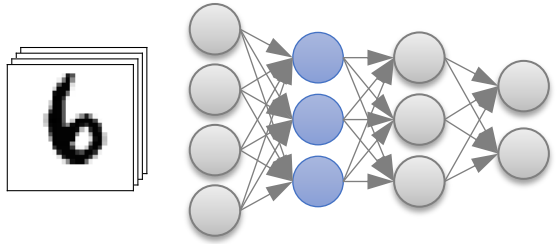
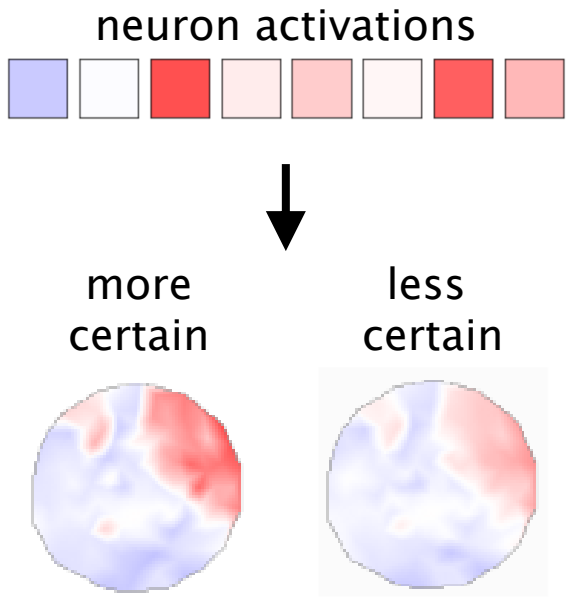
Relation of Activity and Confidence when Training Deep Neural Networks





Valerie Krug, Christopher Olson, Sebastian Stober

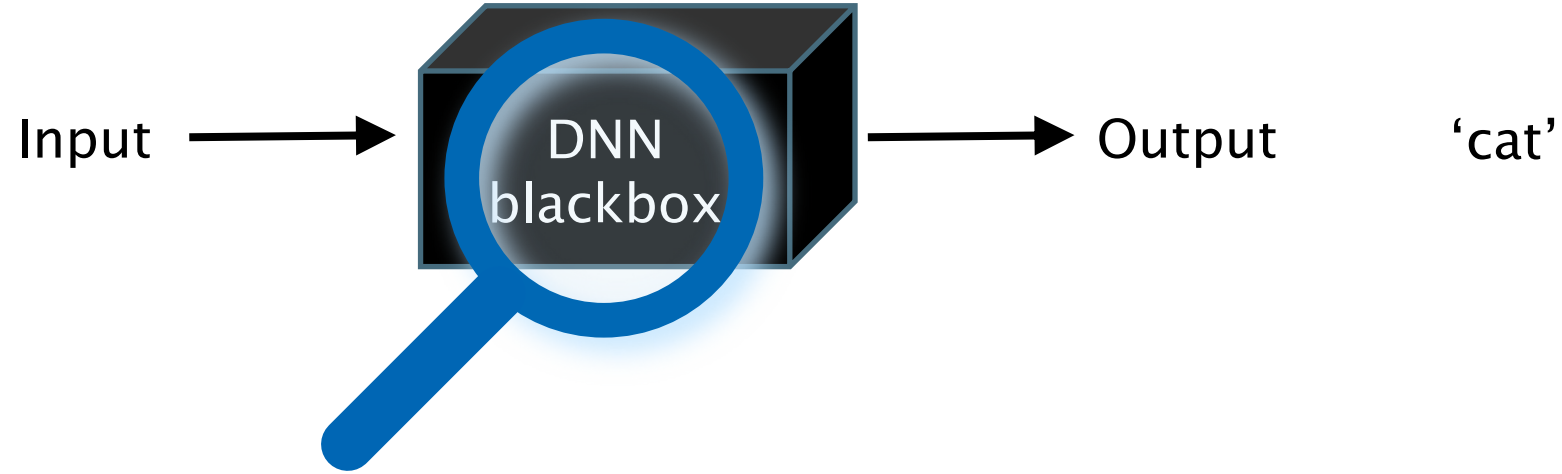
Artificial Intelligence Lab at OVGU Magdeburg

contact: valerie.krug@ovgu.de
web: ai.ovgu.de

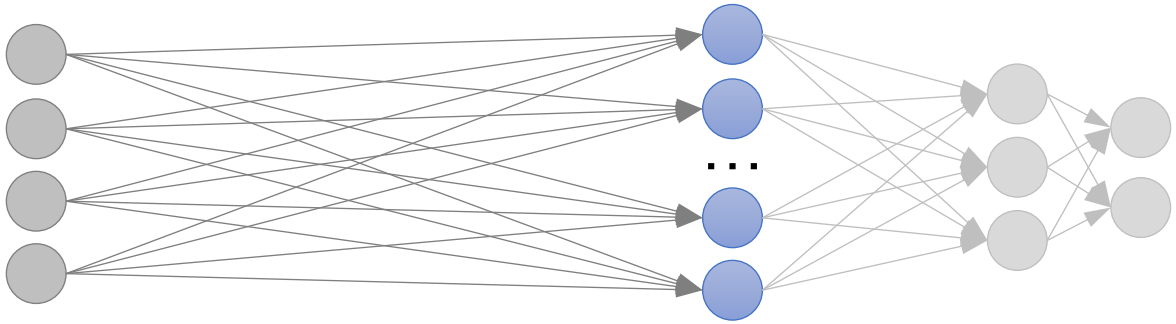
Uncertainty meets Explainability

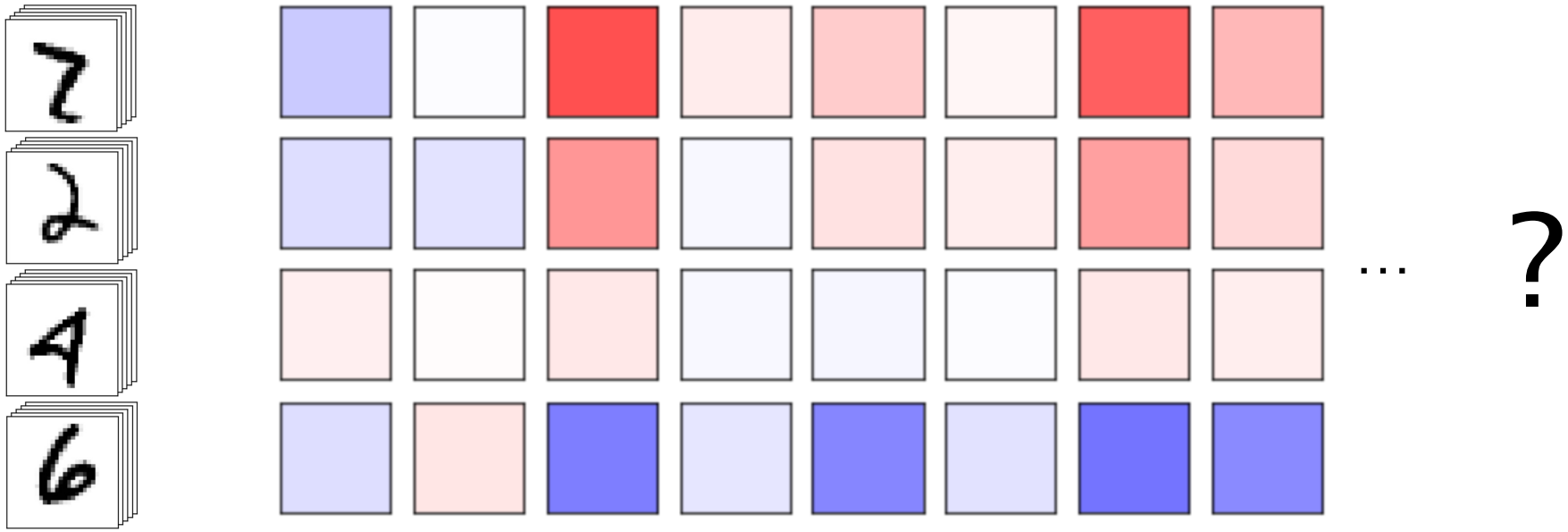
GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

Uncertainty meets Explainability

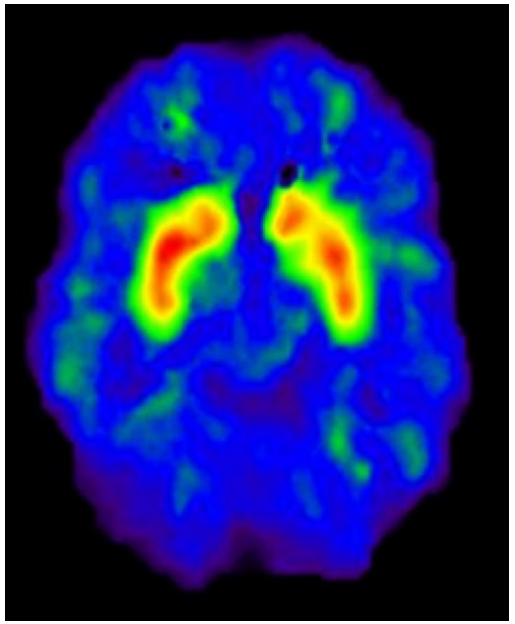


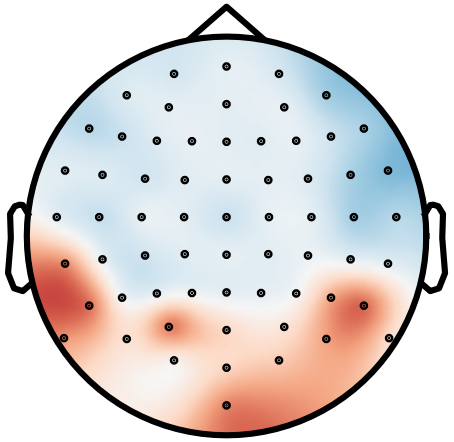
DNN Activations are not Visually Interpretable

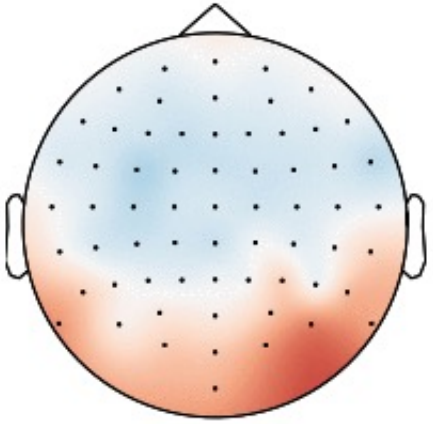




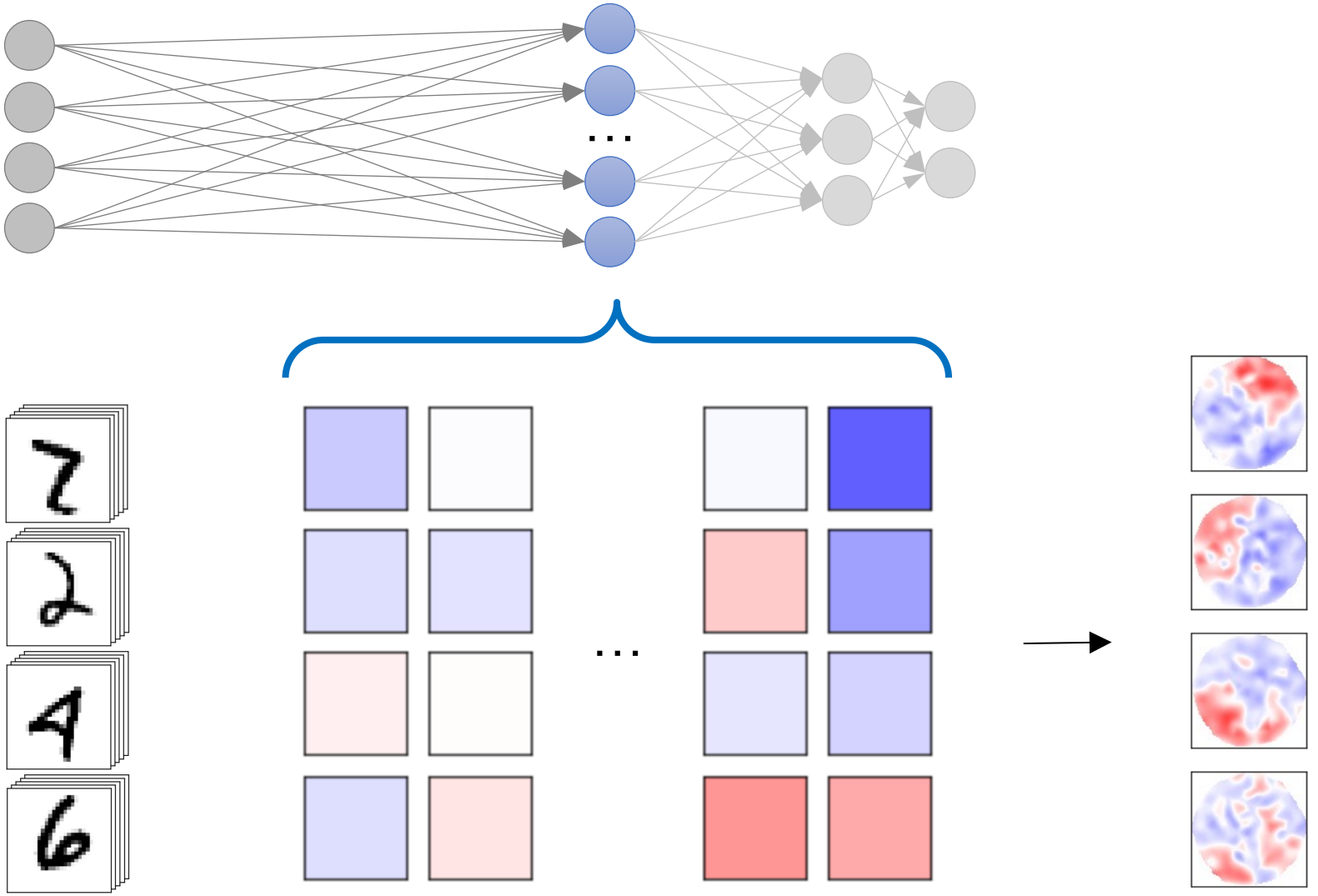
Visualizing Brain Activity



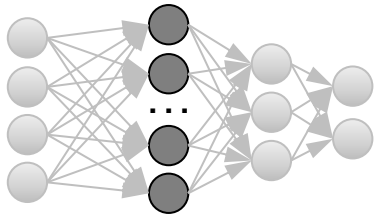




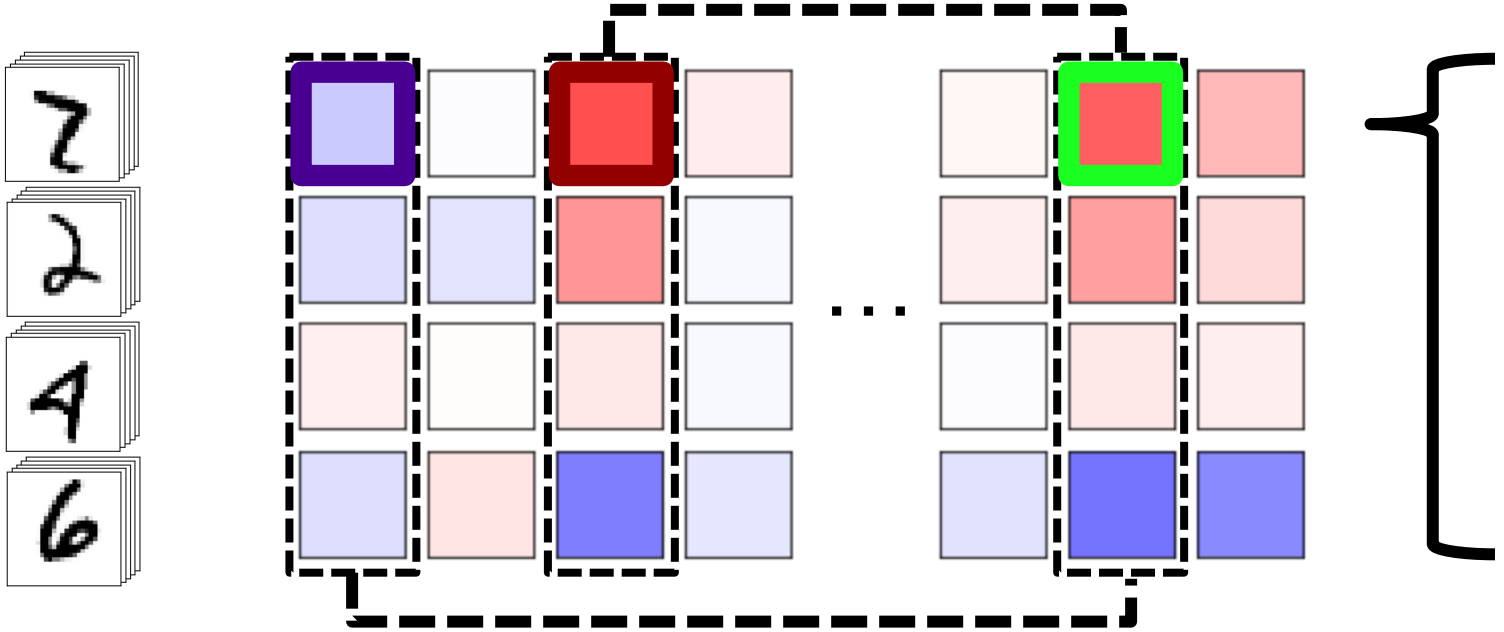
Visualizing DNN Activations



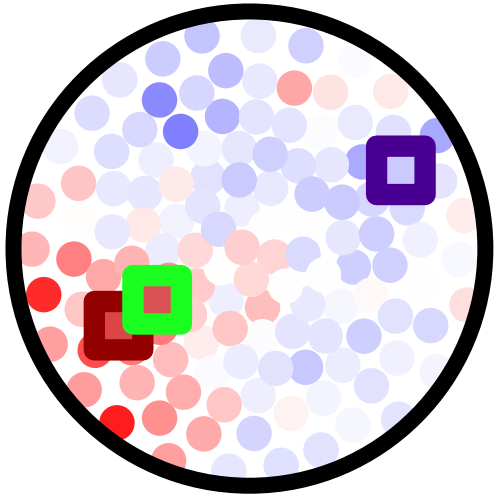
Topographic Activation Maps



high activation similarity across groups
→ small distance in layout

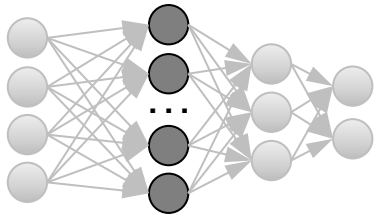


low activation similarity across groups
→ high distance in layout

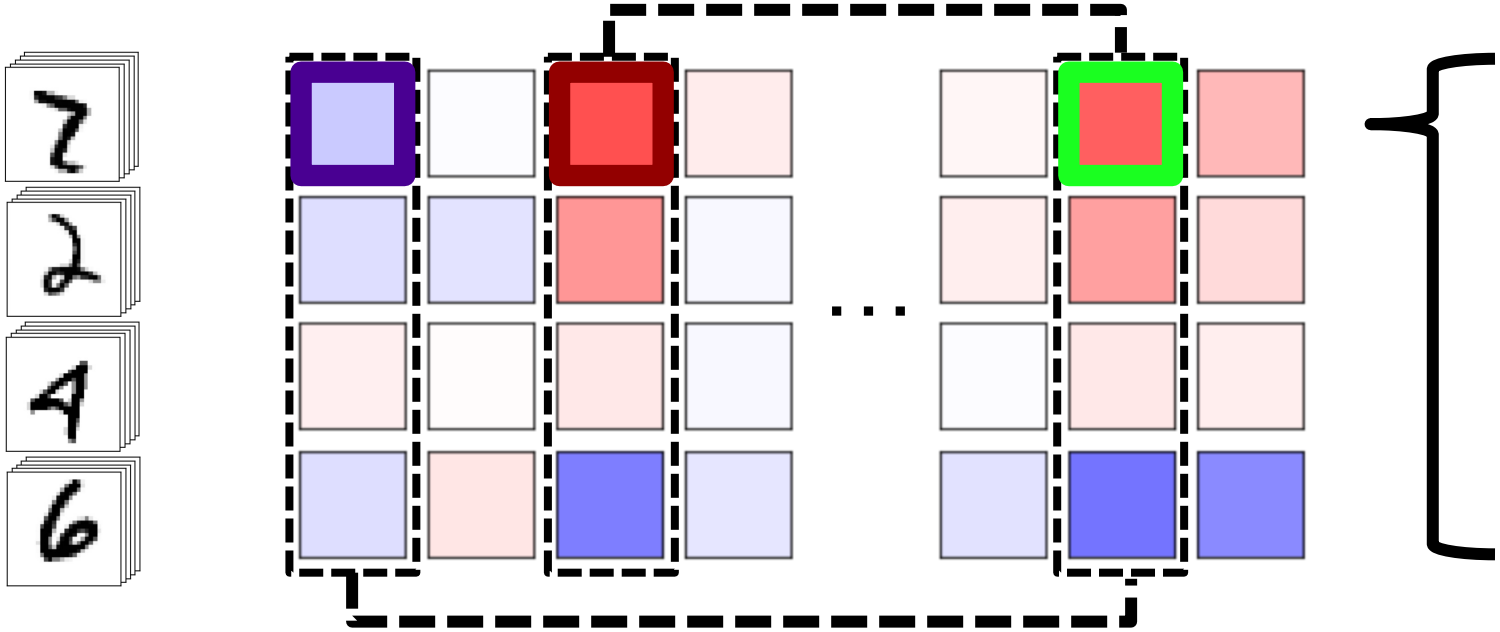


layouted neurons
colored by activation
values for group ‘7’

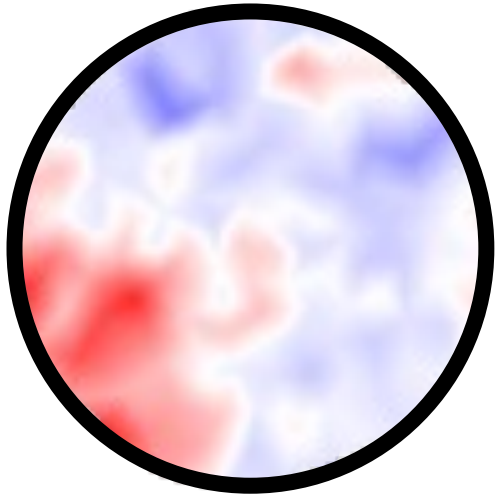
Topographic Activation Maps



high activation similarity across groups
→ small distance in layout



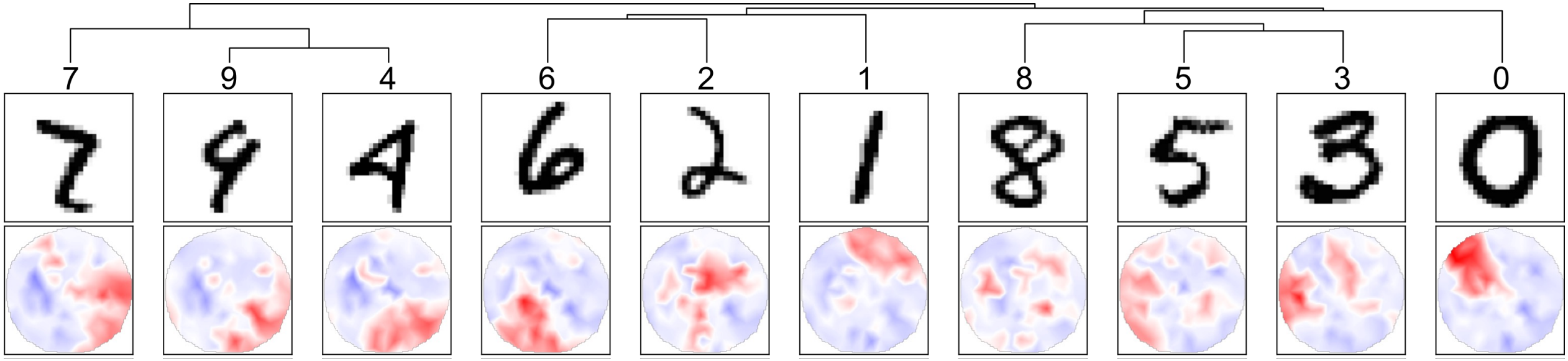
low activation similarity across groups
→ high distance in layout



layouted neurons
colored by activation
values for group ‘7’

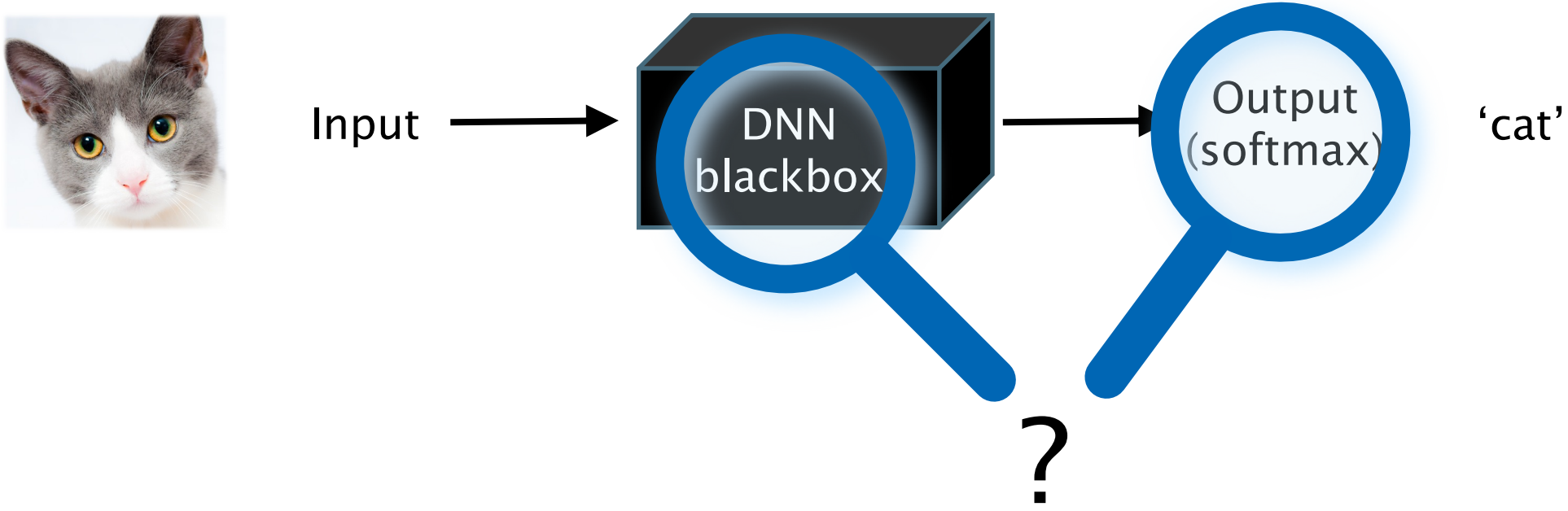
Comparing Groups with Topographic Maps

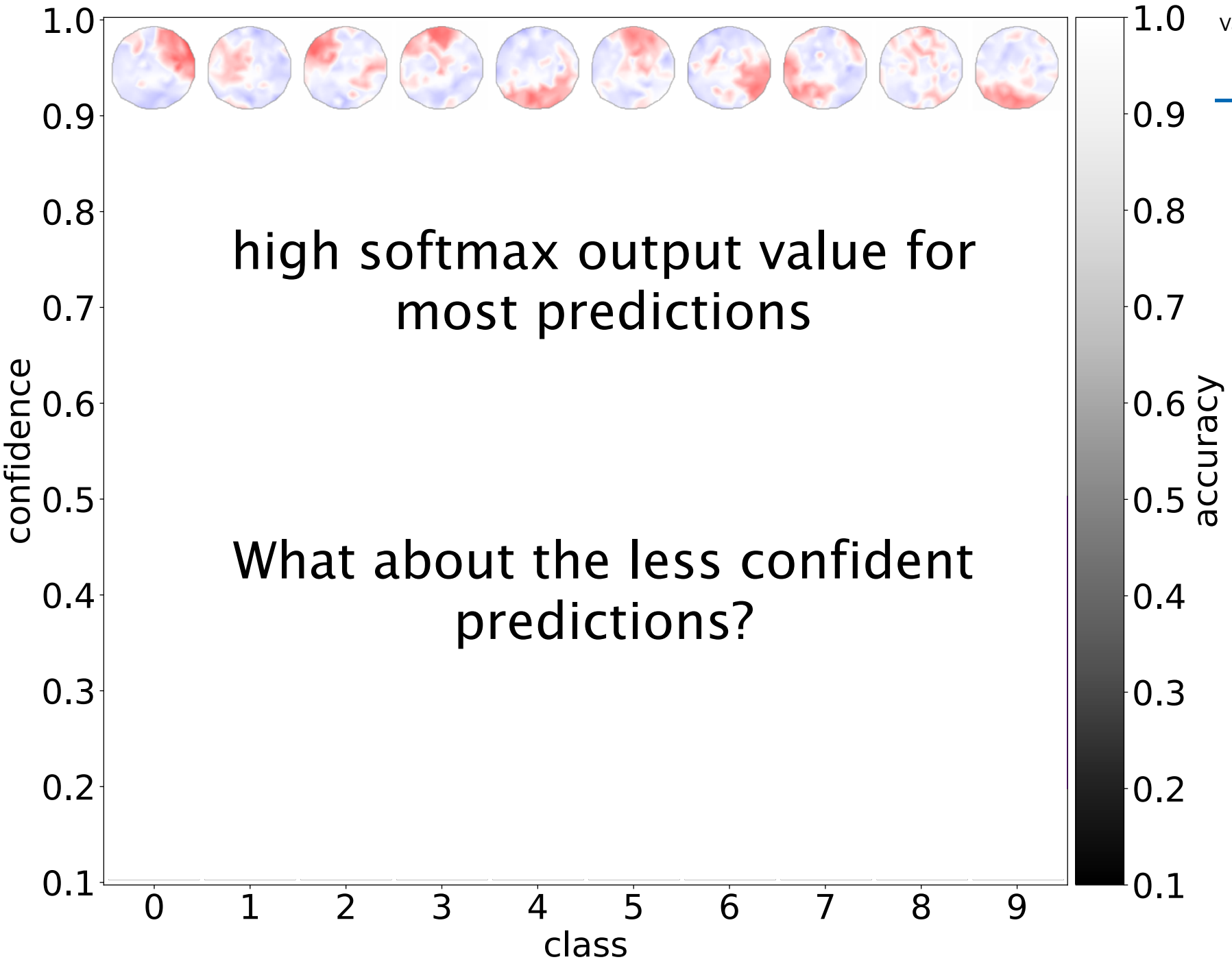
same layout, different activity per group

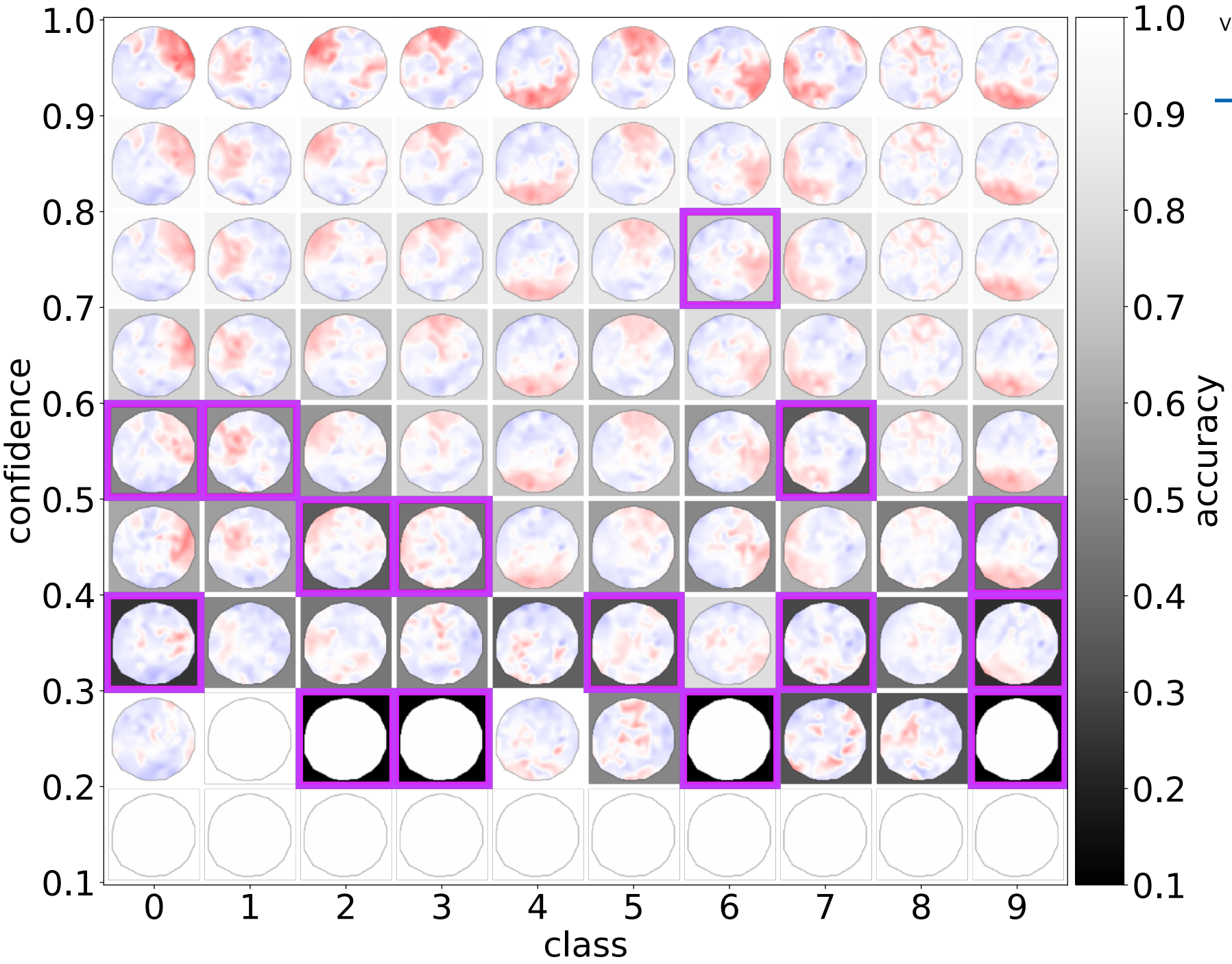


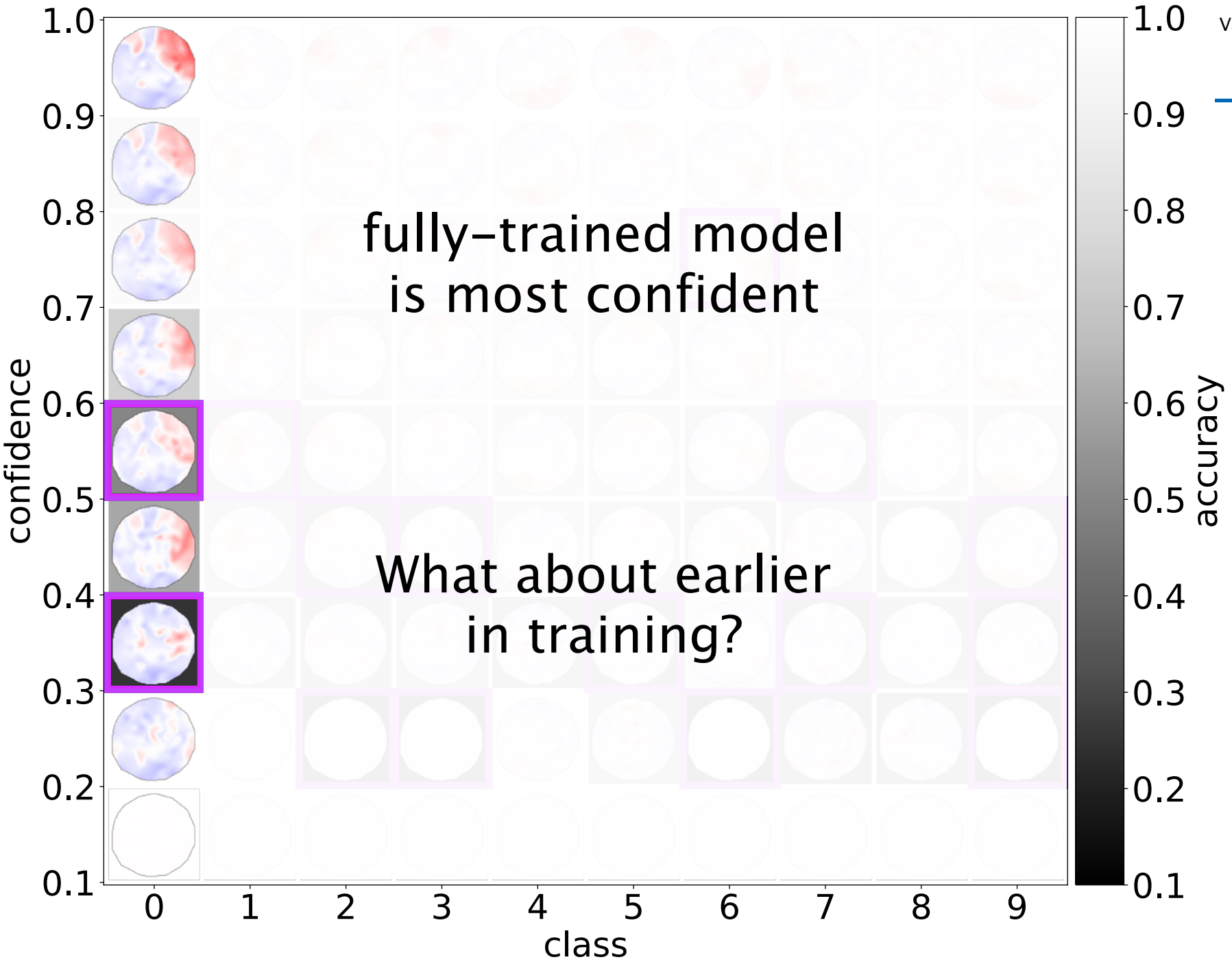
Krug et al. (2023). Visualizing Deep Neural Networks with Topographic Activation Maps. In HHAI 2023: Augmenting Human Intellect, pages 138–152. IOS Press.

Uncertainty meets Explainability



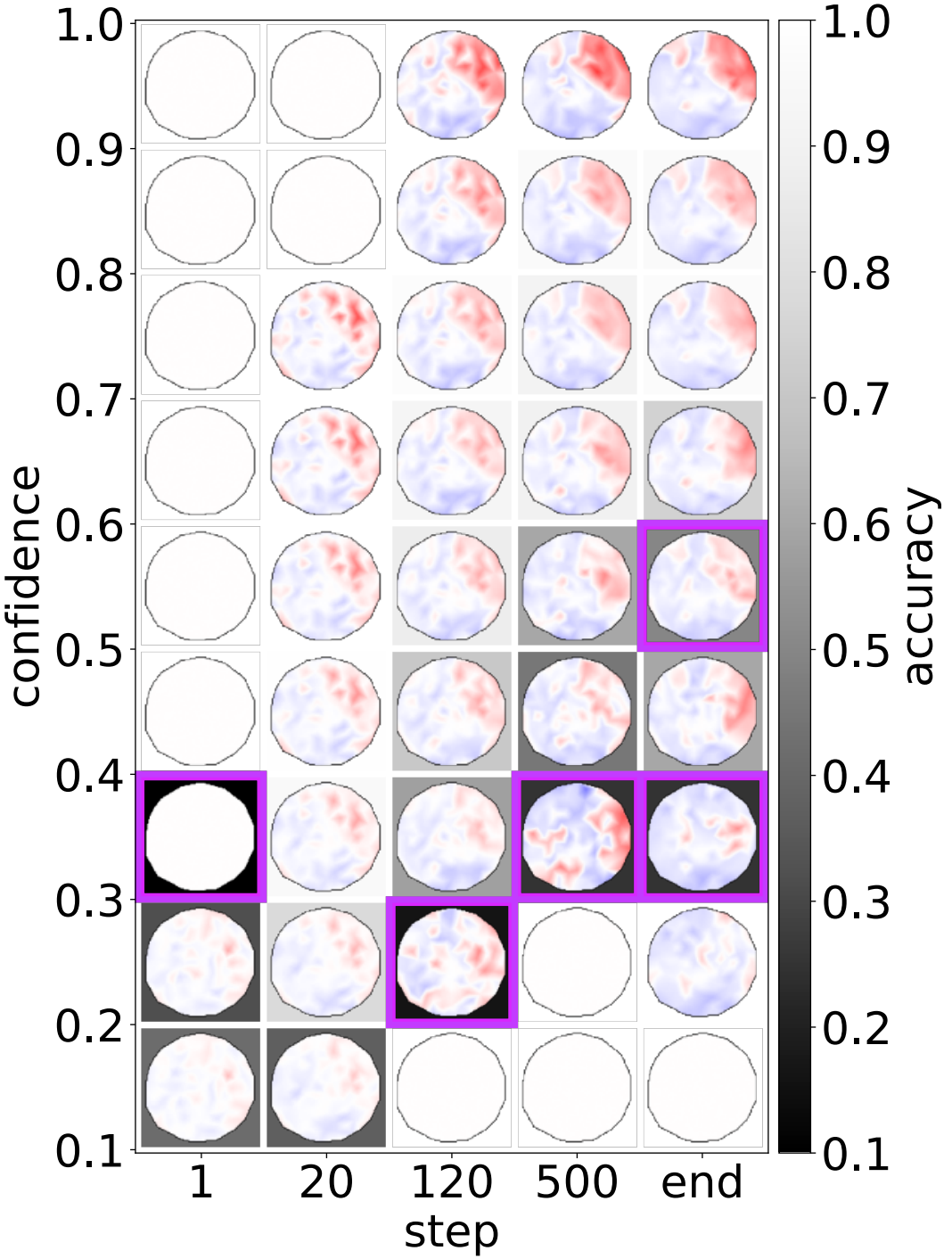

high softmax output value for most predictions
What about the less confident predictions?
confidence
class
accuracy
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
9


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
confidence
0
1
2
3
4
5
6
7
8
9
class
accuracy
overconfidence
confidence
>
accuracy



overconfidence

confidence
>
accuracy



overconfidence

confidence
>
accuracy

Future work

Investigate:

- more complex models
- longer training times (with potential overfitting)
- more sophisticated confidence measures

Acknowledgements

Thanks to

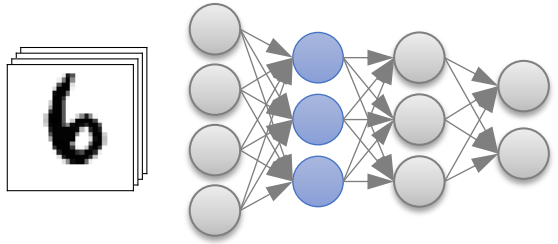
- my co-authors
  Chris and Sebastian
- AI Lab @ OVGU Magdeburg

- BMBF for funding

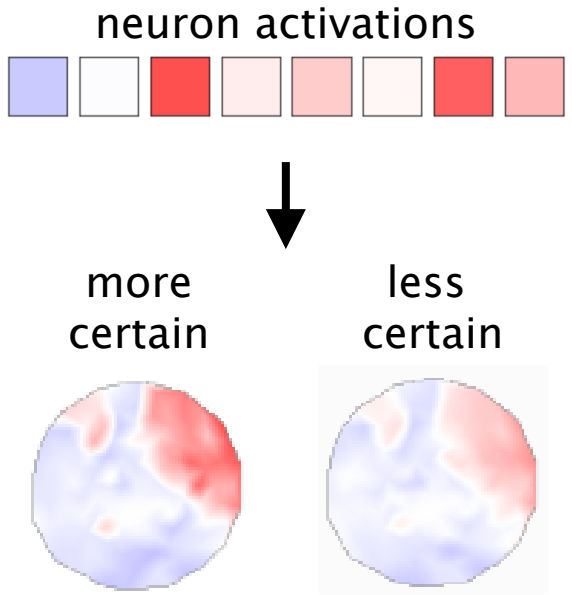
- You for your attention

http://ai.ovgu.de/

Relation of Activity and Confidence when Training Deep Neural Networks



Valerie Krug, Christopher Olson, Sebastian Stober

Artificial Intelligence Lab at OVGU Magdeburg

contact: valerie.krug@ovgu.de
web: ai.ovgu.de