Amar Ali-bey

Montréal, QC · Open to relocation
amar.alibey@gmail.com · (+1) 418-900-8418 · LinkedIn · GitHub · Google Scholar

ABOUT

ML Applied Scientist working on embeddings, re-ranking, and dense retrieval at scale. Deep expertise in designing, training, and shipping architectures end-to-end. Building the core search and video understanding stack at CoactiveAI. Background in image-retrieval, published influential and SOTA retrieval techniques (CVPR, WACV, BMVC).

EXPERIENCE

Machine Learning Applied Research Scientist · *CoactiveAI, Montreal, QC* Feb 2025 – Present

Own the search, ranking, and video understanding systems within the ML research team at a Series-B startup building multimodal search for enterprise media.

Key skills: Search, Reranking, Tagging and Calibration, Facial Recognition, Video Understanding, BM25/Vector DBs.

- **Semantic Search with Negation.** Invented a vector-arithmetic approach that allows exclusion in semantic search (e.g., querying "images of cars" while excluding "red cars") that runs naively on vector DBs, no re-ingestion and no re-ranking.
- **Hybrid Search & Calibration.** Scaled dense retrieval over tens of millions of records. Calibrated raw score distributions of pre-trained encoders (Sentence-Transformer and CLIP) for reliable cross-modal score fusion in production.
- **Reranker for Bias Removal.** Replaced an LLM-in-the-loop bias removal step (the largest inference bottleneck) with a lightweight cross-encoder reranker: up to 100$\times$ cost reduction with no measurable degradation.
- **Celebrity Recognition.** Built the full pipeline from scratch (detection, alignment, embedding, ANN index). 60$\times$ faster and more accurate than DeepFace, especially on underrepresented demographics. Runs as a core production service.
- **Video Understanding.** Redesigned legacy shot-by-shot pipeline with a hierarchical scene-level VLM/LLM architecture for long-form summarization. Developed an LLM-as-a-judge benchmark suite to quantify output improvements on movie data.

SELECTED PUBLICATIONS & OPEN SOURCE

BoQ: A Place is Worth a Bag of Learnable Queries CVPR 2024

First Author · A Transformer-based architecture that learns a set of global queries to aggregate features for visual retrieval. Still the SOTA method in visual place recognition two years after publication. [$\mathbf{90}+$ citations]

MixVPR: Feature Mixing for Visual Place Recognition WACV 2023

First Author · Lightweight feature-mixing global descriptor. Deployed in production by multiple companies for its extreme efficiency. [$\mathbf{340}+$ citations]

GSV-Cities: Toward Appropriate Supervised Visual Place Recognition Neurocomputing 2022

First Author · Large-scale training dataset for place recognition. $5,000+$ downloads; de facto standard for training supervised visual localization models. [$\mathbf{190}+$ citations]

Global Proxy-based Hard Mining for Visual Place Recognition BMVC 2022

First Author · Enabled training contrastive models using tiny batch-sizes. [$\mathbf{15}+$ citations]

Open Source Projects

- **OpenVPRLab**: end-to-end VPR training and benchmarking framework used by the community.
- **NanoCLIP**: highly efficient text-to-image retrieval model bridging DINOv2 and Sentence-Transformers via custom finetuning to produce ultra-compact embeddings.

800+ total GitHub stars across all open-source repositories, widely adopted in academic research and industry.

EDUCATION & ACADEMIC ACTIVITIES

Ph.D. in Computer Vision and Machine Learning 2024

Laval University, Québec, Canada

- **PhD thesis title:** *Deep Representation Learning for Visual Place Recognition.*
- Mentored M.Sc. students ; TA/Lecturer for 5 ML & Robotics courses ; Reviewer for CVPR, ECCV, NeurIPS, TPAMI, IROS.

B.S. in Computer Engineering 2014

École nationale Supérieure d'Informatique (ESI), Algiers, Algeria

- Valedictorian (1st / 200+). Ranked 4th in nationwide competition for PhD fellowship.

TECHNICAL DEPTH

Training: PyTorch, PyTorch Lightning, W&B, HuggingFace, Knowledge distillation, LoRA, ONNX/TensorRT, Quantization.
Serving & Infra: Anyscale, EC2/S3, HNSW, BM25, FAISS.
Languages: Python, C++, SQL · English, French.