-where auto-encoders", *arXiv:1506.02351*, 2015.

GitHub: bit.ly/cnn-dec

ICML 2016
THE 33RD INTERNATIONAL CONFERENCE ON MACHINE LEARNING