Intelligent control of quad-rotor aircrafts with a STM32 microcontroller using deep neural networks

Xiaochun Guan

College of Computer Science and Technology, Zhejiang University of Technology,
Hangzhou, China

Sheng Lou and Han Li

College of Electrical and Electronic Engineering, Wenzhou University,
Wenzhou, China, and

Tinglong Tang

China Three Gorges University, Yichang, China

Abstract

Purpose – Deployment of deep neural networks on embedded devices is becoming increasingly popular because it can reduce latency and energy consumption for data communication. This paper aims to give out a method for deployment the deep neural networks on a quad-rotor aircraft for further expanding its application scope.

Design/methodology/approach – In this paper, a design scheme is proposed to implement the flight mission of the quad-rotor aircraft based on multi-sensor fusion. It integrates attitude acquisition module, global positioning system position acquisition module, optical flow sensor, ultrasonic sensor and Bluetooth communication module, etc. A 32-bit microcontroller is adopted as the main controller for the quad-rotor aircraft. To make the quad-rotor aircraft be more intelligent, the study also proposes a method to deploy the pre-trained deep neural networks model on the microcontroller based on the software packages of the RT-Thread internet of things operating system.

Findings – This design provides a simple and efficient design scheme to further integrate artificial intelligence (AI) algorithm for the control system design of quad-rotor aircraft.

Originality/value – This method provides an application example and a design reference for the implementation of AI algorithms on unmanned aerial vehicle or terminal robots.

Keywords Deployment, Sensor fusion, Microcontroller, On-device machine learning

Paper type Technical paper

1. Introduction

With the new round of development on artificial intelligence (AI), which is typically represented by deep learning, the deep neural networks model has achieved amazing results in pattern recognition and computer vision. Deep neural networks have demonstrated near-human performance in computer vision applications such as image classification, detection and segmentation (He *et al.*, 2016; Goodfellow *et al.*, 2014). Deep learning models are also significant for simultaneous localization and mapping (Zhang *et al.*, 2019a; Lu *et al.*, 2019). The research and development of “AI chip,” a module specially designed to handle a large number of computing tasks in AI applications, has also become the focus of the industry. As there is still a big theoretical gap for AI, there is still a long way to go

for the large-scale production of “AI chip.” There are high demands to deploy deep neural networks on edge devices. How to implement AI algorithms in the existing hardware terminal equipment has important significance. Therefore, the concept of on-device machine learning (ODML) is put forward in the industry. ODML means that AI algorithms are executed locally on a hardware device and the data (including sensor data or signals) used by the algorithms is also generated by the device

© Xiaochun Guan, Sheng Lou, Han Li and Tinglong Tang. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

This work was supported by the Youth Science Foundation Project of Zhejiang Natural Science Foundation (No. LQ19F020006) and the Teaching Reform Research Project of Wenzhou University (No. 18jg29).

Received 29 October 2020

Revised 30 November 2020

7 January 2021

13 February 2021

Accepted 14 February 2021

The current issue and full text archive of this journal is available on Emerald Insight at: <https://www.emerald.com/insight/0143-991X.htm>

Industrial Robot: the international journal of robotics research and application
48/5 (2021) 700–709
Emerald Publishing Limited [ISSN 0143-991X]
[DOI 10.1108/IR-10-2020-0239]

itself. Google launched Tensorflow Lite, which aims to make it easy for engineers to perform machine learning on devices (Warden and Situnayake, 2019). To better realize ODML, model compression and acceleration have become the research hotspot of the summit. For example, Prof Ji Rongrong's team from Xiamen University has done a lot of excellent work on model compression and acceleration. In (Lin *et al.*, 2020a, 2020b), the authors propose a novel filter pruning method by exploring the high rank of feature maps (HRank), the idea behind HRank is that low-rank feature maps contain less information and the corresponding filters are less important, the experiment results show that HRank method leads to significant improvements over the state-of-the-art in terms of floating point operations and parameters reduction, with similar accuracies. Model compression and acceleration offer a promising prospect to facilitate deploying deep neural networks on edge devices. It plays a great role in promoting ODML. However, existing methods are still challenged by the training inefficiency and labor cost in pruning designs, due to missing theoretical guidance of non-salient network components. In (Lai *et al.*, 2018), the authors presents efficient neural network kernels Cortex Microcontroller Software Interface Standard-Neural Network (CMSIS-NN) for Arm Cortex-M central processing unit (CPU). These kernels are developed to maximize the performance and minimize the memory footprint of neural network applications on Arm Cortex-M processors. It can be used as primitives by machine learning frameworks to deploy trained models on Arm Cortex-M CPUs. Furthermore, A high-level inference neural network package named neural network on microcontroller (NNoM), which is written by Dr Jianjia Ma is designed specifically for microcontrollers. NNoM supports a pure C backend and CMSIS-NN backend. NNoM can provide a light-weight, user-friendly and flexible interface for fast deploying the deep neural networks model on microcontrollers.

Nowadays with the development of the global positioning system (GPS)/Beidou system, these systems can provide absolute position and attitude information for aircraft and further promotes the development of unmanned aerial vehicle (UAV) technology. In recent research, UAV plays an important role for cooperative autonomous localization among air-ground robots in a wide-ranging outdoor industrial environment (Zhang *et al.*, 2019b). At present, the main application fields of UAV include street scene shooting, environmental protection monitoring, power line patrol, post-disaster rescue, agricultural plant protection, police anti-terrorism, military reconnaissance and strike, etc. UAV application technology is typically multi-field integrated innovations, it involves a number of key technologies, such as navigation and positioning technology, detection technology, sensing technology and mission payload (Feng *et al.*, 2017). In recent years, UAV technology has been widely concerned by industry and academia. The quad-rotor UAV is a kind of aircraft that uses rotor rotation to generate lift. The quad-rotor UAV is a platform that can take off and land vertically, hover in midair and fly stably. On the UAV platform, equipment such as aerial photography systems, image transmission systems and positioning systems can be carried out to realize specific flight tasks. For example, in the second quarter of 2020, Dajiang company released an "M300 RTK" aircraft for power industry inspection and a "Mavic Air 2"

aircraft for aerial photography or short video creation. Also, nowadays the quad-rotor aircraft control technology has made great progress. Sensors play an important role in the design of a quad-rotor aircraft (Mack *et al.*, 2019; Zhang *et al.*, 2019c; Zou, 2019).

In the past, most of the quad-rotor aircrafts were controlled by a professional aerial model with personal computer (PC) as the remote controller equipment. At present, Dajiang company adopts a mobile phone app for aerial model remote control. It can be seen that the remote controller of quad-rotor aircraft has the trend of miniaturization, portability. The lift of quad-rotor aircraft comes from the high-speed propeller. When the propeller rotates, it has a tangent angle with the air, which will provide an oblique downward thrust to the air. According to Newton's third law, the air will produce a force perpendicular to the surface of the propeller and the direction is oblique upward. By decomposing the oblique upward force in the yaw direction and the perpendicular direction to the yaw, thus one thrust and one lift can be obtained. The performance of this force is that when the propeller rotates clockwise, the air will provide a reverse rotation torque to the propeller. As shown in Figure 1, when the rotation directions of motor M1 and motor M3 are the same, the rotation direction of M1 is opposite to the rotation directions of M2 and the rotation directions of motor M2 and motor M4 are the same, the reverse torques generated by the four propellers are just offset, so that the resultant force of the four rotors in the yaw direction is zero, thus avoiding the occurrence of quad-rotor's self-rotation. Six different spatial motions can be achieved by adjusting the speed of four motors. The six spatial motions are divided into linear motion and angular motion around an axis. The linear motion includes the forward/backward motion on X_b -axis, the left/right motion on Y_b -axis and the vertical up/down motion on Z_b -axis. The angular motion around an axis includes roll motion around X_b -axis, pitch motion around Y_b -axis and yaw motion around Z_b -axis. As shown in Figure 1, the vertical motion diagram is shown. When the lift to the propeller is $F_1 + \Delta F$, $F_2 + \Delta F$, $F_3 + \Delta F$ and $F_4 + \Delta F$, respectively, there are $4F + 4\Delta F > G$, the aircraft will make a straight-line upward motion toward the negative direction of Z_b -axis;

When the lift to the propeller is $F_1 - \Delta F$, $F_2 - \Delta F$, $F_3 - \Delta F$ and $F_4 - \Delta F$ specifically, there are total lift $4F - 4\Delta F < G$ at this time and the aircraft will make a straight-line downward motion toward the positive direction of Z_b -axis.

Figure 1 Schematic diagram of the lifting movement

Stable flight is the premise for the aircraft to perform specific tasks. Stable flight includes the stability of aircraft attitude and the accuracy of the flight path, so it needs an effective flight control system to achieve. In reference (Zou, 2019; Lu *et al.*, 2016), a cascade proportional integral derivative (PID) control system based on angle and angular velocity is designed. It can achieve the stability of aircraft attitude in flight. On the basis of the cascade PID attitude control loop composed of angle and angular velocity, another cascade PID position control loop based on horizontal position and velocity is added to the system to realize its function with multi-sensor fusion. Altogether a cascade PID control system with an attitude loop as an inner loop and position loop as an outer loop is designed. With the development of the embedded microprocessor, Advanced RISC Machine (ARM) or Digital Signal Processor (DSP) are adopted as the control center of the quad-rotor aircraft (Lv *et al.*, 2014; Wang *et al.*, 2015). In this paper, a quad-rotor aircraft was designed based on ARM microcontroller with multi-sensor fusion. Meanwhile, with the development of AI algorithms, to deploy the deep neural networks model on UAV embedded devices has important application significance. We also propose a method to deploy deep neural networks on the micro controller of the quad-rotor aircraft. One contribution of the work is to provide a reference and guidance to some extent for deploying deep neural networks model on UAV simply, so as to further improve the performance and application range of UAV. The rest of the paper is organized as follows: Section 2, introduces quad-rotor aircraft's hardware framework based on STMicroelectronics' 32-bit microcontroller (STM32). Section 3, introduces quad-rotor aircraft's flight control. Section 4, literates on the deployment of deep neural networks on the quad-rotor aircraft. Section 5, elaborates experiments. Section 6, gives out conclusion remarks.

2. Hardware system of quad-rotor aircraft

The quad-rotor aircraft is a special UAV with four rotor axes. It drives the motors to generate the lift force. By changing the relative speed between different rotors, the value of the single-axis propulsion force can be changed to control the flight trajectory of the aircraft. In this design, the complex quad-rotor aircraft system is divided into six modules, namely, main

controller module, attitude acquisition module, position acquisition module, communication module, motor control module and power module. As shown in Figure 2. The aircraft's physical picture is shown in Figure 5.

The main controller module is for the data processing and command center of the quad-rotor. The attitude acquisition module is equivalent to the human cerebellum and ultimately plays the role of maintaining balance; the position acquisition module is to get the geographical position and altitude position of the aircraft and realize the high hover of the aircraft; the communication module is responsible for man-machine interaction; the motor control module and power module provide the power source for the aircraft and determine the lift and flight time of the aircraft. Some modules will be described in detail below.

2.1 Main controller

The main controller module is responsible for the operation and processing of the data collected by the attitude acquisition module and the position acquisition module, so as to obtain the real-time attitude data and position data of the aircraft. According to the control algorithm, the main controller module is also responsible for outputting the pulse-width modulation (PWM) signal to control the rotation speed of the four motors to achieve attitude correction or flight action. The data acquired by sensors can be sent to the mobile device by the bluetooth communication module, also the mobile device can send the command to the flight. The STM32 microcontroller is adopted in this system. Its working frequency can reach up to 72 MHz and it has strong floating-point data calculation ability. It provides two 12-bit analog-to-digital converters (ADCs), three universal synchronous/asynchronous receiver/transmitter interfaces, two inter-integrated circuits (I²C)/serial peripheral interface at most, three universals 16-bit timers and 1 PWM timer, etc. These peripherals can meet the requirement of the system design.

2.2 Altitude acquisition module

This design adopts the MPU6050 module to collect attitude data. The rotation angular velocity and quaternion attitude data of the aircraft are obtained by the MPU6050 inertial sensor. This sensor integrates a three-axis gyroscope, a three-axis accelerometer and a

Figure 2 Diagram of quad-rotor aircraft hardware system

digital motion processor (DMP). The gyroscope can collect the angular velocity of rotation on three axes. The acceleration can collect the component size of gravity acceleration on the three axes of the body coordinate system, to obtain the angle relationship between the plane of the accelerometer and the ground through the trigonometric function relationship. The MPU6050 module can send the collected data to the main controller by the I²C interface. The DMP in the MPU6050 module is used to filter and fuse the collected attitude data, which can effectively relieve the complex fusion algorithm processing for the main controller. After the data is converted by a 16 bits ADC, the gyroscope and accelerometer, respectively, collect the angular velocity data and gravity acceleration component data on each axis. These original data are used to get the quaternion obtained by DMP hard solution in MPU6050, and then the quaternion is sent to the main controller for processing by the I²C interface. The Euler angle can be get by solving the Euler angle conversion equation as shown in equation (1). In which (q_0, q_1, q_2, q_3) is the quaternion, Φ , θ and Ψ represent the roll angle, pitch angle and yaw angle, respectively, which can be used for attitude control of the flight:

$$\begin{pmatrix} \varphi \\ \theta \\ \psi \end{pmatrix} = \begin{pmatrix} \arctan \frac{2(q_0 q_1 + q_2 q_3)}{1 - 2(q_1^2 + q_2^2)} \\ \arcsin 2(q_0 q_2 + q_1 q_3) \\ \arctan \frac{2(q_0 q_3 + q_1 q_2)}{1 - 2(q_2^2 + q_3^2)} \end{pmatrix} \quad (1)$$

The attitude calculation diagram of MPU6050 is shown in Figure 3.

2.3 Position acquisition module

The NEO-6M module from the UBLOX company is used to obtain the location data of the quad-rotor aircraft in outdoor flight. The module has a small size and high sensitivity, it can be adopted for positioning in the city. In addition to the ceramic antenna, the module also has an Innovative Product development & Engineering solution eXpert (IPEX) interface, which can be connected with an active antenna to improve the speed of satellite searching. NEO-6M module can send positioning data such as altitude and altitude to the main controller by serial port.

In addition to the location of the aircraft, we also need to know the altitude of the aircraft. By continuously monitoring the movement of pixels in the ground image captured by the optical flow sensor, the horizontal displacement data of the aircraft can be measured and then the flight speed of the aircraft is calculated. When the aircraft is flying close to the ground, the altitude of the aircraft is further measured by the HC-SR04

ultrasonic module. The ultrasonic module has high measurement accuracy and small error and its measurement ranging can reach 4.5 m. Although the reflected signal received by the ultrasonic module has certain detection angle requirements, which should be no more than 15° generally. Considering that the attitude angle of the aircraft is limited to $\pm 6^\circ$ during flight, the ultrasonic module is suitable in this design.

To realize the self-stable hovering of the aircraft, the optical flow sensor is used to realize hovering. The optical flow sensor continuously collects the ground image through the camera, monitors the movement of a pixel on the two consecutive images to obtain the trend data of aircraft horizontal displacement, and transmits it to the main controller by the serial port for data processing and control of drift correction.

2.4 Communication module

The bluetooth module is adopted for the aircraft's communication. It has the characteristics of high-cost performance, small size and easy to carry. Bluetooth module can transmit attitude data and positioning data to mobile devices, which is convenient for UAV debugging. At the same time, the mobile app software can be used as the remote-control terminal to send flight commands to the quad-rotor aircraft by bluetooth mode.

2.5 Motor control module

The motor control module includes brushless motors, electric regulators and propellers. Because the flight is a quad-rotor aircraft, there are four motor control modules. Compared with the brush motor, the advantages of the brushless motor are more powerful torque, higher output power and conversion efficiency. The function of the electric regulator is to speed the motor, it can convert the PWM signal into the corresponding speed. The selection of the electric regulator is related to the motor and power supply and mainly depends on the maximum continuous output current index. The maximum continuous output current of the electric regulator should be greater than the maximum peak current of the brushless motor and less than the maximum output current of the battery. The propeller is a component that can directly produce the force and torque required by the quad-rotor motion. Choosing the right propeller can improve the performance of the aircraft. In this design, a 5046 three-blade propeller with a diameter of 5 inches and a pitch of 4.6 inches is selected.

Figure 3 Attitude calculation diagram of MPU6050

3. Flight control of quad-rotor aircraft

The software control system consists of data acquisition processing and cascade PID control. The data collected by the attitude acquisition module and position acquisition module can reflect the accurate motion state of the aircraft, which is also the basis for the motion control of the aircraft. Using the DMP hardware within MPU6050, the filtered quaternion can be obtained and then the quaternion is sent to the main controller, then the Euler angle can be obtained for attitude control by equation (1). This method can effectively reduce the complex fusion algorithm processing of the main controller (Lv, 2018; Sanwale, 2020). When the aircraft is at a fixed height or the inclination angle of the fuselage changes, the change of the view point of the camera on the optical flow sensor will produce the illusion of horizontal movement, thus causing unnecessary control. Therefore, to prevent it, the initial horizontal displacement data obtained are compensated by height compensation and attitude compensation. At the same time, attitude compensation processing is also carried out for the altitude data. Figure 4 shows the flow chart of flight control. In the main function, the interrupt pin of MPU6050 is triggered once every 50 ms by the falling edge signal to executes the flight control program of the aircraft. The priority of the interrupt is set as the lowest to ensure that the interruption of Bluetooth data receiving and sending, optical flow data receiving and

Figure 4 Flow chart of flight control

Figure 5 Aircraft physical picture

ultrasonic data calculation can interrupt this interrupt service. Besides, for the Bluetooth serial port interrupt, it can receive the remote-control command from the mobile app.

PID control algorithm adjusts the flight state of quad-rotors according to the deviation between the expected value and the actual measurement value. In this paper, the discrete position PID algorithm is used to design a single PID loop (Feng *et al.*, 2017). The single closed-loop PID controller is an unstable system for the angle and displacement control of the aircraft while it is a stable system for the diagonal speed control (Shi *et al.*, 2016) and the output of a single PID loop can provide a method to make the measured value approach the target value. Therefore, according to the linearized quad-rotor aircraft control model, a cascade PID control system composed of optical flow position loop, optical flow velocity loop, angle outer loop and angular velocity inner loop is designed. Finally, the PID output of the angular velocity inner loop is used as a part of the PWM assignment to speed the brushless motor.

The optical flow position loop and the optical flow velocity loop constitute a cascade PID controller to control the horizontal displacement of the aircraft. The outer angle loop and the inner angular velocity loop constitute a cascade PID controller to control the flight altitude of the aircraft. At the same time, the attitude control loop and the horizontal position control loop constitute an overall cascade PID controller to control the flight motion of the aircraft. The height loop uses the classical PID controller to control the vertical position. In terms of remote control, the flight status on the horizontal plane

is changed by changing the expected speed value after the motion command is sent by the mobile app.

4. Deployment of deep neural networks on STMicroelectronics' 32-bit microcontroller based on RT-Thread

Real Time-Thread (RT-Thread) is an embedded real-time multi-threaded operating system supporting multi-thread scheduling. It is also an IoT operating system, based on the operating system kernel it includes complete middleware components such as file system and graphics library, and a software platform with low power consumption, security, communication protocol support and cloud connectivity. The RT-Thread operating system is a completely open source, mainly written in C language, easy to understand and easy to transplant. It applies the object-oriented design method to the real-time operating system design, which makes the code style elegant, the architecture clear and the system modular. It has very good tailorability when the 32-bit microcontroller is as its main operating platform. For resource constrained microcontroller system, it can cut out the nano version which only needs 3 KB flash and 1.2 KB RAM by using conventional tools. For the resource rich IoT devices, RT-Thread can use online software package management tools and system configuration tools to realize intuitive and rapid modular cutting, seamless import of rich software packages and realize complex functions such as android graphical interface, touch sliding effect, intelligent voice interaction effect. To deploy deep neural networks on STM32, we can use a high-level inference Neural Network library NNoM, which is available from RT-Thread online packages.

NNoM is a high-level inference Neural Network library written by Dr Jianjia Ma, it is designed specifically for microcontrollers (Ma, 2019). Nowadays, neural networks are wider, deeper and denser. NNoM can provide a light-weight, user-friendly and flexible interface for fast deploying. NNoM uses a layer-based structure, the model structure can seem directly from the codes. It can make the model conversion from other layer-based libs (Keras, TensorLayer, Caffe) to the NNoM model very straight forward. Currently, NNoM supports a pure C backend and CMSIS-NN backend. The model will be quantized through model conversion. Its highlights are as follows: it can convert Keras model to NNoM model with one line of code; it supports complex structures; such as Inception, ResNet, DenseNet and Octave Convolution; it has user-friendly interfaces. It supports onboard evaluation tools, such as runtime analysis, Top-k and confusion matrix.

5. Experiment results

5.1 Flight control of quad-rotor aircraft

First, PID parameters of the altitude control loop are tuned without adding a horizontal position control loop. With the help of two strings, the aircraft is fixed in the air and kept relatively stable, so that it can only rotate around the rope. At this time, the rope coincides with the horizontal axis of the aircraft body coordinate. By bluetooth mode, the angle data is sent to the bluetooth serial port debugging assistant in the PC and after the PID parameters of the pitch direction and roll

direction are tuned, the PID control result chart of the attitude loop in pitch direction is shown in Figure 6. At this time, after taking off, the aircraft can keep its attitude stable, but it cannot keep hovering.

After adding the horizontal position control loop and tuning its PID parameters, the horizontal displacement data of the aircraft are tested after taking off. Figure 7 shows the horizontal displacement results based on the optical flow sensor after the aircraft just takes off. In Figure 7, the blue waveform is the change of the displacement data along with the X-axis direction of the aircraft, and the purple waveform is the change of the displacement data along with the Y-axis direction of the aircraft. It is seen from the figure that the displacement was just zero when the flight is just taken off, then the aircraft with $X_Flow > 0$ deviated more than 2 cm to the X-axis. The aircraft with $Y_Flow > 0$ also deviated more than 2 cm toward the Y-axis. After the adjustment of cascade PID control, there is $X_Flow < 0$ and $Y_Flow < 0$, then the aircraft gradually reaches a hovering state. For Figures 6 and 7, the unit of ordinate is centimeter and the unit of abscissa is second.

For the quad-rotor aircraft, the main controller module is equivalent to the human brain and it is the data processing and command center. The attitude acquisition module is equivalent to the human cerebellum, which ultimately plays the role of maintaining balance; the position acquisition module is to determine the geographical position and altitude position of the aircraft and realize the hovering of the aircraft at the same time; the communication module is for human-computer interaction; the motor control module and power supply module provide the power source for the aircraft and determine the lift size and flight time of the aircraft. Based on the data obtained by each sensor, a cascade PID control system with position control as the outer loop and attitude control as the inner loop is designed. In the future, we can use more advanced algorithms to make better use of the data obtained by sensors, so as to achieve better and stable flight missions.

5.2 Deployment of deep neural networks on STMicroelectronics' 32-bit microcontroller

Handwritten digit classification was used as an example to demonstrate the steps of deployment of deep neural networks on STM32 based on the NNoM software package, the Mixed National Institute of Standards and Technology database (MNIST) is selected to train the model. First, RT-Thread Studio is used to build the project engineering framework, configure the kernel and components, and download the software package required by the project and then carry out application development on it. RT-Thread studio is a one-stop RT-Thread IoT operating system development tool. It makes the development of IoT becomes simple and efficient by simple and easy-to-use graphical configuration system and rich software packages and component resources. The packages tool of RT-Thread settings within RT-Thread studio can realize the configuration and tailoring of the whole system.

At first, a deep neural network model can be built and trained by using the MNIST data set based on Tensorflow framework, and then the model structure, weight and bias of the pre-trained model can be stored in open neural network exchange (onnx) data format. The onnx data format is in accordance

Figure 6 PID control process of attitude loop in the pitch direction **Figure 7** Example of horizontal displacement adjustment based on the optical flow sensor

with Google protobuf protocol, which is also supported in RT-Thread.

Before to deploy and run the onnx data format model on a microcontroller based on RT-Thread, the NNoM software package, the protobuf-c software package and the onnx-parser software package should be added and loaded in the RT-

Thread operating system and the model data in onnx data format should be stored into the Secure Digital (SD) card of the microcontroller's development board. Then, the project can be further created based on RT-Thread studio. The trained onnx model data can be used to automatically build the model and load the model parameters based on the application

programming interface (API) in the NNoM software package to realize the network deployment on the embedded device.

Serial port/network data debugger SSCOM V5.13.1 is adopted to demonstrate the result, STMicroelectronics

STLink virtual com port is selected, its setting is 115200 baudrate, 8 bits data width and 1 stop bit. In Figure 8, it first showed the initialization information of the RT-Thread operating system startup, then it printed the structure

Figure 8 Printed structure of the pre-trained deep neural networks deployed on microcontrollers

```

 \ \ / - RT - Thread Operating System / / \ 3.1.0 build Nov 30 2020
2006 - 2018 Copyright by rt-thread team
[SFUD] Find a Winbond flash chip. Size is 16777216 bytes.
[SFUD] w25q128 flash device is initialize success.
msh />
NNOM version 0.2.1
Start compiling model...

```

Layer(#)	Activation	output shape	ops(MAC)	mem(in, out, buf)	mem blk lifetime
#1 Input	-	(28, 28, 1)	(784, 784, 0)	1 - - - - -	
#2 Conv2D	- ReLU	(28, 28, 12)	84k (784, 9408, 36)	1 1 1 - - - -	
#3 MaxPool	-	(14, 14, 12)	(9408, 2352, 0)	1 1 1 - - - -	
#4 Conv2D	- ReLU	(14, 14, 24)	508k (2352, 4704, 432)	1 1 1 - - - -	
#5 MaxPool	-	(7, 7, 24)	(4704, 1176, 0)	1 1 1 - - - -	
#6 Conv2D	- ReLU	(7, 7, 48)	508k (1176, 2352, 864)	1 1 1 - - - -	
#7 MaxPool	-	(4, 4, 48)	(2352, 768, 0)	1 1 1 - - - -	
#8 Dense	- ReLU	(96, 1, 1)	73k (768, 96, 1536)	1 1 1 - - - -	
#9 Dense	-	(10, 1, 1)	960 (96, 10, 192)	1 1 1 - - - -	
#10 Softmax	-	(10, 1, 1)	(10, 10, 0)	1 1 - - - - -	
#11 Output	-	(10, 1, 1)	(10, 10, 0)	1 - - - - -	

```

Memory cost by each block:
blk_0:1536 blk_1:2352 blk_2:9408 blk_3:0 blk_4:0 blk_5:0 blk_6:0 blk_7:0
Total memory cost by network buffers: 13296 bytes

```

Figure 9 One prediction example by the model deployed on the microcontroller

```

Time: 478 tick
Truth label: 5
Predicted label: 5
Probability: 100%
msh />

```


information of the pre-trained deep neural network model by calling function `NNoM_model_create()` in `NNoM`, in the end it displayed the memory occupation information of each layer of the model and the total network memory occupation is 13,296 bytes.

Figure 9 shows one example of prediction based on MNIST using the model deployed on the microcontroller. The tested original handwritten digit is random selected from MNIST, its size is 28×28 . It is roughly represented by ASCII code “” and “.” for better visibility. It is shown that the deep neural network deployed on the microcontroller can accurately predict this handwritten digit.

In this part, we give out the steps to deploy the deep neural networks on STM32 based on an AI-related package. It is meaningful for the low power and low performance machine vision market in the future. Just as mentioned in (Lin et al., 2020a, 2020b), the era of always-on tiny machine learning on IoT devices has arrived. The number of devices based on always-on microcontrollers is increasingly rapidly at a historical rate. We can deploy other AI models on STM32 in the future based on new neural network development kits such as STM32Cube.AI and RT-AK, which can empower the development of edge AI and endpoint AI, to implement low performance machine vision application such as simple speech recognition, gesture recognition, person detection, colored object recognition on IoT devices. It can broaden the application scenarios of UAV and enhance the application functions of UAVs.

6. Conclusion

This paper provides powerful hardware support for the design of a simple quad-rotor UAV based on the rich resources of the STM32 microcontroller and the fusion of multiple hardware sensor modules. Combined with the development of AI, the deployment of deep neural networks on the controller of UAV is also studied. The flight mission control of quad-rotor aircraft is realized based on the multi-sensor fusion method. The system integrates sensor modules such as attitude acquisition module, GPS position acquisition module, optical flow sensor, ultrasonic sensor and bluetooth communication module. These sensors are used to obtain the attitude data, positioning data and altitude data of the quad-rotor aircraft. Based on the linearized flight control model, the cascade PID control system with horizontal position control as the outer loop and attitude control as the inner loop is designed and implemented based on the data fusion of multi-sensor. The actual flight test of the aircraft shows that it can achieve hover control and flight speed adjustment of the aircraft and realize the mobile device terminal's control by bluetooth mode. This paper can provide a simple design scheme for the future quad-rotor aircraft remote control mode. This work can provide a simple reference and guidance for deploying deep neural networks model on UAV, it is helpful to improve the performance and application range of UAV. The deployment of deep neural networks based on NNoM provides a good reference scheme for the development of machine learning algorithm on microcontrollers. In future

work, we will do more on the deployment of compressed deep neural network models on microcontrollers.

References

- Feng, X.Y., Fan, H.G. and Xin, L. (2017), *Design of Quadrotor Unmanned Aerial Vehicle*, Tsinghua University Press.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014), “Generative adversarial networks”, *NIPS*, pp. 2672-2680.
- He, K.M., Zhang, X.Y., Ren, S.Q. and Sun, J. (2016), “Deep residual learning for image recognition”, *IEEE Conference on Computer Vision Pattern Recognition*.
- Lai, L.Z., Suda, N. and Chandra, V. (2018), “CMSIS-NN: efficient neural network kernels for arm Cortex-M CPUs”, available at: <https://arxiv.org/pdf/1801.06601.pdf> (accessed 10 August 2020).
- Lin, J., Chen, W.M., Lin, Y., Cohn, J., Gan, C. and Han, S. (2020a), “MCUNet: tiny deep learning on IoT devices”, *NIPS*, available at: <https://arxiv.org/abs/2007.10319> (accessed 10 August 2020).
- Lin, M.B., Ji, R.R., Wang, Y., Zhang, Y.C., Zhang, B.C., Tian, Y.H. and Ling, S. (2020b), “HRank: filter pruning using high-rank feature map”, *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1526-1535.
- Lu, S., Xu, D., Ouyang, W.L. and Wang, X.G. (2019), “Unsupervised collaborative learning of keyframe detection and visual odometry towards monocular deep SLAM”, *IEEE/CVF International Conference on Computer Vision*, pp. 4301-4310.
- Lu, J.Y., Zeng, X.Y., Wang, J.Y. and Wang, Y. (2016), “Design and implementation of dual loop PID controller for quadrotor aircraft”, *Journal of Hebei University of Technology*, Vol. 45 No. 5, pp. 16-22.
- Lv, Y.H. (2018), “Design and algorithm research of autonomous flight four rotor system based on STM32 MCU”, *Liaoning University of Engineering and Technology*, pp. 33-42.
- Lv, Q., Ni, P.P., Wang, G.S. and Liu, F. (2014), “Design of quadrotor's control system based on DSP”, *Advanced Materials Research*, Vol. 989-994, pp. 3216-3219.
- Ma, J.J. (2019), “Neural network on microcontroller (NNoM). [online]”, available at: <http://packages.RT-Thread.org/detail.html?package=nnom> (accessed 10 August 2020).
- Mack, B., Noe, C., Rice, T., Ahn, I.S. and Wang, J. (2019), “Distributed localization and control of quadrotor UAVs using ultra-wideband sensors”, *IEEE Aerospace Conference*.
- Sanwale, J., Trivedi, P., Kothari, M. and Malagaudanavar, A. (2020), “Quaternion-based position control of a quadrotor unmanned aerial vehicle using robust nonlinear third-order sliding mode control with disturbance cancellation”, *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, Vol. 234 No. 4, pp. 997-1013.
- Shi, C.X., Hu, X.L., Jiang, S.C., Zhang, B.Q., Gao, J.Q. and Liu, Y.H. (2016), “Research on Cascade PID control system of quadrotor UAV”, *Journal of Binzhou University*, Vol. 32 No. 4, pp. 10-16.

- Wang, M.S., Chen, J. and Dai, J.S. (2015), "Design of quadrotor control system based on arm processor", *Instrumentation Technology*, Vol. 8, pp. 18-21.
- Warden, P. and Situnayake, D. (2019), *TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers*, O'Reilly.
- Zhang, L., Ma, Y.J., Liu, G.Q. and Yu, Y. (2019c), "Multi-sensor fusion quadrotor attitude and altitude estimation", *Proceedings of the 2019 International Conference on Robotics Systems and Vehicle Technology*, pp. 131-137.
- Zhang, J.H., Gui, M.P., Wang, Q.C., Liu, R.Y., Xu, J.Z. and Chen, S.Y. (2019a), "Hierarchical topic model based object association

- for semantic SLAM", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 25 No. 11, pp. 3052-3062.
- Zhang, J.H., Liu, R.Y., Yin, K.J., Wang, Z.Y., Gui, M.P. and Chen, S.Y. (2019b), "Intelligent collaborative localization among air-ground robots for industrial environment perception", *IEEE Transactions on Industrial Electronics*, Vol. 66 No. 12, pp. 9673-9681.
- Zou, Z.T. (2019), "Research on flight control system of quadrotor UAV", *Qingdao University*, pp. 31-35.

Corresponding author

Xiaochun Guan can be contacted at: guanxc@wzu.edu.cn