



Guest Editorial: Special Issue on Performance Evaluation in Computer Vision

Daniel Scharstein¹ · Angela Dai² · Daniel Kondermann³ · Torsten Sattler⁴ · Konrad Schindler⁵

Accepted: 4 March 2021 / Published online: 28 April 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

As the field of computer vision is growing and maturing, performance evaluation has become essential. Most sub-areas of computer vision now have established datasets and benchmarks allowing quantitative evaluation and comparison of current methods. In addition, new benchmarks often stimulate research into the particular challenges presented by the data. Conversely, important areas lacking high-quality datasets and benchmarks might not receive adequate attention by researchers.

The deep learning revolution has made datasets and performance evaluation even more important. Learning-based methods not only require large, well-designed training datasets but also well-defined loss functions, which are usually designed to optimize established performance measures. This creates an implicit bias based on the availability of datasets and the definition of performance metrics.

In this special issue we sought all types of contributions broadly relating to performance evaluation in computer vision, including:

- Manuscripts introducing new performance evaluations or benchmarks, in particular in areas where quantitative evaluation is challenging or subjective
- Manuscripts surveying, evaluating, or comparing existing benchmarks or datasets
- Manuscripts addressing pros and cons of performance evaluations and datasets, as well as recommendations for best practices

We received 38 initial submissions for this special issue. Of these, 30 were considered to be within the scope of the special issue and underwent a rigorous peer review process. Each paper received at least 3 reviews, for a total of 148 reviews by experts in the field—we greatly thank all of them for their effort! In the end, we accepted 18 papers for publication, all of which went through at least one round of revisions.

The accepted papers span the following topics: datasets, benchmarks and evaluations, robustness, novel metrics, and bias in datasets.

Several papers introduce new benchmarks for various domains, including wide-baseline matching (Jin et al. 2021), non-rigid structure from motion (Jensen et al. 2021), single-object tracking (Fan et al. 2021), multi-object tracking (Dendorfer et al. 2021), deraining (Li et al. 2021), and inpainting of censored images (Black et al. 2021).

Other papers contribute novel datasets for action recognition (Weinzaepfel and Rogez 2021) and anomaly detection (Bergmann et al. 2021), as well as a technique for improved reference pose generation for the benchmarking of visual localization methods (Zhang et al. 2021).

Two papers address the evaluation of robustness (Kamann and Rother 2021; Shekar et al. 2021).

✉ Daniel Scharstein
schar@middlebury.edu

Angela Dai
angela.dai@tum.de

Daniel Kondermann
dk@quality-match.com

Torsten Sattler
torsten.sattler@cvut.cz

Konrad Schindler
schindler@ethz.ch

¹ Department of Computer Science, Middlebury College, Middlebury, VT 05753, USA

² Department of Informatics, Technical University of Munich, Boltzmannstraße 3, 85748 Garching, Germany

³ Quality Match GmbH, Langer Anger 7-9, 69115 Heidelberg, Germany

⁴ Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University in Prague, Jugoslávských partyzánů 1580/3, 160 00 Praha 6, Czech Republic

⁵ Photogrammetry and Remote Sensing, ETH Zurich, Stefano-Franscini-Platz 5, 8093 Zurich, Switzerland

Several papers contribute novel metrics and taxonomies for the evaluation of different tasks, including multi-object tracking (Luiten et al. 2021), semantic segmentation (Yan et al. 2021), visual place recognition (Zaffar et al. 2021), image quality assessment (Ding et al. 2021), conditional image generation (Benny et al. 2021), and embodied exploration (Ramakrishnan et al. 2021).

Finally, one paper addresses the problem of bias in facial datasets (Georgopoulos et al. 2021).

We thank all authors for their contribution to this special issue. We hope that the datasets, benchmarks, and metrics developed by the accepted papers as well as the general methodologies for creating benchmarks will be of value to the community.

References

- Jin, Y., Mishkin, D., Mishchuk, A., Matas, J., Fua, P., Yi, K. M., & Trulls, E. (2021). Image Matching Across Wide Baselines: From Paper to Practice. *International Journal of Computer Vision*, 129, 517–547. <https://doi.org/10.1007/s11263-020-01385-0>
- Jensen, S. H. N., Doest, M. E. B., Aanæs, H., & Del Bue, A. (2021). A Benchmark and Evaluation of Non-Rigid Structure from Motion. *International Journal of Computer Vision*, 129, 882–899. <https://doi.org/10.1007/s11263-020-01406-y>
- Fan, H. et al. (2021). LaSOT: A High-quality Large-scale Single Object Tracking Benchmark. *International Journal of Computer Vision*, 129, 439–461. <https://doi.org/10.1007/s11263-020-01387-y>
- Dendorfer, P., Ošep, A., Milan, A., Schindler, K., Cremers, D., Reid, I., Roth, S., & Leal-Taixé, L. (2021). MOTChallenge: A Benchmark for Single-Camera Multiple Target Tracking. *International Journal of Computer Vision*, 129, 845–881. <https://doi.org/10.1007/s11263-020-01393-0>
- Li, S. et al. (2021). A Comprehensive Benchmark Analysis of Single Image Deraining: Current Challenges and Future Perspectives. *International Journal of Computer Vision*, 129, 1301–1322. <https://doi.org/10.1007/s11263-020-01416-w>
- Black, S., Keshavarz, S., & Souvenir, R. (2021). Evaluation of Inpainting and Augmentation for Censored Image Queries. *International Journal of Computer Vision*, 129, 977–997. <https://doi.org/10.1007/s11263-020-01403-1>
- Weinzaepfel, P. & Rogez, G. (2021). Mimetics: Towards Understanding Human Actions Out of Context. *International Journal of Computer Vision*, <https://doi.org/10.1007/s11263-021-01446-y>
- Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., & Steger, C. (2021). The MVTEC Anomaly Detection Dataset: A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection. *International Journal of Computer Vision*, 129, 1038–1059. <https://doi.org/10.1007/s11263-020-01400-4>
- Zhang, Z., Sattler, T., & Scaramuzza, D. (2021). Reference Pose Generation for Long-term Visual Localization via Learned Features and View Synthesis. *International Journal of Computer Vision*, 129, 821–844. <https://doi.org/10.1007/s11263-020-01399-8>
- Kamann, C., & Rother, C. (2021). Benchmarking the Robustness of Semantic Segmentation Models with Respect to Common Corruptions. *International Journal of Computer Vision*, 129, 462–483. <https://doi.org/10.1007/s11263-020-01383-2>
- Shekar, A. K., Gou, L., Ren, L. & Wendt, A. (2021) Label-Free Robustness Estimation of Object Detection CNNs for Autonomous Driving Applications. *International Journal of Computer Vision*, 129, 1185–1201. <https://doi.org/10.1007/s11263-020-01423-x>
- Luiten, J., Ošep, A., Dendorfer, P., Torr, P., Geiger, A., Leal-Taixé, A., & Leibe, B. (2021) HOTA: A Higher Order Metric for Evaluating Multi-object Tracking. *International Journal of Computer Vision*, 129, 548–578. <https://doi.org/10.1007/s11263-020-01375-2>
- Yan, J., Zhong, Y., Fang, Y., Wang, Z. & Ma, K. (2021) Exposing Semantic Segmentation Failures via Maximum Discrepancy Competition. *International Journal of Computer Vision*. <https://doi.org/10.1007/s11263-021-01450-2>
- Zaffar, M., Garg, S., Milford, M., Kooij, J., Flynn, D., McDonald-Maier, K., & Ehsan, S. (2021) VPR-Bench: An Open-Source Visual Place Recognition Evaluation Framework with Quantifiable Viewpoint and Appearance Change. *International Journal of Computer Vision*. <https://doi.org/10.1007/s11263-021-01469-5>
- Ding, K., Ma, K., Wang, S., & Simoncelli, E. P. (2021) Comparison of Full-Reference Image Quality Models for Optimization of Image Processing Systems. *International Journal of Computer Vision*, 129, 1258–1281. <https://doi.org/10.1007/s11263-020-01419-7>
- Benny, Y., Galanti, T., Benaim, S., & Wolf, L. (2021) Evaluation Metrics for Conditional Image Generation. *International Journal of Computer Vision*. <https://doi.org/10.1007/s11263-020-01424-w>
- Ramakrishnan, S. K., Jayaraman, D., & Grauman, K. (2021) An Exploration of Embodied Visual Exploration. *International Journal of Computer Vision*. <https://doi.org/10.1007/s11263-021-01437-z>
- Georgopoulos, M., Oldfield, J., Nicolaou, M. A., Panagakis, Y., & Pantic, M. (2021) Mitigating Demographic Bias in Facial Datasets with Style-Based Multi-Attribute Transfer. *International Journal of Computer Vision*. <https://doi.org/10.1007/s11263-021-01448-w>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.