Traffic Sign Detection and Estimation using U-Net Convolutional Neural Network

Jeremy Cheung
jeremy.cheung@ryerson.ca
EE8204 Final Report
Department of Mechanical and Industrial Engineering
Ryerson University, Toronto, Canada.

***Abstract*—In this study, the main focus will be to detect and identify the location of traffic signs within a picture, with the goal being to minimize computation time and increase accuracy. The U-Net is a convolutional neural network algorithm for precisely and accurately segmenting images compared to traditional methods of forward propogation Convolutional Neural Networks (CNN) or windowing. This paper advances and builds on existing knowledge to introduce a novel solution with an already existing algorithm for detecting a traffic sign and estimate the boundary within and input image. Estimating the boundary of traffic signs within an image and classification of a shape class prediction, solved using a CNN. The proposed algorithm and method using a U-Net CNN is compared aganist existing CNN methods in terms of computational time, and accuracy of detection.**

I. Introduction

With the advancement of automobile technology, traffic becomes more and more complicated and confusing for human drivers. In Canada alone, there were over 1600 fatal casualties in 2019 alone, and a total of 140,801 injuries from motor related incidents. [1] Accurate traffic sign detection is an integral step to the overall goal of the fully autonomous vehicle. Different traffic signs have different shapes, sizes, patterns and colours that drivers learn to recognize and obey when getting their license. In order for a fully autonomous vehicle to function as expected, the vehicle must recognize and detect that there is a traffic sign, then interpret the instructions on the sign and make the correct decision accordingly. Each step of this problem requires additional study but the focus of this study will be to recognize and detect a traffic sign.

Recognizing traffic signs consists of three main steps, segmentation to detect signs within the image, template and shape matching and lastly using the neural network to classify the specific sign. [2] In other words, detection, tracking and recognition/ classification of traffic signs. Once the detection of the regional localization problem has been determined, the feature extraction can begin. Feature extraction can be further broken down into filtering, feature segmentation and feature extraction to selection. For accurate traffic sign detection, the machine learning must recognize the traffic signal regardless of external challenges such as weather or image quality.

II. Literature Review

Traffic sign detection has been studied and tested across multiple research papers in the past using CNNs, windowing, and other types of algorithms. Traditional methods of detecting traffic signs consist of using colours, shapes or words to determine the traffic sign. Training and testing a neural network traffic sign recognition model requires a large labelled dataset. The German Traffic Sign Benchmark (GTSRB) and German Traffic Sign Detection Benchmark (GTSDB) holds more than 50000 images with over 40 classes and has been used in studies by Changzhen. [3] Changzhen et al. proposed using a Region Proposal Network (RPN) with the CNN to detect traffic signal locations within a video. [4] The dataset consists of classified pictures of traffic signs with uneven lighting, sign tilt as well as partially visible traffic signs. The GTRSB dataset competition winning team utilized a combination of a single CNN with a multilayer perception (MLP)with a 99% accuracy across all subsets. [6]

Traffic sign recognition is a major component in computer vision estimation and plenty of different algorithms and mathematical techniques have already been shown to be effective in this problem. [10] Maldonado et al. used support vector machines (SVM) to detect the traffic signs starting with the colour segmentation using thresholding to determine a specific colour profile and eliminate outlines. Kartikeyan et al. prposed using a Gabor Filter and edge detection in order to determine the location of the traffic sign within the image for further feature extraction. [5] The gabor filter would give the highest reponse at the edges and where the picture changes which would greatly reduce the amount of the data that would

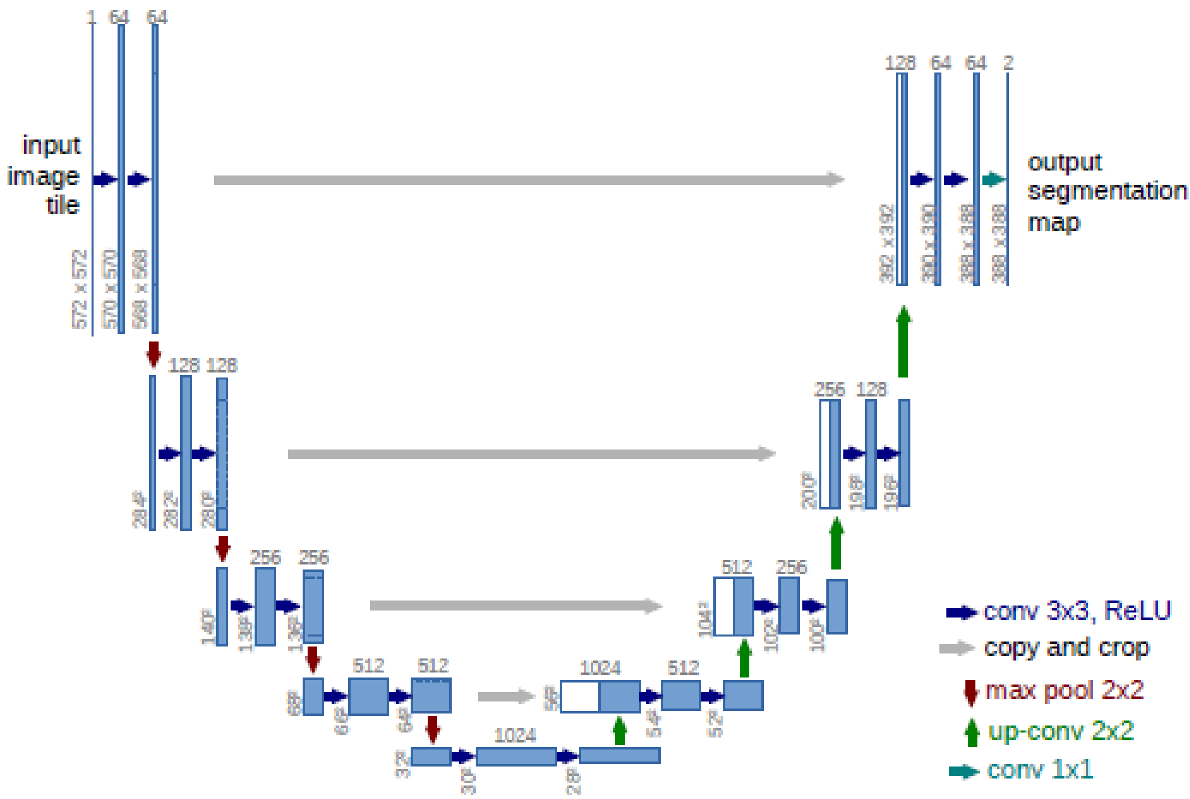

Fig. 1. Network Architecture of U-Net

be needed to be processed. Biswas et al. [8] proposed using a Circular Hough Transform (CHT) for detecting a prohibitory sign, with a 98% performance accuracy out of 501 signs. [11] Using CNNs to recognize traffic signs has also been studied many times using slightly different algorithms. Zang et al. used a local binary pattern (LBP) and the AdaBoost classifier to extract regions of interest then two convolutional layers for the recognition with an accuracy of over 99% on prohibitory traffic signs. [13] In 2016, Lin et al. proposed combining CNNs and Support Vector Machine (SVM) to employ image classifcation with an accuracy of 72% on road arrows. Girshick et al. used a regional CNN approach where the program classifies each region in its specific SVM to provide faster and higher percision. Kakade et al. proposed using a textural base preprocessing to refine the edge conditions or contours of the image then create segmentation maps. [15]

The use of a U-Net CNN has mainly been used for medical imaging thus far. The network architecture of U-Net follows a fully convolutional network architecture with the major modification being that the upsampling uses a large number of feature channels which allow the network to carry forward information to deeper layers. [12] More recently, Yang et al proposed a modified version of the U-Net algorithm the RCNN- Unet utilizing a recurrent neural network followed by a U-Net for geographical road identification tasks. [7]

III. Problem Statement

The goal is to utilize a U-Net Convolutional Neural Network to segment the traffic sign from traffic images faster and more accurately then SVM and other mathematical methods. The goal of this is to use less data to build and develop a model with similar accuracy to the sliding window algorithm. The GTSRB dataset was used for training and testing, with 39209 images in total among 42 classes. Within the german traffic sign benchmark dataset there are a total of 42 different classes, consisting of hazard signs, stop signs, speed limits, directional arrows and more.

IV. Model

The original network architecture utlizied in PAPER OLLENBERG consisted of a contracting encoder portion followed by an expansive decoder portion. The initial encoder portion of the path consists of two 3x3 convolutions, followed by a rectified linear unit (ReLU) activation, then a 2x2 maxpooling with a stride of 2. At each step the number of feature channels are doubled. The decoder upsampling portion then consists of a 2x2

Fig. 2. Examples of the Meta Data Traffic Signs found in German Traffic Sign Dataset

TABLE I
MODEL HYPER PARAMETERS

Hyper parameters	Model
Epochs	30
Optimizer	Adam
Learning Rate	0.00001

convolution that halfs the feature channels followed by two 3x3 convolutions, all with a ReLU function. The concatenation of the feature map is from the decoder path and is necessary to maintaining the border pixels at each layer. The final layer is then used to map the component feature to the desired number of classes.

The models hyper parameters are provided by Table 1. The training was done on a ryzen 5 3.6 GHz machine with a NVIDIA GTX 3070 GPU.

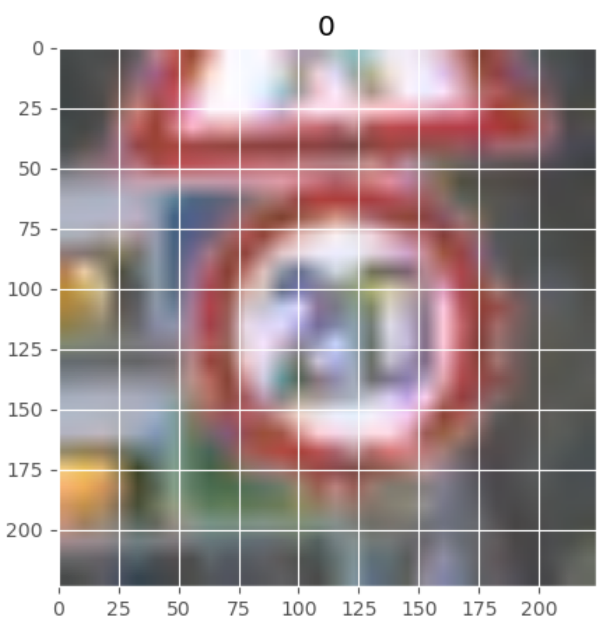

Fig. 3. A sample image from the dataset: the classifer 0 refers to all images of the speed limit of 20 km/h

The loss function used in this model is the binary cross entropy loss function also none as log loss used in segmentation problems. The log loss function

V. RESULTS

The results of the model run on the Kaggle data set are shown to be relatively inconsistent. Notice that the pictures keep the relative structure and also remove all the other parts of the structure. When compared to the U-Nets results completed by Jing et al. there is a significant difference between the model created in this paper and the model done in their paper. [16]

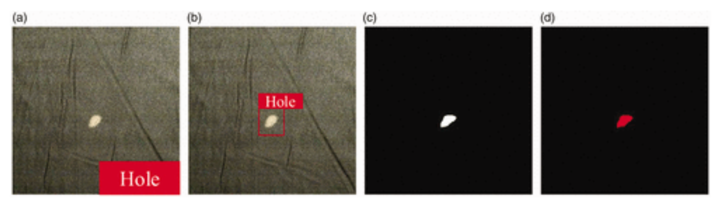

Fig. 4. Defect mask detection completed by Jing et al. [16]

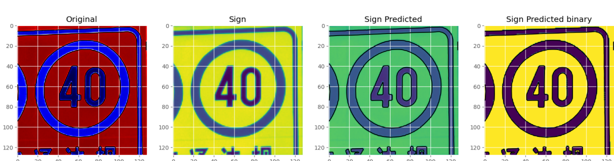

Fig. 5. Traffic Sign Example Result

The current model does not have enough outliers in the data set, which does not allow for a good comparison between datasets and other models. The dataset it is trained on has been preprocessed and classified data, with the image of the traffic clearly visible by the model.

VI. CONCLUSION

This paper explored the possibility of using the U-Net architecture as the first segmentation step in traffic data sign detection and classification. The U -Net is able to remove significant amounts of data from the image without losing features that could be used in edge detection, colour or light. This model is tested and trained on a single dataset however in order to increase the accuracy of the model more images and videos can be used to train the dataset in the future. Future study would include using the U-Net architecture with another neural network or classifier to and comparing it with other classifiers to determine the overall accuracy of the dataset with and without the preprocessing U-Net.

References

[1] Transport Canada, "Canadian Motor Vehicle Traffic Collision Statistics: 2019,' *Online* https://tc.canada.ca/en/road-transportation/statistics-data/canadian-motor-vehicle-traffic-collision-statistics-2019

[2] J. Torresen, J. Bakke, and L. Sekanina, "Efficient Recognition of Speed Limit Signs," in *2004 IEEE Intelligent Transportation Systems Conference* (Washington D.C , 2004).

[3] X. Changzhen, W. Cong, M. Weixin, and S. Yanmei, "A Traffic Sign Detection Algorithm Based on Deep Convolutional Neural Network," in *2016 IEEE International Conference on Signal and Image Processing* (North China University of Technology, 2016).

[4] J.R.R. Uijlings, K.E.A Van de Sande, T. Gevers, and A.W.M. Smeulders, "Selective Search for Object Recognition," in *2013 International Journal of Computer Vision* (University of Amsterdam, Amsterdam, The Netherlands 2013).

[5] D. Karthikeyan, C. Enitha, S. Bharathi, and S. Durkadevi, "Traffic Sign Detection and Recognition using Image Processing," in *International Journal of Engineering Research and Technology. NCICCT - 2020 Conference Proceedings* (Kongunadu College of Engineering and Technology, 2020).

[6] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German Traffic Sign Recognition Benchmark:A multi-class classification competition," in *The 2011 International Joint Conference on Neural Networks* (Ruhr-Universitat Bochum and Unversity of Copenhagen, 2011).

[7] X. Yang, X. Li, Y. Ye, R. Lau, X. Zhang and X. Huang, "Road Detection and Centerline Extraction Via Deepn Recurrent Convolutional Neural Network U-Net," in *IEEE Transactions on Geoscience and Remote Sensing, VOL. 57, NO. 9, Sept 2019* (Harbin Institute of Technology, Shenzhen 518055, China, 2019).

[8] R. Biswas, H. Fleyeh and M. Mostakim, "Detection and Classification of Speed Limit Traffic Signs," (BRAC University and Dalarna University).

[9] H.S. Lee, and K. Kang, "Simultaneous Traffic Sign Detection and Boundary Estimation using Convolutional Neural Network," in *IEE Transactions on Intelligent Transportation Systems*

[10] S. Maldonado-Bascon, S. Lafuente-Arroyo,P. Gil-Jimenez, H. Gomez- Moreno and F. Lopez-Ferraras "Road-Sign Detection and Recognition Based onSupport Vector Machines," in *IEEE Transactions on Intelligent Transportation Systems, V. 8, NO. 2, 2007*

[11] F. Lin, Y. Lai, L. Lin, and Y. Yuan, "A Traffic Sign Recognition Method Based on Deep Visual Feature," in *2016 Progress In Electromagnetic Research Symposium (PIERS), Shanghai, China* (Tongji University and Fudan University, 2020).

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Internation Conference on Medical Image Computing and Computer- Assisted Intervention* (Springer, 2015, pp. 234- 241).

[13] D. Zang, J. Zhang, D. Zhang, M. Bao, J. Cheng and K. Tang, "Traffic sign detection based on cascaded convolutional neural networks," in *2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)* (2016, pp. 201-206).

[14] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition* (UC Berkeley, 2014).

[15] A. Kakade and J. Dumbali, "Identification of Nerve in Ultrasound Images Using U-net Architecture," in *2018 International Conference on Communication, Information & Computing Technology (ICCICT)* (Vellore Institute of Technology 2018).

[16] J. Jing, Z. Wang, M. Ra¨tsch, and H. Zhang, "Mobile-Unet: An efficient convolutional neural network for fabric defect detection" in *Textile Research Journal 0(00) 1–13* (College of Electronics and Information, Xi'an Polytechnic University, China & College of Engineering, Reutlingen University, Germany).