Self-supervised image denoising with deep neural networks

Abstract

This project studies deep learning based image denoising approaches without the requirement of clean targets in the same domain for model training. For this purpose, self-supervised and meta-learning methods will be conducted. This study will reach effective and practical image denoiser, being evaluated on a few datasets especially those containing real-world images.

1 Introduction

Removing noise from degraded images to recover high quality ones, known as image denoising, is a fundamental task in computer vision. It has been a classic research area yet remains active nowadays (Gu & Timofte, 2019). It not only greatly affects user experience in practical applications, but also plays a very important role for subsequent computer vision tasks such as classification and recognition (Gu & Timofte, 2019).

A widely accepted yet simple image degradation model is $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}$ where $\boldsymbol{x}$ refers to the uncorrupted image, $\boldsymbol{y}$ represents the degraded image and $\boldsymbol{n}$ is the additive noise (Gu & Timofte, 2019; Zhang, Zuo, Chen, Meng, & Zhang, 2017). Several kinds of noises has been widely studied, including additive white Gaussian noise (AWGN), Poison noise, and salt-and-pepper noise (Gu & Timofte, 2019).

The biggest challenge in image denoising is the loss of information during degradation, making this problem highly ill-posed (Gu & Timofte, 2019; Lee, Cho, Kim, & Kim, 2020). As a result, prior knowledge is required to compensate the lost information to recover high quality image (Gu & Timofte, 2019). This can be the prior modelling of either the images or noise (Chen, Chen, Chao, & Yang, 2018).

A variety of models have been proposed for image prior representation, including some state-of-art ones such as non-local self-similarity-based methods BM3D or WNNM (Valsesia, Fracastoro, & Magli, 2019; Zhang et al., 2017). The most popular and classic one is BM3D, which serves as a benchmark in image denoising (Chen et al., 2018). However, there are some major disadvantage of these models. First, they mostly rely on human knowledge. Second, they only utilise the information of a single input image (Chen et al., 2018).

In recent years, deep neural networks (DNNs) have revolutionised traditional methods and became the state-of-art technology on most tasks of computer vision (Gu & Timofte, 2019). In terms of image denoising, a variety of DNN models have been proposed, attracting increasing attentions attributed to its performance. Models based on Convolutional Neural Networks (CNNs) achieved significance, such as RED (Mao, Shen, & Yang, 2016) or DnCNN (Zhang et al., 2017). More recent technologies are also introduced to image denoising, such as Generative Adversarial Networks (GANs) (Chen et al., 2018), Graph Neural Networks (GNNs) (Valsesia et al., 2019), and meta-learning (Lee et al., 2020).

2 Related work

There are several CNN-based image denoising models. RED-Net was proposed in (Mao et al., 2016) for denoising in different noise levels with a single model. It is a very deep architecture (up to 30 layers in experiment) which consists of convolutional and deconvolutional layers, with skip connections between each convolutional layer and its symmetric deconvolutional one. The skip connections helps passing gradients in back-propagation to alleviate gradient vanishing problem. DnCNN (Zhang et al., 2017) combines residual learning strategy and batch normalisation into feed-forward convolutional neural network to improve the final model metrics and also to accelerate the training process. DnCNN learns the residual (i.e. difference between the noisy and clean image) instead of the clean image itself. This model outperforms state-of-art methods such as BM3D, WNNM and TNRD, and can be effectively extended to more general image denosing tasks such as blind Gaussian denoising. A recent study (Valsesia et al., 2019) introduced graph neural networks to CNN-based image denoising architecture. Because of the local nature of convolutional operation, CNN-based model is unable to exploit non-local similarity patterns which had been proven to be significant by previous model-based methods. As a response, GraphCNN was proposed by incorporating the Edge Conditioned Convolution

(ECC), a graph convolutional layer, to create non-local receptive field. This method improved the metrics on average, but did not beat existing methods on some categories in their experiment.

Some attempts has been made to loosen the requirement for training data. Chen et al. (2018) proposed a GAN-CNN Based Blind Denoiser (GCBD) model, arguing that in real world applications noise is not ease to obtain and is usually more complex than Gaussian noise, thus the models trained for knowing noises might significantly decrease in performance. To solve this problem, the GCBD utilise a two-stage model. First, a GAN is trained to model the noise distribution. Second, the noise sampled from previous step are paired with real images, resulting in a proper training dataset for deep CNN based denoising models. Though this method does not require noisy and clean image pairs for training, it still needs clean images. Another method named Noise2Noise (Lehtinen et al., 2018) learns a denoising model with noisy data only, based on an basic observation that the loss function will not be affected by the change of distribution of targets as long as it remains the same expectation. This model can be trained in the absence of clean images yet achieves comparable performance, if not better, of models trained from dataset with clean targets. The capability of Noise2Noise was demonstrated by experiments with various noises and images, but they only covered synthetic noises. Another drawback of this method is that it needs different noisy observations of the same image. Going one step further, Noise2Void (Krull, Buchholz, & Jug, 2019) implements image denoising with only a single noisy input, and can be applied to many existing neural network architectures. In this approach, the author introduce the blind-spot network, trained by patches extracted from noisy image with the center pixel masked. Experiments were carried out on both synthetic and real noisy images, however, the results on real data were evaluated by human vision due to the absence of ground truth. A latest research (Lee et al., 2020) proposed a two-phase denoiser which first utilised an arbitrary pre-trained denoiser g and augmented the available patches at self-supervision stage by adding random noise to the output of g. To gain benefit from supervised learning on large labelled datasets, a meta-learning approach was employed for fast adaptation to test inputs.

Many image datasets can be used for denoising model evaluation with synthetic noises, such as Set14 (Zeyde, Elad, & Protter, 2012) and BSD500 (Arbelaez, Maire, Fowlkes, & Malik, 2011). Some real world data are available as benchmark. For example, Darmstadt Noise Dataset (DND) (Plotz & Roth, 2017) contains 50 pairs of images captured by consumer cameras at different ISO values, with the low-ISO ones as ground truth.

3 Methodology

This research will be conducted based on several state-of-art approaches and experiments. Several existing studies have given us useful guidelines in choosing model architecture. First of all, CNN has been proved to be effective and successful in image denoising (Mao et al., 2016), so this study will focus on CNN-based approach. Second, residual learning and batch normalisation are great methods to improve the final model metrics (Zhang et al., 2017), so they will be included in this study. Finally, graph neural network, as one of the newest technology, has been successfully introduced into image denoising recently (Valsesia et al., 2019). GNN has great potential capability of exploiting non-local features, and will be studied with special attentions in this research.

Training strategy is another important aspect of this research. This study aims to effective and practical image denoising method without dependency on large in-domain labelled training dataset. For this purpose, self-supervised approaches such as Noise2Noise (Lehtinen et al., 2018) and Noise2Void (Krull et al., 2019) are promising candidates. Moreover, meta-learning has been proved to be effective to improve the model metrics and speed up model inference (Lee et al., 2020), and thus will be investigated in this study. GAN-based noise modelling (Chen et al., 2018) is another measure to alleviate the lack-of-data problem, which can be introduced to the two-phase denoiser (Lee et al., 2020) for patch generation with more realistic noise distributions.

4 Timeline and milestones

Task	Deadline
Final decision on the topic, create research questions	1 week
Literature review	3 weeks
Research proposal draft	1 week
Prototyping	4 weeks
First round of testing and analysis	4 weeks
Model improvement	4 weeks
Second round of testing and analysis	4 weeks
Write and present final results	4 weeks

5 Research resources

Deep learning models are computational intensive, and recent models to be studied such as GAN, GNN, meta-learning are even more so. Proper hardware sets includes Intel i7 series CPU, an Nvidia Titan X series or 2080Ti GPU, and 1T disk storage. Tensorflow or Pytorch will be chosen as the modelling framework in consideration of many existing works having been implemented in either framework.

6 Planned outcomes and risks

The expected outcomes of this research are practical image denoising systems that are effective on real world environments and applications. It will utilise state-of-art technologies for novel solutions at this aim, and will improve the metrics (PSNR) on particularly real world image datasets such as DND (Plotz & Roth, 2017).

The biggest risk of this research is that the expected improvement cannot be achieved for certain methods. To reduce such risk, more methods could be included and combined, and flexibility of up to 4 weeks will be introduced into schedule.

Reference

Arbelaez, P., Maire, M., Fowlkes, C., & Malik, J. (2011). Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell. 33*(5), 898–916. doi:10.1109/TPAMI.2010.161

Chen, J., Chen, J., Chao, H., & Yang, M. (2018). Image blind denoising with generative adversarial network based noise modeling. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Gu, S., & Timofte, R. (2019). A brief review of image denoising algorithms and beyond. In S. Escalera, S. Ayache, J. Wan, M. Madadi, U. Güçlü, & X. Baró (Eds.), *Inpainting and Denoising Challenges* (pp. 1–21). Cham: Springer International Publishing.

Krull, A., Buchholz, T.-O., & Jug, F. (2019). Noise2void - learning denoising from single noisy images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Lee, S., Cho, D., Kim, J., & Kim, T. H. (2020). Self-supervised fast adaptation for denoising via meta-learning. *arXiv preprint arXiv:2001.02899*.

Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., & Aila, T. (2018). Noise2noise: Learning image restoration without clean data. In *International Conference on Machine Learning* (pp. 2965–2974).

Mao, X.-J., Shen, C., & Yang, Y.-B. (2016). Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *Proceedings of the 30th International Conference on Neural Information Processing Systems* (pp. 2810–2818). NIPS'16. Barcelona, Spain: Curran Associates Inc.

Plotz, T., & Roth, S. (2017). Benchmarking denoising algorithms with real photographs. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Valsesia, D., Fracastoro, G., & Magli, E. (2019). Image denoising with graph-convolutional neural networks. In *2019 IEEE International Conference on Image Processing (ICIP)* (pp. 2399–2403).

Zeyde, R., Elad, M., & Protter, M. (2012). On single image scale-up using sparse-representations. In J.-D. Boissonnat, P. Chenin, A. Cohen, C. Gout, T. Lyche, M.-L. Mazure, & L. Schumaker (Eds.), *Curves and Surfaces* (pp. 711–730). Berlin, Heidelberg: Springer Berlin Heidelberg.

Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, *26*(7), 3142–3155.