Vision-based MAV Navigation in Underground Mine Using Convolutional Neural Network

Sina Sharif Mansouri, Petros Karvelis, Christoforos Kanellakis, Dariusz Kominiak, and George Nikolakopoulos
Robotics Team, Department of Computer, Electrical and Space Engineering, Luleå University of Technology,
Luleå SE-97187, Sweden,
Emails:{sinsha, petkar, chrkan, darkom, geonik}@ltu.se

***Abstract*—This article presents a Convolutional Neural Network (CNN) method to enable autonomous navigation of low-cost Micro Aerial Vehicle (MAV) platforms along dark underground mine environments. The proposed CNN component provides on-line heading rate commands for the MAV by utilising the image stream from the on-board camera, thus allowing the platform to follow a collision-free path along the tunnel axis. A novel part of the developed method consists of the generation of the data-set used for training the CNN. More specifically, inspired from single image haze removal algorithms, various image data-sets collected from real tunnel environments have been processed offline to provide an estimation of the depth information of the scene, where ground truth is not available. The calculated depth map is used to extract the open space in the tunnel, expressed through the area centroid and is finally provided in the training of the CNN. The method considers the MAV as a floating object, thus accurate pose estimation is not required. Finally, the capability of the proposed method has been successfully experimentally evaluated in field trials in an underground mine in Sweden.**

***Index Terms*—Mining Aerial Robotics, Deep Learning for Navigation, MAV**

I. Introduction

In the past decade, significant progress has been made in the development and deployment of Micro Aerial Vehicles (MAVs). These platforms are employed in a wide range of applications for monitoring, exploring an unknown area and minimizing service times, such as underground mine inspection [1], infrastructure inspection [2], and search and rescue operations [3]. Moreover, deployments of MAVs can provide high impact on the mine industry, as these platforms are able to realize an autonomous and unsupervised inspection of unreachable, dark, complex, and dangerous locations. Since, autonomous MAVs could directly enter in the areas and perform their instructed tasks, without the need of pilots and reduce the human exposure to danger environments (e.g. blind openings, areas after blasting, etc.), while providing valuable information, such as images, gas levels, dust levels, 3D models, etc of the environment under exploration.

However, underground mines are challenging environments for deploying MAVs, due to the general lack of illumination, narrow passages, wind gusts and dust, while Figure 1 depicts the flying MAV in the visited under study underground mine, where is also visible the lack of natural illumination and the existence of uneven surfaces. In such cases, the functionalities for the obstacle detection and collision avoidance are essential components for successful and safe autonomous MAV navigation, especially towards the deployment in real-world environments. Generally, in order to provide a stable and reliable autonomous navigation, the MAVs should be equipped with high-end and expensive components and sensor suits. However, the long-term operation of these platforms, in harsh environments, such as an underground mine, degrades their performance and integrity over time.

In this article, a regression Convolutional Neural Network (CNN) method is proposed to enable the autonomous navigation with a low-cost platform in unknown dark underground mines. Initially, the collected data-sets are prepossessed in order to obtain the depth information by utilizing the work reported in [4]. The depth image is then segmented into regions and the region with the highest depth is extracted, while in the sequel, the centroid of this region is calculated. This is a novel way to generate multiple training data-sets for the CNN, when an absolute reference is not available and the access to the field is limited. In the next step, the images are resized from 512×512 to 128×128 pixels, converted to gray scale and the CNN is trained based on images and centroid information. The trained regression CNN provides heading rate commands by extracting the centroid position in the horizontal axis of the image plane from a looking forward camera (Figure 1). It should be highlighted that for obtaining an accurate heading angle estimation, especially for a low-cost navigation system, is a challenging task. Thus, in the proposed method, the obtained centroid position, is converted to a heading rate for the platform. Moreover, the MAV is considered as a floating object with a constant altitude, while an obstacle avoidance algorithm, based on potential fields, is implemented as a higher level collision avoidance scheme. Finally, the proposed method is evaluated experimentally in a real underground mine environment in Sweden.

A. Background & Motivation

Autonomous navigation in unknown environments is highly coupled with collision avoidance and obstacle detection. Moreover, obstacle detection and navigation based on vision based techniques for MAVs received significant attention and in different application scenarios [5]. Visual stereo or monocular camera systems are able to provide depth measurements for

This work has been partially funded by the European Unions Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 730302 SIMS.

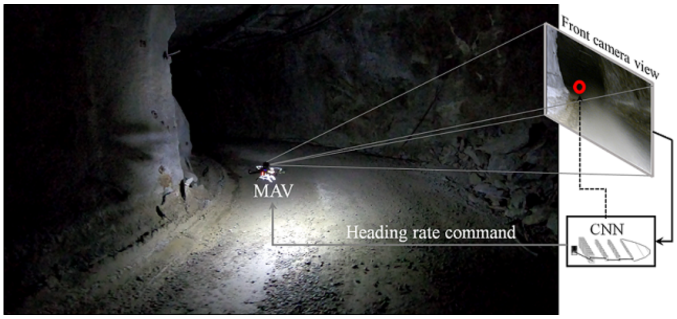

Fig. 1: The proposed regression CNN approach for extracting the centroid with the highest depth based on a forward looking camera to navigate autonomously in underground mines. **Supplementary Video:** https://youtu.be/WKHEvcovXqk

obstacle avoidance, while the obstacle detection methods, based on a monocular camera, in the corresponding literature are based mainly either on computer vision algorithms or on machine learning methods.

Towards computer vision based obstacle detection methods, in [6], a mathematical model to estimate the obstacle distance to the MAV was implemented for collision avoidance. However, the method provided poor results at high velocity and low illumination environments. In [7] it was described the combination of multiple vision based components towards autonomous aerial vehicles. The proposed system used multiple sensing modalities for localization, mapping and obstacle free navigation. The obstacle avoidance scheme was consisted of 3 stereo cameras for $360°$ coverage of the MAV's surroundings in the form of a pointclouds. However, the proposed method relied on sufficient illumination, landmark extraction and high processing on-board power. In general, the performance of the computer vision-based algorithms mainly relies on the surrounding environment with good distinctive features and good illumination and lighting conditions [6]. Furthermore, these methods require a high computation power to process the images and extract landmarks, factors that could limit the usage of these methods in real-life underground mine applications.

Additionally, machine learning methods, such as CNN are gaining more attention in various machine vision tasks, due to the state of the art performance. However, these methods require a large amount of data and a high computation power for training. Eventually and after the training, the CNN can be used for enabling an autonomous navigation with much lower computation power, especially when compared to the training phase. The works using CNN for navigation, such as [8], [9], utilized the image frame of on-board camera to feed the CNN for providing heading commands to the platform. These methods formulated the problem as a classification task with a fixed number of classes e.g. move straight, turn left or turn right and required data-sets with assigned labels in the training phase. In [10] the autonomous MAV navigation in an outdoor urban environments was considered where the yaw-rate command was a regression problem and the collision prediction was addressed as a binary classification problem. Moreover, in [11] a real-time obstacle avoidance, by a regression CNN that predicted the distance to the collision, based on a monocular looking forward camera was proposed and the network was trained based on real-distance data-sets. These works have been evaluated and tuned in out-door and indoor environments and with a good illumination for the camera, thus providing rich data about the surrounding of the platforms. Moreover, in these cases a ground truth reference was available for labeling the data-sets. Furthermore, preliminary and limited studies of MAV navigation in an underground mine using CNN was presented in [12], however the method was evaluated in offline collected data-sets from two underground tunnels, without the MAV in the loop.

B. Contributions

Based on the aforementioned state of the art, the main contributions of this article are provided in this section. The first and major contribution of this work is the development of the regression CNN for providing heading rate commands for the MAV to navigate towards open space areas. Moreover, for training the network, the region with highest depth of the data-sets is extracted by depth map estimation of the scene in an offline procedure. For the best of our knowledge, this is the first work that considers generating data in an offline approach by depth map estimation by utilizing a single image, when the absolute reference is not available and access to the field for collecting data-sets is limited, which is one of the main restrictions, specially in real-life applications. Moreover the trained regression CNN requires less computation power when compared to depth map estimation methods and provides online performance for enabling autonomous navigation of the MAV.

The second contribution, stems from the development and evaluation of the low-cost MAV for autonomous navigation in unknown dark underground tunnels, while accurate pose estimation is not available and the platform operates as a floating object. The proposed method is evaluated in an underground mine and at a $790\,\mathrm{m}$ depth without a natural illumination, while the corresponding experimental results demonstrate the performance of the proposed method towards the establishment of an autonomous MAVs in deep underground mines. The following link https://youtu.be/WKHEvcovXqk provides a video summary of the system.

As a final contribution this work will release the collected data-set with open source access, providing data-sets from underground mine tunnel areas that are not easily accessible for the robotics community, thus enabling further developments in the field.

C. Outline

The rest of the article is structured as follows. Initially, Section II presents the problem formulation of the proposed method and describes the developed components for the MAV. Then, the algorithm for the centroid extraction and the corresponding CNN implementation are presented in Section III. The experimental setup and the extended experimental evaluation of the proposed method in an underground mine are presented in Section IV, while the article concludes by summarizing the findings while presenting some directions for future research in Section V.

II. System Architecture

The MAV is considered as a floating object , which does not depend on the position estimation on x and y-axes, while only the estimation of velocities, attitudes, and altitude are required. The state of the system is $X = [z, v_x, v_y, v_z, \phi, \theta]^\top$.

Furthermore, the Nonlinear Model Predictive Control (NMPC) [13] is implemented to track the desired altitude $z_{d,x}$ and desired velocities $[v_{d,x},\ v_{d,y}]^\top$ and generate the corresponding thrust and attitude commands $[T_d, \phi_d, \theta_d]^\top$ for the low level controller, which is responsible for generating the motor commands $[n_1, \dots, n_4]^\top$ for the MAV.

Moreover, the classical potential field method [1] is implemented to generate a linear x-axis and y-axis velocity commands $[v_{d,x},\ v_{d,y}]^\top$, in order to avoid collisions to the walls or any other obstacles standing in the way of the MAV, while the heading rate commands $\dot{\psi}_d$ are provided from the regression CNN to move towards open spaces. The potential field algorithm uses range measurements $R = \{r_i | r_i \in \mathbf{R}^+, i \in \mathbf{Z} \cap [-\pi, \pi]\}$ from a 2D lidar, placed on top of the MAV, to obtain repulsive velocity commands for both the x and y-axis when flying close to obstacles, while the attractive velocity command is given a constant value on the x-axis.

The overall control structure is presented in Figure 2. The ϕ and θ are provided from the Inertial Measurement Unit (IMU) measurements $[a_x, a_y, a_z, w_x, w_y, w_z]^\top$ through an Extended Kalman Filter (EKF), where a_x, a_y, a_z, w_x, w_y, and w_z are linear and angular accelerations along each axis. The v_x and v_y are calculated from a down-ward optical-flow sensor and the single beam lidar provides altitude z estimation. Moreover, the image stream from the looking forward camera and the estimated states are indicated as I and $\hat{}$ respectively.

III. Methodology

A. Centroid Extraction

In this work the CNN has been trained using information of the open space along the tunnel in sequential frames. This information is expressed through the extraction of the centroid of the identified free tunnel space. The overall concept is based on the depth map estimation of the scene using a single acquired image, while an image can be expressed using the atmospheric scattering model [14] as follows:

$$I(u, v) = J(u, v) \cdot tr(u, v) + A[1 - tr(u, v)] \tag{1}$$

where $I(u, v)$ is the observed image, $J(u, v)$ is the original image from the captured scene, tr is the transmission map, (u, v) are the pixel coordinates, where $u = 0, ..., m - 1$ and $v = 0, ..., n - 1$ with m the width and n the height of the image, and A is the color atmospheric light. The transmission map tr can be defined as:

$$tr(u, v) = e^{-\beta D(x,y)} \tag{2}$$

with $d(x, y)$ to be the scene depth image for (x, y) pixel coordinates and β is the scattering coefficient. A widely used method to extract the transmission map is the Dark Channel Prior (DCP), proposed by [15], in order to estimate the depth map of an image:

$$\begin{aligned} tr(u, v) &= 1 - \omega \Big[\frac{I^{dark}(u, v)}{A} \Big], \\ I^{dark}(u, v) &= \min_{C \in R,G,B} \big[\min_{i_p \in \Omega(u,v)} I^c(i_p) \big] \end{aligned} \tag{3}$$

where ω is a number controlling the desired level of restoration with 1 the highest possible value, $I^{min}(u, v)$ is the dark channel, $\Omega(u, v)$ is a patch of 15×15 pixels centered on (u, v), I^C is the color channel of the image I and i_p represents the index of the pixel of $\Omega(u, v)$. [4] proposed a MultiLayer Perceptron (MLP) module to refine the transmission map outcome targeting the application of image dehazing, as it follows:

$$tr'(u, v) = MLP[tr(u, v)] \tag{4}$$

where MLP is a Multilayer Perceptron that has been trained on a number of images in order to refine the transmission map $tr(u, v)$ to $tr'(u, v)$. In this work, the training data-set for the CNN has been generated by applying Equation 4 for the collected images. The processed images reflect the refined transmission maps and are perceived as a means for estimating the scene depth from single images, without requiring ground truth depth measurements. The method does not rely on accurate depth metrics, since the extracted transmission maps can be used for calculating the free area along the tunnel.

Moreover, every estimated depth map image is segmented into several clusters using the *K-means* algorithm [16]. The cluster with the maximum average intensity (which represents the cluster of pixels with high depth values) is isolated and the centroid [17] of this cluster is calculated $\hat{u}$, while $\hat{v}$ expresses the centroid of the tunnel.

B. CNN Architecture and Training

The Deep Learning (DL) methods require large amount of data in order to train the model, however proper data-sets are not available in all the real-life application scenarios. Thus, in this article, the centroid information from the collected underground mine data-sets is extracted offline for training the CNN. In this way, data-sets without depth information can be used and it is not needed for collecting data-sets with depth information, specially when there is a limited access to the field. The overview of the proposed method is depicted in Figure 3.

Figure 4 depicts the architecture of the proposed CNN [18], while it receives a fixed-size image as an input and provides the centroid position of the tunnel open space. A CNN is composed of an input layer, a number of hidden layer and an output layer. The main advantage of these Neural Networks is

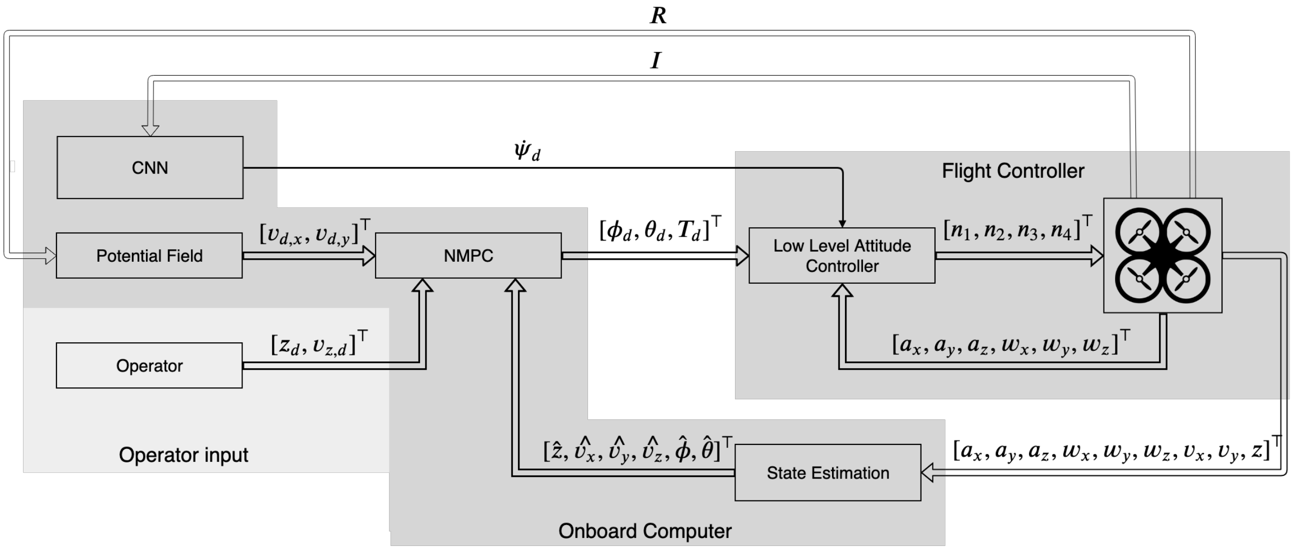

Fig. 2: Control scheme of the proposed method. The NMPC generates thrust and attitude commands, the low level controller generates motor commands, and the CNN provides heading rate commands, while the state estimation is based on IMU, optical flow, and one beam lidar.

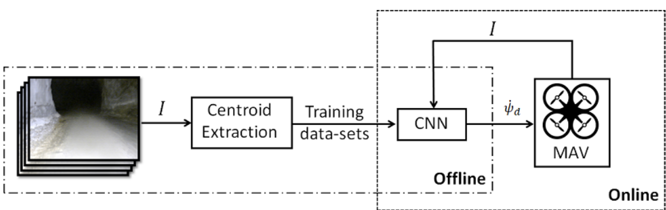

Fig. 3: Overview of the proposed method, while the training data-sets are generated offline by centroid extraction method and the CNN provides online heading rate commands, based on images from the looking forward camera.

the fact that do not rely on feature extraction from the images but the features are extracted automatically and learned during the training process. This is achieved via the convolution operation that is employed with different type of filters in the initial image and allows a number of features to be extracted from the initial image. Furthermore, the convolution operation reduces the number of parameters due to weight sharing. Finally the extra layer of pooling simplifies the output by nonlinear down sampling using e.g. the Rectified Linear Units (ReLUs) function.

In order to reduce the number of the initial inputs and the computation time for the training process, grayscale acquired images have been utilized. The reason for doing this was bi fold: a) first the object recognition process based on gray-scale images can outperform recognition based on color images [19], and b) the RGB sensors do not provide any extra information about the mine environments.

The input layer of the CNN is a matrix of 128×128, followed by a sequence of two 2D convolutional and pooling layers as feature extractors, a fully connected layer to interpret the features, and with a dropout layer to reduce the over fitting [20] and finally an output layer with a *sigmoid* activation [21] to provide outputs between 0 and 1. Depending on the MAV equipped camera, the input image of the CNN can have different sizes, however to reduce the computational power, the image stream of the camera is resized to 128×128 pixels, while for offline centroid extraction, the data-set is resized to 512×512 pixels for providing better information, then the centroid position $\hat{u}$ is mapped to [0,1] ($[0, 511] \rightarrow [0, 1]$) for training the CNN.

Thus, the output o_{cnn} of the CNN is a continuous value between $[0 - 1]$ for representing the centroid position $\hat{u}$, e.g. 0, 0.5, and 1 are the location of the centroid in the left corner ($\hat{u} = 0$), center ($\hat{u} = 63$) and right corner ($\hat{u} = 127$) of the image with 128×128 pixels respectively. Then the output of the CNN is mapped to the heading rate command ($[0, 1] \rightarrow [-0.2, 0.2]$ rad/sec), where the heading rate of $-0.2\,\text{rad/s}$, $-0.0\,\text{rad/s}$, and $0.2\,\text{rad/s}$ corresponds to the centroid in the left, center, and right corner in the image plane. The algorithm 1 provides the overview of the proposed method for generating heading rate commands. The CNN has been implemented in Python by Keras [22] as a high-level neural network Application Programming Interface (API). The loss function is Mean Absolute Percentage Error (MAPE), the optimization is based on an Adam optimizer [23], the learning rate is 0.001, and the learning rate decay is 5×10^{-6} over each update. Finally a workstation has been utilized, equipped with an Nvidia GTX 1070 GPU for the training of the network with 200 epochs and 150 steps per epoch, while the trained network is evaluated online on the on-board MAV main processing unit.

To train the CNN, the data-sets collected from moving the camera by an operator in different directions and from flights in the underground mine are used. For training the CNN 5067

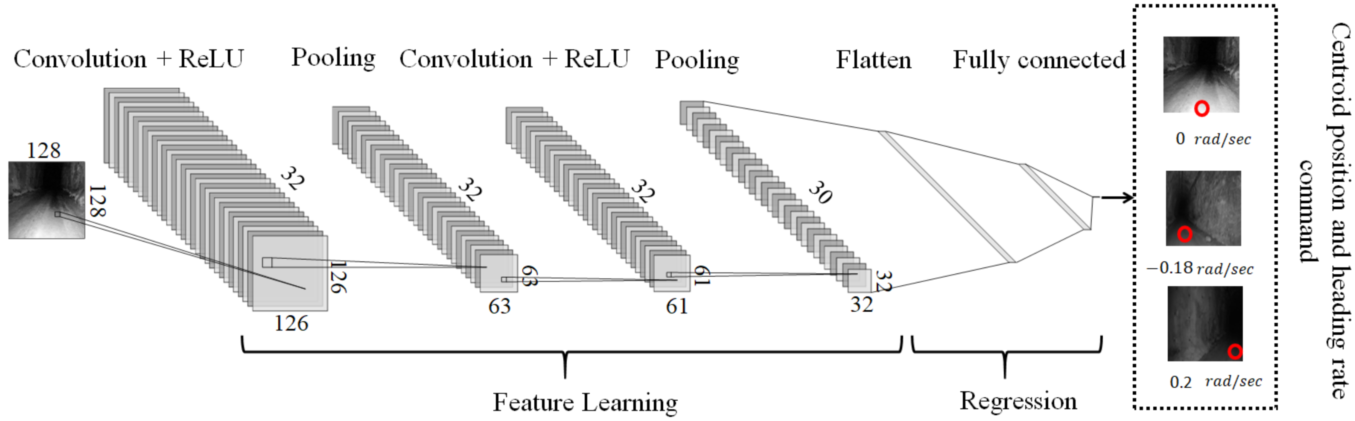

Fig. 4: Architecture of the proposed CNN for the estimation of the centroid with highest depth.

Algorithm 1 Calculate heading rate command.

Input: I
Output: $\dot{\psi}_d$

$I^{m\times n\times 3} \rightarrow I^{128\times 128\times 1}$ //converting the RGB image with $m\times n$ pixels to gray scale and resizing to 128×128 pixels
CNN($I^{128\times 128\times 1}$) $\rightarrow o_{cnn} \in [0,1]$ //Output of the CNN
$\dot{\psi}_d = \frac{o_{cnn}-0.5}{2.5}$ //Mapping $[0,1] \rightarrow [-0.2, 0.2]$ rad/sec

Fig. 5: The developed ROSflight based quad-copter [25] equipped with 2D and one beam lidars, optical flow, PlayStation camera and LED bars.

images corresponding to $50\,\mathrm{m}$ tunnel length are selected, while the data-set is shuffled and 70% are used for training and 30% for the validation in an offline procedure. The trained network provides MAPE of 10.4% and 12.9% on training and validation data-sets.

IV. RESULTS

This section describes the experimental setup and experimental evaluation of the proposed novel CNN method for sending heading rate commands at a MAV, autonomous flying at a deep underground mining environment. The following link provides a video summary of the overall results: https://youtu.be/WKHEvcovXqk.

A. Experimental Setup

In this work, a low-cost quadcopter has been utilized in an underground mine, which was developed at Luleå University of Technology based on the ROSflight [24] flight controller. The developed platform is presented in [25], while Figure 5 depicts the platform equipped with playstation camera.

B. Experimental Evaluation

The performance of the proposed CNN method is evaluated in an $790\,\mathrm{m}$ deep underground mine, while the underground tunnels did not have strong corrupting magnetic fields, their morphology resembled an S shape environment with small inclination. The dimensions of the area where the MAV navigates autonomously were $6(\text{width})\times 4(\text{height})\times 20(\text{length})\mathrm{m}^3$.

The ROSflight based quad-copter is equipped with a PlayStation Eye camera, which is operated at $10\,\mathrm{fps}$ and with a resolution of $640 \times 480\,\mathrm{pixels}$, while the LED light bars provide $460\,\mathrm{lux}$ illumination in $1\,\mathrm{m}$ distance. The desired altitude was set to $1\,\mathrm{m}$, the constant $v_{d,y}$ of $0.0\,\mathrm{m/s}$, and varying $v_{d,x}$ from $0.3\,\mathrm{m/s}$ up to $0.6\,\mathrm{m/s}$ were provided for the platform.

Figure 6 shows the heading rate command generated by the CNN module, that due to the narrow width of the tunnel and the corresponding camera field of view, small heading angle rotations results to replacement of the centroid in another direction, thus the heading rate commands are generated in different signs frequently.

In Figure 7 some examples from the on-board image stream during the autonomous navigation are depicted, while the centroids obtained from the centroid extraction method and the CNN are compared. Moreover, it should be highlighted that the CNN only estimates the $\hat{u}$ centroid position and the input image of the CNN is resized, however for comparison in the following figure it is assumed that the $\hat{v}$ position of the centroid and resolution of the image are the same in both cases.

V. CONCLUSIONS

This article presented a CNN method to enable the autonomous aerial navigation in unknown dark underground mines. The image haze removal method was used to extract the centroid of a highest depth information for training the CNN

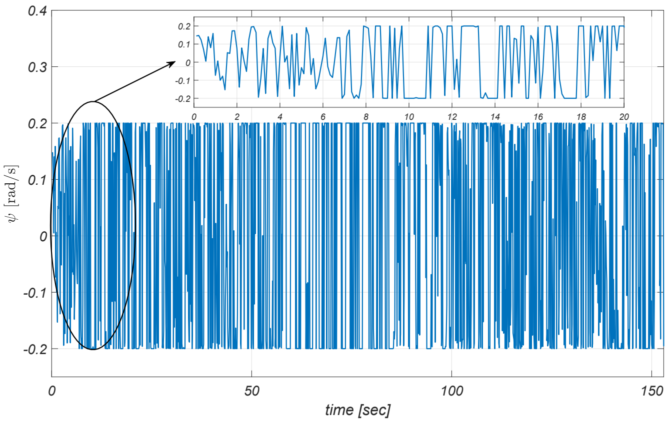

Fig. 6: The heading rate commands generated from the utilized CNN.

Fig. 7: Comparison of $\hat{u}$ from CNN and centroid extraction method from MAV on-board camera. The CNN estimation is indicated by a blue circle, while the centroid extraction method is indicated by a red circle.

from underground mine collected data-sets. The regression CNN provided a heading rate commands for the MAV, while the platform was considered as a floating object. In the presented approach, the potential fields, based on a 2D lidar scanning, were used to avoid collision to walls. The framework has been evaluated in field trials inside a dark underground mine tunnel located in Sweden and has successfully managed to navigate autonomously without colliding to the wall surfaces.

VI. Acknowledgement

We would like to thank Boliden AB for providing the access to underground mine and for giving the opportunity to evaluate the proposed methods in real life challenging conditions and Wolfit AB for the valuable guidance and support during all our mine visits.

References

[1] C. Kanellakis, S. S. Mansouri, G. Georgoulas, and G. Nikolakopoulos, "Towards Autonomous Surveying of Underground Mine Using MAVs," in *International Conference on Robotics in Alpe-Adria Danube Region*. Springer, 2018, pp. 173–180.

[2] S. S. Mansouri, C. Kanellakis, E. Fresk, D. Kominiak, and G. Nikolakopoulos, "Cooperative coverage path planning for visual inspection," *Control Engineering Practice*, vol. 74, pp. 118–131, 2018.

[3] P. Oettershagen, T. Stastny, T. Mantel, A. Melzer, K. Rudin, P. Gohl, G. Agamennoni, K. Alexis, and R. Siegwart, "Long-endurance sensing and mapping using a hand-launchable solar-powered uav," in *Field and Service Robotics*. Springer, 2016, pp. 441–454.

[4] S. Salazar-Colores, I. Cruz-Aceves, and J.-M. Ramos-Arreguin, "Single image dehazing using a multilayer perceptron," *Journal of Electronic Imaging*, vol. 27, no. 4, p. 043022, 2018.

[5] C. Kanellakis and G. Nikolakopoulos, "Survey on computer vision for uavs: Current developments and trends," *Journal of Intelligent & Robotic Systems*, pp. 1–28, 2017.

[6] S. Saha, A. Natraj, and S. Waharte, "A real-time monocular vision-based frontal obstacle detection and avoidance for low cost uavs in gps denied environment," in *2014 IEEE International Conference on Aerospace Electronics and Remote Sensing Technology*. IEEE, 2014, pp. 189–195.

[7] F. Valenti, D. Giaquinto, L. Musto, A. Zinelli, M. Bertozzi, and A. Broggi, "Enabling computer vision-based autonomous navigation for unmanned aerial vehicles in cluttered gps-denied environments," *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 3886–3891, 2018.

[8] A. Giusti, J. Guzzi, D. C. Ciresan, F.-L. He, J. P. Rodríguez, F. Fontana, M. Faessler, C. Forster, J. Schmidhuber, G. Di Caro *et al.*, "A machine learning approach to visual perception of forest trails for mobile robots." *IEEE Robotics and Automation Letters*, vol. 1, no. 2, pp. 661–667, 2016.

[9] L. Ran, Y. Zhang, Q. Zhang, and T. Yang, "Convolutional neural network-based robot navigation using uncalibrated spherical images," *Sensors*, vol. 17, no. 6, p. 1341, 2017.

[10] A. Loquercio, A. I. Maqueda, C. R. del Blanco, and D. Scaramuzza, "Dronet: Learning to fly by driving," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 1088–1095, 2018.

[11] A. Kouris and C.-S. Bouganis, "Learning to fly by myself: A self-supervised cnn-based approach for autonomous navigation," in *Intelligent Robots and Systems (IROS), 2018 IEEE/RSJ International Conference on*, 2017.

[12] S. S. Mansouri, C. Kanellakis, G. Georgoulas, and G. Nikolakopoulos, "Towards MAV navigation in underground mine using deep learning," in *IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2018.

[13] E. Small, P. Sopasakis, E. Fresk, P. Patrinos, and G. Nikolakopoulos, "Aerial navigation in obstructed environments with embedded nonlinear model predictive control," in *2019 European Control Conference (ECC)*. IEEE, 2019.

[14] F. Cozman and E. Krotkov, "Depth from scattering," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 1997, pp. 801–806.

[15] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2010.168

[16] S. Theodoridis and K. Koutroumbas, *Pattern Recognition, Fourth Edition*, 4th ed. Orlando, FL, USA: Academic Press, Inc., 2008.

[17] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[19] H. M. Bui, M. Lech, E. Cheng, K. Neville, and I. S. Burnett, "Using grayscale images for object recognition with convolutional-recursive neural network," in *IEEE Sixth International Conference on Communications and Electronics (ICCE)*, 2016, pp. 321–325.

[20] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[21] A. Saxena, "Convolutional neural networks (cnns): An illustrated explanation," *URL https://xrds. acm. org/blog/2016/06/convolutional-neural-networks-cnns-illustrated-explanation/. Last updated*, pp. 06–29, 2016.

[22] F. Chollet *et al.*, "Keras," https://github.com/fchollet/keras, 2015.

[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[24] J. Jackson, G. Ellingson, and T. McLain, "ROSflight: A lightweight, inexpensive MAV research and development tool," in *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*, June 2016, pp. 758–762.

[25] S. S. Mansouri, M. C. Arranz, C. Kanellakis, and G. Nikolakopoulos, "Autonomous MAV navigation in underground mines using darkness contours detection," in *12th International Conference on Computer Vision Systems (ICVS 2019)*, 2019.