

Cara Van Uden

website: caravanuden.com
email: cvanuden@stanford.edu
github: @caravanuden

Education

MS in Computer Science

Stanford University | Sept 2021 - June 2023 (expected)

- **Concentration:** Artificial Intelligence
- **Teaching assistant:** Data Management and Data Systems (1 term)

BA in Computer Science & BA in Cognitive Science

Dartmouth College | Sept 2015 - June 2019

- **Thesis:** "Comparing brain-like representations learned by vanilla, residual, and recurrent CNN architectures". Oral defense with high honors.
- **GPA:** 3.87/4.00, summa cum laude
- **Teaching assistant:** Introduction to Computer Science (6 terms), Foundations of Applied Computer Science (1 term)

Papers

- **Van Uden, C. E.***, Nastase, S. A.*, Connolly, A. C., Feilong, M., Hansen, I., Gobbin, M. I., & Haxby, J. V. (2018). "Modeling semantic encoding in a common neural representational space." *Frontiers in Neuroscience*.

Projects

- Data quality validation (Python, PySpark, Deequ)
- Visual similarity product clustering for browse (Python, Faiss, BIRCH, PCA)
- Image classification API with ResNet50 (Python, Keras, Redis, Docker)
- Google text normalization with seq2seq LSTM encoder/decoder (Python, Keras)
- Spotify playlist recommendation with neural collaborative filtering and clustering (Python, Keras, Tensorflow)

Awards

- Wayfair Hackathon Finalist (2021)
- Wayfair Hackathon Winner (2019)
- Phi Beta Kappa (2019)
- High Honors Thesis in Computer Science (2019)
- Neukom Award for Outstanding Undergraduate Research in Computational Science (2019)
- Academic Award in Cognitive Science (2019)

Experience

Machine Learning Engineer II, Exact Matching

Wayfair | Feb 2020 - Aug 2021

- Built and deployed machine learning pipelines at scale (SQL, Google Cloud Storage, BigQuery, Python, PySpark, SparkML, Tensorflow, Docker, Airflow). Moved pipelines from on-premises to Google Cloud.
- Pipelines preprocess data, extract features from text and imagery, and perform model prediction for match generation.
- Was sole technical contributor for first six months. Pipelines generated \$180M annually in GRS during tenure on team.

Data Scientist I, Visual Similarity

Wayfair | Aug 2019 - Feb 2020

- Incorporated environmental imagery into the visually similar product recommendation pipeline (SQL, Python, Tensorflow, Keras, Airflow).
- Reduced product coverage gap by >50%. Decreased pipeline runtime by 20%.

Research Intern, Computational Cognitive Neuroscience

Dartmouth College | Jan 2018 - June 2019

- Used hyperalignment (fMRI data alignment technique) and forward encoding models to predict neural responses to naturalistic video stimuli across people (Python). Our models demonstrated improved spatial specificity and model performance compared to previous single- and between-subject methods (paper).
- Compared the representations learned by different CNN architectures to those of the human brain's ventral visual stream (Python, PyTorch). Found preliminary evidence that recurrent and deep residual CNNs learn more brain-like representations than feedforward models; achieved state-of-the-art neural response prediction performance in late-stage visual areas (thesis).

Research Intern, Center for Cognitive Brain Imaging

Carnegie Mellon University | Summer 2018

- Decoded the content of neural representations of emotion in open-source resting state fMRI data from 330 people (MATLAB).
- Found significant differences in duration and frequency of emotion among clinical groups compared to controls.

Research Intern, Biomedical Data Science

Dartmouth College | Jan 2016 - June 2017

- Developed a CNN/LSTM ensemble for estimating high-risk substance use from Instagram data. Built pipelines that extracted word and sentence embeddings from captions and comments (Torch).
- Ensemble was able to estimate the risk of alcohol abuse, and uncovered social media data characteristics that are associated with high-risk alcohol use.

Data Science Intern, Translational Data Science

Celgene | Summer 2016

- Built an exploratory data analytics and visualization tool for analyzing gene expression and drug response data (R, SQL).
- Used site-wide by scientists for exploratory target deconvolution and validation in translational drug development for blood and bone marrow cancers.