Marco Mistretta

PhD Student in Artificial Intelligence

Florence, Italy
+39 327 4716224
marco.mistretta@unifi.it
github.com/marcomistretta
marco-mistretta-0b02a021a

Professional Summary

PhD researcher in *Artificial Intelligence* at *MICC, University of Florence*, under the **supervision** of **Prof. Andrew D. Bagdanov** and **Prof. Marco Bertini**. Published as **first author in top-tier conferences** such as *ICLR 2025 (main conference)*, *ECCV 2024 (main conference)* and *NeurIPS 2023 (workshop)*. Research focuses on *Computer Vision*, *Multimodal Vision-Language Models*, and *Incremental Learning*. Seeking a **2025 internship** to contribute to innovative teams and apply expertise to real-world challenges.

Education

Nov 2023 – Present **PhD Student in Artificial Intelligence**, *University of Florence, Italy*
Topic: Multimodal Vision-Language Models, Incremental Learning, Prompt Learning.
Supervisors: Prof. Andrew D. Bagdanov, Prof. Marco Bertini

Sep 2021 – Jul 2023 **Master's in Artificial Intelligence**, *University of Florence, Italy*
Thesis: "RE-Tune - Incremental Fine-Tuning of Biomedical Vision-Language Models"
Supervisors: Prof. Andrew D. Bagdanov, Prof. Marco Bertini Final Grade: 110/110L

Sep 2018 – Sep 2021 **Bachelor's in Computer Engineering**, *University of Florence, Italy*
Thesis: "Scarlatti-Gen - AI-Driven Sonata Generation Using Weighted Graphs and CNNs"
Supervisors: Prof. Paolo Frasconi, Prof. Simone Conforti Final Grade: 105/110

First-Author Publications

- **Mistretta M.***, Baldrati A.*, Agnolucci L.*, Bertini M., Bagdanov A. D. (2025). "**Cross the Gap**: *Inter-modal CLIP Representations Are Superior for Intra-modal Tasks*." (**ICLR 2025 main conference**). We show that the common practice of individually exploiting the text or image encoders of VLMs is highly suboptimal for intra-modal tasks like image-to-image retrieval. We argue that this is inherently due to the CLIP-style inter-modal contrastive loss that does not enforce any intra-modal constraints, leading to what we call *intra-modal misalignment*.
Code: github.com/miccunifi/Cross-the-Gap **Paper**: https://openreview.net/forum?id=VVVfuIcmKR
- **Mistretta M.***, Baldrati A.*, Bertini M., Bagdanov A. D. (2024). "**KDPL**: *Improving Zero-shot Generalization of Learned Prompts via Unsupervised Knowledge Distillation*." (**ECCV 2024 main conference**). Proposed KDPL, a novel unsupervised approach for improving the generalization of learned prompts without labeled data, achieving significant performance gains across multiple benchmarks.
Code: github.com/miccunifi/KDPL **Paper**: arxiv.org/abs/2407.03056
- **Mistretta M.**, Bagdanov A. D. (2023). **RE-tune**: "*Incremental Fine Tuning of Biomedical Vision-Language Models for Multi-label Chest X-ray Classification*." (**Workshop *Medical Imaging NeurIPS***). Introduced RE-tune, a novel approach for fine-tuning Multimodal Biomedical VLMs in incremental learning scenarios for multi-label chest disease diagnosis, achieving efficient, privacy-preserving classification.
Website: neurips.cc/virtual/2023/82487 **Paper**: arxiv.org/abs/2410.17827

Reviewer Roles

- **Conferences**: NeurIPS 2024, ICLR 2025
- **Journal**: IEEE Transactions on Multimedia

Research Interests

- **Multimodal Learning**: Combining vision and language for richer model understanding.
- **Prompt Learning**: Optimizing prompts to effectively guide AI models for improved task performance.
- **Few-Shot Learning**: Adapting AI to new tasks with limited examples.
- **Incremental Learning**: Enabling models to continuously learn without forgetting previous knowledge.
- **Contrastive Self-Supervised Learning**: Extracting insights from unlabeled data.

Teaching and Mentoring

Jan 2024 – Present
Teaching Assistant, *University of Florence*
- Delivering interactive lessons on C/C++ and Python to over 200 bachelor students.

Apr 2024 – Sep2024
Thesis Co-Supervisor, *University of Florence*
- *"Mitigating Catastrophic Zero-shot Forgetting in CLIP via Distillation of Low-Rank Adapters from Learned Prompts"*, Proposed a novel method to efficiently few-shots fine-tune CLIP models that mitigates catastrophic forgetting and preserves zero-shot capabilities, based on distilling learned prompts in LoRa adapters.

Jan 2020 – Nov 2020
Student Ambassador, *University of Florence*
- Mentoring students on exams projects, internships, and career development.

Technical Skills

Programming	Python, Java, C++, MATLAB, R
Frameworks	PyTorch, TensorFlow, NumPy, OpenCV
Miscellanous	Git, Docker, Bash
	GitHub Profile: github.com/marcomistretta
	KDPL Repository: github.com/miccunifi/KDPL

Personal Interests

Music
- **Flute Player**: Played transverse flute for 10 years in local bands and orchestras.
- **Chorister** at "Coro Giovanile Effetti Sonori" (National Competition Winner)
- **Guitarist and Songwriter** for "I Green Clouds" (Provincial Competition Winner)

Gaming: Enjoying strategic and cooperative games that enhance problem-solving and teamwork.

Languages

English	Professional working proficiency
Italian	Native