Using Machine Learning to Find Interesting Features of Cell Migration

Office of Information Technology at the
National Institute of Dental and Craniofacial Research,
NIH
Mentor: John Prue, NIDCR CIO

CIVIC DIGITAL FELLOWSHIP

NIH

Isaac Robinson
Yale College
Computer Science and Music

The Problem

- Researchers at NIDCR CBS spends hours at a time looking at videos of cells moving over two decades to find 'interesting features'
- What would machine learning mark as interesting?
- Applications to cancer research, difference in what's interesting between healthy and cancerous cells

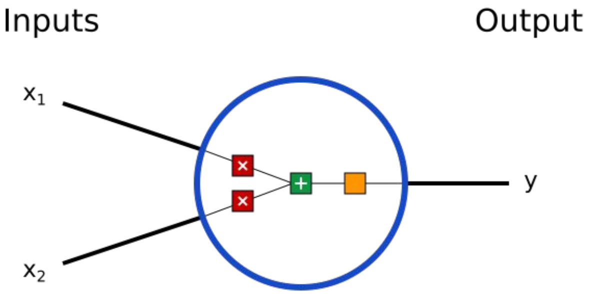
CIVIC DIGITAL FELLOWSHIP

Why the Problem is Hard

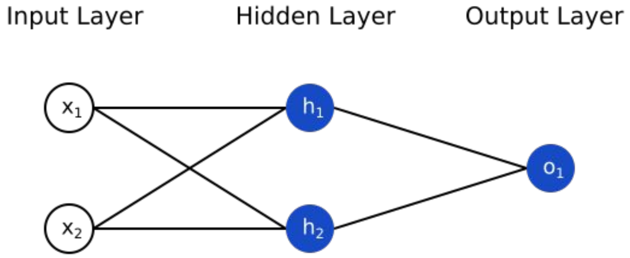
- 'Interesting' is not a quantitative target
- 'Cellular migration' is also not a quantitative target, so interesting with respect to what?
- Machine learning benefits from lots of data, but biologists don't collect that volume
- So machine learning model must:
    - Be unsupervised
    - Have 'interesting' emerge organically from the model

Background: What is a Neural Network

- Layers of 'neurons' that do a calculation and pass their information forward to the next layer
- Layers in between input and output are called 'hidden' layers, deep learning has multiple hidden layers



A single neuron multiplying inputs by weights and adding



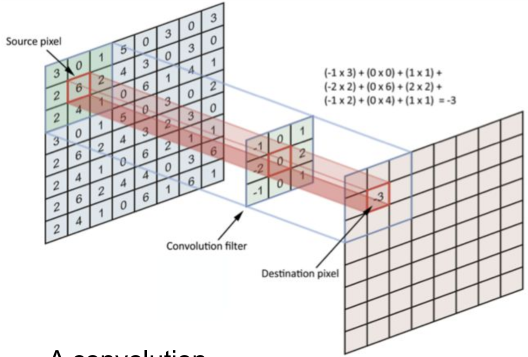
Input neurons passing data forward to hidden neurons, which then pass their data to an output neuron

Source: Towards Data Science [1]

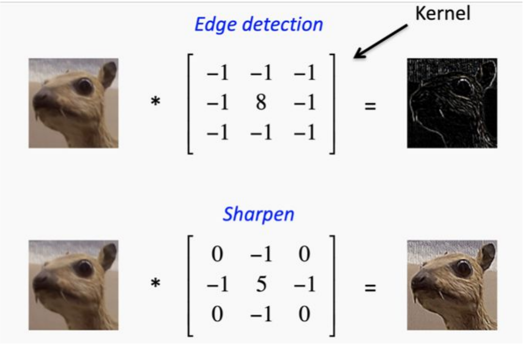
CIVIC DIGITAL FELLOWSHIP

Background: Convolutional Neural Network

- Learn kernels to convolve against input image and extract features
- Many sequential convolutional layers create more and more complex features, extracting features from previous features
- These 'deep features' are hard to interpret from a human perspective



A convolution

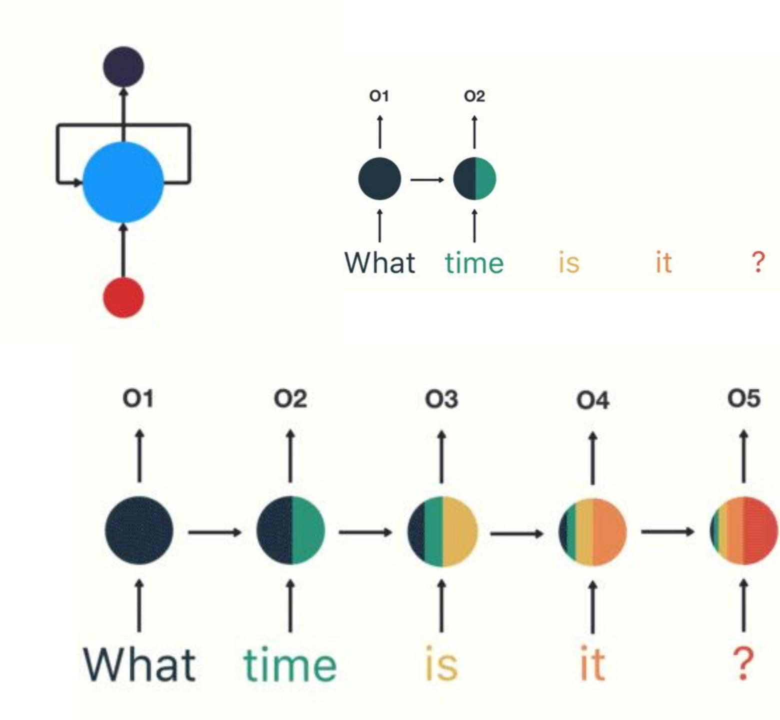


Effect of convolving filters against an image

Source: Towards Data Science [2]

Background: Recurrent Neural Network

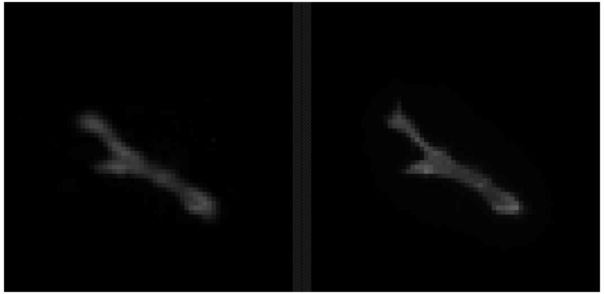
- Process input, produce output and hidden state
- Process next input along with hidden state from previous input via same network
- Keep passing hidden state forward to collect a numerical 'summary' of the inputs seen
- Very difficult to interpret final hidden state from a human perspective



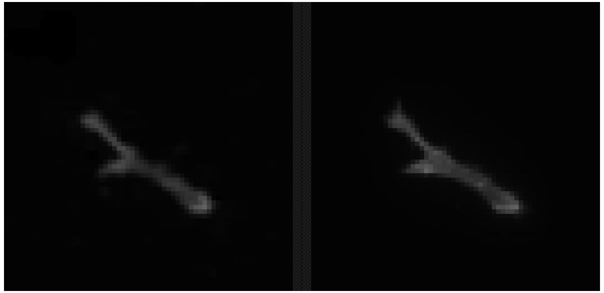
Source: Towards Data Science [3]

Convolutional-Recurrent Neural Network with Attention

- Get visual features from images via convolutional network, pass features to LSTM[6] RNN[3] and allow network to attend[7,8] to images in the sequence differently
- Attention weights can be used as a measure of relative importance for each image in a sequence
- As target, generate last image via DCGAN[4,5] image generator in order to capture all visible aspects of migration, including morphology and motion



Generated via dot product based attention[8,9] (prediction on left and real image on right)



Generated via locality based attention using the concat method[8,9]. Notice the improved shape of the bottom left point.

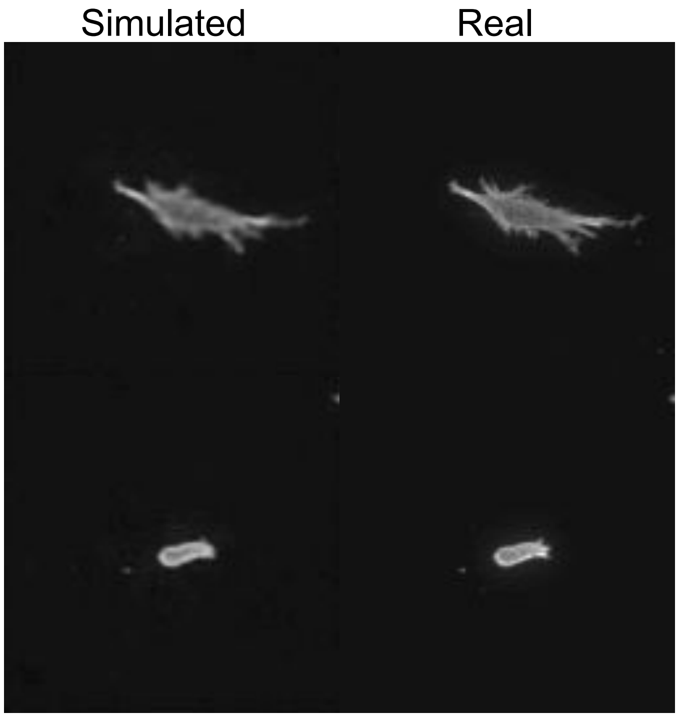
CIVIC DIGITAL FELLOWSHIP

Kullback-Leibler Divergence

- What CRNN+Attention lacks:
    - Want average attention to be distributed evenly across input sequence
    - Want attention to only spike when very important
- Apply secondary loss to neural network based off Kullback-Leibler (KL) divergence of attention from uniform distribution
    - Used in variational autoencoders[10], so there's precedence
- But found that due to location-based attention, found that sequences shifted by one had similar attention distribution shapes, not what we want

Convolutional Transformer

- Get more robust attention mechanism by using a Transformer[11]
- Predict next sequence of images instead of single next image
    - Better captures cellular motion
- But much more complex so potentially will overfit more on small dataset

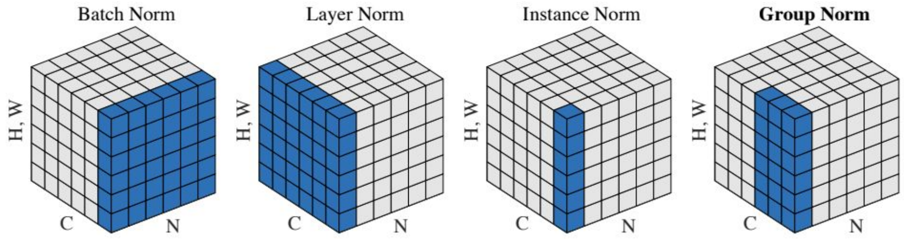


Example predictions of cell motion for a short subsequence. (left is simulation and right is real)

CIVIC DIGITAL FELLOWSHIP

Batch Normalization in Convolutional Sequential Networks

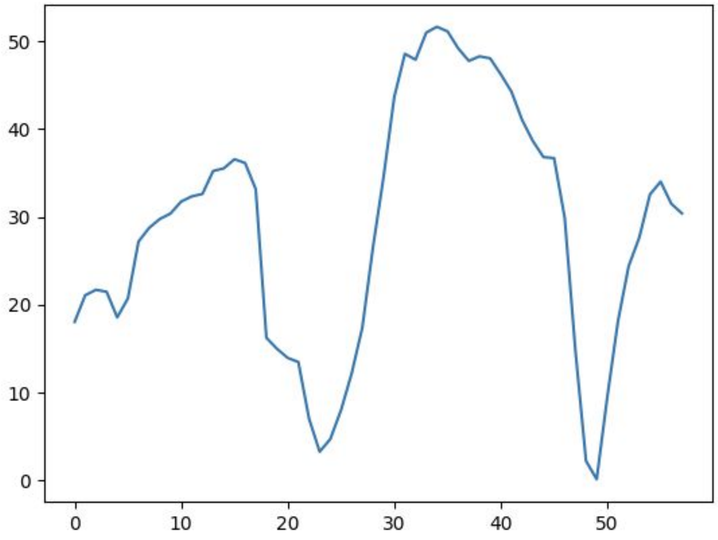
- Batch normalization is essential for modern convolutional neural networks[12] (CNN)
- Since CNNs don't have a concept of time, hard to seperate temporal and batch dimensions and still train effectively
- To predict sequence with Transformer, need to use teacher forcing[13], which when generating messes up batch normalization exposed to temporal dimension
- Switch to group normalization[14] in CNN feature extractor and DCGAN[4,5]-based image generator solves problem since normalizes orthogonally to batch-temporal dimension



CIVIC DIGITAL FELLOWSHIP

Normalizing Transformer Attention

- Transformer attention still is on average unevenly distributed across input sequence
  - Instead of using KLD loss, normalize after training
- Normalize by smallest value in each sequence so relative importance within sequence captured
- Use log of attention probability as measure of information content in event[15]

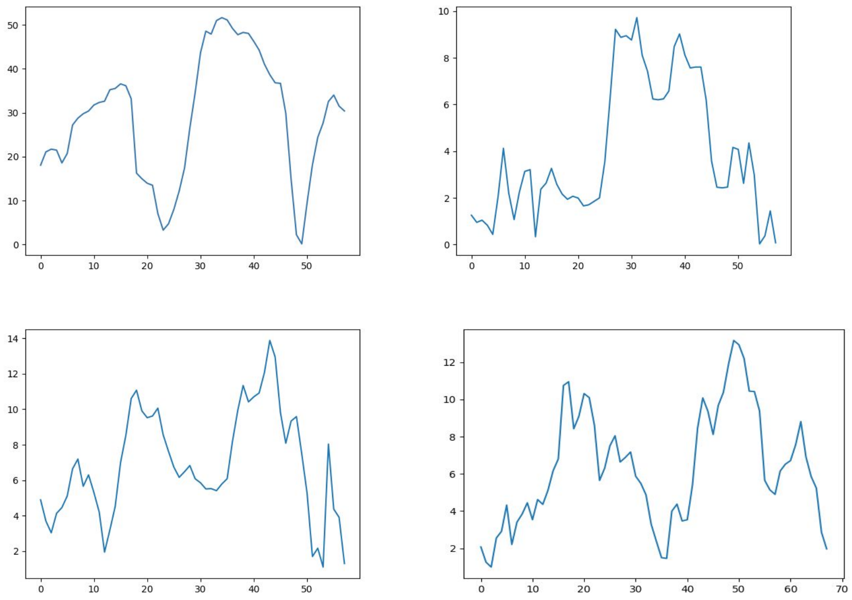


Graph of attention-based importance, importance on y-axis and frame number (time) on x-axis

CIVIC DIGITAL FELLOWSHIP

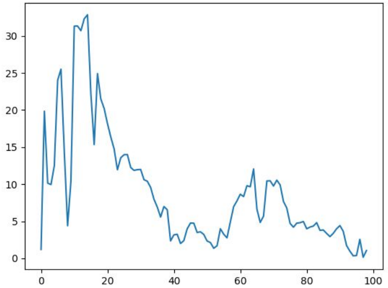
Attention from Different Networks

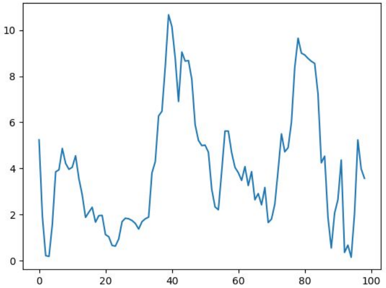
- Ideally, attention-based importance rating would be similar for different networks
- In practice, differences between networks are significant
- Suggests directly using attention weights as measures of importance may not be optimal

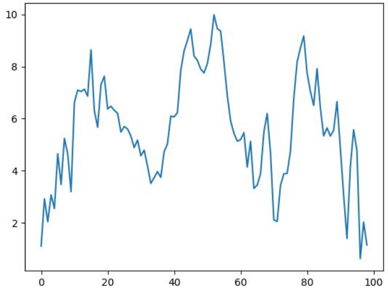


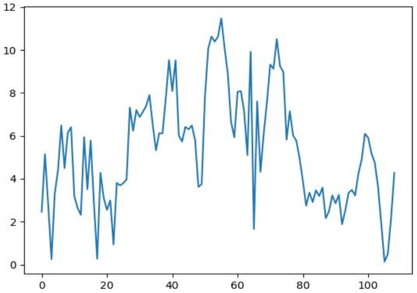
Graphs of importance for one video from different networks. Top left: large network trained for 1000 epochs, top right: small network trained for 500 epochs, bottom left: small network trained for 1000 epochs, bottom right: small network trained for 1000 epochs on shorter sequences

Worse Attention from Different Networks









Next Steps

- Better to get important subsequences
  - Classify importance based on degree of attention variation in short image sequence, likely more stable across different networks
- Use LIME[16] and classification of important or not based off thresholding attention variance to see why a sequence is important
- Use Deep Continuous Clustering[17] (or other unsupervised clustering technique) to group importance sequences for better understanding of what the network thinks is important

Limitations

- Only suitable for in-sample analysis, not a general model of cell migration
- Based entirely on videos so any features aren't necessarily meaningful from a proteomics or genomics perspective
    - And some of the model is memorizing changes that couldn't be predicted solely from observing previous motion
- Due to small dataset size, any features found aren't necessarily generalizable
- Still requires human interpretation of features either via direct analysis of important frames or via inspecting LIME[16] output
- Work is pre-publication, pre-peer review, and not quite finished

CIVIC DIGITAL FELLOWSHIP

Thank You To:

- John Prue - NIDCR CIO, Mentor
- Andrew Doyle - NIDCR CBS
- Ken Yamada - NIDCR CBS
- Rachel Dodell, Chris Kuang, Hillary Mclauchlin - Coding it Forward
- Jess Mazerik - NIH OD
- Everyone at NIDCR OIT

Citations

(In order of appearance)

1. Zhou, Victor. "Machine Learning For Beginners: An Introduction To Neural Networks". *Towards Data Science*, 2019, https://towardsdatascience.com/machine-learning-for-beginners-an-introduction-to-neural-networks-d49f22d238f9.
2. Stewart, Matthew. "Simple Introduction To Convolutional Neural Networks". *Towards Data Science*, 2019, https://towardsdatascience.com/simple-introduction-to-convolutional-neural-networks-cdf8d3077bac.
3. Nguyen, Michael. "Illustrated Guide To Recurrent Neural Networks". *Towards Data Science*, 2018, https://towardsdatascience.com/illustrated-guide-to-recurrent-neural-networks-79e5eb8049c9.
4. Inkawhich, Nathan. "DCGAN Tutorial — Pytorch Tutorials 1.1.0 Documentation". *Pytorch.Org*, 2019, https://pytorch.org/tutorials/beginner/dcgan_faces_tutorial.html.
5. Radford, Alec et al. "Unsupervised Representation Learning With Deep Convolutional Generative Adversarial Networks". *ICLR 2016*, 2016.
6. Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation* 9.8 (1997): 1735-1780.
7. Bahdanau, Dzmitry et al. "Neural Machine Translation by Jointly Learning to Align and Translate." *CoRR* abs/1409.0473 (2014): n. pag.
8. Weng, Lilian. "Attention? Attention!". *Lil'log*, 2018, https://lilianweng.github.io/lil-log/2018/06/24/attention-attention.html.
9. Luong, Thang et al. "Effective Approaches To Attention-Based Neural Machine Translation". *Proceedings Of The 2015 Conference On Empirical Methods In Natural Language Processing*, 2015, Accessed 1 Aug 2019.
10. Doersch, Carl. "Tutorial on Variational Autoencoders." *ArXiv* abs/1606.05908 (2016): n. pag.
11. Vaswani, Ashish et al. "Attention Is All You Need." *NIPS* (2017).
12. Ioffe, Sergey and Christian Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift." *ArXiv* abs/1502.03167 (2015): n. pag.
13. Vaswani, Ashish et al. "Attention Is All You Need." *NIPS* (2017).
14. Wu, Yuxin, and Kaiming He. "Group Normalization." Lecture Notes in Computer Science (2018): 3–19. Crossref. Web.
15. Wikipedia contributors. "Log probability." *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 24 Mar. 2019. Web. 1 Aug. 2019.
16. Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "'Why Should I Trust You?'" Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16 (2016): n. pag. Crossref. Web.
17. Shah, Sohil Atul and Vladlen Koltun. "Deep Continuous Clustering." *ArXiv* abs/1803.01449 (2018): n. pag.

and cell images courtesy of Andrew Doyle at NIDCR CBS