Convolutional Neural Networks Do Not Rely On Object Features Which Drive Human Overt Attention

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

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

International Interdisciplinary Computational Cognitive Science Summer School

(ECVP)

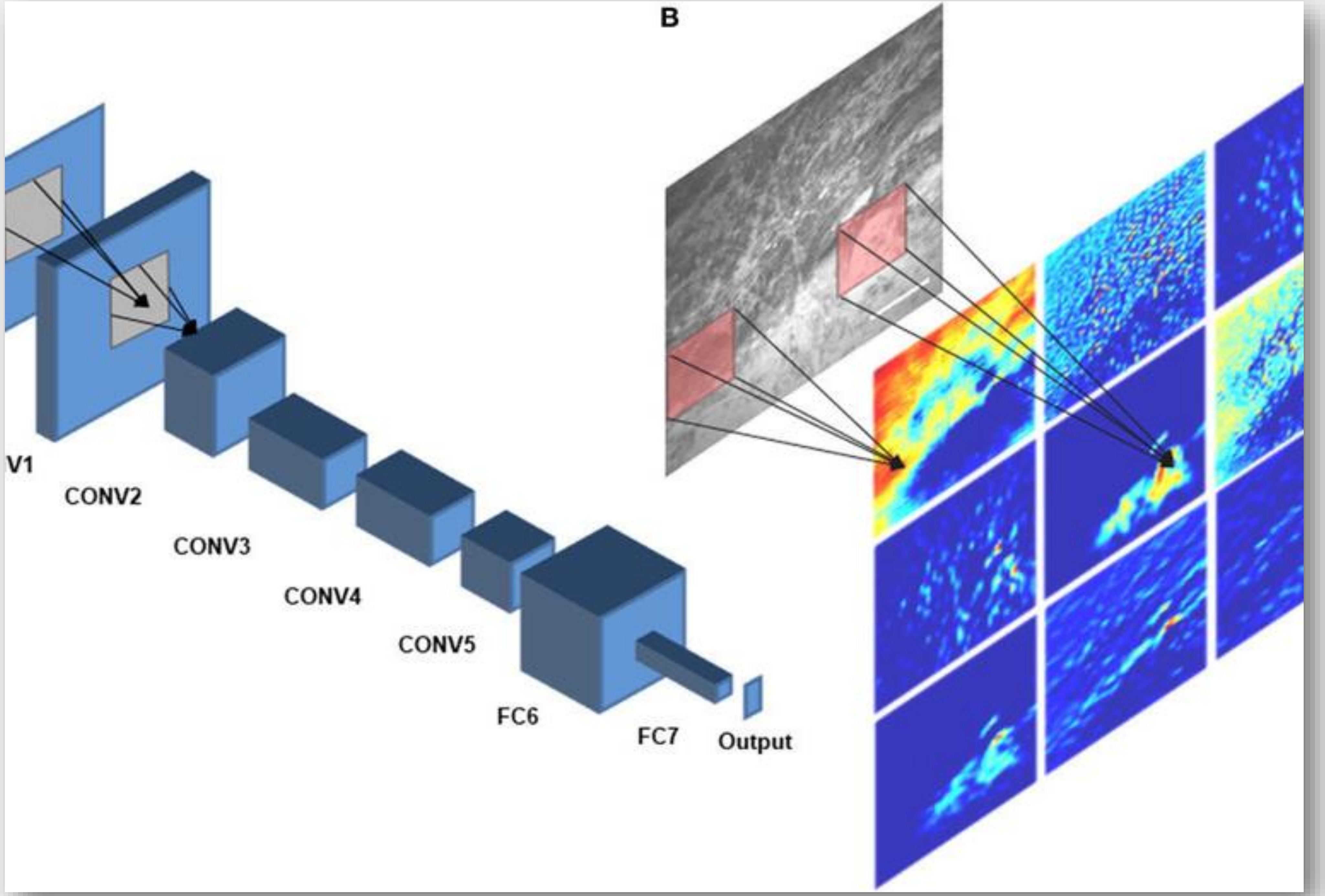
August 2021

Object Recognition in Human Brain

- One of key ability of the brain is Object Recognition; rapid, accurate, robust to variation in object appearance, etc.

- Our main goal is to understand the neural mechanisms of object recognition in humans.

- Deep Convolutional Neural Networks (DCNNs), not only outperformed computer vision algorithms in many applications, but are among the ***best models*** of human object recognition.



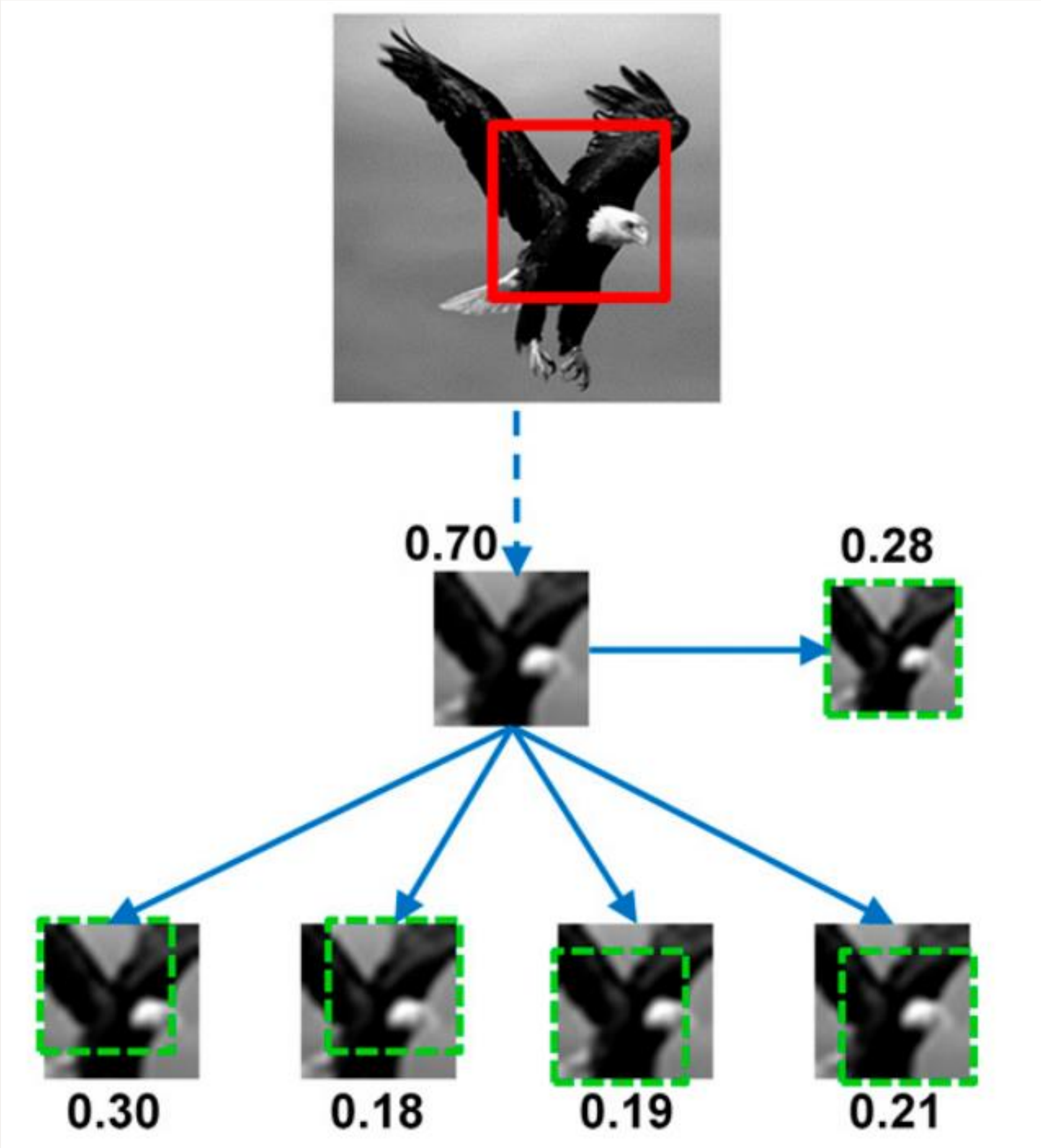
Object Recognition in Human Brain

- Observed that humans rely on specific (diagnostic) object regions (called MIRCs: Minimal Recognizable Configurations) for accurate recognition (Ullman et al., 2016). They remain relatively consistent (invariant) across variations.

- But DCNN models use selected view-specific (non-invariant) features across variations. (Karimi-Rouzbahani et al., 2017).

- Humans rely on specific sets of object parts, referred to as MIRCs (diagnostic features). In other words, some specific object parts were considered more informative than others (Ullman et al., 2016).

diagnostic feature



Main idea

DCNNs did not show such sensitivity to identical diagnostic features;

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

Two Critical Questions

01

As far as variation in DCNNs has led to different diagnostic features,

Will diagnostic features be different for different exemplars of the same category (exemplar variation)?

- Do DCNNs rely on semantically similar diagnostic features from different exemplars of the same object category?

Two Critical Questions

02

What properties do diagnostic features have?

Are they semantically or driven by low-level image statistics?

- Could diagnostic features found for DCNNs be predicted by low-level image statistics?

Question 1

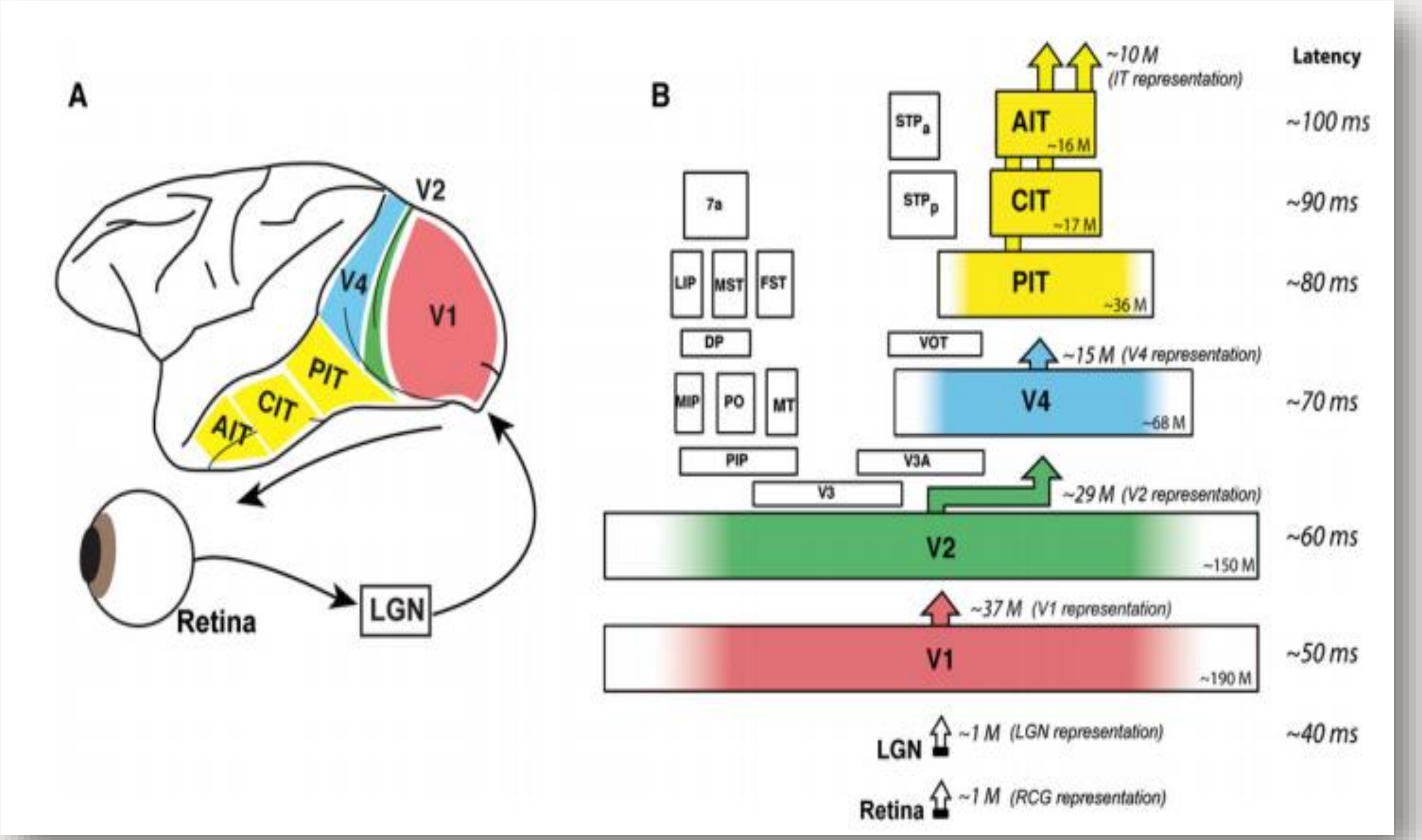
Do DCNNs Rely on Semantically Similar diagnostic features From Different Exemplars of The Same Object Category ?

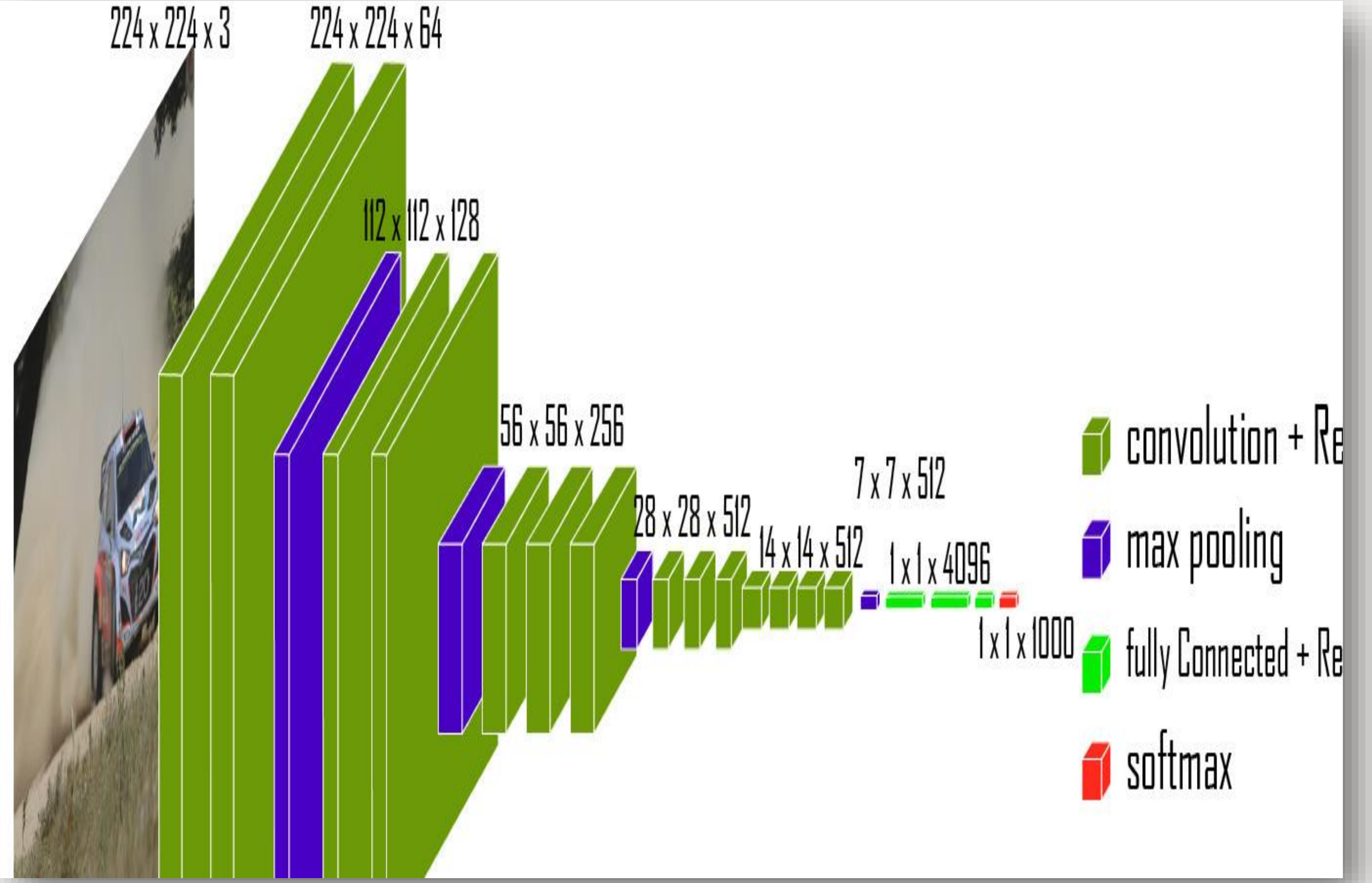
Which DCNN Have We Used and Why?

(Dicarlo et al., 2012)

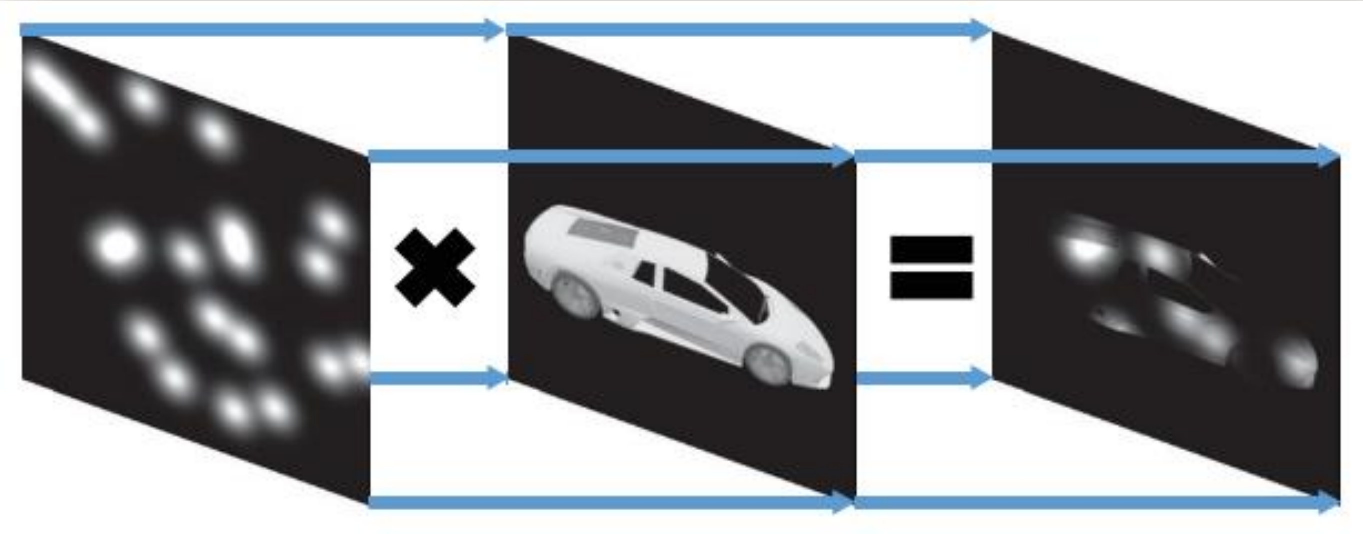
We obtained diagnostic features from VGG16 (Simonyan et al., 2015)

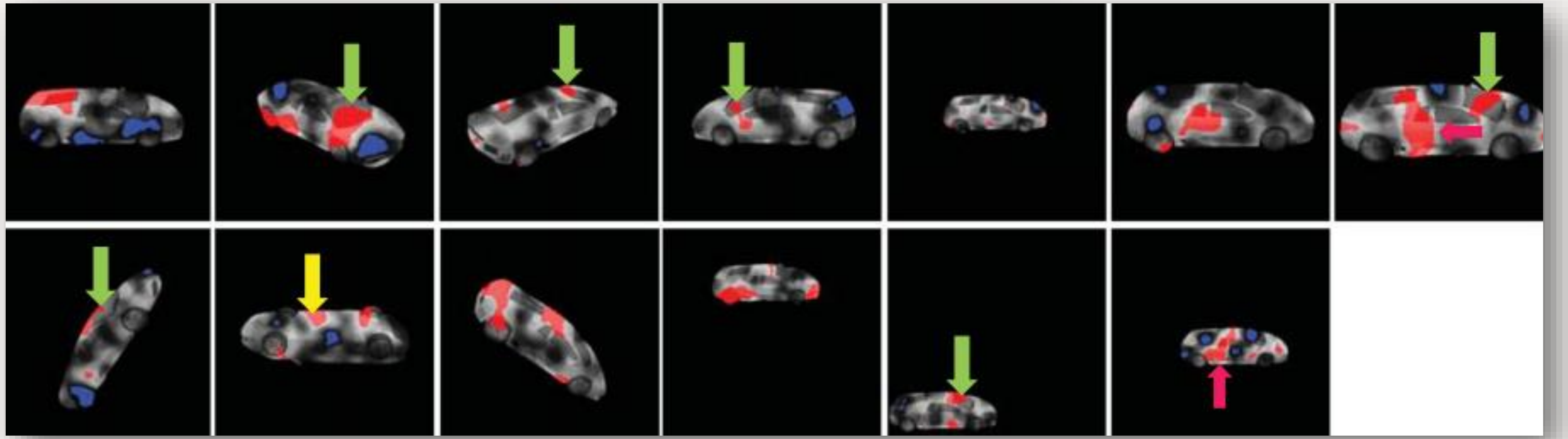
- **One of the most brain-like DCNNs** (Schrimpf et al., 2018).





Use Bubbles Method to Obtain diagnostic features



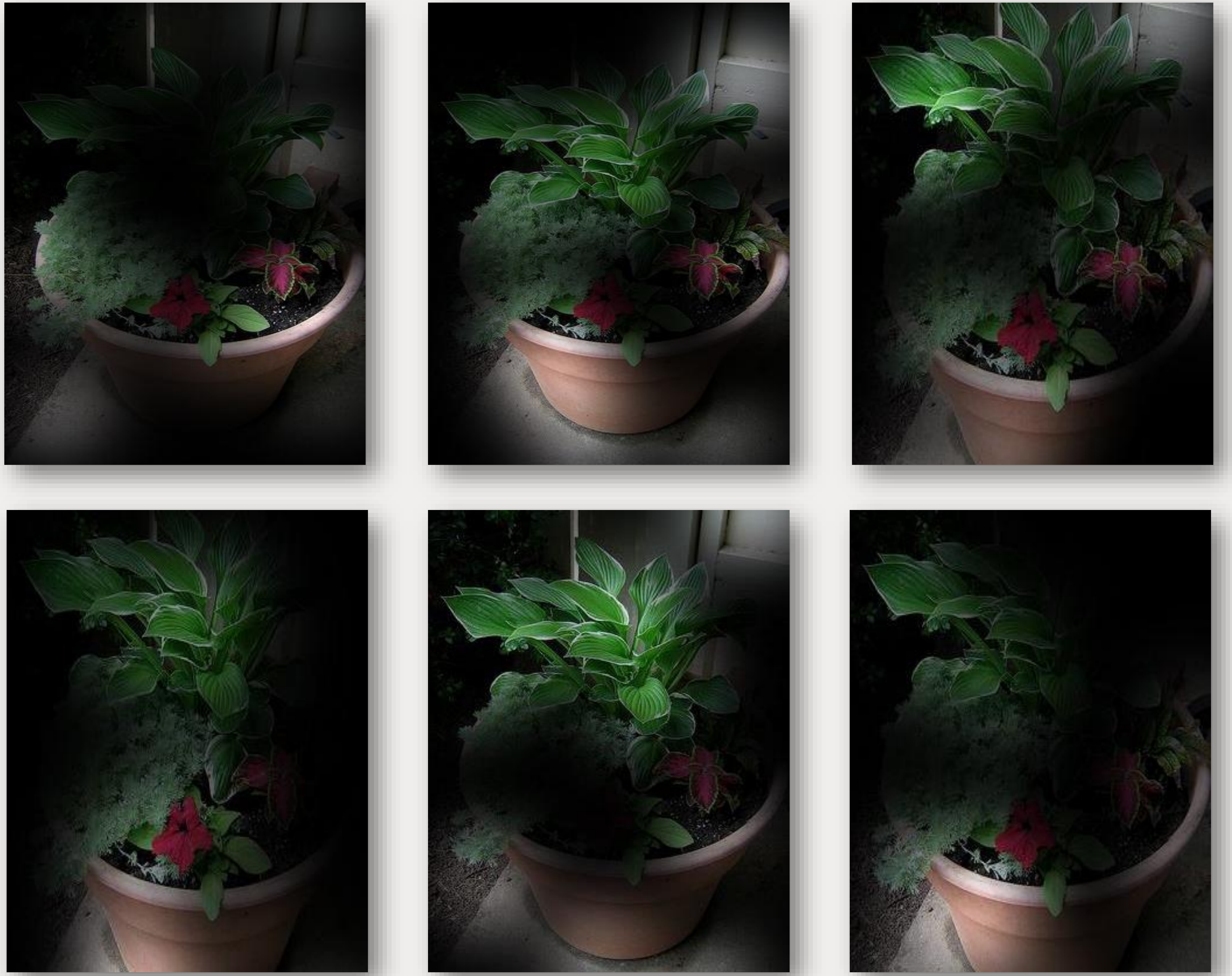


- **We used the well-established Bubbles method** (Gosselin et al., 2001)**.**

- **From our previous work, we found diagnostic features based on human answers to the mask images; We run a lot of trials to calculate model diagnostic features (e.g. 1000 trials).**

- **As an advantage to previous procedures, which detected diagnostic features from preselected discrete image parts, Bubbles sweeps the whole image using continuous masks** (Karimi-Rouzbahani et al., 2017)**.**

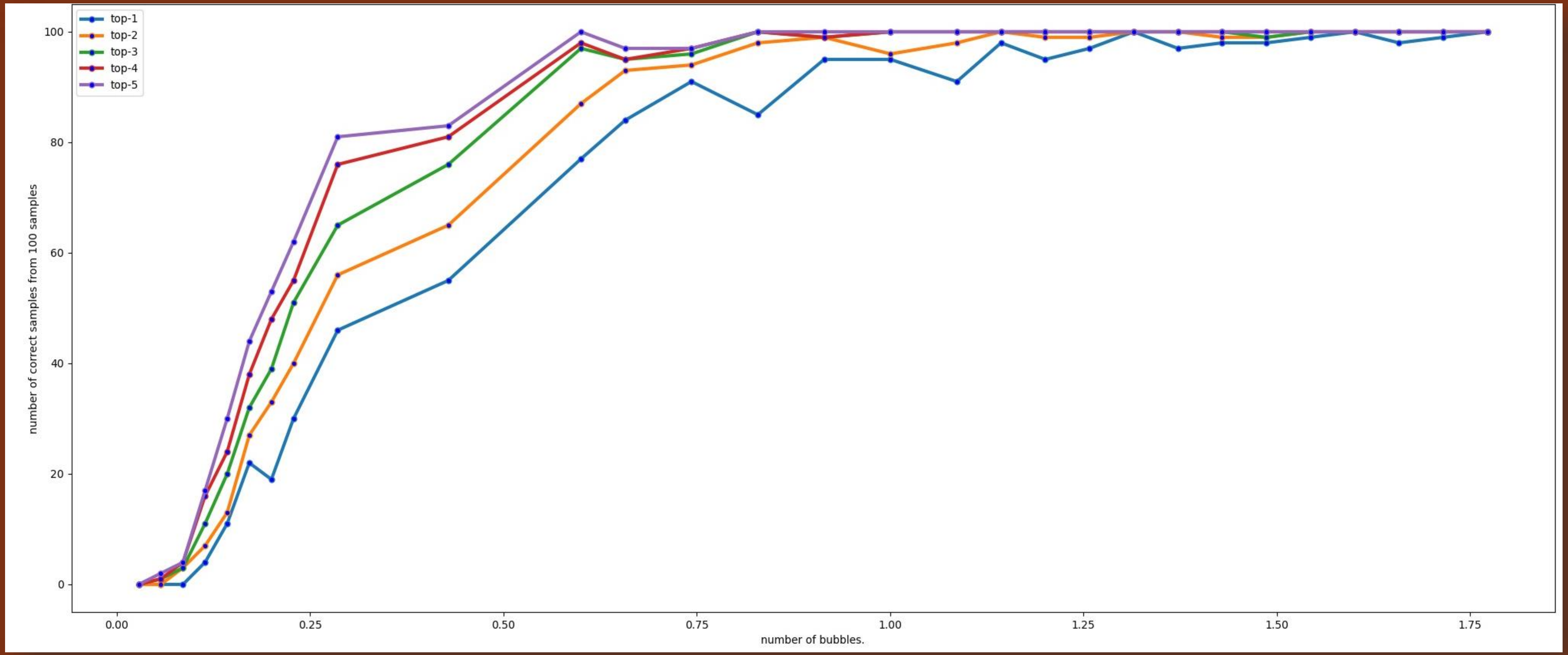
Sample Masked Images for The Current Study



Calculation of Psychometric Function to Choose a Proper Bubble Size

- We obtained model of accuracy as a function of bubble size.
- We chose the bubble size which led to 50% accuracy for each image.

Acc



Bubble Size

Imagenet Large Scale Visual Recognition Challenge

01

The DCNN trained on ImageNet dataset as the largest dataset for object recognition (Deng et al., 2009).

02

Images from test set of ImageNet used here.

03

Diagnostic features obtained for 12 semantically distinct categories (Car, Goldfish , Hamer , Violin, Elephant , Pot, Sewing Machine, Ladybug, Pineapple, Hat, Iron, Hand Blower) each with 16 exemplars.

Results

- **Green regions led to correct recognition of the object category by the DCNN.**
- **Results showed clearly different diagnostic features for distinct exemplars from the same object category.**
- **This reflects the highly variable nature of feature extraction in DCNNs.**
- **Which potentially facilitating recognition under exemplar variations.**

Semantically distinct diagnostic features were Selected for Semantically Similar Exemplars

- <u>Point1</u>: Even though we had background in our images, all of the diagnostic features included the object.
- <u>Point2</u>: Diagnostic features were semantically different.

Question 2

Could diagnostic features found for DCNNs be predicted by low-level image statistics?

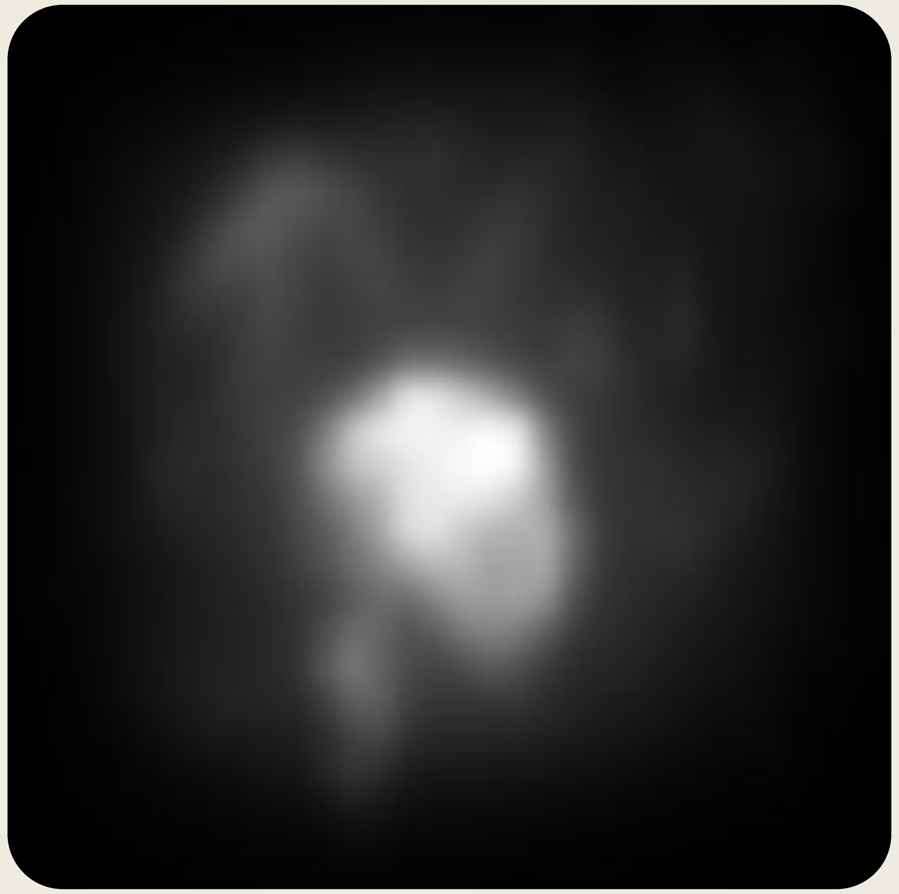
Prediction: __NO__.

➡ Because diagnostic features should contain categorical information to be useful for recognition

- We wondered if the diagnostic features were simply **salient segments** of an image as detected by computational models of saliency.

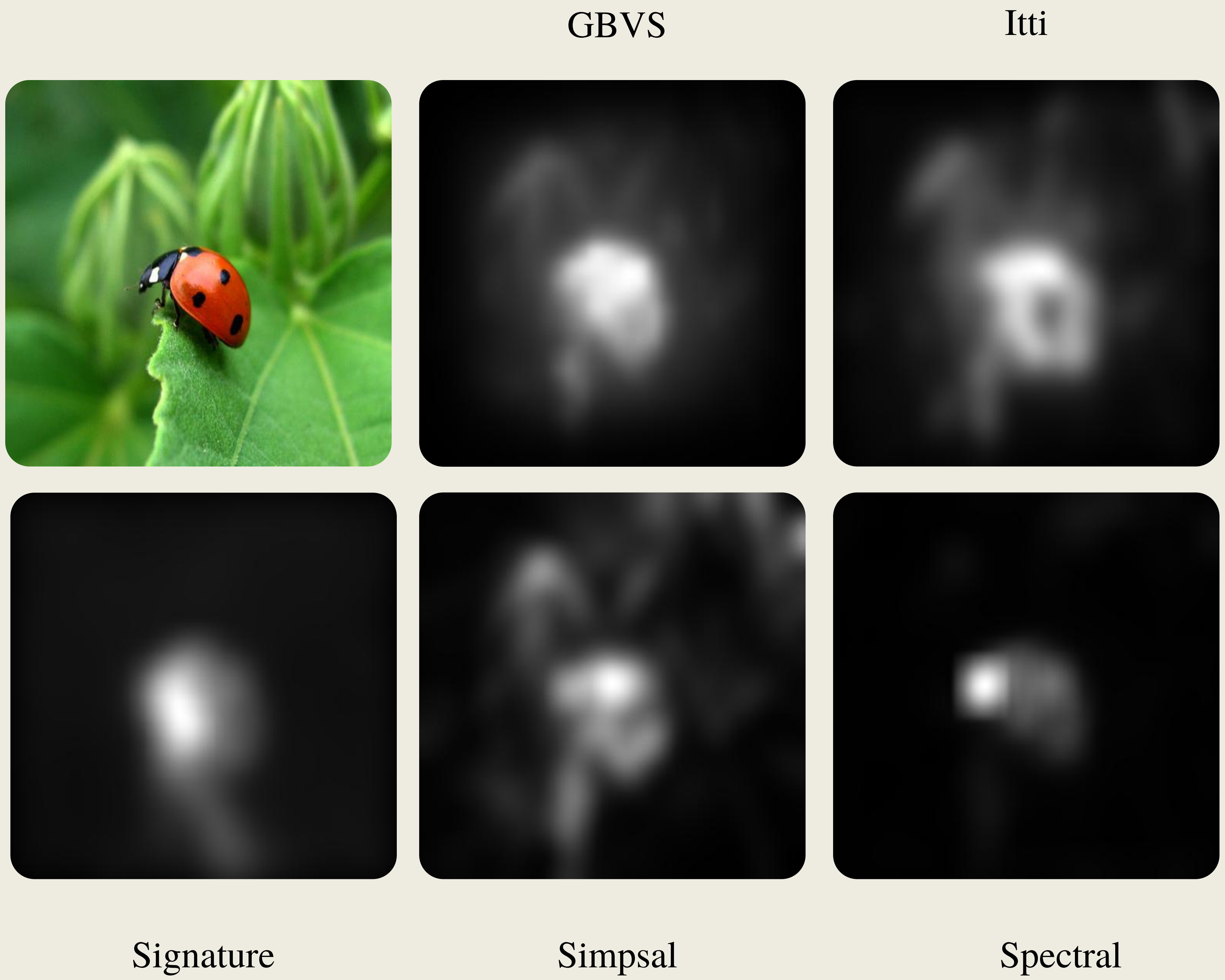
- These models use local **low-level image statistics** (e.g. color, orientation, contrast) to predict the most outstanding (salient) parts of image.

- They have also predicted the location of human overt attention (gaze) on the image (Kimura et al., 2013).



- They can indicate image areas that are visually (rather than semantically) distinct from other areas.

- We obtained the salient segments of all the images in our dataset using 5 of the most brain-plausible saliency models (‘Itti’, ‘GBVS’, ‘Simpsal’, ‘Spectral’, ‘Signature’) (Kimura et al., 2013)



Question 2:

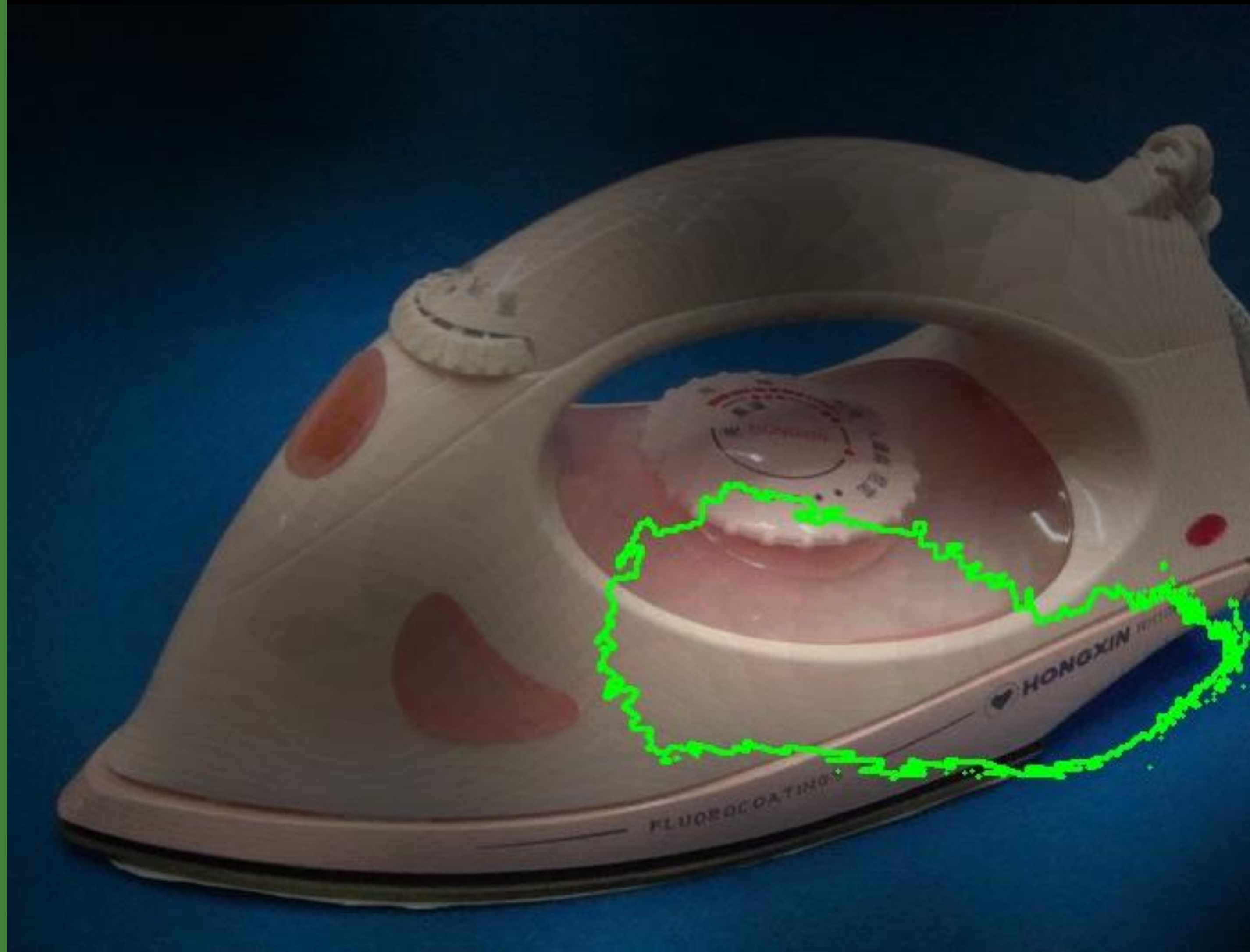
- Could diagnostic features found for DCNNs be **predicted by low-level image statistics**?

Salient Regions

Results:

- Diagnostic features obtained from the DCNN and the salient regions obtained from the saliency models were **qualitatively** and **quantitatively** different.

diagnostic features Given from The DCNN Model



Qualitatively different diagnostic features were Obtained from Saliency Model and the DCNN Model

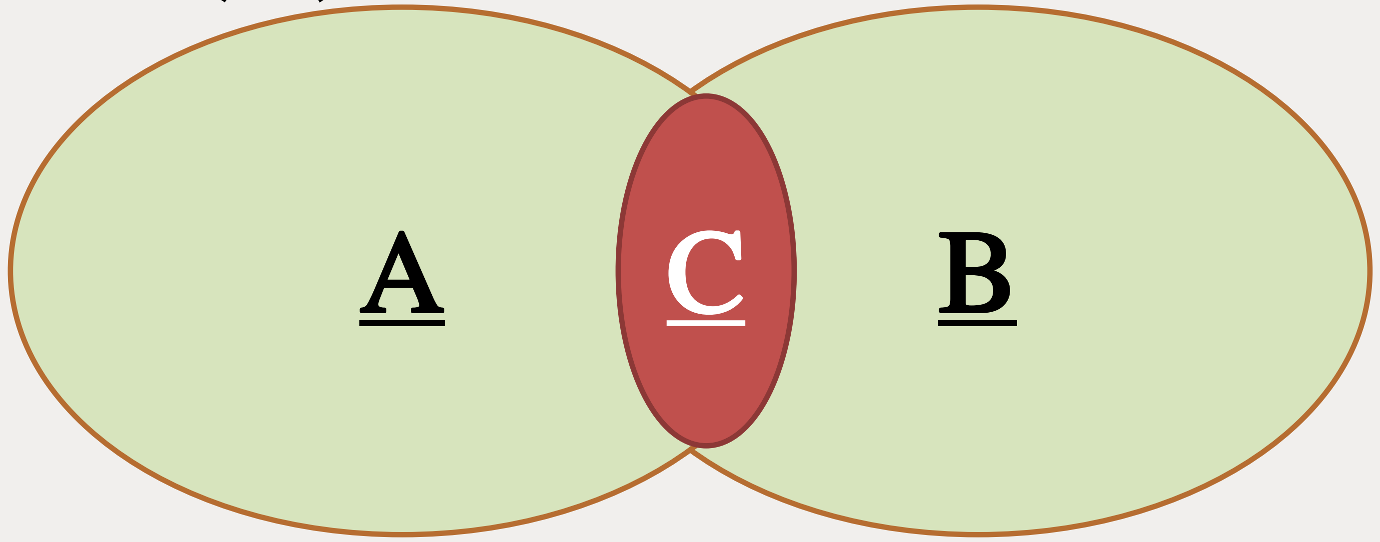
The DCNN diagnostic features

Salient Regions

- <u>Point1</u>: Unlike the DCNN, some of the salient regions from saliency model, are outside of the object.
- <u>Point2</u>: The diagnostic features and salient regions are qualitatively different.

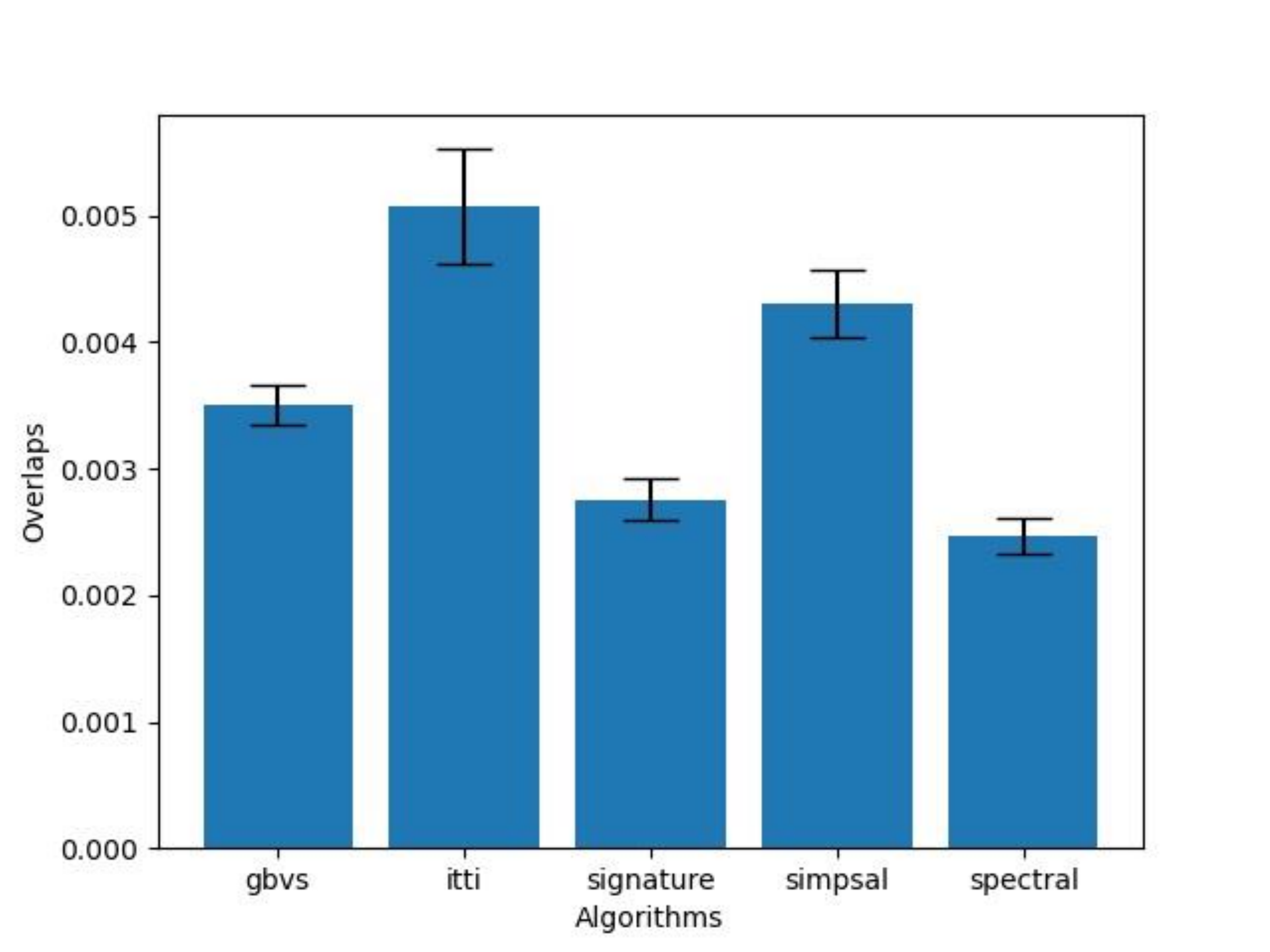
Quantitatively Comparison of diagnostic features (A), and Salient Regions (B)

- We calculated the overlap of the DCNN diagnostic features and Salient Regions using following equation:
- Non parametric permutation test was used to evaluate the significance ($p<0.05$) of overlap



$$Overlap = \frac{C}{A+B}$$

Overall Overlap



This suggests that, rather than relying on salient low-level image statistics, DCNNs may rely on object segments which probably contain semantic category information relevant for object recognition.

Random Permutation Test Showed NO Significant Overlap Between diagnostic features and Salient Regions!

Ongoing work:

Do human use the same diagnostic features as DCNNs models ?

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

References

[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] Deng, 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).

[8] Kimura, A., Yonetani, R. and Hirayama, T., 2013. Computational models of human visual attention and their implementations: A survey. IEICE TRANSACTIONS on Information and Systems, 96(3), pp.562-578.

[9] Itti, L., Koch, C. and Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on pattern analysis and machine intelligence, 20(11), pp.1254-1259.

Thank You !

MohammadHossein NikiMaleki, Hamid Karimi-Rouzbahani

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

ECVP - August 2021
43rd European Conference on Visual Perception