Interpreting ANN's that Predict Illusory Motion using Network Science

Levi Burner, COMBINE Symposium, Spring 2021, University of Maryland

Motivation

- Studying the human visual system can reveal
  - Important principles of brain function
  - Bio-inspired and robust computer vision algorithms
- In-vivo study of the visual cortex is difficult
  - Topology of visual cortex has been probed
  - Identification of function is still difficult
- Artificial Neural Networks (ANN) are much easier to inspect
  - Need to show information is processed in a similar manner

Background

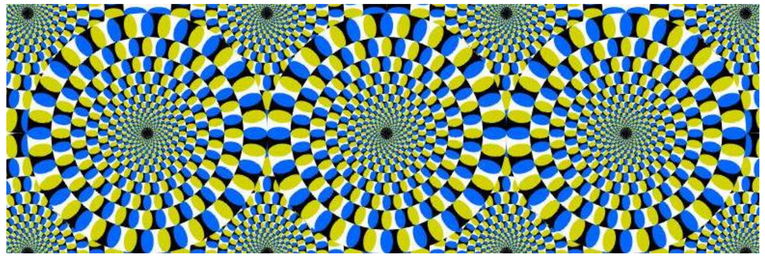

Fig 1: Optical illusions can be used to understand the visual system. The Rotating Snakes pattern elicits a sense of rotation [1].

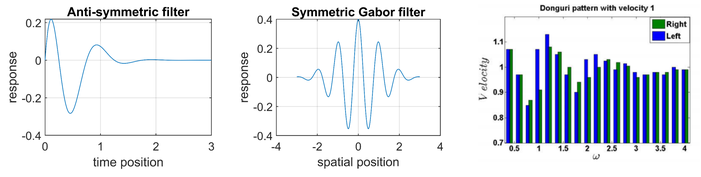

Fig 2: The Snake Illusion may be due to interaction between small-motions of the eye and anti-symmetric temporal filtering (left) and symmetric spatial filtering (center) in the visual cortex. Simulation of the filters results in biased estimates of image velocity (right) [1].

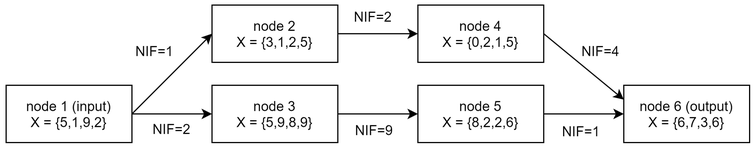

Fig 3: Neural Information Flow (NIF) is assigned to links of a network based on the values of nodes after many evaluations. The link score is based on the mutual information in the values at each node [3].

- An attribution map is created from input to output by:
  - Finding all paths from input to output
  - For each path multiply NIF scores along path
  - The formula is:

$$A_{ij} = \sum_{p \in \mathbb{P}} \prod_{l \in L(p)} NIF(l)$$

Network Description

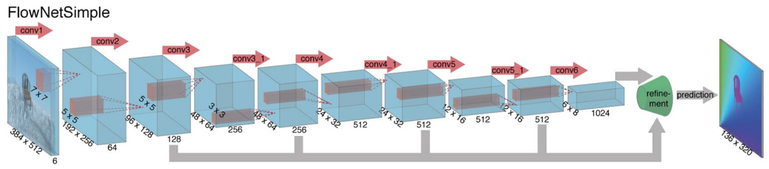

Fig 4: CNN named "FlowNetS", was modified to accept multiple frames [2]. Output is the image velocity (optical flow).

- In Convolutional Neural Networks
  - Nodes are connected by directed edges
  - Nodes have real values that are a function of the values of nodes connected by incoming directed edges
- Network trained with high framerate version of MPI-Sintel [4]
- Network trained to penalize latency

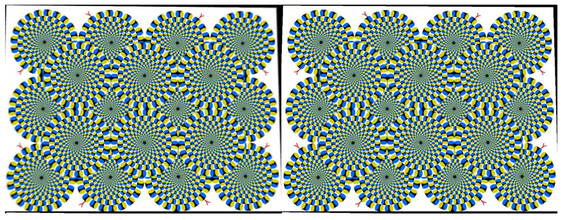

Fig 5: To simulate eye motion, input is series of slightly shifted snake illusion's. The true displacement between every pixel is known.

Research Questions

- Does the internal processing of the CNN support the signal processing based explanation for the Snake Illusion?
- Does NIF show ANN's are emphasizing information as predicted by the signal processing hypothesis?
- What other feature attribution methods are appropriate?
- Do the attributions make sense for non-illusory scenes?

Preliminary Results

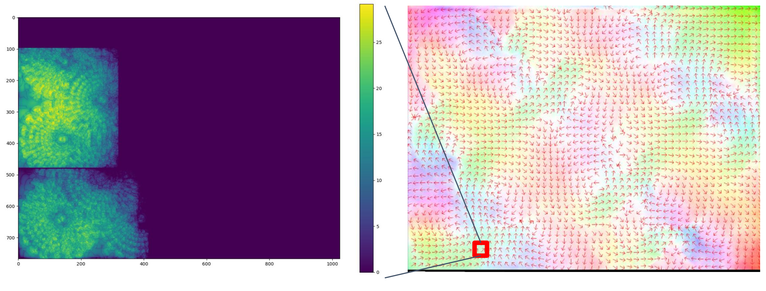

Fig 6: (left) NIF attribution of input to a single output flow vector. Network was evaluated on a sequence of illusion images like Fig 5. (right) Biased flow estimate of network that perceives the snake illusion (swirls). Output flow pixel considered highlighted in red.

Discussion

- Preliminary results do not support the spatial filter hypothesis
  - However attribution was high far from output flow vector, this does not make sense
  - Test on other scenes needed to ensure attribution is sensical
- Neural Information Flow results are likely inaccurate
  - Sample size was low (15 frame batches)
  - Information measure used by NIF is very approximate
- NIF is just one method of neural network interpretation
  - NIF may not be the most appropriate
  - Comparison to other attribution methods is needed
- FlowNetS NIF net has 2.5 million nodes and 13.5 million links
  - Limits attribution methods used due to network size
- Network is purely feedforward
  - Can not capture influence of recurrence in visual cortex
  - Cannot be recurrent or it will be infeasible to calculate saliency with current methods

Research Plan

- Systematic test of attribution methods
  - Determine if attribution method results are reasonable

Fig 7: (left) a scene with a single moving object. (right) a "sensible" attribution map that NIF should predict if it is working correctly [5].

- Grid search over illusory pattern parameters
  - Spatial scale
  - Speed of motion
  - Apply quantitative test of similarity between attribution and spatial filtering hypothesis.
- Refine the NIF algorithm
  - Use better measure of information to assign NIF scores

References/Acknowledgements

[1] Fermüller, C., Ji, H., & Kitaoka, A. (2010). Illusory motion due to causal time filtering. Vision Research, 50(3), 315–329.
[2] Dosovitskiy, A., Fischery, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., … Brox, T. (2015). FlowNet: Learning optical flow with convolutional networks. Proceedings of the IEEE International Conference on Computer Vision, 2015 Inter, 2758–2766.
[3] Davis, B., Bhatt, U., Bhardwaj, K., Marculescu, R., & Moura, J. M. F. (2020). On Network Science and Mutual Information for Explaining Deep Neural Networks. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2020-May(1), 8399–8403.
[4] G, J. J. F., Michael, W., & Andreas, B. (2017). Slow Flow : Exploiting High-Speed Cameras for Accurate and Diverse Optical Flow Reference Data. 3597–3607.
[5] Maczyta, L., Bouthemy, P., & Meur, O. Le. (2019). Unsupervised Motion Saliency Map Estimation Based on Optical Flow Inpainting. Proceedings - International Conference on Image Processing, ICIP, 2019-Septe, 4469–4473.

This research was supported in part by NSF award DGE-1632976