Deep Convolutional Neural Networks Rely on Distinct Semantic Features of Same-Category Objects for Recognition

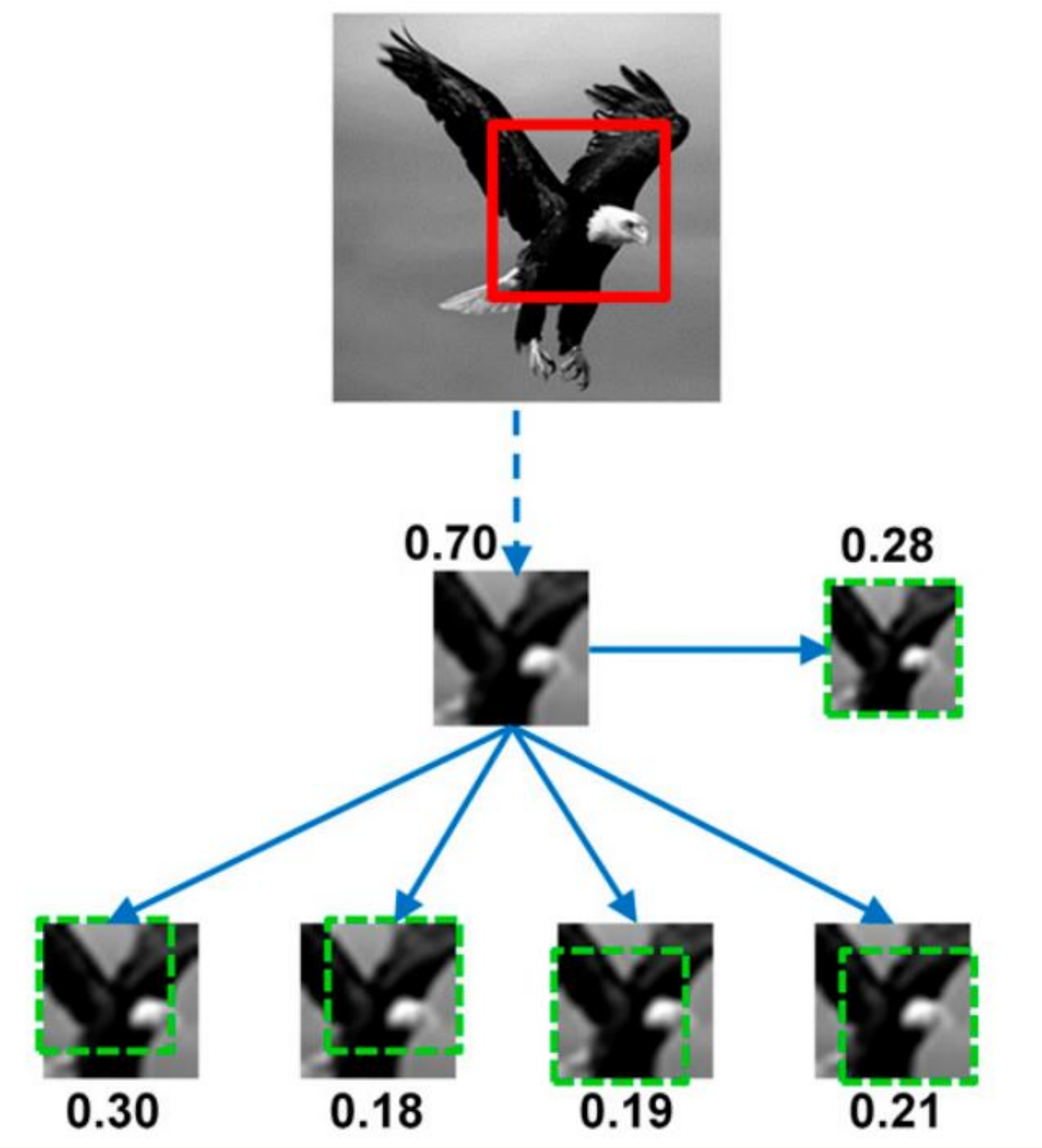
MohammadHossein NikiMaleki(1), Hamid Karimi-Rouzbahani(2,3)

- Faculty of Computer Science and Engineering, Shahid Beheshti University, Iran
- Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, UK
- Department of Computing, Macquarie University, Australia

Understanding Vision - summer 2021

Object Recognition in Human Brain

- One of key ability of brain is Object Recognition; rapid, accurate, with variation.
- Observed that humans relied on specific (diagnostic) object regions (called MIRCs) for accurate recognition (Ullman et al., 2016).
- They remain relatively consistent (invariant) across variations.
- Feed-forward feature-extraction models use selected view-specific (non-invariant) features across variations. This suggests that models can develop different strategies (Karimi-Rouzbahani et al., 2017).

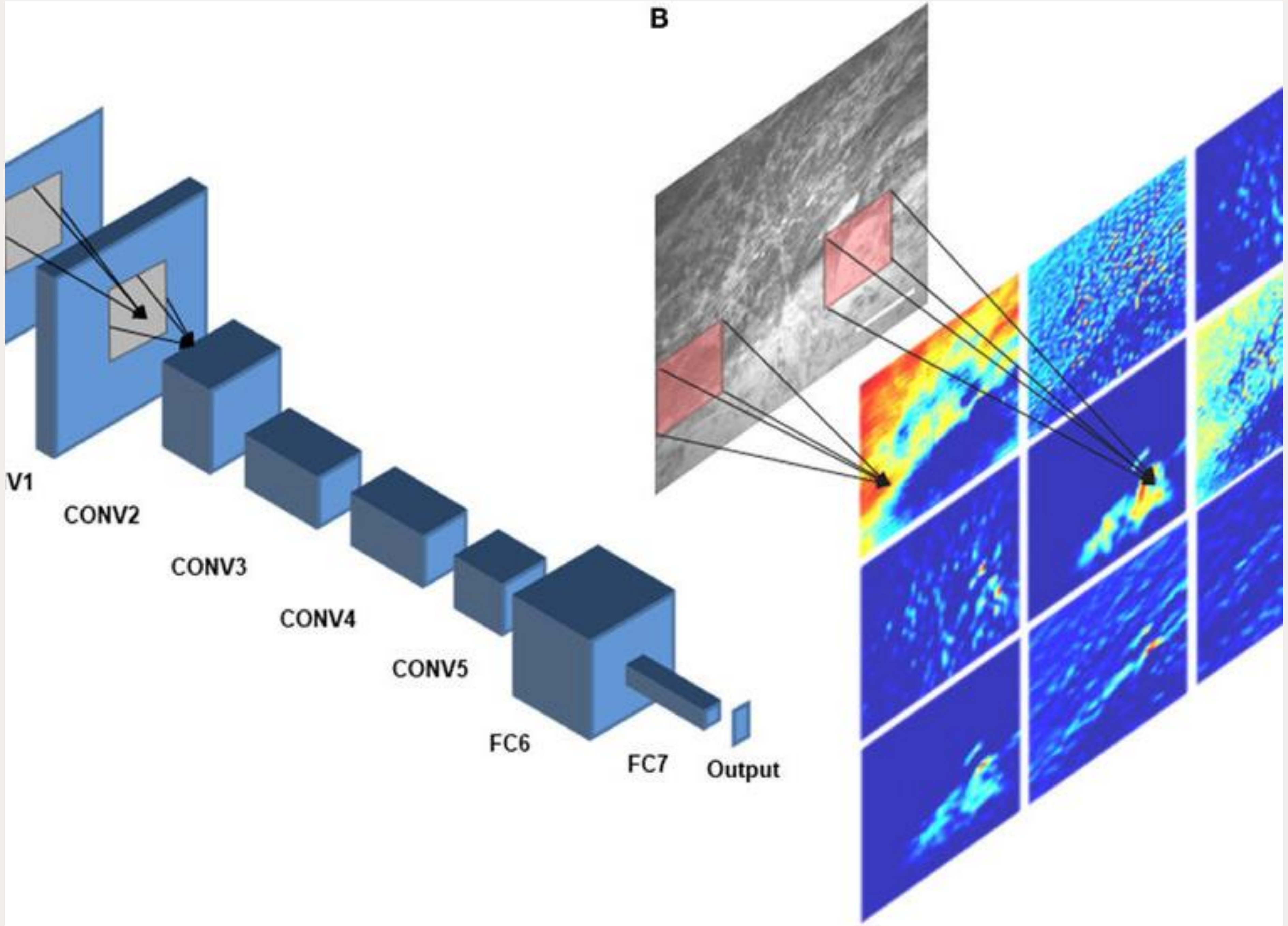


MIRC

Humans rely significantly on specific sets of object parts (i.e. visual features or simply features), referred to as Minimal Recognizable Configurations (MIRCs). In other words, some specific object parts were considered more informative (Ullman et al., 2016).

Deep Convolutional Neural Networks (DCNN)

- outperformed computer vision algorithms in many applications especially in "Object Recognition".



Main idea

DCNNs did not show such sensitivity to identical MIRCs;

It remains unclear if humans and DCNNs use similar strategies for object recognition.

Two Critical Questions

01

Do DCNNs rely on semantically similar MIRCs from different exemplars of the same object category?

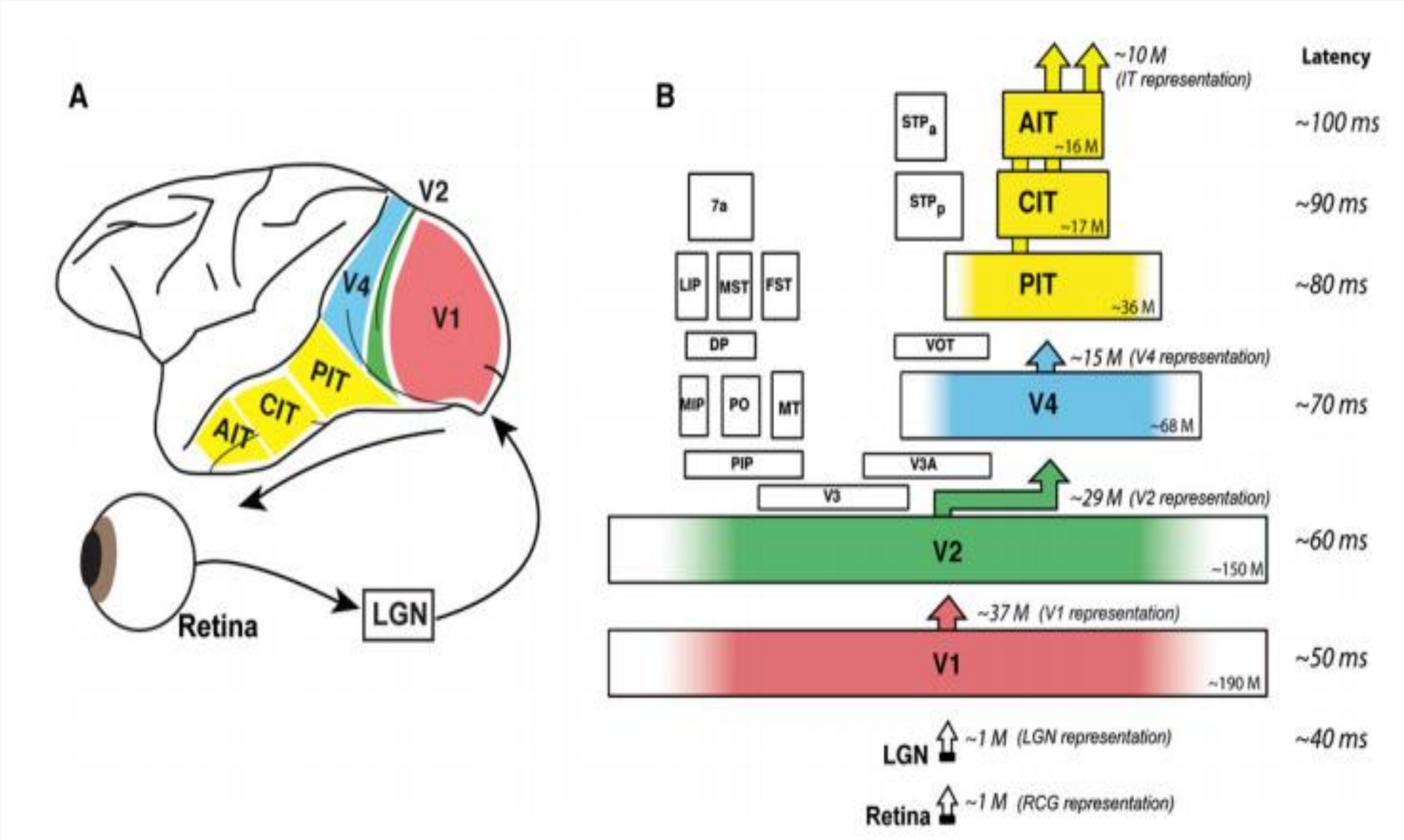
02

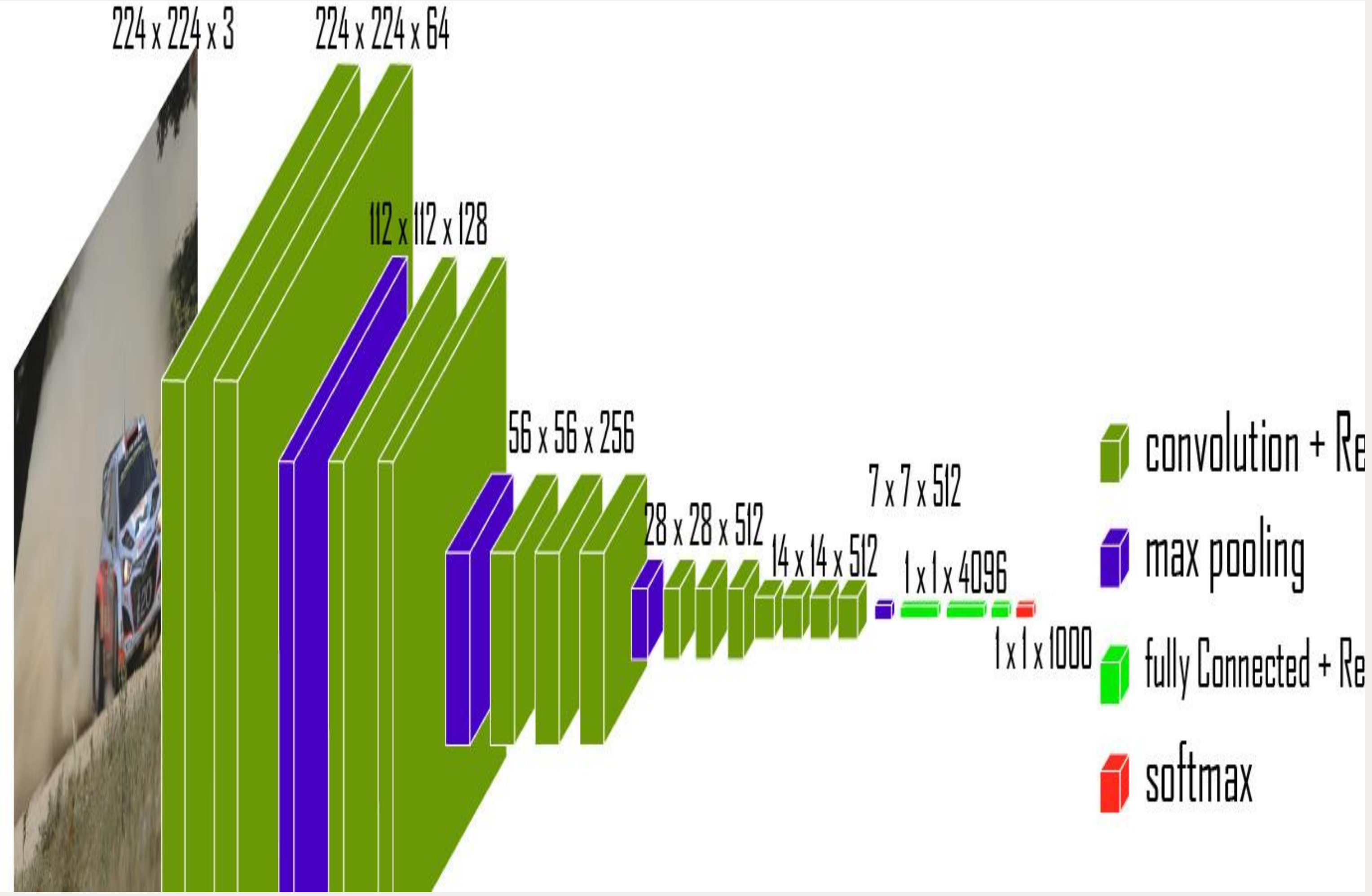
Do human use the same MIRCs as DCNN's model ?

Researching Q1

We obtained MIRCs from VGG16 (Simonyan et al., 2015)

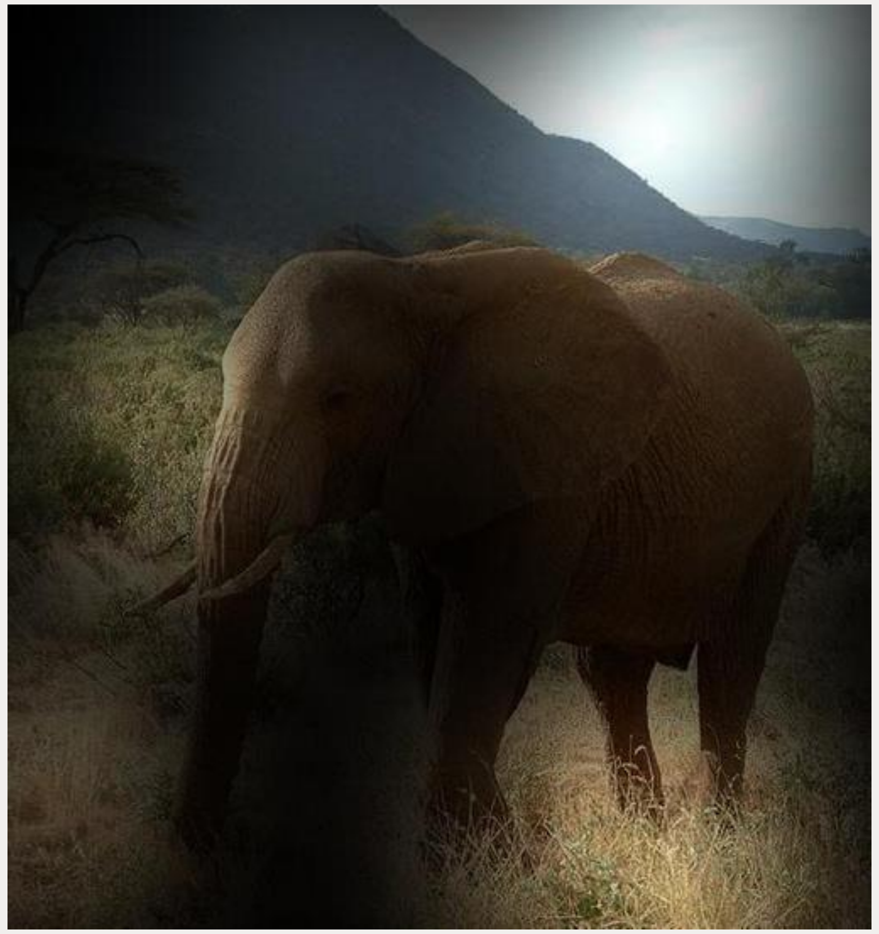
- **One of the most brain-like DCNNs** (Schrimpf et al., 2018) and (Dicarlo et al., 2012).





Obtaining MIRCs

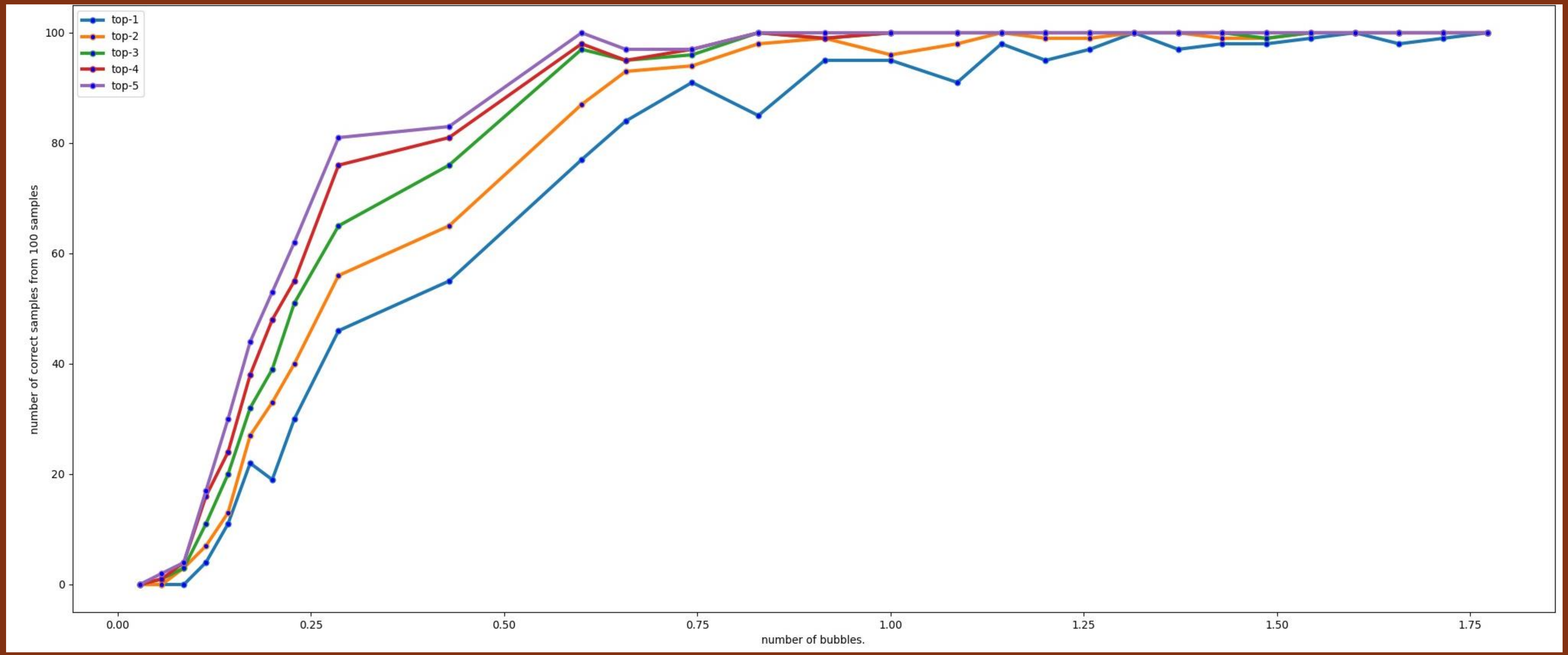
- **We used the well-established Bubbles method** (Gosselin et al., 2001).
- **As an advantage to previous precedures, which detected MIRCs from preselected discrete** image parts (Karimi-Rouzbahani et al., 2017), **Bubbles sweeps the whole image using continuous masks.**
- **Allowing data-driven evaluation of all pixels; so we can control the bias in background.**




Lervia

NOTE

We considered wide range of bubbles and numbers, and calculated the psychometric function of the model so we can control the accuracy of network to <u>50%</u> in top-1.



Imagenet Large Scale Visual Recognition Challenge

01

We choose images from test set of naturalistic ImageNet dataset as a widest and largest dataset for object recognition [6].

02

Test set classified using VGG16 DCNN.

03

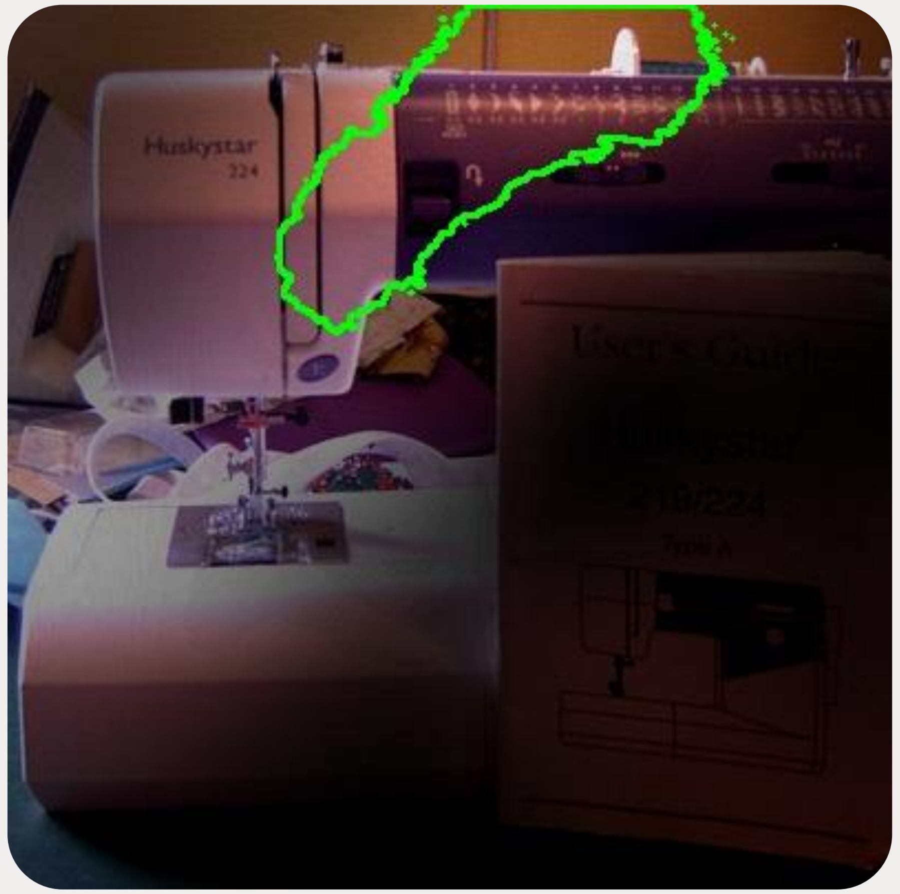
MIRCs given from 12 different categories each with 16 examplars.

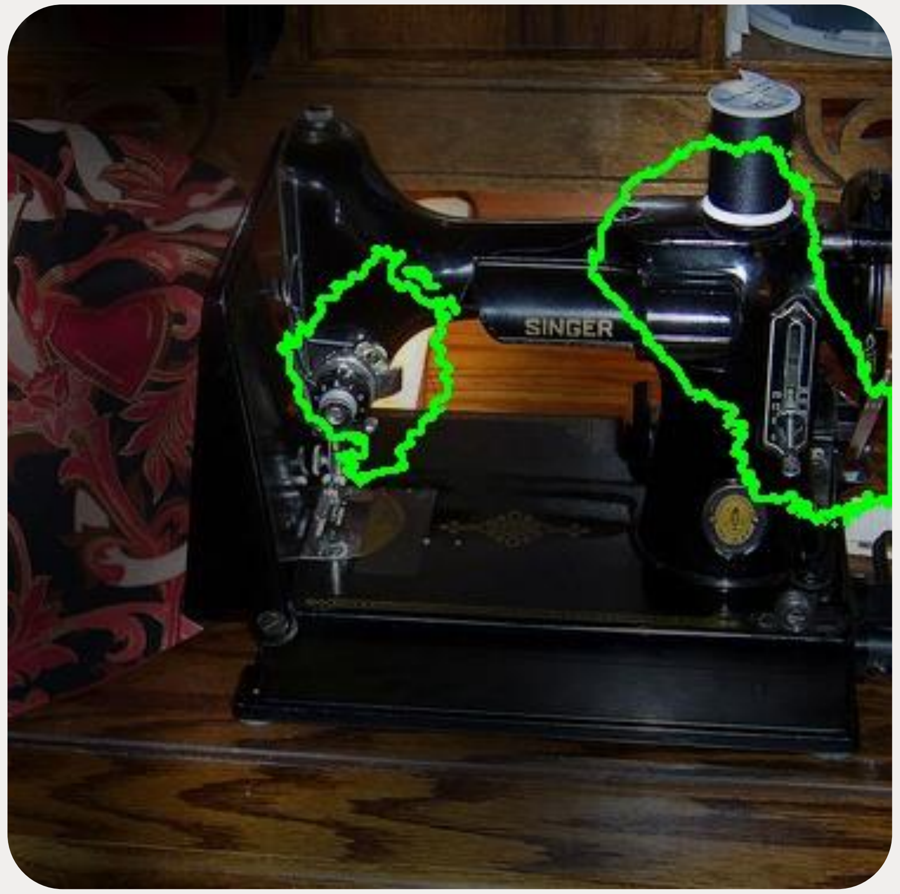
Choose 12 different semantic unrelevant categories like **fruits**, **animals**, **furnitures**, **vehicles**, **plants**, **tools**, etc.

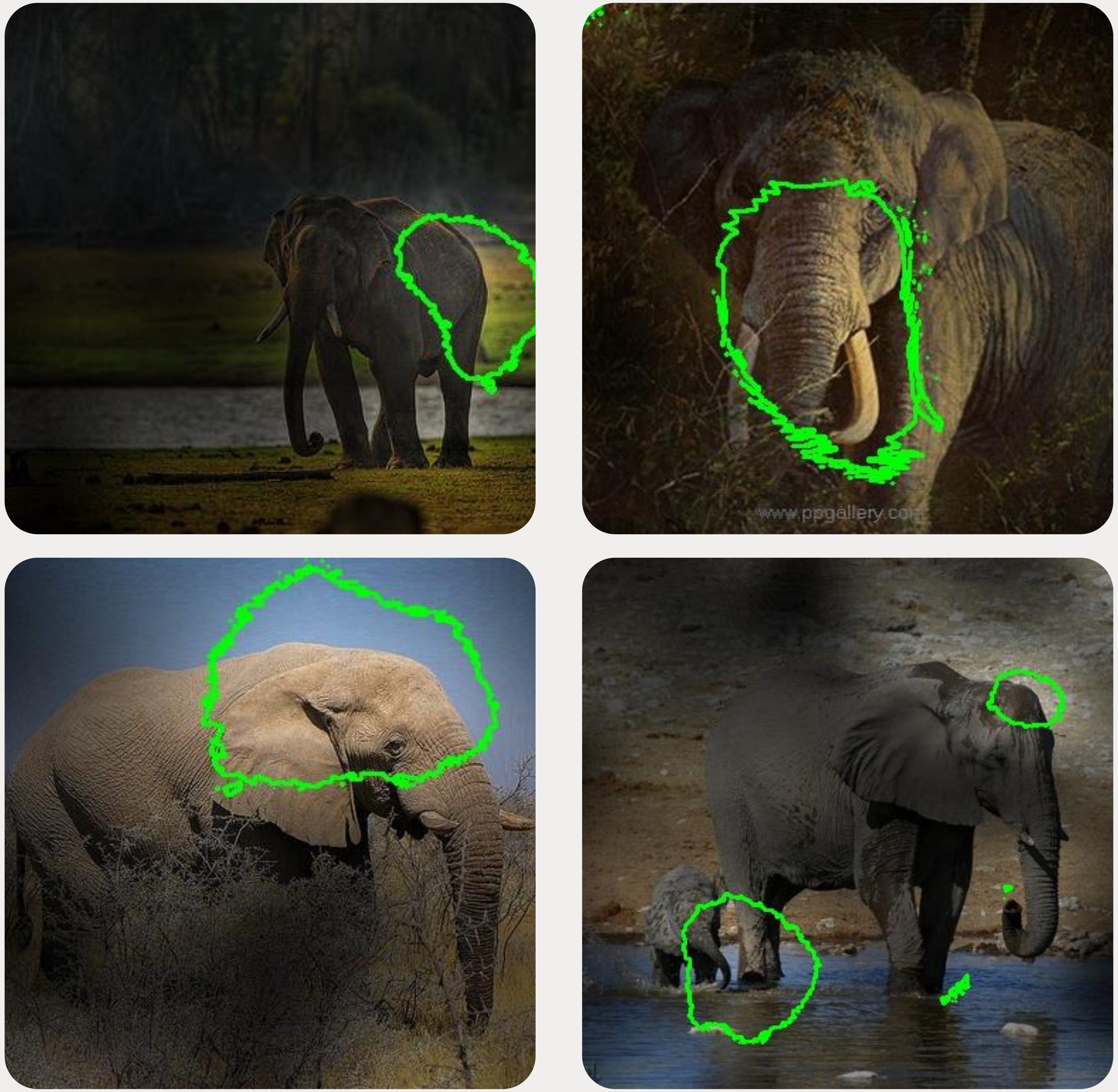
- **Car**
- **Goldfish**
- **Hammer**
- **Violin**
- **Elephant**
- **Pot**
- **Sewing Machine**
- **Ladybug**
- **Pineapple**
- **Hat**
- **Iron**
- **Hand Blower**

Results

- **Showed green regions which led to correct recognition of the object by DCNN.**
- **Results showed clearly different MIRCs for distinct examplars from the same object category.**
- **Reflecting the highly variable nature of feature extraction in DCNNs.**
- **Potentially facilitating recognition under object variation.**


Huskystar
224


SINGER



Researching Q2

Do human use the same MIRCs as DCNN's model ?

We are collecting human data to quantitatively compare to our DCNN results.

Thank You !

MohammadHossein NikiMaleki, Hamid Karimi-Rouzbahani

mh.nikimaleki@gmail.com
hkarimi265@gmail.com

Understanding Vision - summer 2021

Refrences

[1] Ullman, S., Assif, L., Fetaya, E. and Harari, D., 2016. Atoms of recognition in human and computer vision. Proceedings of the National Academy of Sciences, 113(10), pp.2744-2749.

[2] H. Karimi-Rouzbahani, N. Bagheri, and R. Ebrahimpour, "Invariant object recognition is a personalized selection of invariant features in humans, not simply explained by hierarchical feed-forward vision models," Scientific Reports, vol. 7, no. 1, 2017.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–14, 2015.

[4] M. Schrimpf, J. Kubilius, H. Hong, N. J. Majaj, R. Rajalingham, E. B. Issa, K. Kar, P. Bashivan, J. Prescott-Roy, F. Geiger, K. Schmidt, D. L. K. Yamins, and J. J. Dicarlo, "Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like?," 2018.

[5] Gosselin, F. and Schyns, P.G., 2001. Bubbles: a technique to reveal the use of information in recognition tasks. Vision research, 41(17), pp.2261-2271.

[6] eng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., 2009, June. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.

[7] DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How does the brain solve visual object recognition? Neuron 73, 415–434 (2012).