

How do we Make Artificial Intelligence More Human-Like?

Luke D. Baumel¹, Mikayla Cutler¹, Dr. Nicholas Baker², Dr. George K. Thiruvathukal³, Dr. Silvio Rizzi⁴

¹Undergraduate Cognitive/Behavioral Neuroscience Major, Loyola University Chicago, Chicago, IL ²Department of Psychology, Loyola University Chicago, Chicago, IL

³Department of Computer Science, Loyola University Chicago, Chicago, IL ⁴Argonne Leadership Computing Facility, Argonne National Laboratory, Lemont, IL

Key Question: How do we build artificial intelligence (AI) systems that are sensitive to the holistic properties of an object's shape?

Where AI Has Failed Before

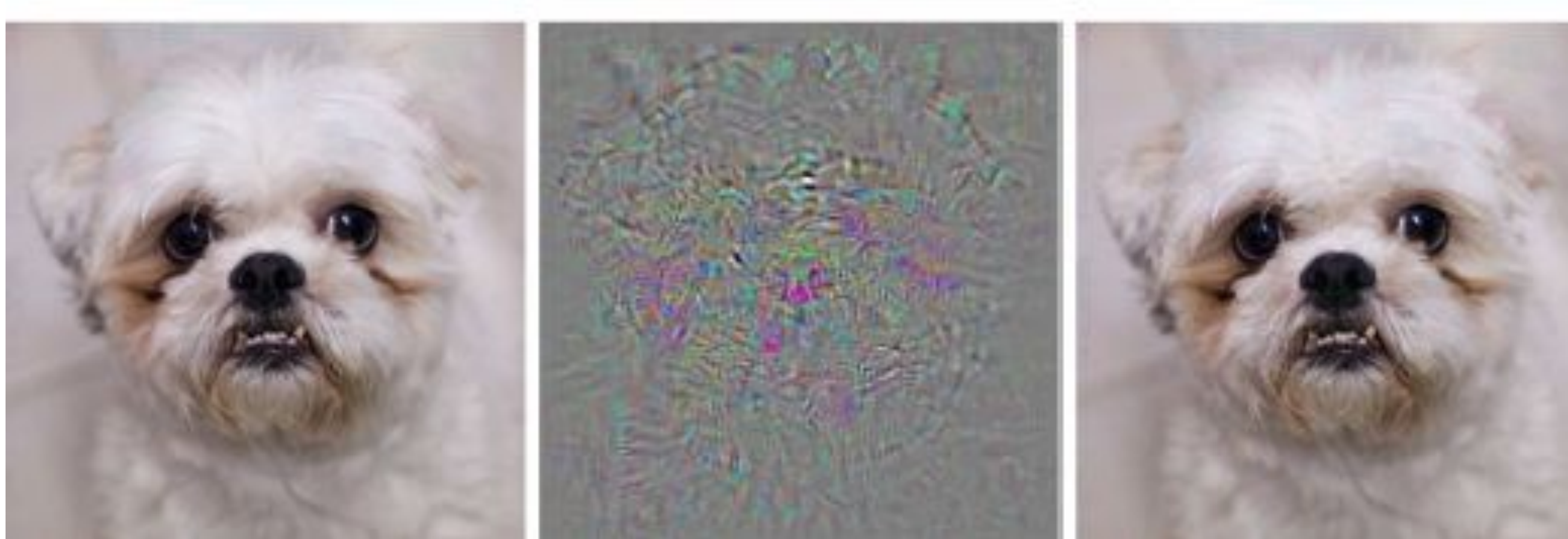


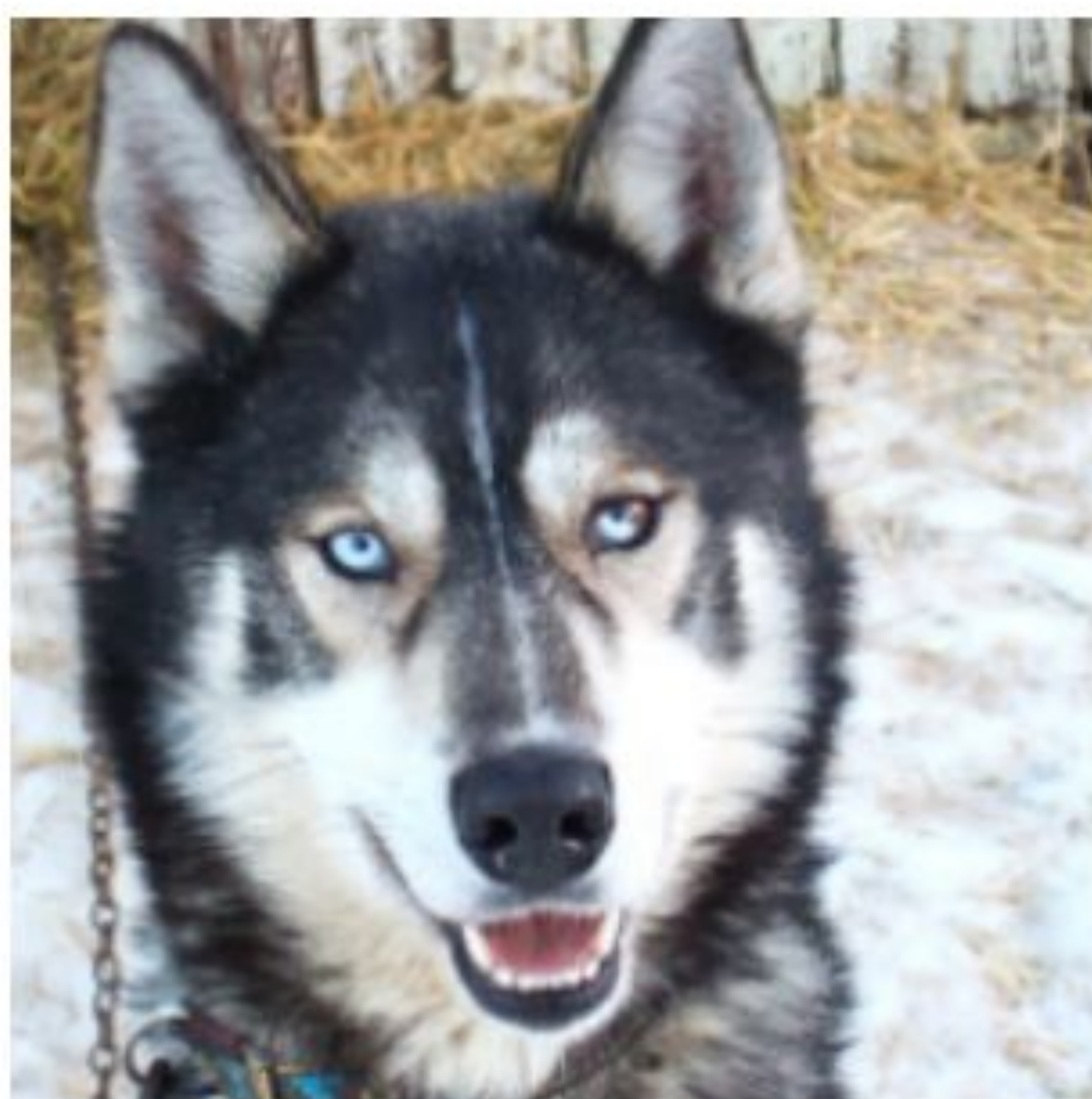
Image incorrectly classified when only a few pixels were changed (Szegedy 2014)



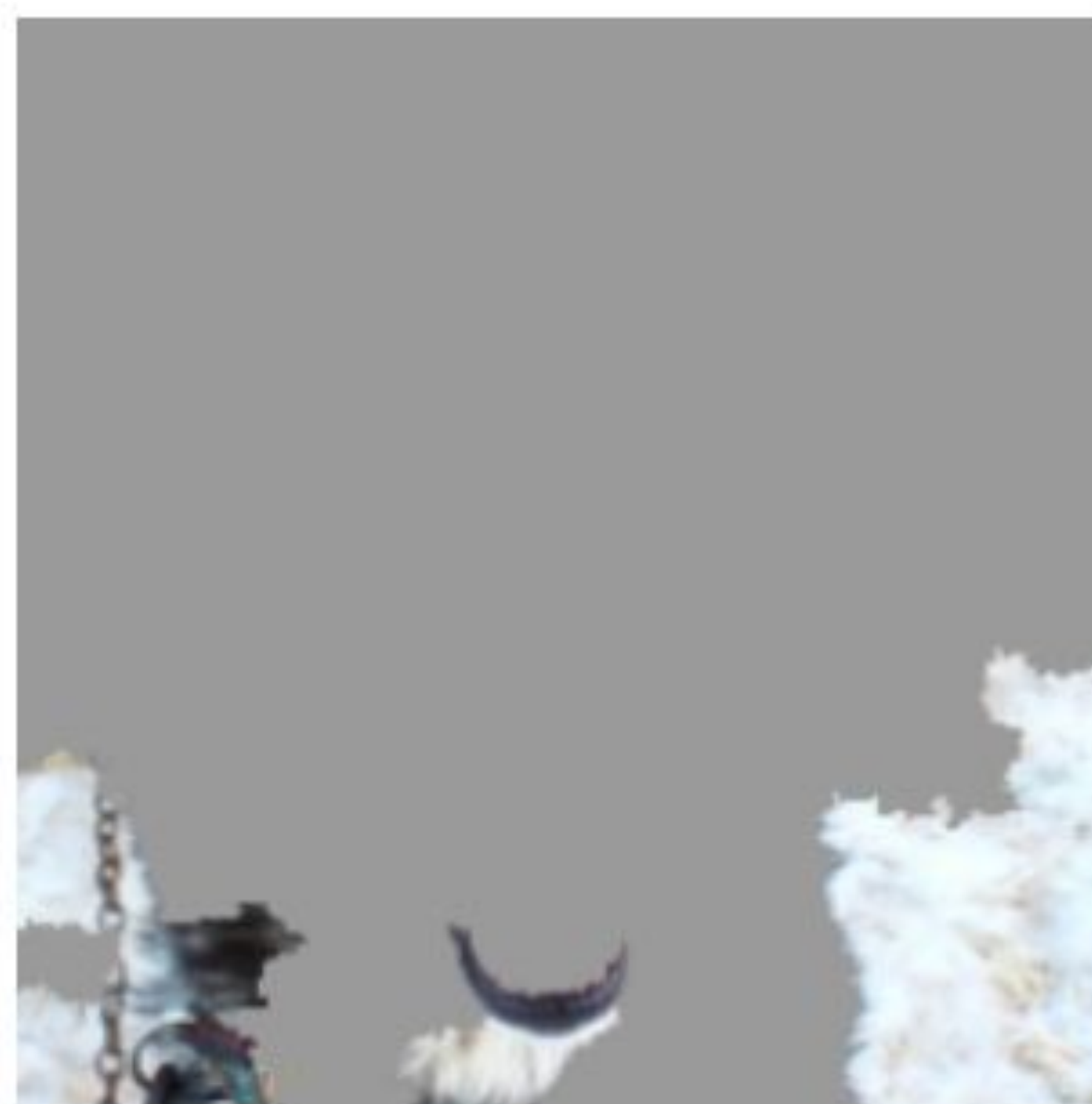
Goose
(0.006%)

Web site
(10.13%)

Glass goose incorrectly categorized as a website (Baker 2018)



(a) Husky classified as wolf



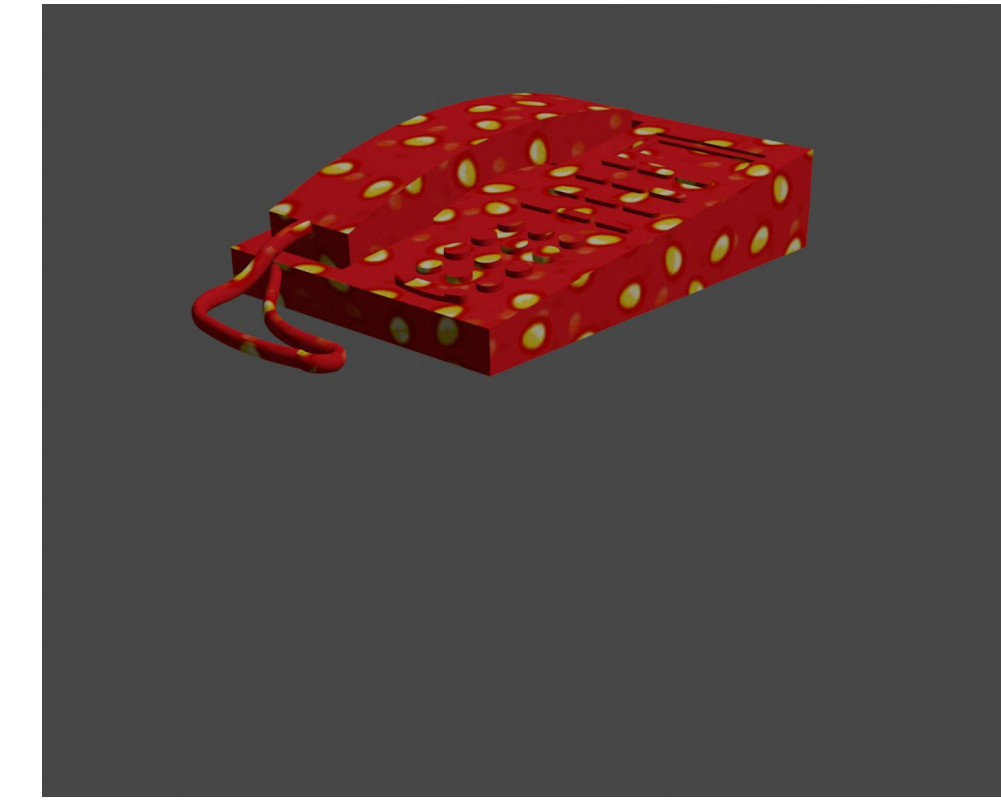
(b) Explanation

Husky misclassified as a wolf due to background (Tulio Ribeiro et al. 2016)

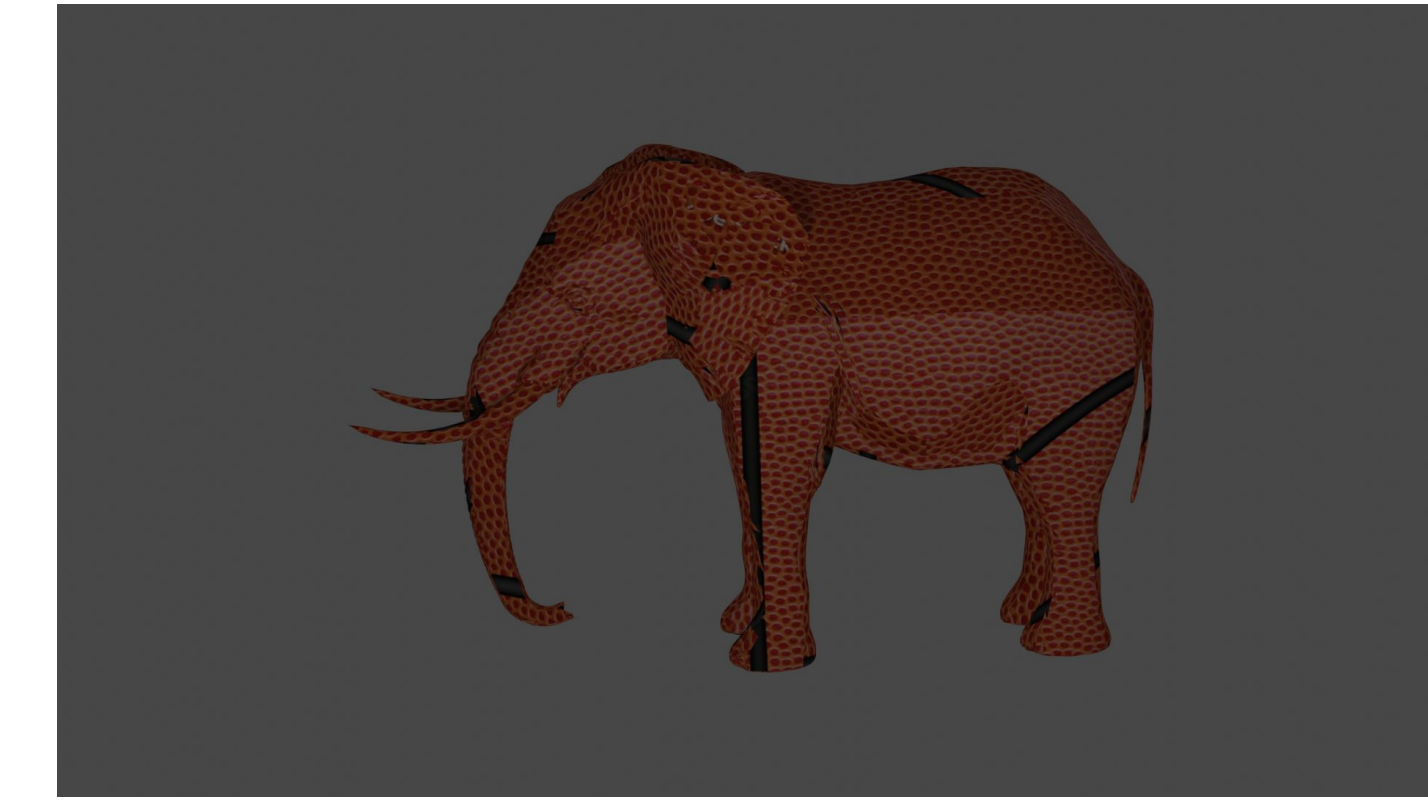
Our Methodologies

- Stimuli creation via scripting in Blender. Capture 12 viewpoints of 3-dimensional stimulus objects with mapped uncharacteristic texture
- Test Deep Convolutional Neural Networks (DCNNs) with stimuli and compare results to stimuli with original or no texture
- Networks will be trained through fine-tuning

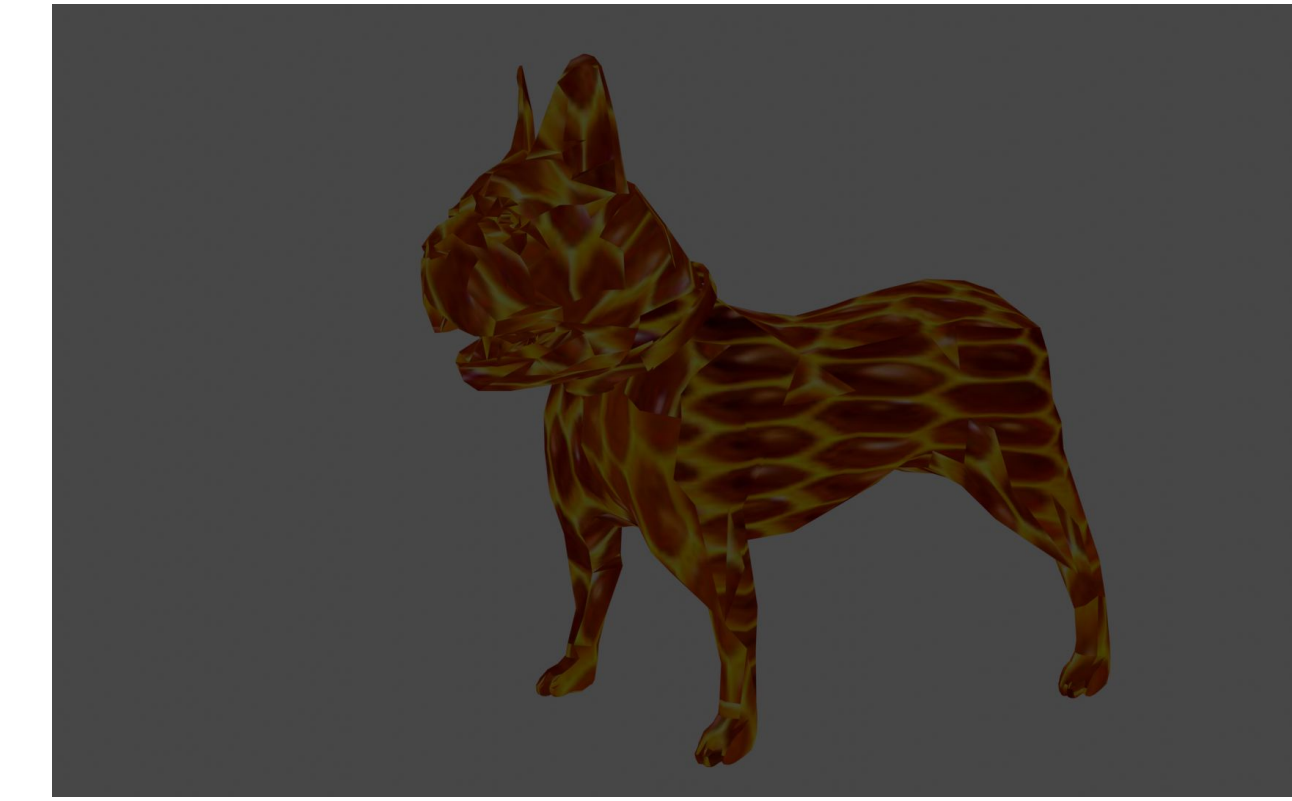
Sample Stimuli and Data



Strawberry wrapped phone
Strawberry: 20.7% confidence
Phone: 0.22% confidence



Basketball wrapped elephant
Basketball: 0.05% confidence
Elephant: 0.02% confidence



Honeycomb wrapped dog
Honeycomb: 0.06% confidence
Dog: 0.01% confidence

Preliminary Conclusions

In comparison to Baker 2018:

- No increased preference for shape in 2-dimensional pictures of 3-dimensional models in comparison to 2-dimensional pictures of objects when both have uncharacteristic texture
- Categorization performance is still poor compared to pictures of objects with their natural texture and background
- This performance suggests preference for texture, including background, over shape in image classification

Next Steps

- Fully automate the stimuli creation process via script
- Compile large dataset of 12 different 2-dimensional angle views of 3-dimensional objects with non-diagnostic texture for the use of other researchers in the field
- Continue testing to verify that DCNNs lack use of texture in image classification
- Begin training networks with our novel stimuli

Impacts

- Improve technology used by physicians in reading scans
- Improve methods used by autonomous vehicles
- Improve technology for image classification-related fields

References and Acknowledgements

Baker, N., Lu, H., Erlikhman, G., & Kellman, P. J. (2018). Deep convolutional networks do not classify based on global object shape. *PLOS Computational Biology*, 14(12). <https://doi.org/10.1371/journal.pcbi.1006613>

Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I, Fergus R. Intriguing properties of neural networks. arXiv preprint arXiv:1312.6199. 2013 Dec 21.

Tulio Ribeiro, M., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*. <https://doi.org/10.18653/v1/n16-3020>

Thank you to our mentors Dr. Nicholas Baker, Dr. George K. Thiruvathukal, and Dr. Silvio Rizzi for their support of our research and to the Carbon Family for funding our research through the Carbon Fellowship.

