# Jesús M. Rodríguez-de-Vera

PhD Candidate, Mathematics & Computer Science | University of Barcelona

jesusmolrdv@gmail.com | jmrodriguezdevera.com | /in/jesusmolrdv | jesusmolrdv | Google Scholar

*My research in computer vision and multimodal deep learning asks what knowledge a model is actually missing to succeed at a given task, then builds whatever closes that gap. That question took me from fine-grained recognition under long-tailed data (Dining on Details, LOFI) to generating the domain-specific data models were missing in the first place: in Precision at Scale, a collaboration with NVIDIA's AI Technology Center, we generated million-image real and synthetic domain datasets that beat ImageNet-1K pretraining by up to 21%. I also co-develop unified visual representation and generation models (CVPR 2026, WACV 2026).*

## EDUCATION

- **PhD in Mathematics & Computer Science (FPU Fellow, ranked #1 in Spain)** *Oct 2023 – May 2028 (Expected)*
  *Supervisors: Prof. Petia Radeva [] & Dr. Bhalaji Nagarajan []* University of Barcelona
- **M.S. in Artificial Intelligence (GPA: 9.51/10)** *Sep 2021 – Jan 2023*
  *Ranked #2 of 36 students* Universitat Politècnica de Catalunya
- **B.S. in Computer Science (GPA: 9.15/10) & B.S. in Mathematics (GPA: 8.38/10)** *Sep 2015 – Jul 2020*
  *Computer Science: Award for Academic Excellence, #2 of 96 graduates* University of Murcia

**Selected schools:** ELLIS Summer School on Autonomous Driving, Barcelona (2026) · ELLIS Winter School on Foundation Models (FoMo), Amsterdam (2026) · International Computer Vision Summer School (2023)

## RESEARCH EXPERIENCE

- **University of Barcelona, AIBA Research Group & Computer Vision Center** *Mar 2022 – Present*
  *Deep Learning Researcher* Barcelona, Spain
  - Co-led Precision at Scale with the **NVIDIA AI Technology Center (NVAITC)**: an LLM, VLM and diffusion pipeline (LLM concepts, Re-LAION-5B retrieval, TensorRT-accelerated Stable Diffusion, deduplication, OOD filtering) that builds million-image domain datasets ( code). Models and datasets released on Hugging Face; 400K PaS-Food image–caption pairs serve as the training set of the Dishcovery challenges at CVPR 2025 and 2026.
  - Co-developed unified self-supervised generative–discriminative models on VQ tokens: synthetic positive pairs sampled from the model's own token predictions **(WACV 2026)**, and generative tokens aligned with a DINOv2 semantic codebook **(CVPR 2026)**.
- **Brown University, SCALE Lab** *Sep 2025 – Nov 2025*
  *Visiting Research Fellow (3-month research stay; host: **Prof. Sherief Reda**), funded by the FPU mobility grant* Providence, RI, USA
  - Prototyped an agentic multimodal-LLM pipeline for driving-scene understanding (Qwen-VL calling detection and segmentation experts as tools, PEFT-adapted), benchmarked on DriveBench and OmniDrive.
- **Biyectiva Technology - AI Startup** *Feb 2019 – Dec 2021*
  *Computer Vision Developer (Part-time)* Murcia, Spain
  - Shipped real-time object detection for Android and led the computer-vision analysis system of the Go5D biomechanics product (3D scanners, infrared cameras), with a FIFA Medical Centre of Excellence.

## PUBLICATIONS

C=Conference, CO=Conference (Oral), J=Journal, S=Under Review, †=Equal Contribution

**[S.1]** Estepa, I. G., Nagarajan, B., **Rodríguez-de-Vera, J.M.**, Radeva, P., & Asano, Y. M. (2026). Conditioned to See: Contrast-Induced Elicitation Latents for Visual Perception in Image Diffusion Models. *Under review*.

**[C.1]** Estepa, I. G., **Rodríguez-de-Vera, J.M.**, Nagarajan, B., & Radeva, P. (2026). Learning from Semantic Dictionaries: Discriminative Codebook Contrastive Learning for Unified Visual Representation and Generation. *CVPR 2026*.
**Contribution:** *Aligns VQ tokens with a DINOv2 codebook: 76.7% linear probe, FID 9.62, 48.7% faster training than MAGE.*

**[C.2]** Estepa, I. G.†, **Rodríguez-de-Vera, J.M.**†, Sarasúa, I., Nagarajan, B., & Radeva, P. (2026). Conjuring Positive Pairs for Efficient Unification of Representation Learning and Image Synthesis. *WACV 2026 (Round 1 Early Accept)*.
**Contribution:** *Self-generated synthetic positive pairs; beats MAGE (FID 9.61 vs. 11.1) with 60.8% fewer H100 GPU-hours.*

**[J.1]** **Rodríguez-de-Vera, J.M.**†, Estepa, I. G.†, Sarasúa, I., Nagarajan, B., & Radeva, P. (2026). Precision at scale: Domain-specific datasets on-demand. *Pattern Recognition*, 171, 112236. [code]
**Contribution:** ***[RES Outstanding Paper Award 2025]*** *Up to +21% over ImageNet-1K pretraining; +3.3% fine-tuning over the order-of-magnitude larger IN-21K.*

**[C.3]** **Rodríguez-de-Vera, J.M.**, Estepa, I. G., Nagarajan, B., & Radeva, P. (2025). CLIP-DoRA: Weight-Decomposed Low-Rank Adaptation for Efficient Vision-Language Models. In *Deep Learning Innovations and Applications Workshop at IJCNN 2025*.
**Contribution:** *New few-shot state of the art across 11 datasets; competitive medical image segmentation with 1.5% of the trainable parameters.*

**[CO.4]** **Rodríguez-de-Vera, J.M.**, Estepa, I. G., Bolaños, M., Nagarajan, B., & Radeva, P. (2024). LOFI: LOng-tailed FIne-Grained Network for Food Recognition. In *1st MetaFood Workshop at CVPR 2024*.
**Contribution:** *Long-tailed detection and segmentation; +4.6 mAP over the previous best on UECFood-256.*

**[CO.5]** **Rodríguez-de-Vera, J.M.**, Villacorta, P., Estepa, I.G., Bolaños, M., Sarasúa, I., Nagarajan, B., & Radeva, P. (2023). Dining on Details: LLM-Guided Expert Networks for Fine-Grained Food Recognition. In *8th MADiMa Workshop at ACMMM 2023*.
**Contribution:** *Expert heads grouped by BLIP-2/LLM descriptions; state of the art on Food-101 without extra training data.*

**[C.6]** Estepa, I. G., **Rodríguez-de-Vera, J.M.**, Nagarajan, B., & Radeva, P. (2023). Good Fences Make Good Neighbours. In *4th VIPriors Workshop at ICCV 2023*.
**Contribution:** *Synthetic latent interpolations replace unreliable neighbour positives in contrastive SSL; beats NNCLR.*

**[J.2]** **Rodríguez-de-Vera, J.M.**, Bernabé, G., García, J.M., Saura, D., & González-Carrillo, J. (2022). Left ventricular non-compaction cardiomyopathy automatic diagnosis using a deep learning approach. *Computer Methods and Programs in Biomedicine*.
**Contribution:** *Deep-learning LVNC diagnosis: AUC 0.94 on 277 patients.*

**[J.3]** **Rodríguez-de-Vera, J.M.**, González-Carrillo, J., García, J.M., & Bernabé, G. (2021). Deploying deep learning approaches to left ventricular non-compaction measurement. *The Journal of Supercomputing*.
**Contribution:** *A full 15-slice MRI study segmented in under 0.1 s on an NVIDIA RTX 2080 Ti.*

## TALKS & PRESENTATIONS

- **Invited talk:** *Precision at Scale* (RES Outstanding Paper Award), plus a WACV 2026 poster. JURES 2026, Madrid.
- **Orals:** *LOFI*, MetaFood Workshop @ CVPR 2024; *Dining on Details*, MADiMa Workshop @ ACMMM 2023.
- **Posters:** *Learning from Semantic Dictionaries*, CVPR 2026; DLBCN 2023–2025; ACMCV 2023–2024.

## HONORS AND AWARDS

- **RES Outstanding Paper Award (2025):** Spanish Supercomputing Network, for Precision at Scale.
- **FPU Fellowship (2023):** Spanish Ministry of Universities; **ranked #1 nationwide**.
- **Best Essay Award (2023):** ICVSS; **#1** of ∼25 teams (∼100 selected PhD students).
- **Spanish Society for Academic Excellence (2021–):** member; **top 1–2%** of graduates in Spain.

## TECHNICAL SKILLS

- **Computer Vision, Multimodal LLMs & Generative Models:** Self-supervised learning; VLMs/MLLMs (CLIP, BLIP-2, Qwen-VL) and PEFT (LoRA/DoRA); diffusion (FLUX, Qwen-Image, Stable Diffusion) and masked token generation (MAGE/VQGAN); detection, segmentation, long-tailed recognition.
- **Synthetic Data Generation & Data Curation:** LLM, VLM and diffusion pipelines for million-image real and synthetic datasets; curation at scale (Re-LAION-5B retrieval, dedup, OOD filtering); VLMEvalKit.
- **Compute & Efficiency:** Multi-node training on up to 32 **NVIDIA H100/A100** GPUs (MareNostrum 5, Leonardo); built the group's **SLURM** GPU cluster (A40/H100) from scratch; **vLLM** serving (FP8, tensor parallel); **TensorRT**, **FlashAttention-v2**, **RAPIDS cuML**; C++ PyTorch operators (merged into Forest R-CNN); **Python**, PyTorch.
- **Languages:** Spanish (Native), English (C2 Proficient)

## TEACHING, MENTORING & SERVICE

- **Mentoring:** MSc student on uncertainty-guided LoRA for ViTs, with Dr. B. Nagarajan (2026).
- **Organizing:** Dishcovery Challenge @ **CVPR** 2025 & 2026 (MetaFood; **1.5K+ submissions**); UB International Winter & Summer Schools "Demystifying AI" (2025).
- **Teaching:** Computer Vision (BSc, UB, 2024); Transformers (MSc guest lecture); Generative AI (UB schools, 2025).
- **Reviewer (2023–present):** NeurIPS, CVPR, ICCV, ECCV, WACV, IJCNN, IBPRIA, Pattern Recognition, IEEE TMM.