

# Simone Rossetti

Senior Deep Learning Scientist • PhD in Computer Science Engineering • Startup Founder

📍 Rome, Italy | @ simone.rossetti@live.com | ☎ +39 3491059384 | LinkedIn | GitHub | Website

## SUMMARY

Specialist in *multimodal feature learning, vision language alignment, and structured visual perception*. Over four years of experience leading AI research teams and developing scalable, research-grade and production-ready models. Published at *top-tier venues* and involved in *EU-funded multidisciplinary research*. Strong focus on bridging theory and application through *weakly- and self-supervised learning, large-scale training, and multimodal system design*, with growing interest in *Vision Language Action models and agent-oriented AI*.

## PROFESSIONAL ACHIEVEMENTS

- Co-founded an *AI research startup* in *agri-tech*, leading the development of *multimodal, agent-oriented AI systems* for *micro-farming management and decision support*.
- Led and mentored a *multidisciplinary R&D team*, defining *roadmaps* and establishing *reproducible ML workflows, dataset curation practices, and evaluation protocols*.
- Designed and optimized *large-scale vision and vision-language models* on *multi-GPU infrastructures*, supporting *experimentation, benchmarking, and efficient fine-tuning*.
- Built *end-to-end multimodal data pipelines*, spanning *dataset engineering, experiment tracking, and foundation model evaluation*.
- Co-authored *EU-funded research proposals* in *sustainable agri-tech and smart automation*.
- Published *first-author research* at *top-tier venues* including *NeurIPS, ICCV, and ECCV* on *weakly supervised, self-supervised, and structure-aware semantic segmentation*.

## PROFESSIONAL EXPERIENCES

### Co-Founder & AI Research Engineer

Sep 2021 - Present

DeepPlants S.r.l. • *AI research startup in agri-tech and intelligent automation*

### AI Research Fellow

Jan 2021 - Oct 2021

Sapienza Università di Roma • *Research grant at DIAG, focused on computer vision and AI*

### ICT Application Developer

Jun 2019 - Apr 2020

VIK School S.r.l. • *Development of accessible digital learning platforms*

## EDUCATION & TRAINING

### PhD in Computer Science Engineering

Nov 2021 - Jan 2025

Sapienza Università di Roma • *GPA: N/A • Final grade: Excellent • Advisors: Pirri F.; Amerini I.*

### MSc in Artificial Intelligence and Robotics

Oct 2019 - Oct 2021

Sapienza Università di Roma • *GPA: 28.2/30 • Final grade: 110/110 Cum Laude*

### BSc in Computer Engineering

Oct 2015 - Mar 2019

Università degli Studi Roma Tre • *GPA: 23.3/30 • Final grade: 95/110*

## ACADEMIC CONTRIBUTIONS

- Rossetti, S., Gatti, P., Palleschi, D. (2025-2026, Ongoing). *CABBAGE: Comprehensive Agricultural Benchmark Backed by AI-Guided Evaluation*.
- Rossetti, S. (2025, UNITesi). *Reducing supervision in semantic segmentation through advancements in bayesian prior modelling*.
- Rossetti, S., Pirri, F. (2024, NeurIPS). *Unsupervised Hierarchy-Agnostic Segmentation: Parsing Semantic Image Structure*.
- Rossetti, S., Samà, N., Pirri, F. (2023, arXiv). *Removing supervision in semantic segmentation with local-global matching and area balancing*.
- Samà, N., David, E., Rossetti, S. et al. (2023, ICCV). *A new large dataset and a transfer learning methodology for plant phenotyping in Vertical Farms*.
- Rossetti, S. et al. (2022, ECCV). *Max pooling with vision transformers reconciles class and shape in weakly supervised semantic segmentation*.
- Rossetti, S., Zharkynbek, T., Pirri, F. (2021, YouTubeVOS). *Video Instance segmentation Challenge 2021 with YoloV4+1Tr*.

## LANGUAGES

- Italian - Native (C2)
- English - Fluent (C1)

## TECHNICAL SKILLS

### Expertise Areas

Multimodal feature learning, vision language alignment and grounding, weakly- and self-supervised learning, structured visual perception, semantic and instance segmentation, foundation model benchmarking

### Vision & Multimodal Models

Vision Transformers, vision language models and contrastive pretraining, segmentation foundation models, multimodal encoders and decoders, masked autoencoding, contrastive and clustering-based learning, efficient fine-tuning and distillation

### Language & Agentic Models

Large language models and encoder-decoder architectures, multimodal prompting and instruction tuning, vision language reasoning, tool-augmented and agent-oriented model design, retrieval-augmented pipelines

### Training, Scaling & Optimization

Large-scale multimodal training, distributed training, multi-GPU optimization, scalable inference, experiment tracking, evaluation protocols, reproducibility-oriented research workflows

### Engineering & Research Tooling

Python, PyTorch and Lightning, Hugging Face ecosystem, Docker, Linux, Git, SQL, multi-GPU environments, dataset curation and pipeline engineering

## VENUES

### Poster Presentations

NeurIPS '24, ICCV '23, ECCV '22

### Networking & Partnerships

EU R&I Days '25

### Advanced Trainings

ICVSS '22, DeepLearn '22