Modelling response to trypophobia trigger using intermediate layers of ImageNet networks

Piotr Woźnicki[1], Michał Kuźba[2,3], Piotr Migdał[4,†]

[1]Medical University of Warsaw
[2]Faculty of Mathematics and Information Science, Warsaw University of Technology
[3]Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw
[4]p.migdal.pl - AI consulting and workshops
[†]contact: pmigdal@gmail.com

Trypophobia

Trypophobia is a condition where a person experiences an intense and disproportionate fear towards clusters of small holes. It has been suggested in the literature that aversion to clusters is an evolutionarily prepared response to stimuli that resemble parasites and infectious disease[1]. It is known that threatening cues in the environment receive priority in visual attention processing and automatically capture attention[2]. We state the hypothesis that trypophobia cues can be effectively recognized by a simplified neural network with massively reduced number of layers and parameters.

Figure 1: The holes in lotus seedheads elicit feelings of discomfort or repulsion in some people.

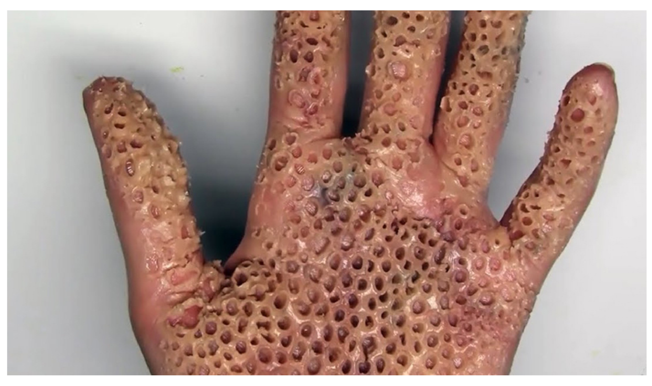

Figure 2: A trypophobic reaction to a parasitic or infectious disease threat could be potentially advantageous.

Dataset and experiment design

The training dataset that we use consists of 6000 trypophobia triggering and 10000 neutral images. The test set contains 500 trypophobic and 500 neutral images. It can be found on GitHub [6].
We train a few different models based on popular CNN architectures, including VGG16, Inceptionv3 and ResNet50 for the task of classifying image as trypophobic/neutral. We also experiment with using pretrained weights from ImageNet and freezing convolutional blocks. For further experiments we choose the best performing VGG16 model[7]. We create four new models by using only respectively the first 4, 3, 2 and 1 convolutional blocks and experiment with the design of the top layers.

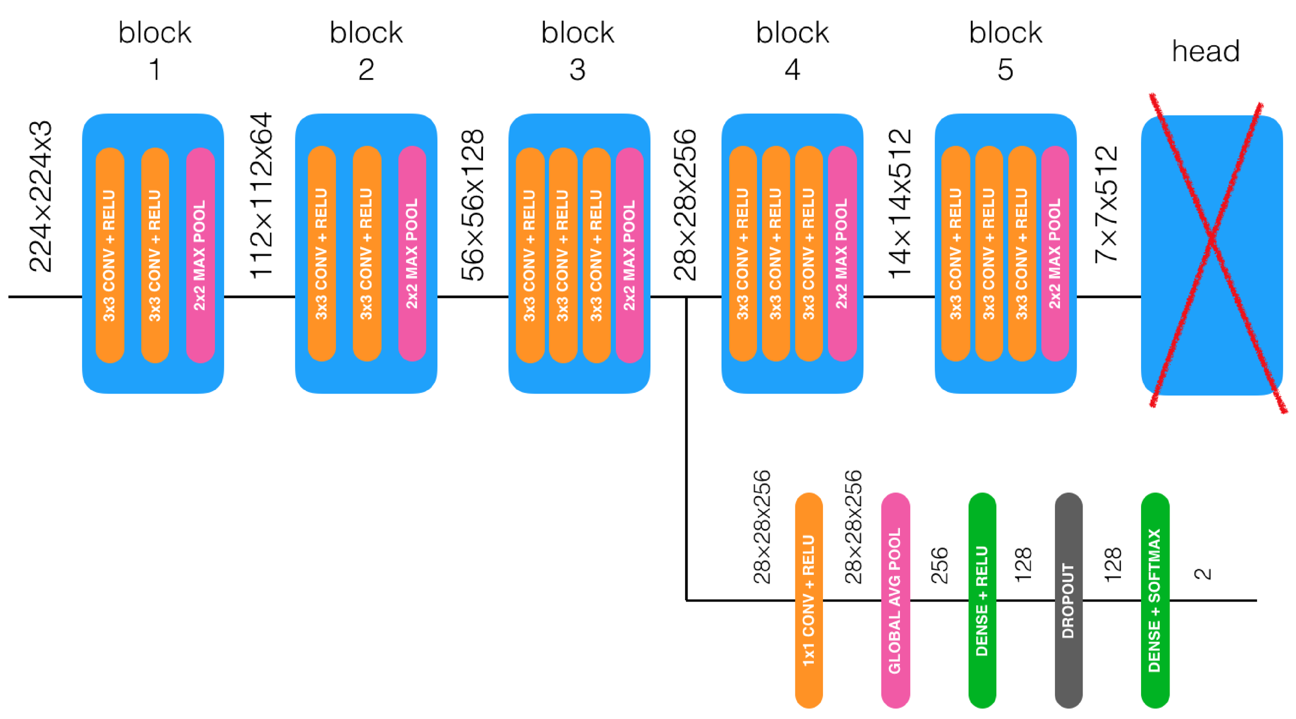

Figure 3: Model architecture: after k convolutional blocks of ImageNet-pretrained VGG16, we use a fully convolutional head. We use only 1x1 convolution, to avoid introducing new spatial features. The size of hidden layer (here: 128) and the dropout rate (here: 0.5) are our hyperparameters.

Results

We report accuracy ranging from 82.8% to 94.0%, as compared to 94.5% achieved by the full VGG16 model. In our experiments we have consistently seen networks with 4 and 3 convolutional blocks performing comparably to the full VGG16 and being more stable while training and less prone to overfitting.

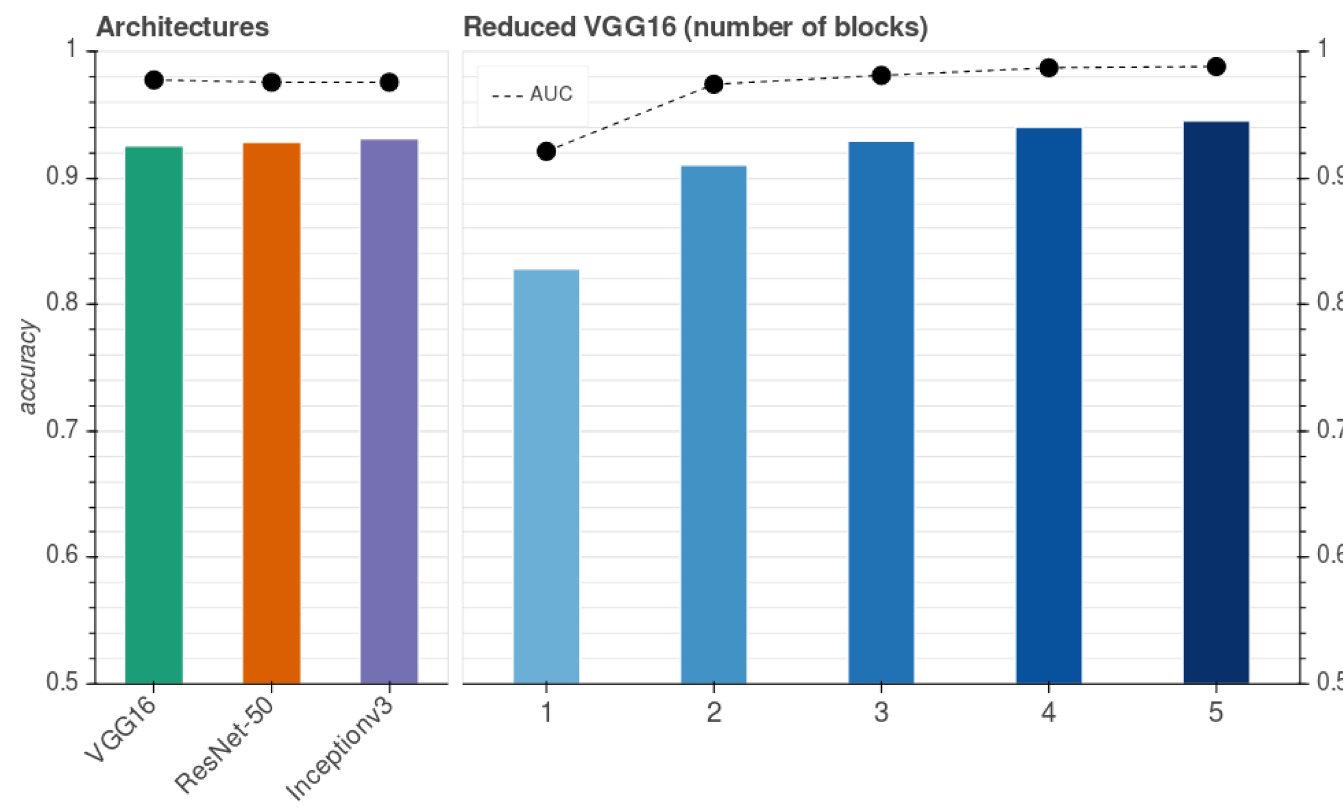

Figure 4: Comparison of various image classification architectures (left). Accuracy of dissected variants of VGG16 (right).

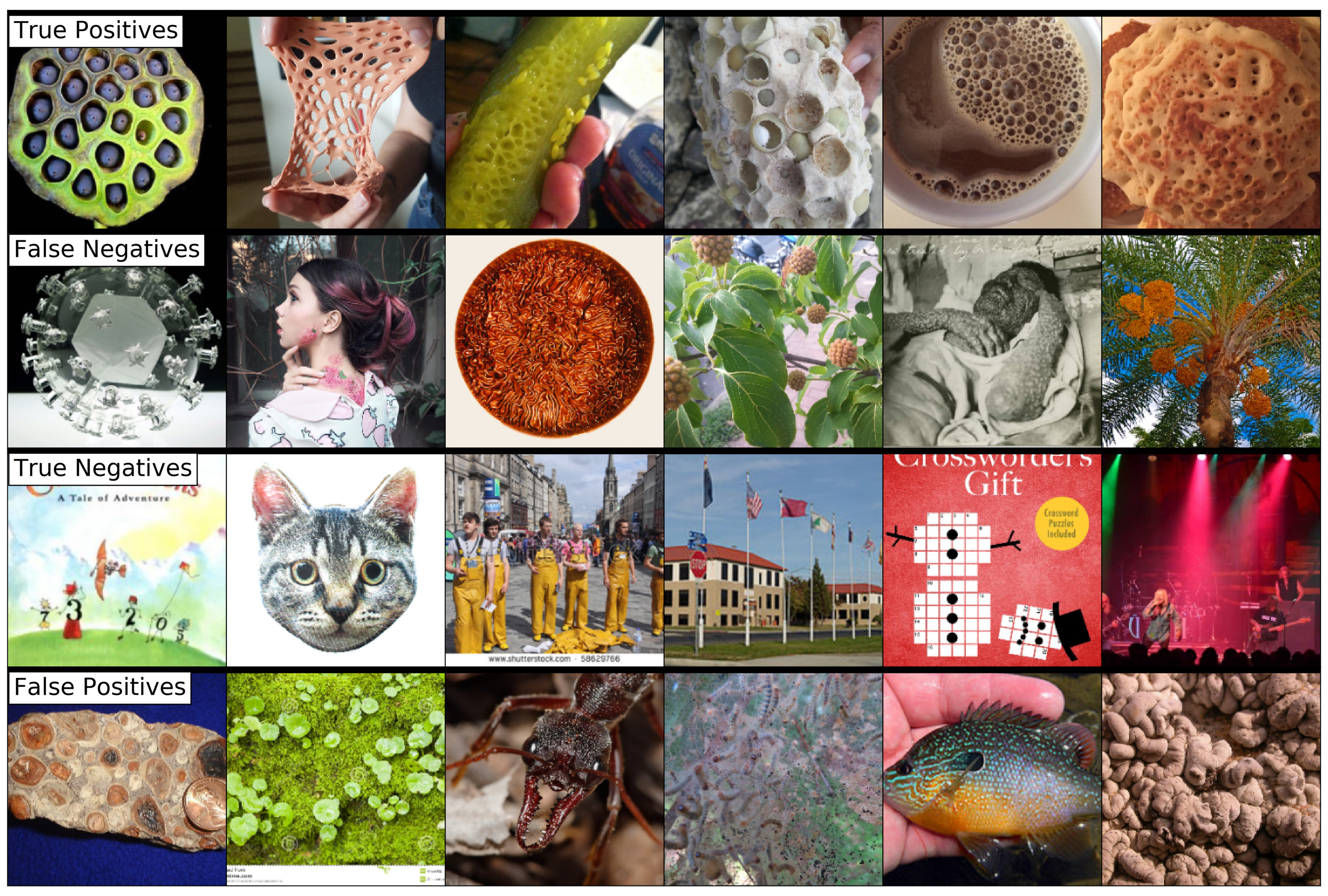

Figure 5: Selected images from the test dataset, grouped by predicted vs actual class, for the task of binary classification (positive = trypophobia pattern present).

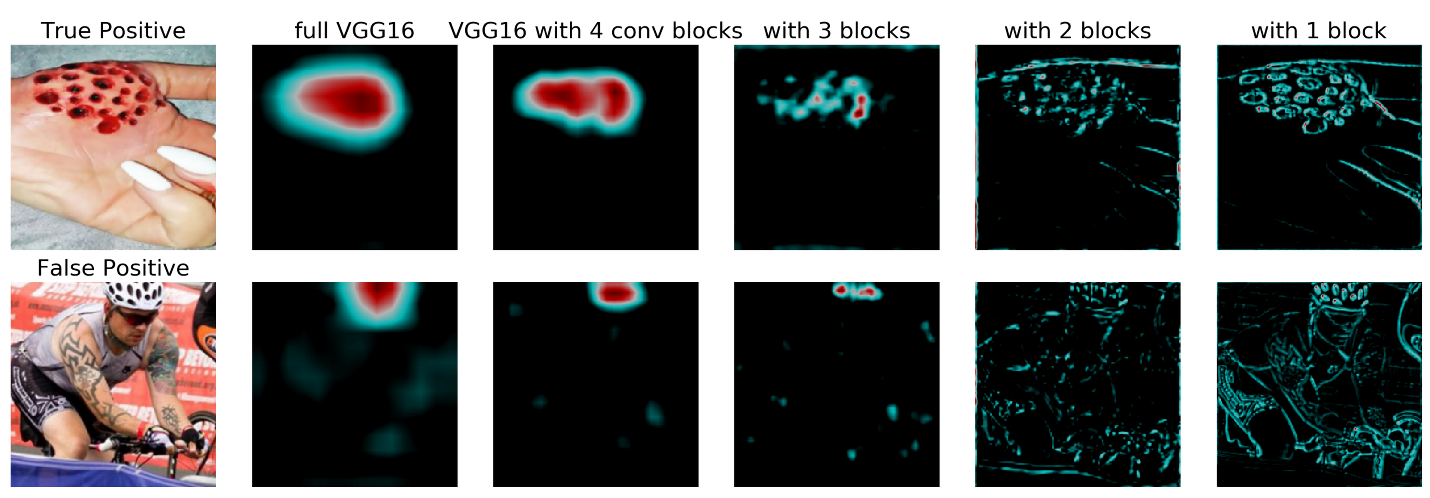

Figure 6: Gradient-weighted Class Activation Mapping[8] for two images classified as trypophobic.

Acknowledgements

This work was financially supported by NCN Opus grant 2016/21/B/ST6/0217.

References

1. Martínez-Aguayo JC, Lanfranco RC, Arancibia M et al, Trypophobia: What Do We Know So Far? A Case Report and Comprehensive Review of the Literature. Front Psychiatry. 2018;9:15. doi:10.3389/fpsyt.2018.00015
2. Cole, G. G., & Wilkins, A. J. (2013). Fear of Holes. Psychological Science, 24(10), 1980–1985. doi:10.1177/0956797613484937
3. Lecun Y, Bottou L, Bengio Y et al, Gradient-based learning applied to document recognition, Proceedings of the IEEE, 1998 (2278-2324)
4. Olah, Chris and Mordvintsev, Alexander and Schubert, Ludwig, Feature Visualization, Distill (2017), https://distill.pub/2017/feature-visualization
5. Shirai R., Banno, H. Ogawa, H. Atten, Trypophobic images induce oculomotor capture and inhibition, Percept Psychophys (2018). doi:10.3758/s13414-018-1608-6
6. Puzio A, Uriasz G, dataset available at: https://github.com/cytadela8/trypophobia
7. Simonyan K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs] 2014.
8. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Gradcam: Visual explanations from deep networks via gradient-based localization. (2016)