cs512 Project Proposal

Department of Computer Science
Illinois Institute of Technology
November 10, 2020

1 Project details

Project Title: Deep Convolutional Architecture for Image Segmentation

Student 1

Name: Yash Patel

ID: A20451170

Student 2

Name: Chandni Patel

ID: A20455322

2 Main Research Paper

V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481-2495, 1 Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.

3 Problem Statement

Using semantic pixel-wise segmentation create deep fully convolutional neural network architecture. The main purpose will be to understand spatial relationship between classes like road, sidewalk from road scenes through modelling appearances (road, building) and shapes (car, pedestrians) and also retaining the boundary information from the extracted image representation.

4 Approach

The Network will be created using Python, TensorFlow, Keras. Thefollowing approach will be followed:

1. The core trainable engine will be containing encoder network and corresponding decoder network which will consist of a hierarchy of decoders one corresponding to each encoder and then be followed by pixel-wise classification layer using Softmax.

2. The architecture of the encoder network will have 13 convolutional layers which is identical to VGG16 network.
3. The fully connected layers of VGG16 will be removed as they make SegNet encoder network significantly smaller and easier to train.
4. The architecture of decoder network will map the low-resolution encoder feature maps to original input resolution feature maps for pixel-wise classification using upsampling of lower resolution input feature maps.
5. To perform non-linear upsampling in decoder, we will be using pooling indices from max-pooling step of the corresponding encoder for accurate boundary localization.
6. The upsampling maps will be convolved with trainable filters to produce dense feature maps in decoder because upsampling maps are spared.
7. Proposed architecture will be compared with widely adapted FCN, DeepLab-LargeFOV, DeconvNet architecture to achieve good segmentation performance based on accuracy.
8. It will be designed such that efficiency increases in terms of memory and computational time during inference when compared to competing architectures and also reduce number of trainable parameters without sacrificing performance.

5 Data

Dataset: Cambridge-driving Labeled Video Database (CamVid)

Description: The database provides ground truth labels in order to associate each pixel with one of 32 semantic labels like road, car, pedestrian, sidewalk, trees, sky, building, etc in image frames. The dataset consists of 367 training and 233 testing RGB images at 360 x 480 resolution.

6 Reference

G. Brostow, J. Fauqueur, and R. Cipolla, “Semantic object classes in video: A high-definition ground truth database,” *Pattern Recognit. Lett.*, vol. 30, no. 2, pp. 88–97, 2009.

H. Noh, S. Hong, and B. Han, “Learning deconvolution network for semantic segmentation,” *in Proc. IEEE Int. Conf. Comput. Vis*., 2015, pp. 1520–1528.

K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv:1409.1556*, 2014.

S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” *arXiv preprint arXiv:1502.03167*, 2015.

V. Badrinarayanan, A. Handa, and R. Cipolla, “SegNet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling*,” arXiv preprint arXiv:1505.07293*, 2015.

7 Team Member Responsibilities

1. Understanding model architecture – Both team members
2. Model creation – Yash Patel
3. Model training – Chandni Patel
4. Model evaluation – Chandni Patel
5. Result analysis – Chandni Patel
6. Final Paper – Both team members
7. Presentation – Both team members