

Research Questions

Amartya Dutta

1 Relevant Papers

My area of interest lies in Computer Vision. Specifically, my current research statement is Weakly Supervised Semantic Segmentation. There are several interesting papers in that particular domain of Computer Vision. Three very relevant papers are listed below.

The first paper very relevant to this topic is titled, Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. The paper's authors are Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. It was submitted as an ICLR Workshop Poster and published in 2014.

A second paper would be Learning Deep Features for Discriminative Localization. The authors of the paper are Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. This paper was submitted to CVPR and was published in the year 2016.

Another very relevant article is Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. The authors of the paper are Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. This was submitted to ICCV and was published in 2017.

2 Research Questions

The challenge that Weakly Supervised Semantic Segmentation tries to solve is to remove the need for manual annotations of ground truth segmentation masks to train segmentation models. Therefore, researchers started trying out different methods to see if the existing models, such as the Classifiers could assist in the segmentation task.

The first paper, Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps, aimed to answer the research question of how to interpret the internal representations of deep convolutional neural networks used for image classification. Specifically, it aimed to find the answer to two main questions. These are, understanding what features the network has learned to recognize in the images and visualizing what parts of an image are most important for the network's decision-making process. Finding the answers to these questions would enable one to understand which regions of an object are usually detected and hence could be used in detecting the entire object.

In Learning Deep Features for Discriminative Localization, the authors aimed to answer the research question of how to localize the discriminative image regions that are responsible for deep convolutional neural networks' predictions. The goal was to be able to identify the regions of an image that contribute most to the network's classification decision. This information would eventually help to perform image classification and object localization.

The third paper, Grad-CAM aimed at answering the research question of how to generate visual explanations for the predictions made by deep convolutional neural networks (CNNs) used for image classification. It improved upon the concept of the Class Activation Map introduced in the previous

paper. It tried to further provide insights into how can we ensure that the visual explanations are accurate and reliable.

3 Evaluation of Research Questions

Research Questions are always fun to ask, however, they need to meet certain requirements to qualify as a good research question. And only then does the problem get sufficient attention from the research community. Otherwise, it is most often not worthwhile to pursue such a problem. Therefore there are certain evaluation metrics.

The first paper was Interesting because it was the first to shed light on the internal representations of deep convolutional neural networks used for image classification. Everyone used Convolutional neural networks for classification however no other work had previously pondered upon the question of what and how exactly are objects detected using classifiers. Therefore, they implemented a method they termed as Saliency Map. Therefore, their question was also answerable because it was quite specific. Furthermore, it was something none had yet answered in the Vision community, so it was also repeatable. They generated standard metric results on benchmark datasets, thus making their work also measurable. Furthermore, the set of experiments necessary to achieve the results that they had was not too time-consuming. It was the idea formulation that was the biggest hurdle. Thus, their work was appropriately scoped too.

The second paper took it one step forward and tried to use the Convolutional Neural Networks' spatial properties to detect the important regions of an image. This was therefore interesting because there was a gap in the existing literature regarding this idea. In their paper, they introduced the concept of a Class Activation Map (CAM), which effectively discovered the regions of highest interest in an object. Accordingly, their research question was also answerable because it was quite specific. Furthermore, it was something of immense interest and importance to the researchers in this field, so it was also repeatable. They also generated standard metric results on benchmark datasets, thus making their work measurable. Furthermore, their work was appropriately scoped too because it includes a set of feasible experiments.

The final paper improved upon the idea of the Class Activation Map. They introduced an explainable method that made the previous work much more generalized and intuitive. Therefore it was interesting because not only did it fill a gap in the existing literature but also resulted in the development of a lot of improved methods in the future. Similarly, by providing insight into visual explanations for the predictions made by deep convolutional neural networks, the question was answerable because it was quite specific and repeatable. Furthermore, they also generated standard metric results on benchmark datasets, thus making their work measurable. Finally, the set of tasks required to achieve the results was feasible, making their question appropriately scoped as well.

4 Research Goal

Since I have been working on this area for a few months now, I have come across some shortcomings which I believe could be potential research questions I might be working towards.

One of the questions tries to find answers to how Saliency Maps are different from CAMs. Adding to it, we want to find out if Saliency Maps are able to detect regions of an object that CAMs are unable to and if so under what conditions. This problem is something there is no clear answer to yet. CAMs are most often used for the task of Object Segmentation, primarily because Saliency Maps are noisy. However, the aim is to conduct experiments to discover if certain regions of interest

are discovered. Therefore, the question is interesting and also answerable because it is quite specific. Furthermore, it is a question many in the field would be interested in knowing the answers to thus making it repeatable. The evaluation and the results of the experiment would involve standard metrics for Image segmentation and on benchmark datasets, hence it is measurable. Finally, though an extensive experiment to find answers to this question may span over a few months, it would be appropriately scoped.

Another question we want to find answers to is whether the Non-Discriminative regions of an object can be detected under certain conditions. Usually, CAMs and Saliency Maps detect the Discriminative regions of the object. Therefore, we want to perturb the detection method in order to find insights into situations when even the Non-Discriminative regions would get detected. Successfully answering this would eventually lead to better performance in Weakly Supervised Semantic Segmentation. Thus, the research question we raise is interesting, answerable, and also repeatable. The data used be again the same benchmark datasets and the results generated would be based on standard metrics for such tasks. Therefore, this question is measurable too. While extensive experiments like this might span over a few months, it is still appropriately scoped for my Masters's Thesis.