Machine Learning Engineer Nanodegree

Capstone Proposal

Tapish Rathore

Proposal : Kaggle's Carvana Image Masking Challenge

Domain Background

Pixel-level semantic segmentation is the task of assigning a class label to each pixel in an image. It involves classifying an object in an image as well as detecting its location with the highest possible resolution. Humans are exceptionally good at this task. For example, when we see a pigeon sitting on a tree, we have no difficulty in distinguishing which part comprises the tree and which the pigeon, even if we only have a single image and are not allowed to move. There are several use cases for pixel level segmentation in the field of computer vision, such as detecting road signs[1], quantification of tissue volumes[2], study of anatomical structure[3] and land cover classification[4]. It has been used extensively in medical imaging[5] and autonomous car driving[6]. With recent advancements in training CNNs to learn image features for detecting and classifying objects in images[7], efforts have been made to use CNNs for semantic segmentation as well[8]. This has resulted in several state-of-the-art CNN architectures such as Mask R-CNN[9] and SegNet[10].
The author is also personally motivated to work in this domain as he has been involved in designing a computer vision pipeline for classification of food grains[11] in the past and had implemented a non-learning approach to segmentation. Now he would like to explore it from a deep learning perspective.

Problem Statement

The problem statement is specified on Kaggle's webpage for the Carvana Image Masking Challenge[12]. The challenge is to develop an algorithm which distinguishes each pixel in a given image as either belonging to a car or the background. The effectiveness of the algorithm will be calculated using the mean of Dice coefficient[13] over all images. The Dice coefficient is a metric to compare the pixel-wise agreement between the ground truth and the mask predicted by the algorithm. A potential solution for the given task is to train a Convolutional Neural Network on the given training data to generate an image mask classifying each pixel. The problem statement can also be expanded to analyze if the algorithm is robust enough to segment cars in images of other datasets (explained below).

Datasets and Inputs

The dataset used in the competition is provided on the competition website[14].
However, I would like to test the final algorithm on images of cars from other datasets to test the robustness of the algorithm. I would also like to analyze the causes of the algorithm being less or more robust by examining the properties of the training set given in the Kaggle competition. For example, the images given for the competition are shot in a studio. Does that affect detection of cars in outdoor images? If so, why?
I have compiled car images and ground truth segmentation masks from the Pascal VOC Database[15] – specifically from the TU Darmstadt Database[16] and the TU Graz-02 Database[17]. I will be sharing these through a Google Drive link[18].

Solution Statement

Using deep learning architectures, we can learn features of a car from the training data and produce a model which can produce a reasonably good segmentation mask. The selection of the model architecture will depend on factors like training data size and amount of computational resources available. The segmentation masks produced would be compared to the ground truth using Dice coefficient. This is a measurable and replicable process.

Benchmark Model

The benchmark model can be a default CNN model which has been used previously for pixel-wise semantic segmentation such as FCN-AlexNet[19]. It can be trained on the competition dataset and compared with the proposed solution using Dice coefficient.

Evaluation Metrics

The Dice coefficient is the proposed evaluation metric for this task. It can be calculated using the formula –

$$\frac{2 * |X \cap Y|}{|X| + |Y|}$$

where X is the set of pixels predicted as belonging to the car by the proposed solution and Y is the ground truth. When both X and Y are empty, the Dice coefficient is assumed to be 1.

Project Design

The first phase of the project would involve a literature review to determine a suitable CNN architecture, keeping in mind constraints in terms of training data and computational resources. Larger CNN architectures might overfit on the dataset and would require a lot of time to train.
The second phase of the project would be concerned with implementation details. Preprocessing techniques for training data and image augmentation techniques applied to increase the number of images in the dataset will be required. A number of hyper-parameters would need to be tuned by performing extensive parameter search. If the model fails to converge, an extensive analysis of the

training error logs would be required to recognize the cause of failure. If the model converges, it will be compared to the benchmark model and submissions from other teams in the competition to determine if it should be further improved. The model will also be tested on the external dataset and analyzed to determine causes of its performance by examining its activations for each layer.

References

1. S. Maldonado-Bascon, S. Lafuente-Arroyo, P. GilJimenez, H. Gomez-Moreno, and F. LopezFerreras, "Road-sign detection and recognition based on support vector machines," *Intelligent Transportation Systems, IEEE Transactions on, vol. 8, no. 2, pp. 264–278, Jun. 2007. [Online]. Available: http://ieeexplore.ieee.org/xpls/ abs_all.jsp?arnumber=4220659*
2. Larie SM, Abukmeil SS. 1998. "Brain abnormality in schizophrenia: a systematic and quantitative review of volumetric magnetic resonance imaging studies." *J. Psychol.* 172:110–20
3. Worth AJ, Makris N, Caviness VS, Kennedy DN. 1997. Neuroanatomical segmentation in MRI: technological objectives. *Int. J. Pattern Recognit. Artif. Intell.* 11:1161–87
4. C. Huang, L. Davis, and J. Townshend, "An assessment of support vector machines for land cover classification," International Journal of remote sensing, vol. 23, no. 4, pp. 725–749, 2002
5. D. L. Pham, C. Xu, and J. L. Prince, "A survey of current methods in medical image segmentation," *Annual Review of Biomedical Engineering, vol. 2, no. 1, pp. 315–337, 2000, pMID: 11701515*. [Online]. Available: http:// dx.doi.org/10.1146/annurev.bioeng.2.1.315
6. Treml, Michael, et al. "Speeding up semantic segmentation for autonomous driving." *NIPSW* 1.7 (2016): 8.
7. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.
8. Zhao, Feng et al. "A survey on deep learning-based fine-grained object classification and semantic segmentation." *International Journal of Automation and Computing 2017.*
9. Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross B. Girshick "Mask R-CNN." *Clinical Orthopaedics and Related Research, March 2017.*
10. V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *arXiv:1511.00561, 2015.*
11. Nebulaa Innovations, [Online] *http://www.nebulaa.in/*
12. Kaggle's Carvana Image Masking Challenge, [Online] *https://www.kaggle.com/c/carvana-image-masking-challenge*
13. Dice coefficient, [Online] https://en.wikipedia.org/wiki/S%C3%B8rensen%E2%80%93Dice_coefficient

14. Kaggle's Carvana Image Masking Challenge Dataset, [Online] *https://www.kaggle.com/c/carvana-image-masking-challenge/data*
15. Pascal Object Recognition Databease Collection, [Online] *http://host.robots.ox.ac.uk/pascal/VOC/databases.html*
16. TU Darmstadt Database, [Online] *http://www.mis.informatik.tu-darmstadt.de/leibe*
17. TU Graz-02 Database, [Online] *http://www.emt.tugraz.at/~pinz/data/GRAZ_02/*
18. Compiled Car Images and Ground truths [Online] *https://drive.google.com/drive/folders/0B0Te0p-dLjuqc1VRZEp0d3k1Q3c?usp=sharing*
19. Long, J., Shelhamer, E., and Darrell, T. "Fully convolutional networks for semantic segmentation." *CoRR, abs/1411.4038, 2014*