Connecting low-level image processing and high-level vision for degraded image classification

Group 16 Project 1 Proposal

Changyue Su
The University of Hong Kong
3035534854
scytina6@connect.hku.hk

Zhui Ding
The University of Hong Kong
3035533044
u3553304@connect.hku.hk

Kehan Long
The University of Hong Kong
3035771323
longkhan@connect.hku.hk

Haozhou Xu
The University of Hong Kong
3035771086
xhz1@connect.hku.hk

Qi Yijiazhen
The University of Hong Kong
3035773096
u3577309@connect.hku.hk

1. Introduction

Recent years have witnessed tremendous and robust development in Computer Vision areas, especially in Image Classification [9]. The rapidly emerging image classification techniques have an increasing impact on our life. Some presentational applications are auto-driving, satellite object identification, medical imaging and machine controls.

However, most current image classification models only focus on high-level vision, regardless of low-level features [6]. Low-level image processing can play a crucial role in image classification tasks by extracting fundamental image primitives. Connecting low-level image processing to high level vision is promising.

An interesting direction is to use low-level vision to enhance image classification on degraded images [5]. Common state-of-the-art image classification methods largely rely on annotated datasets of high-quality images. Their performance on low-quality images is significantly degraded [7]. However, real-world vision data usually suffer from low resolution, noise, occlusion and motion blur. As a result, it is meaningful to tackle low-quality image classification problems. A low-level processing model can be used to improve image quality as well as extract low-level features for future high-level processing [6]. We believe it is feasible to combine low-level image processing with high-level CNN into a framework for enhancing degraded image classification.

2. Related Works

2.1. High level image classification models

Classical image classification models are AlexNet [4], VGG [5] and ResNet [6], respectively. These three methods are commonly used for the classification task. They can achieve good enough results on various datasets.

2.2. Low-quality image classification

Many works have been done to mitigate the drop in accuracy for low-quality image classification [5]. One direction is to transfer this problem into two-stage tasks [3]. First, restore and enhance the degraded images, and then apply image classification. Another method is to make domain adaptation [12]. By using adversarial learning or kernelized training, this approach aims at matching the marginal distribution of low- and high-quality images [5]. A more intuitive solution is fine-tuning the degraded images [6].

3. Methodologies

Our overall goal is to,

1) implement the three classical image classification models on the selected dataset.

2) verify that classical CNN image classification methods suffer significant degradation on low-quality images.

3) implement a model/method that combines low-level image processing with image classification networks to achieve better low-quality image classification results.

4) conduct an ablation study to improve the framework in 3) for better-degraded image classification performance.

3.1. Baseline Models

We would like first to train and test several baseline models, AlexNet [4], VGG [10] and ResNet [11], on normal and degraded image datasets. The result for degraded image classification is predicted to be worse than normal image classification for classical models.

3.2. Low-level and high-level vision models

Several works have been done for connecting low-level processing and high-level classification tasks. We plan to implement one of the following and conduct an ablation study to improve the framework.

WaveCNet [1]: Due to aliasing effects from the frequently employed down-sampling operation, CNNs for image classification in deep networks may have weak noise-robustness. To preserve the fundamental object structures and prevent noise propagation, we propose integrating CNNs with wavelets by replacing the down-sampling operations with discrete wavelet transform (DWT). It can provide better noise-robustness and adversarial robustness. The suppression of the aliasing effect is also an advantage.

Group-wise Inhibition-based Feature Regularization [2]: Focusing more on the most discriminative regions while ignoring the auxiliary features when learning, CNN is susceptible to degraded images with even very modest alterations, which results in a lack of feature diversity for final judgment. By using group-wise inhibition, we suggest dynamically suppressing significant CNN activation values. CNN is finally directed to acquire richer discriminative features hierarchically for robust classification.

Deep Denoising & High-Level Vision Tasks [3]: We propose a jointly trained network connecting two modules for image denoising and various high-level vision tasks, which are conventionally addressed independently in computer vision. It improves robustness and generalization.

3.3. Datasets

To perform the baseline models training and evaluation, we decide to choose the Oxford 17 flowers dataset [13], which has 17 categories of common flowers and 80 images for each category.

For an additional part of this project, we plan to use ImageNet [8] as the high-quality dataset and ImageNet-C [7] as the low-quality dataset. ImageNet-C [7] is an open-source data set that consists of algorithmically generated corruptions (blur, noise) applied to the ImageNet [8] test set.

The dataset will be divided into training set, validation set, and test set at random, with training and validation set used for setting hyperparameters and the test set for assessing performance.

4. Work distribution and timelines

By Sep 16:
- Submit proposal
- Finish work distribution and get familiar with the whole project ideas

By Sep 30:
- Review classical image classification models (AlexNet, VGG and ResNet) and their implementations
- Set up basic frameworks for training the three classical models

By Oct 14:
- Implement and train the classical models
- Finish classical models' evaluations and results analysis

Oct 30:
- Submit midterm report
- Choose suitable models and dataset for the extended part
- Understand the ideas in the advance model papers
- Set up frameworks for training

By Nov 12:
- Implement the low-quality image classification methods and train models
- Conduct ablation study

By Nov 22:
- Analyze experiment results and prepare final presentation

By Nov 30:
- Write the final report

We are separated into two sub-groups, group 1 (Ding Zhui, Qi Yijiazhen, Su Changyue) and group 2 (Xu Haozhou, Long Kehan). Group 1 will be responsible for training two baseline models together with the analysis of one basic model, while group 2 will train one baseline model and conduct analysis on two baseline models. To implement and analyze the low-quality classification method, group 2 will try to build and training the newly proposed model, and group 1 will do the following analysis.

References

[1] Qiufu Li, Linlin Shen, Sheng Guo, and Zhihui Lai. WaveCNet: Wavelet Integrated CNNs to Suppress Aliasing Effect for Noise-Robust Image Classification. arXiv:2107.13335v1 , 2021. 2

[2] Haozhe Liu, Haoqian Wu, Weicheng Xie, Feng Liu, and Linlin Shen. Group-wise Inhibition based Feature Regularization for Robust Classification. arXiv:2103.02152, 2021. 2

[3] Ding Liu, Bihan Wen, Xianming Liu, Zhangyang Wang, and Thomas S. Huang. When Image Denoising Meets High-Level Vision Tasks: A Deep Learning Approach. arXiv:1706.04284, 2017. 1, 2

[4] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Communications of the ACM* 60.6 (2017): 84-90. 1

[5] Yang Wang, Yang Cao, Zheng -Jun Zha, Jing Zhang and Zhiwei Xiong. Deep Degradation Prior for Low-Quality Image Classification. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 11046-11055, 2020. 1

[6] Steven Diamond, Vincent Sitzmann, Frank Julca-Aguilar, Stephen Boyd, Gordon Wetzstein, Felix Heide. Dirty Pixels: Towards End-to-End Image Processing and Perception. arXiv:1701.06487, 2021. 1

[7] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. arXiv preprint arXiv:1903.12261, 2019. 1, 2

[8] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee, 2009. 2

[9] Wei Zhang, Shengnan Hu, Kan Liu, and Zhengjun Zha. Learning compact appearance representation for video-based person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(8):2442–2452, 2018. 1

[10] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014. 1

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recog- nition, pages 770–778, 2016. 1

[12] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10):1345–1359, 2009. 1

[13] Maria-Elena Nilsback and Andrew Zisserman. 17 category flower dataset, 2006. 2