



## Editorial for Special Issue on Computer Vision in the Wild

Mei Chen<sup>1</sup> · Cha Zhang<sup>1</sup> · Katsushi Ikeuchi<sup>1</sup>

Accepted: 20 October 2021 / Published online: 6 November 2021

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

Our topic for this special issue was motivated by the exciting development in the field of computer vision in recent years – the tsunamic volume, the blazing speed. From veteran vision researchers to recent adopters, there is a shared sense of optimism and eagerness in the air, that this technology is coming of age, ready for the real world, in the wild. What applications can it already enable? What tasks can it reliably conduct? What are the potential barriers to adoption, is it the accuracy, is it the efficiency, or is it the adaptability?

The community echoed our enthusiasm, and responded with 29 submissions ranging a wide array of topics. Each submission was assigned three or more experienced reviewers, and only when all reviewers recommended "accept as is" did the editors decide accordingly. After a rigorous review and revision process despite the pandemic, we are elated to present to the community 20 accepted manuscripts for this special issue. The problems under investigation include classic vision tasks such as object detection, re-identification, segmentation, tracking, pose estimation, as well as more deep learning specific treatments such as few shot learning, unsupervised domain adaptation, efficient neural architecture search, and more.

Suffice to say that, perhaps unsurprisingly, much of the state of the art is still *towards* computer vision in the wild rather than already computer vision in the wild. That is exactly what this special issue aims to serve, as a forum to bring front and center practical considerations and solutions for real world applications.

In classic vision tasks such as object recognition and segmentation, there are still open problems to be addressed in order to apply to in the wild situations. For instance, in "*Compositional Convolutional Neural Networks: A Robust and Interpretable Model for Object Recognition under Occlusion*", the authors design an occlusion robust recognition model by combining part-based models and deep neural networks to achieve occlusion-awareness and interpretability; in "*OCNet: Object Context for Semantic Segmentation*", a more effective context aggregation scheme is proposed for semantic segmentation.

When applying vision algorithms to image sequences, difficulties arise due to various in the wild scenarios to ensure image continuity. To name a few, in "*Learning Regression and Verification Networks for Robust Long-term Tracking*", the method combines the merits of target matching and classification to address the challenges in long-term-tracking; in "*AdaFuse: Adaptive Multiview Fusion for Accurate Human Pose Estimation in the Wild*", it tackles the problem of occlusion for human pose estimation by simplifying the conventional 'point-point' correspondence process to 'point-line' correspondence calculation considering the sparsity of the heatmap features.

---

✉ Mei Chen  
Mei.Chen@microsoft.com

<sup>1</sup> Redmond, WA, USA

The space of in the wild applications is rich in diversity. When adopting existing techniques to different domains or settings, a plethora of issues come to light. Such as in "*Unsupervised Domain Adaptation in the Wild via Disentangling Representation Learning*", a new architecture is designed to allow unsupervised domain transfer that considers possible categorical imbalance between source and target domains; in "*Progressive DARTS: Bridging the Optimization Gap for NAS in the Wild*", the authors make the important observation that there exists an optimization gap between what happens during search time and how the resulting final network is utilized and proposes a "progressive" search technique to alleviate the gap; in "*Learning Adaptive Classifiers Synthesis for Generalized Few-Shot Learning*", the few-shot and many-shot classifiers are synthesized using an attention-

based mechanism so the final classifier works well for both few-shot and many-shot categories.

Whether you are looking for insights for particular vision tasks such as multi-object tracking or object re-identification, or for ideas on certain application scenarios such as zero-shot or unsupervised setting, there is an article sharing keen observations and useful techniques with thoughtful and extensive validation and ablation.

Happy reading.

Mei, Cha, Katsushi  
2021

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