Supplementary Materials
Learning Deep Features for Scene Recognition using Places Database

Bolei Zhou[1], Agata Lapedriza[1,3], Jianxiong Xiao[2], Antonio Torralba[1], and Aude Oliva[1]

[1]Massachusetts Institute of Technology
[2]Princeton University
[3]Universitat Oberta de Catalunya

1 Experimental setup

In SUN397 experiment [7], the training size is 50 images per category. Experiments are ran on 10 Splits of train set and test set given in the dataset.

In MIT Indoor67 experiment [6], the training size is 100 images per category. Experiment is ran on 1 split of train set and test set given in the dataset. On the 10 splits randomly generated by ourself, the classification accuracy is $69.10\% \pm 1.62\%$

In the Scene15 experiment [3], the training size is 50 images per category. Experiments are ran on 10 random splits of train set and test set.

In the SUN Attribute experiment [5], the training size is 150 images per attribute. The report result is average precision. The splits of train set and test set are given in the paper.

In Caltech101 and Caltech256 experiment [1, 2], the training size is 30 images per category. The experiments are ran on 10 random splits of train set and test set.

In Stanford Action40 experiment [8], the training size is 100 images per category. Experiments are ran on 10 random splits of train set and test set. The reported result is classification accuracy.

In UIUC Event8 experiment [4], training size is 70 per category and the testing size is 60 images per category. The experiments are ran on 10 random splits of train set and test set.

2 Visualization

More visualization of the units in the ImageNet-CNN and Places-CNN is attached.

References

[1] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, 2007.

[2] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. 2007.

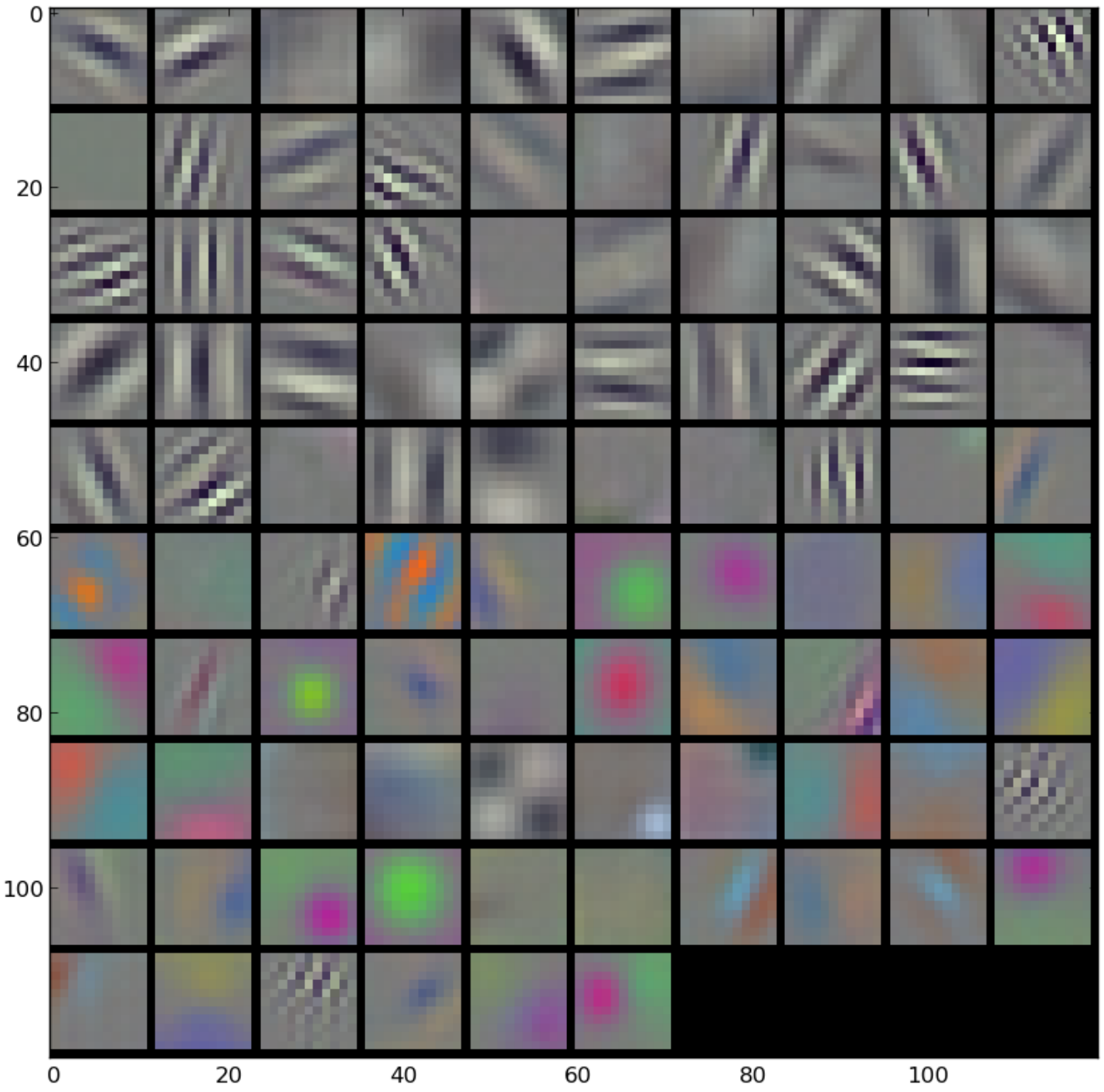
[3] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proc. CVPR*, 2006.

[4] L.-J. Li and L. Fei-Fei. What, where and who? classifying events by scene and object recognition. In *Proc. ICCV*, 2007.

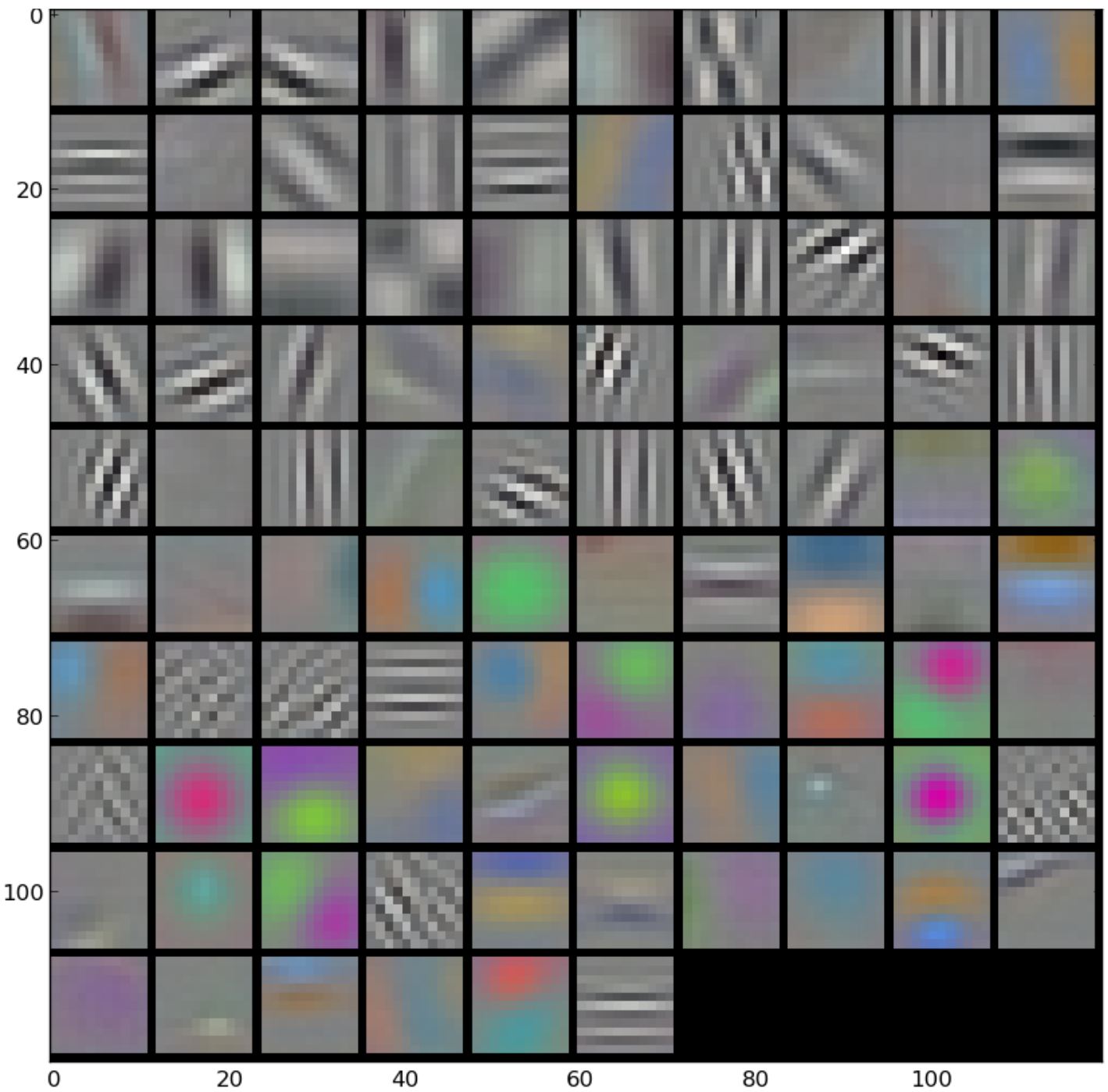
[5] G. Patterson and J. Hays. Sun attribute database: Discovering, annotating, and recognizing scene attributes. In *Proc. CVPR*, 2012.

[6] A. Quattoni and A. Torralba. Recognizing indoor scenes. In *Proc. CVPR*, 2009.
[7] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *Proc. CVPR*, 2010.
[8] B. Yao, X. Jiang, A. Khosla, A. L. Lin, L. Guibas, and L. Fei-Fei. Human action recognition by learning bases of action attributes and parts. In *Proc. ICCV*, 2011.

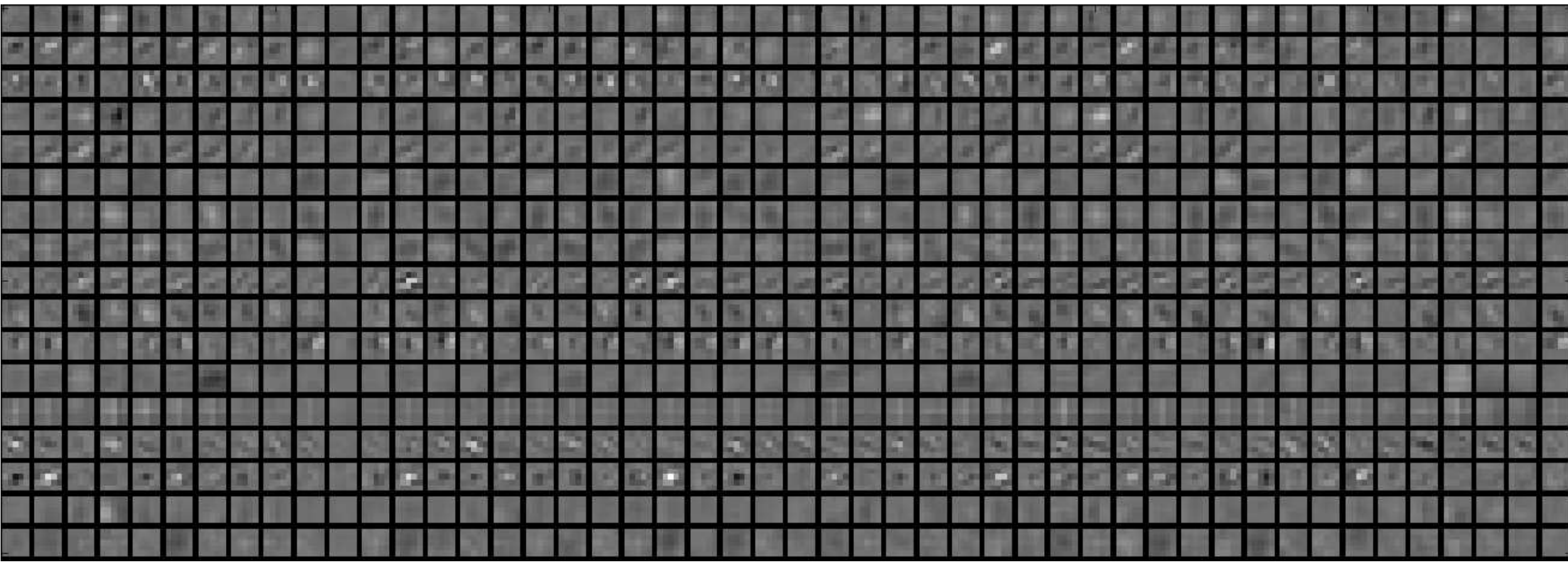
ImageNet-CNN Conv 1 units



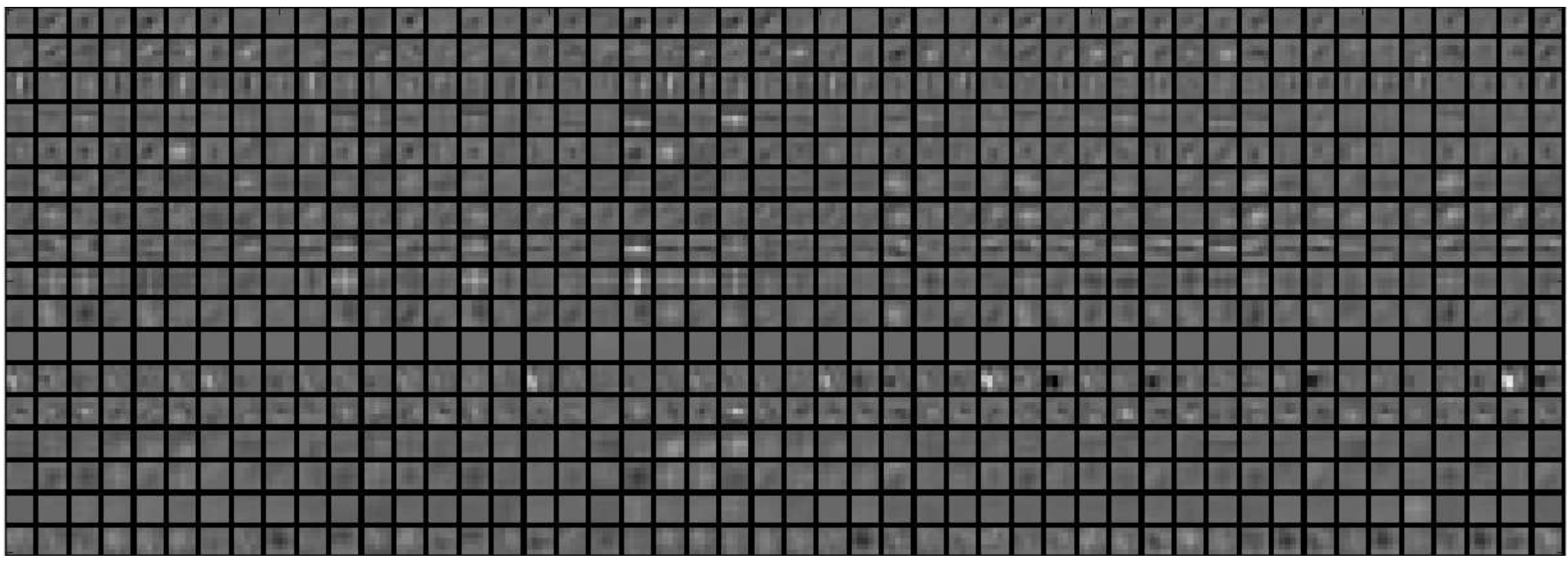
PLACES-CNN Conv 1 units



ImageNet-CNN Conv 2 units

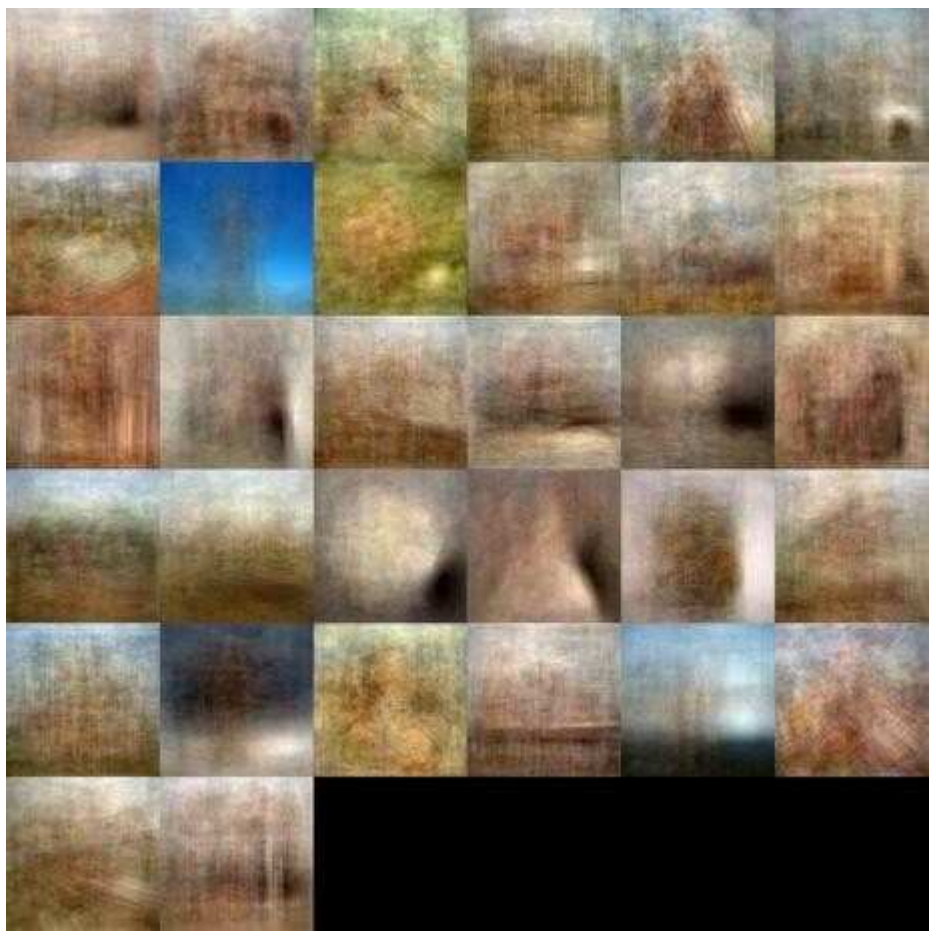


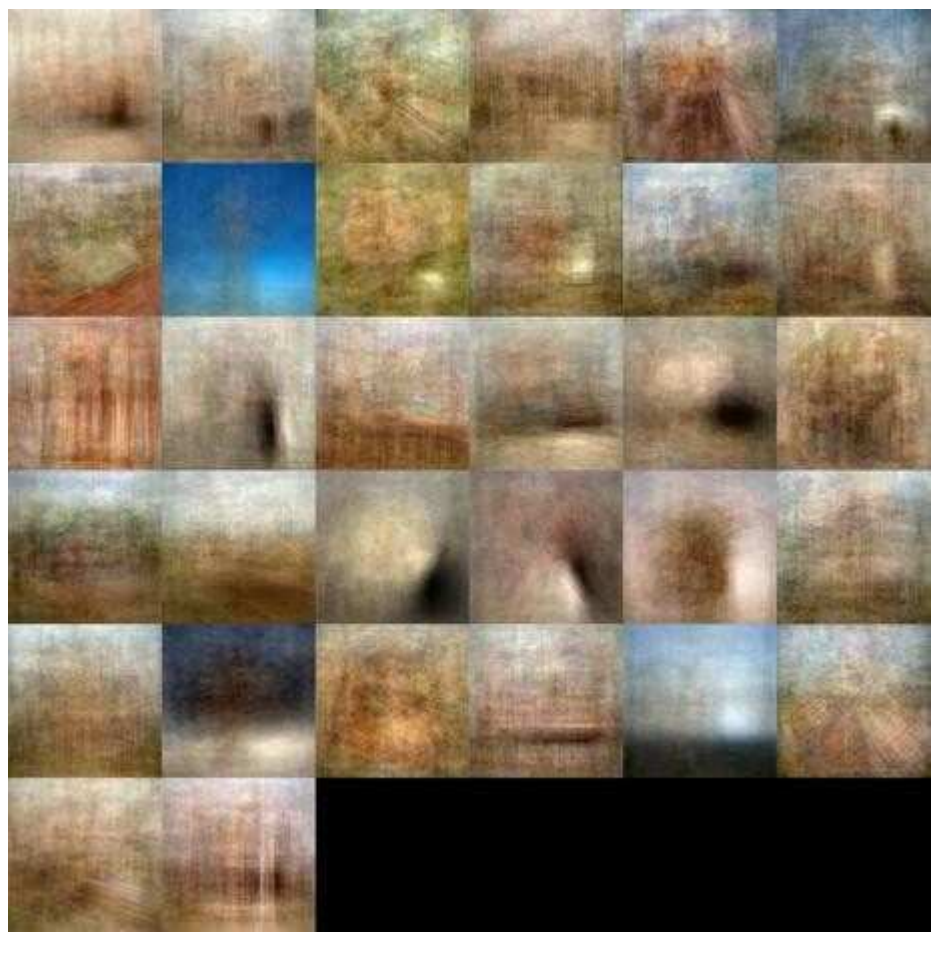
PLACES-CNN Conv 2 units

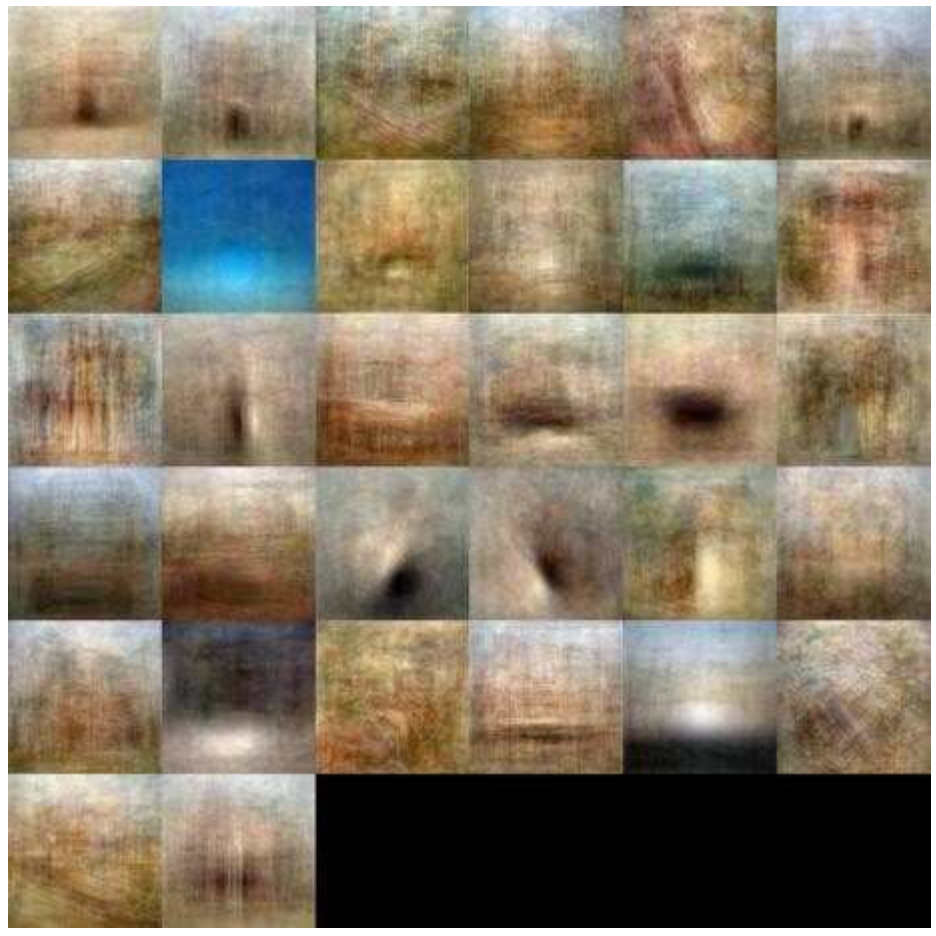


ImageNet-CNN Pool 2 units

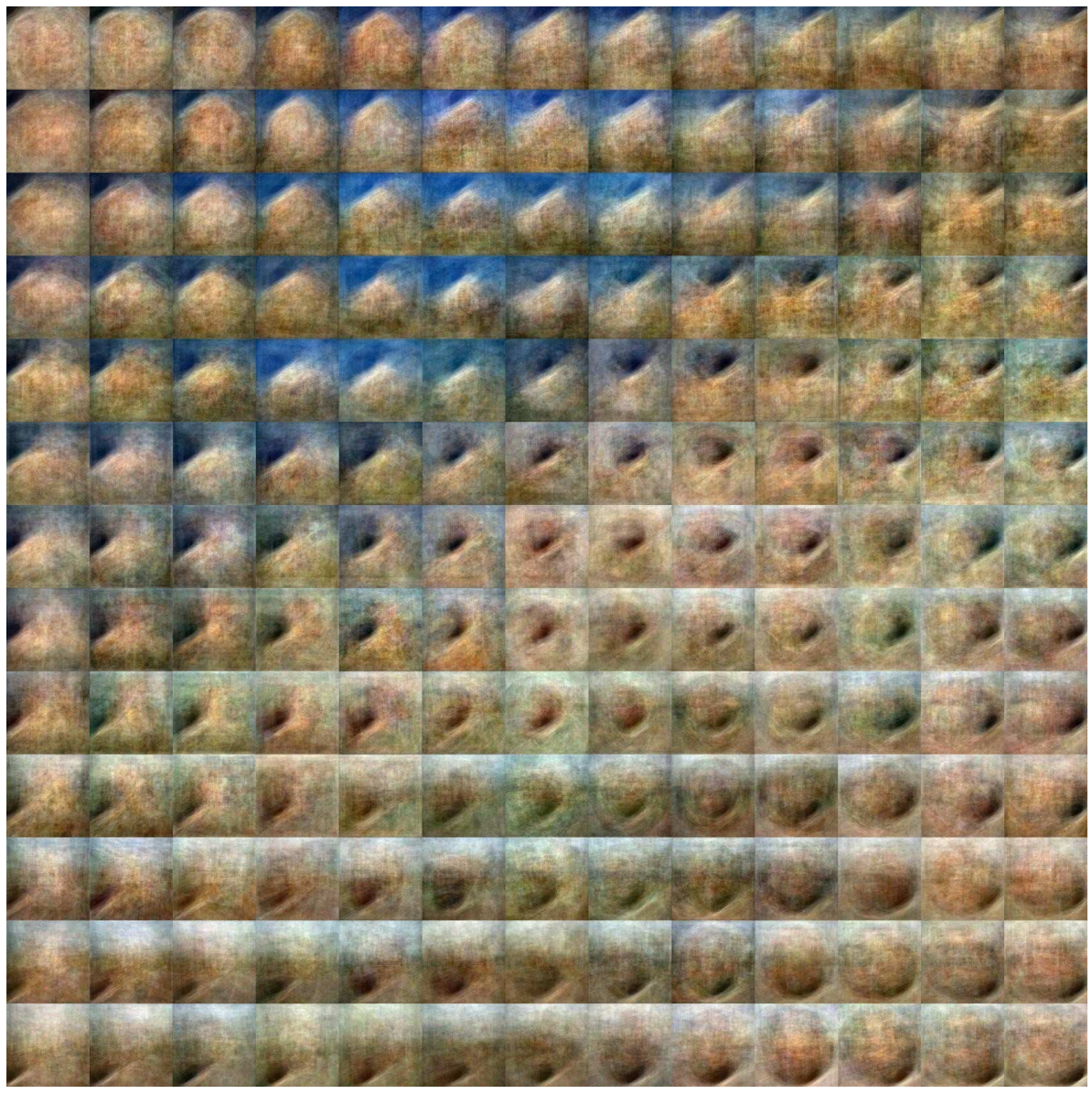
PLACES-CNN Pool 2 units



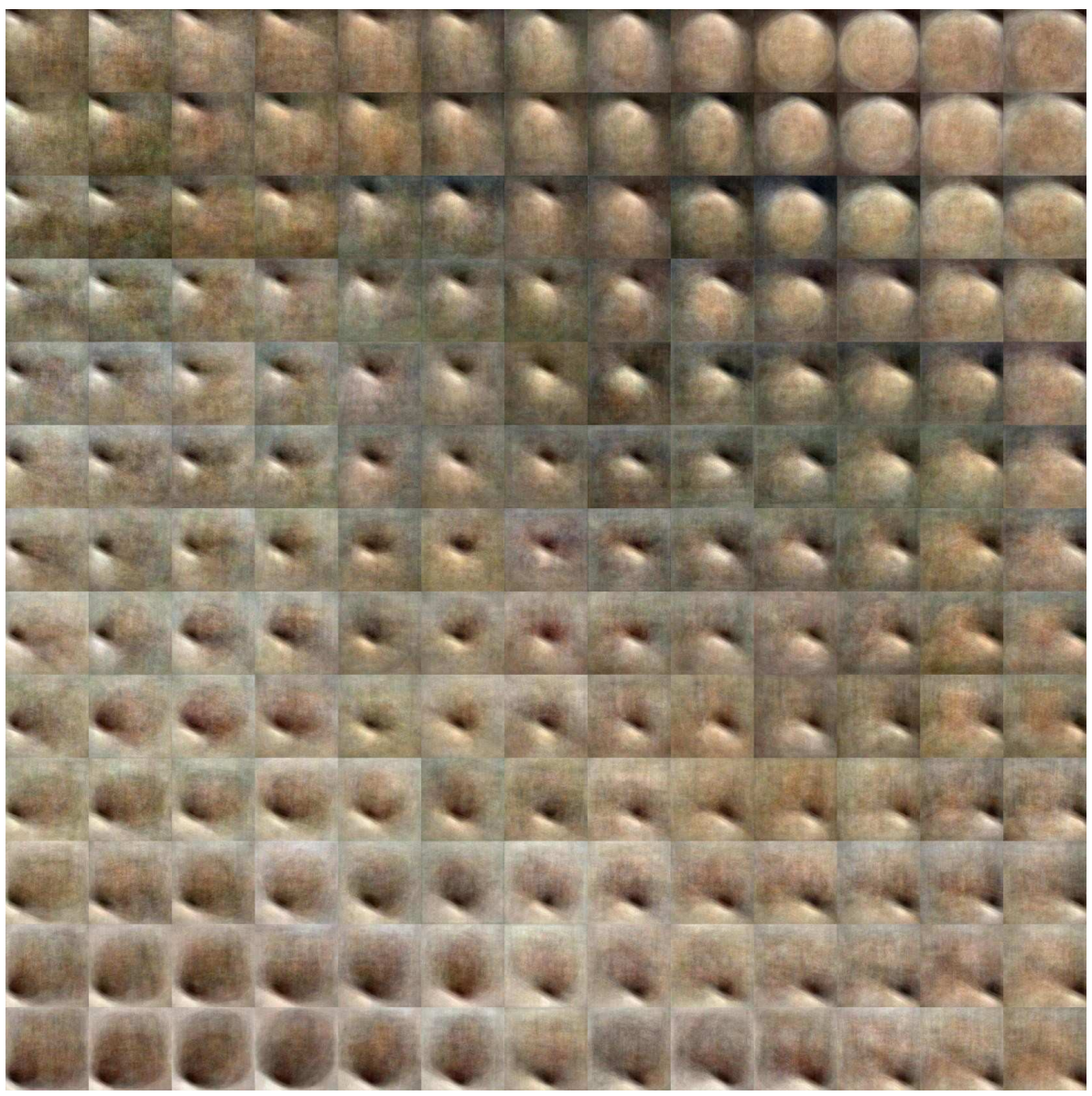


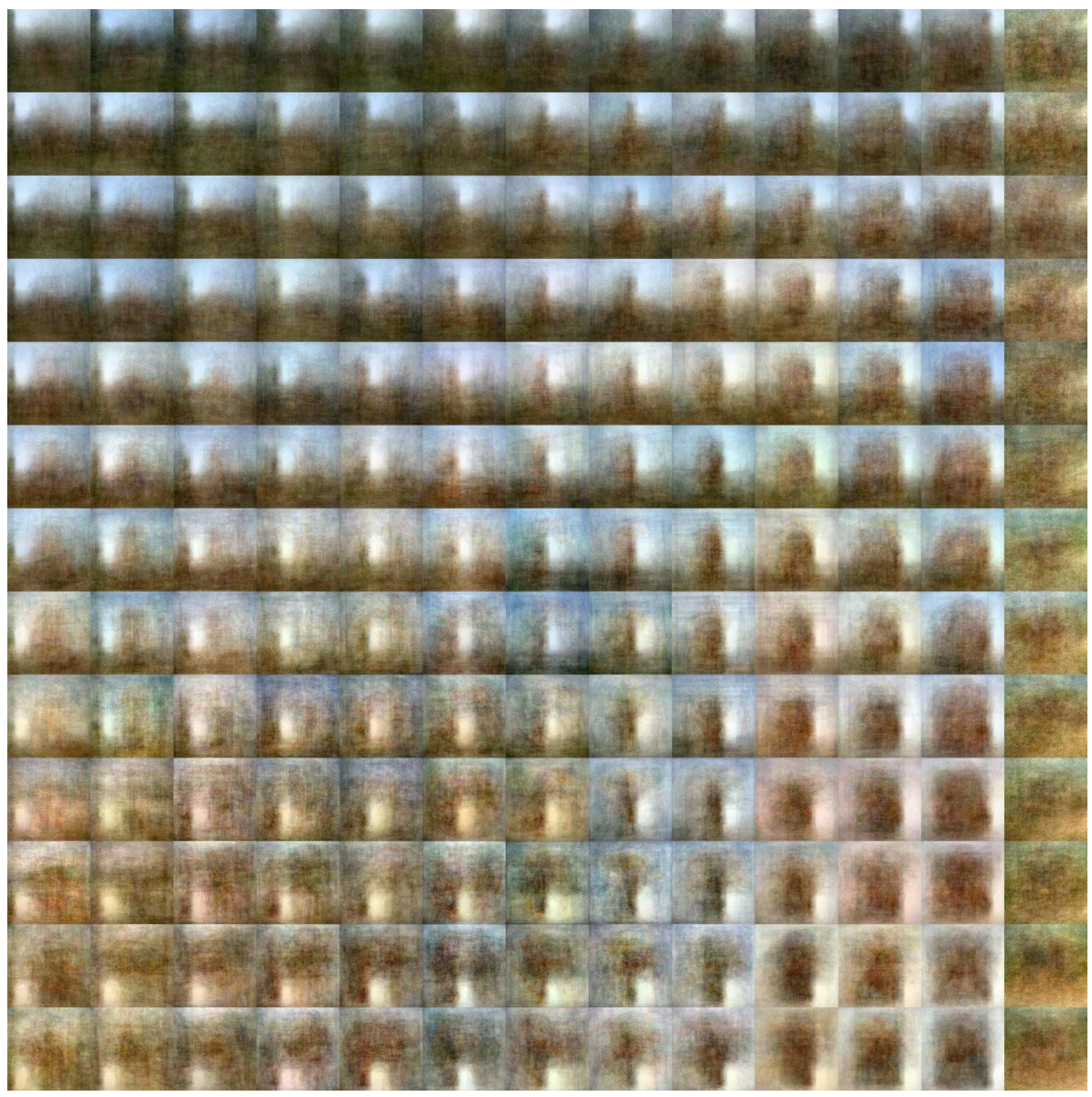


ImageNet-CNN Pool 2 units

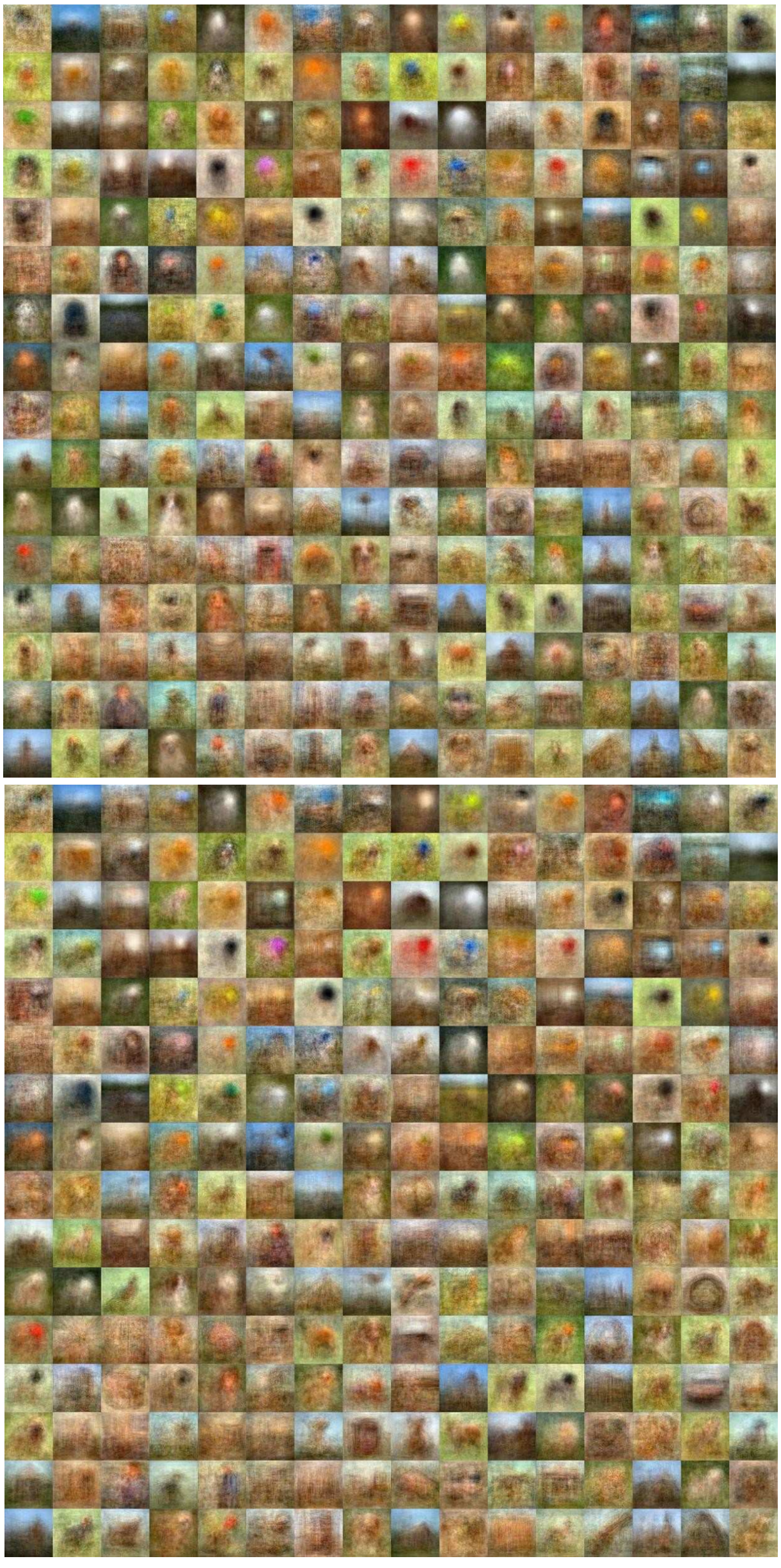


PLACES-CNN Pool 2 units

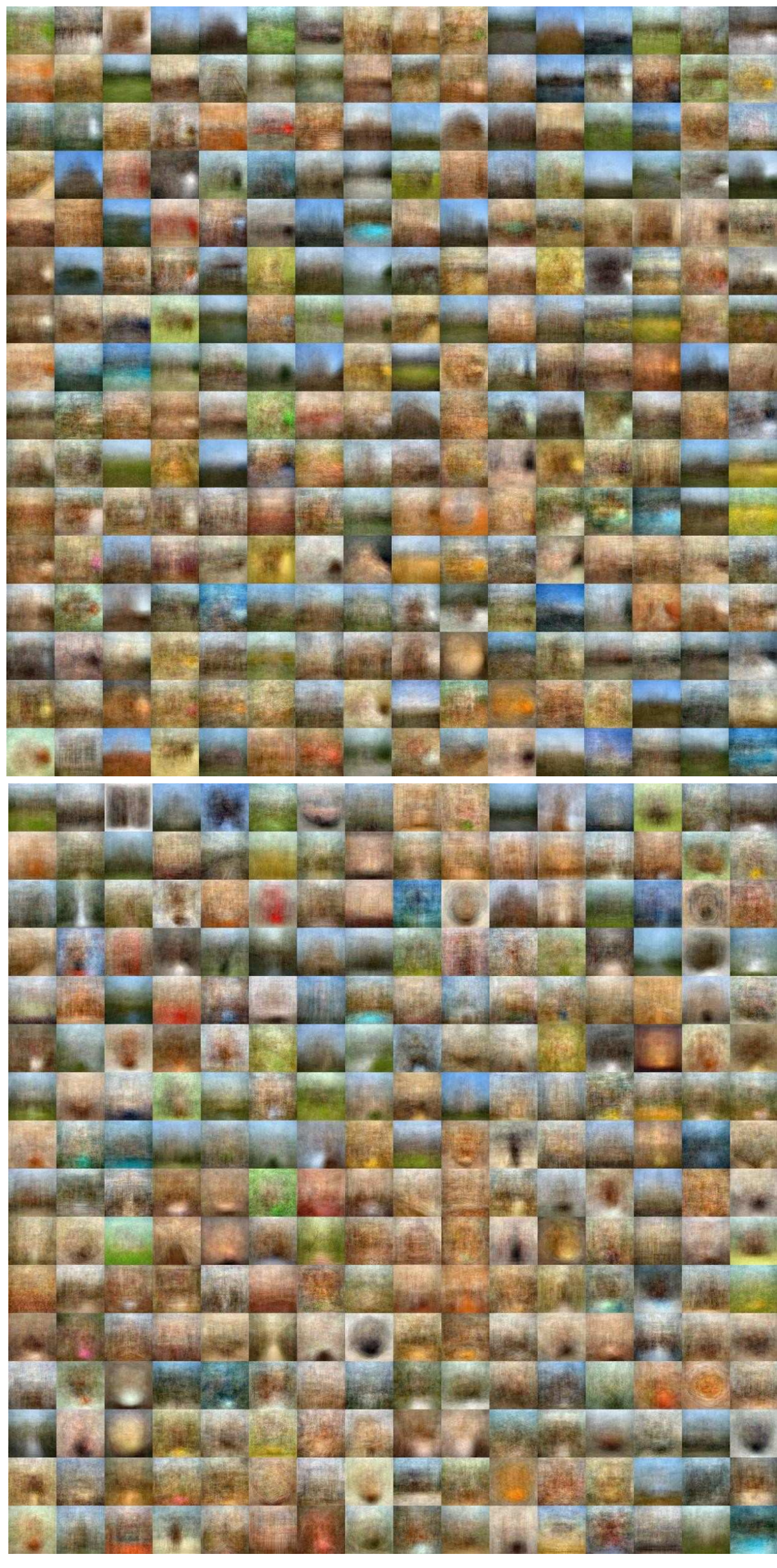




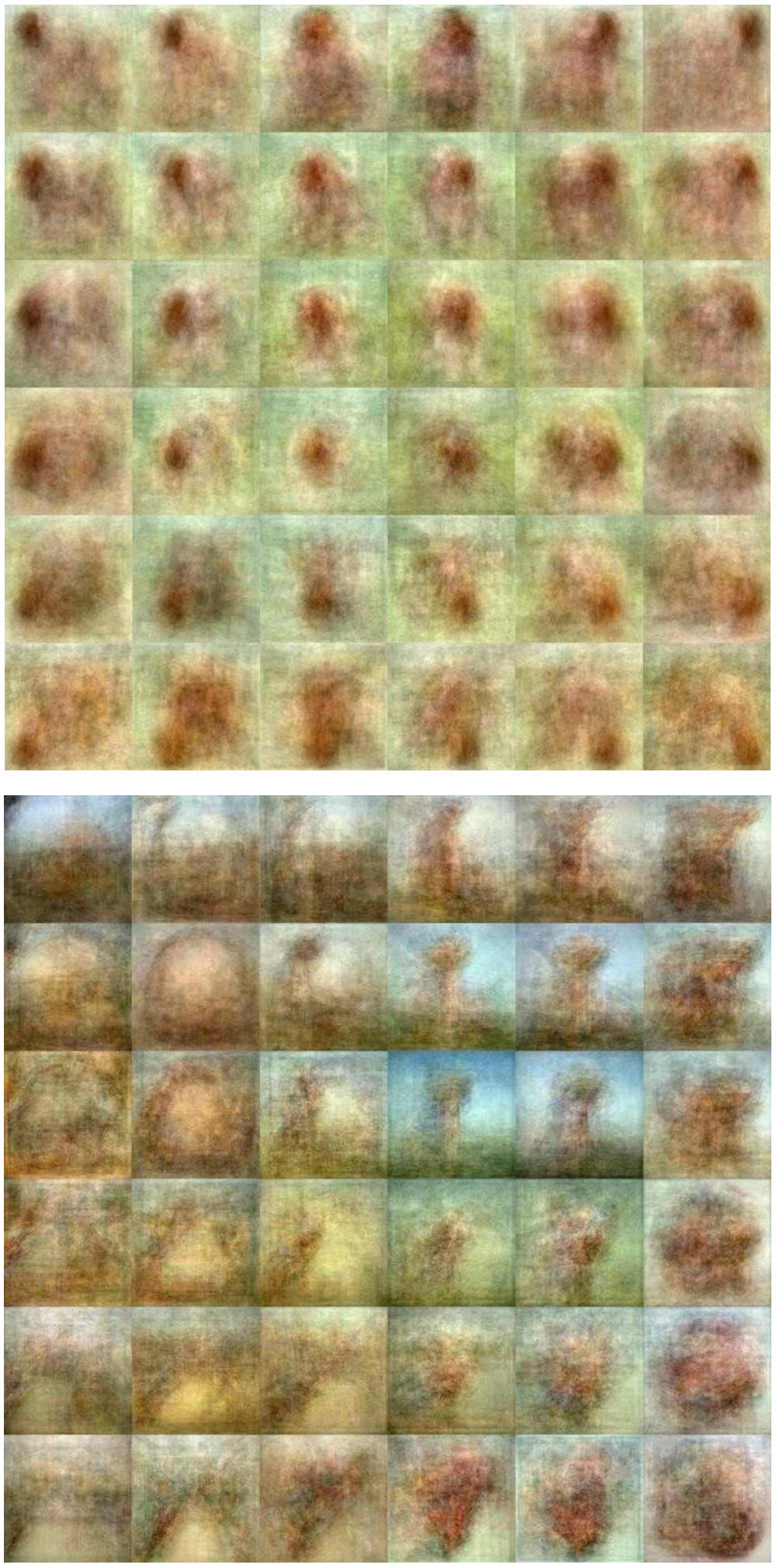
ImageNet-CNN Pool 5 units



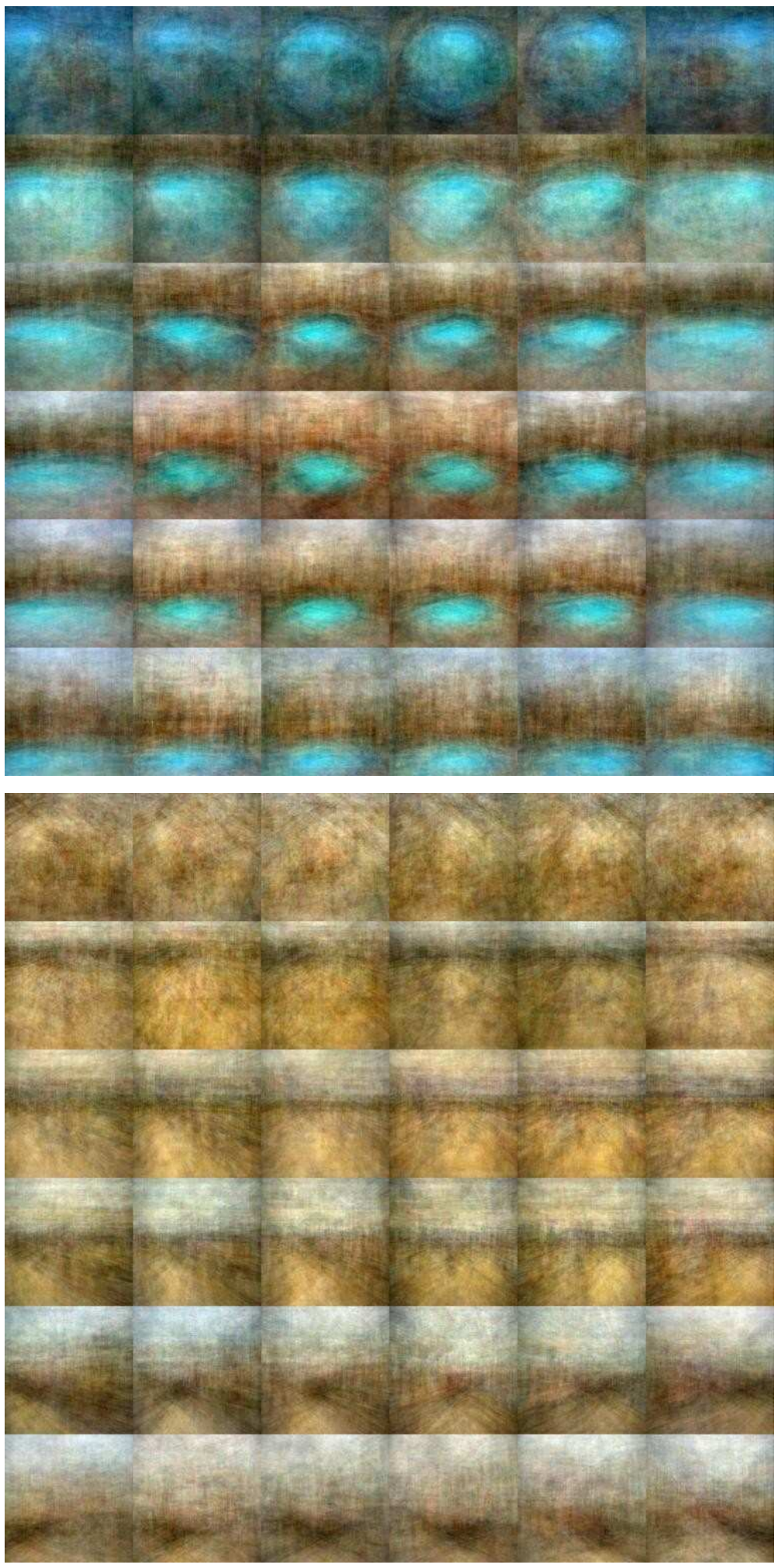
PLACES-CNN Pool 5 units



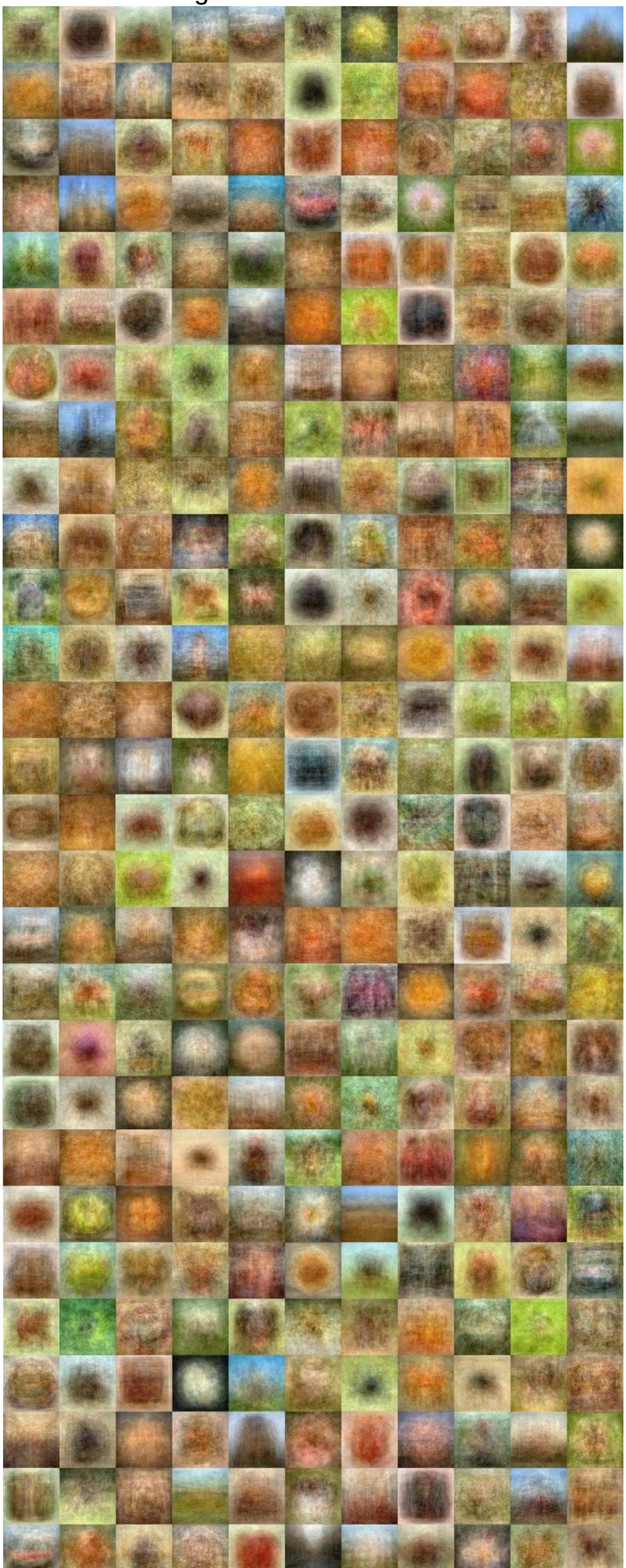
ImageNet-CNN Pool 5 units



PLACES-CNN Pool 5 units



ImageNet-CNN FC 7 units



PLACES-CNN FC 7 units