Review on “Deep Watershed Transform for Instance Segmentation”

Authors: M. Bai, R. Urtasun

# Short Summary

This paper discusses combining the classical watershed transform with deep learning for the task of instance segmentation. The intuition behind this revolves around using image gradients to create energy basins that correspond to homogeneous areas in an image. However, the traditional method of performing the watershed transform can lead to over-segmentation of an image which is difficult to resolve without good heuristics. The author instead proposes to learn the energy map of the transform such that each energy basin corresponds to a single object instance with dividing ridges at the same height in the energy domain. The net result would be an easily trained end-to-end model that operates in constant runtime irrespective of the number of object instances. The proposal, dubbed the Deep Watershed Transform, describes two networks: the Direction Network and the Watershed Transform Network.

The Direction network accepts an RGB image as an input with the binarized segmentation (from PSPNet) added as a fourth channel and is responsible for learning the direction of descent of the watershed energy. It does so by parameterizing a unit vector describing the direction of the nearest boundary. The architecture of this network is inspired by VGG-16 with modifications for pooling, multi-scale information aggregation and up sampling. The Watershed network accepts the two-channel unit vector map output of the Direction Network as input and applies a discretized transformation map with K =16 possible energy values.

The model achieves more than double the mean AP (test set) on Cityscapes compared to state-of-the-art (achieving 19.4% versus the previous best of 8.9%).

# Main Contributions

* Proposed the Deep Watershed Transform, a deep neural model that learns to apply the watershed transformation without over-segmenting results.
* Proposed architecture of the Direction and Watershed transform networks respectively
* Doubled the performance of the previous state-of-the-art on Cityscapes instance segmentation
* Demonstrates how model performance can be further improved by changing semantic segmentation source and using instance ordering techniques

# High-Level Evaluation of Paper

The focus of this paper is the Deep Watershed Transformation and it conveys relevant information concisely. I liked that the authors opted to present the material in this manner instead of proposing multiple variations of the model and performing excessive experiments. With that said, I did not gain a comprehensive understanding of how the Watershed transformation works but rather a general intuition of what it accomplishes which I think is sufficient for a paper of this nature. In particular, I was confused by the section on Energy Cut and Instance Extraction which used terminology that was not previously described.

# Discussion on Evaluation Methodology

The paper made special note of the metrics used in the evaluation and specifically discussed how the AP may not be an accurate indicator of performance. I liked that they referenced an alternative, the mean weighted coverage score, and provided both results to the reader. Furthermore, they also provided a good explanation why the alternative is more suited for the task. The actual evaluation performed is done on the Cityscapes instance segmentation dataset and is compared to other well-established results. Performance is also compared on a per-class basis demonstrating that this new model performs better on most classes. As minor extensions, the paper also shows how performance is altered by using a different semantic segmentation source (LRR vs PSPNet) two instance ordering techniques. Oddly, only the AP scores are given for both extensions instead of both AP and muCov.

# Possible Directions for Future Work

Investigate better semantic segmentation sources and instance ordering techniques to investigate how to improve model performance further.