An Optimised Deep Neural Network Approach for Forest Trail Navigation for UAV Operation within the Forest Canopy

B. G. Maciel-Pearson, T.P. Breckon
{b.g.maciel-pearson, toby.breckon}@durham.ac.uk

Department of Computer Science, Durham University, UK

Abstract

Autonomous flight within a forest canopy represents a key challenge for generalised scene understanding on-board a future Unmanned Aerial Vehicle (UAV) platform. Here we present an approach for automatic trail navigation within such an environment that successfully generalises across differing image resolutions - allowing UAV with varying sensor payload capabilities to operate equally in such challenging environmental conditions. Specifically, this work presents an optimised deep neural network architecture, capable of state-of-the-art performance across varying resolution aerial UAV imagery, that improves forest trail detection for UAV guidance even when using significantly low resolution images that are representative of low-cost search and rescue capable UAV platforms.

1 Introduction

Scene understanding within unstructured environments with varying illumination conditions are critical for autonomous flight within the forest canopy. Growing interest in solving this challenge has motivated researchers to investigate the use of Deep Neural Networks (DNN) to identify trail images for UAV navigation. However, in order to train such a DNN, a large volume of labelled data is required, which is challenging to obtain due to the target task in hand (i.e. sub canopy UAV operation).

The work of [3] gathered data by using a head-mounted rig with three cameras worn by a human trail walker, allowing their proposed DNN architecture to identify the direction of the trail in a given view - $\{left, right, forward\}$. A similar approach is followed by [7] whereby a wide-baseline rig is used, also with three cameras mounted to gather data which they used to augment the dataset of [3] (IDSIA dataset). As a result, the approach presented by [7] is capable of estimating both lateral offset and trail direction. In both cases [3, 7], the authors, follow the common practice of dataset augmentation, via affine image transformations, which adds extra computation without any performance guarantees.

Alternatively, synthetic data, from virtual environment models, could potentially replace or at least supplement hard-won real environment data [5, 6, 9]. However, the significant discrepancy between synthetic data and real-world data often results in models that are trained only on synthetic environment examples not being able to directly transfer this knowledge to real-world operating tasks [2, 8, 5, 1].

Even when training a DNN using only real-world data, it is important to observe that models trained on a limited domain-specific dataset often fail to generalise successfully. In addition, since common DNN architectures require the dataset to be formed from fixed resolution images [4], models commonly fail to generalize across domains.

Our work here is closely related to [3, 7, 2] and demonstrates that the same trail direction required for automous UAV navigation can be acquired by using imagery gathered by a single forward-facing camera (Figure 1). This is due the fact that the center of the forward-facing camera usually shows the trail ahead. Additionally, we demonstrate that a trail can be identified in unseen parts of a forest by training the model with data gathered across varying devices and camera resolutions. This not only facilitates more general data gathering but also eliminates the need for synthetic data and augmentation. As result, the same model can be used by UAV with different sensor payload capabilities.

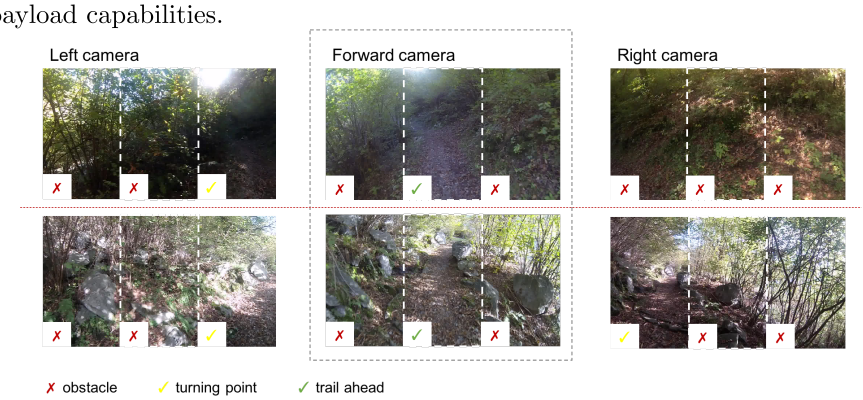

Figure 1: Comparison of three way image cropping performed on varied camera view

2 Method

Here we were motivated by the three class problem presented by [3] in which an estimation of the trail direction, $\{left, right, forward\}$, is achieved by processing an image triplet of left/right/forward camera views via a DNN. In contrast to [3, 7], our approach uses only a single forward facing camera view which more representative of an operational UAV case. This image view is then itself cropped into $\{left, right, forward\}$ which can be labelled for trail presence/absence (Figure 1).

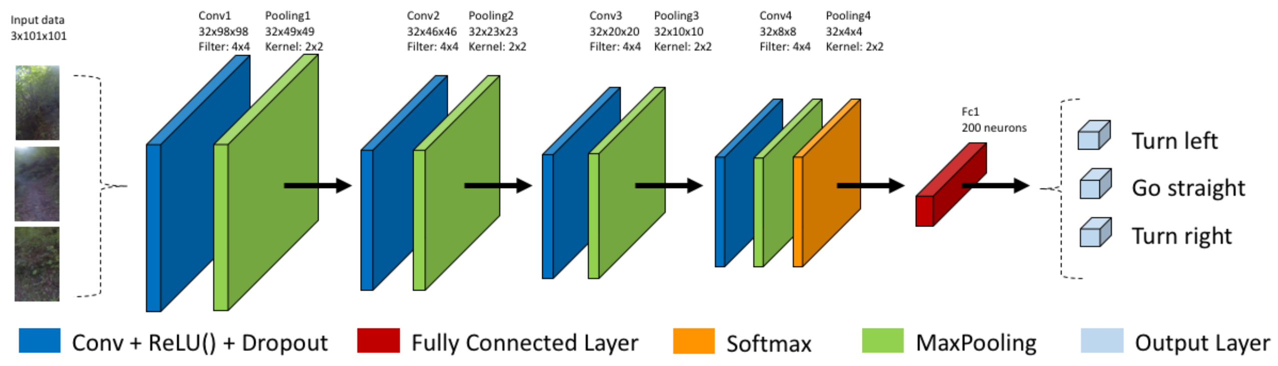

Figure 2: An outline of our DNN architecture - based on [3]

Using the architecture of [3] (illustrated in Figure 2), we evaluate varying image resolution, the use of additional data augmentation (DA) and activation function (*tanh() / ReLU()*). DNN training uses a gradient descent optimiser, random weight initialisation with zero node biases and is performed over 90 epochs with a 0.05 reduction in learning rate per epoch (decay rate: 0.95). For both training and testing we use the high-resolution (752×480) IDSIA dataset (from [3]) and a low-resolution (106×240) Urpeth Burn (UB) dataset, gathered locally. For training 45,097 high-resolution and 32,017 low-resolution image were used, while for testing 12,251 high-resolution and 5,152 low-resolutions images were used. Further data augmemtation (mirror, translation & rotation) was performed on a copy of this original dataset. For simplicity of reporting, we define NA as non augmented data obtained results and DA as data augmented obtained results (Table 1).

3 Results

Our experimental results are divided into three sets:- (1) image triplet approach of [3] with differing activation functions (*tanh()/ReLU()* - Table 1 upper 2 sets), (2) our proposed approach (single forward view image, split into three views - Table 1 middle sets in bold) and (3) the impact of high/low/varied image resolutions on performance (Table 1 lower sets). Overall we see what the use of the ReLU() activation outperforms tanh() and our approach gives high levels of accuracy without the need for data augmentation outperforming the prior reported results in [3] (*in fact no significant improvement was achieved by data augmentation*). Although our approach fails to generalise when trained with high-resolution images on to low resolution images, it achieves 82% accuracy when low-resolution images are added to the training dataset and achieves 78% accuracy for training and testing on low-resolution images only.

Method	Activation Function	Training Dataset	Testing Dataset	Training Loss	Training Accuracy	Test Loss	Test Accuracy
Giusti *et al.* [3] [DA]	*tanh()*	High	High	1.04	0.46	1.07	0.47
Giusti *et al.* [3] [NA]	*tanh()*	High	High	1.01	0.41	1.02	0.48
Giusti *et al.* [3] [DA]	*ReLU()*	High	High	0.02	1.00	2.97	0.59
Giusti *et al.* [3] [NA]	*ReLU()*	High	High	0.00	1.00	2.56	0.72
Our Approach [DA]	***ReLU()***	**High**	**High**	**0.18**	**0.92**	**0.26**	**0.89**
Our Approach [NA]	***ReLU()***	**High**	**High**	**0.07**	**0.97**	**0.31**	**0.89**
High Resolution[NA]	*ReLU()*	High	Low	0.17	0.94	1.00	0.44
Low Resolution [NA]	*ReLU()*	Low	Low	0.40	0.86	0.58	0.78
Varied Resolutions [NA]	***ReLU()***	**High+Low**	**Low**	**0.24**	**0.93**	**0.51**	**0.82**
Varied Resolutions [NA]	***ReLU()***	**High+Low**	**High**	**0.40**	**0.83**	**0.64**	**0.76**

Table 1: Results showing varying performance across High and Low image dataset combinations.

4 Conclusion

In this paper, we present an alternative method to gather and process UAV imagery that improves the level of accuracy for trail navigation under forest canopy based on the use of a single forward facing camera view instead of the triplet view approach of [3]. Our approach also performs well across varying image resolutions and increases the capability of low-cost UAV platforms with limited payload capacity. Future work will include additional aspects of UAV perception and control targeting end-to-end autonomy across this and other challenging operating environments.

References

[1] Paul Christiano, Zain Shah, Igor Mordatch, Jonas Schneider, Trevor Blackwell, Joshua Tobin, Pieter Abbeel, and Wojciech Zaremba. Transfer from simulation to real world through learning deep inverse dynamics model. *CoRR*, abs/1610.03518, 2016.

[2] Dhiraj Gandhi, Lerrel Pinto, and Abhinav Gupta. Learning to fly by crashing. *CoRR*, abs/1704.05588, 2017.

[3] Alessandro Giusti, Jerome Guzzi, Dan C. Ciresan, Fang-Lin He, Juan Pablo Rodriguez, Flavio Fontana, Matthias Faessler, Christian Forster, Jurgen Schmidhuber, Gianni Di Caro, Davide Scaramuzza, and Luca M. Gambardella. A machine learning approach to visual perception of forest trails for mobile robots. 1:1–1, 01 2015.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *CoRR*, abs/1406.4729, 2014.

[5] Klaas Kelchtermans and Tinne Tuytelaars. How hard is it to cross the room? - training (recurrent) neural networks to steer a UAV. *CoRR*, abs/1702.07600, 2017.

[6] Matthias Mueller, Vincent Casser, Neil Smith, Dominik L. Michels, and Bernard Ghanem. Teaching uavs to race using ue4sim. *CoRR*, abs/1708.05884, 2017.

[7] Nikolai Smolyanskiy, Alexey Kamenev, Jeffrey Smith, and Stan Birchfield. Toward low-flying autonomous MAV trail navigation using deep neural networks for environmental awareness. *CoRR*, abs/1705.02550, 2017.

[8] Lei Tai and Ming Liu. Deep-learning in mobile robotics - from perception to control systems: A survey on why and why not. *CoRR*, abs/1612.07139, 2016.

[9] Tianhao Zhang, Gregory Kahn, Sergey Levine, and Pieter Abbeel. Learning deep control policies for autonomous aerial vehicles with mpc-guided policy search. *CoRR*, abs/1509.06791, 2015.