

Aiding In-the-wild Deep Learning-based Computer Vision Tasks via Pseudo-data and Dynamic Adaptive Learning Strategies

Jonathan Samuel Lumentut, Bina Nusantara University

Abstract

In recent times, deep learning approaches have been vastly utilized for various computer vision tasks. Their implementation in computer vision spans the work of the image capturing stage, such as image signal processing (ISP) and image restoration, up to the downstream task, such as depth estimation, 3D reconstruction, object recognition, segmentation, etc. Although deep learning models are richly available today, their utilization in real-world, in-the-wild cases still faces major limitations.

Running common deep learning models on real-world, in-the-wild data cases may lead to the performance drop due to the following challenging factors, namely: (i) the model's dependency on pre-defined training data, which can be obsolete at times, (ii) the limited ability to acquire reliable labels for that training dataset, and (iii) the unknown domain gap characteristic between the training data itself and the in-the-wild test data.

The following research topics, which I conducted during my previous study, namely (1) light field computational photography, (2) content-based image restoration, and (3) 3D human-body reconstruction (3DHR), presented similar challenges to the above when it came to handling in-the-wild data. To answer these challenges, I opted to embrace pseudo-data exploration and dynamic adaptive learning strategies. Exploring pseudo-data can be seen as a way to mimic real-world datasets, which are helpful for training deep learning models, except that they are provided cheaper.

On the other hand, a dynamic adaptive learning strategy is an approach that can invoke the adaptability of deep learning models while handling the test data unknown to the model's knowledge, e.g., real-world in-the-wild test data. My research employed both methods (pseudo-data and dynamic-adaptive learning), and they have proven useful in enhancing the capability of various deep-learning models in running various computer vision tasks, especially in handling in-the-wild data.

Moreover, the combined strategies above can further refine (adapt) the deep-learning models at test time (inference stage), helping to reduce the domain gap between the training data and the uncertainty characteristics of in-the-wild test target data. Quantitatively, these strategies gave huge performance boosts during the benchmarking process.

The gains above came with the extra benefit of having only a small computational footprint (fast runtime) thanks to the framework that is designed agnostically without modifying the pre-trained models. In terms of future directions, I would like to take the opportunity to work on my expertise above according to the team's objective while embracing the trend of multimodal strategies.