TensorMask: A Foundation for Dense Object Segmentation

Abraham Jose

1 Summary

The author is proposing Tensormask, a dense sliding-window instance segmentation framework used for image segmentation. The model treats the dense instance segmentation as a prediction task over 4d tensor and that explicitly captures the geometry of object and use novel 4D operations which improves the performance from the baseline dense instance segmentation algorithms. Tensormask was able to achieve performances close to 'detect-then-segment' systems such as mask-RCNN and successfully was able to gap the performance between the dense instance segmentation and state-of-the-art Mask-RCNN. The model was tested in MS-COCO instance segmentation dataset.

2 Good points

Author was able to use multiple the high dimensional tensor that creates a pyramid of feature maps in 4D. They uses a structured high dimensional tensor to represent images in a set of densely shifting window. Each pixel in the 4D tensor is a geometric structure with it's own spatial dimension which ensures that as many possibilities are considered as compared to 2000 bounding box in RPN networks. Natural representation and aligned representations which is intuitive and is being eventually converted to natural representation which is intuitive and easy to parse information.

3 Weak points

Using high dimensional tensor for extraction of the feature will result in high requirements for computations. Also, the optimization in a high dimensional feature vector that represents the model will be hard and it is directly proportional to the complexity of the data being handled. So the optimization is really hard in this case and even if the natural representation of the tensor is intuitive, training such a network will be hard and takes longer period of time. Also, Tensormask is 4x slower than Mask-RNN model which is important as we consider the framework for real-time, real-world capability. Matching the Detectron code to reflect on Tensormask's implementation details(FPN average fusion, 1K warmup and l1 box loss) might not be the best way to compare the performance. It might cause different effects in both networks.

4 Questions

The computational complexity required for the model is crucial to compare with Faster-RCNN and among the inference time, the tensormask model performs worse. Could you comment on the performance and adaptability of this framework in real-world real-time scenarios?

5 Ideas

Using Tensormask and using the high-dimensional vector to represent the input image and compare it with the representational ability of Mask-RCNN's to decompose and to represent the image.