

# Mikhail Breslav

Redmond, WA

✉ [mikhail.breslav@gmail.com](mailto:mikhail.breslav@gmail.com)

## Summary

Applied Scientist (Software Engineer and Researcher) with a PhD in Computer Vision and Machine Learning and 5+ years of industry experience successfully translating research into impactful products. Most recently, developed a novel safety tool to validate the perception subsystem of our autonomous vehicle, resulting in significant cost savings and improved safety. Previously, designed and productionized a code recommendation engine utilized by millions of VS Code users, increasing developer productivity. Seeking a collaborative role in (or adjacent to) the application of Machine Learning to high-impact large-scale problems.

## Education

### Boston University

*Ph.D Computer Science, Advisor: Margrit Betke*

**Boston, MA**

*8/2010–9/2016*

### The Pennsylvania State University

*M.S Electrical Engineering, GPA 3.80/4.00*

*B.S Electrical Engineering, GPA 3.60/4.00*

**University Park, PA**

*8/2008–6/2010*

*8/2004–5/2008*

## Work Experience

### Cruise

*Senior Systems Engineer*

**Remote from WA**

*11/2023–Current*

- Led the implementation of a physics-based testing tool designed to identify safety critical failures in the perception subsystem of our [autonomous vehicles](#).
- Demonstrated tool's ability to reduce testing costs as well as perception development time.
- Enabled metrics (generated by the tool) to be consumed by integrating it into existing metric data pipelines.
- Designed closed course and simulated tests necessary for validating the safety of the AV in scenarios where agents are very close to the AV.
- Worked collaboratively to deliver test collections used to validate AV behavior and inform leadership of deployment risks.

### Microsoft

*Senior Applied Scientist*

**Redmond, WA**

*9/2019–7/2023*

- Initiated and led R&D for a code recommendation feature from prototype through production. The feature is now a part of the IntelliCode extension for [VS Code \(34M+ downloads\)](#) and is in preview for [Visual Studio](#).
- Developed data processing pipelines which ingest millions of source code files, parse them, extract key semantics, and aggregate results into a code recommendation model which gets deployed (Python).
- Leveraged Azure Data Factory and Azure Batch to run pipelines on a mixture of compute, including pools with 50+ nodes in order to parallelize workloads.
- Created and published Python packages to avoid duplicating code across activities.
- Collaborated on API design to enable data from the model to be served by a service (Python, Azure Kubernetes).
- Contributed to the client library (npm package & Typescript) consumed by clients like the IntelliCode extension.
- Implemented a greedy tree-based pattern mining algorithm to quickly find patterns in parse trees and leverage them to improve recommendation quality.
- Leveraged a vulnerability scanner to ensure recommended code snippets are secure.
- Analyzed customer telemetry and survey feedback to inform stakeholders of user engagement and sentiment. Proposed new improvements and directions to continue satisfying our customers.
- Benchmarked large language models from OpenAI (e.g Codex) on their ability to perform specific code generation tasks.

### HP Inc.

*Data Scientist & Research Engineer*

**Palo Alto, CA**

*9/2017–6/2019*

- Conceived of and prototyped a knowledge-based product recommendation engine.
- Researched the problem of rank aggregation and its solutions.
- Implemented the genetic algorithm in Python, C++, and Java for the special case where members are rankings.
- Analyzed data coming from HP devices using Spark and Databricks. Processed tables with millions of rows and visualized results using matplotlib and d3.js.
- Contributed a major bug fix to RankAggreg, the open source rank aggregation library for R.

## HP Inc.

Computer Vision Research & Development Engineer

Palo Alto, CA

9/2016–9/2017

- Evaluated CNN codes coupled with SVMs and Random Forests for object recognition problems where little training data was available.
- Developed algorithms for the evaluation and improvement of existing OCR engines which led to two granted patents.
- Developed C++ Qt-based modules allowing developers to access a new camera sensor.
- Improved performance of an existing segmentation algorithm by re-implementing it in CUDA.

## HERE (Formerly Nokia)

Intern

Berkeley, CA

5/2014–8/2014

- Used C++ to implement and evaluate geometric algorithms for generating depth maps from large ( $\approx 10^7$ ) lidar-based point clouds. Extended classes from the Point Cloud Library (PCL) to develop custom modules. Delivered a presentation showing the feasibility of a commercial application that uses the generated depth maps.

## MIT Lincoln Laboratory

Intern

Lexington, MA

6/2010–8/2010

- Used GNU Radio to implement an end to end communications system over software defined radios.

## Lockheed Martin

Intern

Owego, NY

6/2008–8/2008

- Wrote algorithms to integrate elevation data into a large Matlab-based signal location system.

## Research Experience

---

### Image & Video Computing Group, Boston University

Research Assistant

Boston, MA

8/2010–8/2016

Implemented and evaluated novel computer vision and machine learning based algorithms for the problem of 2D and 3D pose estimation of flying animals in multi-view datasets. Published and presented research results at leading peer-reviewed computer vision conferences.

### Deep Learning

- Demonstrated that deep learning could be applied to pose estimation of moths, a niche biology application where only small datasets exist (typically). Published a tutorial style paper for biologists to gain exposure to deep learning.
- Implemented transfer learning in Caffe where a VGG 16 network was adapted to output the 2D pose of a Moth.

### Graphical Models, Features, Clustering, and Linear Classifiers

- Developed an approach that improved upon traditional part based models, like Pictorial Structures, by leveraging parts that are automatically discovered from *unannotated* regions of training images.
- Gained experience with a variety of features (HOG, SIFT, SIFT variants based on BOW and Spatial Pyramid BOW), clustering algorithms (greedy, k-means, affinity propagation, discriminative clustering), and classifiers (LDA, SVM).
- Used Markov Random Fields to model the problem of 3D pose estimation over time. Generated 3D pose estimates of bats in low resolution videos.

### Camera Calibration and Multi-View Geometry

- Gained experience with traditional camera calibration approaches and have implemented routines to: estimate the fundamental and essential matrices from point correspondences, decompose the fundamental and essential matrices into pairs of camera matrices, and triangulate 2D points to obtain 3D reconstructions.
- Helped setup multiple thermal infrared cameras for a bat colony capture in Texas.

### Image Processing Lab, The Pennsylvania State University

Research Assistant

University Park, PA

8/2008–6/2010

Setup and operated a new Endobronchial Ultrasound System (EBUS) for our lab, and designed novel underwater experiments to evaluate existing 3D voxel based reconstruction algorithms.

## Computer Skills

---

**Programming:** Python, C++, SQL, Typescript, C, Java, C#, CUDA, LaTeX, Git, Vim      **OS:** Unix, Mac, Windows

**Libraries:** Pandas, Scipy, Numpy, OpenCV, PCL, Boost, OpenGL      **3D Graphics:** Blender, Unity3D

**Deep Learning:** Pytorch, TensorFlow, Keras      **Big Data:** Spark, Azure, AWS, Databricks

## Publications

---

D. Drain, C. Hu, C. Wu, **M. Breslav**, N. Sundaresan. "Generating Code with the Help of Retrieved Template Functions and Stack Overflow Answers". *arXiv* 2021.

**M. Breslav**, T. L. Hedrick, S. Sclaroff, M. Betke. "Automating Image Analysis by Annotating Landmarks with Deep Neural Networks". *arXiv* 2017.

**M. Breslav**, T. L. Hedrick, S. Sclaroff, M. Betke. "Discovering Useful Parts for Pose Estimation in Sparsely Annotated Datasets". *WACV* 2016.

**M. Breslav**, N. W. Fuller, S. Sclaroff, M. Betke. "3D Pose Estimation of Bats in the Wild". *WACV* 2014.

**M. Breslav**, N. W. Fuller, M. Betke. "Vision System for Wing Beat Analysis of Bats in the Wild". *ICPR* workshop, November 2012.

X. Zang, **M. Breslav**, W. E. Higgins. "3D Segmentation and Reconstruction of Endobronchial Ultrasound". *SPIE Medical Imaging*, 2013.

## Professional Service

---

CVPR Reviewer (2012 & 2019)

## Patents

---

Patent: 17894569, Software development context history operations (Pending: 2024)

Patent: 17219886, Source code generation using code templates with neural transformers (Pending: 2022)

Patent: 11328167, Optical Character Recognitions Via Consensus Of Datasets (Granted: 2022)

Patent: 11163993, Image Alignments Via Optical Character Recognition (Granted: 2021)

## Languages

---

**English:** Native Speaker   **Russian:** Fluent   **Spanish:** Basic