UDC 629.735.018.7.015.3.025.1

I.O. Boryndo, V.M. Sineglazov
(National Aviation University, Ukraine)

Application of deep learning hybrid convolutional neural networks in visual navigation systems

In this paper review of modern applications for navigation algorithms was done. As a result, algorithmic shortcomings were identified and the usage of convolutional neural networks was proposed. Within the research the qualitative analysis of modern architectures of convolutional neural networks was carried out and their separate shortcomings at use in systems on the basis of aircraft hardware was shown.

I. Introduction

Nowadays, embedded systems ad-hoc designed to work in relative isolation are being replaced by AI-based sensor nodes that acquire information, analyse and interpret it, and utilize it to communicate with the environment and with one other. The "ultimate" IoT node will be able to navigate the surroundings independently while simultaneously perceiving, analysing, and comprehending it.

Complex aircrafts and completely autonomous unmanned aerial vehicles (UAVs) are ideal embodiments for this type of smart sensors: because to their speed and agility, they are able to rapidly gather data from both their onboard sensors and a multitude of devices placed in the surrounding environment. Additionally, aircraft might undertake sophisticated onboard analytics to pre-select vital data before transferring it to centralized computers. For unmanned aerial vehicles (UAVs), the small size of drones is ideal for both indoor applications where they must operate safely near humans (for surveillance, monitoring, ambient awareness, interaction with smart environments, etc.) and for densely populated urban areas where they can exploit complementary sense-act capabilities to interact with their surroundings (e.g., smart-building, smart-cities, etc.)[1].

The one of traditional approaches to autonomous navigation of aircraft or UAV is the so-called localization-mapping-planning cycle which is based on estimating UAV velocity using either offboard (e.g., GPS) or onboard (e.g., visualinertial) sensors, constructing a local 3D map of the environment, and planning a safe pathway across it. However, these solutions are prohibitively costly for systems with limited computing resources. Recent research has shown that considerably simpler algorithms based on convolutional neural networks (CNNs) are enough to enable basic reactive navigation of drones, even in the absence of an environment map. Nonetheless, their computational and power requirements exceed the budget of existing drone navigation engines, which are based on basic, low-power microcontroller units (MCUs).

Visual navigation is now one of a multitude of image recognition tasks. Utilizing convolutional neural networks is the key to solving these problems; nevertheless, a number of variables must be addressed, including the utilization platform, the accuracy of the goal, and the performance of the system. The majority of contemporary varieties of convolutional neural networks are significantly constrained

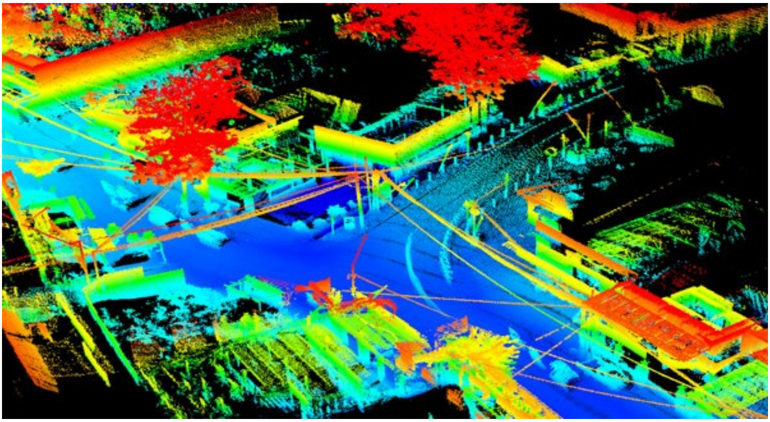

Fig. 1. Example of landscape-based visual analysis of urban surroundings.

because to their learning speed and complexity, and they also demand a high-quality training sample. Over time, the architectural complexity of convolutional neural networks has increased in order to solve such problems and improve result quality and accuracy, which has led to the emergence of new issues when further structural enrichment of convolutional neural networks encounters hardware limitations. In such circumstances, hybrid convolutional neural networks become vital. Multiple convolutional neural networks may be linked to build a hybrid convolutional neural network in order to improve overall performance and precision. In this article, we will discuss and show the findings of research into the use and architecture of hybrid convolutional neural networks, as well as their optimization and training.

II. Applications of hybrid convolutional neural networks within visual navigation

Visual navigation is a complicated system of interconnected algorithms for picture identification, processing, and enhancement. Since the primary process of implementing visual navigation is dependent on the identification of the surrounding environment, which is the raw graphical data by itself, the recognition process may be performed using convolutional neural networks. Different forms of navigational systems might be subdivided, resulting in a variety of tasks that need be performed using neural networks. The primary forms of visual-based navigation strategies include[3]:

• landscape-based visual analysis (it based on recognition the different in heights of surface, mapping the 3D map that is used to construct optimal flight-path) that is showed in figure 1;

• key-fractured visual navigation (visual identification of known object whom navigational data could be used by passing by aircraft or UAV).

The fundamental approach for localization and control is based on the concept of a vanishing point obtained from the picture received. Processing the received picture, identifying line characteristics, and scaling pixel offsets and pixel counts on a

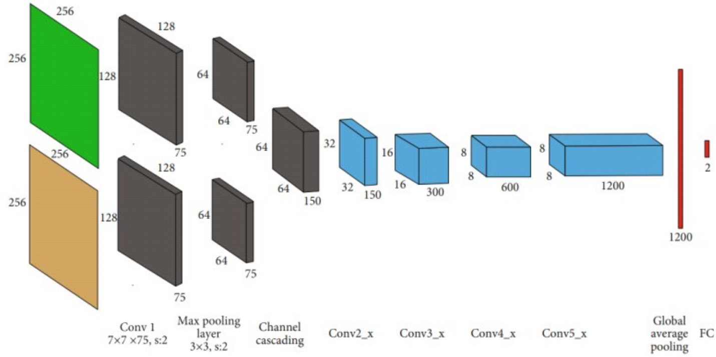

Fig. 2. Hybrid convolutional neural network simplified structural scheme.

line to real lengths determines intersections, dead-ends, and their closeness to the drone. Knowing the distance to an intersection or dead-end allowed the aircraft to descend in a timely manner in order to perform correct turning maneuvers. However, during the implementation of such a method, the following constraints must be considered:

- this method is highly dependent on visual clarity and lighting; many features and edges whose detection is crucial for identifying dead-ends or intersections may not be highly noticeable due to poor contrast, reflections, very florescent lighting (sunlight coming through introspective side), or occlusions due to environmental entities;
- scalability issue where adding more capabilities to the image processing components to handle a broader variety of situations did not scale or generalize well to the much increased number of conceivable scenarios that must be handled;

III. Topology analysis of modern convolutional neural network

Before delving into the specifics of the implementation of hybrid CNN systems for visual navigation, it is important to examine CNNs themselves. Today, convolutional neural network (CNN) is the fundamental technique for processing graphical input and extracting features. It is a prominent deep learning architecture. By analyzing a lot of prepared graphical input data, these neural networks may automatically extract the features or representation properties due to their unique architecture. One component of CNN extracts characteristics, while another processes and classifies them based on the original job requirements[2].

Similarly, a hybrid convolutional neural network is the merging of two or more distinct convolutional neural networks that are set and organized to function in pair (serial or parallel) and tackle a certain job or range of tasks.

The concept behind HCNN design is to execute the single-responsibility principle, which requires each system component (such as CNNs, classification/recognition algorithms, and input/output data processors) to do just one specified duty. Therefore, complicated specialized tasks may be broken down into needed dataflow phases that must be implemented. Each step will be carried out using its corresponding ingredient.

Having these reduced tasks makes training neural networks simpler and improves their accuracy and performance. As an example, the following networks may be merged: densely connected convolutional neural network in conjunction with squeezing and excitation convolutional neural network based on ResNeXt. Due to the global information held at the SE-CNN[5, 6] structure and the overall performance of DenseNet, it has great potential. While merging networks, the following factors must be taken into account:

• input data variables include starting scale, resolution, channels count, and identification task type;

• the first CNN's output must be suitable to the second CNN, and the original target data must be preserved;

• The topologies of both CNNs must be flexible and able to include layers such as normalization layers, residual blocks, dropout layers, 1x1 convolution layers, and so on.

In this method, any CNN type may be matched. It depends largely on the particulars of tasks and input data factors (e.g. image resolution, scale, color channels, number of training samples, etc.).

IV. Choosing the optimal hybrid convolutional neural network for visual navigation implementation in aircraft conditions

As the majority of visual navigation tasks are utilized on UAVs and aircrafts that are severely constrained in terms of computational speed and power consumption, it is necessary to select a convolutional neural network and architecture that can be easily streamlined and performed on peripherals with limited resources. As an example, we choose devices based on the STM32 architecture with 4-cores operating at nominal rates of 1.9(1.67) GHz, 1.1GB of LPDDR3X RAM, and an Adreno 440 GPU. Potential CNN architectures should not be too complex in terms of layer count and should have robust feedback links across structural levels. There are a few known architectures among contemporary convolutional neural networks that might be used for our needs:

- EfficientNet B0 & EfficientNet B3;
- MobileNet CNN & MobileNet V2 CNN;
- DroNet LVS 2;
- InceptionResNet V2 CNN;
- DenseNet201 & DenseNet169 & DenseNet121;
- Channel-boosted 2OR CNN;
- ResNet101;

It was anticipated that the Top-1 accuracy rate would stay the same throughout the conversion between TF and TFL for all architectures, as a difference would indicate implementation differences and impede the subsequent metric evaluations. The models with the best accuracy were DenseNet121 and DenseNet201 [4], both of which achieved 98.18 percent. We would like to point out that the accuracy depends on the application, therefore a lesser accuracy does not always indicate that the design should be abandoned.

Training			Inference - Fixed16					
			Original Dataset				HiMax Dataset	
Dataset	Max-Pooling	Data Type	EVA	RMSE	Accuracy	F1-score	Accuracy	F1-score
Original	3 × 3	Float32	0.758	0.109	0.952	0.888	0.859	0.752
	3 × 3	Fixed16	0.746	0.115	0.946	0.878	0.841	0.798
	2 × 2	Float32	0.766	0.105	0.945	0.875	0.845	0.712
	2 × 2	Fixed16	0.795	0.097	0.935	0.857	0.873	0.774
Original + HiMax	3 × 3	Float32	0.764	0.104	0.949	0.889	0.927	0.884
	3 × 3	Fixed16	0.762	0.109	0.956	0.894	0.918	0.870
	2 × 2	Float32	0.747	0.109	0.964	0.916	0.900	0.831
	2 × 2	**Fixed16**	**0.732**	**0.110**	**0.977**	**0.946**	**0.891**	**0.821**

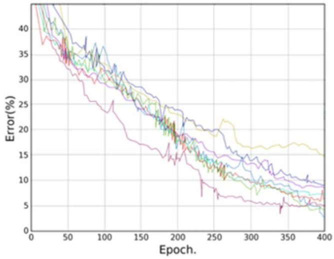

Fig. 3. DroNet accuracy results and learning process comparison of HCNN architectures

The only assessed metric that differed significantly between the TF model executed on a computer and the TFL model executed on a mobile device was inferential time. Due to structural variations in between two environments, there was a disparity in the latency of distinct processes. MobileNet was the TF model with the shortest inference time, while MobileNetV2 was the TFL model with the shortest inference time.

V. Conclusions

In this paper we've done the review of modern implementation ways of visual navigation in pair with different onboard aircraft systems and hardware used both for UAVs and aircrafts. In the result of visual navigation system analysis, the main sub-tasks were excluded. Based on the criteria that is related to the type of visual navigation approach, it's running hardware and usage approach, the classification list of recognition tasks and their implementation details was listed. To solve them we're recommended to use hybrid convolutional neural network systems as it's the best tool for image recognition and processing at this moment. To apply this neural network for image processing, the basic review of convolutional neural network essentials was considered, including resulting structural idea, the number or core layers and their configuration parameters.

References

[1] Iveta Mrazova, Marek Kukacka. "Hybrid convolutional neural networks" IEEE International Conference on Industrial Informatics 10.1109/INDIN.2008.4618146.

[2] Chaitanya Nagpal and Shiv Ram Dubey. “A Performance Evaluation of Convolutional Neural Networks for Face Anti Spoofing”.

[3] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” in Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014, pp. 675–678.

[4] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014.

[5] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440..

[6] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, Enhua Wu. “Squeeze-and-Excitation Networks”.