Towards MAV Navigation in Underground Mine Using Deep Learning

Sina Sharif Mansouri[1], Christoforos Kanellakis[1], George Georgoulas[1], and George Nikolakopoulos[1]

***Abstract*— The usage of Micro Aerial Vehicles (MAVs) is rapidly emerging in the mining industry to increase overall safety and productivity. However, the mine environment is especially challenging for the MAV's operation due to the lack of illumination, narrow passages, wind gusts, dust, and other factors that can affect the MAV's overall flying capability. This article presents a method to assist the navigation of MAVs by using a method from the field of Deep Learning (DL), while considering a low-cost platform without high-end sensor suits. The presented DL scheme can be further utilized as a supervised image classifier that has the ability to process the image frames from a single on-board camera and to provide mine tunnel wall collision prevention. The efficiency of the proposed scheme has been experimentally evaluated in two underground tunnel environments that were used for data collection, training, and corresponding testing under multiple flying scenarios with different cameras configurations and illuminations.**

Fig. 1: Photo from an underground tunnel in Boden, Sweden for indicating the lighting conditions in a mine tunnel, while the effect of an artificial light source being also evident.

I. INTRODUCTION

Micro Aerial Vehicles (MAVs) are platforms that have received great attention during the last decade due to their mechanical simplicity, agility and hovering ability [1]. These platforms have the potential to provide leading solutions in a wide range of applications, especially in hostile or challenging environments, such as underground mines [2]. Thus, the MAVs can provide access to unreachable, complex, dark and dangerous locations for the monitoring personnel, while minimizing service times. Overall, the deployment of MAVs can have a high impact on the mine's operation, production, and safety. However, as depicted in Figure 1, low illumination, narrow passages, uneven surfaces, and dust, are conditions commonly found in underground mining tunnels. These dark and featureless environments challenge the state estimation schemes, since range sensors and cameras do not yield sufficient information [3]. Therefore, there is a need to develop advanced control, navigation and perception modules to compensate for these challenges, towards the establishment of autonomous aerial platforms in underground areas.

One of the main research challenges in mine navigation is the definition of a proper heading (yaw) for the MAV, since the mine is an environment where there is lack of visual and geometric features and there is a general absence of illumination. Thus, to successfully navigate in a tunnel, the MAV has to identify the direction of the flight in unknown areas, while avoiding the surrounding walls, without depending on accurate state estimation. This article proposes a module, for a low-cost MAV platform, for navigating in low illumination environments, equipped with an on-board camera and an LED light bar. In the presented approach and independently to the type of the on-board camera, the images are resized to 128×128 pixels to reduce noise and computation time, while the RGB images are converted to gray-scale, since the encapsulated color information is redundant. In the sequel, these images are fed to a Convolutional Neural Network (CNN) [4], probably the most prominent member of the Deep Learning (DL) family. The CNN is used for classifying the images into three categories: *left*, *middle* and *right*. This categorization is critical for preventing potential collisions onto walls, while the heading of the MAV can be corrected based on the observed image; as an example, if the image is categorized as *right* then the heading should go to left. As it will be presented, the network is trained in two different underground tunnel environments and evaluated in different scenarios.

In the related literature, there have been many works that addressed the navigation problem in 2D and 3D environments, while the three main exploration methods have been the entropy based [5], the frontier based [6], and the information gain based exploration [7]. The entropy gain and information gain based methods, compute regions that reduce the map uncertainty, based on current information on the map. Furthermore, in the frontier based exploration approaches, the exploration frontiers are computed as the discrete boundary between the unknown regions of the current map, while these methods have been successfully applied in 2D environments [8]. Towards a 3D exploration, in [9]

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

[1] Robotics Team, Department of Computer Science, Space and Electrical Engineering, Luleå University of Technology, Luleå SE-97187, Sweden.
Corresponding author Email: sinsha@ltu.se

it has been proposed a stochastic differential equation-based algorithm to enable exploration in indoor environments. This method resulted from the evaluation of existing methods of frontier based exploration, however, there were no requirements of dense representation of the free space. In [10], random trees were generated to find the best branch, while the method was evaluated in indoor environments and the paths were calculated on-line, while the occupancy map of the perceived environment was conducted. These exploration methods require in general a high computation power to process the images, to calculate the best next point and to accurately localize and store the previous information of the map in order to avoid revisiting the area. This fact limits the usage of these methods in large-scale structures, while at the same time there have been very few works that considered the navigation problem in dark tunnels. In [3], the authors addressed the problem of estimation, control, navigation and mapping problems for autonomous inspection of tunnels using aerial vehicles with the overall approach to be validated through field trials, however, the high-end sensor suit that was utilized has limited the applicability of the overall method.

Furthermore, there are few works using machine learning techniques for the problem of navigation in in-door and out-door environments, mainly due to the fact that these methods require a large amount of data and a high computation power for training, in most cases a CNN, which is an off-line procedure. However, after the training the CNN can be used for enabling an autonomous navigation with much lower computation powers. The works using CNN for navigation, such as [11], [12], [13], utilized the image frame of on-board camera to feed the CNN for providing heading commands to the platform. These works have been evaluated and tunned in out-door environments and with a good illumination with the camera and thus providing rich data about the surrounding of the platforms.

Based on the aforementioned state of the art, the main contribution of this article is threefold. Firstly, this article establishes a method towards the navigation in dark and unknown environments, using a low-cost MAV platform, equipped with a single camera and a LED light bar. The proposed method classifies the images based on a CNN to identify the tunnel walls and void space, while providing information regarding the direction of the camera, which can be left, center or right. This information as a future step can be used further for correcting the heading of the MAV and for avoiding collisions to the mine walls without depending on localization information. Secondly, the proposed novel method has been evaluated for the first time ever in field trials and more specifically, two underground tunnels were visited with different dimensions, while the trained network was evaluated in multiple scenarios with different camera configurations and illuminations to prove the robustness of the method in unknown mine tunnels. The final contribution stems from the fact that datasets were gathered from manual flights in order to evaluate the CNN when there are uncertainties in the height of the camera and faster motions. It should be also noted that this article is accompanied with the public release of all collected datasets from the tunnel environments, to provide the research community open access data and by that enabling further developments towards the envisioned autonomous flying in the dark.

The rest of the article is structured as it follows. Initially, the problem formulation of the proposed method is presented in Section II, followed by the CNN implementation in Section III. In Section IV multiple evaluations based on collected datasets with triple camera setup and manual flights with MAV are presented, followed by a corresponding comparison and discussion. Finally Section V concludes the article by summarizing the findings and offering some directions for future research.

II. Problem Formulation

In general, an underground environment is considered harsh for the operation of MAVs, thus it poses multiple challenges, like low illumination, narrow passages, dust, wind gusts and short line-of-sight. Usually, the aerial platforms are built with high-end and expensive components, to reach increased levels of autonomy that can provide stability and reliability in their operation, while the long-term operation of these platforms, in such environments, degrades their performance and integrity over time.

Therefore, the overall aim of mining companies is to consider the aerial vehicles as consumables that can be instantly replaced. Therefore, low-cost solutions are lately introduced that can accomplish the task equally reliable. Generally, the state estimation is the core of MAVs, providing the basis for planners to build on top and fulfill exploration tasks. In underground tunnels, localization schemes become unreliable with low-cost sensors, therefore advanced algorithms need to be developed to compensate for their inefficiency and accomplish the task successfully. Towards this direction, this article, identifies the position estimation issues, tackles the inspection problem from another perspective, while proposing a generic method for wall collision prevention towards the navigation around the tunnel, without depending on accurate localization schemes. More specifically, the proposed method incorporates a high-level planning module that distinguishes the walls from the tunnel axis, using a CNN and a single on-board camera. The output of the developed module can be used by the low-level controller of the platform to generate heading commands for avoiding the corresponding obstacles, including the mine walls.

An overview of the proposed concept is depicted in Figure 2, where the overall trajectory of the MAV, the tunnel axis and the heading direction of the MAV are also presented. To avoid collisions, while navigating around the tunnel, the platform should correct its heading commands according to the information provided by the wall detection module. In this inspection scenario, the height of the platform is constant and the platform is commanded to move forward following the tunnel axis. At this point, it should be highlighted that this work focuses on the development and validation of a wall collision prevention module based on machine learning and it is considering as the first fundamental step towards autonomous MAV enabled surveillance in underground mining.

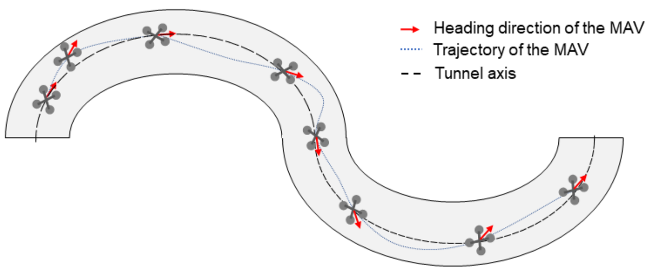

Fig. 2: Top-view concept image of a mine tunnel with a MAV and the corresponding resulted corrections in the heading.

Inspired by [12], in this article, a CNN is used as an image classifier. In this approach, three classes are considered that correspond to three different scenarios that are necessary for remaining in the tunnel axis, while the camera is looking in the direction of the movement. Each image from the on-board camera is classified to mutually exclusive categories: *left*, *middle* or *right*. The *left* and *right* categories correspond to the left and right walls, while the *middle* category corresponds to the tunnel axis. For the described problem, the overall system diagram is presented in Figure 2.

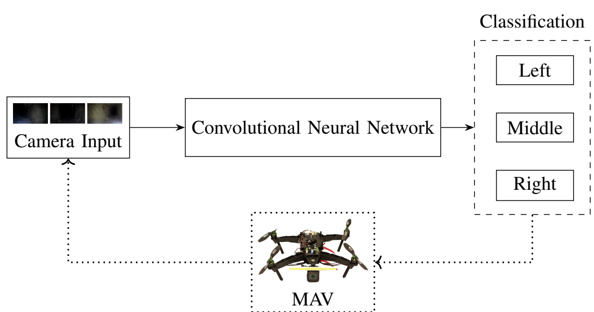

Fig. 3: The overall proposed architecture for the recognition of the MAV's heading; it should be highlighted that the overall closed loop process from the classification to the MAV heading commands are not studied in this article (dotted lines).

III. Convolutional Neural Network for Image Classification

The CNN [4] presented in Figure 4 receives a fixed-size image as an input and outputs one out of three categories for each image. Like most of other types of neural networks, a CNN is composed of an input layer, an output layer, and many hidden layers in between. These layers perform operations that alter the data with the intent of learning features specific to the task. The main difference of these novel architectures is that they do not rely on tedious feature engineering processes, instead the features are learned during the training process. In the case of CNNs this is basically achieved via the convolution filters, each of which learns to be activated by certain features of the image. An extra advantage of the convolution connections is that they dramatically reduce the number of parameters, especially when compared to a fully connected architecture, due to weight sharing. Apart from the convolutions nonlinearities are also part of the architecture which lately are in most cases ReLUs. The use of ReLUs allows for faster and more effective training by mapping negative values to zero and maintaining positive values. An extra layer, which is not part of the classic NNs id the pooling layer. Pooling simplifies the output by performing nonlinear down-sampling, while at the same time reducing the number of parameters that the network needs to learn. These operations are repeated over a large number of layers, with each layer learning to identify different features.

The input layer of the CNN is a matrix of 128×128, followed by a number of hidden layers and ending with a layer with three output neurons equipped with softmax activation functions (to provide outputs that sum to one). The input image can have different size depending on the on-board camera, however for the proposed scheme it should be resized to 128×128 pixels. Moreover, in [14] it was shown that the object recognition based on gray-scale images can outperform recognition based on RGB images, which is an ideal situation since the mine environments are dark and the RGB sensors do not provide any extra information about the environment. Thus, the images from the cameras are converted to gray-scale, which also reduce the noise and the computation time in the training phase. For each input, the CNN provides the probability of an image class, which can be *Left*, *Middle*, *Right* image. For the presented results, the CNN was trained on a workstation equipped with an Nvidia GTX 1070 GPU with 50 epochs, a selected initial learning rate of 1^{-4} and solved by the stochastic gradient descent with momentum optimizer.

IV. Experiments and Results

A. Data collection

For collecting the data sets, the setup depicted in Figure 5 was used, which consists of three mounted cameras with separate LED light bars pointing towards the field of view of each camera. For evaluating the method with uncertainties in the camera model, different types of cameras were used, including the GoPro Hero 3, the GoPro Hero 4 and the Foxeer Box. During the dataset collection for the training phase, the resolution of the cameras was fixed to 1920x1080 pixels, while before each run, the cameras were exchanging positions in a varying configuration to generalize the applicability of the method, while avoiding the dependency on a specific camera model. Furthermore, the light bars were calibrated to provide an equal illumination power. During the collection of the data sets, the setup can be carried by a person or installed on a manual flying MAV, while special care should be considered in guaranteeing that the middle camera is always looking towards the tunnel axis. In the presented approach, the cameras were recording the video with 60 Frame Per Seconds (FPS), however the videos were

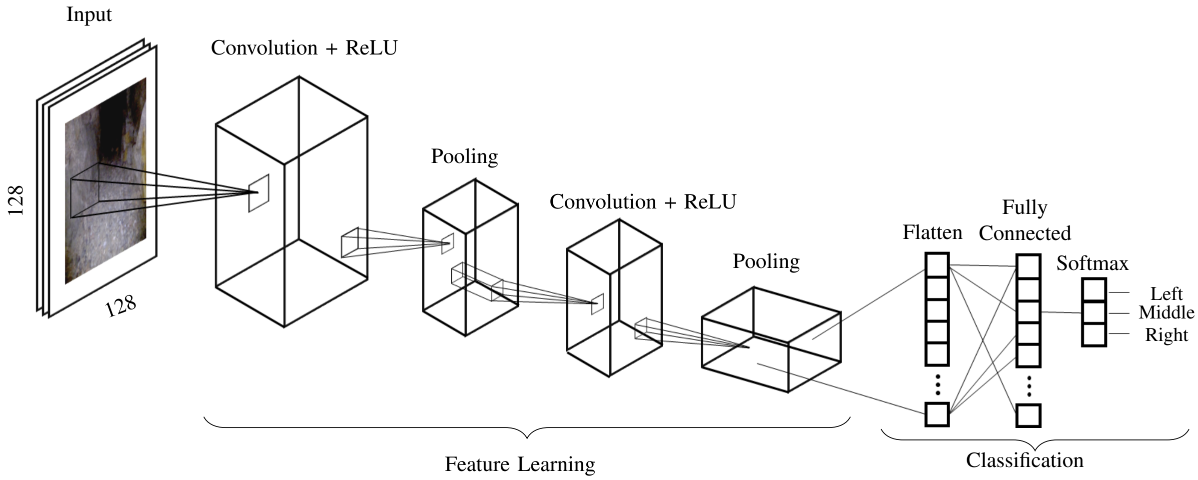

Fig. 4: Architecture for the proposed CNN for heading clasification.

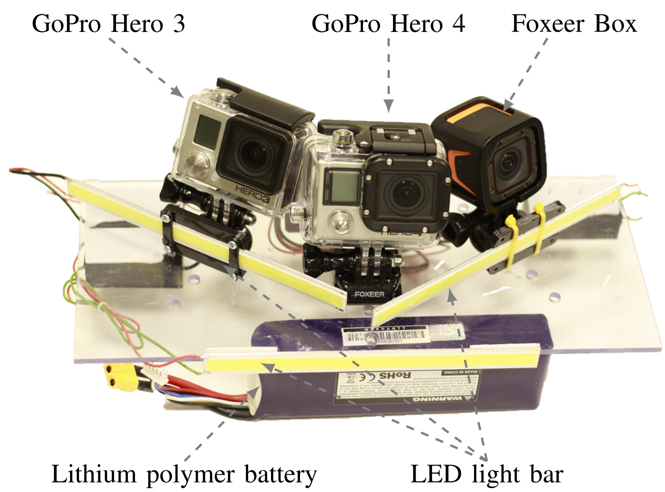

Fig. 5: Top view of the setup for obtaining datasets.

down-sampled to 10 FPS and converted to sequential images, in order to reduce the redundancy of the images, however faster FPS could be selected without a loss of generality. In the sequel, the images were resized and converted to a gray-scale mode and labeled based on the direction of the camera and as a last step, the images were fed to the CNN.

For collecting data for the proposed training, two underground locations were visited. The first one located in Luleå, Sweden and the second one located in Boden, Sweden. The testing areas had the following dimensions $400 \times 2.5 \times 3 \text{ m}^3$ and $50 \times 2 \times 2 \text{ m}^3$ respectively. Both environments were completely dark with no external illumination was available and with uneven stone surfaces. Furthermore, the areas did not have a strong corrupting magnetic field, however, small particles were floating on the air. Figure 6 depicts sample images of different areas of each location, which are part of multiple data sets gathered during the trials.

B. Train and Evaluate the CNN

For training and evaluating the CNN, multiple scenarios were defined, while the main purpose of the following scenarios was to evaluate the performance of the CNN with different cameras and multiple underground environments.

1) Luleå mine scenario: In the first scenario, the CNN was trained with 6066 images from the Luleå underground tunnel location and evaluated in the Boden underground tunnel location datasets. The camera configuration for the training phase was GoPro Hero3, Foxeer Box and GoPro Hero 4 from left to right respectively and the LED light bars had an illumination power of 6447 lux. During the dataset collection for the testing phase of the network, the Foxeer Box and GoPro Hero 4 exchanged their positions. The accuracy on the Luleå underground tunnel location was 97.74%, while for the Boden underground tunnel location was 96.27%. The CNN was tested on the Boden images which were not used during training. In Figure 7 some samples of input data used for evaluating the CNN from the Boden underground tunnel location are depicted, where the correct class is written in blue text above each image.

2) Boden underground tunnel scenario: In this scenario, the Boden underground tunnel location dataset was used for training the CNN, while the dataset from the Luleå underground tunnel were used for evaluating the network's performance. In this case, the tunnel is shorter than the one in the previous case, thus the training datasets are smaller in size. In this case 1616 images were used for the training phase, while the CNN had the same architecture as discussed in Section III. The camera configuration was GoPro Hero 3, GoPro Hero 4 and Foxeer Box from left to right respectively with an illumination power of 5407 lux. The trained network had an accuracy of 99.6% on the training dataset and 65.2% on the evaluation data set respectively. This drop of accuracy in comparison to the Luleå underground tunnel scenario was expected as the Boden underground tunnel location was shorter in tunnel's length, thus the network was trained with less (most probably not sufficiently enough) information.

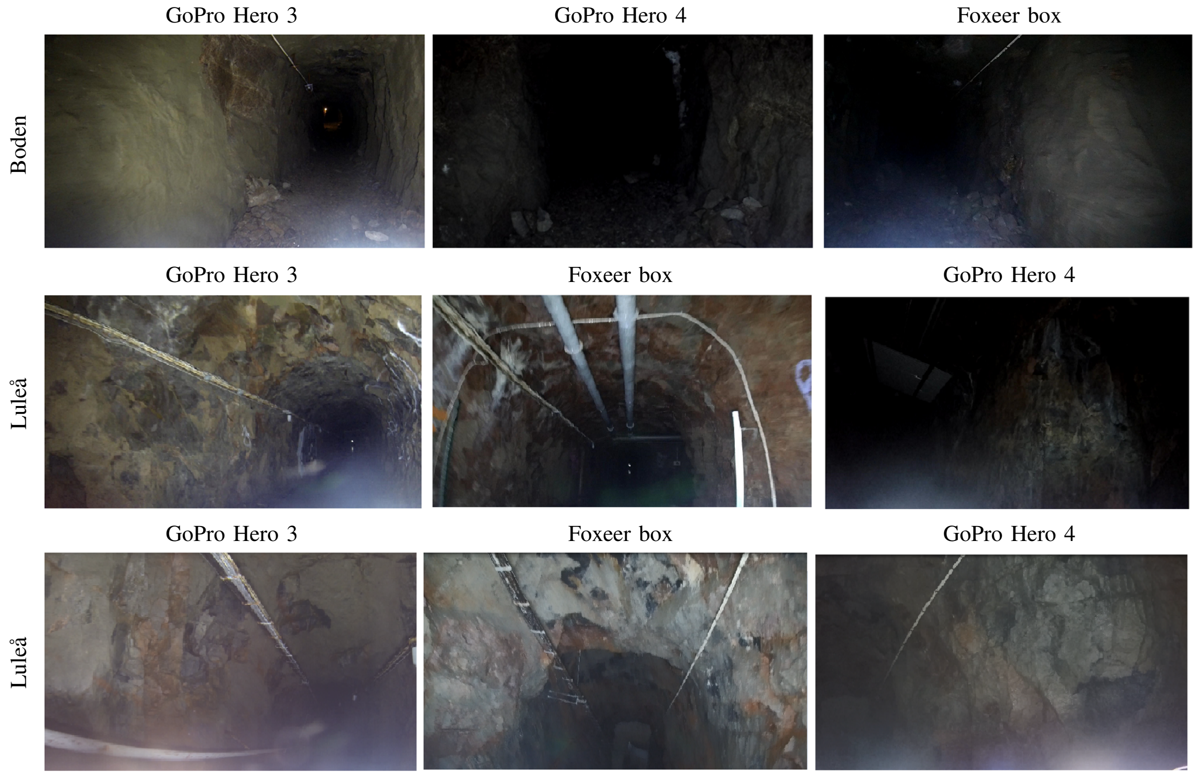

Fig. 6: Examples of collected images used for training the CNN. The left, center and right images are from cameras looking toward left, center and right respectively.

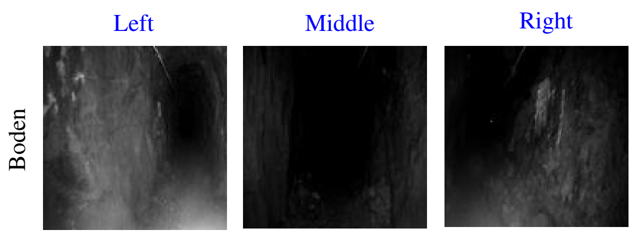

Fig. 7: Examples of images for testing the CNN with the Luleå underground tunnel dataset, where the class of images from the CNN is written above the images and with the CNN to be trained from the Boden underground tunnel datasets.

Figure 8 provides some samples of the mismatch and correct classification of the CNN. The correct class of images is written in green in the right side of images, while the output of CNN is written above each image. The red text means mismatch class, while the blue text is the correct class for each image.

3) Effect of Illumination and flight: In this scenario the CNN from the Luleå underground tunnel was selected as it had a higher accuracy in classification of the images. The CNN was trained by Luleå underground tunnel location data and it is evaluated with the same camera configuration and environment, while different levels of illuminations are used ranging from 6447 lux to 3547 lux. The accuracy of the CNN based on illumination levels is depicted in Figure 9, where it is shown that even though the training and testing environments are the same, the accuracy of the CNN is directly related to illumination, where smaller illumination levels lead to less accurate results. However, the accuracy reduction is different for each camera, where GoPro Hero 3 located on the left position of the setup has the highest drop and the Foxeer Box located in the middle position on the setup has the lowest changes.

Furthermore, manual flights of a custom made and low-cost MAV have been performed,to evaluate the CNN ability to identify the walls and correct the heading. The platform was equipped with a forward-looking camera and a LED light bar. The CNN was evaluated with datasets collected with different cameras on the platform (one camera flying each time), such as GoPro Hero 3, Foxeer Box and Intel RealSense camera (R200). However, due to safety and narrow areas the MAV mainly was looking forward the tunnel axis. The video of the evaluation can be reached at (https://youtu.be/uJFvTGnrPAY), where it can be seen that the CNN is quite accurate except in cases when the MAV is very close to the ground, a fact that shows the impact of the height in the accuracy of the classification. Additionally, the datasets obtained from the RealSense were more difficult to categorize in the proposed CNN scheme, that may result from the fact that no training datasets from this camera were used. However, more tests are needed to evaluate the results with different camera models in order to reach safer conclusions.

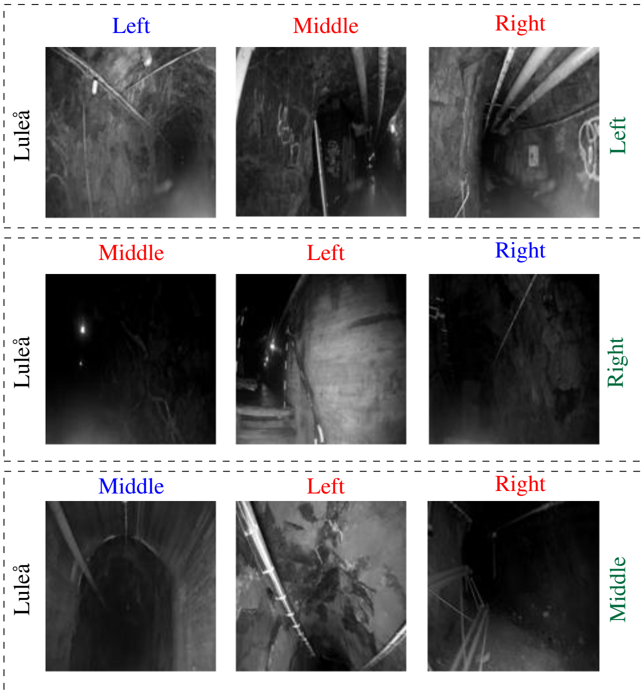

Fig. 8: Examples of images for testing the CNN with the Luleå underground tunnel dataset, trained using the Boden underground tunnel datasets. The class of images from the CNN is written above the images (red-mismatch class, blue-correct class) and the correct label of each image is written in the right with green.

V. Discussion and Conclusions

In this article, the problem of navigation in mine environments was studied to enable, safe and collision free inspection with MAVs. In the proposed approach a CNN was used to classify the images from the camera to *left*, *middle*, *right*. The trained network was evaluated in multiple scenarios and in most of the cases the CNN provided high accuracy. It was shown that the accuracy of the CNN depends on the number of training data and the illumination of the environment (the CNN accuracy decreases when the illumination is reduced). In future work, the network should be evaluated in a closed loop system with the MAV equipped with the camera and a LED light bar to correct the heading of the platform and enable autonomous navigation in a mine tunnel. Toward this direction the platform should be equipped with sensors such as sonars to keep its distance from ceiling and ground constant, while the CNN should provide commands for going forward on the tunnel axis. Additionally, the network should be evaluated in longer and more complex mine environments with multiple branches in the tunnel. Finally, the effect of the illumination in the training and testing stage of the CNN should be studied, as it is shown that the results are sensitive to the lighting of the environment.

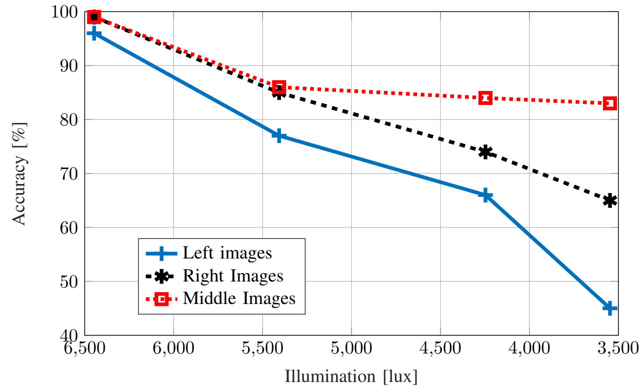

Fig. 9: The accuracy of the trained network in relation to the illumination in the same environment.

References

[1] C. Kanellakis and G. Nikolakopoulos, “Survey on Computer Vision for UAVs: Current Developments and Trends,” *Journal of Intelligent & Robotic Systems*, pp. 1–28, 2017.

[2] C. Kanellakisa, S. S. Mansouri, G. Georgoulas, and G. Nikolakopoulos, “Towards autonomous surveying of underground mine using mavs,” *27th International Conferenceon Robotics in Alpe-Adria-Danube Region*, 2018.

[3] T. Özaslan, G. Loianno, J. Keller, C. J. Taylor, V. Kumar, J. M. Wozencraft, and T. Hood, “Autonomous navigation and mapping for inspection of penstocks and tunnels with MAVs,” *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1740–1747, 2017.

[4] 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.

[5] W. Burgard, M. Moors, C. Stachniss, and F. E. Schneider, “Coordinated multi-robot exploration,” *IEEE Transactions on Robotics*, vol. 21, no. 3, pp. 376–386, 2005.

[6] J. A. P. Fentanes, R. F. Alonso, E. Zalama, and J. G. García-Bermejo, “A new method for efficient three-dimensional reconstruction of outdoor environments using mobile robots,” *Journal of Field Robotics*, vol. 28, no. 6, pp. 832–853, 2011.

[7] S. Bhattacharya, N. Michael, and V. Kumar, “Distributed coverage and exploration in unknown non-convex environments,” in *Distributed Autonomous Robotic Systems*. Springer, 2013, pp. 61–75.

[8] A. Cowley, C. J. Taylor, and B. Southall, “Rapid multi-robot exploration with topometric maps.” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 1044–1049.

[9] S. Shen, N. Michael, and V. Kumar, “Stochastic differential equation-based exploration algorithm for autonomous indoor 3d exploration with a micro-aerial vehicle,” *The International Journal of Robotics Research*, vol. 31, no. 12, pp. 1431–1444, 2012.

[10] A. Bircher, M. Kamel, K. Alexis, H. Oleynikova, and R. Siegwart, “Receding horizon next-best-view planner for 3d exploration,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 1462–1468.

[11] S. P. Adhikari, C. Yang, K. Slot, and H. Kim, “Accurate natural trail detection using a combination of a deep neural network and dynamic programming,” *Sensors*, vol. 18, no. 1, p. 178, 2018.

[12] 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.

[13] 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.

[14] 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.