Recurrent issues with deep neural networks of visual recognition

Tim Maniquet (timotheemaniquet@gmail.com)
Brain & Cognition Research Unit, 106 Tiensestraat
Leuven, Belgium

Hans Op de Beeck (hans.opdebeeck@kuleuven.be)
Brain & Cognition Research Unit, 106 Tiensestraat
Leuven, Belgium

Andrea Ivan Costantino (andreaivan.costantino@kuleuven.be)
Brain & Cognition Research Unit, 106 Tiensestraat
Leuven, Belgium

Abstract

The human ventral stream is equipped with recurrent connectivity allowing it to deal with the noise and uncertainty of visual inputs. Including recurrence in Deep Neural Networks (DNNs) is a promising way of modelling this non-feedforward connectivity. However, just like the role of recurrent processing in the brain remains elusive, it is unclear how making DNNs recurrent makes them more human-like. Here, we put to the test a wide range of DNN models equipped with various recurrent connections. We compared them to human behaviour facing challenging object recognition, and found recurrent model performance and consistency with humans to be mediated by size. Moreover, we found recurrent DNN confusion matrices to be less similar to that of humans than feedforward ones. These findings give perspective on the implementation of recurrence and the benchmarks used to assess it.

Keywords: recurrent processing; recurrent neural networks; object recognition

Introduction

Recurrent processing in the ventral stream allows to deal with the computational difficulty of solving real-world object recognition. Recurrence dynamically processes information, and is especially implicated in the challenging scenarios of recognition, when sensory inputs are incomplete or otherwise ambiguous. In line with this, it has been shown that including recurrent connections in the architecture of DNNs made them more resilient to visually challenging conditions (Loke et al., 2022; Thorat, Aldegheri, & Kietzmann, 2021; Sörensen, Bohté, de Jong, Slagter, & Scholte, 2023). Simultaneously, recurrent DNNs have been shown to develop more brain-like representations than feedforward ones (Kar, Kubilius, Schmidt, Issa, & DiCarlo, 2019; Kietzmann et al., 2019).

While it is agreed that recurrence serves the resolution of complex visual inputs, there remains much to learn about the representational details carried out by recurrent signals. This is due in part to the high number of existing recurrent connections to investigate, and in part to the virtually infinite number of possible challenges that these connections could be involved with. There is a wealth of perceptual phenomena potentially explained by different types of recurrence. In this study, we ask whether we can use DNNs to distinguish between these different types of recurrent processing. We bring together models from the CORnet (Kubilius et al., 2018) and B (Spoerer, McClure, & Kriegeskorte, 2017) architectures with variations of lateral and feedback connectivity, and try them against human participants on a challenging object recognition task. We also include large, feedforward-only VGG models as controls for the influence of size in task performance (Simonyan & Zisserman, 2015).

We found no task-specific dissociation between types of recurrence. Furthermore, we observed recurrence to not change information processing in a different way than model size did. Finally, we report a decrease in confusion matrix correlation with humans for recurrent models as compared to feedforward ones.

Table 1: Models used in the study.

Model	Nb. parameters	Recurrence
C (CORnet Z)	1.5m	/
C V1-V1	1.6m	Lateral V1-V1
C IT-IT	1.6m	Lateral IT-IT
CL (CORnet RT)	4.7m	Lateral
CT	11.5m	Top-down
CLT	11.5m	Lateral + top-down
CS (CORnet S)	52.9m	Lateral + skip
B	8.3m	/
BL	17.2m	Lateral
BT	12.6m	Top-down
BLT	21.5m	Lateral + top-down
VGG11	9.4m	/
VGG16	134.3m	/

Methods

Stimulus set & task Images of objects from 8 categories were chosen from online databases (Lin et al., 2014; Zhou et al., 2018). 16 different visual manipulations were applied onto every image, each consisting in a variations of clutter, occlusion or phase scrambling. These three classes of manipulations are known to trigger recurrent processing (Seijdel et al., 2021; Rajaei, Mohsenzadeh, Ebrahimpour, & Khaligh-Razavi, 2019; Tang et al., 2018). Both human participants and DNN models were presented with a categorisation task on the 8 chosen object categories, which included backward masking for humans.

DNN modelling We investigated thirteen distinct DNN models, chosen to represent a range of complexities, from basic feedforward structures to more sophisticated recurrent networks. Within the CORnet models (hereafter renamed *C*), we included the foundational C Z, C RT (hereafter called *C* and *CL*, respectively) and C S, as well as the custom-built C V1-V1, C IT-IT, CT and CLT. We included all four B models, as well as VGG11 and VGG16 (see Table 1 for more details).

Results

Model performance

Average accuracy varied significantly across models, with VGG16 significantly higher than all others ($p < 0.001$, Tukey test, see Figure 1). While more recurrent models seem to perform higher than their baseline counterparts, this improvement is largely explained by the increase in size brought by adding recurrence, as shown by the large correlation between average accuracy and model size (Pearson's correlation $r = 0.74$, $p = 0.003$).

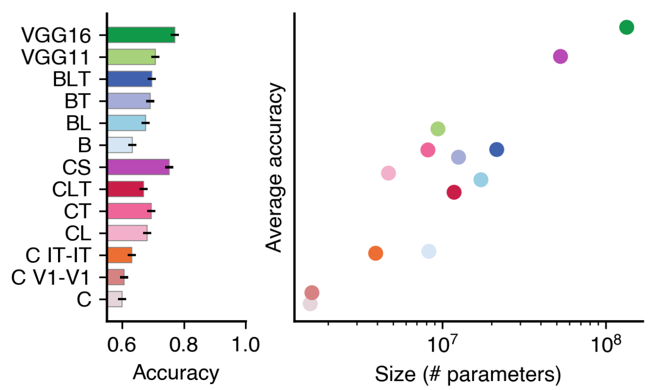

Figure 1: **Model performance correlates with size**. Left: grand average performance per model. Error bars represent 95% confidence intervals. Right: model average accuracy per model size, in number of parameters.

Task consistency

We then checked which models were more consistent with human participants by correlating patterns of accuracy across our 16 visual conditions. We found an overall high correlation (> 0.65), indicating an agreement on task difficulty (see Figure 2). While recurrent models correlate more than their feedforward counterparts, VGG16 surpasses all with a correlation of 0.85. Generally, model size seems to drive the consistency of models with humans (Pearson's correlation $r = 0.83$, $p = 3.9e-04$), in a way comparable to how it drives performance.

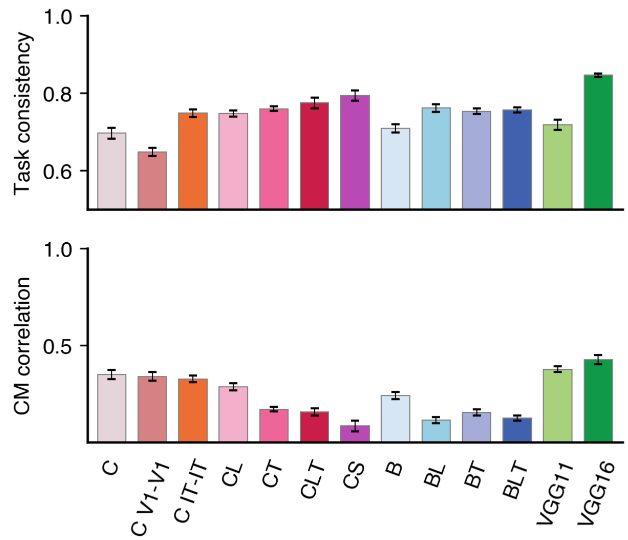

Figure 2: **Recurrence helps consistency and impairs confusion matrix correlation**. Pearson's R correlation scores between models and humans in task difficulty (top) and confusion matrix (bottom, error bars represent 95% confidence intervals).

Confusion matrices

We finally looked at confusion matrices to compare category-level representations between humans and models. Our results indicate that recurrence in models *decreases* the correlation with humans, with baseline feedforward models outperforming recurrent ones within the B and C families (see Figure 2). The feedforward VGG11 and VGG16, on the other hand, follow a similar trend as in previous analyses and outperform all others.

Conclusion

Our findings highlight the limitations of recurrence as implemented in our models. On the one hand, performance and task consistency analyses suggest that recurrent DNNs are equally good as their size-matched feedforward unwrapped equivalent. On the other hand, confusion matrix comparisons highlight a discrepancy in the information processing strategies brought by recurrent connectivity compared to humans. This is especially striking when considering that recurrent models are shown to better match patterns of activity in the ventral stream, which suggests that recurrent DNNs could at the same time be better models of the human brain and worse models of human behaviour than feedforward ones.

Acknowledgments

T.M. was funded by FWO (Fonds Wetenschappelijk Onderzoek) through a fundamental research doctoral fellowship (11C9122N). H.O.B. was supported by the KU Leuven Research Council (C14/21/047; AKUL/19/05) and Methusalem funding (METH/24/003) by the Flemish Government & KU Leuven.

References

Kar, K., Kubilius, J., Schmidt, K., Issa, E. B., & DiCarlo, J. J. (2019, June). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature Neuroscience*, *22*(6), 974–983. doi: 10.1038/s41593-019-0392-5

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K. A., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019, October). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences*, *116*(43), 21854–21863. doi: 10.1073/pnas.1905544116

Kubilius, J., Schrimpf, M., Nayebi, A., Bear, D., Yamins, D. L. K., & DiCarlo, J. J. (2018, September). *CORnet: Modeling the Neural Mechanisms of Core Object Recognition.* bioRxiv. doi: 10.1101/408385

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014* (pp. 740–755). Cham: Springer International Publishing. doi: 10.1007/978-3-319-10602-1$_{4}8$

Loke, J., Seijdel, N., Snoek, L., van der Meer, M., van de Klundert, R., Quispel, E., ... Scholte, H. S. (2022, November). A Critical Test of Deep Convolutional Neural Networks' Ability to Capture Recurrent Processing in the Brain Using Visual Masking. *Journal of Cognitive Neuroscience*, *34*(12), 2390–2405. doi: 10.1162/jocn$_{a0}1914$

Rajaei, K., Mohsenzadeh, Y., Ebrahimpour, R., & Khaligh-Razavi, S.-M. (2019, May). Beyond core object recognition: Recurrent processes account for object recognition under occlusion. *PLOS Computational Biology*, *15*(5), e1007001. doi: 10.1371/journal.pcbi.1007001

Seijdel, N., Loke, J., van de Klundert, R., van der Meer, M., Quispel, E., van Gaal, S., ... Scholte, H. S. (2021, July). On the Necessity of Recurrent Processing during Object Recognition: It Depends on the Need for Scene Segmentation. *Journal of Neuroscience*, *41*(29), 6281–6289. doi: 10.1523/JNEUROSCI.2851-20.2021

Simonyan, K., & Zisserman, A. (2015, April). *Very Deep Convolutional Networks for Large-Scale Image Recognition* (No. arXiv:1409.1556). arXiv. doi: 10.48550/arXiv.1409.1556

Sörensen, L. K. A., Bohté, S. M., de Jong, D., Slagter, H. A., & Scholte, H. S. (2023, June). Mechanisms of human dynamic object recognition revealed by sequential deep neural networks. *PLOS Computational Biology*, *19*(6), e1011169. doi: 10.1371/journal.pcbi.1011169

Spoerer, C. J., McClure, P., & Kriegeskorte, N. (2017). Recurrent Convolutional Neural Networks: A Better Model of Biological Object Recognition. *Frontiers in Psychology*, *8*, 1551. doi: 10.3389/fpsyg.2017.01551

Tang, H., Schrimpf, M., Lotter, W., Moerman, C., Paredes, A., Ortega Caro, J., ... Kreiman, G. (2018, August). Recurrent computations for visual pattern completion. *Proceedings of the National Academy of Sciences*, *115*(35), 8835–8840. doi: 10.1073/pnas.1719397115

Thorat, S., Aldegheri, G., & Kietzmann, T. C. (2021, November). Category-orthogonal object features guide information processing in recurrent neural networks trained for object categorization. *arXiv:2111.07898 [cs]*.

Zhou, B., Zhao, H., Puig Fernandez, X., Xiao, T., Fidler, S., Barriuso, A., & Torralba, A. (2018, December). Semantic Understanding of Scenes Through the ADE20K Dataset. *arXiv*.