

Ali Arbabian

a.arbabian@gmail.com

aliarbabian.com
linkedin.com/in/a-arbabian

SUMMARY

Machine Learning Researcher with **6+ years of experience** building and deploying production ML systems. Currently working on representation learning at **Meta AI** for content similarity understanding. Previously led computer vision systems processing $\sim 100\text{M}$ frames daily in production retail environments and conducted applied research in medical imaging. ML focus areas include **deep learning representations for image/video/audio, self-supervised learning, and similarity search.**

WORK EXPERIENCE

Meta

Menlo Park, CA

Applied Research Scientist

June 2025 – Present

Similarity Understanding Team — building image, video, and audio similarity signals at Meta scale.

- **Similarity Search at Scale:** Develop foundational image, video, and audio representation models that support billions of content comparisons daily, serving as core signals for copyright detection, integrity enforcement, LLM/VLM training-data curation, and recommender-system originality scoring. Index representation vectors with ANN methods such as IVF, SQ, PQ, and HNSW, balancing speed, storage efficiency, and accuracy.
- **Self-Supervised Learning:** Train self-supervised image representations using contrastive and non-contrastive SSL objectives, including SimCLR, Barlow Twins, VICReg, and SIGReg. Study how entropy regularization affects retrieval performance through extensive experiments.
- **Image Retrieval with Global and Local Features:** Proposed a ground-up overhaul of the production image-matching system by jointly extracting global and local features in a single model. Improved computational efficiency by eliminating redundant feature extraction while addressing hard partial-match cases where global representations failed to capture salient visual details.
- **Retrieval Evaluation:** Develop evaluation methods for retrieval tasks with “needle-in-a-haystack” positives ($\sim 10^{-8}$ prevalence). Collaborate with data scientists on a two-stage sampling strategy that reduces measurement variance and labeling requirements from $O(n \times m)$ to $O(k)$ while preserving high-confidence estimates.

Zippin

San Carlos, CA

Senior Machine Learning Engineer

April 2023 – May 2025

Tracking Team — technical lead for person-tracking services in checkout-free retail platform.

- **Multi-Camera Person Tracking:** Designed, deployed, and scaled a distributed tracking system utilizing edge and cloud compute, processing live feeds across 50+ stores with 99.99% uptime ($\sim 100\text{M}$ frames/day) on GCP Kubernetes clusters.
- **Segmentation with Synthetic Data:** Fine-tuned SAM2 model for person detection with zero-shot generalization to unseen store layouts and camera views. Engineered cost-effective groundtruth generation combining a prompted video segmentation model with classic object detector outputs, scaled training data 5x (200K to 1M frames), validated with ablation study showing +6pp mAP improvement.

Machine Learning Engineer

January 2022 – April 2023

- **Distributed Task Queue:** Architected a distributed task queue for cross-camera ReID matching using Redis with a competing-consumers pattern, achieving $\sim 10,000$ messages/sec throughput with autoscaling workers.
- **Graph Visualization:** Designed and implemented a real-time graph visualization system for tracking service observability, significantly reducing MTTR for tracking issues.

Enlitic

Deep Learning Research Engineer

Applied research engineer for medical imaging AI.

San Francisco, CA

October 2020 – September 2021

- **Shortcut Learning Mitigation:** Discovered model was exploiting a spurious correlation of X-ray scanner type as a diagnostic prior due to acquisition bias in dataset; applied adversarial training of a discriminator module to reduce device predictability from 80+% to random chance while maintaining downstream diagnostic accuracy.
- **Multi-task Medical Vision:** Designed a custom U-Net-based CNN for simultaneous disease classification and localization across 50+ conditions, achieving 92% AUC-ROC with 30% faster inference; delivered the first U.S.-developed Chest X-ray ML model to receive regulatory approval to commercialize in Japan.
- **Batch Processing Pipeline:** Engineered an ETL pipeline for 10TB+ of medical imaging data with lossless LZ4 compression (3x faster compression, 8x faster decompression) and metadata-based dataset partitioning for reproducible research splits.

MetaOptima Technology

Machine Learning Engineer

Applied research engineer for a dermatology diagnostic platform.

Vancouver, BC

September 2019 – May 2020

- **Visual Search:** Designed a lightweight skin lesion visual search architecture compressed to 5MB with quantization, achieving 0.94 MRR with sub-20ms on-device inference across longitudinal patient records.
- **Color Constancy:** Developed a statistical color constancy algorithm to normalize longitudinal lesion images, improving clinician diagnostic consistency by 35–40% in blind A/B tests; optimized for browser and mobile inference.
- **Model Robustness:** Introduced a feature consistency training approach improving model robustness to common image corruptions (JPEG artifacts, noise, blur) by 27%, alongside a systematic evaluation framework for reliability across varying image quality conditions.

EDUCATION

University of British Columbia (UBC)

Bachelor of Science (B.Sc.) in Computer Science

Vancouver, BC

Graduated May 2019

SELECTED PROJECTS

cinematic

Semantic visual search engine for movie frames.

Web Application

September 2024 – December 2024

- Built a pipeline for scraping and indexing 150K movie frames into SQLite, with VLM-generated captions and LLM embeddings enabling natural language search queries. Experimented with CLIP-style models that align image and text embeddings but they struggled with OCR which was an important usecase.
- Optimized kNN search to sub-100ms query times across 150K frames using bit-quantized candidate vectors reranked with higher-precision vectors; entire app runs on a \$7/mo VPS.

Chess Theory Tree Visualizer

Tree visualization mapping chess matches to theoretical openings.

Data Visualization

October 2022

- Matched opening sequences from 83K professional chess games to a standardized ECO code, and visualized the state-space graph to identify popular lines and their counter-moves.

TECHNICAL SKILLS

- **ML & Data:** PyTorch (Distributed), TensorFlow, OpenCV, scikit-learn, Pandas, Polars, NumPy, matplotlib
- **Infrastructure:** GCP, AWS, Azure, Kubernetes, Docker, Helm, RabbitMQ, PubSub
- **Databases:** Redis, Faiss, PostgreSQL, MySQL, SQLite