EAI 6020: AI System Technologies

Week 2:
Training PSPNet for Semantic Segmentation on Traffic Post Images

Submitted To:
Prof. Siddharth Rout, Faculty Lecturer

Submitted By:
Abhilash Dikshit
Murtaza Vora
Gunjan Paladiya

Academic Term: Winter 2024
Northeastern University, Vancouver, BC,Canada
Master of Professional Studies in Analytics

March 02, 2024

Comparison Parameters, Vanishing and Exploding Gradients, and Regularization in Training a Deep CNN Model using PSPNet on Traffic Images

I. Introduction:

This report aims to investigate the impact of various parameters, including initialization methods, learning rates, batch sizes, and regularization techniques, on the training of a deep convolutional neural network (CNN) model using PSPNet (Pyramid Scene Parsing Network) for semantic segmentation of traffic images. The focus is on understanding how these factors influence the training process, particularly in addressing issues such as vanishing and exploding gradients and improving the generalization ability of the model through regularization.

II. Comparison Parameters:

Several parameters play a crucial role in training deep CNN models effectively. These include:

1. Initialization Methods: The choice of initialization method for model parameters (e.g., Xavier, He initialization) can significantly impact the convergence speed and final performance of the model.

2. Learning Rates: The learning rate determines the step size during gradient descent optimization. Setting an appropriate learning rate is essential for ensuring stable training and preventing issues such as divergence or slow convergence.

3. Batch Sizes: Batch size affects the stability and efficiency of training. Larger batch sizes often lead to faster convergence but may require more memory and computational resources.

4. Regularization Techniques: Regularization methods such as L1/L2 regularization, dropout, and batch normalization are employed to prevent overfitting and improve the generalization ability of the model.

III.   Vanishing and Exploding Gradients:

Vanishing and exploding gradients are common issues encountered during training deep neural networks, particularly in architectures with many layers. Vanishing gradients occur when gradients become extremely small as they propagate backward through the network, leading to slow convergence or stagnation in learning. Exploding gradients, on the other hand, occur when gradients become excessively large, causing numerical instability and divergence during optimization.

IV.   Regularization Techniques:

Regularization techniques are employed to address overfitting and improve the generalization ability of deep CNN models. Some commonly used regularization methods include:

1. L1/L2 Regularization: Penalizes large weights by adding a regularization term to the loss function, encouraging simpler models and reducing overfitting.

2. Dropout: Randomly drops a fraction of neurons during training to prevent co-adaptation of features and encourage robustness.

3. Batch Normalization: Normalizes activations within each mini-batch to stabilize training and accelerate convergence.

V.   Conclusion:

In conclusion, the training of a deep CNN model using PSPNet for semantic segmentation of traffic images requires careful consideration of various parameters, including initialization methods, learning rates, batch sizes, and regularization techniques. Proper selection and tuning of these parameters are essential for ensuring stable training, addressing issues such as vanishing and exploding gradients, and improving the generalization ability of the model. Experimentation and thorough analysis of different configurations are crucial for achieving optimal performance in training deep CNN models for semantic segmentation tasks.

<u>PART B</u>

Topic: Training PSPNet for Semantic Segmentation on Traffic Post Images

I. Introduction:

This report documents the process of training a semantic segmentation model using PSPNet (Pyramid Scene Parsing Network) on a dataset consisting of traffic post images. The goal of this project is to develop a model capable of accurately segmenting images to identify various classes of objects present in the scene, such as sky, trees, roads, buildings, and foreground objects.

II. Data Preparation:

The dataset used for training consists of images captured from traffic posts, along with corresponding annotations that specify the class labels for different regions within the images. These annotations were converted into semantic segmentation maps, where each pixel is assigned a class label representing the object category it belongs to. The dataset was then split into training and validation sets for model training and evaluation.

III. Model Configuration:

To adapt the PSPNet model for our dataset, several modifications were made to the configuration file. These modifications included adjusting the normalization method to Synchronized Batch Normalization (SyncBN) for improved stability during distributed training. Additionally, the number of classes in the model's decoding and auxiliary heads was modified to match the number of classes in our dataset.
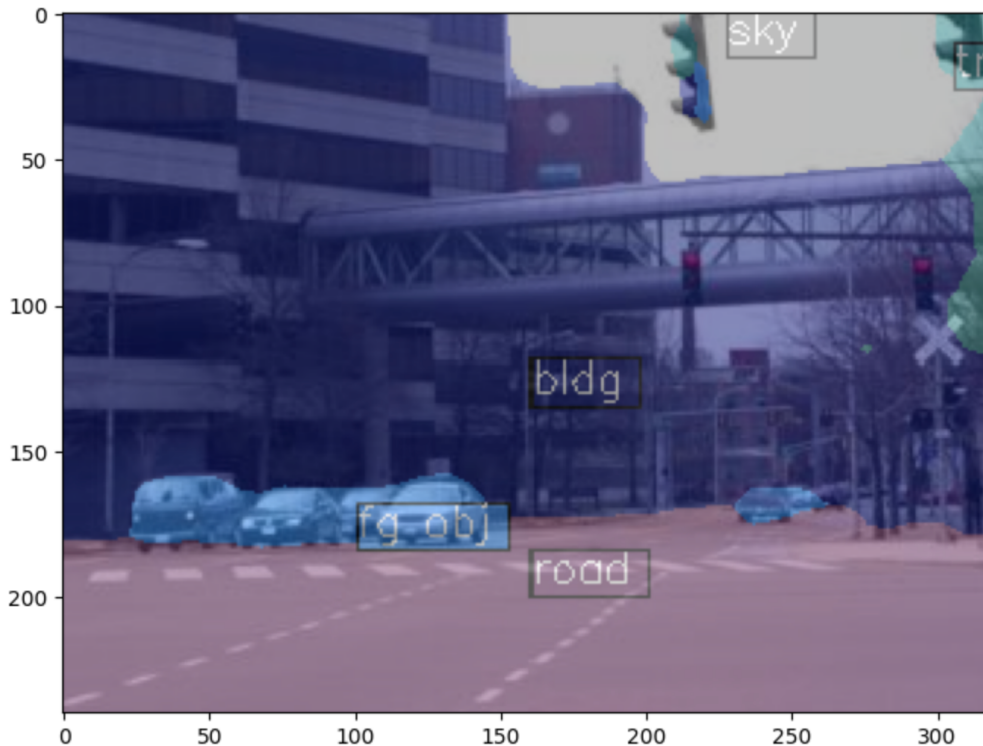
IV. Training Process:

The model was trained using the modified configuration and pre-trained weights obtained from a Cityscapes dataset. Training was performed for a total of 200 iterations, with validation conducted every 200 iterations to monitor the model's performance. The training process involved data augmentation techniques such as random resizing, cropping, and flipping to enhance the model's robustness and generalization ability.

V. Results:

After training, the model achieved satisfactory results in segmenting traffic post images, accurately delineating various objects and regions within the scene. The segmentation maps generated by the model provided valuable insights into the spatial distribution of different classes of objects, enabling better understanding and analysis of the traffic post images.

VI. Inference and Visualization:

The trained model was evaluated on sample images from the dataset using inference. The model demonstrated its ability to accurately segment objects within the images, producing visually appealing results. These segmentation maps can be further analyzed and utilized for various applications, such as traffic monitoring, urban planning, and infrastructure management.

VII. Conclusion:

In conclusion, the training of PSPNet on traffic post images for semantic segmentation proved successful, with the model exhibiting promising performance in accurately segmenting objects within the scenes. The trained model can serve as a valuable tool for analyzing and understanding traffic post images, enabling better decision-making and planning in urban environments.

VIII. Code File

https://colab.research.google.com/drive/1_0Jf5IcJJiwO9oz4Ma6T9uWZ2KwXgpD5?usp=sharing

IX. References

1. Open-Mmlab. (n.d.). GitHub - open-mmlab/mmsegmentation at master. GitHub. https://github.com/open-mmlab/mmsegmentation/tree/master

2. Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2016, December 4). Pyramid Scene Parsing network. arXiv.org. https://arxiv.org/abs/1612.01105