Visual perception of liquids: insights from deep neural networks

Jan Jaap R. van Assen, Shin’ya Nishida, Roland W. Fleming

1 NTT Communication Science Laboratories, 2 Kyoto University, 3 Justus-Liebig-University Giessen

This work has been supported by JSPS KAKENHI (JP15H05915) and an ERC Consolidator Award (ERC-2015-CoG-682859: "SHAPE").

**Network architecture**

- Convolutional layer
- Fully-connected layer

**Human performance vs Network performance**

- Dataset: 10 scenes (varying key interactions)
- 16 viscosities (0.001 Pa·s to 10 Pa·s)
- 5 optical variations (illumination, viewpoint)
- 10 scenes (varying key interactions)
- 20 frame sequences (30 fps)
- 5 optical variations (illumination, viewpoint)

**Network similarities**

- 100 networks
- 1-Pearson correlations
- Spearman correlations

**Conclusions**

- Relatively shallow feedforward DNNs are able to predict the perceived viscosity very well, including the estimation errors.
- This network makes similar estimation errors as human observers and outperforms most individual observers compared to the human mean.
- We identified a range of image cues that are represented across nodes and layers, which form the foundation for viscosity estimation.
- Clusters in our image metric space show a clear computational structure of which we found eight clusters of different networks, which suggest that there are multiple solutions to the image-based viscosity estimation problem.