

Alessio Tonioni, Ph.D.

 alessio.tonioni.91@gmail.com
 alessio-tonioni

 <https://alessiotonioni.github.io/>



Employment History

- 2019 – Now ◇ **Research Scientist Manager**, Google Zurich.
Part of Google XR, developing ML technologies for products with a focus on computer vision and multi-modality. Developed and launched several ML models in production (e.g., in Google Lens, Drive, Translate) and part of the development of experimental and upcoming products (e.g., Gemini Live for XR devices). Experience as: IC, TL and people manager. Currently managing 7 people between research scientists and software engineers.
Internal roles:
- Jul. 2019 - Dec. 2019: Visiting researcher
 - Jan. 2020 - Oct. 2021: Research scientist (L4)
 - Nov. 2021 - Jul 2022: Senior research scientist (L5)
 - Jul. 2022 - May 2024: Senior research scientist manager (L5)
 - May 2024 - Now: Staff research scientist manager (L6)
- 2019 – 2019 ◇ **Post-Doc.** University of Bologna. Computer Vision Lab.
1 year post-doc position under Prof. Luigi Di Stefano, interrupted after 6 months.

Education

- 2015 – 2018 ◇ **Ph.D., Computer Science and Engineering** University of Bologna.
Thesis title: *Computer Vision and Deep Learning for retail store management*.
PhD scholarship sponsored by Centro Studi S.R.L. on the use of computer vision techniques in retail and wholesale stores.
- 2018 ◇ **Visiting Researcher** University of Oxford.
Six months scolarship in the Torr Vision Group of Prof. Philip Torr working on depth estimation for autonomous driving under the Street Wise grant sposored by FiveAI.
- 2013 – 2015 ◇ **M.Sc. Computer Engineering** University of Bologna.
Final grade: 110/110 with Honors.
Thesis title: *Automatic Learning of a multi-scale 3D keypoint detector*.
- 2010 – 2013 ◇ **B.Sc. Computer Engineering** University of Bologna.
Final grade: 110/110 with Honors
Thesis title: *Study and implementation of algorithms for the 3D reconstruction of rooms and the automatic navigation of autonomous drones*.

Skills

- | | |
|----------------|--|
| Languages | ◇ Motherlanguage Italian. Full Proficiency in English. B2 in German. |
| Coding | ◇ Python, C++, L ^A T _E X, ... |
| ML Development | ◇ Tensorflow, Pytorch, TPU development, |
| Misc. | ◇ Academic research, managing, teaching, training, consultation. |

Research Publications

Journal Articles

- 1 S. Morkva, M. Oechsle, **A. Tonioni**, M. Hutter, and V. Patil, "Mosaic-gs: Monocular scene reconstruction via advanced initialization for complex dynamic environments," *arXiv preprint arXiv:2601.05368*, 2026.
- 2 G. Comanici, E. Bieber, M. Schaeckermann, *et al.*, "Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities," *arXiv preprint arXiv:2507.06261*, 2025.
- 3 A. Kukleva, E. Simsar, **A. Tonioni**, *et al.*, "Refam: Attention magnets for zero-shot referral segmentation," *arXiv preprint arXiv:2509.22650*, 2025.
- 4 C. Zaccagnino, F. Quattrini, V. Pippi, S. Cascianelli, **A. Tonioni**, and R. Cucchiara, "Autoregressive styled text image generation, but make it reliable," *arXiv preprint arXiv:2510.23240*, 2025.
- 5 L. Zanella, M. Mancini, Y. Wang, **A. Tonioni**, and E. Ricci, "Training-free online video step grounding," *arXiv preprint arXiv:2510.16989*, 2025.
- 6 M. Comi, Y. Lin, A. Church, **A. Tonioni**, L. Aitchison, and N. F. Lepora, "Touchsdf: A deepsdf approach for 3d shape reconstruction using vision-based tactile sensing," *IEEE Robotics and Automation Letters*, 2024.
- 7 P. Z. Ramirez, A. Cardace, L. De Luigi, **A. Tonioni**, S. Salti, and L. Di Stefano, "Learning good features to transfer across tasks and domains," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 8, pp. 9981–9995, 2023.
- 8 M. Segu, **A. Tonioni**, and F. Tombari, "Batch normalization embeddings for deep domain generalization," *Pattern Recognition*, vol. 135, p. 109 115, 2023.
- 9 M. Poggi, **A. Tonioni**, F. Tosi, S. Mattoccia, and L. Di Stefano, "Continual adaptation for deep stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 4713–4729, 2021.
- 10 D. De Gregorio, **A. Tonioni**, G. Palli, and L. Di Stefano, "Semiautomatic labeling for deep learning in robotics," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 2, pp. 611–620, 2019.
- 11 O. Rahnama, T. Cavallari, S. Golodetz, *et al.*, "Real-time highly accurate dense depth on a power budget using an fpga-cpu hybrid soc," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 66, no. 5, pp. 773–777, 2019.
- 12 **A. Tonioni** and L. Di Stefano, "Domain invariant hierarchical embedding for grocery products recognition," *Computer Vision and Image Understanding*, vol. 182, pp. 81–92, 2019.
- 13 **A. Tonioni**, M. Poggi, S. Mattoccia, and L. Di Stefano, "Unsupervised domain adaptation for depth prediction from images," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 10, pp. 2396–2409, 2019.
- 14 **A. Tonioni**, S. Salti, F. Tombari, R. Spezialetti, and L. D. Stefano, "Learning to detect good 3d keypoints," *International Journal of Computer Vision*, vol. 126, pp. 1–20, 2018.

Conference Proceedings

- 1 M. Comi, **A. Tonioni**, M. Yang, *et al.*, "Snap-it, tap-it, splat-it: Tactile-informed 3d gaussian splatting for reconstructing challenging surfaces," in *3DV*, 2025.
- 2 V. Pippi, F. Quattrini, S. Cascianelli, **A. Tonioni**, and R. Cucchiara, "Zero-shot styled text image generation, but make it autoregressive," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 7910–7919.

- 3 C. Plizzari, **A. Tonioni**, Y. Xian, A. Kulshrestha, and F. Tombari, “Omnia de egotempo: Benchmarking temporal understanding of multi-modal llms in egocentric videos,” in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 24129–24138.
- 4 V. Udandarao, N. Parthasarathy, M. F. Naeem, *et al.*, “Active data curation effectively distills large-scale multimodal models,” in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 14422–14437.
- 5 J. Xie, **A. Tonioni**, N. Rauschmayr, F. Tombari, and B. Schiele, “Test-time visual in-context tuning,” in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 19996–20005.
- 6 O. F. Kar, **A. Tonioni**, P. Poklukar, A. Kulshrestha, A. Zamir, and F. Tombari, “Brave: Broadening the visual encoding of vision-language models,” in *ECCV*, 2024.
- 7 B. Korbar, Y. Xian, **A. Tonioni**, A. Zisserman, and F. Tombari, “Text-conditioned resampler for long form video understanding,” in *ECCV*, 2024.
- 8 M. Shahbazi, L. Claessens, M. Niemeyer, *et al.*, “Inserf: Text-driven generative object insertion in neural 3d scenes,” in *arXiv preprint arXiv:2401.05335*, 2024.
- 9 E. Simsar, **A. Tonioni**, Y. Xian, T. Hofmann, and F. Tombari, “Lime: Localized image editing via attention regularization in diffusion models,” in *WACV*, 2024.
- 10 E. Simsar, **A. Tonioni**, Y. Xian, T. Hofmann, and F. Tombari, “Uip2p: Unsupervised instruction-based image editing via cycle edit consistency,” in *arXiv preprint arXiv:2412.15216*, 2024.
- 11 C. Tsalicoglou, F. Manhardt, **A. Tonioni**, M. Niemeyer, and F. Tombari, “Textmesh: Generation of realistic 3d meshes from text prompts,” in *2024 International Conference on 3D Vision (3DV)*, IEEE, 2024, pp. 1554–1563.
- 12 M. Shahbazi, E. Ntavelis, **A. Tonioni**, *et al.*, “Nerf-gan distillation for efficient 3d-aware generation with convolutions,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 2888–2898.
- 13 E. Simsar, **A. Tonioni**, E. P. Ornek, and F. Tombari, “Latentswap3d: Semantic edits on 3d image gans,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 2899–2909.
- 14 F. Tosi, **A. Tonioni**, D. De Gregorio, and M. Poggi, “Nerf-supervised deep stereo,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 855–866.
- 15 D. M. Arroyo, **A. Tonioni**, and F. Tombari, “Pargan: Learning real parametrizable transformations,” in *arXiv preprint arXiv:2211.04996*, 2022.
- 16 P. Z. Ramirez, D. M. Arroyo, **A. Tonioni**, and F. Tombari, “Unsupervised novel view synthesis from a single image,” in *arXiv preprint arXiv:2102.03285*, 2021.
- 17 F. Yagubbayli, Y. Wang, **A. Tonioni**, and F. Tombari, “Legoformer: Transformers for block-by-block multi-view 3d reconstruction,” in *arXiv preprint arXiv:2106.12102*, 2021.
- 18 R. Spezialetti, D. J. Tan, **A. Tonioni**, K. Tateno, and F. Tombari, “A divide et impera approach for 3d shape reconstruction from multiple views,” in *2020 International Conference on 3D Vision (3DV)*, IEEE, 2020, pp. 160–170.
- 19 P. Z. Ramirez, **A. Tonioni**, S. Salti, and L. D. Stefano, “Learning across tasks and domains,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 8110–8119.
- 20 **A. Tonioni**, O. Rahnama, T. Joy, L. D. Stefano, T. Ajanthan, and P. H. Torr, “Learning to adapt for stereo,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9661–9670.
- 21 **A. Tonioni**, F. Tosi, M. Poggi, S. Mattoccia, and L. D. Stefano, “Real-time self-adaptive deep stereo,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 195–204.

- 22 P. Z. Ramirez, **A. Tonioni**, and L. Di Stefano, “Exploiting semantics in adversarial training for image-level domain adaptation,” in *2018 IEEE International Conference on Image Processing, Applications and Systems (IPAS)*, IEEE, 2018, pp. 49–54.
- 23 **A. Tonioni**, E. Serra, and L. Di Stefano, “A deep learning pipeline for product recognition on store shelves,” in *2018 IEEE International Conference on Image Processing, Applications and Systems (IPAS)*, IEEE, 2018, pp. 25–31.
- 24 **A. Tonioni** and L. Di Stefano, “Product recognition in store shelves as a sub-graph isomorphism problem,” in *Image Analysis and Processing-ICIAP 2017: 19th International Conference, Catania, Italy, September 11-15, 2017, Proceedings, Part I 19*, Springer, 2017, pp. 682–693.
- 25 **A. Tonioni**, M. Poggi, S. Mattoccia, and L. Di Stefano, “Unsupervised adaptation for deep stereo,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1605–1613.
- 26 F. Tosi, M. Poggi, S. Mattoccia, **A. Tonioni**, L. Di Stefano, et al., “Learning confidence measures in the wild,” in *Proceedings of 28th British Machine Vision Conference 2017 (BMVC 2017)*, The British Machine Vision Association and Society for Pattern Recognition, 2017, pp. 1–13.

Patents

- 1 **A. Tonioni**, B. KORBAR, F. Tombari, A. Zisserman, and Y. Xian, *Text conditioned video resampler for video understanding*, US Patent App. 18/949,777, May 2025.
- 2 M. Poggi, T. Fabio, S. Mattoccia, L. Di Stefano, and **A. Tonioni**, *Method to determine the depth from images by self-adaptive learning of a neural network and system thereof*, US Patent App. 17/781,783, May 2024.
- 3 D. M. Arroyo, F. Tombari, and **A. Tonioni**, *Image transformation using interpretable transformation parameters*, US Patent 11,599,980, Mar. 2023.
- 4 M. Segù, F. Tombari, and **A. Tonioni**, *Domain generalization via batch normalization statistics*, US Patent App. 17/909,545, Apr. 2023.

Miscellaneous Experience

Side Projects

- 2013 ◊ “Blam!”, a user powered comic strip aggregator for windows phone.
- 2022 ◊ Informal help with the software development of OCHA AnticiPy: Access data for anticipating humanitarian risk