NEURAL INFORMATION PROCESSING SYSTEMS

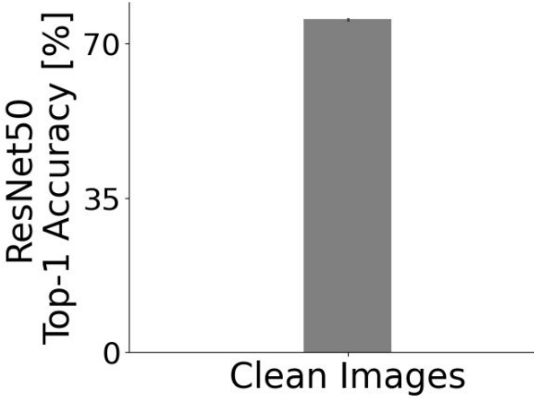
Explicitly Modeling Subcortical Vision with a Neuro-Inspired Front-End Improves CNN Robustness

Lucas Piper, Arlindo L. Oliveira, Tiago Marques

inesc id lisboa · TÉCNICO LISBOA · BCRP Breast Cancer Research Program · Champalimaud Foundation · FACULDADE DE MEDICINA LISBOA

Motivation: CNNs are vulnerable to a wide range of perturbations

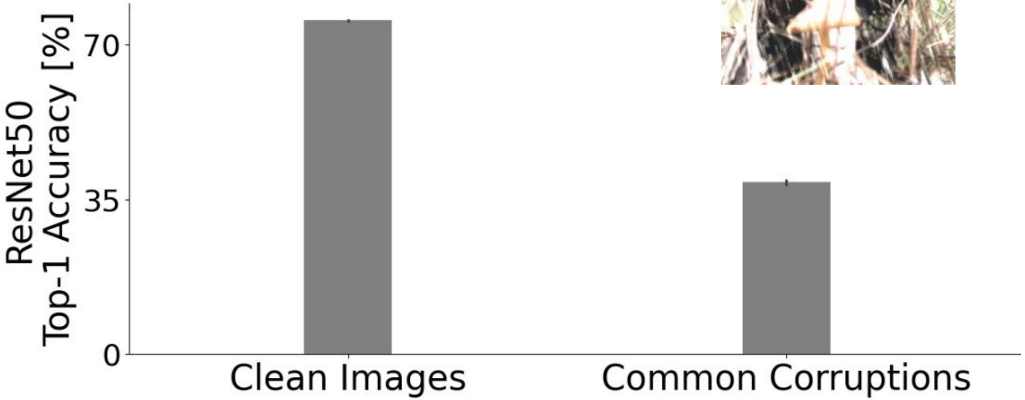
ImageNet



Motivation: CNNs are vulnerable to a wide range of perturbations

ImageNet

ImageNet-C[1]

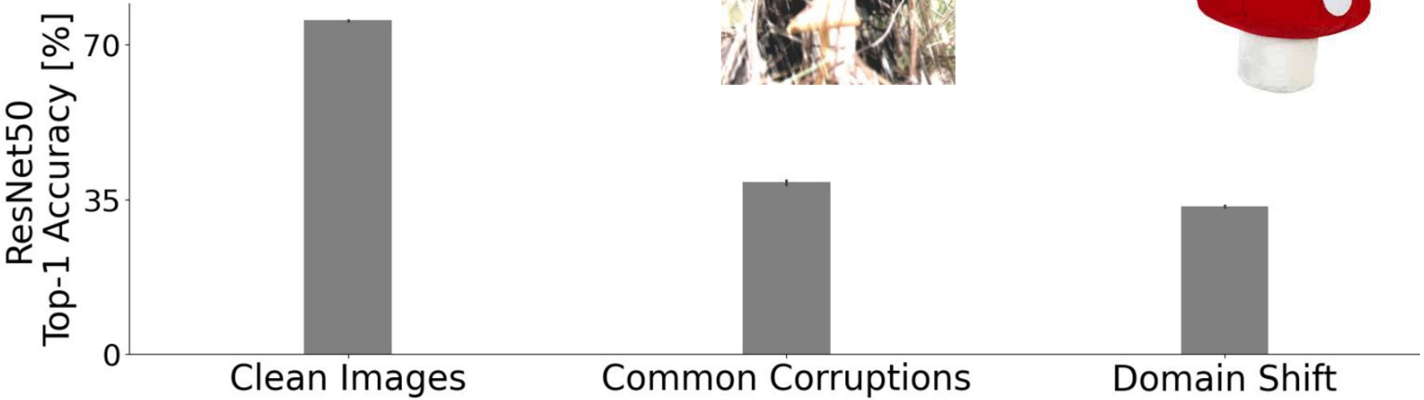


Motivation: CNNs are vulnerable to a wide range of perturbations

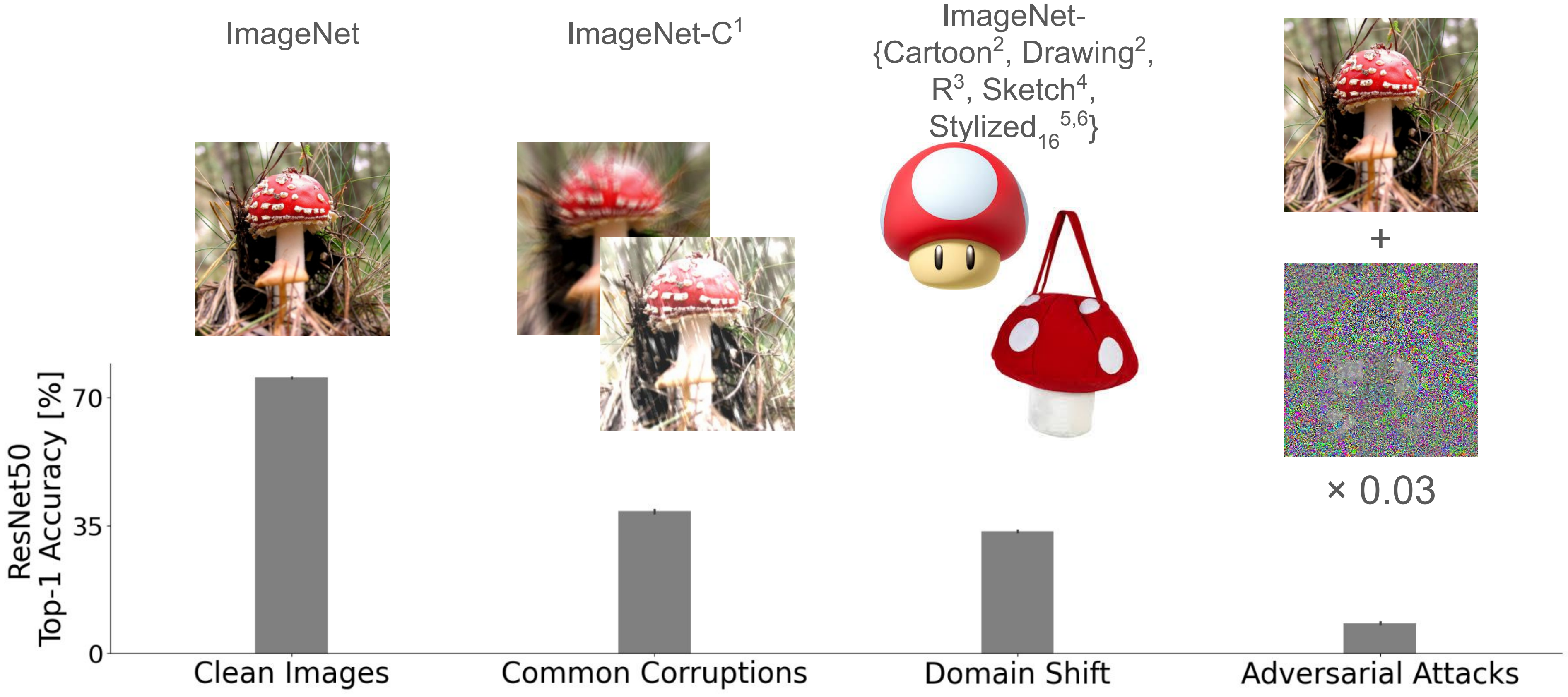
ImageNet

ImageNet-C[1]

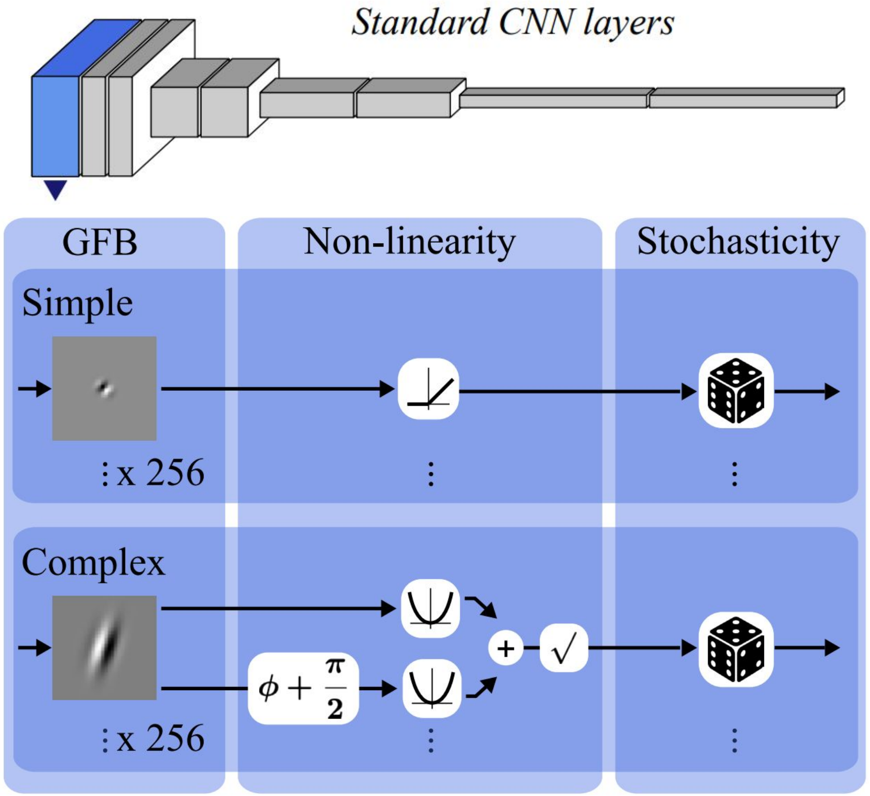
ImageNet-{Cartoon[2], Drawing[2], R[3], Sketch[4], Stylized$_{16}$[5,6]}



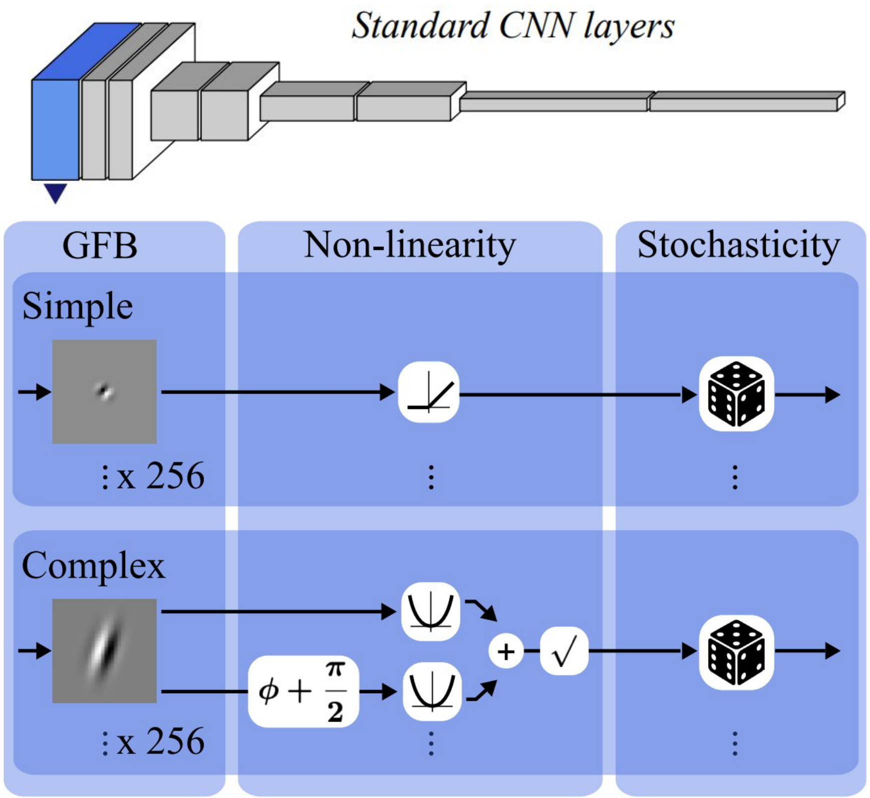
Motivation: CNNs are vulnerable to a wide range of perturbations



Related Work: VOneNets[7] model V1 in front of a CNN back-end

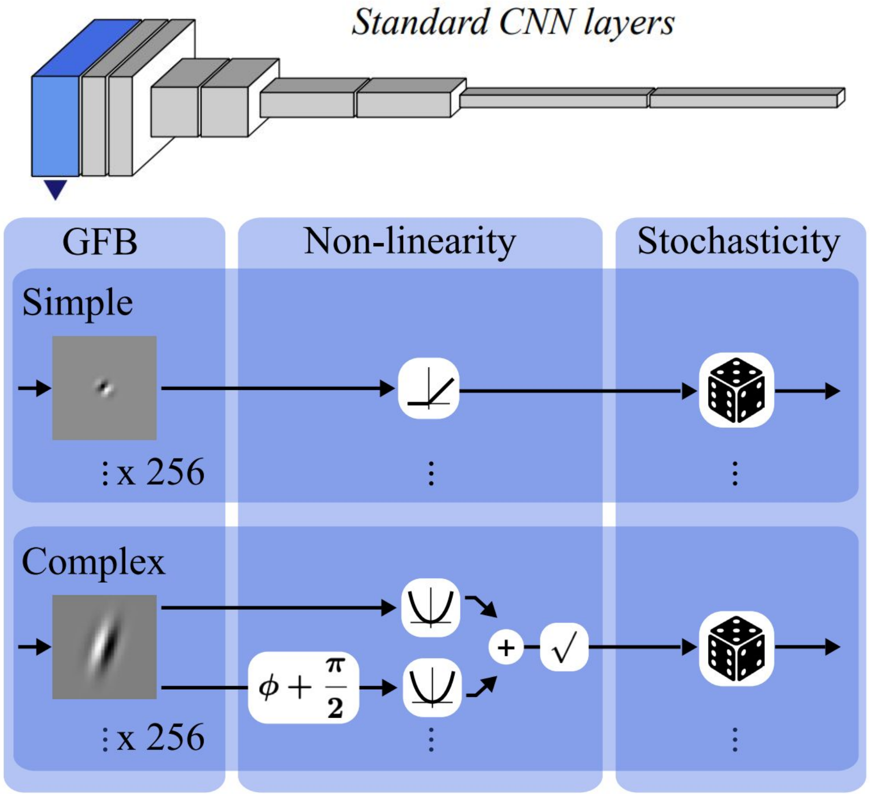


Related Work: VOneNets[7] model V1 in front of a CNN back-end



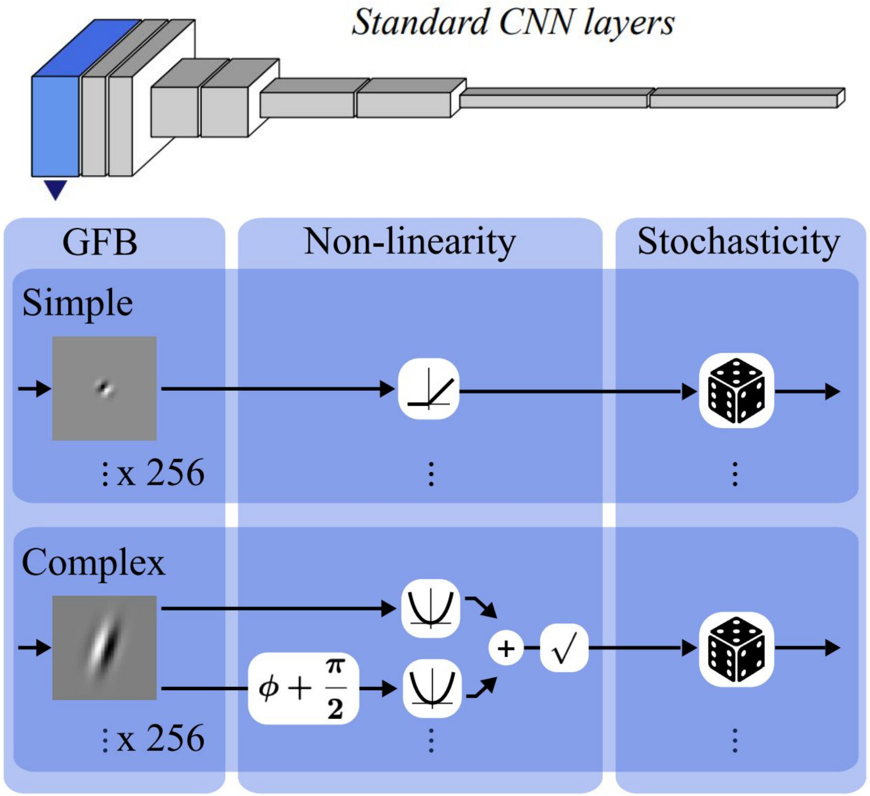
- Fixed-weight front-end block

Related Work: VOneNets[7] model V1 in front of a CNN back-end



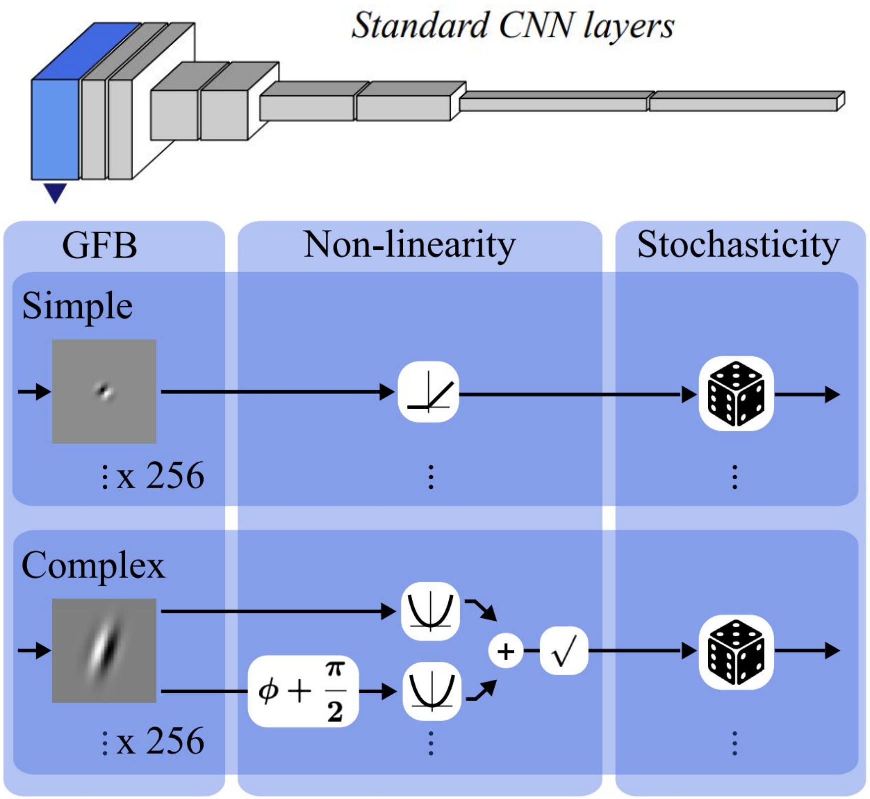
- Fixed-weight front-end block
- Based on Linear-Nonlinear-Poisson model

Related Work: VOneNets[7] model V1 in front of a CNN back-end



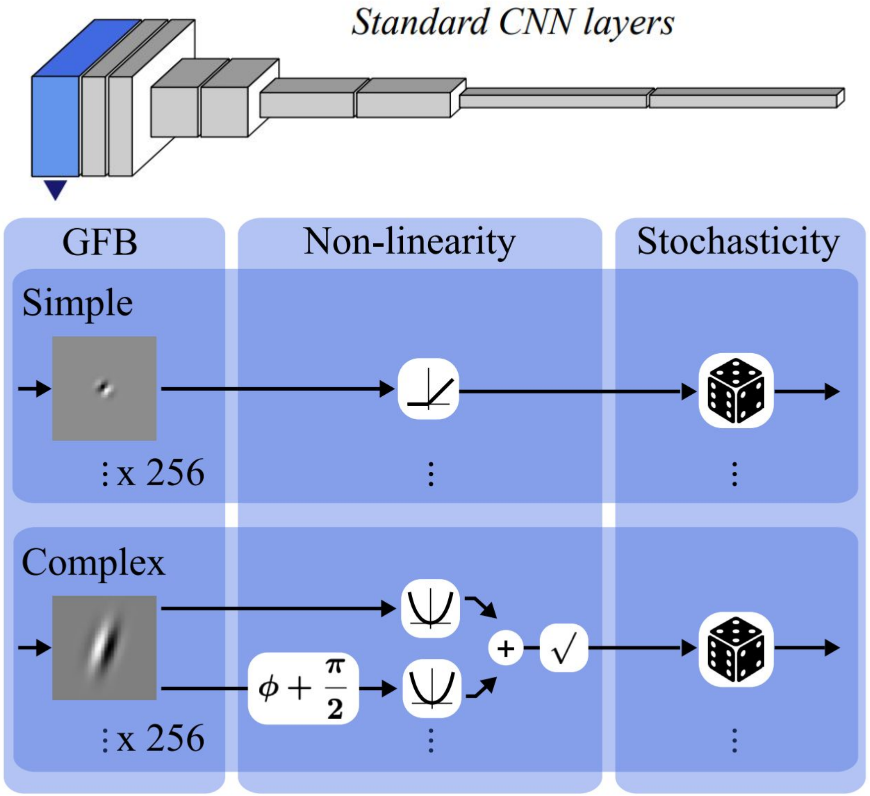
- Fixed-weight front-end block
- Based on Linear-Nonlinear-Poisson model
- Biologically-constrained Gabor filter bank

Related Work: VOneNets[7] model V1 in front of a CNN back-end



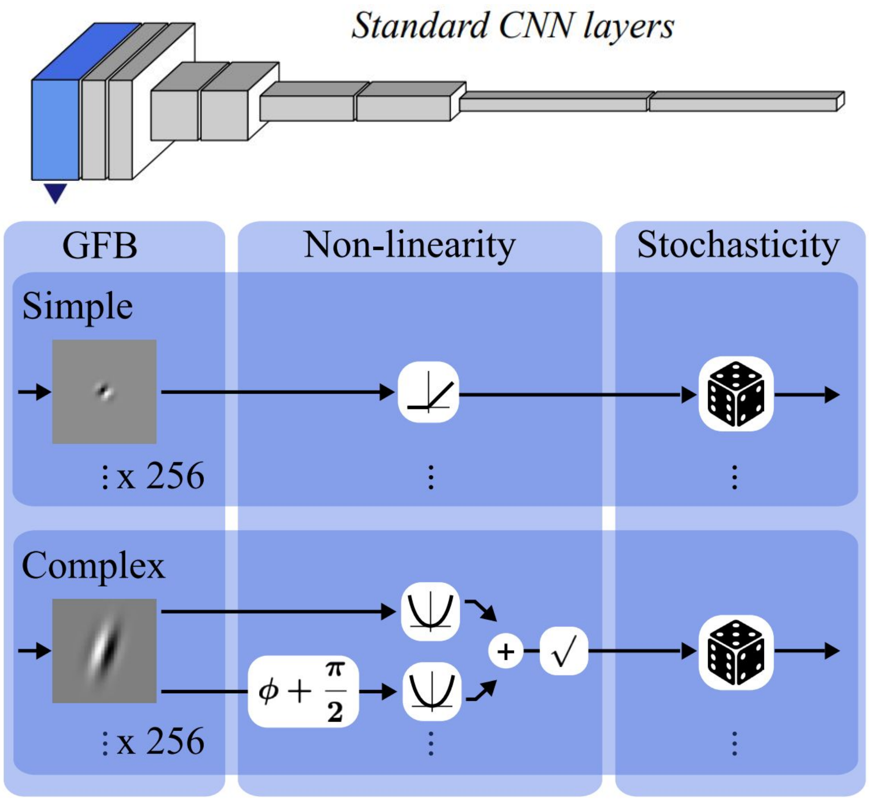
- Fixed-weight front-end block
- Based on Linear-Nonlinear-Poisson model
- Biologically-constrained Gabor filter bank
- Simple and complex cell nonlinearities

Related Work: VOneNets[7] model V1 in front of a CNN back-end



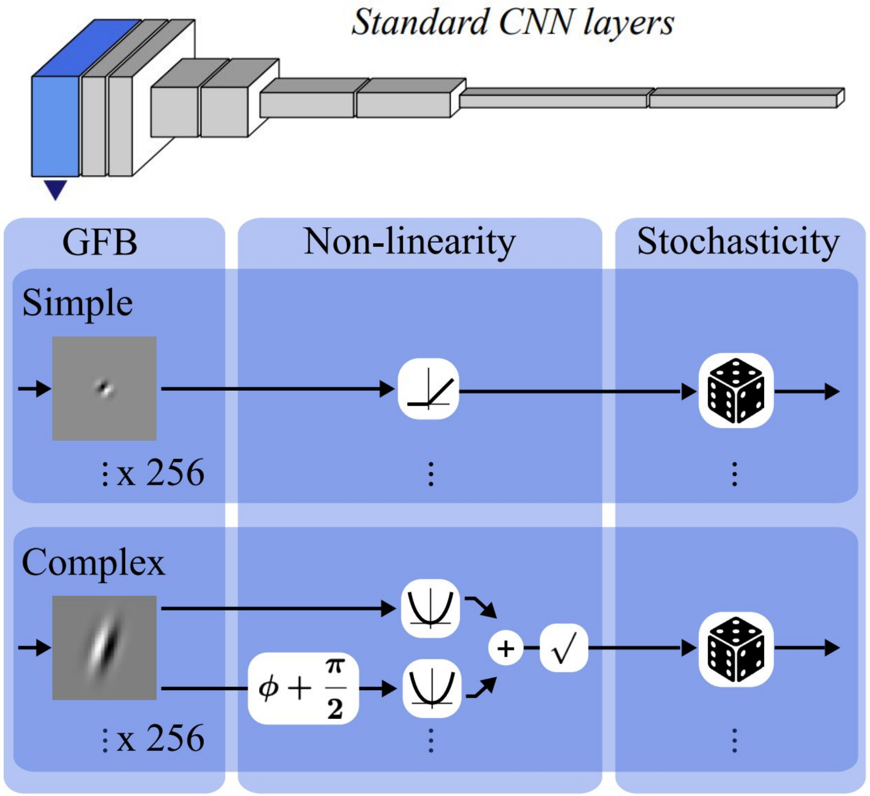
- Fixed-weight front-end block
- Based on Linear-Nonlinear-Poisson model
- Biologically-constrained Gabor filter bank
- Simple and complex cell nonlinearities
- V1 neuronal noise generator

Related Work: VOneNets[7] model V1 in front of a CNN back-end



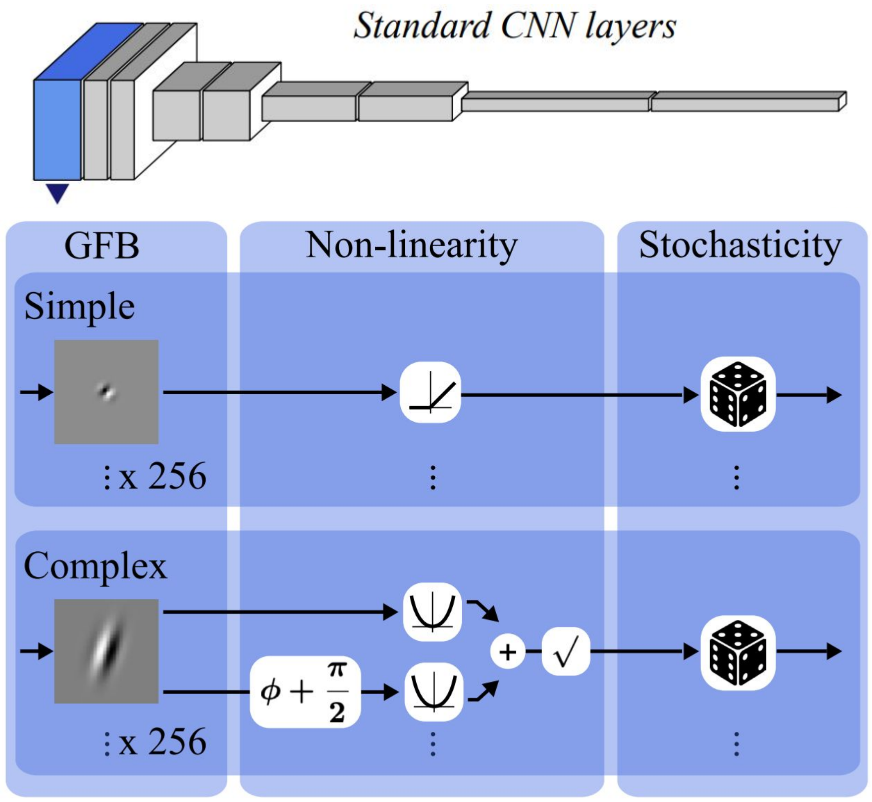
- Fixed-weight front-end block
- Based on Linear-Nonlinear-Poisson model
- Biologically-constrained Gabor filter bank
- Simple and complex cell nonlinearities
- V1 neuronal noise generator
- Mapped to V1 neuronal populations

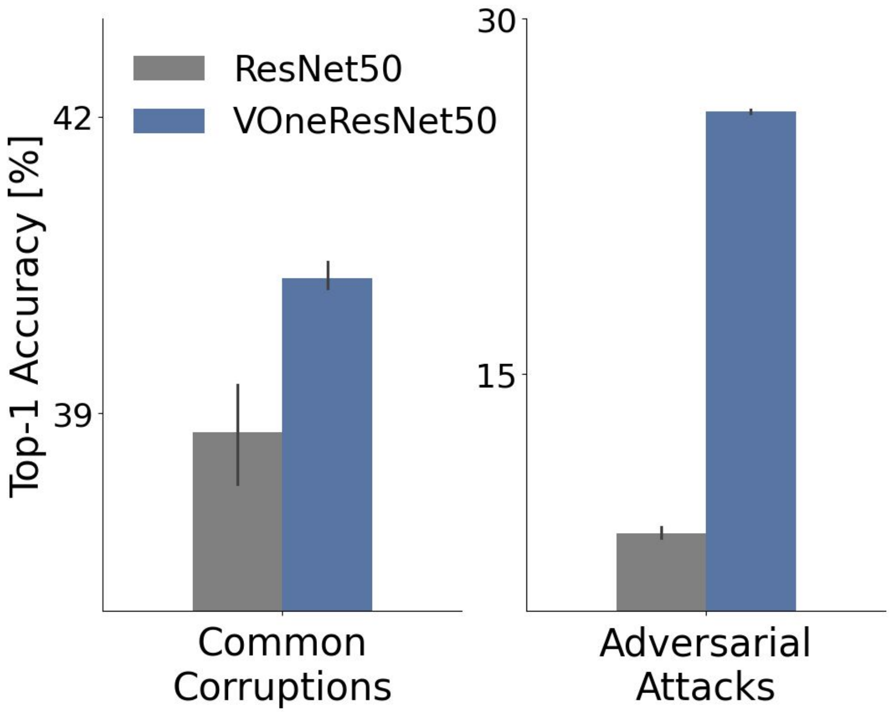
Related Work: VOneNets[7] model V1 in front of a CNN back-end



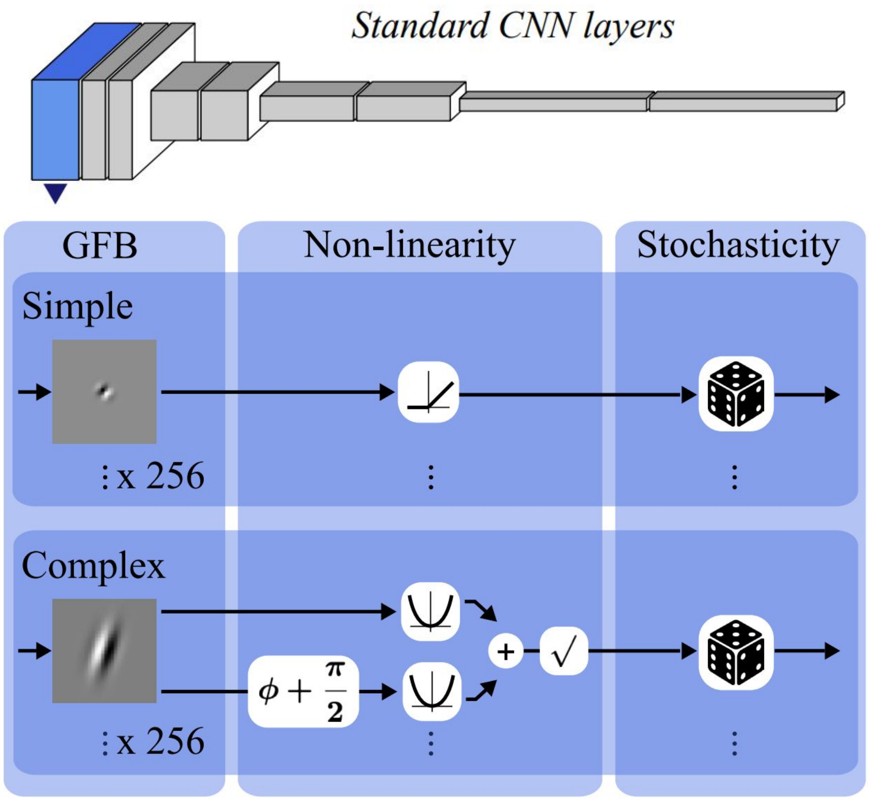
- Fixed-weight front-end block
- Based on Linear-Nonlinear-Poisson model
- Biologically-constrained Gabor filter bank
- Simple and complex cell nonlinearities
- V1 neuronal noise generator
- Mapped to V1 neuronal populations
- Adaptable to any back-end architecture

Related Work: VOneNets[7] model V1 in front of a CNN back-end



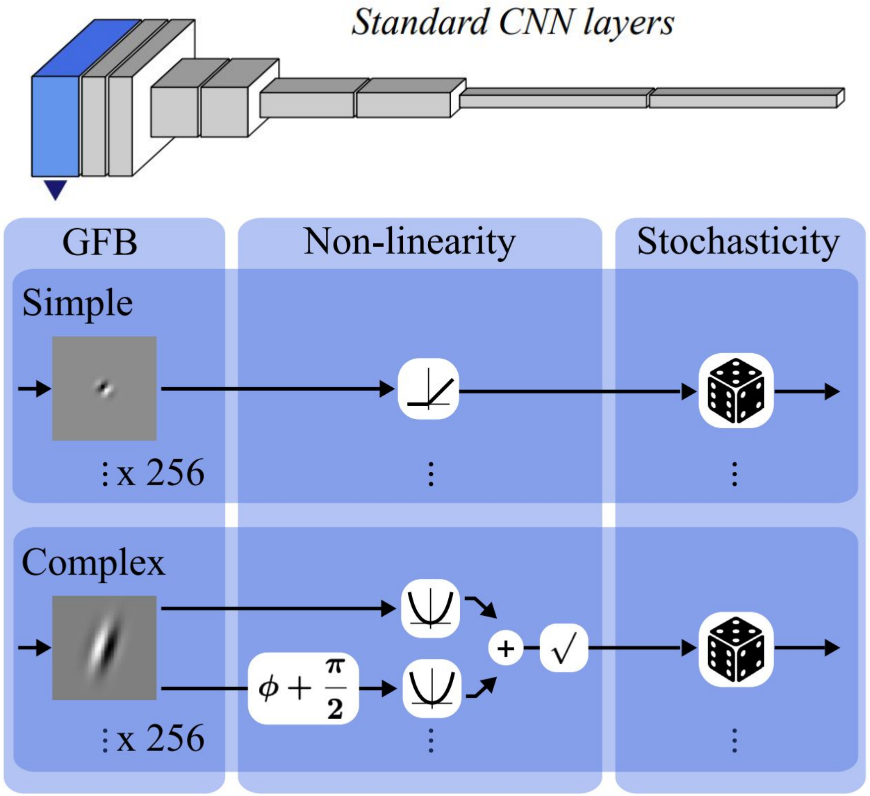


Related Work: VOneNets[7] model V1 in front of a CNN back-end



- Single-stage linear-nonlinear model with no extraclassical properties

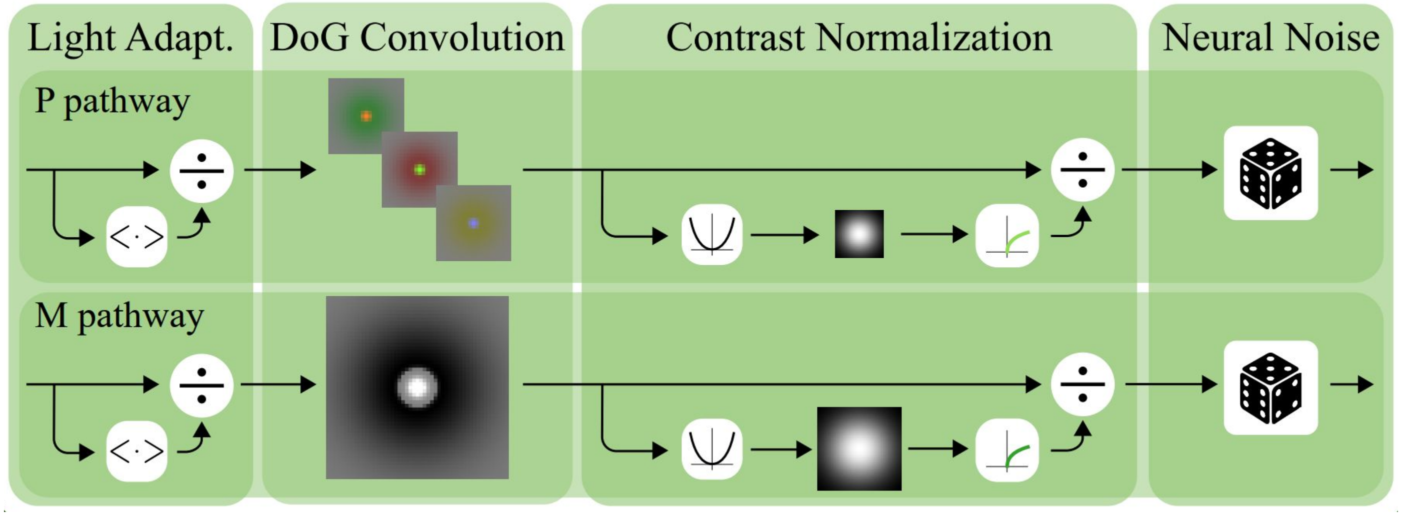
Related Work: VOneNets[7] model V1 in front of a CNN back-end



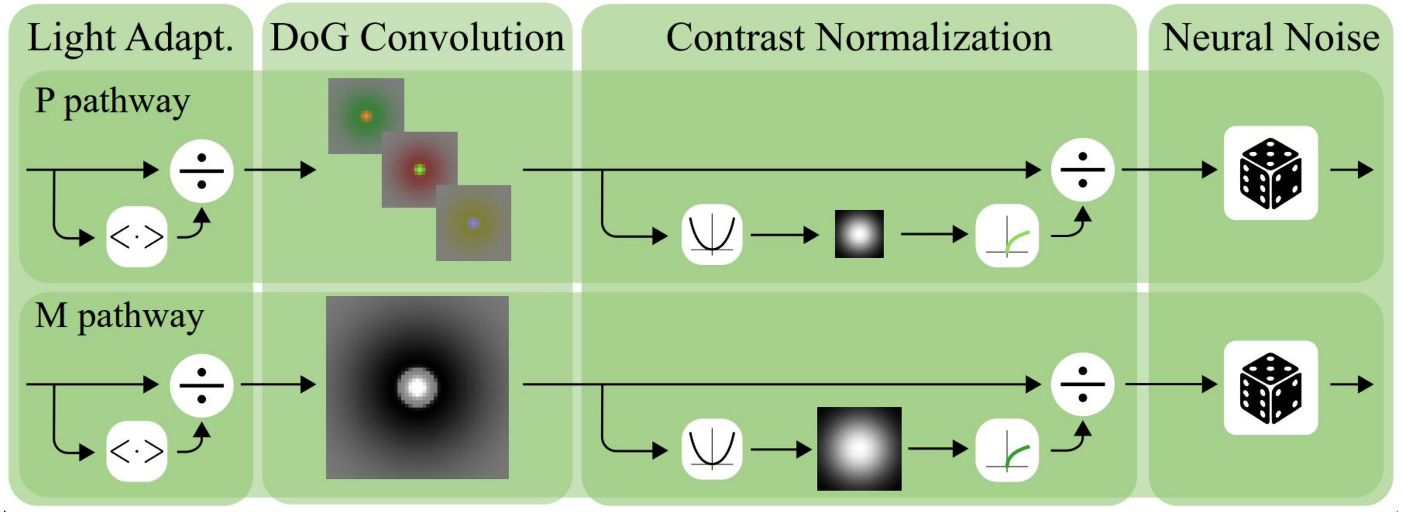
- Single-stage linear-nonlinear model with no extraclassical properties
- No explicit subcortical modeling (retina and LGN)

Can the explicit modeling of the retina and the LGN further improve model robustness?

Methods: the SubcorticalBlock models the retina and the LGN

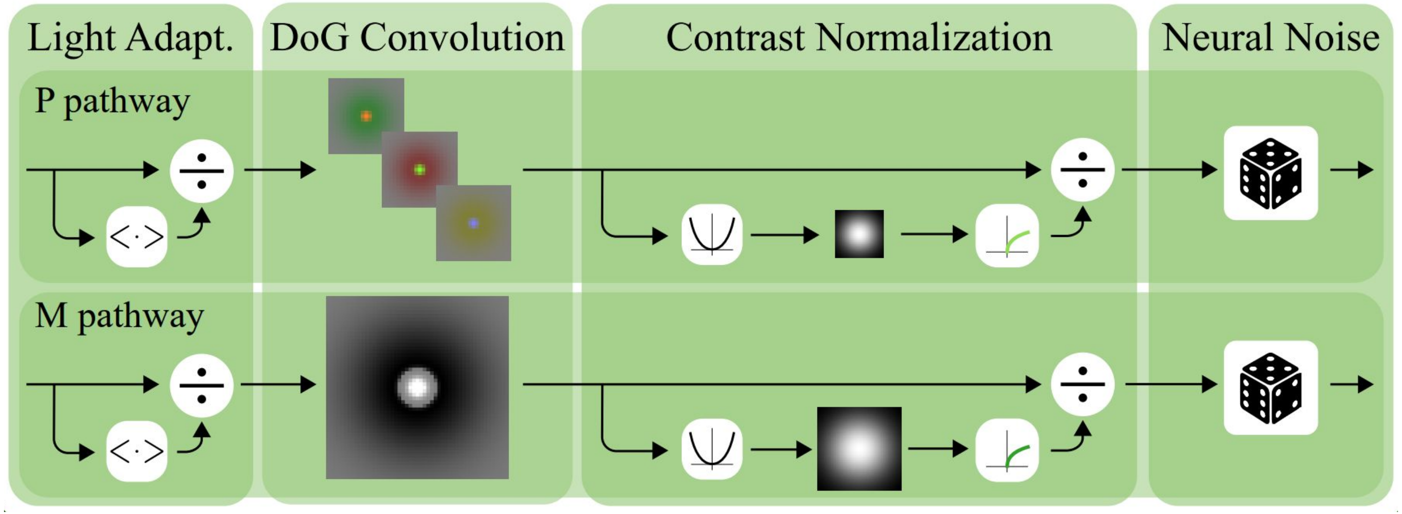


Methods: the SubcorticalBlock models the retina and the LGN



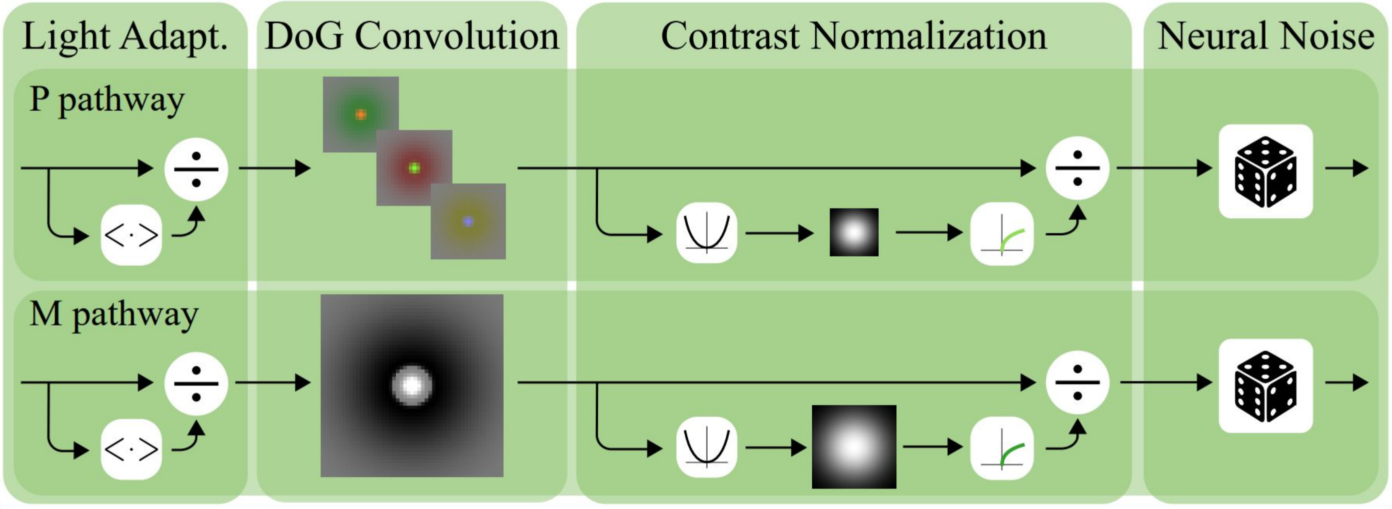
- Cascading linear-nonlinear model

Methods: the SubcorticalBlock models the retina and the LGN



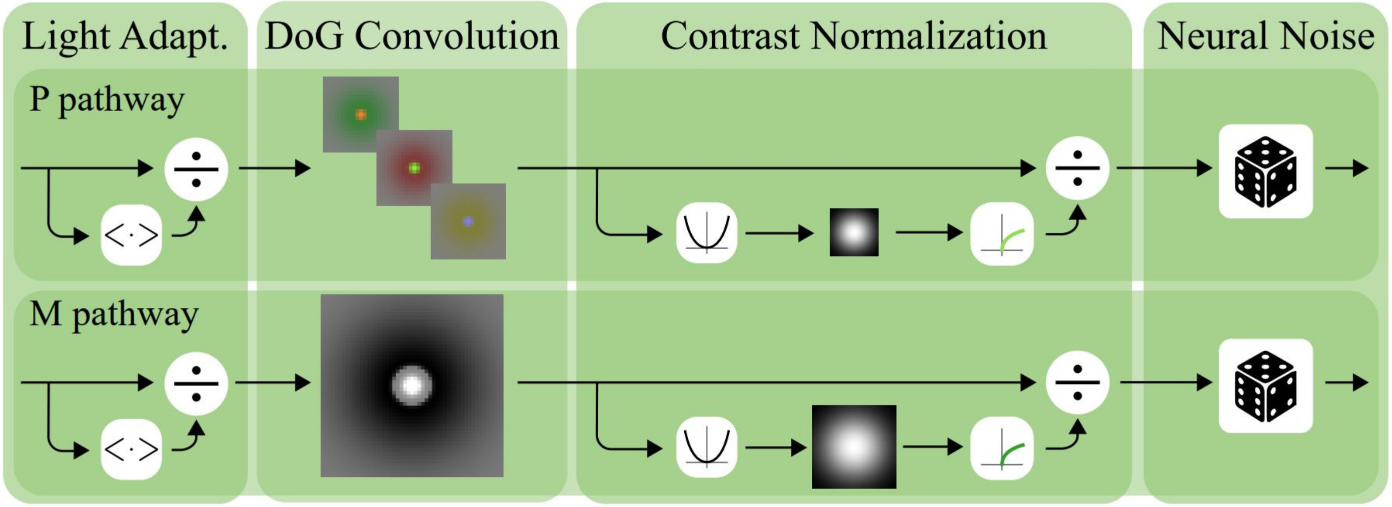
- Cascading linear-nonlinear model
- Separate pathways for P and M cells

Methods: the SubcorticalBlock models the retina and the LGN



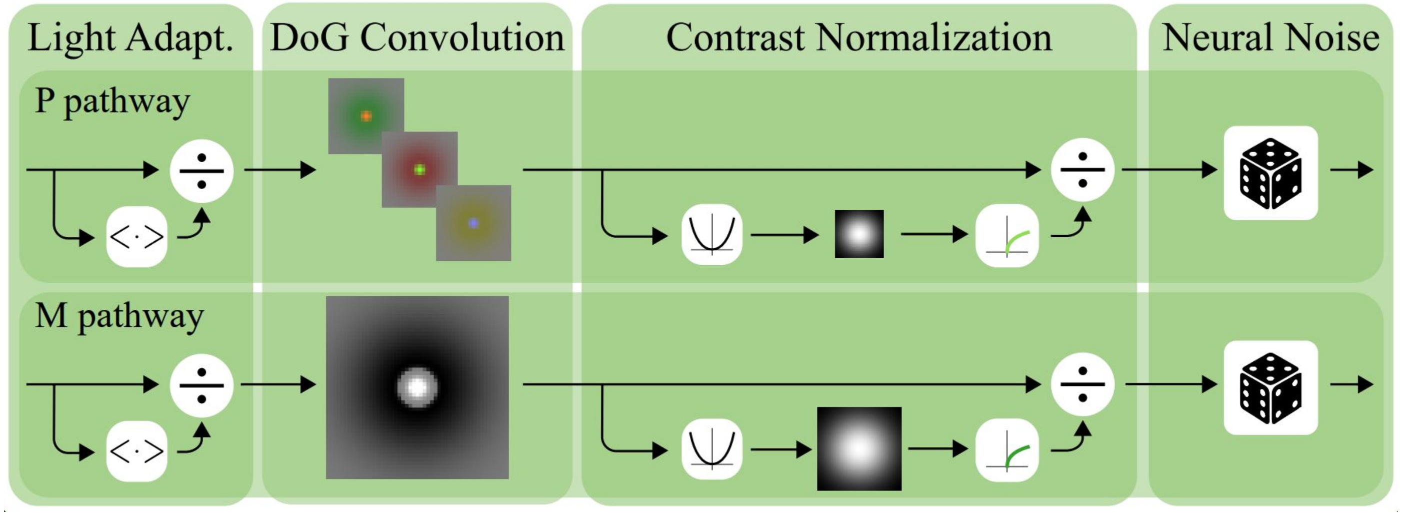
- Cascading linear-nonlinear model
- Separate pathways for P and M cells
- Spatial summation over the RF

Methods: the SubcorticalBlock models the retina and the LGN



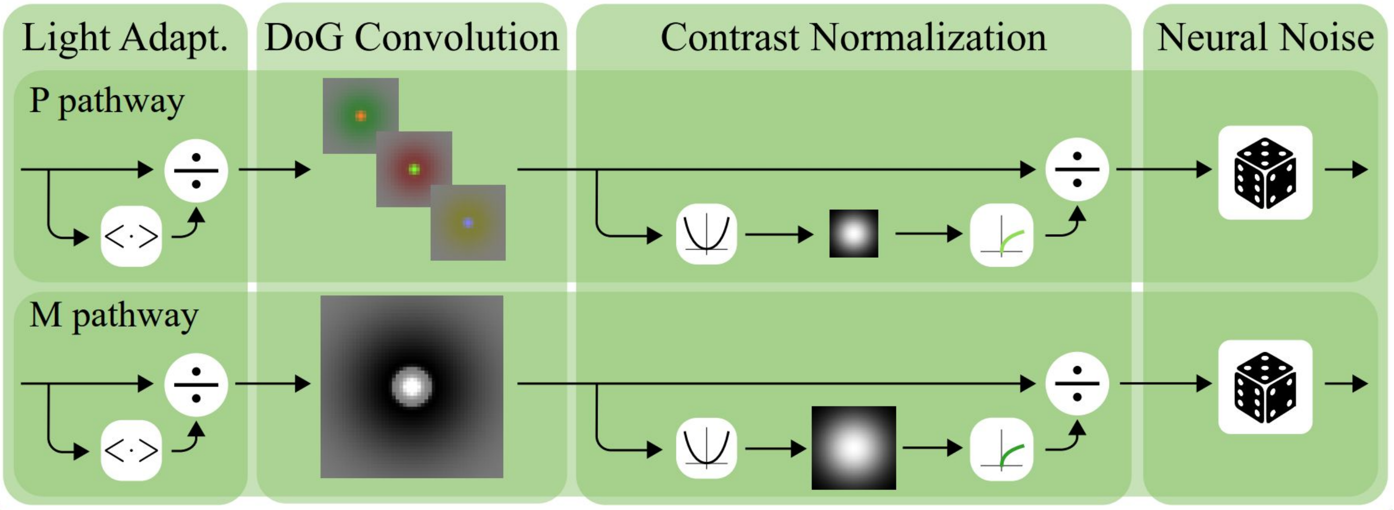
- Cascading linear-nonlinear model
- Separate pathways for P and M cells
- Spatial summation over the RF
- Biological color vision

Methods: the SubcorticalBlock models the retina and the LGN



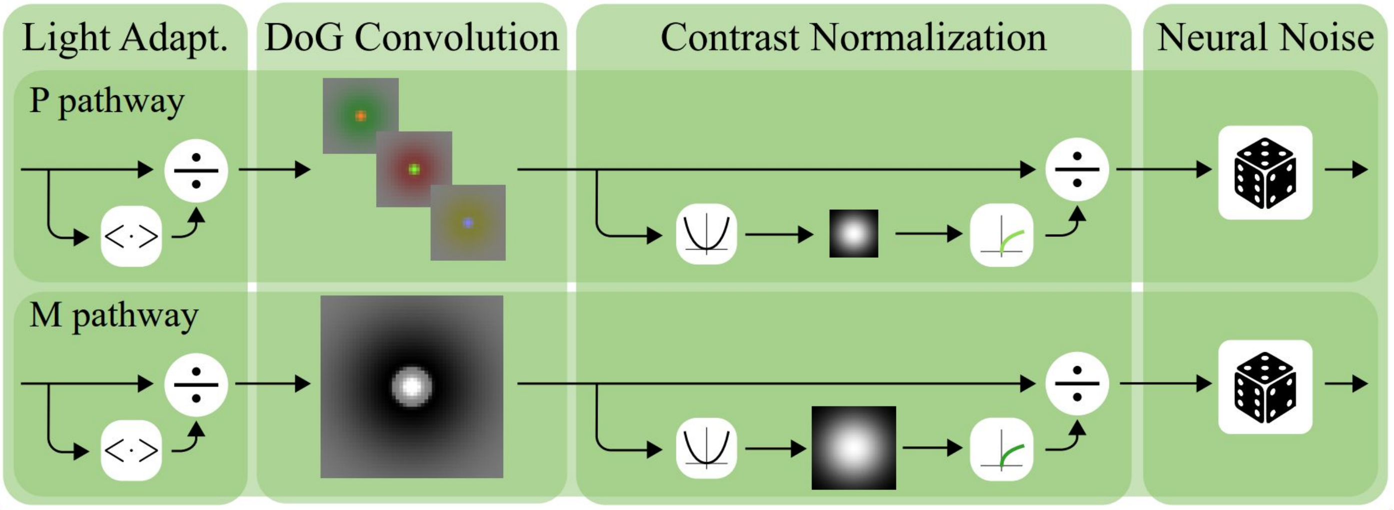
- Cascading linear-nonlinear model
- Separate pathways for P and M cells
- Spatial summation over the RF
- Biological color vision
- Luminance and contrast normalization

Methods: the SubcorticalBlock models the retina and the LGN



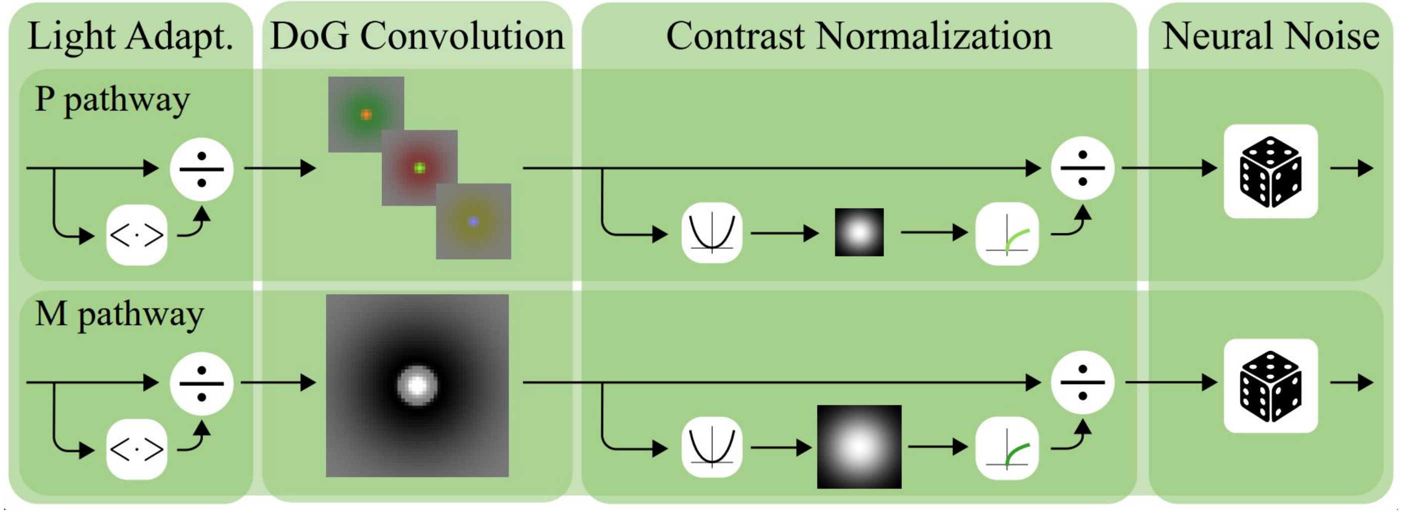
- Cascading linear-nonlinear model
- Separate pathways for P and M cells
- Spatial summation over the RF
- Biological color vision
- Luminance and contrast normalization
- Subcortical noise generator

Methods: the SubcorticalBlock models the retina and the LGN

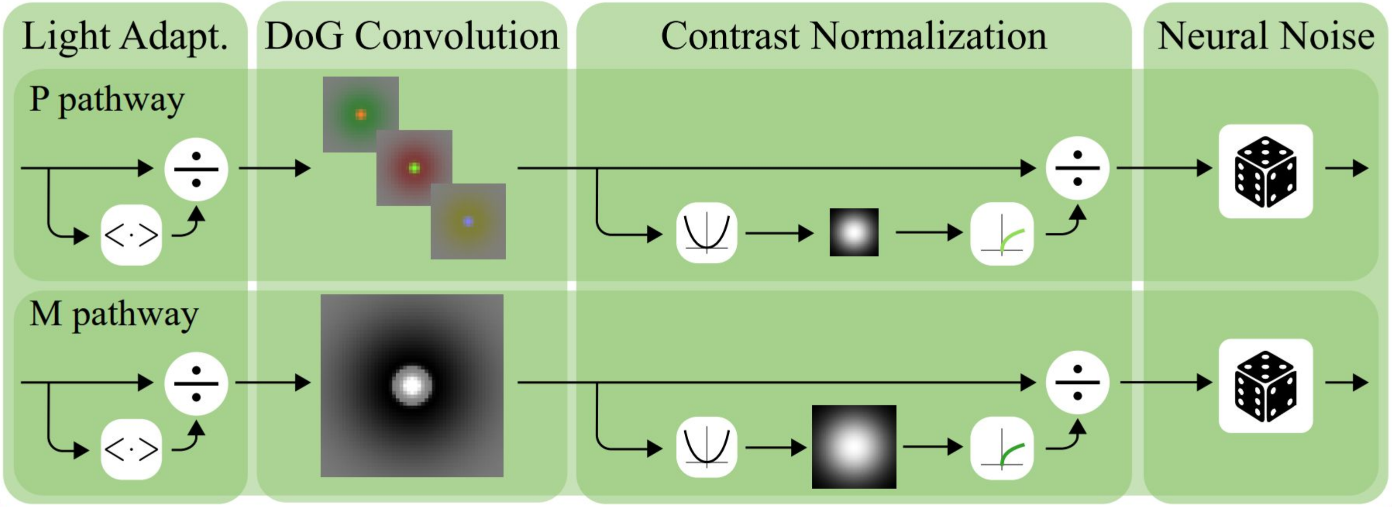


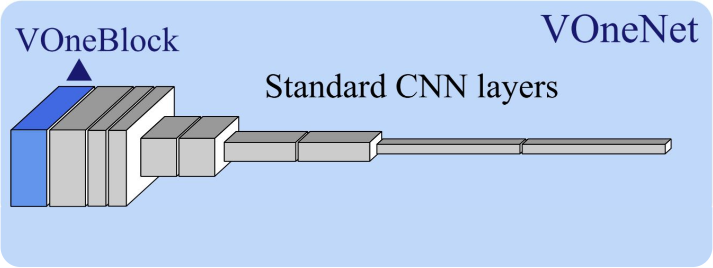
- Cascading linear-nonlinear model
- Separate pathways for P and M cells
- Spatial summation over the RF
- Biological color vision
- Luminance and contrast normalization
- Subcortical noise generator
- Parameters tuned to mimic subcortical response properties

Methods: the SubcorticalBlock models the retina and the LGN

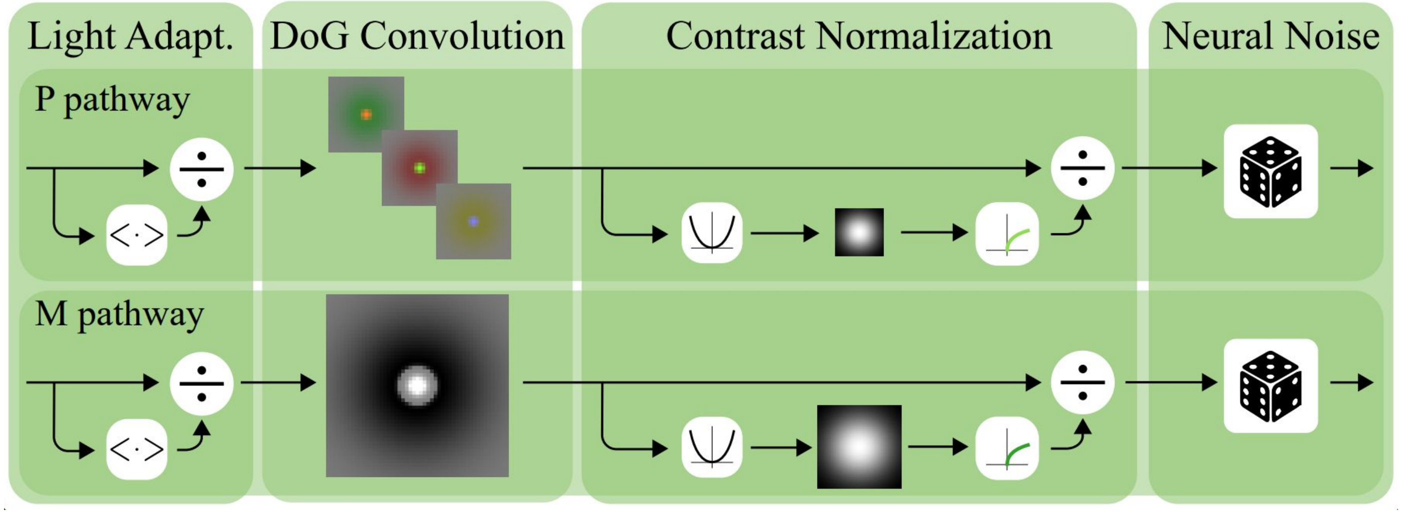


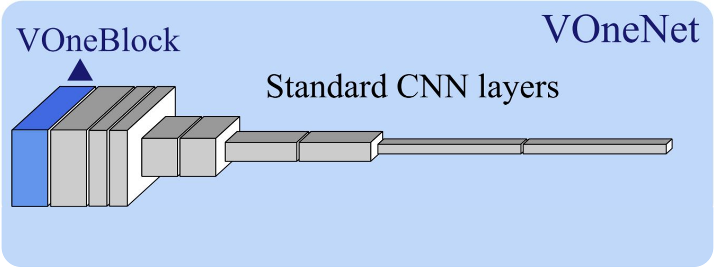
Methods: the SubcorticalBlock models the retina and the LGN

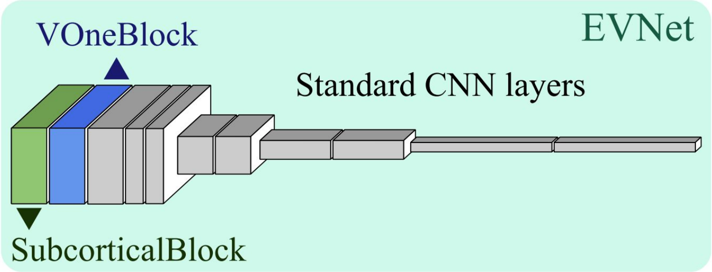




Methods: the SubcorticalBlock models the retina and the LGN

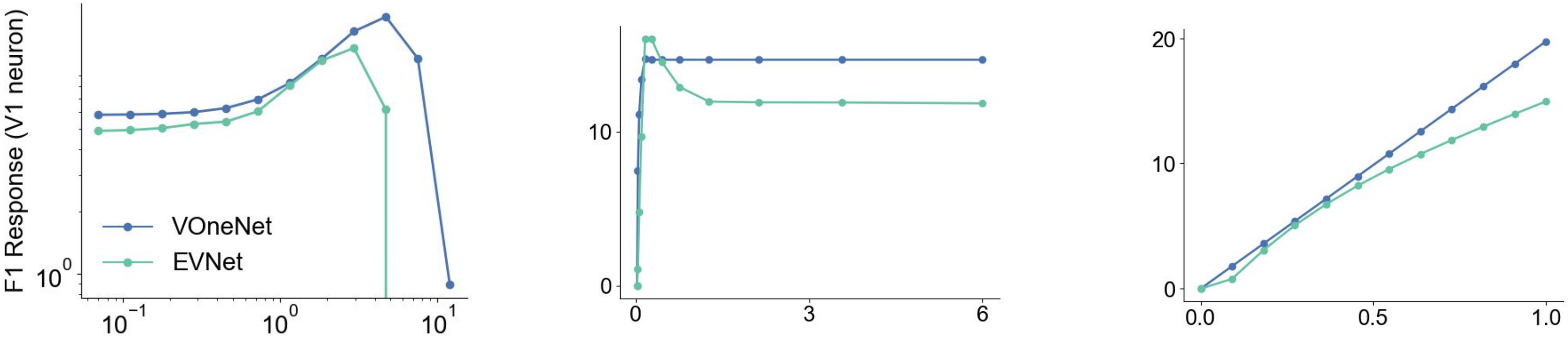




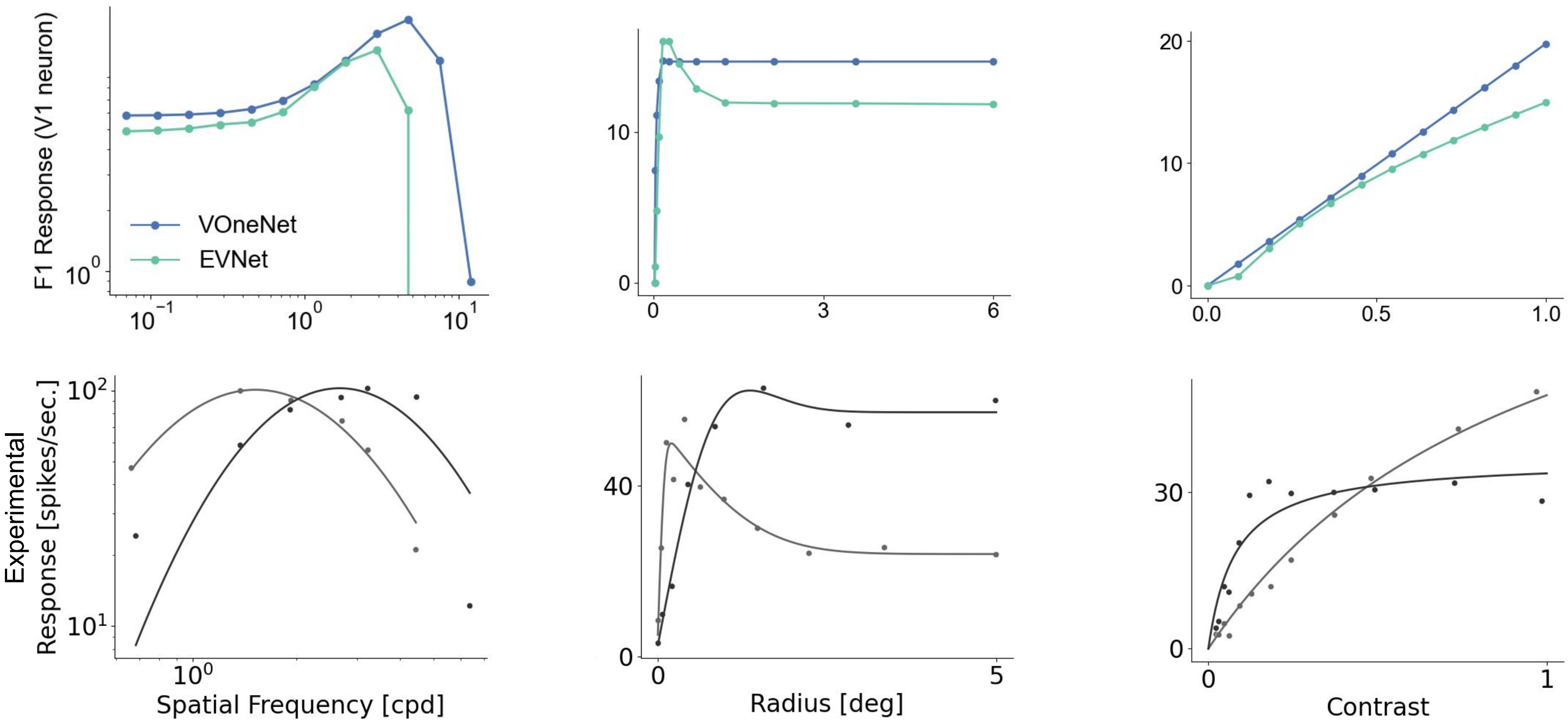


Results: EVNets mimic hallmark V1 response properties

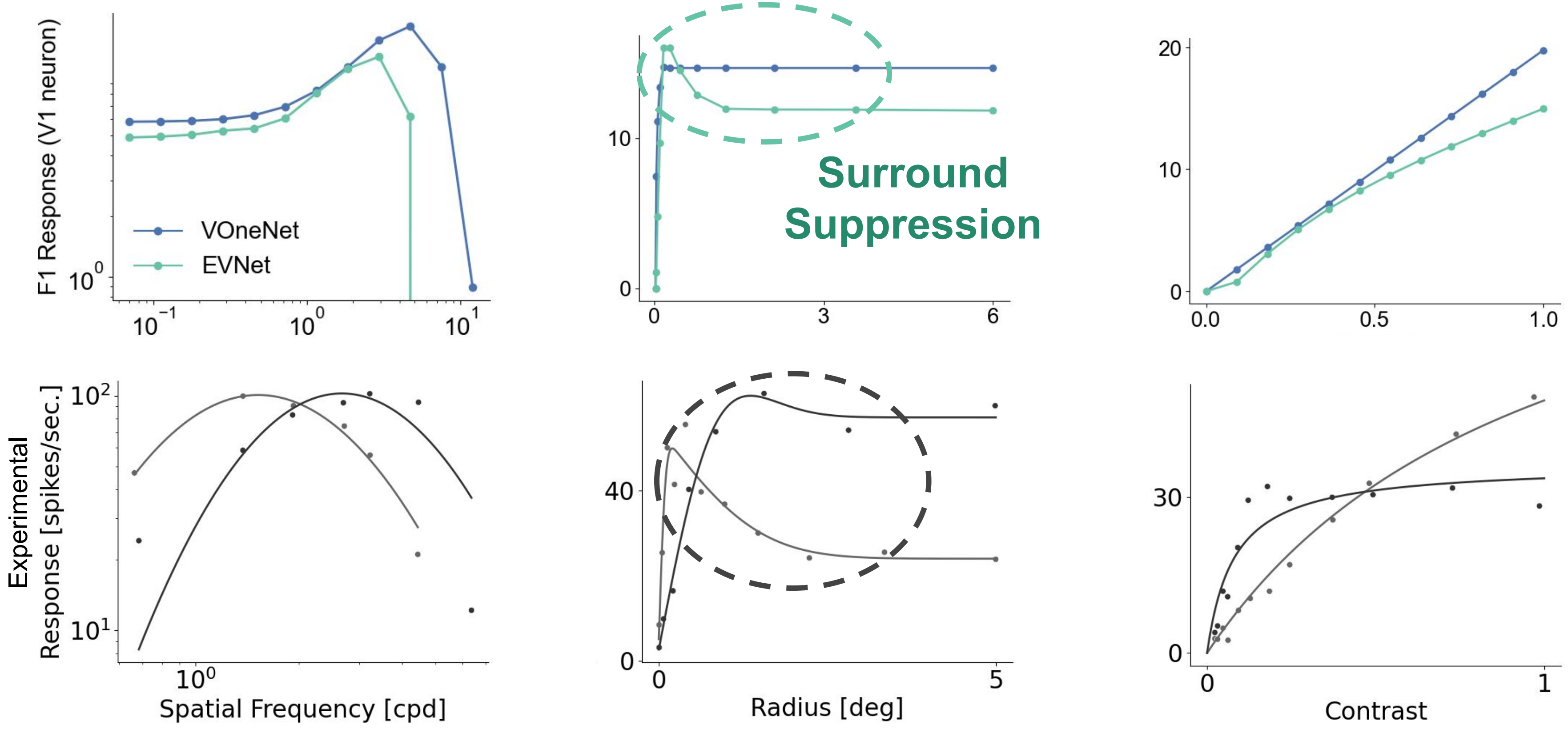
Results: EVNets mimic hallmark V1 response properties



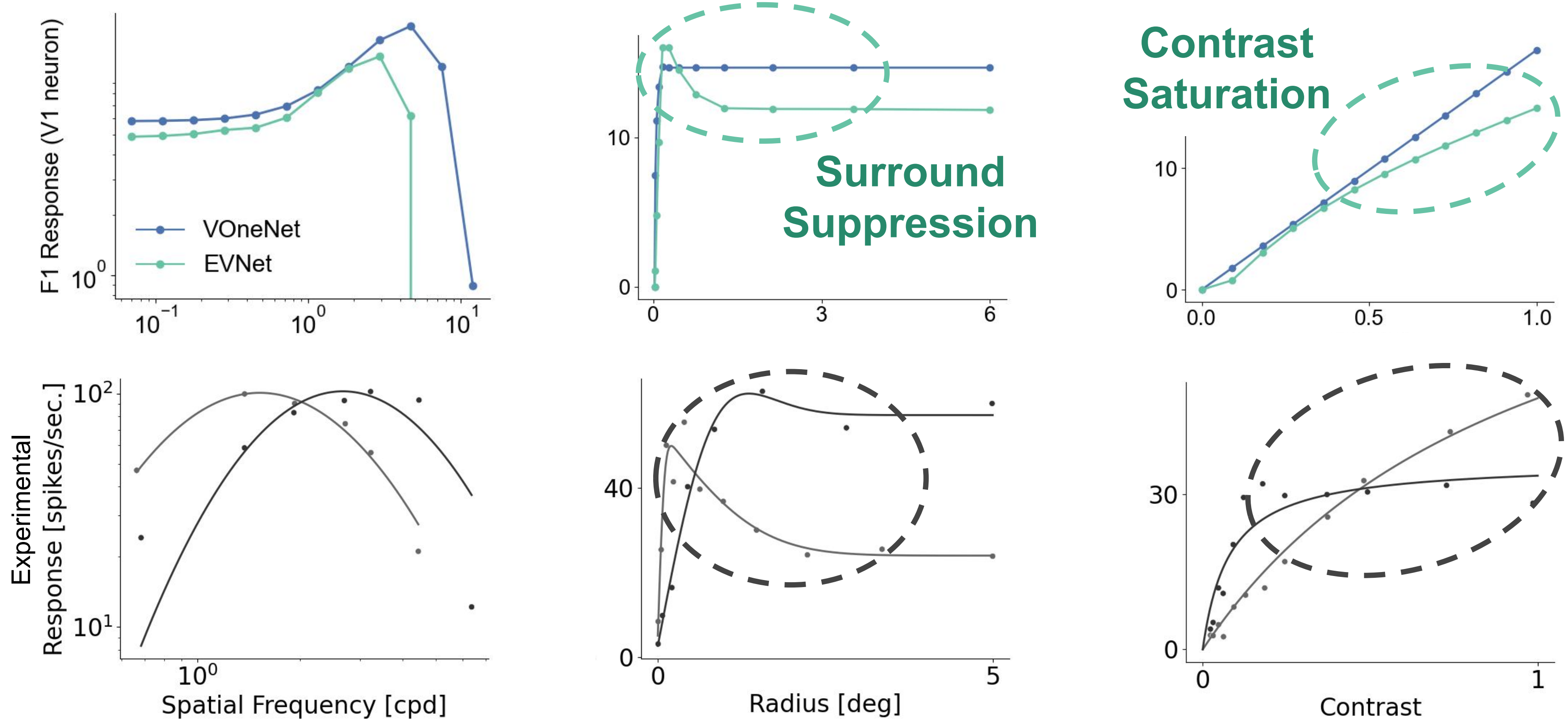
Results: EVNets mimic hallmark V1 response properties



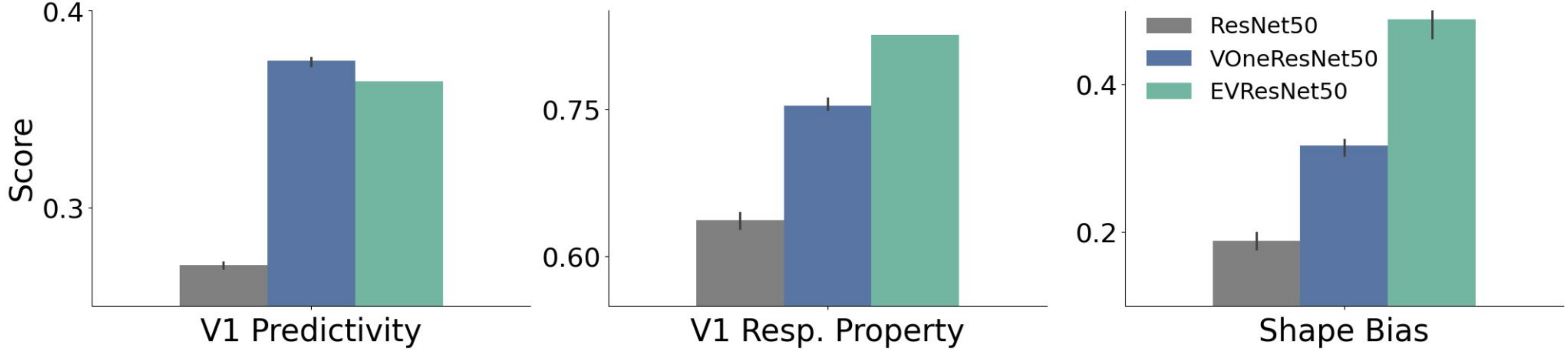
Results: EVNets mimic hallmark V1 response properties



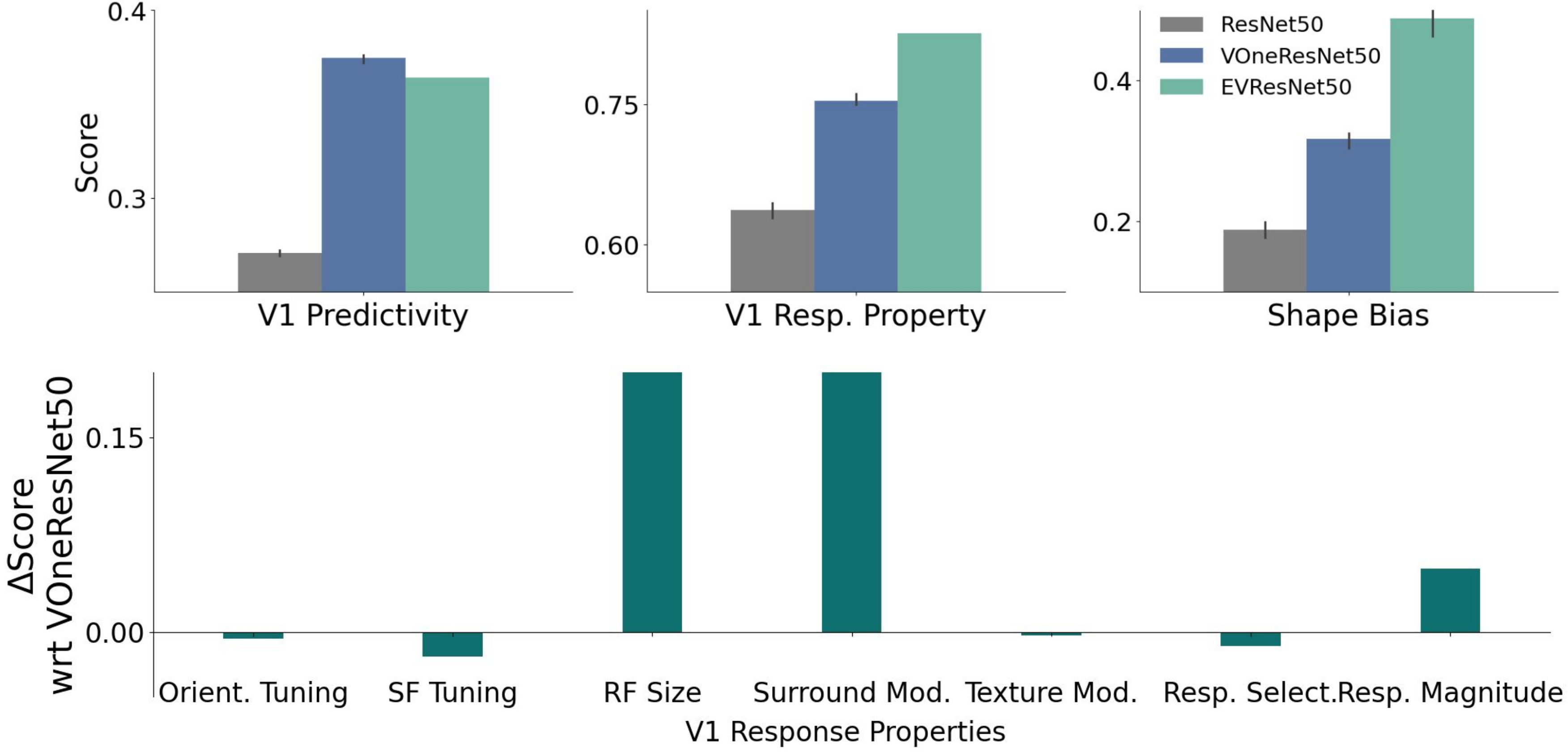
Results: EVNets mimic hallmark V1 response properties



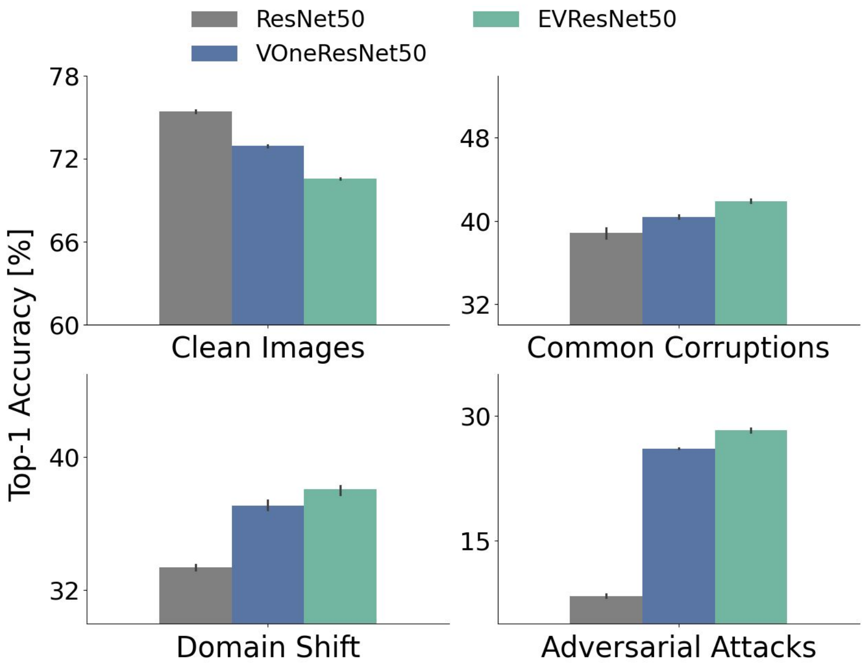
Results: EVNets improve primate vision alignment



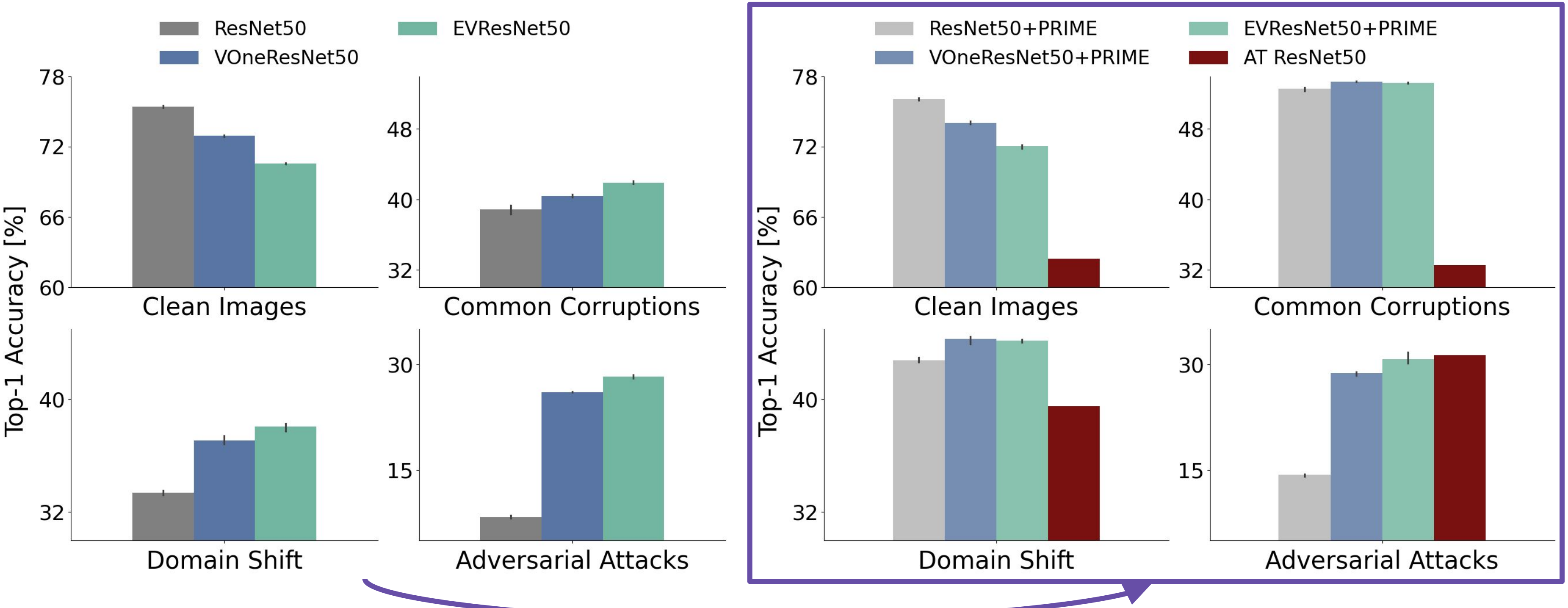
Results: EVNets improve primate vision alignment



Results: EVNets improve perturbation robustness and provide cumulative gains with PRIME[8] augmentation



Results: EVNets improve perturbation robustness and provide cumulative gains with PRIME[8] augmentation



+ PRIME Data Augmentation

Conclusion

Conclusion

- EVNets improve neural and behavioral alignment with primate vision

Conclusion

- EVNets improve neural and behavioral alignment with primate vision
- EVNets improve robustness across a diverse battery of evaluations, including adversarial attacks, common image corruptions, and domain shifts

Conclusion

- EVNets improve neural and behavioral alignment with primate vision
- EVNets improve robustness across a diverse battery of evaluations, including adversarial attacks, common image corruptions, and domain shifts
- EVNets trained with a state-of-the-art data augmentation technique yield additive improvements in robustness

Conclusion

- EVNets improve neural and behavioral alignment with primate vision
- EVNets improve robustness across a diverse battery of evaluations, including adversarial attacks, common image corruptions, and domain shifts
- EVNets trained with a state-of-the-art data augmentation technique yield additive improvements in robustness

Limitations

Conclusion

- EVNets improve neural and behavioral alignment with primate vision
- EVNets improve robustness across a diverse battery of evaluations, including adversarial attacks, common image corruptions, and domain shifts
- EVNets trained with a state-of-the-art data augmentation technique yield additive improvements in robustness

Limitations

- EVNets reinforce the trade-off between robustness and clean performance

Conclusion

- EVNets improve neural and behavioral alignment with primate vision
- EVNets improve robustness across a diverse battery of evaluations, including adversarial attacks, common image corruptions, and domain shifts
- EVNets trained with a state-of-the-art data augmentation technique yield additive improvements in robustness

Limitations

- EVNets reinforce the trade-off between robustness and clean performance
- The SubcorticalBlock only capture "mean" LGN response

Conclusion

- EVNets improve neural and behavioral alignment with primate vision
- EVNets improve robustness across a diverse battery of evaluations, including adversarial attacks, common image corruptions, and domain shifts
- EVNets trained with a state-of-the-art data augmentation technique yield additive improvements in robustness

Limitations

- EVNets reinforce the trade-off between robustness and clean performance
- The SubcorticalBlock only capture “mean” LGN response
- EVNets do not improve alignment across all neural benchmarks

Conclusion

- EVNets improve neural and behavioral alignment with primate vision
- EVNets improve robustness across a diverse battery of evaluations, including adversarial attacks, common image corruptions, and domain shifts
- EVNets trained with a state-of-the-art data augmentation technique yield additive improvements in robustness

Limitations

- EVNets reinforce the trade-off between robustness and clean performance
- The SubcorticalBlock only capture “mean” LGN response
- EVNets do not improve alignment across all neural benchmarks

Future Work

Conclusion

- EVNets improve neural and behavioral alignment with primate vision
- EVNets improve robustness across a diverse battery of evaluations, including adversarial attacks, common image corruptions, and domain shifts
- EVNets trained with a state-of-the-art data augmentation technique yield additive improvements in robustness

Limitations

- EVNets reinforce the trade-off between robustness and clean performance
- The SubcorticalBlock only capture “mean” LGN response
- EVNets do not improve alignment across all neural benchmarks

Future Work

- Adapt VOneBlock GFB to accommodate subcortical processing

Conclusion

- EVNets improve neural and behavioral alignment with primate vision
- EVNets improve robustness across a diverse battery of evaluations, including adversarial attacks, common image corruptions, and domain shifts
- EVNets trained with a state-of-the-art data augmentation technique yield additive improvements in robustness

Limitations

- EVNets reinforce the trade-off between robustness and clean performance
- The SubcorticalBlock only capture "mean" LGN response
- EVNets do not improve alignment across all neural benchmarks

Future Work

- Adapt VOneBlock GFB to accommodate subcortical processing
- Add normalization to the VOneBlock

Conclusion

- EVNets improve neural and behavioral alignment with primate vision
- EVNets improve robustness across a diverse battery of evaluations, including adversarial attacks, common image corruptions, and domain shifts
- EVNets trained with a state-of-the-art data augmentation technique yield additive improvements in robustness

Limitations

- EVNets reinforce the trade-off between robustness and clean performance
- The SubcorticalBlock only capture “mean” LGN response
- EVNets do not improve alignment across all neural benchmarks

Future Work

- Adapt VOneBlock GFB to accommodate subcortical processing
- Add normalization to the VOneBlock
- Allow task-driven fine tuning of neuro-inspired weights

References

[1] Dan Hendrycks and Thomas Dietterich. “Benchmarking Neural Network Robustness to Common Corruptions and Perturbations”. In: International Conference on Learning Representations. 2019.
[2] Tiago Salvador and Adam M Oberman. “ImageNet-Cartoon and ImageNet-Drawing: two domain shift datasets for ImageNet”. In: ICML 2022 Shift Happens Workshop. 2022.
[3] Dan Hendrycks, Steven Basart, Norman Mu, Saurav Kadavath, Frank Wang, Evan Dorundo, Rahul Desai, Tyler Zhu, Samyak Parajuli, Mike Guo, Dawn Song, Jacob Steinhardt, and Justin Gilmer. “The Many Faces of Robustness: A Critical Analysis of Out-of-Distribution Generalization”. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). 2021, pp. 8320–8329.
[4] Haohan Wang, Songwei Ge, Zachary Lipton, and Eric P Xing. “Learning Robust Global Representations by Penalizing Local Predictive Power”. In: Advances in Neural Information Processing Systems. 2019, pp. 10506–10518.
[5] Robert Geirhos, Kantharaju Narayanappa, Benjamin Mitzkus, Tizian Thieringer, Matthias Bethge, Felix A. Wichmann, and Wieland Brendel. “Partial success in closing the gap between human and machine vision”. In: Advances in Neural Information Processing Systems. Ed. by A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan. 2021.
[6] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, and Wieland Brendel. “ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness.” In: International Conference on Learning Representations. 2019.
[7] Joel Dapello, Tiago Marques, Martin Schrimpf, Franziska Geiger, David Cox, and James J DiCarlo. “Simulating a Primary Visual Cortex at the Front of CNNs Improves Robustness to Image Perturbations”. In: Advances in Neural Information Processing Systems. Ed. by H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin. Vol. 33. Curran Associates, Inc., 2020, pp. 13073–13087.
[8] Apostolos Modas, Rahul Rade, Guillermo Ortiz-Jiménez, Seyed-Mohsen Moosavi-Dezfooli, and Pascal Frossard. “PRIME: A Few Primitives Can Boost Robustness to Common Corruptions”. In: Computer Vision – ECCV 2022. Ed. by Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner. Cham: Springer Nature Switzerland, 2022, pp. 623–640. ISBN: 978-3-031-19806-9.