Programming Assignment - 1

Advanced Computer Vision CAP6412

February 20, 2019

1 Overview

In this assignment your goal is to implement the idea from Jakab et. al. [4] for 2D key-point detection. In [4], the authors proposed an unsupervised approach for detecting object landmarks. The authors choose the input image pairs by performing some transformations and have experimented on facial images. Your goal is to reproduce the paper and perform experiments on facial images. Apart from this, you will also modify the proposed network architecture and experiment on a different dataset.

This assignment consists of two parts:

1. Part-1: Implement the method proposed by the authors for key-point detection on CelebA dataset [5] and reproduce the results. The source code for this implementation is available at [1] which you can refer to, but you must implement this on your own. You can re-use any existing code from other sources which has also been used in this repository [1]. But all the code written in this repository has to be your own code.

2. Part-2: In this part your goal is to make changes to the network architecture and experiment with a different dataset. You are free to choose any changes you want but it should not be very trivial like, adding one more conv layer to the network. Some suggestions are, adding inception blocks [6], resnet skip blocks [3], attention blocks [7], etc. You can also adapt from well known existing networks. You have to perform your experiments on ShapeNet dataset [8] for three categories; car, plane, and chair. You can download the dataset from,

   (a) https://storage.googleapis.com/discovery-3dkeypoints-data/cars_with_keypoints.zip

   (b) https://storage.googleapis.com/discovery-3dkeypoints-data/planes_with_keypoints.zip

   (c) https://storage.googleapis.com/discovery-3dkeypoints-data/chairs_with_keypoints.zip

   More details regarding the dataset can be found at,
   https://github.com/tensorflow/models/tree/master/research/keypointnet

2 Evaluation

1. Write-up : Clarity, structure, references

2. Implementation : Quality of implementation

3. Evaluation and Results : Insights, discussions, analysis

3 Submission

You will submit a final report on your project which will be in a research paper format. The report should be 6-8 pages long and use this template [2] for consistency and you will not have to worry about the font size and formatting. The following is a suggested structure for your report which you should follow,

1. Title and name of the student
2. Abstract
3. Introduction: Problem introduction and motivation behind the changes
4. Background and related work : Relevant literature, discuss at least 4-5 most relevant works
5. Method : Proposed implementation with details
6. Experiments : Your experiments, evaluation plan and the results
7. Discussion : Discussion and insights on your results
8. Conclusion : Possible future direction
9. References

You will submit the following items as a single (zip) file [your_name_assignment_1.zip],

1. Report : A pdf file for your final report. Please don't forget to put your name on the report and try to avoid doc files.
2. Source code : Implementation of your project. Please do not submit datasets along with this as some of the students did for assignment 0. This submission should not be bigger than 10MB.
3. Supplementary items [optional] : This includes additional results which can not be added to your report, such as videos, demos, etc.

4 Important Dates

1. Start of the project : Feb 21, 2018
2. Submission : March 24, 2018

References

[1] Paper template. https://github.com/tomasjakab/imm/.

[2] Paper template. http://cvpr2019.thecvf.com/files/cvpr2019AuthorKit.zip.

[3] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[4] T. Jakab, A. Gupta, H. Bilen, and A. Vedaldi. Unsupervised learning of object landmarks through conditional image generation. In *Advances in Neural Information Processing Systems*, pages 4020–4031, 2018.

[5] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738, 2015.

[6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008, 2017.

[8] J. Wu, C. Zhang, T. Xue, B. Freeman, and J. Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In *Advances in neural information processing systems*, pages 82–90, 2016.