ECVP - August 2021
NikiMaleki_COM_Poster_1044

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

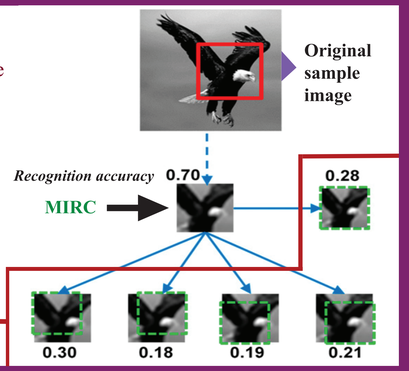
Introduction

- Deep Convolutional Neural Networks (DCNNs) are among the most accurate models of human object recognition.
- Humans rely on specific diagnostic parts of objects for accurate recognition (called MIRCs (Minimal Recognizable Configuration) by Ullman et al., 2016).
- Humans use those diagnostic parts relatively consistently/invariantly across variations (in-plane rotation, size and translation; Karimi-Rouzbahani et al., 2017).
- DCNNs, however, rely on relatively inconsistent/distinct diagnostic parts across variations of the same objects (Karimi-Rouzbahani et al., 2017).
- This suggests that, as opposed to humans, DCNNs seem to rely on different mechanisms for recognition under variations.
- We address two questions in this study:
  1- Are diagnostic parts different for different exemplars of the same category (exemplar variation)?
  2- What properties do diagnostic parts have? Could they be predicted by low-level image statistics?

Diagnostic object parts

MIRCs (diagnostic parts) were defined as smallest patches of an image which lead to its correct recognition.

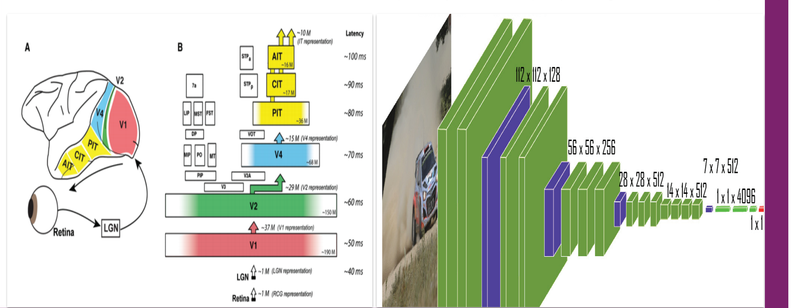
Any smaller patchs leads to a considerable drop in accuracy.
(Ullman et al., 2016)



Mechanisms of visual object recognition in brain and DCNN

Hierarchical feature extraction in Inferior Temporal Cortex (ITC; Dicarlo et al., 2012)

VGG-16: it has shown very similiar representations to the monkey ITC (Schrimpf et al., 2018) & (Simonyan et al., 2015)

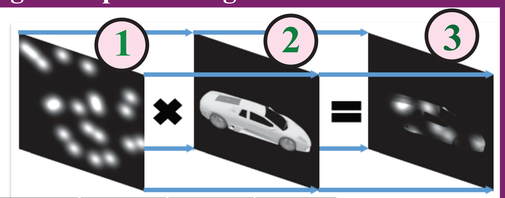


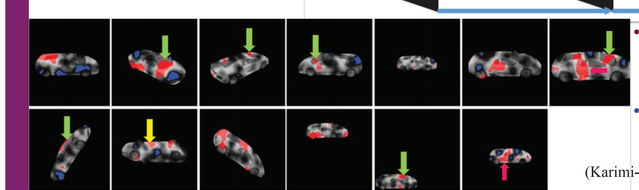
Methods: detecting diagnostic parts using Bubbles method

1- Random mask

2- Sample image

3- Masked image

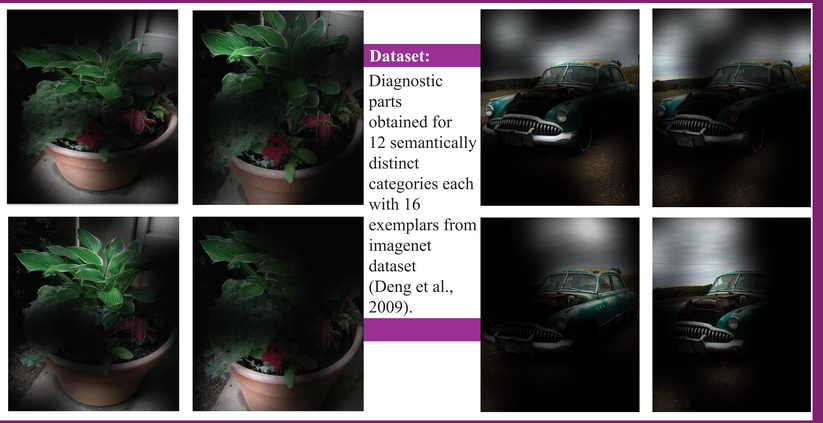




- *Red regions led to correct recognition (diagnostic)*
- *Blue regions led to incorrect recognition (anti diagnostic)*

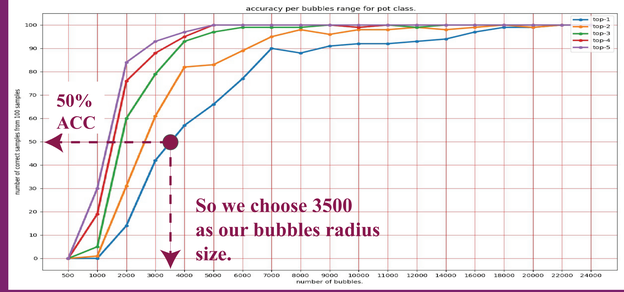
(Karimi-Rouzbahani et al., 2017)

Methods: the dataset and sample masked images



Dataset:
Diagnostic parts obtained for 12 semantically distinct categories each with 16 exemplars from imagenet dataset (Deng et al., 2009).

Methods: psychometric function for the DCNN

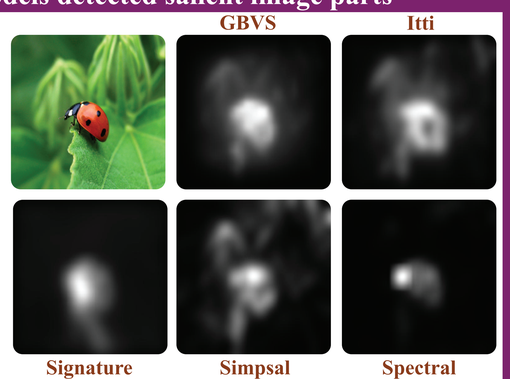


- Psychometric function was used to determine the bubble size which led to 50% accuracy. This was needed for using Bubbles analysis.

Methods: five saliency models detected salient image parts

What is Saliency Model?

- Saliency models find the most salient parts of an image. (Itti et al., 1998.)
- We want to see whether our diagnostic parts are predictable by low-level image statistics (i.e. salient parts).
- Salient parts were obtained from all images in our dataset using 5 of the most brain-plausible saliency models. (Kimura et al., 2013)



Results Q1: are diagnostic parts different for different exemplars of the same category?

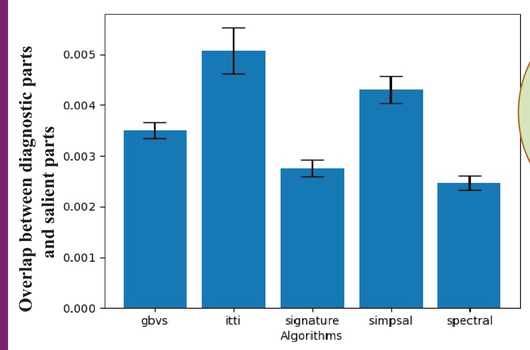
- Diagnostic parts were semantically different for semantically similiar object images.

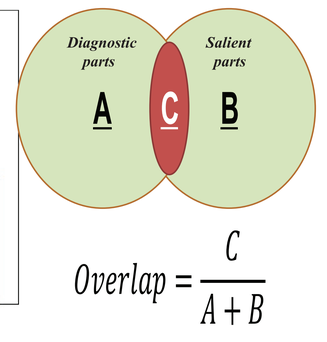
Results Q2: are diagnostic parts predicted by salient image parts?

- Visual inspection shows **NO** overlap between the DCNN diagnostic parts and salient parts.
- Even though we had background in images, **ALL** of the diagnostic parts included the object.

Results Q2: are diagnostic parts predicted by salient image parts?

- We calculated the overlap between diagnostic parts and salient parts of our 5 saliency models using the following equation:





Summary

- **Results Q1:** ***semantically distinct*** diagnostic parts were found for ***semantically similiar*** objects.
  - This reflects the highly variable nature of feature extraction in DCNNs, potentially facilitating recognition under exemplar variations.
- **Results Q2:** diagnostic parts obtained from the DCNN and the salient parts obtained from the saliency models were qualitatively and quantitatively different.
  - Random permutation test showed NO significant overlap between the DCNN diagnostic parts and salient parts.
- This suggests that, rather than relying on salient low-level image statistics, DCNNs may rely on object parts which probably **contain semantic category information** relevant for object recognition.

Acknowledgement

This research was funded by UK Royal Society's Newton International Fellowship NIF \R1\192608 to H.K-R.

The British Academy
The Royal Society
The Academy of Medical Sciences
Newton International Fellowships

References

[1] Ullman, S., Assif, L., Fetaya, E. and Harari, D., 2016. Proceedings of the National Academy of Sciences, 113(10), pp.2744-2749.
[2] H. Karimi-Rouzbahani, N. Bagheri, and R. Ebrahimpour, Scientific Reports, vol. 7, no. 1, 2017.
[3] K. Simonyan and A. Zisserman, 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, 2018.
[5] Gosselin, F. and Schyns, P.G., 2001. 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. 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. IEICE TRANSACTIONS on Information and Systems, 96(3), pp.562-578.
[9] Itti, L., Koch, C. and Niebur, E., 1998. IEEE Transactions on pattern analysis and machine intelligence, 20(11), pp.1254-1259.