Density of Top-Layer Codes in Deep Convolutional Neural Networks Trained for Face Identification

Connor J. Parde[1], Y. Ivette Colón[1], Matthew Q. Hill[1], Rajeev Ranjan[2], Carlos Castillo[2], and Alice J. O'Toole[1]

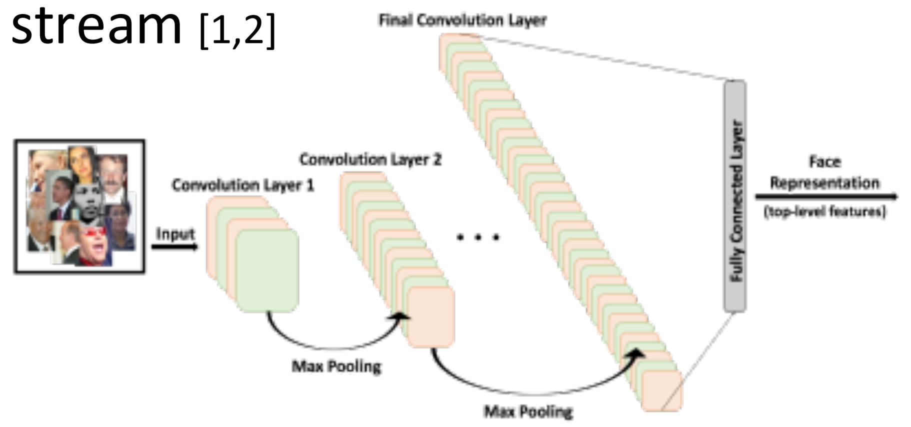
[1]The University of Texas at Dallas, [2]University of Maryland

THE UNIVERSITY OF TEXAS AT DALLAS

UNIVERSITY OF MARYLAND

Deep Convolutional Neural Networks (DCNNs)

- Robust across image conditions (view, illumination, etc.)
- Modeled after primate ventral visual stream [1,2]
  - Early layers -> V1
  - Intermediate layers -> V4
  - Final layers -> IT [3,4]



- Image information (e.g, viewpoint) retained despite identity training [5]
- Distribution of information in top-level code?
  - Identity, gender (subject variable) [6], viewpoint (image variable) [5]

Features in Primate Visual System

- Low-level visual features retinotopic & semantically interpretable (e.g., edge detectors) [7]
- Face identification -> high-level vision, categorical codes [8]
  - Link between visual receptive fields and features less clear
- Categorical representation of faces not well understood
  - Neurons tuned to features [9]
  - CNNs "directions in a space" interpretable/meaningful features [10]

Goal

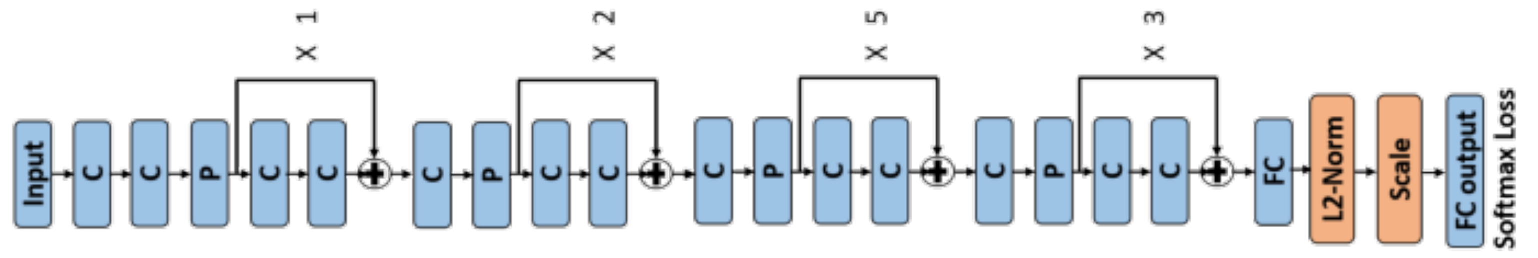
Probe distribution of identity, gender, viewpoint across individual units in DCNN top layer

Approach

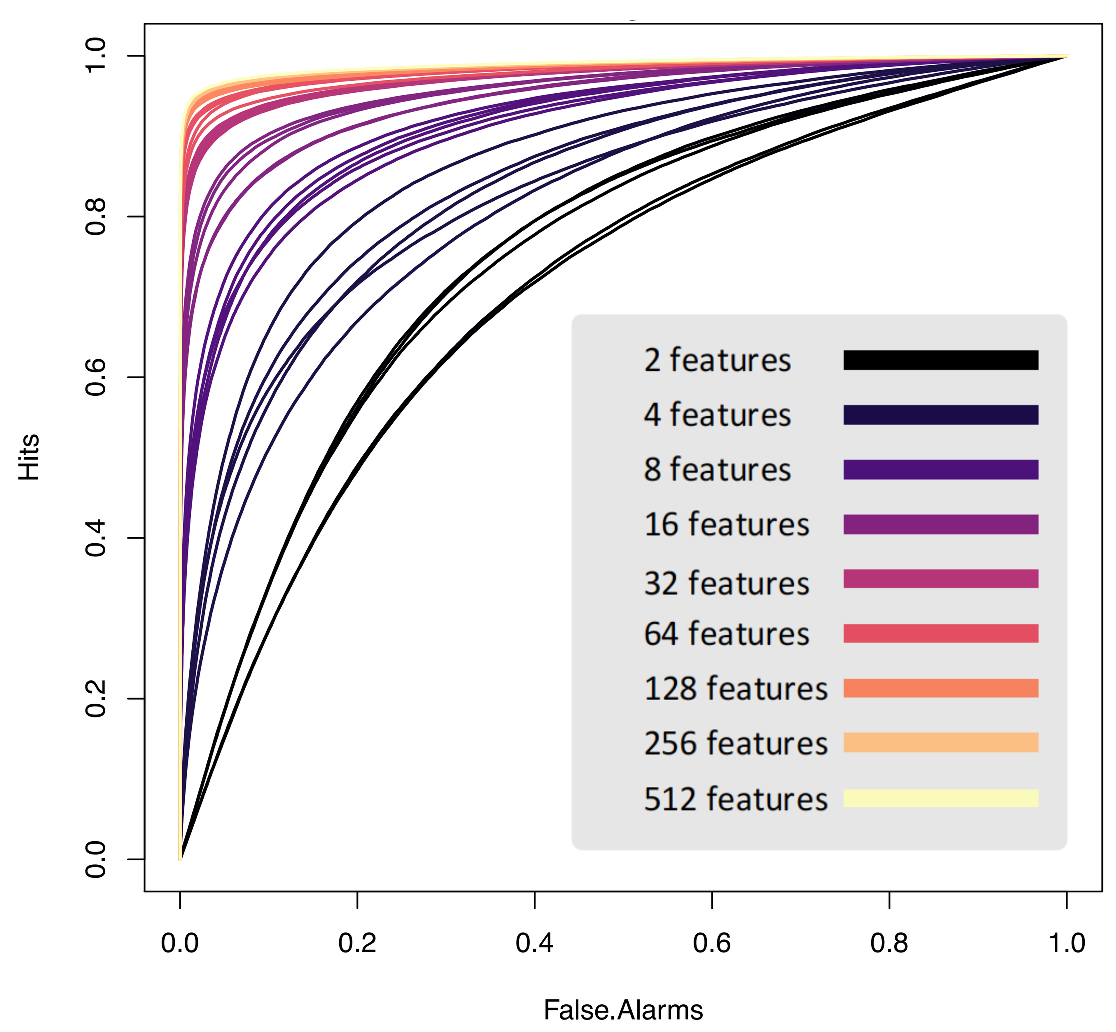
- Random feature deletion
- Re-evaluate identity, gender, viewpoint information in subspaces of varying dimensionalities

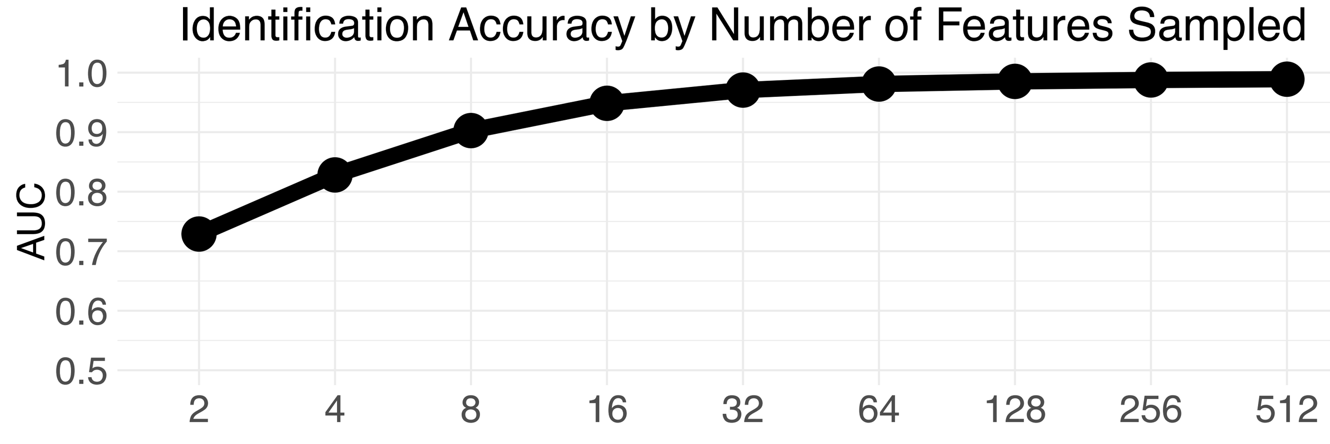
Network

- High-performing face identification network [11]
  - 101 –layered ResNet [12], trained on nearly 6 million images of 58,000 identities
  - 512-dimensional final output representation



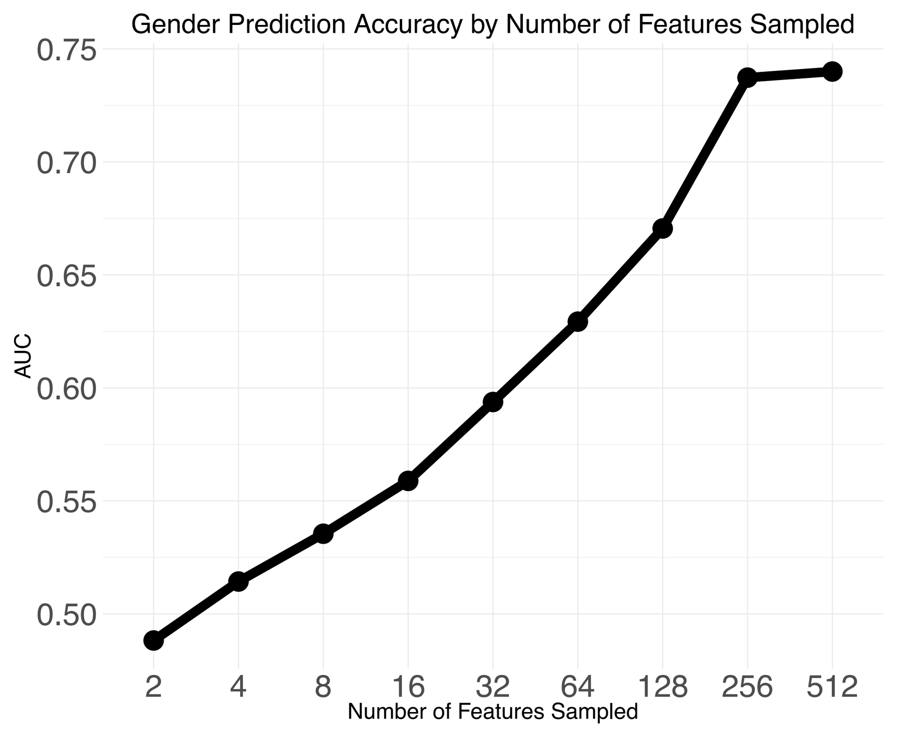
Subspace: Identification





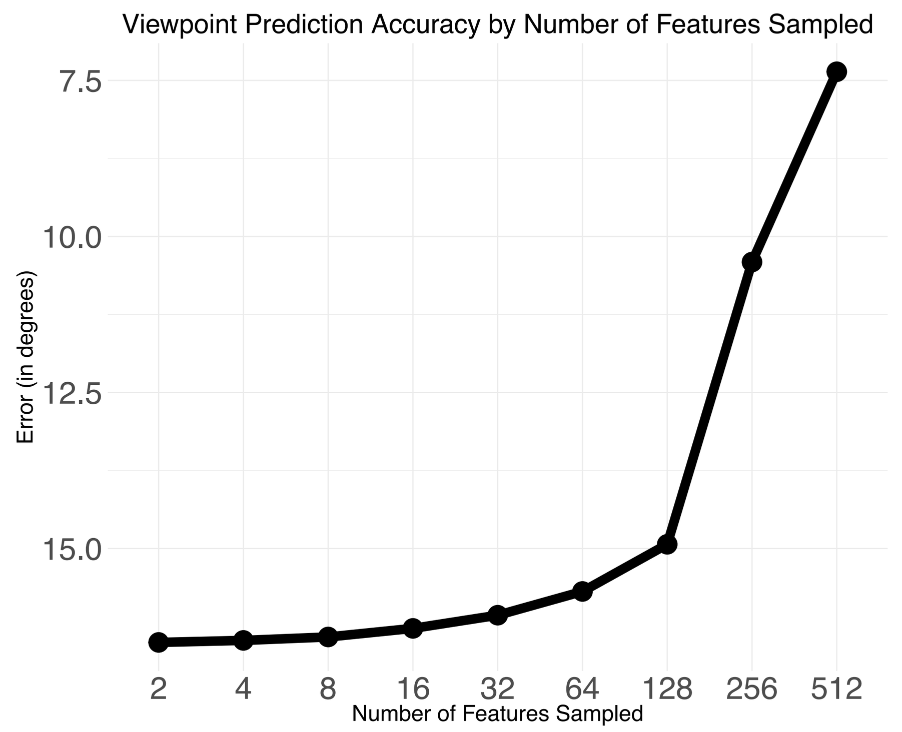
- Identification accuracy maintained when sampling very few randomly chosen features

Subspace: Gender



- Gender prediction accuracy decreases gradually with size of sampled features

Subspace: Viewpoint



- Viewpoint prediction accuracy declines sharply when fewer features sampled

Methods

- randomly sampled sets of features from output representation
- computed ROC curves of each sample on face-identification, gender classification, and viewpoint prediction tasks using the "in the wild" IJB-C dataset [13]
  - > 141,332 images, 3,531 identities

Conclusions

- Top-level feature units spread identity information efficiently
  - All units contain identity information, random combinations of units lead to good performance
- Coding of gender and viewpoint make sense in context of [6]
  - Hierarchy of clustering according to identity/image variables
- Robust code for face identity
- Many sources of identity information, many solutions

References

1. Fukushima, K. (1988). Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural networks*, *1*(2), 119-130.
2. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105
3. Hong, H., Yamins, D. L., Majaj, N. J., & DiCarlo, J. J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature neuroscience*, *19*(4), 613.
4. Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, *19*(3), 356.
5. Parde, C. J., Castillo, C., Hill, M. Q., Colon, Y. I., Sankaranarayanan, S., Chen, J. C., & O'Toole, A. J. (2017, May). Face and image representation in deep CNN features. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* (pp. 673-680). IEEE.
6. Hill, M. Q., Parde, C. J., Castillo, C. D., Colon, Y. I., Ranjan, R., Chen, J. C., ... & O'Toole, A. J. (2018). Deep Convolutional Neural Networks in the Face of Caricature: Identity and Image Revealed. *arXiv preprint arXiv:1812.10902*.
7. Gross, C. G., Bender, D. B., & Rocha-Miranda, C. E. D. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science*, *166*(3910), 1303-1306.
8. Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual review of neuroscience*, *19*(1), 109-139.).
9. Freiwald, W. A., Tsao, D. Y., & Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *Nature neuroscience*, *12*(9), 1187.
10. Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2013). Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*.
11. Ranjan, R., Sankaranarayanan, S., Castillo, C. D., & Chellappa, R. (2017, May). An all-in-one convolutional neural network for face analysis. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* (pp. 17-24). IEEE.
12. Y. Wen, K. Zhang, Z. Li, and Y. Qiao. A discriminative feature learning approach for deep face recognition. In European Conference on Computer Vision, pages 499–515. Springer, 2016
13. Maze, B., Adams, J., Duncan, J. A., Kalka, N., Miller, T., Otto, C., ... & Grother, P. (2018, February). IARPA janus benchmark-c: Face dataset and protocol. In *2018 International Conference on Biometrics (ICB)* (pp. 158-165). IEEE.

Acknowledgements

This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2014-14071600012. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.