Semantic Segmentation using DensePose: Graph Convolutional Networks

Category: Semantic Segmentation

Team: Pierre Nowicki - SUNet ID: npierre9

1. Problem to investigate

Semantic segmentation is a fundamental problem in computer vision, and aims to assign an object class label to each pixel in an image. It has numerous applications including autonomous driving, augmented- and virtual reality and medical diagnosis.

An inherent challenge in semantic segmentation is that pixels are difficult to classify when considered in isolation, as local image evidence is ambiguous and noisy. Additionally, human part segmentation and dense pose estimation require details at the instance level which involve the sequential process of detection, segmentation and estimation.

The DensePose task involves simultaneously detecting people, segmenting their bodies and mapping all image pixels that belong to a human body to the 3D surface of the body.

[1] and [2] strategies revolve around finding dense correspondence by partitioning the surface through regression or region-based CNN approaches to the appearance of a human and thus instances details.

2. Proposed approach

Graph Convolution Network have been used to capture dependencies generated from non-Euclidean domains or complex relationships and interdependence between objects.

For example, [3] investigates image co-segmentation where each input image is over-segmented into a set of superpixels. Those superpixels categorisation feed a weighted graph representing spatial adjacency and both intra-image and inter-image feature similarities.

Similarly, [4] models the global context of the input feature by modelling two orthogonal graphs in a single framework. The first component models spatial relationships between pixels in the image, whilst the second models interdependencies along the channel dimensions of the network's feature map.

Thus, I would like to apply a similar approach of GCN to the task of DensePose classification using the COCO data source and assess if a graph structure can capture better the link between the semantic segmentation (person) and the instance-level segmentation.

2.1. Data Source

The DensePose-COCO dataset, is a large-scale ground-truth dataset with image-to-surface correspondences manually annotated on 50k COCO images, collecting more then 5 million manually annotated correspondences.

2.2. Evaluation of results

Dense pose estimation requires human part segmentation and pose estimation, where one predicts continuous part labels of each human body.

Thus,the metrics to compare the efficiency of the proposed architecture with classical CNN architecture like ResNet will be at two level:

- For semantic segmentation (person), we generate a multi-person mask and evaluate standard mean intersection over union (mIoU).
- For instance-level performance, the Average Precision based on part (APp) for multi-human parsing evaluation uses part-level pixel IoU of different semantic part categories within a person instance.

References

[1] R. A. Güler, N. Neverova, I. Kokkinos, Densepose: Dense human pose estimation in the wild, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7297–7306.

[2] L. Yang, Q. Song, Z. Wang, M. Jiang, Parsing r-cnn for instance-level human analysis, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[3] A. Hati, S. Chaudhuri, R. Velmurugan, Image co-segmentation using graph convolution neural network, 2018.

[4] L. Zhang, X. Li, A. Arnab, K. Yang, Y. Tong, P. H. S. Torr, Dual graph convolutional network for semantic segmentation (2019). `arXiv:1909.06121`.