Towards the development of autonomous wheelchair

Mihai Crăciunescu
Department of Automatic Control and Applied Informatics
Faculty of Automatic Control and Computer Science
Bucharest, Romania
mihaiacr@gmail.com

Ștefan Mocanu
Department of Automatic Control and Applied Informatics
Faculty of Automatic Control and Computer Science
Bucharest, Romania
stefan.mocanu@upb.ro

Diana Baicu
Department of Automatic Control and Applied Informatincs
Faculty of Automatic Control and Computer Science
Bucharest, Romania
diana.baicu10@gmail.com

Radu Dobrescu
Department of Automatic Control and Applied Informatics
Faculty of Automatic Control and Computer Science
Bucharest, Romania
rd_dobrescu@yahoo.com

Maria Cîrciumaru
Department of Automatic Control and Applied Informatincs
Faculty of Automatic Control and Computer Science
Bucharest, Romania
c.maria_93@yahoo.com

***Abstract*—Precise and fast indoor navigation represent well-known challenges that, if solved, can improve many aspects of our lives. In this paper we propose a complex system architecture dedicated to precise indoor navigation in unknown environments with the final purpose of developing an autonomous electric wheelchair. The desired outcome is to use computer vision to detect floors and doors with high accuracy and speed and compute a safe path to be followed. Our solution is based on fully convolutional neural networks (FCN) for image segmentation of corridor images to identify the floor area further used for navigation.**

***Keywords*— FCN, autonomous vehicle, indoor navigation component, image segmentation**

I. Introduction

Indoor environments represent a challenge for autonomous robots, given the fact that scene understanding, as well as precise localization within a map, are hard to be achieved. However, the navigation task is rather complex, and can therefore be divided into several sub-tasks, such as obtaining the map of the environment, overall positioning within the map, issuing a movement command based on the available space, while considering a fixed destination point, as well as responding to elements and changes in the scene.

One of the most relevant applications for indoor navigation dedicated to personal use is represented by autonomous wheelchairs. These devices are very important for people who exhibit various forms of physical or motion disabilities and autonomous navigation both in indoor and outdoor environments would represent a major plus.

In this paper, the early stages of achieving autonomous navigation with a wheelchair system through an indoor environment are presented. Thus, two of the fundamental issues are addressed, namely the segmentation of the navigable corridor, regardless of the surroundings, as well as the detection of commonly encountered elements, such as doors. For the first task, a fully convolutional network architecture has been proposed, whereas for the second task a detection network has been trained within NVIDIA's Deep Learning GPU Training System on a custom created dataset.

The rest of paper is organized as follows: section II is dedicated to related work, section III presents the Proposed System, section IV is dedicated to Results and Discussions while in Section V some Conclusions are drawn.

II. Related Work

This section is dedicated to some discussions related to methods used today for wheelchairs indoor navigations as well as the work and progress into deep neural network field for image segmentation and object detection that can make indoor navigation an easier and safer task.

There are many existing approaches when it comes to autonomous wheelchair navigation in unknown environments. Some approaches for indoor wheelchair navigation use LIDAR sensors usually alongside other sensors or cameras to map the environments. A navigation method is presented in paper [1], using mainly a LIDAR, a microcontroller, rotary encoders implemented on Robotic Operating System (ROS) to determine a path to a user-defined destination. The system described in this paper is capable of adapting to the changes that occur in an environment such as an obstacle along the path way. First a map is constructed by manually driving the wheelchair through the entire building, followed by localization through Adaptive Monte Carlo Localization and other ROS packages that run together to obtain the shortest path while avoiding obstacles.

Paper [2] is also based on LIDAR equipment for a laser scan and simultaneous localization and mapping (SLAM) to make a 3D point cloud scene mapping which goes through the process of semantic segmentation using deep learning to classify each point into an object category that is likely to be found in a building.

Similar works are presented in [3] that also have a laser scanner system and other sensors for navigation like depth or usual camera. One of the main concerns comes with the position of the equipment so that it is comfortable enough for the human sitting in the chair and at the same time it can take good measurements plus it is more expensive than an image

2379-0482/19/$31.00 ©2019 IEEE
DOI 10.1109/CSCS.2019.00101

sensor/digital camera. Existing autonomous vehicles are already used today in places like airports or hospitals [4].

Our previous work relies on a computer vision system. In paper [5] we proposed an autonomous wheelchair architecture based on two cameras and sensors which can navigate inside a building with corridors delimited either by walls or transition lanes. Image processing based on various color information is regularly used for complex tasks as presented in [6]. Our system uses the images from the cameras, each placed on the left and right side and applies a series of filters, edge detection and transformation techniques to detect the lane lines alongside the corridor. Knowing the lines coordinates, their angle from the horizontal can be computed and it will show the wheelchair orientation error, considering that the ideal position is the one where the wheelchair is in the middle of the corridor and parallel to the walls. Also considering the lanes position in the images, the lateral distance of the wheelchair from the walls is analyzed to calculate the lateral error, meaning the deviation of the wheelchair from the center of the corridor. The control strategy computes the output steering angle based on the lateral error and angle error from the desired trajectory. There are other works in the domain that use one or multiple cameras and computer vision techniques to detect the corridor [7][8][9]. In the field of indoor navigation there are applications intended for drones/small aerial vehicles that use CNN for training and a single camera to navigate [a] or transfer learning for indoor positioning [b].

The grown interest in Machine Learning ML algorithms led to their use to solve or improve some real-life situations. This application is intended for disabled people in a wheelchair and means to improve their comfort by automating some of the task like finding a route and navigating autonomously through a corridor. In this manner the person is not forced to control the wheelchair all the time through the joystick. Current literature reveals the fact that CNNs are a good choice when it comes to image classification [10]. Thus, the task of corridor recognition, as it will be demonstrated in this paper, can be achieved with high accuracy with image segmentation using Fully Convolutional Neural Networks FCN. In paper [11] the authors show a great improvement in accuracy by using FCNs for semantic segmentation. The paper demonstrated that by building fully convolutional networks can improve the previous results in semantic segmentation (30% relative improvement to 67.2% mean IU on 2012 on PASCAL-VOC). Semantic segmentation aims to label each pixel in the image and a very computational efficient method for this purpose is presented in [11] as fully convolutional networks that take an arbitrary sized input image and create an output by using end-to-end, pixel-to-pixel operation simultaneously. Furthermore, they demonstrated an increased output precision by learning on end-to-end skip architecture which ads skips between the layers to combine the semantic and spatial information.

Drastically improved performance of image semantic segmentation using FCN increased their use for other purposes, especially for self-driving vehicles. An example is presented in paper [12] for using a convolutional patch network with fully connected layers to incorporate the spatial information with which they obtained state of the art results in road detection and urban scene understanding. The time results for this architecture enables them to be incorporated in real-time object detection. A similar architecture (named SqueezeDet) is presented in [13] for object detection with respect to all the constraints required by real time autonomous driving tasks. A fast road detection solution presented in paper [14] analyses the response time and precision for different patch sizes while in [15] the authors are performing roads detection based on images taken by UAVs.

The typical architecture can be been divided into branches, each concentrating on an optimization for a specific purpose. A derived architecture is presented in paper [16] as the problem of detecting the road boundaries with high accuracy is solved by having a Siamese FCN with two streams to process the original RGB images and to contour maps. Paper [17] further develops the concept of convolutional networks into full-resolution residual networks for semantic segmentations in street views. Also based on FCN, the study ads an up-convolutional part for a high-resolution segmentation. Paper [18] analyses the advantages and disadvantages of FCNN4SS giving the real-world environment imperfections.

III. PROPOSED SYSTEM

In this chapter, a proposal for the early stages of achieving autonomous navigation with a wheelchair system through an indoor environment is presented. Thus, two challenges are addressed, namely the segmentation of the navigable corridor, regardless of the surroundings, and the detection of commonly encountered elements, such as doors. For this purpose, a FCN architecture was chosen, implemented and deployed on the system.

A. Corridor segmentation

Fully convolutional networks (FCNs) have previously been successfully implemented for various semantic segmentation tasks, including tracking objects and precisely delimiting borders in high-resolution images [19]. In the same time, for a human, the first step in navigating through an environment is represented by clearly recognizing the accessible regions. Therefore, an autonomous wheelchair system should also be able to compute a command considering the total navigable points within a captured image. In general terms, the first part of the network consists of an encoder block, responsible with extracting relevant features and reducing the dimensions of the feature map, while increasing its depth. The encoder block it is then followed by a decoder, which groups the resulted features, while segmenting the desired classes.

The implemented network has a total of 27 layers, accepting as an input RGB images with 320x640 resolution, as one can see in **Error! Reference source not found.**.

The output of every convolution or deconvolution layer is normalized by applying a batch normalization, which used the mean and the variance of the values in the current mini-batch. This procedure enables a faster training of the network, since convergence can be achieved faster, but also the learning rates can be higher. At the same time, the dropout procedure has been employed in certain parts of the network, after the outputs of the convolutional layers have been normalized. Even though the two procedures are often considered to be antagonistic, as proposed in [20], the carried experiments were proven to give better results when using the two methods combined.

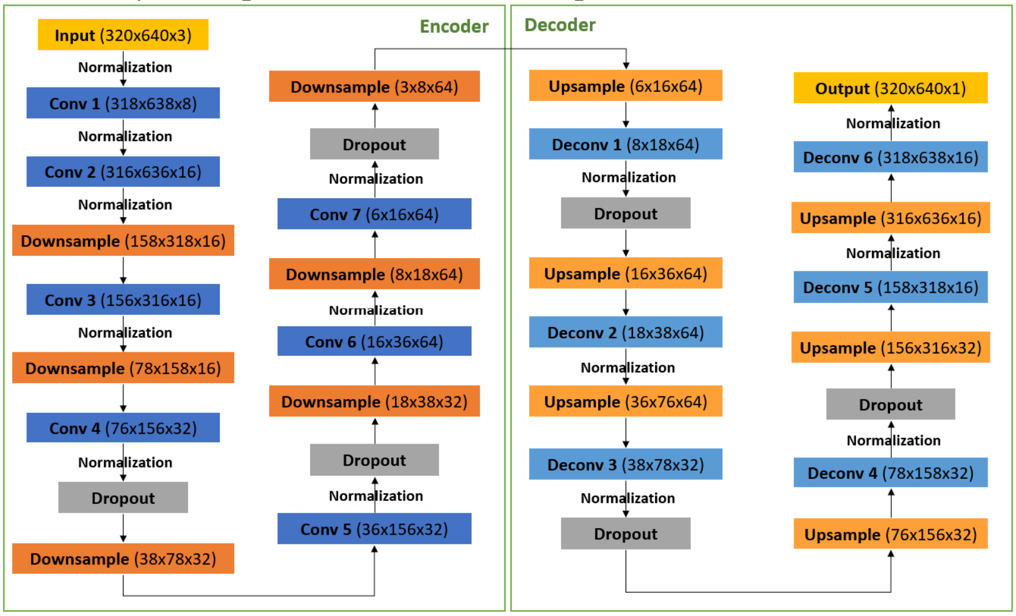

Fig. 1. Convolutional Network Architecture

Below, in **Error! Reference source not found.** the deployment result of an early version of the network can be seen, where the batch normalization technique has not been yet added in the architecture, whereas the dropout has been set with rates between 20 and 30%. **Error! Reference source not found.** depicts the results of normalization combined with the same dropout rate, which was used after every convolution and deconvolution layer.

Fig. 2. Early version result of the FCN

Fig. 3. Result of the FCN with normalized layers

Error! Reference source not found. shows a trade-off between the two, with the structure presented above, as well as a dropout rate slightly reduced (10-20%). **Error! Reference source not found.** also displays the computed angle for moving forward, in straight direction based on the corridor shape. It was considered that the desirable position when moving through such a corridor is at its centre, in the case that it is not occluded with any obstacle. The overall direction vector is determined by transforming the resulting corridor mask into polar coordinates and computing the average angle. Since the wheelchair is equipped with two independently controlled motors, the resulting angle would then be transformed to a rotation command for each of the motors.

Fig. 4. Final FCN result with direction arrow

For the training procedure, a part of the dataset from [21] has been used and enhanced to approximately 1000 training pictures of size 160x80 pixels. For each additional training image, a ground truth mask of the navigable floor has been manually created. **Error! Reference source not found.**a and 5b display a few samples taken from the dataset, the first representing the original image, while the second is the binary mask. For a reasonable performance to be obtained, between 30 and 40 epochs were necessary. The training was carried on a NVIDIA GeForce GTX 1070 and took 25 to 30 minutes in average to complete.

Fig. 5a. Ground truth corridor image with associated mask

Fig. 5b. Ground truth corridor image with associated mask

B. Doors detection

A different approach was used for the second problem, given that the main concern was to create a network easy to deploy on the wheelchair, which is equipped with a Jetson TX2 platform. Therefore, the natural alternative was that of implementing a slightly modified version of NVIDIA's DetectNet, which is based on the reference network model GoogleNet [22]. The detection network uses as a starting point the weights of GoogleNet, pre-trained on ImageNet large dataset, and then fine-tuned on custom created dataset. The main focus was to develop a system capable to identify doors in a scene, information which is will be combined in further iterations with other sensor data (such as LIDAR) to detect if a door is open or closed.

For the creation of the dataset, an online annotation tool has been used [23], which exports the created labels in popular datasets formats. The labels have been pre-processed so that they correspond to the format expected by the framework, which includes the category of the detected object and the coordinates of the left-most point of the detections box, as well as the coordinates of the right-most point. An example of a labelled image can be seen in **Error! Reference source not found.**. The dataset contained a total of 150 images, with various door models from various indoor locations, such as lecture halls, corridors, office buildings or dormitories.

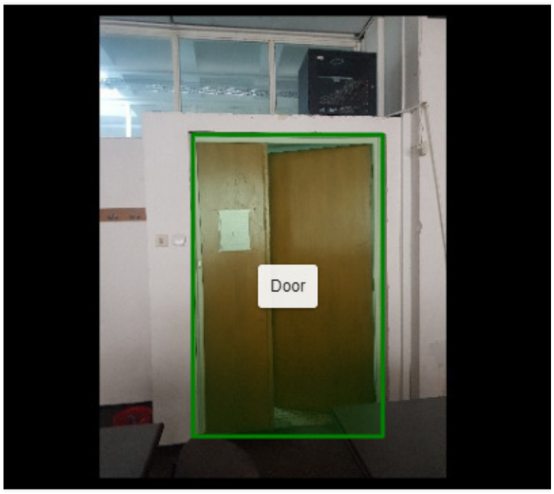

Fig. 6. Door label

The data set was divided into training and validation sections, with the validation containing approximately 15% of the total samples. The training has been carried in two stages, the first consisting of 400 epochs, obtaining a mean average precision under 5%. A test image can be observed in **Error! Reference source not found.**, where the label is visibly shifted from the desired location. For this reason, the model has been further trained on an additional 600 epochs, throughout this process the mean average precision increasing to 72%. As an illustration, **Error! Reference source not found.** presents a good detection of the same picture previously assessed. Training has been carried in a Cloud instance, having as base hardware an NVIDIA Tesla K8 with 12GB of graphical memory. Total training time took approximately 6 hours to complete.

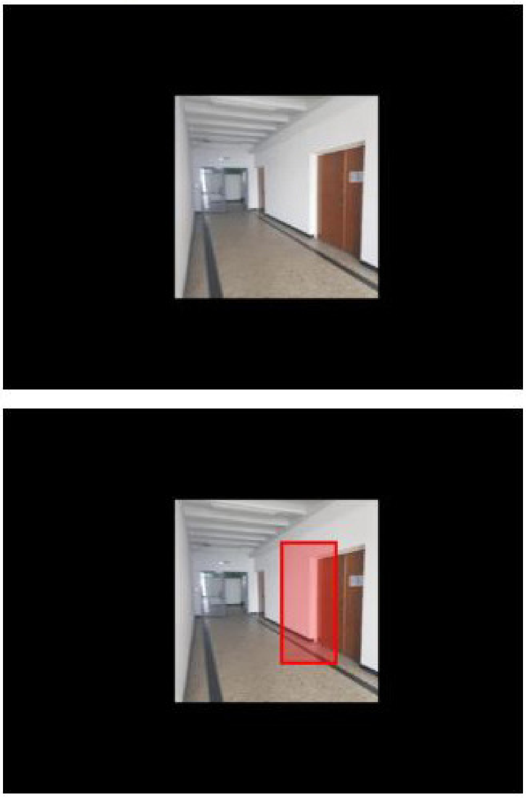

Fig. 7. Early stage of door detection result

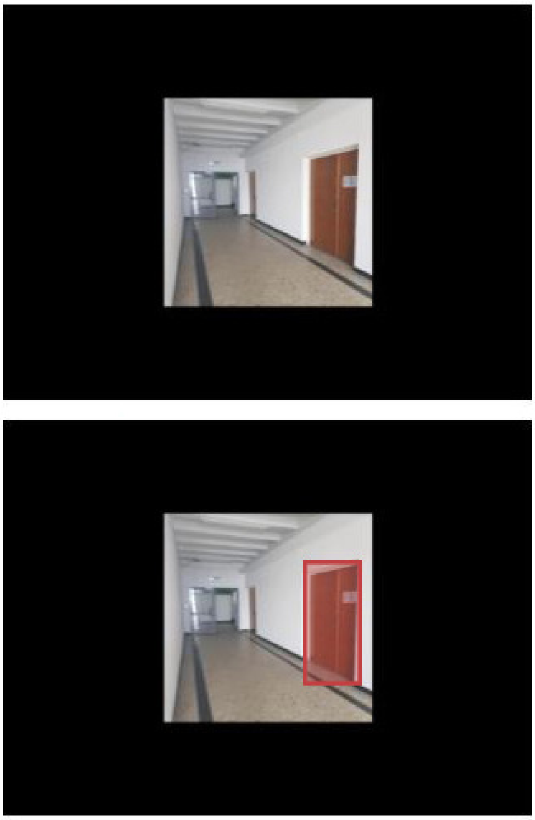

Fig. 8. Final door detection result

IV. RESULTS AND DISCUSSIONS

When discussing the performance on the segmentation results, as mentioned in the previous section, the Intersection over Union (*IoU*) metric has been used.

$$IoU = \frac{P \cap G}{P \cup G}$$

where P represents the area predicted as corridor in an image and G is the associated mask. Ideally, the IoU score should be 0.5, as it can be derived from the above definition.

The metrics have been computed on the entire data set, and the resulted average intersection over union score has been of approximately 0.47.

Therefore, the segmentation task has been carried on a dataset of approximately 1000 images, the model has been trained for 40 epochs to obtain the afore mentioned performance. Should the performance need further improvement, an alternative would be that of further labelling and adding images in order to obtain a more complex dataset. Increasing the number of training epochs could also be a solution, but the data must be monitored nearly so that overfitting does not occur.

Regarding the performance of the door detection network, the chosen metric has been the mean average precision (*mAP*), which can be defined as it follows:

$$Prec = \frac{TP}{TP + FP} \qquad Rec = \frac{TP}{TP + FN}$$

$TP - true\ positive\ results$
$FP - false\ positive\ results$
$FN - false\ negative\ results$

$$mAP = \frac{1}{n} \sum_{Rec_i} Prec(Rec_i)$$

The *Prec* (precision) factor represents how good the prediction is, whereas the *Rec* (recall) factor shows how well the found positives are fitting.

Since after a training of 600 epochs on 150 images, a mean average precision of 72% has been obtained, the detection task can be further improved by labelling and adding supplimentary data. Whether upon integrating the network in the complete system, this step can be further carried.

Moreover, besides the two presented networks, to the wheelchair system, other sensors and subsystems need to be included in order to navigate autonomously indoors. Thus, the first action which naturally follows is the integration of spatial information from a two-dimensional LIDAR. This will enable it to conclude whether a door that has been identified, and through which the system must pass, is open or closed. Furthermore, the segmentation of the corridor will be combined with the mapping and localization techniques.

V. CONCLUSIONS

In this paper, two early steps of the autonomous navigation process have been presented. On the one hand, a fully convolutional network was implemented and deployed for segmenting the floor of a frame containing an indoor scene. The role of the network is that of outputting a movement command based on the free space on the corridor, as well as the overall direction given by the navigation module. On the other hand, a detection network for outputting the bounding boxes of the encountered doors has been constructed using a fast-development framework. Its purpose is only that of signaling the presence of a door in the vicinity of the wheelchair. The implementation on the mobile platform should be enhanced with spatial information to allow reliable positioning when entering the required room.

Using the presented networks as the image processing layer of the wheelchair, and adding the other compulsory modules, such as actuation, control, high-level decision making, an autonomous system can be further developed.

ACKNOWLEDGMENT

This work was partially supported by the Romanian Ministry of Education and Research under grant 78PCCDI/2018-CIDSACTEH

REFERENCES

[1] Grewal, H., Matthews, A., Tea, R., George, K., "LIDAR-based autonomous wheelchair", 2017 IEEE Sensors Applications Symposium (SAS), Glassboro, NJ, USA , DOI: 10.1109/SAS.2017.7894082

[2] Chen, J., Kim, P., Cho Y.K., Ueda, J., "Object-sensitive potential fields for mobile robot navigation and mapping in indoor environments", 2018 15th International Conference on Ubiquitous Robots (UR), Honolulu, HI, USA, DOI: 10.1109/URAI.2018.8441896

[3] Grewal, H.S., Jayaprakash, N.T., Matthews, A., Shrivastav, C., George, K.," Autonomous wheelchair navigation in unmapped indoor environments", 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Houston, Texas, USA, DOI: 10.1109/I2MTC.2018.8409854

[4] Armitage, S., "The opportunities and challenges of autonomous vehicles at an airport", Airport Access Ideas Forum, https://www.globalairrail.com/images/events/2018/AAIF/Case_Studies/AAIF18%20Case%20Studies-Heathrow.pdf

[5] Nicolae, M., Mocanu, Ş., Crăciunescu, M., Dobrescu, R., "Trajectory Control of an Autonomous System dedicated to Assisted Living", 2018 22nd International Conference on System Theory, Control and Computing (ICSTCC), Sinaia, Romania, DOI: 10.1109/ICSTCC.2018.8540770

[6] Popescu, D., Ichim, I., Gornea, D., Stoican, F., "Complex Image Processing Using Correlated Color Information", Proceedings 17th International Conference, ACIVS 2016, Lecce, Italy, October 24–27, 2016, Lecture Notes in Computer Science 10016, pp. 723-734.

[7] Pasteau, F., Krupa, A., Babel, M., "Vision-based assistance for wheelchair navigation along corridors", 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, DOI: 10.1109/ICRA.2014.6907505

[8] Khalilullah, K.M.I., Ota, S., Yasuda, T., Jindai, M., "Development of robot navigation method based on single camera vision using deep learning", 2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), Kanazawa, Japan, DOI: 10.23919/SICE.2017.8105675

[9] Pasteau, F., Narayanan, V.K., Babel, M., Chaumette, F., "A visual servoing approach for autonomous corridor following and doorway passing in a wheelchair", in Robotics and Autonomous Systems, vol.75, part A, January 2016, pp. 28-40, https://doi.org/10.1016/j.robot.2014.10.017

[10] Jaswal, D., Sowmya, V., Soman. K.P., "Image Classification Using Convolutional Neural Networks", in International Journal of Advancements in Research & Technology, Volume 3, Issue 6, June-2014, p. 1661, ISSN 2278-7763, DOI: 10.14299/ijser.2014.06.002

[11] Shelhamer, E., Long, J., Darrell, T., "Fully Convolutional Networksfor Semantic Segmentation", in IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 39 , Issue 4 , April 1, 2017, DOI: 10.1109/TPAMI.2016.2572683

[12] Brust, C., Sickert, S., Simon, M., Rodner, E., Denzler, J. "Convolutional Patch Networks with Spatial Prior for Road Detection and Urban Scene Understanding", in Proceedings of the 10th International Conference on Computer Vision Theory and Applications - Volume 2: VISAPP, (VISIGRAPP 2015), ISBN 978-989-758-090-1, pp. 510-517, DOI: 10.5220/0005355105100517

[13] B. Wu et al., "SqueezeDet: Unified Small Low Power Fully Convolutional Neural Networks for Real-Time Object Detection for Autonomous Driving", Proc. IEEE CVPRW, vol. 7, pp. 446-454, 2017, DOI: 10.1109/CVPRW.2017.60

[14] Mendes, C.C.T., Frémont, V., , D.F., "Exploiting fully convolutional neural networks for fast road detection", 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, DOI: 10.1109/ICRA.2016.7487486

[15] Ichim, L., Popescu, D., "Road Detection and Segmentation from Aerial Images Using a CNN Based System", 2018 41st International Conference on Telecommunications and Signal Processing (TSP), Athens, Greece, 4-6 July 2018, pp 294-297

[16] Wang, Q., Gao, J., Yuan. Y., "Embedding Structured Contour and Location Prior in Siamesed Fully Convolutional Networks for Road Detection", IEEE Transactions on Intelligent Transportation Systems, Vol. 19 , Issue: 1 , January, 2018, DOI: 10.1109/TITS.2017.2749964

[17] Pohlen, T., Hermans, A., Mathias, M., Leibe, B., "Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes",2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, DOI: 10.1109/CVPR.2017.353

[18] Romera, E., Bergasa, L.M., Arroyo, R., "Can we unify monocular detectors for autonomous driving by using the pixel-wise semantic segmentation of CNNS?", 2016, arXiv preprint arXiv:1607.00971.

[19] Sherrah, J., "Fully Convolutional Networks for Dense Semantic Labelling of High-Resolution Aerial Imagery", 2016, CoRR, arxiv.org/abs/1606.02585

[20] Li, X., Chen, S., Hu, X., Yang, J.," Understanding the Disharmony between Dropout and Batch Normalization by Variance Shift", (2018), CoRR, arxiv.org/abs/1801.05134

[21] Yang, S., Maturana, D., Scherer, S., "Real-time 3D scene layout from a single image using Convolutional Neural Networks", 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, DOI: 10.1109/ICRA.2016.7487368, dataset is available here: http://theairlab.org/cmu-corridor-dataset/

[22] C. Szegedy, et al., "Going deeper with convolutions," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015 pp. 1-9, DOI: 10.1109/CVPR.2015.7298594

[23] Labelbox, app.labelbox.com, Internet resource, last accessed October 2018