CSC413 Final Project Proposal

Zhuoheng Huang(1004727842), Junming Zhang(1003988982), Yuchen Tong(1003534669)

1 Abstract

In this project we plan to re-implement and evaluate two deep learning algorithms for image restoration with very deep convolutional networks. We adopt the Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections (REDNet) by Mao et al. [1] and the Residual Non-local Attention Networks (RNAN) by Zhang et al. [2]. Mao proposes a REDNet consists of multiple deconvolutional and convolutional layers. The convolutional layers detect features of the input image and reduce the noises and corruption. The deconvolutional layers recover image details. To speed up the training and obtain a better local extreme, the skip connection is applied to avoid gradient vanishing, back-propagate the signal directly to bottom layers, and send the image features to deconvolutional layers from convolutional layers, so to simplify the model training and improve the image restoration quality. Zhang proposes a RNAN containing local and non-local attention blocks to capture details about long-range dependencies between pixels. Each attention block has the trunk branch to extract hierarchical features, and then those features are re-scaled by the local and non-local mask branch. While the local branch focuses more on the local structures and the non-local branch concentrates more on long-range dependencies. The very deep network for RNAN is trained by residual local and non-local attention learning to promote the representation ability. Experiments for each network demonstrate comparable to or even better results than the cutting-edge methods recently, therefore it is worth comparing both methods by implementing the two algorithms.

2 Introduction

Image restoration is a process to recover a corrupted image to its high-quality version, which is known as an ill-posed problem because the image degradation is irreversible. Some mathematical models are applied before. Recently, deep convolutional neural networks have shown credible abilities in image restoration for a wide range of problems without assuming a specific task. It is interesting to ask, can a single deep CNN solve multiple levels of corruptions? Does a deeper CNN generally solve the image restoration problem better? To observe the performance of deep CNN on image restoration, we find REdNet proposed by Mao et al. who asked the questions above and provided a deep CNN based framework for image restoration with only one single very-deep network model. This network consists of multiple convolutional and deconvolutional layers. Their operators are connected symmetrically with skip-layer connections to obtain a superior local minimum and accelerate convergence process in training.
Later on, Zhang et al. reviewed deep CNN algorithms in image restoration field, including Mao's work. He mentioned about three common unsatisfactory issues for deep CNN based frameworks before his work:

1. The receptive field size is too small to detect long-range dependencies between pixels and in the whole image.
2. The ability to solve the corruptions and noises in distinctive regions is limited.
3. Channel-wise features are treated equally, and the treatment cannot flexibly tackle all kinds of problems.

To tackle those issues, Zhanget al. proposes the very deep residual non-local attention networks (RNAN), constructed by local and non-local attention blocks as the fundamental unit of the very deep network. Each block has trunk and mask branches. The trunk branch has the residual block to extract hierarchical features. The receptive field size is enlarged by feature down-scaling and up-scaling with large stride convolution and deconvolution. The mask branch contains non-local block to collect non-local information.

3 Related Works

Since we re-implement, evaluate and compare two deep CNN based approaches in this project, the extensive work we introduce in this section is mainly about these two approaches. To check the related work for the two approaches investigated in this project, please review papers for two algorithms in the reference.

1. **Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections (REDNet).** A very deep network implemented by Mao et al. [1], to verify if a single CNN can address image restoration problem, and if a deeper CNN generally works better. This network has only one single very deep CNN, constructed by symmetrically-linked convolutional and deconvolutional operators by skip-connections.

2. **Residual Non-local Attention networks (RNAN).** A very deep network implemented by Zhang et al. [2], to address three problems in previous deep CNN for image restoration: small receptive field size, restricted abilities on distinct corruptions, and lack of flexibility in channel-wise feature extraction. This network consists of local and non-locak attention blocks, with trunk and local mask branch in each block.

4 Method

The final project applies two methods on image restoration and comparisons will be made to investigate in the performance of hyper-parameters and the network models. The method mainly being applied is the very-deep convolutional encoder-decoder networks with symmetric skip connections proposed by Mao et al. [1], in which a neural network is decomposed into convolutional and deconvolutional blocks with skip connections for every two layers to build a bridge between convolutional layers and their mirrored deconvolutional counterparts. This model does not contain any fully-connected layers and the benefit of it is that the skip connections have element-wise correspondence, which can be very important in pixel-wise prediction problems. In addition, such model could be feasible for different levels of corruption representing varying levels in the task of image denoising and multiple scaling parameters in image super-resolution.

Besides, we suggest employing the residual non-local attention networks as the second method in comparison with the first network model. It was initially proposed by Zhang et al. [2] and share some similarities to the first method in that residual pattern is applied as well and no fully-connected layer is implemented either. However, the second network pattern is different in the overall structure. The network is composed of 8 residual local and 2 non-local attention blocks, where for each block there are 2 residual blocks in the beginning and end, and a trunk and mask branch are established in between, whose sub-output feature maps undergo matrix multiplication and element-wise addition with the sub-output map generated from the first two residual blocks. The residual blocks in this network are simplified compared to their traditional forms for better contribution to image super-resolution. The non-local blocks are set up because the receptive field size of the mask branch is much larger than that of the trunk branch and it cannot cover the whole features at a time, while the non-local blocks address this issue in a pleasant way with non-local mixed attention.

To evaluate how single deep CNN works on the general image restoration problems with no assumptions, we plan to implement the two algorithms listed above. Furthermore, we will compare the performance of the two algorithms by investigating proper metrics, to evaluate how RNAN exceeds REDNet based on Zhang's design.

The possible contributions from our projects are summarized in the following:

- Implement the state-of-art image restoration algorithms with deep CNN framework, one consists of symmetric convolutional and deconvolutional layers as encoders and decoders, and linked with skip-layer connections, another consists of local and non-local attention blocks with trunk and mask branches.

- Verify if a single deep CNN can address the image restoration problem, and if a deeper CNN performs better in general.

- Reproduce the experimental results from reference, and compare the performance of two algorithms, and find the strength and weakness of both approaches.

what is this? be specific

In detail, to conduct our comparisons and studies, we intend to use BSD200 dataset and some other famous image collections of real-world objects and scenes. With multiple image deforming techniques we would generate deformed and damaged images as the inputs and their origianl images as the ground truths, thereby targeting different patches for different restoration methods to be tested like image denoising, demosaicing, super-resolution and even inpainting. These input-output pairs will be applied onto the above two methods. For the evaluations, mIOU loss and cross entropy loss will be compared conditioned on different hyper-parameters to evaluate the performance of two methods. In addition, we will set up some loss standard as the benchmark for predictions being "well-generated" and test out the number of epochs, iteration or batch sizes needed to achieve the pre-set loss value. It is an effective measurement on the model efficiency and may be extended for further studies and improvements on the original network models.

How to compute mIoU and cross entropy for image restoration tasks?

5 Summary

How to evaluate the performance? What metric?

Which image restoration task?

Our goal for this project is to investigate the image restoration problem by re-implementing and analyzing the performance of two existing deep learning algorithms:

GPU?

- Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections (REDNet). [1]
- Residual Non-local Attention networks (RNAN). [2]

run grammar check

We will use BSD200 and some other famous image collections which consist of real-world objects and scenes as our dataset. These images will be considered as ground truth while the corresponding deformed images will be generated by performing various image deforming techniques as inputs. To compare the performance of these two algorithms conditioned on different hyper-parameters, both mIOU loss and cross-entropy loss will be used separately during training. The trained models are expected to successfully perform various restoration methods such as image denoising, demosaicing, super-resolution, and even inpainting. Lastly, we will develop an effective measurement of the model efficiency for bench-marking.

What is this? be specific

References

[1] Xiaojiao Mao, Chunhua Shen, and Yu-Bin Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

[2] Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. Residual non-local attention networks for image restoration. 2019.