

## Drone Monitoring System to Detect Human Posture Using Deep Learning Algorithm

Udin Komarudin<sup>1</sup>, Azuan Ahmad<sup>2,\*</sup>, Widyatama<sup>3</sup>, Iman Hafifi Zainal Abiddin<sup>4</sup>,  
Madiah Mohd Saudi<sup>5</sup>

<sup>1</sup>Widyatama University

<sup>2</sup>CyberSecurity and Systems (CSS) Research Unit, Faculty of Science and Technology (FST), Universiti Sains Islam Malaysia (USIM), 71800 Nilai, Negeri Sembilan, Malaysia;

<sup>3</sup>Widyatama University

<sup>4</sup>CyberSecurity and Systems (CSS) Research Unit, Faculty of Science and Technology (FST), Universiti Sains Islam Malaysia (USIM), 71800 Nilai, Negeri Sembilan, Malaysia

<sup>5</sup>CyberSecurity and Systems (CSS) Research Unit, Faculty of Science and Technology (FST), Universiti Sains Islam Malaysia (USIM), 71800 Nilai, Negeri Sembilan, Malaysia

<sup>2</sup>azuan@usim.edu.my

**Article History:** Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 20 April 2021

**Abstract:** Artificial Intelligence is taking part in many forms of daily human life. Many innovations have been embedded in technology to ensure they reach the same or even higher than human capability. One of the trending innovations in the market is image recognition. Image recognition is now implementing machine learning and deep learning for better image detection. Usual issues come for any image recognition technology about the low ability to detect various object classes. Too much research on image recognition also led to difficulty in applying the best algorithms. Although there are many technologies regarding image recognition, there is still not much work on human pose detection. Therefore, this paper proposed an application for detecting human pose related to drone control using a deep learning algorithm. This research aims to review the type of deep learning algorithm for human pose detection, develop an enhanced algorithm based on deep learning algorithm for human pose classification, and evaluate the proposed human pose classification algorithm's performance based on accuracy using drone technology. This paper proposes using Convolutional Neural Network (CNN) as a selected deep learning algorithm suitable for pattern recognition. This research expects an accurate result for detecting the human pose involving controlling the drone.

**Keywords:** Image recognition, deep learning, drone, pose

### 1. Introduction

Nowadays, image recognition spreads drastically in the world of technology. Many image recognitions implement machine learning in their system, especially when related to high load processing and large image data pixels. An effort to make an output of specific data to be much more accurate and reliable is why this technology intensely studied and implemented. Much research has been on image recognition to be applied for video and image processing. The enhancement made on the Unmanned Aerial Vehicle (UAV), or publicly known as a drone, can be seen as the research results.

For academic purposes, image recognition has made a significant contribution for the student with learning disabilities. This new technology has provided a suitable learning process for them. The present apps powered by computer vision that offer text-to-speech options significantly impact those who suffer impaired vision and dyslexia to 'read' the content. Moreover, image recognition brings traditional teaching into an advanced way of teaching.

Therefore, this paper is conducted to leverage drone technology usage, seeing its multi-usage in the daily human lives. In the future, the drone will not only be a benefit for the industry sector but also for humankind. The implementation of drone technology will be vast and cover many aspects that will ease human lives.

This paper is organized as follows: Section 2 explains the literature review, Section 3 discusses the methods, Section 4 presents the findings, and finally Section 5 with conclusions and future work.

### 2. Literature Review

#### 2.1 Image Recognition

Image recognition has been a highlighted issue being discussed in the present world. This is due to many technology improvements relating to video and image processing for research purposes, industrial (example: agriculture, fishery, automotive, and filming), medical, and security surveillance. Image recognition is the process

of identifying and detecting objects or features in digital images or videos. Machine learning and deep learning are the best approaches to image recognition.

### 2.2 Deep Learning

Deep learning (also known as deep structured learning or hierarchical learning) is an aspect of artificial intelligence (AI) that emulates the learning approach in gaining certain types of knowledge. A machine learning class is a cascade of multiple layers of nonlinear processing for feature extraction and transformation. The output that came from the previous result will be used as an input by the next layer. It can be learned either in supervised (e.g., classification) and/or unsupervised (e.g., pattern analysis) manners.

Deep learning can be applied to supervised and supervised learning but preferably unsupervised learning because it is faster and usually has better accuracy. Firstly, the process for deep learning will include understanding the problem and justifying whether deep learning is a good fit. Secondly, identifies relevant data sets and prepares them for analysis. Thirdly, choose the type of deep learning algorithm to use. Fourthly, train the algorithm on a large amount of labeled data. Finally, test the model's performance against unlabeled data. A simplified step is concluded in Figure 1.

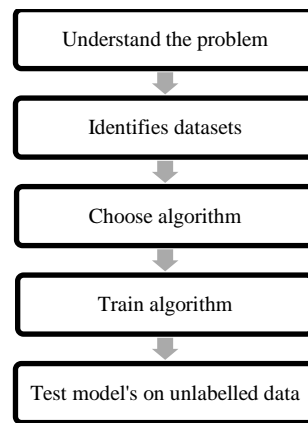


Figure 1 Convolutional Steps

### 2.3 Convolutional Neural Network (CNN)

The convolutional neural network is a class of deep neural networks. CNNs are comprised of multi-layer of a fully connected network. It is a powerful image processing using deep learning to perform certain tasks involving image and video recognition. CNN consists of multi-layer, input layer, convolution layers, pooling layers, rectified Linear Unit layers, and fully connected layers [1]. Figure 2 shows a simplified illustration of CNN architecture [2].

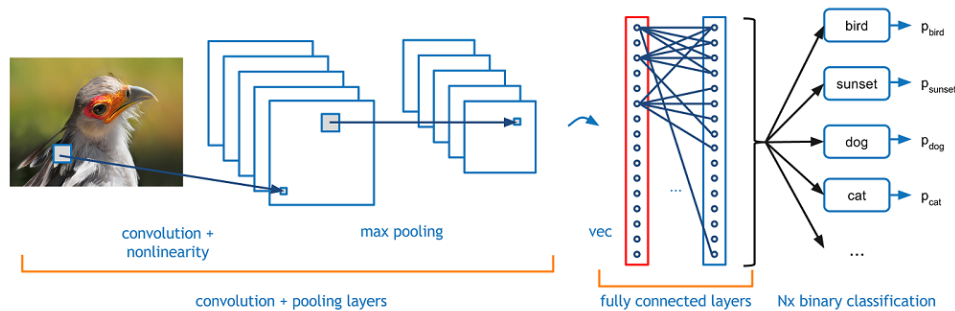


Figure 2 Architecture of CNN

### 2.4 Unmanned Aerial Vehicle (UAV)

Drone technology is widely used for many purposes, either for any industry, organization, and even personal use. To get a suitable UAV or drone that suits demand and requirement, this research should know the types and abilities of the drone want to use. Here is a list of a different drone with different capability [3]:

- i. Unmanned Aerial Vehicle for surveillance (UAVS) of fixed wings is used for long-distance purposes [4].
- ii. Quadcopter Unmanned Aerial Vehicle (UAV) is used for short-time purposes and surveillance [5].
- iii. iHex-rotor or oct-rotor is used for critical purposes such as to carry expensive sensors [6].
- iv. iTwin and single rotors are used for low power consumption purposes [7].

UAV is going towards more stability and can off according to vertical takeoff landing (VTOL).

### 2.5 Related Work

This section is discussing on related works that being done on machine learning applications on UAV.[8] discussed precision in agriculture. Some method has been proposed from the paper: retrieving plant pigment

concentrations, estimating vegetation height from UAVs and hydrological modeling based on UAV terrain data. The first part combines physical and statistical methods in exploiting the information content. It was also utilizing the optical-to-thermal canopy radiative transfer model (4SAIL). The second part estimates the vegetation height from UAV, it uses the structure from motion (SfM) and sub-decimeter resolution optical UAV data. The last part was utilizing digital surface models derived from a UAV SfM workflow for hydrological modeling. This proposed method provides a better visualizing crop system. Next, it develops higher-level products on vegetation height and biomass. It also performed an improvement in drone battery life. Having an automated operation of flight and navigation controls. Finally, they have success work in a growing range of lightweight sensors for deployment. There is one issue regarding this research paper: it has lacked in translating sensor retrievals into useable and georeferenced suitable for precision agriculture application.

In a research paper from [9], they are implementing deep learning-based software which uses a CNN algorithm on a platform named "You Only Look Once" (YOLO), where it combines deep learning algorithms and advanced GPU technology. From their research, it shows an improvement of autonomous flying capabilities and the operational safety of the aircraft, but it has a low capability in detection and classification of various object classes.

Work by [10] is proposing an investigation on the deep learning approach for image recognition. Two major algorithms have been explained clearly in the paper. The first one is CNN and the second one is DBN. They have chosen the SVM-KNN algorithm as a benchmark. CNN and DBN have different capabilities when been applied for image recognition. The paper shows that CNN has a specialty in pattern recognition tasks while DBN can learn a hierarchical representation of the training data. One issue highlighted is for DBN models, it provides only a fixed input size. Therefore, extra work needs to perform like resizing and cropping the image to make it fix with the input size, but the process of truncating the object can reduce the classification performance.

[11] proposed a method for dealing with and taking advantage of noise in the image recognition task. To perform the proposed idea, they came out with two methods. The first one is augmenting training data with some expected noise. This method is said to be the simplest approach. Second, train some classification on undistorted data prior to data denoising. This method enables the possibility of finetuning the final model. It also produces sufficient image quality, allowing the transfer of further layers from pre-existing models and trained on undistorted data. One issue regarding this paper is if there is too high an amount of noise or wrong type of distortion, it will negatively affect the classification of performance.

[12] works are focusing on data parallelism. They have implemented two parallelism strategies. The first one is butterfly synchronization, and the second one is a lazy update. They also implement data augmentation as this process can improve the performance of the network. For the training phase, they were implementing multi-scale training as it works better than single-scale training.

[13] proposed drone monitoring for flooding zone. The method used in this paper consist of classifying images to highlight the flooding zone between urban and non-urban areas. Their drone technology can gain high-resolution images and have accuracy in the classification of images.

[14] proposes a technology for shark detection in a shallow coral lagoon. To investigate shark detection, they have made 2 survey blocks, each with 3 parallel transects oriented perpendicular to the cost. What is best in their research is they can successfully achieve fishery-independent density data using UAV. They are also able to investigate population trends and habitat use patterns. This technology also eases assessing the effect of human activities, especially in coastal and shallow habitats.

### **3. Methods**

#### **3.1 Proposed Architecture**

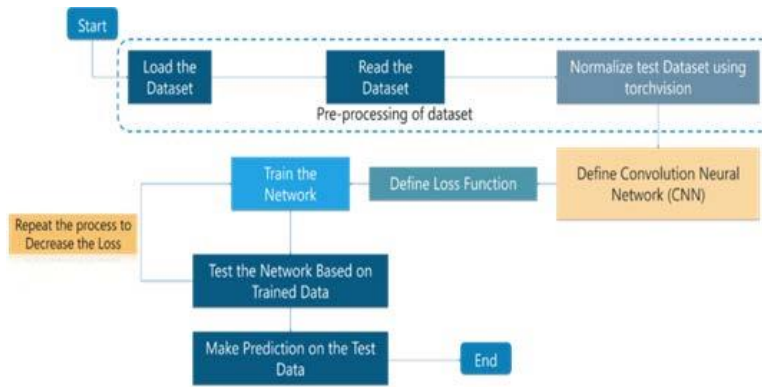


Figure 4: Flowchart of CNN [15]

### 3.2.1 Collecting Dataset

This research drone system will need some amount of dataset to make it work. Therefore, this research divide into two the way collects the dataset. First, online collection, and second, offline collection. Online collection requires us to search and crawl through the internet to find any relevant human behavior data. The offline collection is where take images using the drone to make a group of data set. This research's total dataset will be 6000 for the training image and 4000 for the testing image.

### 3.3 Experiment Setup and Dataset

#### 3.3.1 Drone Selection

This research uses a quadcopter as this drone is much more suitable for surveillance activity, and it can perform a stable movement make it easy to get capture the human pose. For this research, the Tello drone is used as it is friendly for various research related to drones and suitable for a beginner in conducting recognition research using a drone.

#### 3.3.2 System Feature

This research provides video streaming to capture human pose and some other features. The system starts with providing an interface of the video stream and some other functional button:

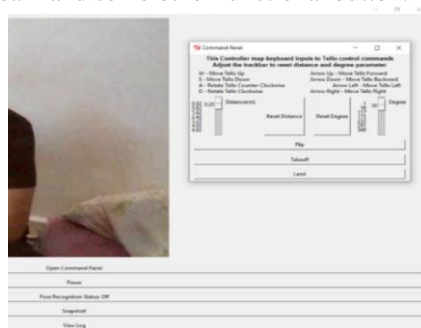


Figure 5: System main interface

Table 1: Details of system feature

No	Button	Description
1	Open Command Panel	This button will pop up a "Command Panel" interface used to control the drone. It consists of five other functional buttons: <ol style="list-style-type: none"> <li>i. Reset Distance – This button is used to reset the distance of the drone flight.</li> <li>ii. Reset Angle – This button is used to reset the turning angle of the drone.</li> <li>iii. Flip – This button is an extra feature to command the drone to perform some stunt movement like flip forward, flip backward, flip right, and left.</li> <li>iv. Takeoff – This button will give the command to the drone to take off.</li> <li>v. Land – This button will give the command to the drone to land on the ground.</li> </ol>
2	Pause	This button will act to freeze to video streaming.

- |   |                         |  |
|---|-------------------------|--|
| 3 | Pose Recognition Status | This button has two functions either to turn it "On" or "Off." This button will detect the human skeleton or known as a key point to start recognizing the human pose and give the drone command using the human pose. |
| 4 | Snapshot                | This button is used to take the desired picture from the video streaming screen.   |
| 5 | View Log                | This button is used to view the log of drone usage.  |



Figure 6: Some Pose Recognition

This section explains how the pose recognition being setup.

### 3.3.3 Dataset Preparation

The research started with the collection of the dataset of human pose. This research requires a huge number of datasets to get a precise classification of recognition. Therefore, dataset collected by CMU Panoptic Studio is being used combined with my dataset to get an excellent recognition result. From the total dataset of 10000, 60% will be separate for training, and another 40% will be saved for testing purposes.

### 3.3.4 Training of dataset

Once the dataset is ready, they will now be trained by using the CNN algorithm. As you can see in the illustrated features, the dataset collected earlier will go through four processes: a detailed explanation in Chapter two. The four processes will be:

i. Convolution:

Two-layer of the convolutional process are implemented for this research. Here, the image will be filtered as the kernel slides through every pixel of the image to produce an activation map.

ii. Pooling:

The pooling process is where the image pixel's complexity will be reduced, making image identification much easier.

iii. Relu:

This is where the activation function will take place. All negatives values will be removed while all the positives values will stay remain.

iv. Fully Connected Layer:

The final process is where all those processed images will be matched according to their highest level and translated into categories with labels. of dataset sample

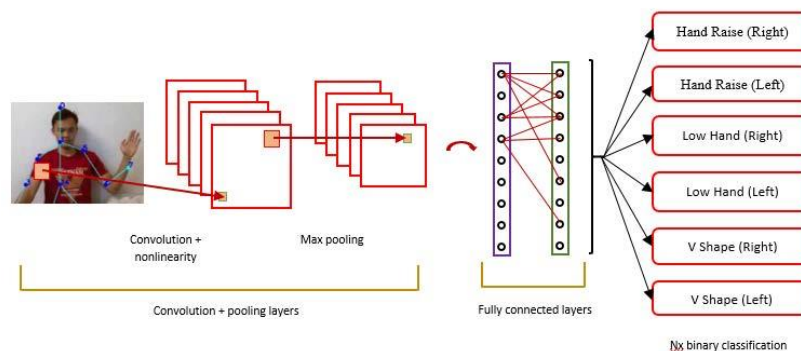





Figure 7: The architecture of CNN

### 3.3.5 recognition using drone system

This research was conducted to control the drone movement by interpreting the human pose. Therefore, for this research to be successful, the system had been set to detect the movement of arm pose. Since the pre-trained model already trained and understands what so-called a key point or human skeleton drew on the GUI picture, it can detect the arm's position, either right or left, by knowing the key point of the arm. The right arm's key point will be numbers 2, 3, and 4, while the left arm will be numbers 5, 6, and 7, where the number sequence followed the position starting from the shoulder, elbow, and wrist, respectively.

To control the drone, give it the right pose to move according to the desired command. In this research, the drone will perform three movements from the recognition of the human pose. The first one is moving forward, the second is moving backward and finally land on the ground safely. For a much flexible movement, two range of angle has been set for each arm. The angle of the shoulder only and the angle range from the elbow to the shoulder. Therefore, the drone still can understand the given pose as long the arm angle is still inside the required range. Following are the details on how the drone differentiates the human pose:

Table 2: Details how drone differentiates the human pose

Pose	Description	Figure
Hand Raise Pose	To command the drone to move forward, need to flatten up an arm like the illustrated sketch. For the right arm, the angle range will be from $-10^{\circ}$ to $40^{\circ}$ . As for the opposite arm, the angle range will be from $140^{\circ}$ to $190^{\circ}$ . While the angle from the elbow to the shoulder, both will need to be below $30^{\circ}$ .	
Low Arm Pose	To make the drone moving forward, need to lower down the arm, as illustrated. The range of angles for the right arm will be from $-60^{\circ}$ to $-20^{\circ}$ . The other arm will be from $200^{\circ}$ to $240^{\circ}$ . While the angle from the elbow to the shoulder, both will need to be below $25^{\circ}$ .	
"V" Shape Arm Pose	The final pose will give a command to the drone to land safely on the ground. The required arm pose is a "V" shape pose like have been illustrated. The range of angle of the right arm for the shoulder will be from $-60^{\circ}$ to $-20^{\circ}$ , and the angle between the elbow to the shoulder will be from $0^{\circ}$ to $90^{\circ}$ . While the left-arm angle for the shoulder will be from $200^{\circ}$ to $240^{\circ}$ , and the angle between the elbow to the shoulder will be from $90^{\circ}$ to $180^{\circ}$ .	

### 3.3.6 Log File

Inside the log file, the usage of the drone in detail can be seen. It records the day, date, and time for the drone is being launched.

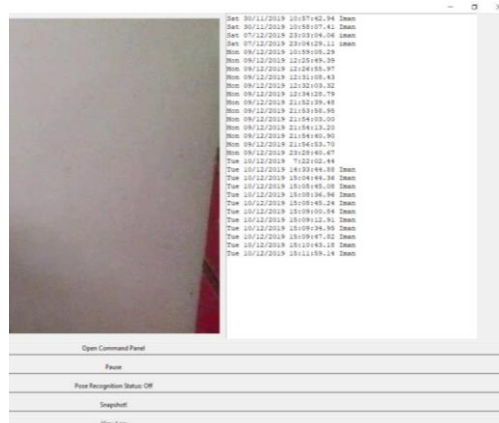


Figure 11: Log File









**4. Findings**

**4.1 Evaluation Based on Accuracy Using Drone Technology**

For each class of classification, a repetition of a total of 100 attempts is done to test the accuracy of the drone and to test whether the drone could perform the desired movement according to the human pose detection.

Table 3 Drone detection desired movement according to the human pose

Command	Right hand	Left hand	Figure
1. Command the drone to move forward:	a. Using right hand		 <p>Figure 12: Arm Raise pose using the right hand</p>  <p>Figure 13: Enlarge image for arm raise pose</p>
	b. Using left hand		 <p>Figure 14: Arm Raise pose using the left hand</p>  <p>Figure 15: Enlargement image for arm raise pose</p>
2. Command the drone to move backward:	a. Using the right hand		 <p>Figure 16: Low hand pose using the right hand</p>  <p>Figure 17: Enlargement image for low hand pose</p>
	b.		

b. Using left hand



Figure 18: Low hand pose using the left hand

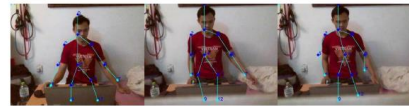


Figure 19: Enlargement image for low hand pose

3. Command the drone to land:

a. Using the right hand



Figure 20: "V" shape hand pose using the right hand



Figure 21: Enlargement image for "V" shape hand pose

b. Using left hand



Figure 22: "V" shape hand pose using the left hand



Figure 23: Enlargement image for "V" shape hand pose

Table 5.7: Summarization of all Attempt

Pose	Number of Attempts	Successful Detection
Hand Raise (Right)	100	95
Hand Raise (Left)	100	96
Low Hand (Right)	100	93
Low Hand (Left)	100	93
V Shape (Right)	100	97
V Shape (Left)	100	96



**Total**

**600**

**570**

Thus, the total accuracy of detection using drone technology is equal to:

$$\frac{95 + 96 + 93 + 93 + 97 + 96}{600} * 100 = 95\%$$

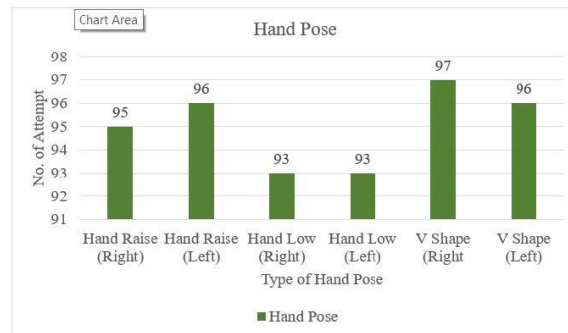


Figure 24: Graph of attempt vs. type of hand pose accuracy result

From the result shown, it can be seen the total number of successful detections for each pose. It gained 95 successful detections for hand raise (right), 96 successful detections for a hand raise (left), 93 successful detections for hand low for right and left, 97 successful detections for 'V' shape (right), and 96 successful detections for 'V' shape (left). From the above result, it reached its best accuracy with a total of 95%. It can be concluded that pose recognition could be best suited with the implementation of the CNN algorithm.

#### 4. Conclusions

As for this research, the chosen algorithm, Convolutional Neural Network (CNN), performed its best in getting astonishing accuracy to be used by the drone for human pose recognition. Mainly from the overall research, the constraints and limitations of this research are the varieties in controlling the drone. This research provides only three movements rather than many other movements that can be done. This is because when adding other movements will cause some conflict during the detection of a pose. It caused some interference while giving a different command to the drone. Another constraint will affect the system itself, where it cannot afford to get much input too fast. Otherwise, it will make the system crash. Furthermore, the Tello drone model used to have a short-life battery, causing slow testing, evaluation, and demonstration. Image recognition has many kinds and variations of implementation, such as detecting the human face. Therefore, some recommendations can be made for future work of this research. Adding the face recognition where an authorized person only can control the drone means a raise in the drone's security level. Other than that, combining an image detection of various objects, thus increasing the drone's functionality instead of just detecting a human pose and face, and an extra advantage by classifying other objects. Next, increase the variety of commands given to the drone. For future recommendations, the ability to command the drone by using a hand gesture is an advantage.

#### Acknowledgment

The authors would like to express their gratitude to Widyatama University, Indonesia, and Universiti Sains Islam Malaysia (USIM) (USIM grant no:) for the funding, support, and facilities provided.

#### References

1. Sharma, N., Jain, V., & Mishra, A. An Analysis of Convolutional Neural Networks for Image Classification. *Procedia Computer Science*, 132(Iccids), 377– 384. 2018 <https://doi.org/10.1016/j.procs.2018.05.198>
2. Deshpande, A. A Beginner's Guide To Understanding Convolutional Neural Networks – Adit Deshpande – Engineering at Forward | UCLA CS '19. 2016
3. Karim, S., Zhang, Y., Laghari, A. A., & Asif, M. R. . Image processing based proposed drone for detecting and controlling street crimes. *International Conference on Communication Technology Proceedings, ICCT, 2017–Octob(May 2018), 1725–1730. 2018* <https://doi.org/10.1109/ICCT.2017.8359925>

4. Pfeifer, Barbosa, Mustafa, Peter, Brenning, & Rümmler, M. . Using Fixed-Wing UAV for Detecting and Mapping the Distribution and Abundance of Penguins on the South Shetlands Islands, Antarctica. *Drones*, 3(2), 39. 2019. <https://doi.org/10.3390/drones3020039>
5. Tatale, O., Anekar, N., Phatak, S., & Sarkale, S. Quadcopter : Design, Construction, and Testing Quadcopter : Design, Construction, and Testing. (March), 0–7. 2019. <https://doi.org/10.18231/2454-9150.2018.1386>
6. Pathak, S., Poudel, R., Maskey, R. K., & Shrestha, P. L. . Design and Development of Hexa-copter for Environmental Research. 11th International Conference on ASIAN Community Knowledge Networks for the Economy, Society, Culture and Environmental Stability, (April 2015).
7. Agarwal, S., Mohan, A., & Kumar, K. . Design, Construction and Structure Analysis of Twinrotor UAV. *International Journal of Instrumentation and Control Systems*, 4(1), 33–42. 2014. <https://doi.org/10.5121/ijics.2014.4103>
8. McCabe, M. F., Houborg, R., & Lucieer, A. . High-resolution sensing for precision agriculture: from Earth-observing satellites to unmanned aerial vehicles. *Remote Sensing for Agriculture, Ecosystems, and Hydrology XVIII*, 9998, 999811. 2016. <https://doi.org/10.1117/12.2241289>
9. Radovic, M., Adarkwa, O., & Wang, Q. . Object Recognition in Aerial Images Using Convolutional Neural Networks. *Journal of Imaging*, 3(4), 21. 2017. <https://doi.org/10.3390/jimaging3020021>
10. Bärar, A.-P., Neagoe, V., & Sebe, N. . Image Recognition with Deep Learning Techniques 2 Model Description. *Recent Advances in Image, Audio and Signal Processing*, 126–132. 2016
11. Jabarullah, N. H., Surendar, A., Arun, M., Siddiqi, A. F., & Krasnopevtseva, T. O. (2020). Microstructural Characterization and Unified Reliability Assessment of Aged Solder Joints in a PV Module. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 10(6), 1028-1034.
12. Wu, R., Yan, S., Shan, Y., Dang, Q., & Sun, G.. Deep Image: Scaling up Image Recognition. 2015. Retrieved from <http://arxiv.org/abs/1501.02876>
13. Popescu, D., & Ichim, L. . Acivs 2004 (advanced concepts for intelligent vision systems). *Computers & Graphics*, 28(3), 460. 2006. [https://doi.org/10.1016/s0097-8493\(04\)00041-x](https://doi.org/10.1016/s0097-8493(04)00041-x)
14. Kiszka, J. J., Mourier, J., Gastrich, K., & Heithaus, M. R. Using unmanned aerial vehicles (UAVs) to investigate shark and ray densities in a shallow coral lagoon. *Marine Ecology Progress Series*, 560(November), 237–242. 2016. <https://doi.org/10.3354/meps11945>
15. Rao, A.. Convolutional Neural Network (CNN) Tutorial In Python Using TensorFlow | Edureka. 2019