RTNet: An image computable model of human choice, response time, and confidence

Medha Shekhar (medha@gatech.edu)
School of Psychology, 654 Cherry Street NW
Atlanta, Georgia, 30332 USA

Farshad Rafiei (farshadrafiei3@gmail.com)
School of Psychology, 654 Cherry Street NW
Atlanta, Georgia, 30332 USA

Dobromir Rahnev (rahnev@psych.gatech.edu)
School of Psychology, 654 Cherry Street NW
Atlanta, Georgia, 30332 USA

Abstract:

Convolutional neural networks show promise as models of biological vision. However, unlike humans, they are deterministic and use equal number of computations for easy and difficult stimuli, which limits their applicability as models of human behavior. Here we develop a new neural network, RTNet, that generates stochastic decisions and human-like response time (RT) distributions. Through comprehensive tests, we show that RTNet reproduces all foundational features of human accuracy, RT, and confidence and does so better than all current alternative models. We further test RTNet's ability to predict human behavior on novel images by collecting accuracy, RT, and confidence data from 60 human subjects performing a digit discrimination task. The responses produced by RTNet for individual novel images correlated with the same quantities produced by human subjects and these correlations were higher than those produced by all competing models. Overall, RTNet is a promising model of human response times that exhibits the critical signatures of perceptual decision making.

Keywords: Deep neural networks, reaction time, perceptual decision making, sequential sampling, confidence

Introduction

Traditional cognitive models of perceptual decisions (Ratcliff & McKoon, 2008) can account for the major features of human perceptual decision making, but do not operate on the level of images and are mostly limited to 2-choice tasks (Rahnev, 2020). On the other hand, convolutional neural networks (CNNs) can achieve human-level performance for novel images (Kriegeskorte, 2015; Kriegeskorte & Golan, 2019) and naturally handle multi-choice categorization tasks. However, unlike humans, traditional CNNs are both deterministic and static, thus always producing the same responses and response times for a given input.

Here we combine modern CNNs with traditional cognitive models to create a model that is image-computable, stochastic, and dynamic. The model, which we call RTNet, features a CNN with noisy weights that processes a given image several times using a different random sample of these weights in each processing step (**Figure 1A**). By sampling from noisy weight distributions, the network's units produce variable responses to the same input, which mimics the randomness of neural responses. After each processing step, RTNet accumulates the evidence or output corresponding to each choice until one of the choices reaches a predefined threshold. Thus, the model has a strong conceptual relationship to race models from the cognitive literature on decision-making (Ratcliff, 1978; Ratcliff & McKoon, 2008) and combines the image-computability of CNNs with traditional models of perception. We compare the behavior of RTNet to that of three other popular dynamic CNNs – Parallel Cascaded Network (CNet; Iuzzolino et al., 2021), a recurrent CNN (BLNet; Spoerer et al., 2020), and Multi-Scale Dense Network (MSDNet; Huang et al., 2017).

Results

We collected data from 60 human subjects who performed an 8-choice digit discrimination task with MNIST images embedded in noise (**Figure 1B**). The experiment was a 2 x 2 design with factors of task difficulty (low vs. high noise; **Figure 1C**) and speed pressure (speed vs. accuracy focus). Each condition consisted of 120 unique images, with each image being presented twice. Thus, each subject completed 960 trials in total. To improve the model correspondence with human data, we trained 60 instances of each model (by changing the random initialization before training) and analyzed the data produced by these 60 instances in equivalent manner to the 60 human subjects.

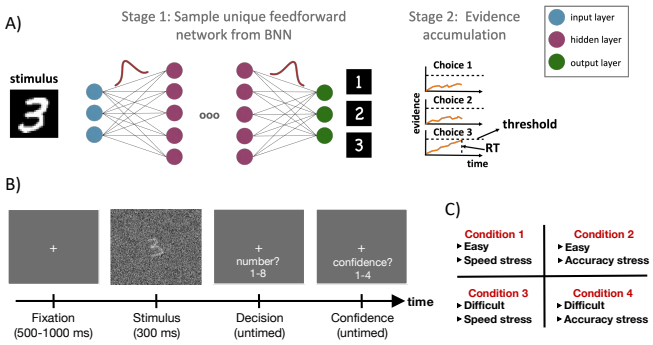

Figure 1: A) RTNet architecture. B) Task. C) The four experimental conditions.

Signatures of human perceptual decision making

We first examined six foundational signatures of human perceptual decision making: 1) Human decisions are stochastic i.e. the same stimulus can elicit different responses on different trials (**Figure 2A**), 2) increasing speed stress decreases accuracy and decreases RT (speed-accuracy trade off; Figure 2B-C), 3) more difficult decisions lead to reduced accuracy and longer RT (**Figure 2B-C**), 4) RT distributions are right-skewed,

and this skew increases with task difficulty (**Figure 2D-E**), 5) RT is lower for correct than for error trials (**Figure 2F**), and 6) confidence is higher for correct than for error trials (**Figure 2G**). We first confirmed that the signatures occur in the human data and then tested the models.

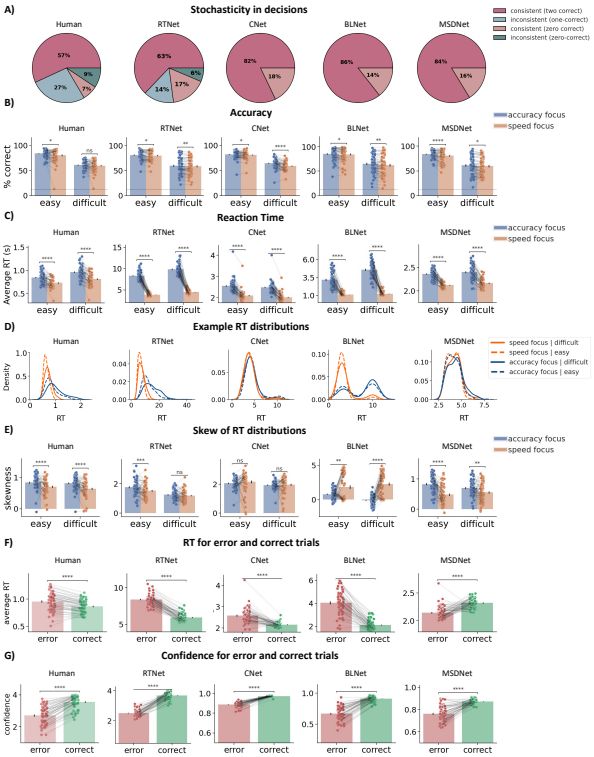

Figure 2: Signatures of human perceptual decisions as shown by our data (left) and the four models – RTNet, CNet, BLNet and MSDNet. RTNet is the only model that reproduces all six signatures of human decision making.

We found that RTNet was the only model that was able to predict all six of these features. Particularly, all the other models failed to capture the observed stochasticity of human decisions (**Figure 2A**), and the shape (**Figure 2D**) and skewness of RT distributions (**Figure 2E**). MSDNet additionally incorrectly predicted slower RTs for correct decisions (**Figure 2F**).

Model predictions of accuracy, RT, and confidence for individual images

Next, we tested whether the accuracy, RT, and confidence for unseen images produced by the networks predict the same quantities in humans. All models, except BLNet, predicted individual human accuracy, RT, and confidence much better than chance (all *p*'s < 0.0001). Critically, RTNet provided substantially better predictions than all other models (**Figure 3A**) for accuracy, RT, and confidence (all but one *p*'s < 0.0001). RTNet's predictions were within 62.5%, 79.6%, and 64.8% of the noise ceiling for accuracy, RT, and confidence, respectively (the noise ceiling was calculated as the average subject-to-group correlation in the human data). These numbers were substantially lower for CNet (16.1%, 20.3%, and 40.5%), BLNet (0%, 64.4%, and 54.1%), and MSDNet (16.1%, 50%, and 51.3%). We also explored how well the models compared to the ability of individual subjects to predict the group human data. We found that RTNet outperformed 73.3%, 100%, and 100% of individual human subjects in predicting the accuracy, RT, and confidence of the rest of group, respectively (all p's < 0.0001; Figure 3B). All other networks were worse than individual subjects in predicting group accuracy (p's < 0.0001). In addition, CNet and MSDNet were worse than individual subjects in predicting RT (p's < 0.0001). In sum, RTNet was the only network that outperformed most individual subjects in predicting all three measures of human performance (accuracy, RT, and confidence).

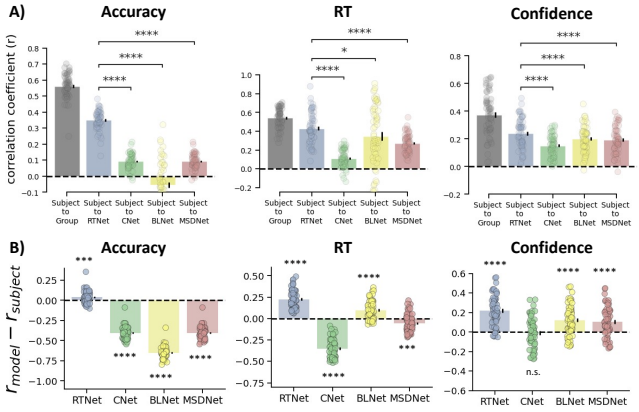

Figure 3: A) Model predictions of accuracy, RT and confidence for novel images and B) Comparing model predictions of group responses to individual subject predictions of group responses.

Conclusion

We developed a new neural network, RTNet, which exhibits the critical features of human perceptual decision making and predicts human accuracy, RT, and confidence on an image-by-image basis. The network provides a better model of human perceptual decisions than the current state-of-the-art networks for generating response times. RTNet thus represents an important step in the use of neural networks as models of human decisions.

Acknowledgments

This work was supported by the National Institute of Health (award: R01MH119189) and Office of Naval Research (award: N00014-20-1-2622). We thank Sashank Varma and Paul Verhaeghen for helpful suggestions about the analyses, as well as Ana Shin and Himanaga Sahithi Pandi for assistance with data collection.

References

Huang, G., Chen, D., Li, T., Wu, F., Van Der Maaten, L., & Weinberger, K. (2017). Multi-Scale Dense Networks for Resource Efficient Image Classification. *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*. https://arxiv.org/abs/1703.09844v5

Iuzzolino, M. L., Mozer, M. C., & Bengio, S. (2021). Improving Anytime Prediction with Parallel Cascaded Networks and a Temporal-Difference Loss. *Advances in Neural Information Processing Systems*, *33*, 27631–27644. https://arxiv.org/abs/2102.09808v4

Kriegeskorte, N. (2015). Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. *Http://Dx.Doi.Org/10.1146/Annurev-Vision-082114-035447*, *1*(1), 417–446. https://doi.org/10.1146/ANNUREV-VISION-082114-035447

Kriegeskorte, N., & Golan, T. (2019). Neural network models and deep learning. *Current Biology*, *29*(7), R231–R236. https://doi.org/10.1016/J.CUB.2019.02.034

Rahnev, D. (2020). Confidence in the Real World. *Trends in Cognitive Sciences*, *24*(8), 590–591. https://doi.org/10.1016/J.TICS.2020.05.005

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Spoerer, C. J., Kietzmann, T. C., Mehrer, J., Charest, I., & Kriegeskorte, N. (2020). Recurrent neural networks can explain flexible trading of speed and accuracy in biological vision. *PLoS Computational Biology*, *16*(10). https://doi.org/10.1371/JOURNAL.PCBI.1008215