

Brais Martinez

✉ brais.mart@gmail.com
www.braismartinez.org

Professional Experience

10/20 - Now **Group Leader/Principal Researcher,** *Samsung AI Center - Cambridge, UK.*

Focus on long-term research that acts as enabler for commercialization.

KPI: commercializations, patents, and top-tier publications.

Responsibilities:

- Technical leadership, focusing on Efficient Vision & Language (Visual Perception, visual LLMs, and Image Generation).
- Management of 4-6 persons.
- Steering strategy and collaborating within wider group (16 HC).

4/19 - 9/20 **Senior Researcher,** *Samsung AI Center - Cambridge.*

Worked on Video analysis and Binary Quantization as an individual contributor and research lead.

6/16 - 3/19 **Research Scientist,** *Amazon - Seattle.*

Part of the **Amazon Go** team, my work was to develop, validate and implement the algorithms behind Amazon's *Just Walk Out* technology, the core technology behind Amazon's brick-and-mortar checkout-free grocery shopping experience.

6/14 - 4/16 **Research Fellow,** *University of Nottingham.*

10/10 - 5/14 **Research Associate,** *Imperial College London.*

Education

06 - 10 **Ph.D. in Computer Science,** *Autonomous University of Barcelona, Spain.*

05 - 06 **MSc. in Computer Vision,** *Autonomous University of Barcelona, Spain.*

98 - 03 **B.S. in Mathematics,** *University of Santiago de Compostela, Spain.*

Projects:

2025 **Multi-modal Graph RAG.** Goal: build a multi-modal RAG feature that enables (V)QA across all documents in a computer. Target: deploy as a DesktopAI feature by end of 2025, possible expansion to internal cloud documents.

2025 **Human Inpainting for Eraser Feature.** Goal: Contribute to a Samsung Research eraser feature by developing a pipeline specific for improving quality and consistency for face, body, and hands during inpainting. Core model is diffusion-based. Target: beat Google solution and commercialize for next Galaxy Z (Flip/Fold). Feature readiness by end of Q1 2026.

- 2025 **On-device Image Gallery VQA** - demo (completed). Leveraged internal technology to compact images into discriminative representations (e.g., 16 tokens) during pre-indexing. At test-time, a RAG VQA ingests the compacted visual information + metadata, packing 40+ images in a 2048 input context length. Supports questions e.g. “Summarize my trip to Mexico”, “where did I see hieroglyphs?”.
- 2024 **Natural Language Gallery Search (S25 commercialization)**. Adds support for Natural Language queries when searching on the phone gallery. Uses a contrastively-trained Vision&Language model.
 - Presented at the Galaxy Unpacked 2025 event [YouTube link](#)
 - Marques Brownlee: “... finding an old photo by describing it is even more cool, and it actually works!” [YouTube link](#),
- 2024 **Gallery Photo Upscaler (S25 commercialization)**. Enables editing a gallery image by zooming in, and then saving at the original resolution. The upsampling relies on diffusion SR running on device (4K×3K under 3.5 sec on S24).
- 2023 **On device LLM**. Leverage distillation, on-device runtime optimizations, and on-device quantization (see EMNLP’24 MobileQuant).

Selected Publications (Full list: [Google Scholar](#))

Image Generation:

- CVPR’25 I. Hadji, M. Noroozi, V. Escorcia, A. Zaganidis, **B. Martinez**, G. Tzimiropoulos, “Edge-SD-SR: Low Latency and Parameter Efficient On-device Super-Resolution with Stable Diffusion via Bidirectional Conditioning”, *IEEE Conference on Computer Vision and Pattern Recognition*, 2025.
- CVPR’25 H. Yang, A. Bulat, I. Hadji, H.X. Pham, X. Zhu, G. Tzimiropoulos, **B. Martinez**, “FAM Diffusion: Frequency and Attention Modulation for High-Resolution Image Generation with Stable Diffusion”, *IEEE Conference on Computer Vision and Pattern Recognition*, 2025.
- ECCV’24 M. Noroozi, I. Hadji, **B. Martinez**, A. Bulat, G. Tzimiropoulos, “You Only Need One Step: Fast Super-Resolution with Stable Diffusion via Scale Distillation”, *European Conference on Computer Vision*, 2024.

vLLMs / LLMs:

- CVPR’25 Y. Ouali, A. Bulat, A. Xenos, A. Zaganidis, I. Maniadis Metaxas, **B. Martinez**, G. Tzimiropoulos, “VladVA: Discriminative Fine-tuning of LVLMs”, *IEEE Conference on Computer Vision and Pattern Recognition*, 2025.
- NeurIPS’24 X. Xu, M. Kim, R. Lee, T. Hospedales, **B. Martinez**, “A Bayesian Approach to Data Point Selection”, *Neural Information Processing Systems*, 2024.
- ECCV’24 Y. Ouali, A. Bulat, **B. Martinez**, G. Tzimiropoulos, “CLIP-DPO: Vision-Language Models as a Source of Preference for Improved Vision-LLMs”, *European Conference on Computer Vision*, 2024.

- EACL'24 H.X. Pham, I. Hadji, X. Xu, Z. Degutyte, J. Rainey, E. Kazakos, A. Fazly, G. Tzimiropoulos, **B. Martinez**, "Graph Guided Question Answer Generation for Procedural Question-Answering". *European Chapter of the Association for Computational Linguistics*, 2024 (Oral).

Visual Perception:

- ECCV'24 Y. Ouali, A. Bulat, **B. Martinez**, G. Tzimiropoulos, "CLIP-DPO: Vision-Language Models as a Source of Preference for Improved Vision-LLMs", *European Conference on Computer Vision*, 2024.
- EMNLP'24 A. Bulat, Y. Ouali, R. Guerrero, **B. Martinez**, G. Tzimiropoulos, "Efficient Vision-Language pre-training via domain-specific learning for human activities", *Empirical Methods in Natural Language Processing*, 2024.
- ICCV'23 Y. Ouali, A. Bulat, **B. Martinez**, G. Tzimiropoulos, "Black Box Few-Shot Adaptation for Vision-Language models". *International Conference on Computer Vision*, 2023.
- ICCV'23 A. Bulat, E. Lozano, **B. Martinez**, G. Tzimiropoulos, "ReGen: A good Generative zero-shot video classifier should be Rewarded". *International Conference on Computer Vision*, 2023.
- ICCV'23 M. Derakhshani, E. Sanchez, A. Bulat, VG. Turrise, C. Snoek, G. Tzimiropoulos, **B. Martinez**, "Bayesian Prompt Learning for Image-Language Model Generalization". *International Conference on Computer Vision*, 2023.
- ECCV'22 N. Dvornik, I. Hadji, H. Pham, D. Bhatt, **B. Martinez**, A. Fazly, A. Jepson, "Graph2Vid: Flow graph to Video Grounding for Weakly-supervised Multi-Step Localization". *European Conference on Computer Vision (Oral)*, 2022.

Efficiency (Architectures, Quantization, Distillation):

- EMNLP'24 F. Tan, R. Lee, L. Dudziak, S. X. Hu, S. Bhattacharya, T. Hospedales, G. Tzimiropoulos, **B. Martinez**, "MobileQuant: Mobile-friendly Quantization for On-device Language Models", *Empirical Methods in Natural Language Processing*, 2024.
- IJCV'24 J. Yang, X. Zhu, A. Bulat, **B. Martinez**, G. Tzimiropoulos, "Knowledge Distillation Meets Open-Set Semi-supervised Learning", *International Journal of Computer Vision*, 2024.
- ECCV'22 J. Pan, A. Bulat, F. Tan, X. Zhu, L. Dudziak, H. Li, G. Tzimiropoulos, **B. Martinez**, "EdgeViTs: Competing Light-weight CNNs on Mobile Devices with Vision Transformers". *European Conference on Computer Vision*, 2022.
- ICLR'21 J. Yang, **B. Martinez**, A. Bulat, G. Tzimiropoulos "Knowledge distillation via softmax regression representation learning", *International Conference on Learning Representations*, 2021.
- ICLR'21 A. Bulat, **B. Martinez**, G. Tzimiropoulos, "High-Capacity Expert Binary Networks", *International Conference on Learning Representations*, 2021.

NeurIPS'21 A. Bulat, J. Perez-Rua, S. Sudhakaran, **B. Martinez**, G. Tzimiropoulos, "Space-time Mixing Attention for Video Transformer", *Neural Information Processing Systems*, 2021.

ICLR'20 **B. Martinez**, J. Yang, A. Bulat, G. Tzimiropoulos, "Training binary neural networks with real-to-binary convolutions", *International Conference on Learning Representations*, 2020.

Service

Area Chair: ICCV'21, WACV'22.

Reviewer: CVPR, ICCV, ECCV, NeurIPS, ICML, ICLR, TPAMI, etc.

Outstanding CVPR'25,'21,'20, AAAI'20, NeurIPS'20, ICLR'22.
Reviewer: