Federico Girella, Ph.D. student

Povegliano Veronese, 37064, Verona, Italy

federico.girella@univr.it
@FedericoGirella
Linkedin
federicogirella.github.io
Google Scholar

About Me

I am a third-year Ph.D. student at the University of Verona (Italy), supervised by Prof. Marco Cristani, with expected graduation in May 2026.
I research joint representations in the Image and Language multi-modal domain, working with deep neural networks such as (Large) Vision and Language Models and Text-to-Image Generative Models. My main body of work focuses on Text-to-Image Retrieval and Generation in the Fashion domain.
During my Ph.D. studies, I have collaborated with Humatics SYS-DAT (Italy), developing AI tools for the fashion industry. I also had the pleasure to join Prof. Yi-Zhe Song at the University of Surrey (United Kingdom), tackling the abstraction gap in sketch and language conditioning for image generation.
In October 2025, I was awarded an oral presentation at the International Conference on Computer Vision 2025 (ICCV25) in Honolulu, Hawaii. Currently looking for full-time opportunities.

Education

Apr. 2025 – Sep. 2025 — **Visiting Ph.D., University of Surrey**, United Kingdom.

2022 – May 2026 — **Ph.D. Computer Science, University of Verona**, Artificial Intelligence.
AI for Vision and Language. Generative Models.

2019 – 2022 — **M.Sc. Computer Science, University of Verona**, Visual Computing.
Thesis title: *Multi-Task Learning: Pose Estimation for Video Analytics*.
Final grade: 110/110L

2016 – 2019 — **B.Sc. Computer Science, University of Verona**, Computer Science and Eng.
Thesis title: *Study on Augmented Reality Functionalities*.
Final grade: 110/110

Publications (full list on Google Scholar)

1. **F. Girella**, D. Talon, Z. Lie, Z. Ruan, Y. Wang, and M. Cristani, "[Oral] LOTS of Fashion! Multi-Conditioning for Image Generation via Sketch-Text Pairing," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025.
2. D. Talon*, **F. Girella***, Z. Liu, M. Cristani, and Y. Wang, "Seeing the Abstract: Translating the Abstract Language for Vision Language Models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025.
3. L. Capogrosso*, **F. Girella***, F. Taioli*, *et al.*, "Diffusion-Based Image Generation for In-Distribution Data Augmentation in Surface Defect Detection," in *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2024.
4. **F. Girella**, Z. Liu, F. Fummi, F. Setti, M. Cristani, and L. Capogrosso, "Leveraging Latent Diffusion Models for Training-Free In-Distribution Data Augmentation for Surface Defect Detection," *21st International Conference on Content-based Multimedia Indexing*, 2024.
5. F. Taioli*, F. Cunico*, **F. Girella***, R. Bologna*, A. Farinelli, and M. Cristani, "Language-enhanced RNR-map: Querying renderable neural radiance field maps with natural language," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023.

Applied Projects

2022-2024 **Humatics SYS-DAT** I collaborated with Humatics SYS-DAT, implementing Deep Learning models as tools for the fashion industry. Some examples include a sales forecasting model used to predict the sales volume of unreleased fashion items, and a specialized text-to-image retrieval system.
Technologies used: *Diffusion Models, Transformers, Vision and Language Models.*

2022-2023 **SAFEPLACE Project** This project aims to boost the integration and development of innovative solutions in the field of IoT Systems for healthy and sustainable living environments. My contribution resulted in the development of Computer Vision systems for the detection of crowds and face masks in challenging environments, such as shops with difficult camera angles and outdoor scenarios. These systems were deployed in real-world locations and are in use to this day.
Technologies used: *Object Detection models.*

Skills

Languages Strong reading, writing, and speaking competencies for Italian and English.

Coding Python, PyTorch, High-Performance Computing training (SLURM), Web Development.

Research Image Generation, Vision and Language Models, Text-to-Image Models, Image-to-Image Models, Diffusion Models, Diffusion Transformer Models.

Miscellaneous Experience

Certifications

2016 **Certified Level C1 in English**. Awarded by Cambridge School of English.