Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)

Connecting Low-Level Image Processing and High-Level Vision via Deep Learning

Ding Liu
University of Illinois at Urbana-Champaign, USA
dingliu2@illinois.edu

1 Overview

The latest development of computer vision has made exciting progress and tremendous impact in our daily lives. In this exciting era of technological advances, deep learning has gained huge popularity as a powerful tool for solving a lot of computer vision problems, and has added a great boost to this already rapidly developing field. Conventionally, the connection between different vision tasks is fragile. For example, low-level image processing and high-level vision tasks are usually coped with separately. However, the inherent relation of feature representations among various tasks should be effectively utilized rather than omitted. My research focuses on **connecting low-level image processing and high-level vision via deep learning**. Specifically, my goal is to design deep learning mechanisms that can efficiently and effectively learn features from low-level image processing and use them to improve the performance of high-level vision tasks.

Low-level image processing tasks, such as image restoration and image enhancement, play a key part in computer vision, as they enable the extraction of fundamental image primitives for further processing. In the past few years, we have witnessed an increasing interest in these fundamental topics from not only research communities but also industry, and substantial progress has been achieved. While the high-level vision research has made significant progress with deep learning in recent years, such as image classification and face recognition, most models are trained, applied, and evaluated on high-quality (HQ) visual data, such as the LFW and ImageNet benchmarks. However, in many real-world scenarios, the performances of visual sensing and analytics are largely jeopardized by low-quality (LQ) visual data acquired from complex unconstrained environments, suffering from various types of degradations such as low resolution, noise, occlusion and motion blur. While some mild degradations may be compromised by sophisticated visual recognition models, their impacts turn much notable as the level of degradations passes some empirical threshold. To tackle such problems, building a model directly on LQ data is usually not robust due to the severe information loss caused by adverse conditions. On the other hand, the model trained on HQ data does not also perform well when tested on LQ data due to the domain mismatch.

I propose to use such low-level image processing techniques as an important frontend to help solve the challenging high-level vision tasks, especially under adverse conditions. My main intuition is to regularize and enhance the feature extracted from LQ data, via injecting auxiliary information from HQ data with low-level image processing techniques. With the help of HQ data, the deep learning model can better discriminate true signals from severe corruptions, and learn more robust feature extractors from LQ inputs.

2 Current Achievement

I have made crucial contributions on solving low-level image processing problems with deep learning, such as image and video super-resolution [Wang *et al.*, 2015; Liu *et al.*, 2016b; 2016a; 2017b], image restoration [Wang *et al.*, 2016b] and image denoising [Liu *et al.*, 2017c].

For tackling low-level image processing problems, I propose a *sparse coding network* for image super-resolution, by combining the merits of sparse representation and deep learning in [Wang *et al.*, 2015; Liu *et al.*, 2016b]. This network is designed with the domain expertise from the conventional sparse coding model, which leads to faster convergence in training and smaller model size. This model is evaluated on a wide range of images, and shows clear advantage over existing state-of-the-art methods in terms of both restoration accuracy and human subjective quality. The performance is further improved by learning a mixture of deep networks in [Liu *et al.*, 2016a]. This technique of super-resolution has been deployed in Adobe's most profitable product, Photoshop, and has been covered in media worldwide. Image super-resolution is reformulated as a recurrent neural network that exploits both low-resolution and high-resolution sigmals jointly in [Han *et al.*, 2018]. In [Liu *et al.*, 2017b], I propose a spatial alignment network to align consecutive low-resolution frames, and a temporal adaptive neural network which can adaptively select the best temporal scale for video super-resolution at much higher speed.

In addition to super-resolution, a deep dual-domain model is devised for restoration of JPEG-compressed images in my work [Wang *et al.*, 2016b]. By leveraging the large learning capacity of deep networks as well as the prior knowledge of the JPEG compression scheme, this model outperforms the latest competing method for around 1dB in PSNR and is 30 times faster. In [Liu *et al.*, 2017c] I design a neural network which conducts convolutions in different spatial scales via downsampling and upsampling operations before

the resulting features are fused together, so that the kernels have a larger receptive field after all the feature contraction. It achieves the state-of-the-art methods denoising performance, demonstrating the strong capability of neural network to capture both local and global contextual information for denoising.

3 Ongoing Research and Future Directions

It becomes the latest trend to jointly study low-level image processing to high-level vision problems, as there is practical need to adopt high-level vision applications, such as surveillance system, intelligent human-computer interaction and autonomous driving, in low-quality imagery data. I expect to apply cutting-edge deep learning techniques to systematically investigate the mutual influence between low-level image processing and high-level vision problems.

I propose a unified deep learning framework to deal with image denoising and high-level vision tasks jointly in [Liu *et al.*, 2017c], which is the first work investigating the benefit of exploiting image semantics simultaneously for image denoising and high-level vision tasks via deep learning. First I quantitatively show that image denoising can enhance the accuracy of image classification and semantic segmentation. Then I show that high-level semantics can be used for image denoising to generate visually appealing results in a deep learning fashion. In [Wang *et al.*, 2016a; Liu* *et al.*, 2018], a transfer learning method is developed to deal with deep learning based visual recognition in very low resolution conditions by exploiting pre-trained features for super-resolution. Extensive experiment results show that pre-training part of neural network models from the super-resolution task is able to maintain a good balance between the feature extraction and the discrimination ability, and achieve impressive performance in the task of object recognition and face identification under adverse conditions. Moreover, I demonstrate that my video super-resolution approach is superior than other recent competing methods in terms of the accuracy of video face recognition under a low resolution setting in [Liu *et al.*, 2017b]. These works reveal the fact that the feature representations learned from low-level image processing tasks can not only be used to generate visually pleasant results, but also provide more semantically faithful information, which benefits high-level vision tasks [Liu *et al.*, 2017a].

For the future work, my goal is to systematically investigate the potential links between low-level image processing problems and high-level vision tasks via deep learning, with the focus on two questions, namely (1) how low-level image processing can help solving high-level vision problems, and (2) how the semantic information from high-level vision tasks can be used to guide low-level image processing. I expect to extend my current transfer learning approach to cope with more types of degradations, such as noise, occlusion and motion blur, and develop a degradation-robust recognition framework, which can be widely applied in modern surveillance system, intelligent robot and autonomous driving.

Network architecture has been shown to play an important role in deep learning models and various types of architecture are designed in a task-specific manner. Another direction of mine is to explore new universal neural network architecture with the help of multi-task learning, in order to handle visual data under various adverse conditions and extract semantically faithful features more effectively and efficiently. I believe my previous research experience and the skills I have accumulated before will put me in the special position to solve this challenging problem and make a unique contribution to the field of computer vision.

References

[Han *et al.*, 2018] Wei Han, Shiyu Chang, Ding Liu, Michael Witbrock, and Thomas S Huang. Image super-resolution via dual-state recurrent neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[Liu *et al.*, 2016a] Ding Liu, Zhaowen Wang, Nasser Nasrabadi, and Thomas Huang. Learning a mixture of deep networks for single image super-resolution. In *Asian Conference on Computer Vision*, pages 145–156. Springer, 2016.

[Liu *et al.*, 2016b] Ding Liu, Zhaowen Wang, Bihan Wen, Jianchao Yang, Wei Han, and Thomas S Huang. Robust single image super-resolution via deep networks with sparse prior. *IEEE Transactions on Image Processing*, 25(7):3194–3207, 2016.

[Liu *et al.*, 2017a] Ding Liu, Bowen Cheng, Zhangyang Wang, Haichao Zhang, and Thomas S Huang. Enhance visual recognition under adverse conditions via deep networks. *arXiv preprint arXiv:1712.07732*, 2017.

[Liu *et al.*, 2017b] Ding Liu, Zhaowen Wang, Yuchen Fan, Xianming Liu, Zhangyang Wang, Shiyu Chang, and Thomas Huang. Robust video super-resolution with learned temporal dynamics. In *International Conference on Computer Vision*, pages 2507–2515, 2017.

[Liu *et al.*, 2017c] Ding Liu, Bihan Wen, Xianming Liu, and Thomas S Huang. When image denoising meets high-level vision tasks: A deep learning approach. *arXiv preprint arXiv:1706.04284*, 2017.

[Liu* *et al.*, 2018] Ding Liu*, Bowen Cheng*, Zhangyang Wang, Haichao Zhang, and Thomas S Huang. Visual recognition in very low-quality settings: Delving into the power of pre-training. In *The Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[Wang *et al.*, 2015] Zhaowen Wang, Ding Liu, Jianchao Yang, Wei Han, and Thomas Huang. Deep networks for image super-resolution with sparse prior. In *International Conference on Computer Vision*, pages 370–378, 2015.

[Wang *et al.*, 2016a] Zhangyang Wang, Shiyu Chang, Yingzhen Yang, Ding Liu, and Thomas S Huang. Studying very low resolution recognition using deep networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4792–4800, 2016.

[Wang *et al.*, 2016b] Zhangyang Wang, Ding Liu, Shiyu Chang, Qing Ling, Yingzhen Yang, and Thomas S Huang. D3: Deep dual-domain based fast restoration of jpeg-compressed images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2764–2772, 2016.