ECE 445

SENIOR DESIGN LABORATORY

FINAL REPORT

Semantic Communications for Unmanned Aerial Vehicles

Team #25

YU LIU (yul9@illinois.edu)
CHANG SU (changsu4@illinis.edu)
CHENHAO LI (cl89@illinis.edu)
TIANZE DU (tianzed2@illinis.edu)

Sponsor: Meng Zhang
TA: Xiaoyue Li

May 10, 2023

Contents

1 Introduction

1.1 Purpose: Problem and Solution

Existing communication systems are mainly based on Shannon's information theory, and they are mostly developed to maximize data-oriented performance indicators, such as communication data rate, while ignoring content-related information or considering only upper-level information [1]. In the process of transmission, once the noise makes some bits in the process of transmission wrong, the meaning of the transmission result will become ambiguous [2]. This not only increases the demand for communication resources but also limits the transmission rate of information.

In this case, people start to think about using semantic communication. Semantic communication breaks through the traditional theoretical framework of Shannon's information theory, making great breakthroughs in reducing communication loss, improving transmission rate and accuracy, and transforming the content of communication into the meaning of information more valuable to human beings, thus fundamentally transforming the existing communication architecture into a more universal intelligent and human-oriented system [3]. In semantic communication, the meaning of the result can be roughly predicted even if some bits are disturbed by interference.

And now, most of the existing image semantic communication technologies rely on the direct transmission of the whole image between the sender and the receiver [4]. The transmission process rather than the semantic understanding algorithm is the bottleneck of its performance. In real life, there are many scenarios that require the unmanned aerial vehicle (UAV)'s overlooking function and the UAV's direct real-time communication with other intelligent devices.

The UAVs currently on the market can only fly and take pictures, and transmit the pictures to the receiver using traditional means of communication [5]. But in many cases, the direct transmission of images from UAV is a huge waste of power and transmission. So, our goal is to develop a UAV technology that allows the UAV to transmit images using semantic communication. More specifically, our UAV can build on the capabilities of existing UAV to process a sample of the image taken, extract specific semantics, and convey its symbolic representation to the target receiver (for example, another UAV or smart device). In this way, all we need to transmit is a sentence instead of a whole picture. We hope this technique will be much faster than transmitting each complete image directly.

Firstly, we'll assemble a powered UAV, complete with batteries and controls, as well as a camera, a transmission module, and Raspberry PI. The controller will control the four propellers used to control the UAV's movement. The controller ensures that the UAV has enough power to carry all of its equipment and is always balanced. During the flight of the UAV, we use the camera of the UAV to shoot images, use the Raspberry PI to process images, and extract the semantic features in the images. Then the encoder encodes these features into digital signals and transmits them to the receiver. The receiver is another smart device, such as a smartphone or computer. The digital signal is transmitted through a physical channel, mainly through a WiFi transmission module. Finally, the receiver's decoder can translate the bits back into semantic information to make sense of the message. We will use a display screen to display the message received by the receiver.

1.2 Functionality

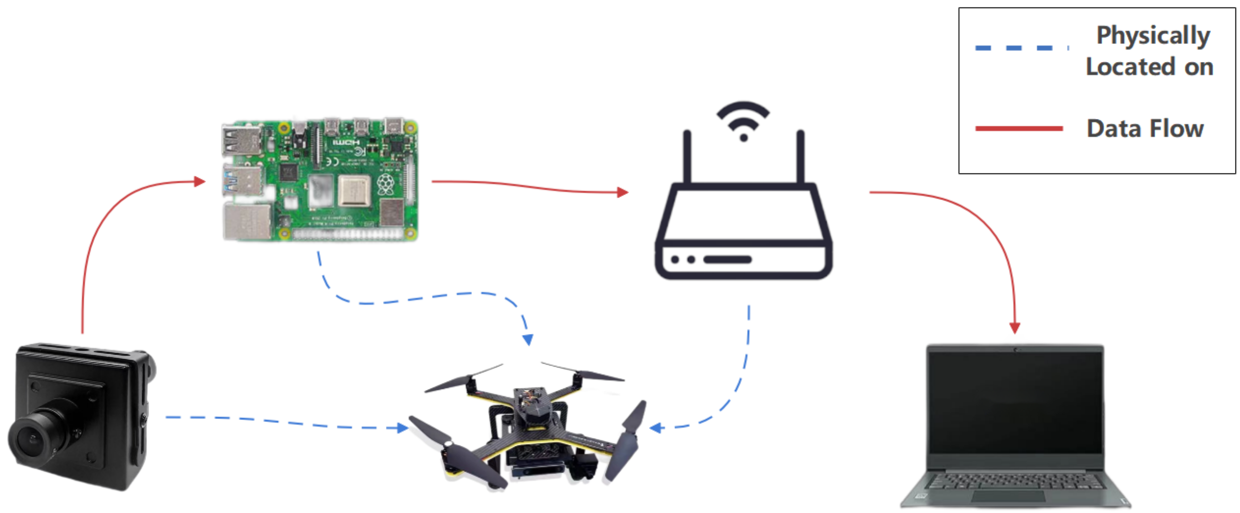

Figure 1: Visual Aid: Data Flow and Physical Location

The UAV carries cameras and microcomputers, such as the Raspberry PI, to move around. The UAV hovers up to at least 5 meters in the air and the camera can take clear photo. The microcomputer understand the video taken by camera and extract useful information (semantics), especially the types of their behavior. Finally, the WIFI chip will must transmit semantic information to the receiver successfully. The computer will show the finally semantic information.

1.3 Subsystem Overview

First, here is our block diagram.

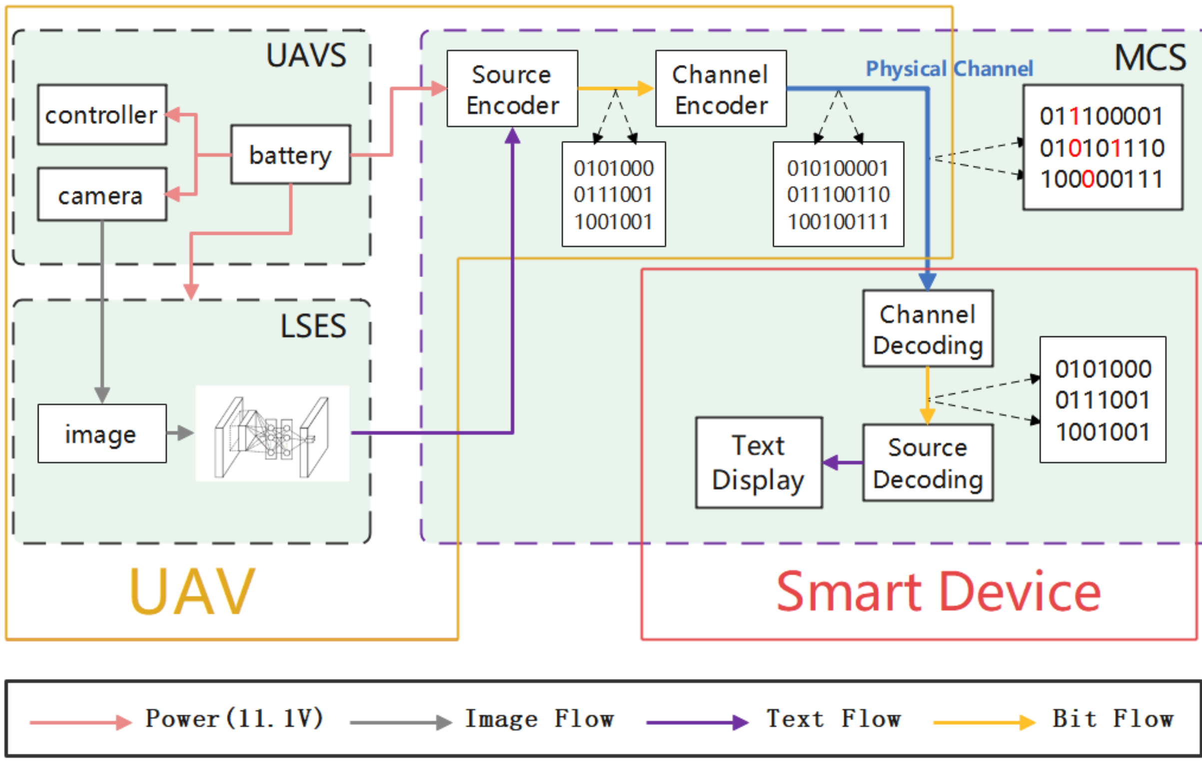

Figure 2: System Block Diagram

1.3.1 UAV Subsystem (UAVS)

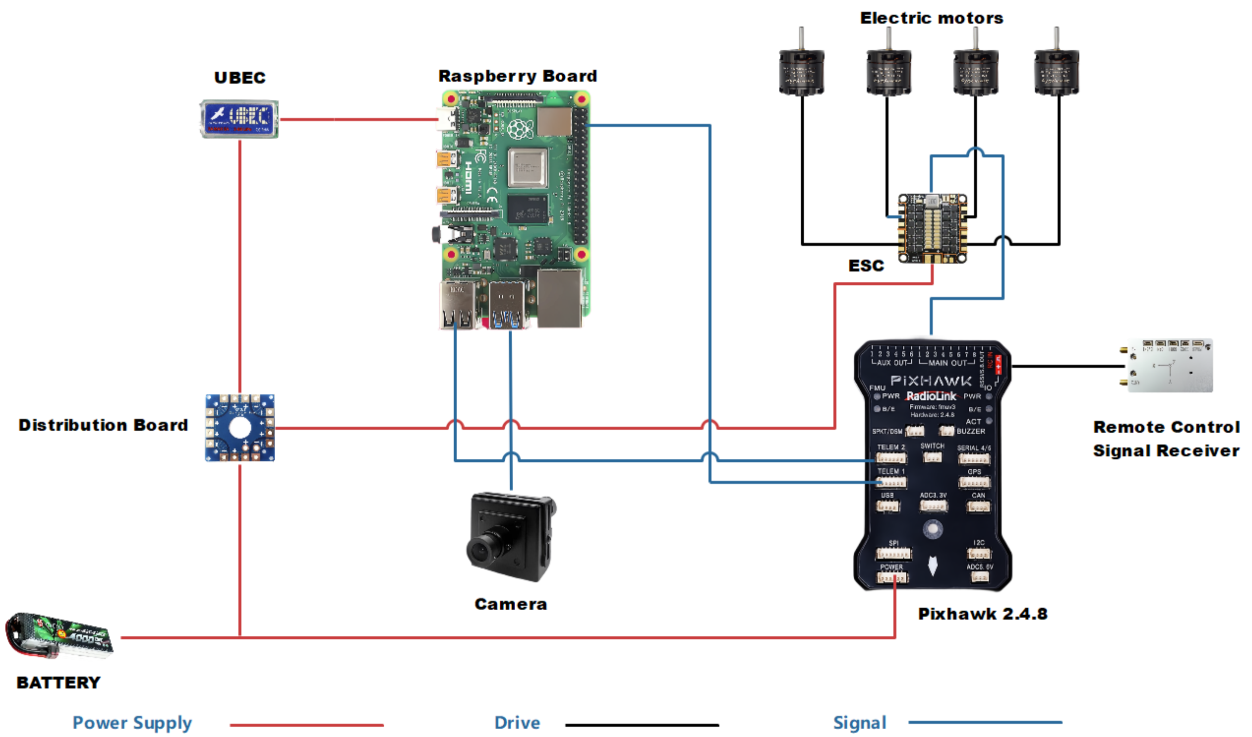

Figure 3: Structure of UAV

The UAV subsystem includes a power module, a controller module, and a camera as shown in figure 3. The power module, which includes a distributor board and a lithium battery, is to supply stable voltage for every device and other subsystems on UAV. The controller module, PIXHAWK, will receive signal and control the four propellers, which are used to control the UAV's movement. The camera will take images, which will be used as input data for lighting semantic extraction subsystems (LSES).

1.3.2 Lighting Semantic Extraction Subsystem (LSES)

Given an video as input, LSES can be designed to analyze the video and generate a descriptive sentence within the scene. LSES could utilize advanced computer vision algorithms and machine learning techniques to accurately detect and identify the people and their actions in the image, providing valuable insights for a range of applications. Our project will focus on some specific domains, especially basketball game at the gym. The semantic information extracted by LSES will serve as the input of the mutual communication subsystem (MCS).

1.3.3 Mutual Communication Subsystem (MCS)

MCS accepts the text information extracted from images by LSES. This subsystem converts text into a bits signal and transmits it to another smart device over a physical channel. MCS consists of two separate parts: the transmitter on UAV and the receivers on smart devices, for example a computer, which are connected by the physical channel. The subsystem includes a source encoder and a source decoder. And finally, the text information will be displayed on a screen. The text has similar semantic information but omits unnecessary words. The text example could be like “there is a player shooting the ball”, and the result could be ”quotafree commentary player shooting quotafree ball quotafree” and can be summary as ”player shooting ball”. Communication should be quick and without losing semantic information.

2 Design

2.1 Equations & Simulations

Besides, the semantic channel capacity of a discrete memoryless channel [4] is expressed as

$$C_s = \sup_{p(Z|X)} \left\{ I\left(X;V\right) - H\left(Z|\,X\right) + \overline{H_S\left(V\right)} \right\},$$

Here $I(X;V)$ is the mutual information between the source, X , and the transmission task, V. Here $p(Z|X)$ is the conditional probabilistic distribution that refers to a semantic coding strategy with the source, X, encoded into its semantic representation, Z, and $H(Z|X)$ means the semantic ambiguity of the coding. $\overline{H_S(V)}$ is the average logical information of the received messages for the task V. Then here we can see that if we could make $\overline{H_S(V)}$ be bigger than $H(Z|X)$, the semantic channel capacity could be always bigger than 0. That means the receiver can handle the semantic ambiguity. For our design, this is easy to accomplish.

The following is the simulation result of our LSES on the computer. This is the accuracy of the trained model in the training set, test set and verification set.

```
100%|
[train] Epoch: 27/210 Loss: 0.003548036794581864 Acc: 1.0
Execution time: 46.642165508819744

100%|
[val] Epoch: 27/210 Loss: 0.2100410066358217З Acc: 0.9523809523809523
Execution time: 3.6507324730046093

100%|
[train] Epoch: 28/210 Loss: 0.0007815782620493614 Acc: 1.0
Execution time: 46.21627769409679

100%|
[val] Epoch: 28/210 Loss: 0.1439463860113404 Acc: 0.9523809523809523
Execution time: 3.761741875903681

100%|
[train] Epoch: 29/210 Loss: 0.0034656953346459468 Acc: 1.0
Execution time: 46.405119572998956

100%|
[val] Epoch: 29/210 Loss: 0.17476579843249448 Acc: 0.9206349206349206
Execution time: 3.660524234175682

100%|
[train] Epoch: 30/210 Loss: 0.0010915619829783816 Acc: 1.0
Execution time: 47.097705852938816

100%|
[val] Epoch: 30/210 Loss: 0.2448815985508039 Acc: 0.9365079365079365
Execution time: 3.523531877901405

100%|
[test] Epoch: 1/210 Loss: 0.0840309256134009 Acc: 0.9735099337748344
Execution time: 7.94139563315548
```

Figure 4: The result of LSES

2.2 Design Alternatives

Since the design document, under the suggestion of the course instructor, we change our application scenario. We will choose a single kind of sport, basketball, and recognize some specific action in the game, for example, passing, laying up, shooting. This change makes our project much more challenging, since we need to analysis video instead of image to understand semantic meaning. We need to change our network. The YOLO is not a good network deal with video. More importantly, we cannot any suitable dataset. Thus, we need to collect and label the dataset by ourselves, which has huge amount of work.

2.3 Design Description and Justification

Figure 5: Initial UAVS

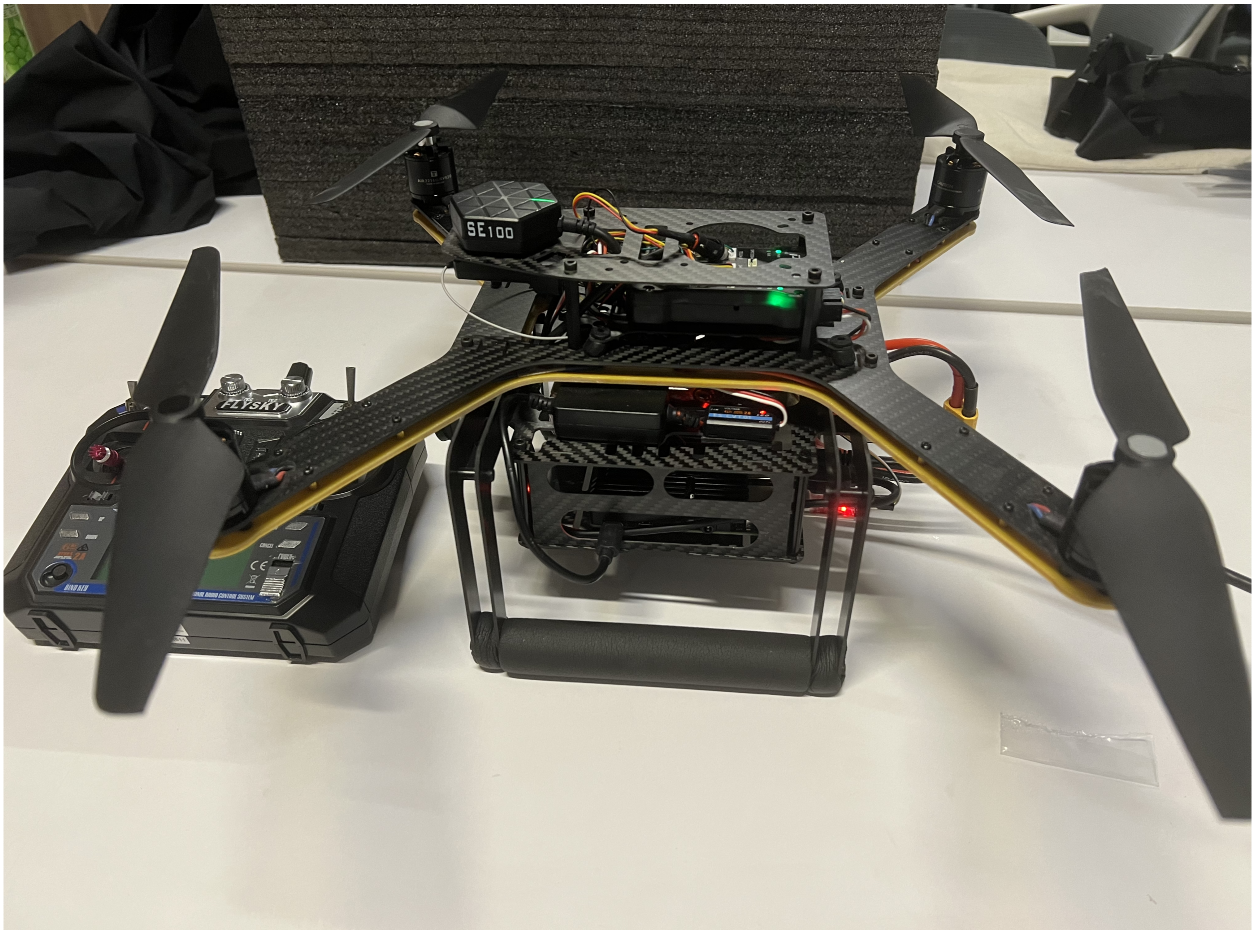

Figure 6: Modified UAVS

2.3.1 UAV mechanical, balance and dynamic Subsystem (UAVS)

We modified a dumb drone, which already has a power module, a controller module as shown in figure 5. Then, we add a camera, a power distributor and a Raspberry Pi on the UAV, which is shown in figure 6. It is also important to ensure that the UAV has enough power to carry all the devices and keep them balanced all the time. The camera will take images with high resolution, which will be used as input data for lighting semantic extraction subsystems (LSES). We modified a dumb UAV. Equipment likes, a camera, Raspberry Pi, and a communication module was added.

2.3.2 Lighting Semantic Extraction Subsystems (LSES)

The model used in our project is 3D convolutional network(C3D network) based on the paper Learning Spatiotemporal Features with 3D Convolutional Networks. This paper proposes an efficient C3D network to extract the spatial-temporal features of videos and extract features from time series better. The usage scenario of C3D network is to do action recognition, which is very similar to our purpose. In particular, it was found that the convolution kernel of 3*3*3 had the best effect and had a revelatory effect.

The architecture of C3D network is like this:

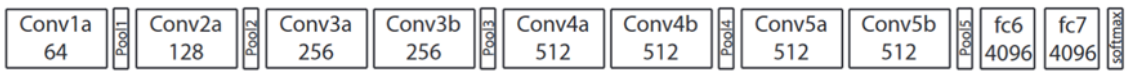

Figure 3. **C3D architecture**. C3D net has 8 convolution, 5 max-pooling, and 2 fully connected layers, followed by a softmax output layer. All 3D convolution kernels are $3 \times 3 \times 3$ with stride 1 in both spatial and temporal dimensions. Number of filters are denoted in each box. The 3D pooling layers are denoted from `pool1` to `pool5`. All pooling kernels are $2 \times 2 \times 2$, except for `pool1` is $1 \times 2 \times 2$. Each fully connected layer has 4096 output units.

Figure 7: Structure of LSES network

From this figure we can see that the network takes the edited video clips as input, and the resolution of all videos is adjusted to 128*171. The video is also split into 16 frames that do not overlap each other and are used as network input. The network has 5 convolutional layers and 5 pooling layers, two fully connected layers and a softmax loss function layer to predict action labels. All convolutional layers have suitable padding and stride to ensure that the input to output of convolutional layers does not change in size. The pooling layer is processed by 2*2*2 kernels to reduce the output by a factor of 8. The least gradient algorithm has a starting learning rate of 0.003.

For our project, I modified the fully connected layers output slim to the number of our labels. To be specific, that is 3. Besides, I changed the probability of the dropout from 0.5 to 0.3 to make the fitting of the model better. Since we are running the code on Raspberry PI, we want to minimize the model size and computational time. Therefore, here I delete two layers of convolutional networks that do not change the output dimension. The reason is that the network accuracy will not decrease much after deleting, but the operation time can be greatly reduced. Here I did the testing, using both the modified network and the pre-modified network on the UCF-101 dataset. The result is that after 20 epochs, the modified network accuracy rate is 76.5%, and the average running time per epoch is 10 minutes and 26 seconds; the pre-modified network accuracy rate is 80.2%, and the average running time per epoch is 13 minutes and 35 seconds. As you can see, although the accuracy doesn't decrease much, the computation time decreases a lot.

2.3.3 Mutual Communication Subsystem (MCS)

MCS accepts the text information extracted from images by LSES. MCS consists of two separate parts: the transmitter on UAV and the receivers on smart devices, which are connected by the physical channel. The transmitter includes a semantic encoder, a semantic decoder.

Driven by deep learning, natural language processing (NLP) has achieved great success in analyzing and understanding large volumes of linguistic text. we also try to use a new perspective for the communication system from the semantic level, and proposes a semantic transmission system based on deep learning, that is DeepSC, for text transfer. On the basis of Transformer, the goal of DeepSC is to minimize semantic errors, restore sentence meaning, not traditional bit or symbol error communication.

The structure of DeepSC is shown in figure 8.

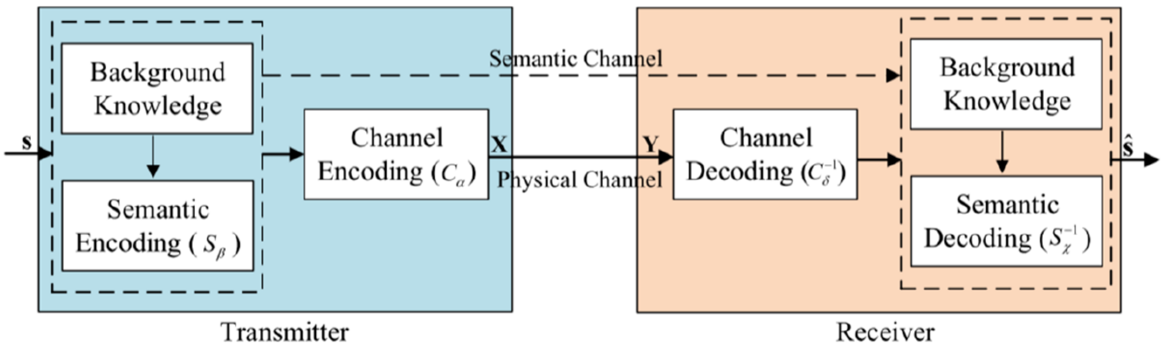

Figure 8: Structure of DeepSC network

2.4 Subsystem Diagrams Schematics

3 Cost and Schedule

3.1 Cost

First, for our labor cost, we assume everybody's hourly wage is ¥100/hour, and we need to work for 10 hours/week for all four people. And we need to do this for the following 10 weeks this semester. So for this part, our fixed development cost is :

$$4 \cdot \frac{20CNY}{hr} \cdot \frac{10hr}{wk} \cdot 10wks \cdot 2.5 = 20000CNY$$

Then, since only one person is needed to operate the drone, we don't need a lot of bulk. For the parts and manufacturing prototype costs, it's estimated ts ¥2946 each:

Part	Vendor	Cost (prototype)	Cost (bulk)
Professional aerial photography UAV (CK10pro)	Taobao	¥1888	¥50
8GB Raspberry Pi (4B; generic)	Taobao	¥728	¥20
200W pixels Monocular Camera (Reshi Technology)	Taobao	¥150	¥20
WiFi module (Small R Technology; MT7620)	Taobao	¥50	¥40
Total	Taobao	¥2816	¥130

Then we add two parts together, our total development cost should be ¥22946.

3.2 Schedule

Week	Yu Liu	Chenhao Li	Chang Su	Tianze Du
3/20/23	Write Design Document 2.1 and 2.2	Write Design Document 2.3 and 2.4	Write Design Document Part 1 and 3	Write Design Document Part 4
3/27/23	Learn the use of the Raspberry Pi	Look for object detection algorithms	Find the appropriate dataset	Purchase the required parts
4/3/23	Simple programming on the Raspberry Pi	Run through object detection algorithms on the computer	Find the appropriate semantic segmentation algorithm	Add parts on the UAV
4/10/23	Run object detection algorithms on the Raspberry Pi	Find the right means of communication for UAV	Run the semantic segmentation algorithm on the computer	Design and construction of UAV balancing systems

4/17/23	Enable communication between drones and other smart devices	Enable communication between drones and other smart devices	Implement the semantic segmentation algorithm on the Raspberry Pi	Design and construction of drone power systems
4/24/23	Carry out the final inspection of the part for which he is responsible	Carry out the final inspection of the part for which he is responsible	Carry out the final inspection of the part for which he is responsible	Carry out the final inspection of the part for which he is responsible
5/1/23	Test flights of UAV, detection and analysis of errors	Test flights of UAV, detection and analysis of errors	Test flights of UAV, detection and analysis of errors	Test flights of UAV, detection and analysis of errors
5/8/23	Prepare for Mock demo	Prepare for Mock demo	Write the Final Report draft	Write the Final Report draft
5/15/23	Detect the overall effectiveness of the project	Detect Lighting Semantic Extraction Subsystems	Detect Lighting Semantic Extraction Subsystems	Detect UAV mechanical, balance and dynamic Subsystem
5/22/23	Write Final Report	Write Final Report	Prepare for Final Presentation	Prepare for Functionality Demonstration Video

4 Requirements and Verification

4.1 UAV mechanical, balance and dynamic Subsystem (UAVS)

4.2 Lighting Semantic Extraction Subsystems (LSES)

4.2.1 Completeness of Requirements

The requirements for LSES subsystem are that it could identify the actions of players on the basketball court with an accuracy of more than 80%. The accuracy value should range from 70 to 90. Besides, the running time of the model should also be small enough. Our expectation is to process a video in 5 seconds.

4.2.2 Appropriate Verification Procedures

First, we did the training model part on the computer. The model divides the dataset taken by us into three parts: training set, test set and verification set. At this stage, we will try our best to adjust the model parameters to make the accuracy of the model on the training set as high as possible and pay attention to the accuracy of the model on the test set and the verification set.

Subsequently, we wrote the inference script to test the validity of the model by using the video file. We selected video files of some test sets and validation sets. We tested the validity of this file on the computer first.

Then we put the script on Raspberry PI and used the drone to shoot video for motion recognition. We will record the accuracy of our predictions.

4.2.3 Quantitative Results

The final completeness of our requirements is very exciting. As the Figure 4 shows, after training, the accuracy of our C3D network can reach 1.0 on the training set, and more than 90% on the verification set and test set.

The accuracy of the predicted results of script can reach more than 90%. On the NVIDIA GeForce RTX 3090 graphics card, our script running time was about 2 seconds per video. This processing speed is capable of processing drone images in real time and predicting results.

On Raspberry PI, the prediction accuracy of the model is roughly the same as that of the computer, at more than 90%, as long as the drone does not shake badly due to air currents, causing the picture to shake or be unclear. However, the Raspberry PI GPU model is VideoCore VI. Due to the limited hardware conditions of Raspberry PI, it takes about 7 seconds to run a script file on Raspberry PI. That speed makes it impossible for PI to process video in real time.

4.3 Mutual Communication Subsystem (MCS)

4.3.1 Completeness of Requirements

The requirements for MCS subsystem are that it could transmit the text output of the LSES subsystem accurately and efficiently. The transmission time should be less than 1 second, and the semantics during the transmission process should not change significantly. We can evaluate the performance of our network by BLEU score.

4.3.2 Appropriate Verification Procedures

We divide the dataset into three parts: training set, test set and verification set. In our test, we will use our test data, which will not be used in our training, to test the model's performance under different epoch. Meanwhile, I will test some sentences used in basketball games, for example, the player shoots the ball to test it on Raspberry PI.

4.3.3 Quantitative Results

The final result of this subsystem is not bad. After training, the BLEU of our DeepSC network can reach 80% on the test dataset, and 70% for our whole system. The result for our test under different epoch is shown in figure

5 Conclusion

In short, we have completed all subsystems well and reached the pre-set high-level requirement:

- The UAV must be able to carry cameras and microcomputers, such as the Raspberry PI, to move around. The UAV can hover up to at least 5 meters in the air and take sample video.
- The UAV must understand its images and extract useful information (semantics), especially the number of people and even the types of their behavior. The UAV should be able to predict the number of people with 80% accuracy and predict the types of behavior with 70% accuracy.
- The UAV must transmit semantic information to the receiver successfully. The time required to transfer each picture should be less than 1s.

5.1 Uncertainties and Future Work

5.2 Ethical Considerations

A number of potential ethical and safety issues had to be considered in our project. First of all, both the UAV and the Raspberry Pi board need to be powered by batteries, which cannot be replaced by other power sources. So the stability and safety of the batteries are an important part of ensuring the success of the project. According to the ECE445 battery safety document[6], we will understand the battery specifications before installing the battery, test the battery circuit packaging, charging and discharging, and the operating temperature, and pay attention to the isolation from other work areas such as the transmission module and the Raspberry Pi board to avoid impact.

In addition to this, as the drone is flying work outdoors, the weather will cause problems such as short circuits in our work area, which will affect our work. Therefore, we will follow IP66 standard to make the working area of the drone waterproof, so that it can operate normally even in the rain and snow.

Also, given that the drone needs to be manned, I need to make sure that the drone can fly safely and avoid collisions with people or other objects to reduce the risk of maneuvering. Currently, I decided to add sensors and GPS modules to determine the location of the drone. In addition, I will use the programming software to achieve special circumstances such as an automatic landing function. These measures maximize the safety of drone maneuvering.

In addition to the manipulation method, when choosing the flight area and time period, we also need to ensure that the drone will not cause threats and interference, and avoid flying in densely populated areas and flight-restricted areas. When flying, we need to comply with local regulations and rules to ensure that my project is operating within legal limits. Because our tests and demonstrations are conducted on the campus of Zhejiang

University, according to the school's guidelines [7], we need to apply to the school in advance before the drone flight.

According to the Institute of Electrical and Electronics Engineers(IEEE) Code of Ethics 1 "to protect the privacy of others, and to disclose promptly factors that might endanger the public or the environment;"[8], we promise that the data set used in the project will seek the permission of the owner, and the collected images will also be cleared after use in order to protect information security. The final result of the project cannot be used in any scenario that infringes on public information and privacy.

According to IEEE Code of Ethics 5 "to seek, accept, and offer honest criticism of technical work, to acknowledge and correct errors" and "credit properly the contributions of others;" [8] We will take the problems that may occur in the whole process of the project seriously and correct it in time. For example, we need to adjust the index and model if the simulation result is not as expected. Also, the references involved and other open-source references will also be correctly referenced and identified with permission.

References

[1] X. Luo, H.-H. Chen, and Q. Guo, "Semantic communications: Overview, open issues, and future research directions," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 210–219, 2022.

[2] D. Deacon, M. Pickering, P. Golding, and G. Murdock, *Researching communications: A practical guide to methods in media and cultural analysis*. Bloomsbury Publishing USA, 2021.

[3] T. Han, Q. Yang, Z. Shi, S. He, and Z. Zhang, "Semantic-preserved communication system for highly efficient speech transmission," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 245–259, 2022.

[4] Z. Qin, X. Tao, J. Lu, and G. Y. Li, "Semantic communications: Principles and challenges," *arXiv preprint arXiv:2201.01389*, 2021.

[5] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Communications magazine*, vol. 54, no. 5, pp. 36–42, 2016.

[6] U. of Illinois ECE445 Course Staff. "Safe practice for lead acid and lithium batteries." (2016), [Online]. Available: https://courses.engr.illinois.edu/ece445zjui/documents/GeneralBatterySafety.pdf (visited on 03/21/2023).

[7] Z. University. "Drone school flight." (2020), [Online]. Available: http://xwfw.zju.edu.cn/wsbsdt.php?cmd=sx_content&blsxm=347300&dxlb=&fwxz=&bjlx=&fwms=&sldz=&fwzt=&zuixin_tuijian=&orderby=&jianhua_flag=&shoulibm=&one_run=&keyword=%E6%97%A0%E4%BA%BA%E6%9C%BA&p=1 (visited on 03/23/2023).

[8] IEEE. "Ieee code of ethics." (2020), [Online]. Available: https://www.ieee.org/about/corporate/governance/p7-8.html (visited on 03/07/2023).