

Ben Wortman

Data Scientist

814.594.4981 | ben.wortman@gmail.com | ben-wortman.com | State College, PA

Education

Masters, Informatics | The Pennsylvania State University | 3.94 Dec 2022

B.S. Data Science, Math Minor | The Pennsylvania State University | 3.75 May 2020

Skills

Languages - Python, SQL, C, Scala, Matlab/Octave, SAS, R, HTML/CSS

Technologies and Select Libraries - AWS, Docker, Pytorch, Tensorflow, Pillow, Pandas, OpenCV, NLP, Scikit-learn, mTurk, Multiprocessing, Cluster Computing, Spark, Hadoop, Data Visualization, XAI, Shapely/LIME, Linux, Pickle, Time Series Analysis, Jyro, Git, VSCode, Jupyter Notebooks, excel (advanced)

Other Skills - Research, Leadership, Communication, Project Management, Public Speaking, Technical Writing

Experience

Data Engineer | AdPredictive June 2022 - Present

Handled ETL for data partners. Automated data ingestion into S3 and Redshift using Apache Airflow. Oversaw legacy systems to identify data leakage and debug runtime errors. Performed analysis on new potential data partners to assess the quality and value of their data in relation to our current offering.

Research Assistant | Penn State University Aug 2020 - May 2022

Automatic recognition of emotion in visual media using Deep Learning. Taught a module on high performance computing and the use of popular deep learning frameworks like Tensorflow and Keras, including a review of popular topics in machine learning so students can get a better idea of what is currently SOTA. Contributed to the writing of several proposals for submission to granting agencies and foundations. Technical lead for our group's motion capture lab.

Machine Learning Intern | Carnegie Mellon SEI AI Division May 2021 - Aug 2021

Conducted a comprehensive review on the current state of satellite object detection for the National Geospatial Agency. Developed a prototype interface for probabilistic object detection and localization for autonomous vehicles (see projects).

Data Engineer Intern | Impact Radius May 2019 - Feb 2020

Built enterprise solutions for determining media spend attribution from walled garden advertising platforms such as facebook and instagram (see projects). Was subsequently hired on part time at the conclusion of my internship to continue my work during the school year.

Sales Executive | Semrush Dec 2016 - Oct 2017

I spent a year working in software sales which taught me valuable lessons in communication and also gave me a healthy understanding of the business context behind decisions in application development.

Select Projects

Emotion Classification with Semantic Embedding Loss Jan 2022 - May 2022

Since emotion labels are not completely independent, I am using semantic embeddings as targets to encode additional information into the model's training. In addition to this I have trained a complementary model using transfer learning from a pretrained Resnet to make predictions on LMA components which have been shown to be correlated with emotion. I have also crowdsourced an additional dataset annotated for human interaction to be processed in a separate channel.

Image De-fencing with Synthetic Data and Adversarial Loss **Oct 2021 - Feb 2022**

For a class project I developed an end-to-end DL model for the simultaneous detection and inpainting of images obstructed by fences. This was achieved by overlaying raw images with existing fence masks and training to minimize SSIM, L2, and Adversarial loss. By doing so I was able to inpaint 80%+ of real occluded images with photorealistic results.

Identifying Gaps in Emotion Models Using Word Embeddings **Oct 2021 - May 2022**

By analyzing the annotations from several large scale emotion recognition datasets, I was able to identify limitations in existing emotion models common in affective computing. Then through the use of word embeddings from Facebook's FastText model, I was able to identify these gaps and produce an updated high coverage emotion model using UMAP reduction, a greedy search, and agglomerative clustering.

Interface for Quantifying Uncertainty in Autonomous Vehicles **May 2021 - Aug 2021**

Prototyped an interface for autonomous drone controllers for the US army. This included training bayesian object detectors and a monte carlo dropout implementation for depth estimation for the simultaneous detection, localization, and uncertainty estimation of objects.

Hidden Patch Attacks on Optical Flow **Jan 2021 - Jun 2021**

A white box adversarial patch attack on optical flow DNNs. I used gradient descent to minimize the similarity between the predicted optical flow in adjacent video frames. By adjusting the alpha value during training I was able to hide these adversarial patches from humans and train them invariant to background increasing the feasibility of this attack. Using zero flow tests I was able to show the effectiveness of these patches on DNN FlowNets.

Probabilistic Attribution for Walled Gardens **Jun 2019 - Sep 2019**

Matching walled garden Ad impressions to customer conversion data appended with census demographic data through a fuzzed decision tree algorithm. I developed a preprocessing step that identified 30% fraudulent TV promo code linked conversions that when filtered out using an SVM dramatically improved predictive performance and reason code certainty (Shapely values)

Publications and Presentations

Wortman, B. (2022). Models of Human Emotion and Artificial Emotional Intelligence. Preprint.

Wortman, B., Wang, J. Z. (2022) HICEM: A High-Coverage Emotion Model for Affective Computing. In Review.
<https://arxiv.org/abs/2206.07593>

Wortman, B. (2021). Hidden Patch Attacks for Optical Flow. ICML 2021 Workshop on Adversarial Machine Learning.
<https://openreview.net/forum?id=7Zc8KHNdih>

Grants

Amazon - \$300k - *Affective and Social Interaction between Human and Intelligent Machines* **FY 2021 - 2023**
PI: Dr. James Wang

XSEDE Allocation - \$40k - *Advancing Computational and Image Understanding Technologies* **FY 2021 - 2022**
PI: Dr. James Wang

Awards and Certifications

Certificate in Machine Learning | Stanford Online **2019**
This covered a broad range of topics in supervised learning, unsupervised learning, and best practices in machine learning

NASA PSGC Graduate Fellowship Award **2021 - 2022**