Alma Mater Studiorum · University of Bologna

department of computer science and engineering

Degree Programme - Artificial Intelligence

PIDNet: A Real-Time Semantic Segmentation network

Mini Project for
Machine Learning for Computer Vision

Presented by:
Asif Khan Pattan

Academic Year 2022 - 2023

Overview

This mini project is the implementation of "PIDNet : A real time semantic segmentation network inspired by PID controllers" using PyTorch. PIDNet is the current state-of-the-art real time semantic segmentation network which outperformed other models on Cityscapes and Camvid dataset. There are 3 variations of PIDNet(small, medium and large) based on the complexity and architecture used. This implementation is mainly focused on PIDNet-medium and PIDNet-small. The Source code is adopted from the authors official github page. In this implementation the PIDNet is fine-tuned on ADE20K dataset. Colab pro is used for training the model along with my personal laptop. This report contains the details of the dataset used along with implementation and fine-tune results.

Contents

Chapter 1

Introduction

1.1 DataSet

The dataset for finetuning is taken from ADE20K which contains more than 20K scene-centric images exhaustively annotated with objects and object parts. Dataset is divided into 20K images for training, 2K images for validation, and another batch of held-out images for testing. There are totally 150 semantic categories included for evaluation, which include stuffs like sky, road, grass, and discrete objects like person, car, bed. Note that there are non-uniform distribution of objects occuring in the images, mimicking a more natural object occurrence in daily scene.

1.2 Experimental Setup

Google Colab pro with 16GB V100 GPU and 50GB RAM is used primarly along with personal laptop with 4GB Nvidia GeForce RTX 3050 and 16GB RAM. SGD is used as optimizer through out the experiments.

Models Comparison

PIDNet-medium architecture is opted for model 1 and trained on colab pro while PIDNet-small architecture is opted for model 2 and trained on personal laptop.

Model 1

- Training image size: 512×512
- Dataset Train: 20,210, Val: 2000
- Classes: 150
- Batch size: 8
- Learning rate: 0.001
- Weight decay: 0.00005
- Epochs: 11
- OHEM: False
- Nasterov : False
- Test image size: 512×512

Model 2

- Training image size: 512×512
- Dataset Train: 6736, Val: 666
- Classes: 21
- Batch size: 4
- Learning rate: 0.001
- Weight decay: 0.0005
- Epochs: 16
- OHEM: True
- Nasterov : True
- Test image size: 512×512

Chapter 2

Results

2.1 Model 1

Figures 2.1 and 2.2 shows training and validation loss curves for model 1. Validation is performed for every 2 epochs so the x axis of figure 2.2 is in range 0 to 5 where each unit is equal to 2 epochs. Model 1 achieved an mIoU of 34 over the validation set.

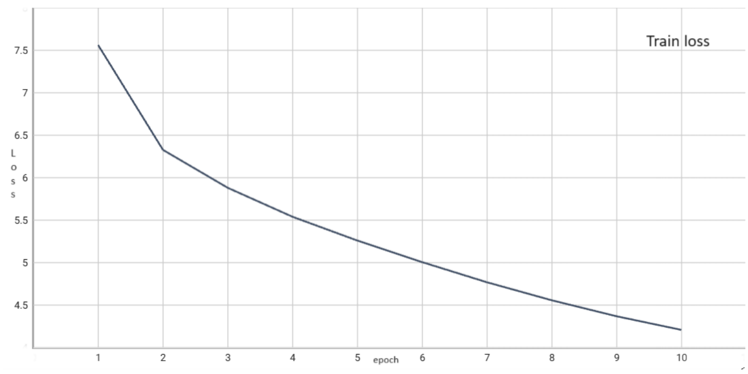

Figure 2.1: Train loss of model 1

2.2 Model 2

Model 2 has achieved much better results than model with with an mIoU of 48.32 over validation set considering only 21 classes and ignoring the rest.
From figures 2.4 and 2.5 we could clearly see that loss has reduced

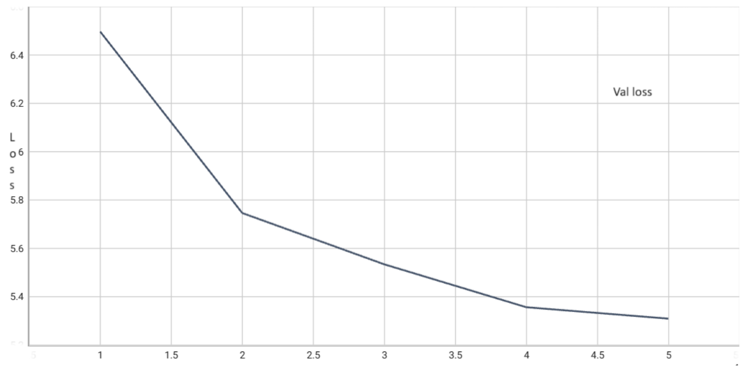

Figure 2.2: Validation loss of model 1

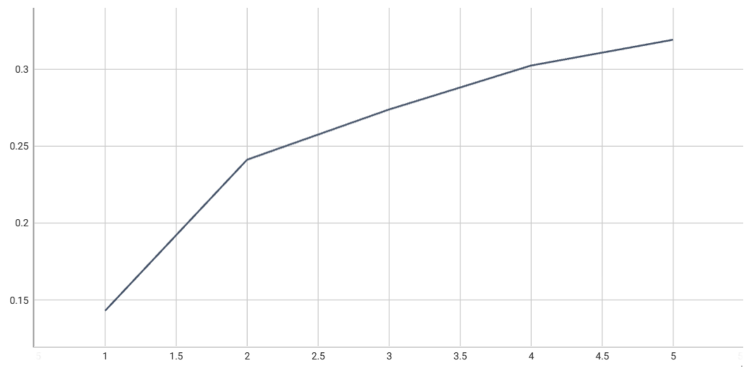

Figure 2.3: Validation mIoU of model 1

drastically in the first 2 epochs and then followed a flat curve with little progress over time.

Below are some of the prediction results of model 2 on validation set images. Figure 2.8 took 3.4sec(first image always longer time due to model initialization) and 2.10 took 0.015 sec and 2.12 took 0.029 secs. This proves the real time performance of PIDNet architecture.

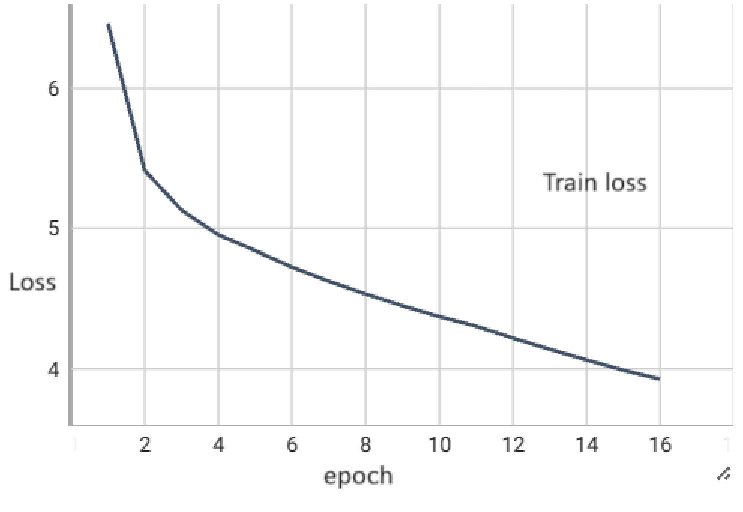

Figure 2.4: Train loss of model 2

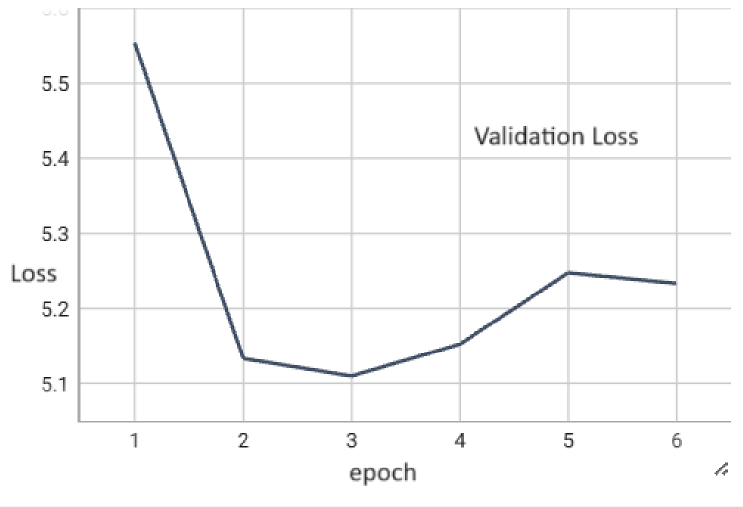

Figure 2.5: Validation loss of model 2

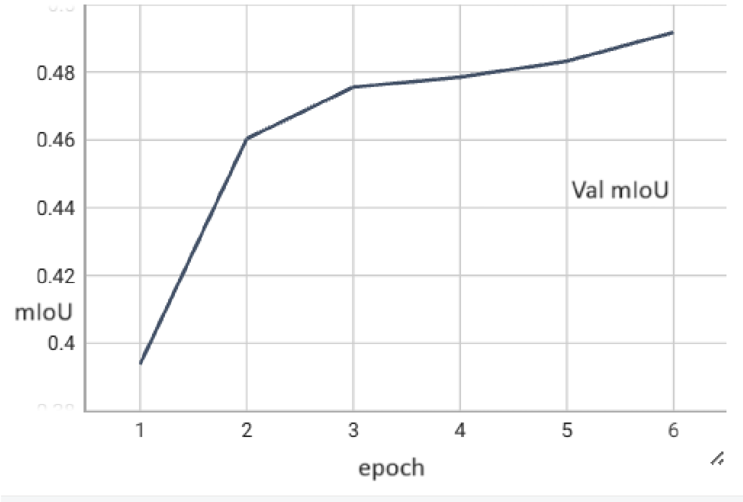

Figure 2.6: Validation mIoU of model 2

Figure 2.7: original image

Figure 2.8: predicted segmentation map

Figure 2.9: original image

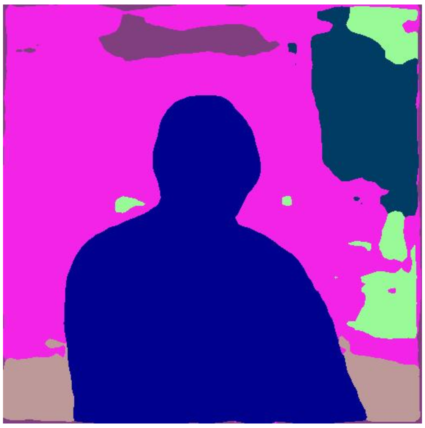

Figure 2.10: predicted segmentation map

Figure 2.11: original image

Figure 2.12: predicted segmentation map

Chapter 3

Conclusions/Future Improvements

Based on the experimental findings, it is evident that achieving at least 50% mIoU (mean Intersection over Union) requires significantly increased training time, particularly when dealing with a large number of classes for segmentation, typically exceeding 50. Model 2, outperforming Model 1, attained an mIoU of 48.32 on the validation set while demonstrating real-time inference speed. Furthermore, it exhibited superior boundary delineation despite fewer epochs.

Future work includes:

- Increasing the training time for the better prediction in case of more than 50 classes.
- Incorporating class-weights derived from the dataset's class distribution could potentially enhance the obtained results.