Poster & Paper

bit.ly/3r4CylX

Exemplary Natural Images Explain CNN Activations Better than State-of-the-Art Feature Visualization

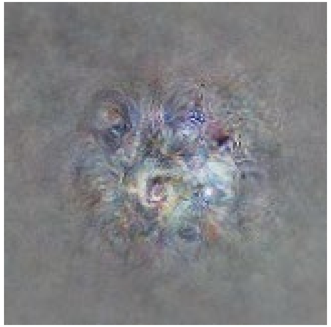
Judy Borowski*, Roland Zimmermann*
Judith Schepers, Robert Geirhos, Tom Wallis†, Matthias Bethge†, Wieland Brendel†

International Conference on Learning Representations (ICLR) 2021

EBERHARD KARLS UNIVERSITÄT TÜBINGEN
imprs-is
Bundesministerium für Bildung und Forschung
ROBUST VISION
DFG

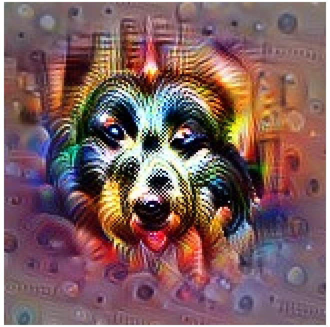
Feature Visualizations

Step 0



Step 4

Step 48



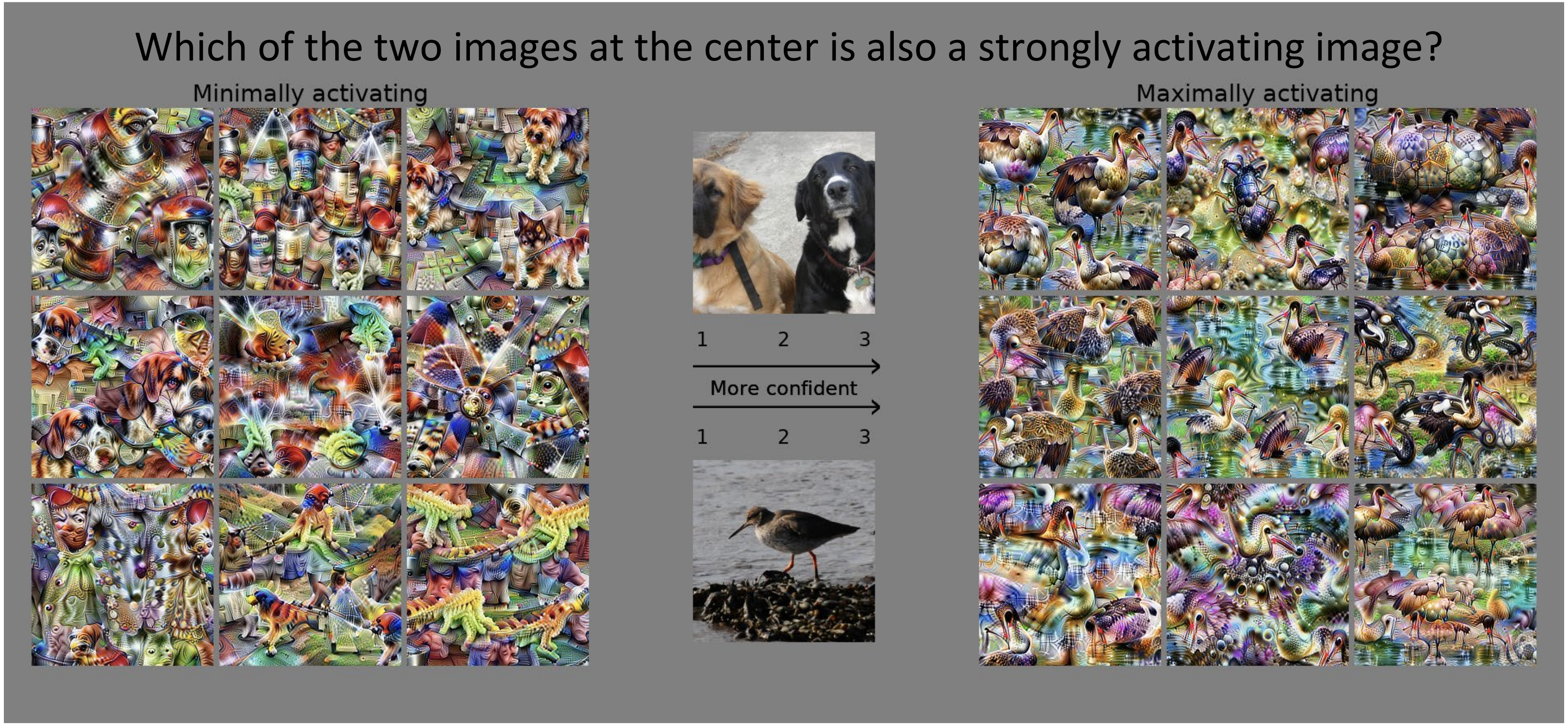
Step 2048

Stronger activation

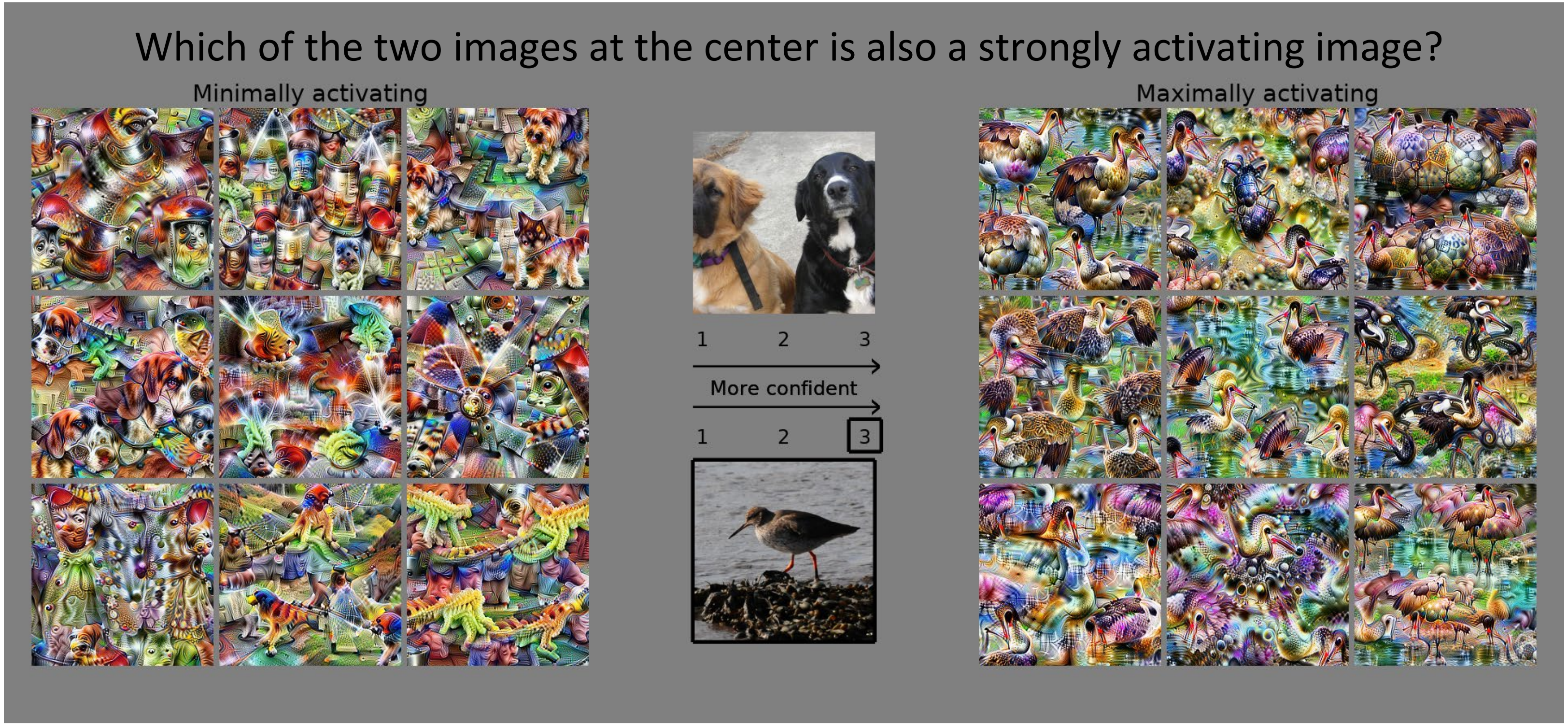
Olah et al. (2017)

Olah et al. (2017)

How helpful are feature visualizations for humans?

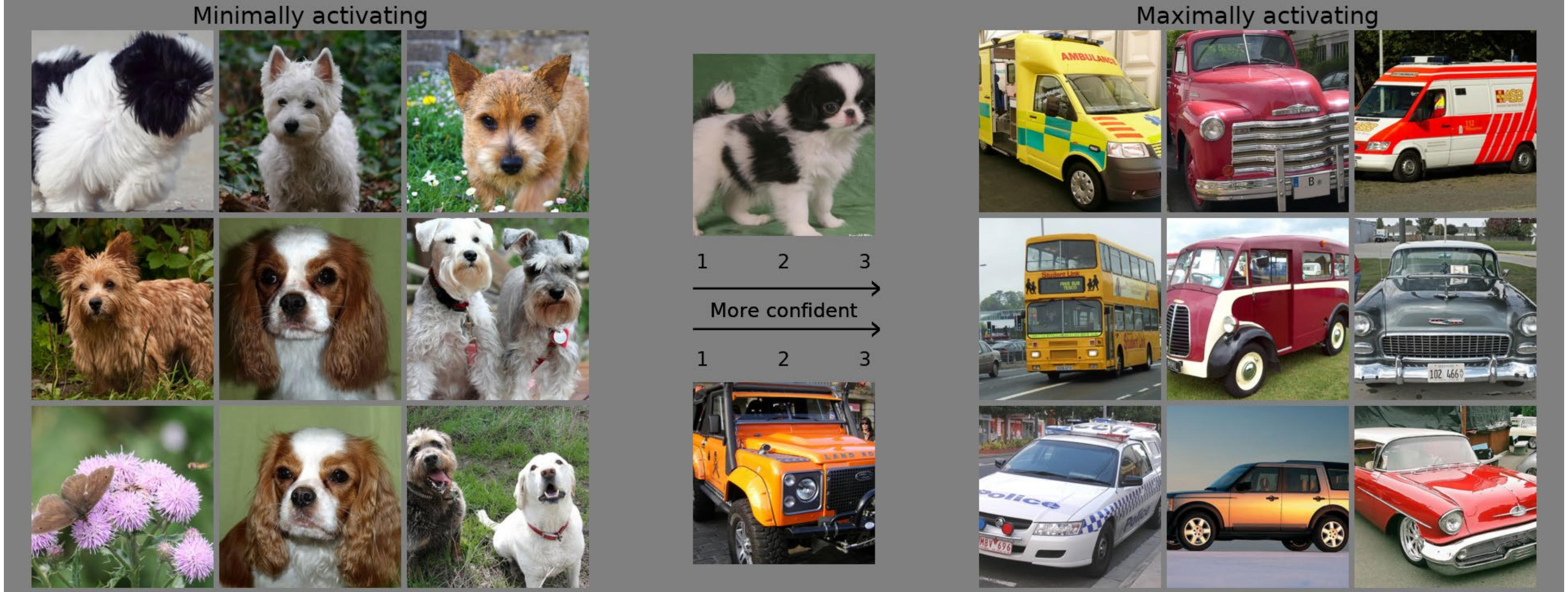


Which of the two images at the center is also a strongly activating image?

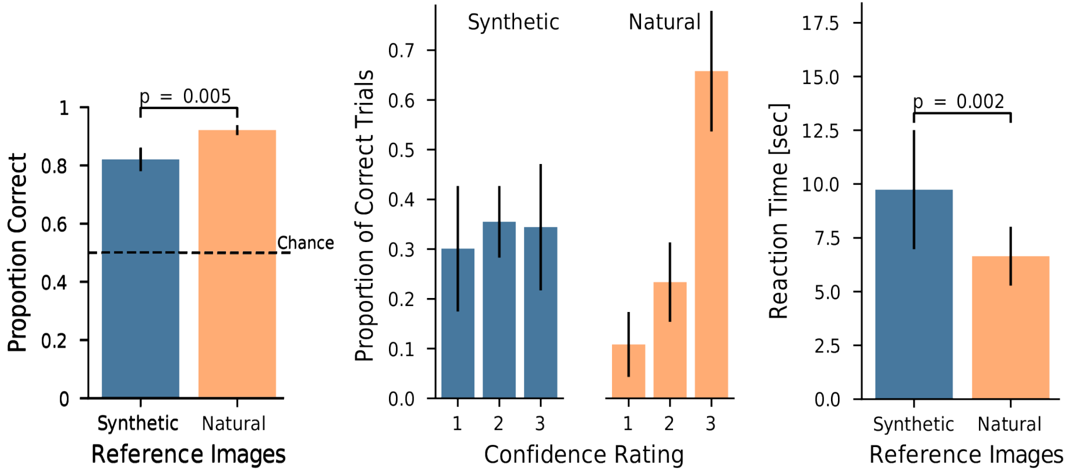


Which of the two images at the center is also a strongly activating image?

Which of the two images at the center is also a strongly activating image?



- Synthetic images provide helpful information about CNN activations
- But exemplary natural images are even more helpful



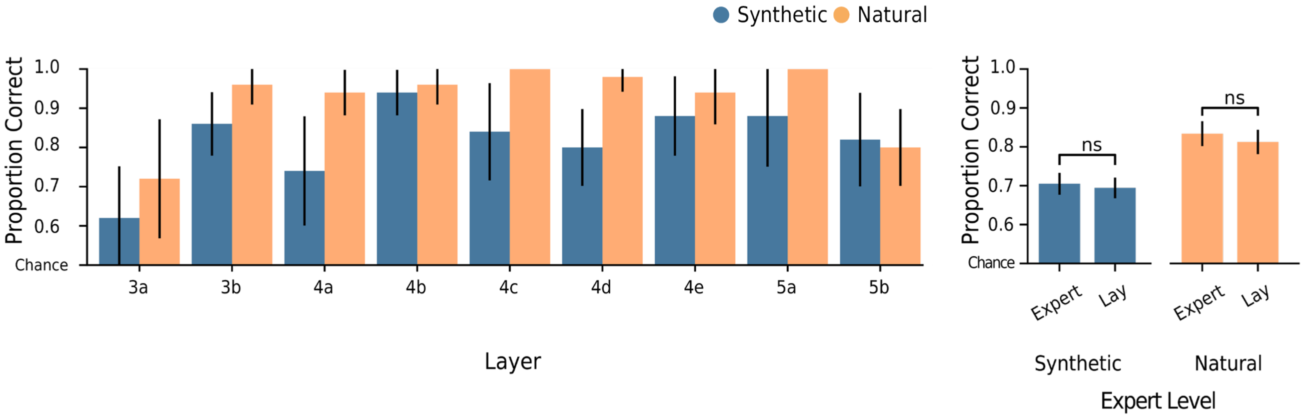
Higher performance

Higher confidence

Faster responses

Natural images more helpful than synthetic images

- Synthetic images provide helpful information about CNN activations
- But exemplary natural images are even more helpful
- Findings hold across various aspects



Natural images more helpful than synthetic images

- Synthetic images provide helpful information about CNN activations
- But exemplary natural images are even more helpful
- Findings hold across various aspects

→ Need for thorough quantitative evaluations of feature vis
→ Interpretability methods should improve over the baseline of natural images

Poster Presentation:
May 4th at 1 and 3 am (PDT)

Poster & Paper

bit.ly/3r4CylX