Monocular Vision Aided Drone Localization in Indoor Corridor

Shahzad Ahmad

Department of Computer Science and Engineering
National Institute of Technology Rourkela

Monocular Vision Aided Drone Localization in Indoor Corridor

Thesis submitted in partial fulfillment

of the requirements for the degree of

Master of Technology

in

Computer Science and Engineering
(*Specialization*: *Computer Science*)

by

Shahzad Ahmad

(Roll Number: 216CS1134)

based on research carried out

under the supervision of

Prof. Pankaj Kumar Sa

May, 2018

Department of Computer Science and Engineering
National Institute of Technology Rourkela

Department of Computer Science and Engineering
National Institute of Technology Rourkela

May 26, 2018

Certificate of Examination

Roll Number: *216CS1134*
Name: *Shahzad Ahmad*
Title of Thesis: *Monocular Vision Aided Drone Localization in Indoor Corridor*

We the below signed, after checking the thesis mentioned above and the official record book (s) of the student, hereby state our approval of the thesis submitted in partial fulfillment of the requirements of the degree of *Master of Technology* in *Computer Science and Engineering* at *National Institute of Technology Rourkela*. We are satisfied with the volume, quality and correctness of the work.

Pankaj Kumar Sa
Principal Supervisor

Carl Friedrich Gauss
External Examiner

Varahamihira
Head of the Department

Department of Computer Science and Engineering
National Institute of Technology Rourkela

Prof. Pankaj Kumar Sa
Associate Professor

May 26, 2018

Supervisor's Certificate

This is to certify that the work presented in the thesis entitled *Monocular Vision Aided Drone Localization in Indoor Corridor* submitted by *Shahzad Ahmad*, Roll Number 216CS1134, is a record of research carried out by him under my supervision and guidance in partial fulfillment of the requirements of the degree of *Master of Technology* in *Computer Science and Engineering*. Neither this thesis nor any part of it has been submitted earlier for any degree or diploma to any institute or university in India or abroad.

Pankaj Kumar Sa

Dedication

Dedicated to family ,professors and friends

Declaration of Originality

I, *Shahzad Ahmad*, Roll Number *216CS1134* hereby declare that this thesis entitled *Monocular Vision Aided Drone Localization in Indoor Corridor* presents my work carried out as a postgraduate student of NIT Rourkela and, to the best of my knowledge, contains no material previously published or written by another person, nor any material presented by me for the award of any degree or diploma of NIT Rourkela or any other institution. Any contribution made to this research by others, with whom I have worked at NIT Rourkela or elsewhere, is explicitly acknowledged in the dissertation. Works of other authors cited in this dissertation have been duly acknowledged under the sections "Reference" or "Bibliography". I have also submitted my original research records to the scrutiny committee for evaluation of my dissertation.

I am fully aware that in case of any non-compliance detected in future, the Senate of NIT Rourkela may withdraw the degree awarded to me on the basis of the present dissertation.

May 26, 2018
NIT Rourkela

Shahzad Ahmad

Acknowledgment

Though an individual contribution, this M.Tech. thesis has actually been benefited in various ways from several people. The list may be finite, however, their continued support is innumerable.

The supervision of Prof. Pankaj Kumar Sa has inspired me to go beyond my limits and improve my skills to analyze the feasible solutions for the research challenges.

The profound insight and the invaluable advise of Dr. Sambit Bakshi has been a source of motivation for nourishing my intellectual maturity.

I am also grateful to research scholar Ram Prasad Padhy for his unfailing support and assistance. Without his superior knowledge and experience, the Project would like in quality of outcomes, and thus his support has been essential.

I especially thank to my friends for their continuous encouragement and generosity throughout my masters research.

Last but not the least, my heartfelt thanks to my family for their unconditional love and support. Words fail me to express my gratitude to my beloved parents, who sacrificed their comfort for my betterment.

May 25, 2018
NIT Rourkela

Shahzad Ahmad
Roll Number: 216CS1134

Abstract

Vision based navigation of unmanned aerial vehicles (UAV) has been an active field of research in the past decade. There are many challenges in making the vision system understand the environment in which it is placed. Such an environment can be either indoors or outdoors depending on the task a hand. In This project we deal with localization of UAV in GPS-denied indoor environment. We consider A.R Parrot drone is our UAV model. Our aim is to localization of drone in corridor during flying time with the help of monocular camera which is attached with drone.

In this project we designed navigation algorithm which will tell us the position of drone in corridor by seeing the images which are taken by drone camera. We create on dataset for localization algorithm by capturing the image of corridor for possible position of drone in corridor. Our localization algorithm implement with help of deep learning models use to perform regression task on the images which are taken by drone camera. We use help of transfer learning take recent deep learning models already trained with Imagenet Challenge data set .We modify deep learning models remove last layer and add own convolution and fully connected layers base on our desired output and train with our created dataset of corridor images.

***Keywords*: *Deep learning* ; *localization*; *UAV*; *corridor*; *convolution* .**

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Nowadays, the technological revolutions in both hardware and software vanishes the gap between science fiction stories and physical reality. Regarding robot technology, sophisticated hardware and software are combined together to develop machines which are capable to behave like human and replicating human activities. Today UAVs are used in many wide applications like surveys , search and rescue operations[7], surveillance, defense application, and path detection of ground vehicles etc. which are very dangerous and difficult to reach to human beings.

However this shows some problems being simple for human beings , but it become very difficult when these simple problems comes down to replicating them on machines. So technology in this field still in required in more advancement in order to gain human like intelligence.

Walking down in corridor for human being is very simple task but it is very difficult for UAVs to walk down in GPS denied indoor corridor environment. Due to unavailability of GPS localization of UAV is still challenge using monocular camera in UAV. In this thesis we develop capability in UAV to navigate autonomously in GPS denied indoor corridor environments.

1.1 Motivation

Autonomous indoor navigation of Unmanned Aerial Vehicles (UAVs) still suffers many problems. One of the main problems is because limitation of GPS signals in indoor environments. The reason behind the above problem is UAVs are capable to carry some more heavy weight and power consuming sensors, such as side cameras and Leaser depth estimator, makes indoor autonomous navigation a difficul task.In This project we propose practical system in which UAV(quadcopter) navigate autonomously GPS denied Indoor corridor environment. We use deep learning model Convolutional Neural Network (ConvNet) to develop localization capability by seeing the images which are capture by drone camera.

1.2 Problem Statement

Unmanned Aerial Vehicles(UAVs) or which are more popularly known as Drones or Quadcopter have pushed their way into modern day problem solving techniques and by far have successfully proved their usability in Defence Activities, Reconciliation Missions and photography. With the improvements in optimized energy utilization during flight, newer dimensions of UAVs deployment are being explored. These may include Drones in agriculture, disaster management, tracking, area mapping, traffic control, crime monitoring and surveillance. Yet again these problems can be solved again taking into account in what type of environment we want to deploy our drone. These environments may have the property of being windy, high altitude(low pressure), low visibility(fog), night conditions or indoor environments. These conditions are not mutually exclusive of each other. Navigation of drones are basically dependent on GPS coordinates, but in indoor environments , getting GPS signal is tough. Instead of flying blind, we make use of the on-board Camera and apply deep learning techniques to sense the environment and make a meaningful flight pattern by avoiding obstacles and also gathering maximum information on the go.

1.3 Unmanned Aerial Vehicles (UAVs)

In last decade, UAVs such as quad-copters are center of interest in robotics and computer vision domain. Generally used as a flying camera for surveillance and recording videos which are impossible for human. Hence safe navigation of these drones depend on-board sensors and cameras to find their positions. The advantages of using UAVs are, it is very dynamically flexible and ability to stay in air.
While there has been so many research done about navigation algorithms for ground vehicles, but these techniques can not be use directly to aerial vehicles:

- The limitation of UAV it can not hold many sensor unlike ground vehicles.
- A UAV six degree of freedom which is more than a ground vehicle so we cannot use 2D navigation algorithm.
- The risk factor associated with UAV is much more than ground vehicle. If navigation algorithm predicted wrong in ground vehicle ,it can be

stopped by user, but if same scenario occurs with UAV then may be it fall on the ground and damage occur.

1.3.1 On-board Electronics

Different type of UAVs have set of devices which are help to give accurate data which can useful for localization ,navigation ,landing and obstacle avoidance .

- **GPS**
  The Global Positioning System[8] was funded by the United States Department of Defence andwas initially designed for the United States military, although in 1980, it was made available forcivilian use.

  GPS works at any coordinate of the world at any point of time in any weather condition. Four satellite are used to compute the position coordinate of UAVs in three dimensional space and also the time use as fourth coordinate in the receiver clock.

- **IMU**
  The IMU is a single unit has consist two sensors, which accounts angular velocity and linear acceleration data and send to main processor. It is used to manage pitch ,roll and yaw information from the UAV.
  A brief introduction of these two sensors follows:

  – **Accelerometer:** An accelerometer is a device that compute acceleration forces which may be static (the force of gravity), or dynamic (vibrating or even moving the accelerometer itself).

  – **Gyroscope:**A Gyroscope is a device that compute the orientation of a device based on the angular momentum of that same device. It is used to find the angular rate of a certain UAV[8].

- **Laser Range Finder**
  This device uses a laser beam to determine the depth estimation of the an object from source camera .The mechanism of this device is sending a laser pulse to target object and computing the depth that it receive the pulse to reflect off the object and bounce back to the emitter[9].

- **Ultrasonic Sensor**
  An Ultrasonic sensor is a device that uses to compute the distance to an object by using sound waves. It computes distance by sending sound wave at a certain frequency and receiving for that sound wave to bounce back. By calculating the elapsed time between the sound wave being

generated and the sound wave bouncing back, it is possible to calculate the distance between the sonar sensor and the object.

- **Pressure Sensor**
  A pressure sensor is generally used to achieve stability to an UAV, allowing the on-board processing to automatically correct and manage a still position while the UAV is air. This is achieved regardless of height and wind pressure[7]. On-board pressure sensors gives unique stability that will automatically correct and manage a still position in the air regardless of height and wind pressure.

- **Magnetometer**
  A Magnetometer is work as a compass able of measuring the strength and direction of a earth's magnetic field. So it is very for navigation of UAVs.

- **Monocular Camera**
  Monocular cameras are light weight , low power consuming ,size and cost they still provide some valuable information which is unmatchable by any other type of sensor.The benefit of monocular camera is the range is virtually infinite while depth estimating devices has limited range[10].

- **Stereo Camera**
  Stereo camera has two or more set of lenses which also provide 3D information of object with help of depth information. It is possible to capture 3D images with these cameras due to the fact that the lenses simulate the human binocular vision.

- **RGB-D Sensor**
  The RGB-D sensor gives the depth information of every pixal in RGB image.The data acquired from these sensors in the form of point cloud. Point cloud is the set of point in three dimensional space.

1.3.2 Quad-copters

In last decade UAVs are the center of attraction in computer vision research domain due their simple movability, mechanical simplicity and many enhancements that have been made in technologies such as: control, stability, batteries and sensors.

Recent researches in drone technology have provides different types of drones. The most common ones are listed below

- **Flapping wing drones**
- **Fixed wing drones**
- **Rotor based drone**

Flapping drone has great advantage due to its light weight structure, small size and ability to move quickly and easily[11] [12]. Fixed wing drones use a gliding mechanism: either built in linear propellers or an external propulsion mechanism. Fixed wings drones are less activity unlike flapping wing drones but it have more flight duration which are very useful in survey and mapping application[13].Rotor Based drone provides an six degree of freedom during its flight from regulating the speed of rotors. The drones have capability to carry more sensors as well as onboard cameras. These camera capture images and video that are very useful in many computer vision research domains such as obstacle detection, tracking, feature extraction. The limitation of camera based method is it depend on external light sources needed to illuminate the environment and on edge and texture features.

Now in present time there are so many option available in UAVs with set of sensors and sophisticated hardware. For this project a quad-copter is sufficient to fulfill our requirement which is related to indoor navigation. The quad-copter provides agile and stable flight and very less maintenance. We consider only those quad-copters which has frontal monocular camera help us to capture video.

A various range of drone is present in market. One of the quad-copter drones is Parrot AR. Drone. It is most affordable , easy to handle and low maintenance.

When compared to other quad-copters such as the AscTec Firefly or the Dji Phantom, the Parrot AR. Drone look best solution for this computer vision project[14] due to several factors:

- **Low price**
- **Open communications protocol:** AR Drone provide HD camera with high quality and transmission rate.
- **Repairability:** AR. Drone is very easy repairable which is a characteristic of great interest since it will be used as a testing platform. Therefore, falls and impacts is certainly not avoidable.

1.3.3 Parrot AR. Drone

The Parrot AR. Drone figure 1.1 was initially popular as a toy and therefore is quite popular and affordable.

Table 1.1: Technical Specifications of the Parrot AR. Drone 2.0

Structure	Carbon Fiber and Outdoor Hull
Weight	400 g
Autonomy	18 min
Sensors	Gyroscope,Accelerometer,Magnetometer,Pressure Sensor, Ultassounds,Vertical Camera for ground velocity
Communications	Wi-Fi
Camera	720p 30fps. Low latency Wi-Fi Transmission
Observations	Fully Repairable. Open and documented communications protocol

Figure 1.1: The AR. Drone quad-copter with the protection hull attached

Parrot A.R Drone is very easy to maintain and replace part well as providing specific features of on-board stabilization and accessible control design. A list of the most useful properties of Parrot A.R Drone are shown on the list below.

- **Front and bottom facing cameras:** The Parrot AR. Drone has two camera first is front facing camera and another is bottom facing camera .First camera use for capturing videos and images while the bottom facing camera is used to se object below the drone which help to drone provide horizontal stabilization and velocity estimation.
- **Automatic stabilization:** With the help of rotors, gyroscope and bottom facing camera drone provides -board stabilization system.
- **Wireless connectivity:** Wireless devices can easily connect to the AR. Drone by drivers such as the ardrone-autonomy.

1.3.4 Quad-copter flight control

The control and dynamics of drone is completely depend by its design. In this design four rotor blades attached to drone body that is arranged in four directions. In this design each pair of rotors turn in the same direction meaning that, as seen in Figure 1.2, the rotor pair 1,4 and 2,3 turn in opposite directions: the former turns anticlockwise and the latter turns clockwise. This architecture provide the ability to drone to stay in air on one place.

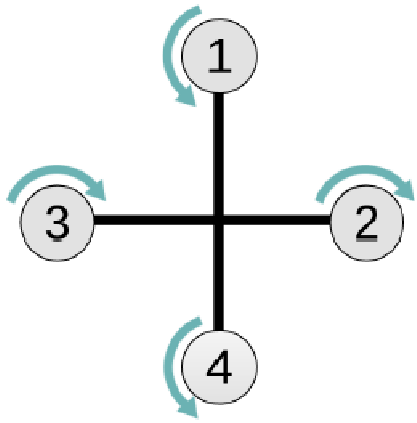


Figure 1.2: Rotation Direction of each of the four rotors found in a quad-copter

The motion of drone depends on a three-dimensional tilt/rotation system around all three perpendicular axes that are X, Y ad Z axis. As shown in Figure 1.3, rotation about each of the three axes allows the drone to move forward/backwards (pitch), left/right(roll) and turn left/right (yaw).

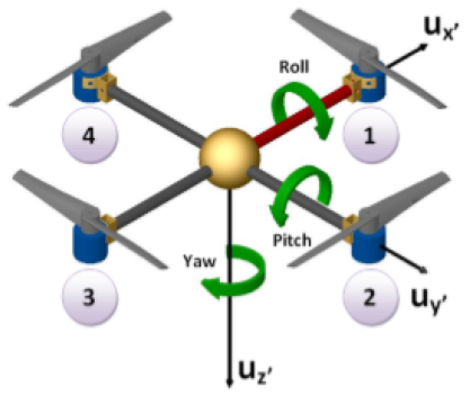


Figure 1.3: The rotation system on a quad-copter: yaw, pitch and roll

To gain linear movement, the rotation speed of corresponding rotors changes relatively to the other two rotors. A quad-copter has the following degree of freedom:

- **Forward/backward linear movement:** To maintain change in speed between the front and the rear rotors gains this movement (+/- pitch).
- **Up/down linear movement:** Maintain the same speed on all four rotors acquire this movement.This means that lower speed for lower height of the drone and higher speed increases the hight in air (+/- height).
- **Left/right linear movement:** To maintain change speed between the right and the left rotors gains this movement (+/- roll).
- **Left/right rotation movement:** To maintain change speed between the diagonal rotors gains this movement (+/- yaw).

The speed and direction of motion of the quad-copter depend on the degree of rotation about their axes and speed of each rotor.

1.3.5 Inertial Measurement Unit (IMU)

The Parrot AR. Drone has an Inertial Measurement Unit (IMU) with six degrees of freedomthat are computed using the given components:

- A 3-axis accelerometer used to compute acceleration along the X,Y and Z axes.
- A gyroscope used to compute rotation in degrees per second, roll about X-axis ,pitch about Y-axis and yaw about Z-axis by angular velocity.
- To estimation of height, stabilization and vertical speed AR. Drone is use an Ultrasound Altimeter that is attached to bottom of the drone.

1.3.6 On-board Processing

Since A.R Drone was designed as a toy hence it has very limited computational power which are not sufficient to execute computer vision algorithms that requires much more computational resources. One possible solution is using a computer connected to the AR. Drone via wireless. The communication path between the AR. Drone and the computer can be seen in Figure 1.4.The batteries of this UAV provide for a 15 minutes backup for flight.

1.4 Related Works

There are some interesting research work has been done on autonomous navigation and flight of UAVs. In this section, we examine related works.

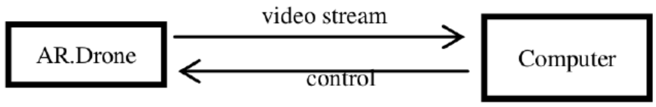


Figure 1.4: Data exchange between the AR.Drone and a Computer

1.4.1 Range Sensor

One possible solution is to use range sensor like range finders , infrared sensor,RGB-depth sensors . Bry et al.[15] use to measure depth of objects in other words depth of object in 2D image using an on-board laser range find to achieve autonomous navigation in GPS-denied environments.

Roberts et al.[16] used four infrared sensors and one ultrasonic sensor and implement fully autonomous flight. But to carry too much sensors for a UAV is too heavy and consume lots of energy. Our proposed work is based on only use of monocular camera which take very less energy and easily attach in to many of quad-copters.

1.4.2 SLAM(Simultaneous localization and mapping)

The concept behind SLAM is Using range sensors or visual sensors and create 3-D map of unknown indoor environments which helps to estimate the position of object and structure of environment ([17],[18],[19],[20]). Bachrach et al.[21] used a laser finder to construct 3-D map for high level SLAM implementation to visualizing unknown indoor environments.

Celik et al[20] proposed autonomous indoor navigation based on SLAM using monocular camera. However,3-D construction is computationally ex-pensive.Therefore it is very difficult to maintain synchronization between UAV and computation due to delay between perception and action. SLAM also less accurate when it use in indoor environment ,which provide insufficient 3-D points that can be traced frame to frame. Our approach does not construct 3-D map. Thus our proposed work is to able minimizing the delay by reacting fast to its currently faced situation.

1.4.3 Stereo Vision

With the help of stereo cameras that is two or more cameras it is possible to compute accurate depth estimation and relative position estimation([22],[23]).But in texture-less environment stereo vision algorithms perform not well ,it is very hard to match features between two images.Another problem is mot easily available quad-copters have only monocular camera so that above approach is not practical to use. Our system shows robust performance in texture-less environments with the help of deep learning

1.4.4 Other Approach

The other approach uses vanishing points. With the help of monocular camera Bills et al.[24] find vanishing point. Vanishing point is the intersection point of 3-d parallel lines in 2-d image plane. Vanishing point help to fly in corridor environment. In staircase environment they try to find center of the staircase. Front- facing short-range sensor also used to avoid collision in corners. Another approach learns control policies from input data. The ALVINN project[25] [25]uses 3-layer artificial neural networks try to mimic a human driver response road and performed autonomous vehicle driving.Ross et al.[26] used a novel imitation learning approach[27], the DAgger Algorithm, and learned a controller strategy that mimic human pilots selection of action from demonstrations of the desired behavior.

1.5 Goals and objective

In this project we will implement a deep network to guide a drone to autonomously navigate in indoor environment. The problem in indoor environment basically lies in the non-availability of GPS coordinates. Contrary to the present age state of the art navigation methods for UAVs, use GPS localizations alone and manual override wherever necessary.The goal of the project is to develop a way to guide drone such that it can navigate inside a in corridors Today indoor navigation is still an unsolved problem, thus our research attempts to solve it.

In our project we will work on a manufactured Drone from Parrot named A.R Drone version 2.0. We will be working with its monocular HD primary Camera with frame rate of 30fps. We shall have an Ubuntu System with a powerful processor (2.7 or higher clocked) and interfaced with ROS(Robot Operating System).

Deep network in our project will take the raw image as active input,captured from the camera at the drone side, and predict angle between the horizontal axis of image and image of bisector of the floor of corridor and the distance between point of intersection of horizontal axis and the bisector of corridor.

Given the complexity of the stated problem, this project needs to be split into simpler problems:

- **Dataset Creation:** Our first objective is dataset creation for our deep learning architecture .We capture images of corridor at different position from drone camera.
- **Deep learning architecture design:** Second objective is design different deep learning models using transfer learning to finding angle prediction of bisector.
- **Localization Algorithm design:** Our third objective is design of localization algorithm Which help us to find position of drone in corridor.
- **Deploy in drone:** Our last objective is to deploy deep learning architecture in by using ROS(Robot operating System).

1.6 Outline of Thesis

- **Chapter 2:**In this chapter we will discus about deep learning basics and deep learning models and its characteristics.
- **Chapter 3:** In this chapter we will discus about our propose work. Our propose work includes proposed localization algorithm, dataset creation, modified deep learning networks which we use to train and find drone position in corridors, and in last discussed about results.
- **Chapter 4:**In this chapter we conclude our thesis work by highlighting the insights and future scope of thesis work.

Chapter 2

Deep Learning

The idea of deep learning comes from the concept of human brain having many layer of neuron. With the help of layer of neuron human brain recognize the pattern at the lower levels and high-level abstractions. Deep learning is nothing but a technique to train deep neural network. Human brain is like deep neural network which consist many layers of neurons that helps to recognize pattern at more abstract levels. This way of representing information in a more abstract way is easier to generalize for the machines.

The shallow network used by conventional learning technique but deep neural network has ability to learn higher level of feature due to it deep layers. Training of deep network is more computationally expensive than the shallow network. Deep neural networks has very large number of parameter due to fully connected layers. Functions compactly represented in n layers may require exponential size when expressed in 2 layers. Formally, it can be proved that a $n\ layers$ network can represent functions compactly but a $(n-1)\ layers$ network cannot represent them unless it has an exponentially large number of hidden units. A so many of factors like faster CPUs, parallel CPU architectures, GPU computing enabled training of deep networks and made it computationally feasible.

Machine learning is very fast developing to be a major branch of computer science with so many applications in science and engineering for many years. A big challenge in machine learning is lack of data. Deep neural networks need lot of training data to build accurate and reliable model. Machine learning applications depends on hand-engineering features where the researcher manually create a data set .the quality of data set decide quality of machine learning model. When quality data are in short supply, the learning models can predicts very poorly on a new domain, even if the learning algorithms are best chosen. Deep Neural Networks can able to learn features from directly raw data. Usual machine learning starts with features engineered manually

2.1 Convolutional Neural Network (ConvNet)

Convolutional Neural Network are similar to ordinary neural networks which are made up of neurons and have learnable weights and biases. The wights are multiplied to input to neuron and optionally follows non linearity. They continue to have loss function on their last layer. The difference between this network and ordinary neural network figure 2.1 is that ConvNet[28] use pixel value as input rather than feature vectors. This makes the forward function work efficiently and reduces parameters. Unlike ordinary neural network ,neurons in Convnet figure 2.2 are arranged in 3 dimension. The two dimension refer to the spatial size of the filter while the third dimension refer to the no of channel of the input. The neurons at one layer is connected to small region of layer above it. And output of one region helps to generalize the output of other regions. Ordinary neural network can not be given pixel value as in that the number of pixel decides the number of neuron in input layer and the spatial information of the image are lost.

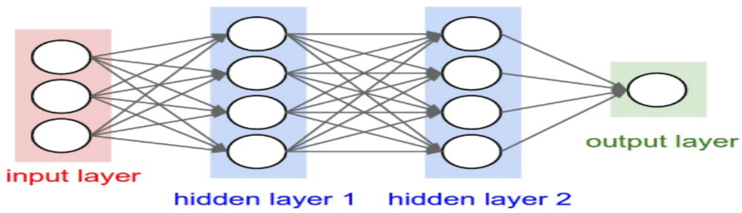


Figure 2.1: Ordinary Neural Network with two hidden layers

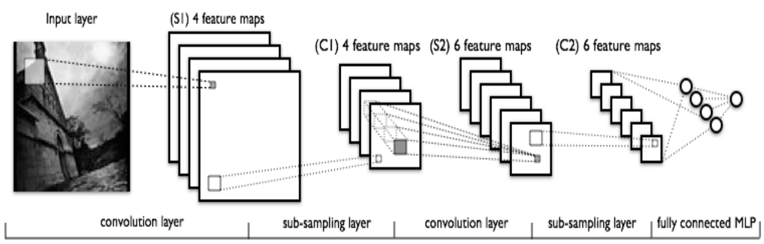


Figure 2.2: Convolution Neural Network

2.1.1 3D Volumes Of Neuron

In a regular Neural Network, neurons are stacked in 2 dimensional layer, but in ConvNet[28] neurons are arranged in 3 dimension. These dimensions width, height as always and depth which refer to third dimension of the activation volume figure 2.3. The neurons at every layer only connected to small region.

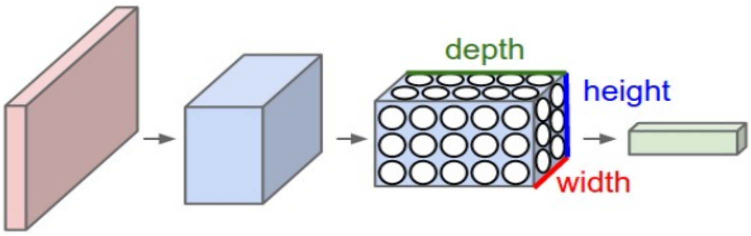


Figure 2.3: 3D volume of neuron

2.1.2 Layers Used in ConvNet

ConvNet are composed of series of several layers and each layer transform one intermediate activation output to other using differentiable function.

Convolutional Layer

This layer forms the basic block of ConvNet. Here parameters are in the form of filter (or kernel) figure 2.4 whose elements are learnable. These filters are small, and access only small portion of activation coming from the previous layer at one point of time but extends through the whole depth. During forward pass, the filter are convolved across the activations coming from previous layer and performs dot product through full depth of activation and produce 2D map of new activation. These new generated activations by different filters are stacked in a layer to form the output of a convolutional layer figure 2.5. The paramets of the filters are so learned such that it detects a particular features like edges, ridges, etc.

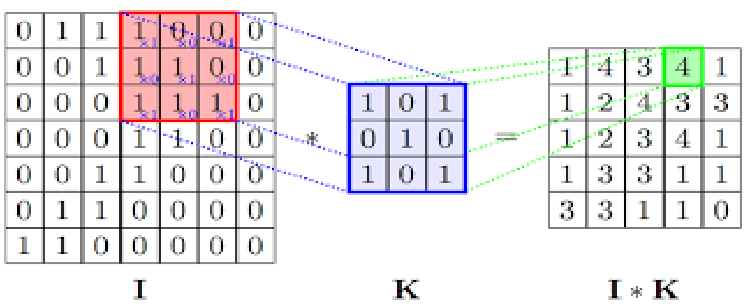


Figure 2.4: Convolution filter

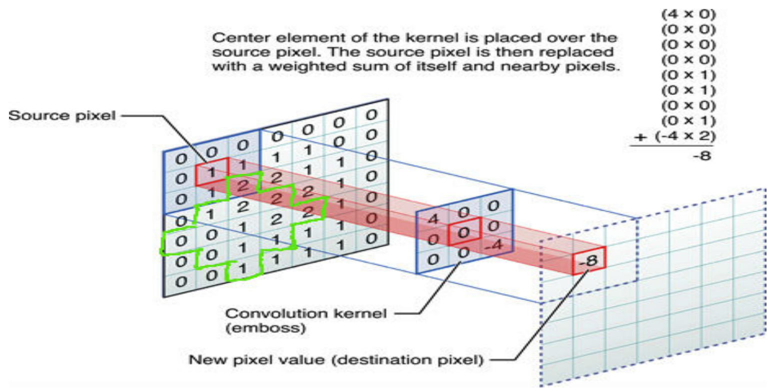


Figure 2.5: Convolution layer

Activation Layer

In neural networks, activation functions are used to enable the network to learn non-linear functions. There are several activation function e.g. Rectified Linear Unit (ReLu) figure 2.7 which performs element wise and produce maximum among the 0 and the activation value i.e. *f(x)=max(0,x)* where x is the activation value.

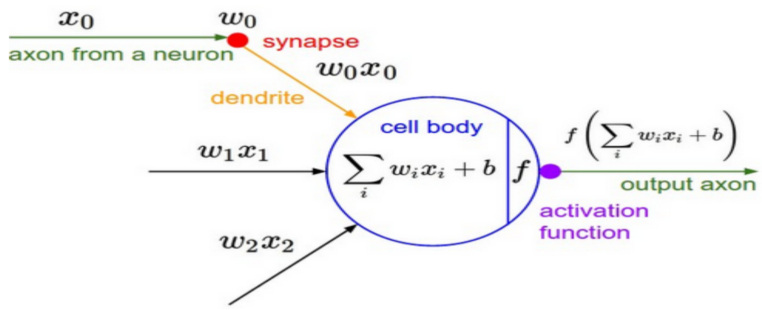


Figure 2.6: Activation function of Neural Network

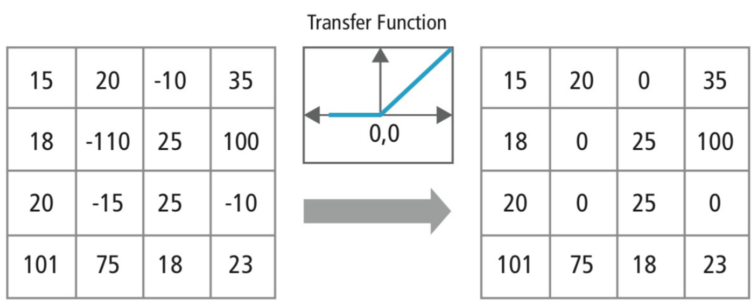


Figure 2.7: ReLU activation function

Tanh is a hyperbolic tangent function which also applied element wise, and produce the hyperbolic tangent of the activation value i.e. $f(x) =$

$\frac{exp^x - exp^{-x}}{exp^x + exp^{-x}}$ where x is the activation value.

Pooling

Pooling layers figure 2.8 are used to reduce the spatial dimension of the activations generated by convolutional layer. this makes computation faster by keeping the prominent features and throwing away the less important features. There are several techniques for pooling like average pooling, max pooling.

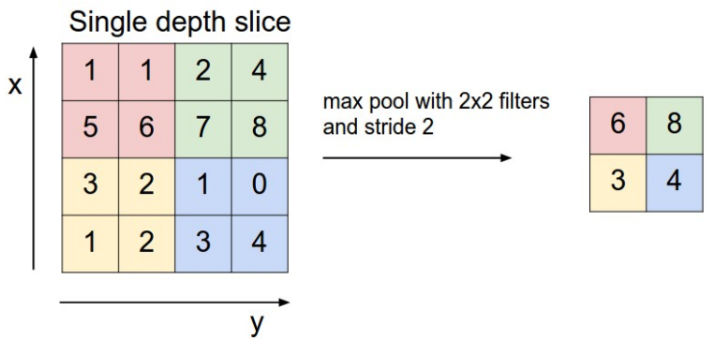


Figure 2.8: Max pool

Fully Coonected Layer

the neurons at the convolutional layer where maintaining spatial feature intact. At the deep layers of ConvNet[28] the model use similar architecture of regular neural network, which here is defined as fully connected layer figure2.9. It is just a linear stack of neurons which takes input all activations from the previous layer and performs dot product by the learnable parameter to produce output.

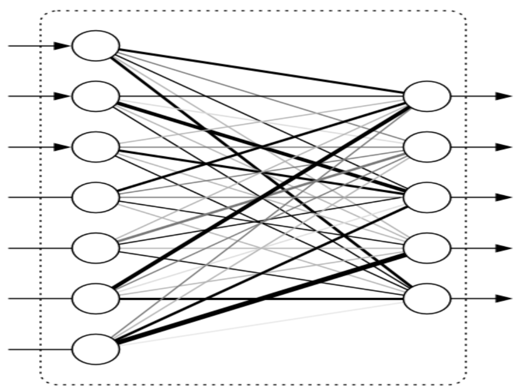


Figure 2.9: Fully connected layer

2.2 Related Work

This section will summarize important recent developments in the field of computer vision and convolutional neural networks. Here are some of the most important networks that have used CNN [1] as their core part in learning model.

AlexNet [1](2012)

The paper, titled "ImageNet Classification with Deep Convolutional Networks", [1] proposed a deep mosel in 2012 termed termed as Alexnet [1] . This model was used in Imagenet Large Scale Visula Recognition Challenge(ILSVRC) [29] in 2012. For the first time in ILSVRC, CNN [1] model was used to achieve a top 5 test error to be 15.4%. Before this top 5 error was 26.2%.
The AlexNet [1] is composed of 5 convolutional layer, max pool layer, dropout layer and fully connected layer. The network was designed to classify among 1000 class labels.

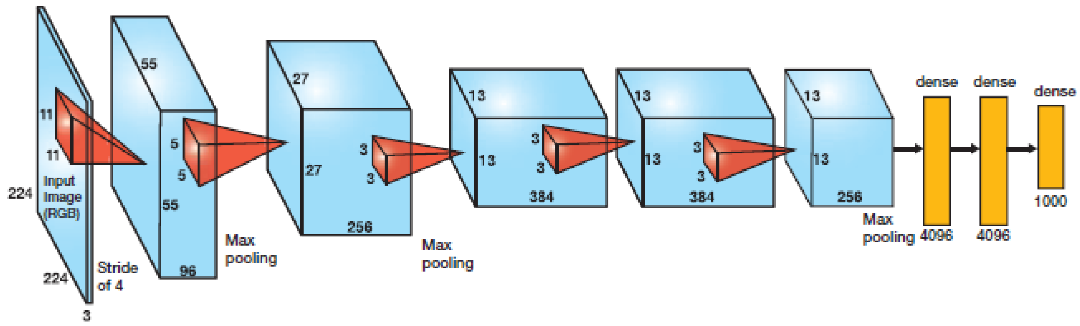


Figure 2.10: AlexNet [1] Model

ZFNet [2](2013)

In 2013, the winner of ILSVRC [29] is a network called as ZFNet [2] which achieved 11.2 % top 5 error rate. This network can be imagined as some fine tuning of the previous winner mode AlexNet [1]. Here, every layer of the trained CNN citekrizhevsky2012imagenet is attached to DeconvNet which has a path back to pixel value. This helps to cross-examine the activation at any point of time.

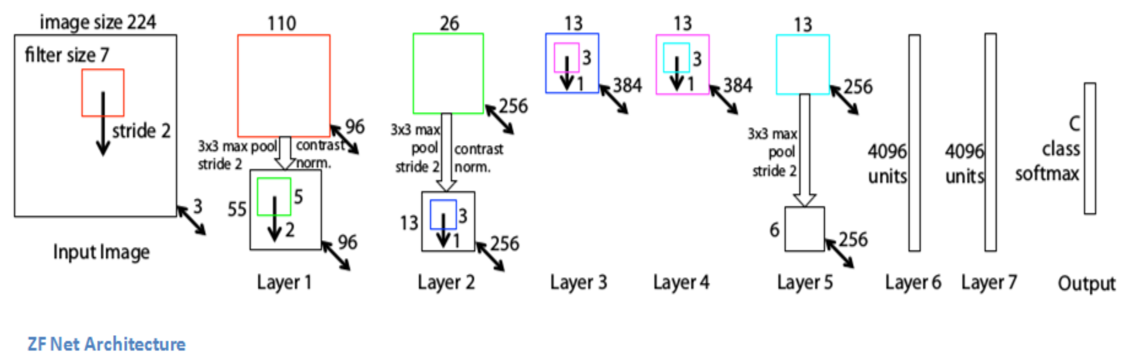


Figure 2.11: ZFNet [2] Model

ResNet [3](2015)

ResNet[3] became the winner of ILSVRC[29]2015 by reducing the top 5 error rate to 3.6%. The concept here is that activation generated by one layer of convolution, like conv-relu-pool, are augmented with the original image for the second layer input.
The main motive of doing this was the layers at any step of deep propagation must get full chance to access the features learned so far as the case till now, but also directly pull out features according to it.

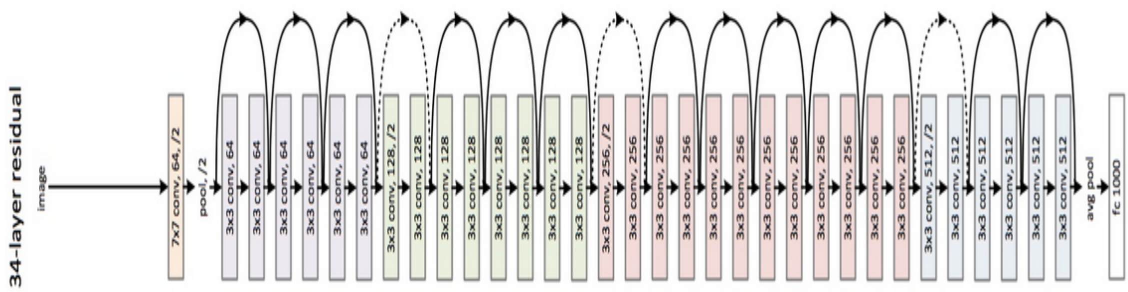


Figure 2.12: ResNet [3] Model

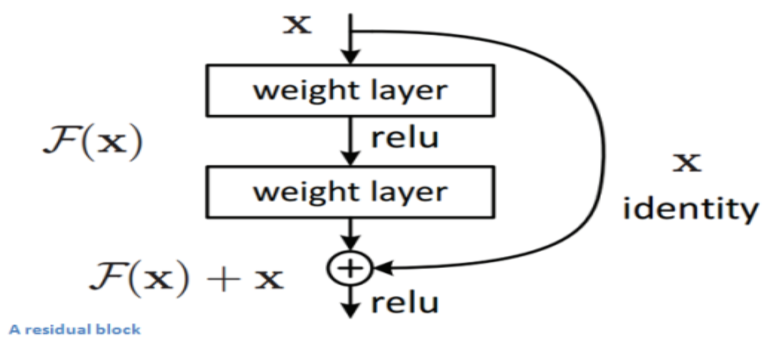


Figure 2.13: ResNet building block

DenseNet [4](2016)

Instead of concatenating the input directly to activation generated at any layer , DenseNet [4] concatenated the activation generated at any layer with the activation of all the previous layer.
In forward propagation, any layer get lower level features along with the higher layer features, and during backward propagation, allows all gradient to reach to respected place easily as all the layer are connected from one another.

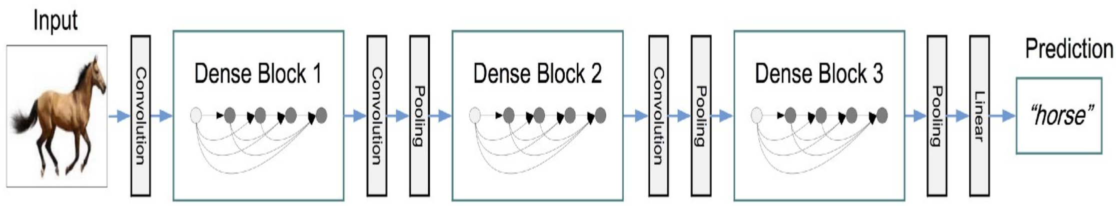


Figure 2.14: Dense Net [4] Model

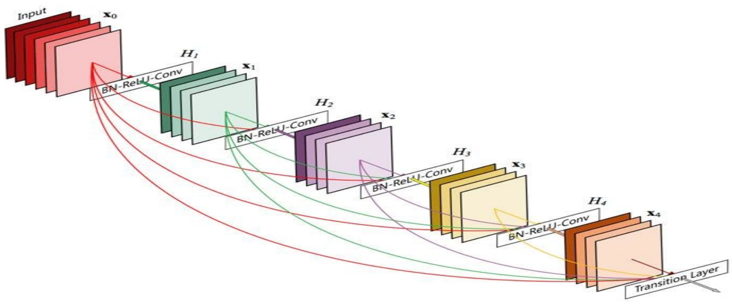


Figure 2.15: Dense block

VGG Net[5]

The VGG network architecture was proposed by Simonyan and Zisserman in their 2014 paper, Very Deep Convolutional Networks for Large Scale Image Recognition[5]. This network is very simple, 3×3 convolutional layers used to increasing depth. Reducing 3D neuron size by max pooling. Two fully-connected layers, each with 4,096 nodes are then followed by a softmax classifier.

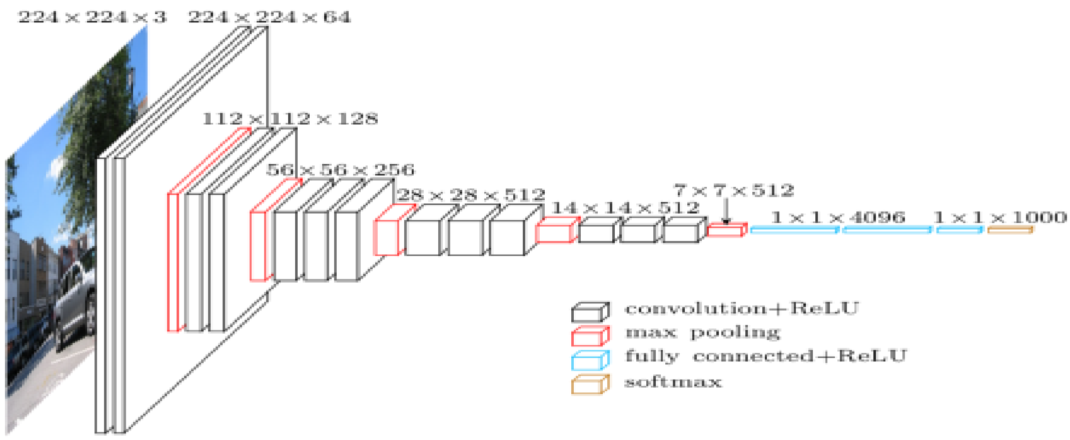


Figure 2.16: VGG Net [5] Model

Inception Net V3[6]

The "Inception" micro-architecture was initially proposed by Szegedy et al. in their 2014 paper, Going Deeper with Convolutions[6] The aim of the inception module is to work as a "multi-level feature extractor" by computing 1×1, 3×3, and 5×5 convolutions within the same module of the network — filters outputs are stacked along the channel dimension and before being fed into the next layer in the network.

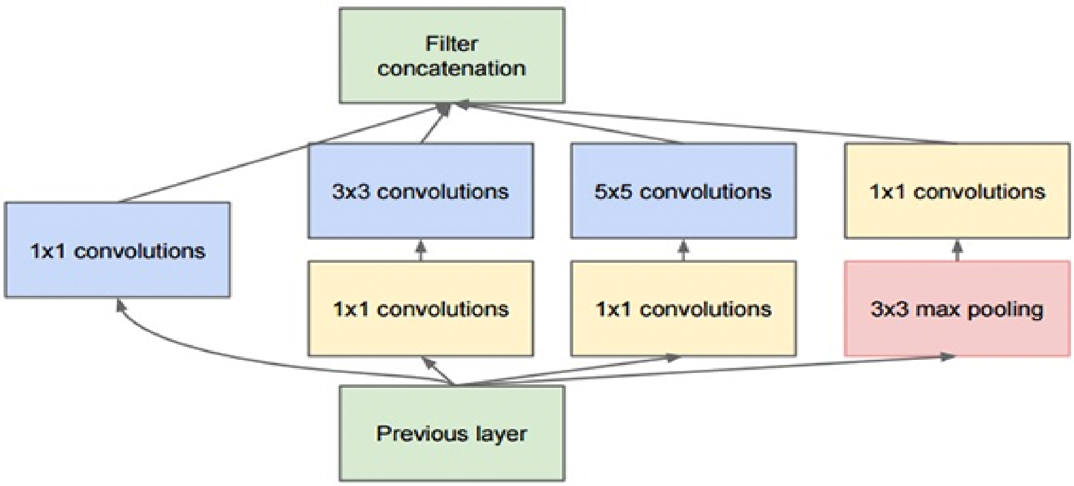


Figure 2.17: Inception Net [6] Module

2.3 Conclusion

We will the above discussed deep learning networks for our project to find drone location in corridor in flight time.Here we will use transfer learning that is take already train network with image net dataset and further train with our data set for corridor images with different position of drone.

Chapter 3

Drone Localization in Indoor Corridors using Deep Learning

Several approaches and there pros and cons were discussed in section .This section we summarize the work we followed to achieve the goal of the experiment. A brief summery of drone navigation is presented along with some images captured from drone camera in order to validate the purpose algorithm.

3.1 Proposed Methodology

Our approach of drone localization is finding the drone position in corridor when drone fly. There are three possible position of drone in corridor, first is left side, second is right side and middle of the corridor and the foundation of our custom dataset . We have presented a navigation algorithm which decides position of drone in the corridor. Also a brief description of deep learning architectures used for the estimation of position of drone is presented here to summarize the working model of our experiment.

3.2 Localization Algorithm

To enable quad-copters to localize autonomously we came up with noble approach that aid a quad-copter localizes itself in indoor scenario with only single monocular camera. In our experiment the input is a RGB image that capture by drone attached camera.

We design localization algorithm which help us to find the position of drone in indoor environment. In our research we consider the corridors as indoor environment. Our purposed algorithm works for autonomous localization of drone in indoor flat corridors. The idea of algorithm is we take help from bisector of floor in corridor for finding position of drone. In corridor we consider three main position of drone which are drone lies left side ,center and right side of corridor. It is easily visible from given figure3.1.

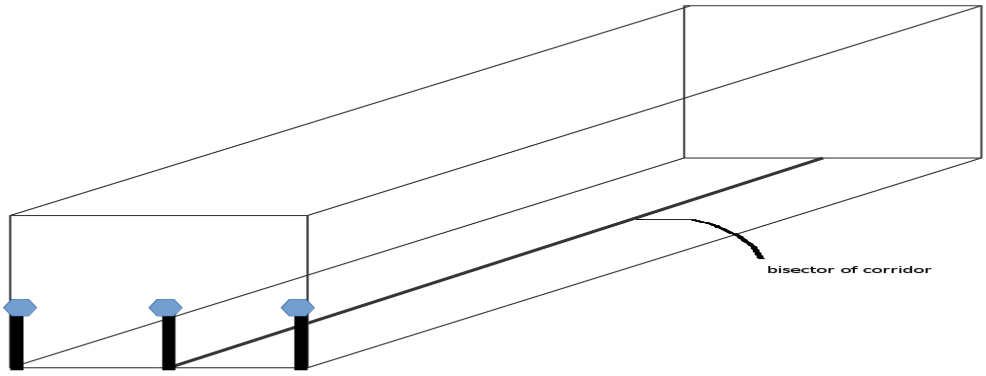


Figure 3.1: bisector of corridor plane and positions of drone

- If our drone lie left side of corridor and we take image from drone camera then image of bisector make angle less than 90° degree from horizontal axis of image in this case we give right shift command to the drone to bring drone in center.

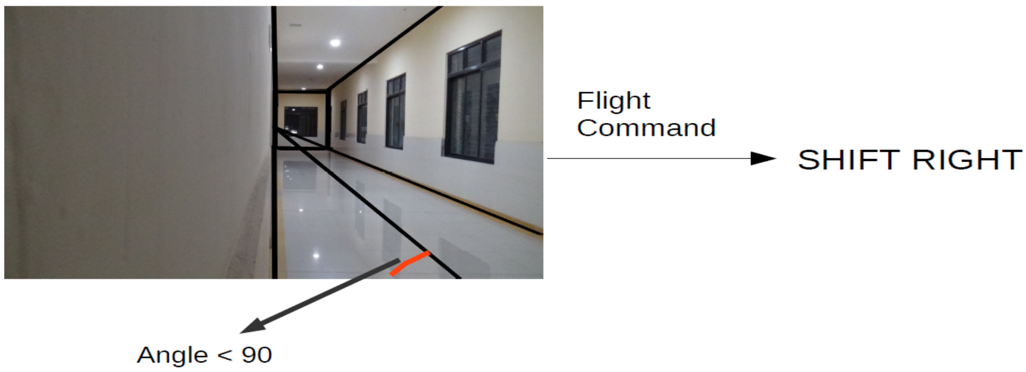


Figure 3.2: Drone lie left side of corridor

- If our drone lie right side of corridor and we take image from drone camera then image of bisector make angle greater than 90° degree from horizontal axis of image in this case we give left shift command to the drone to bring drone in center.

- If our drone lie center of corridor then bisector make angle equal to 90° degree and look toward straight in this case intersection point of bisector and horizontal axis of image lies middle of the image then we give go forward command.

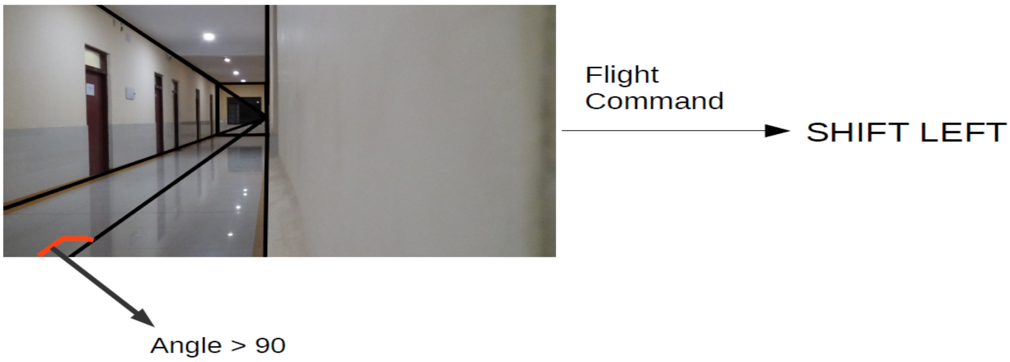


Figure 3.3: Drone right left side of corridor

Figure 3.4: Drone in center of corridor looking straight

Algorithm 1: Localization algorithm

Input: IMAGE,height,width

Result: move command for drone

θ=Algorithm2(IMAGE);

if $\theta < 90$ **then**

  Roll right drone ;

else if $\theta > 90$ **then**

  Roll left drone ;

else

  Pitch forward ;

Algorithm 2: Algorithm for finding angle

```
  Input: IMAGE
  Result: Angle between bisector of plane and horizental axis of image
1 Normalize pixal between 0 to 1;
2 RGB to BGR conversion;
3 Normalize with mean and standard deviation of Image Net dataset.;
4 θ = trained_weights(IMAGE);
5 return θ;
```

In Algorithm 2

- In step 1 normalize the pixels values between 0 to 1 by dividing 255.
- In step 2 convert RGB image BGR used model DenseNet work on BGR images.
- in step 3 normalize image with mean and standard deviation of Image NET dataset RGB mean is [0.406,0.456,0.485] and RGB standard deviation is [0.225,0.224,0.229].
- In step 4 trained_weights is trained Convolution neural network which return predicted value.

3.3 Proposed Architecture

We have use a quite a few number of pretrained deep learning model in ILSVRC since 2012 and their variants. Originally these pretrained model were used majority for classification task. We used transfer learning technique where the original model weight were used and again fine tuned by further training. Also classification layer of all the original models are replaced with convolutional and fully connected layers.

3.4 System Setup

3.4.1 Hardware platform

In our Experiment we used Parrot AR Drone 2.0 figure3.5 attached with monocular facing forward, some ultrasound sensors to keep track of ground altitude and onboard computer. The frontal camera has fish-eye lens capturing wide range of 92 degree. This camera HD 720p camera and has a resolution of 1280x720 pixels. The frame generated at 30 frames per second.

Figure 3.5: Parrot AR Drone 2.0

The quad-copter is attached with 3-axis gyroscope which measure, pitch and roll angle and 3-axis accelerometer to measure acceleration in all three directions. These frames generated from drone are then sent to host machine. Which process these frames, over Wifi which in turn send the motion command to the drone.
Our host machine has the following specification table 3.1.

Table 3.1: System Specification

ITEM	SPECIFICATION
Memory	32 GB RAM
Processor	Intel Xeon(R) CPU E5-2620 v3 @ 3.33 Ghzx12
Graphics	NVIDIA GeForce GT 1080,11GB memory

3.4.2 Software platform

As a software platform we use Ubuntu 14.04. as OS ,Anaconda Jupyter notebook as editor ,Pytorch library for deep learning ,Python Open CV ,and MATLAB for dataset creation.

3.5 Dataset

There is no benchmark dataset of this experiment. Also labeling data manually huge effort and is prone error. Hence we created own dataset capturing around three different location in a corridor featured as left, right, center and each location thee possible scenario i.e. tilted left, tilted right and facing straight. The dataset created in different building, brightness etc. to cover to cover almost all possible scenarios. Also the height of drone at kept at one meter above the ground, which is sufficient to fly in any flat corridor. Shown below some captured image and labeled image for our custom dataset.

3.5.1 Data Set Creation

The second phase of the research we create dataset for our deep architecture. we have covered around 63 corridors. We have used augmentation technique to increase the list of the corridor i,e. we have made our drone to capture images from both the ends of the same corridor.

For converting the video taken by the drone camera into active input for our deep network, we have taken frames from the video of the corridor. The overall approach is take out around as many frames from the video and randomly select it for training and test set. Our drone is capable of extracting 30 frames per second. We presently dealing with around 35k images as trainset and around 600 images for testset. Some of our extracted frames from the video of the corridor are as follws:

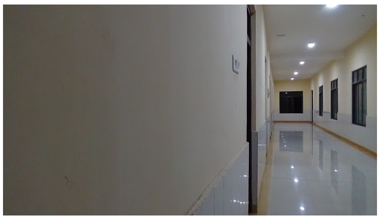


Figure 3.6: Atleft aligned to left

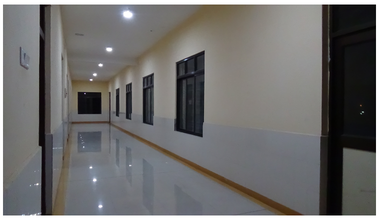


Figure 3.7: at Left aligned to right

Figure 3.8: At left aligned to center

Figure 3.9: At right aligned to left

Figure 3.10: at right aligned to right

Figure 3.11: At right aligned to center

Figure 3.12: At center aligned to left

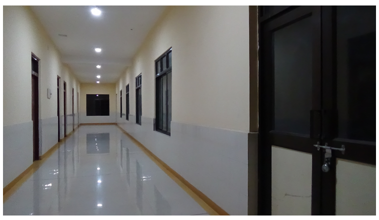


Figure 3.13: at center aligned to right

Figure 3.14: At center aligned to center

The main motive of the data is to capture all possible location of the corridor such that neural network can detect whether the drone is at the left side of the corridor or at middle or at right side. Again there is a possible chance that at left side or right or at middle the drone is not facing straight or simply tilted. To overcome this two challenges we are using our self made concept called line bisector method.

Through some analysis we found that when the drone is at the left side irrespective of its tilted position the angle between the axis of the floor of the corridor and the horizontal axis is greater than 90° while when the drone is at right side fo the corridor the angle is less than 90° ,also when the drone is at the center irrespective of its tilt this angle is 90°. This feature is kept effective as label for the trainset and testset.

Label Data

After capturing images of corridors at different positions of drone draw the bisector of corridor and find angle with horizontal axis of image.This angle treated as ground truth value of images.On the basis of ground truth value we trained our deep learning model for estimation of angle of bisector with horizontal axis of image.
Some labeled images given below:

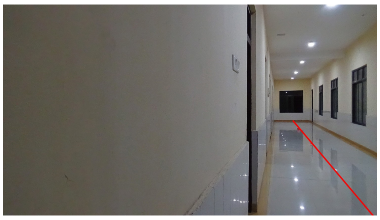


Figure 3.15: Atleft aligned to left

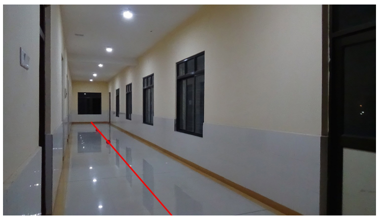


Figure 3.16: at Left aligned to right

Figure 3.17: At left aligned to center

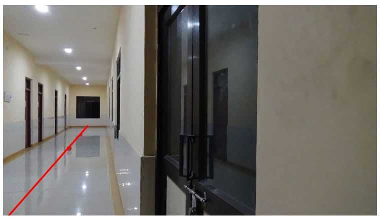


Figure 3.18: At right aligned to left

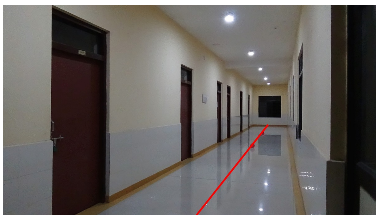


Figure 3.19: at right aligned to right

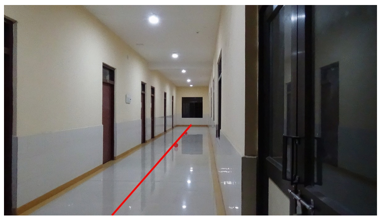


Figure 3.20: At right aligned to center

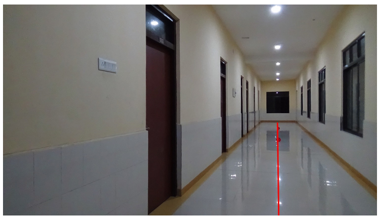


Figure 3.21: At center aligned to left

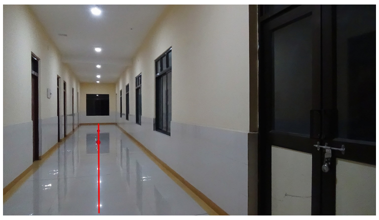


Figure 3.22: at center aligned to right

Figure 3.23: At center aligned to center

Augmentation Technique

We are using augmentation techniques to increase the size of dataset. Primary requirement of deep learning models are large amount of dataset. Here we use augmentation technique first is zoom and flipping about vertical axis of image.

- **Zooming images:** In this we apply zoom technique on image about a point and save images at different zoom views which help us increase number of images.Zooming shown in figure3.24

Figure 3.24: Zooming of image

- **Flip:** In this we apply flip about vertical axis of images and save flipped image.Flip shown in figure3.25

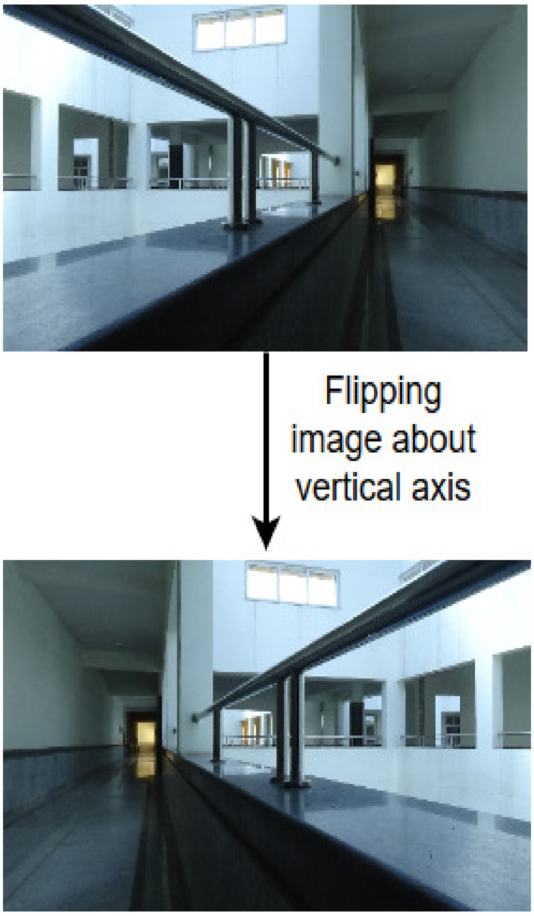


Figure 3.25: Flip of image

3.6 Performance Evaluation

We used several deep learning models to predict the angle between bisector of corridor and horizontal axis of image in image. To train these models we used 35000 images for training and 600 images for testing. Following two subsections summarize the loss function and the result of training and testing process. The angel is measured in degree.

3.6.1 Working Models

Coming to the implementation perspective, we have used eight deep learning models for our regression to predict the angle between the bisector of the corridor and the horizontal axis of image.

- **Alex Net**[1]

- **VGG-16 Net**[5]
- **Res-101 Net**[3]
- **Res-152 Net**[3]
- **Inception Net**[6]
- **Dense-201 Net**[4]
- **Dense-161 Net**[4]

These models are winners of Imagenet Large Scale Visual Recognition Challenge(ILSVRC) in different years and a state of art in present scenario and performing very well around various computer vision challenges. We use our working models with some modification. These deep learning models used for classification problem, but we will use transfer learning technique for using it as regression technique figure3.26. We removed the last fully connected layers and augmented sequential blocks consisting of Batch Normalization followed by ReLU and Convolution Layers also one FC layer is also added at end. Modified architecture of used model is given table3.2 where
CONV2D(input channel,output channel,kernal size) shows convolution layer structure.

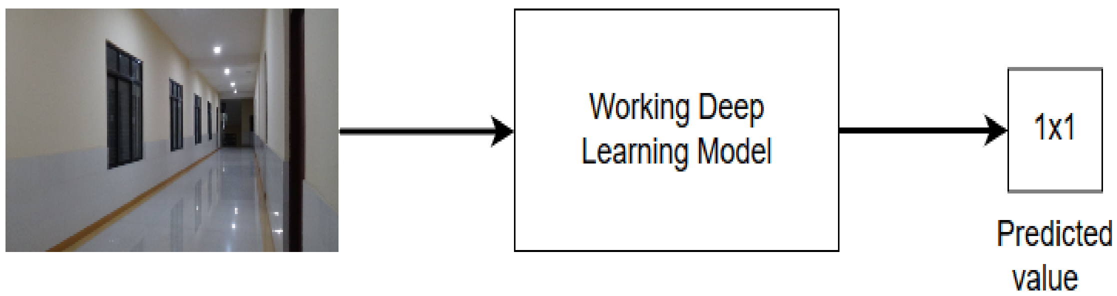


Figure 3.26: Use deep learning model for solving regression task

Table 3.2: Architecture of used modified deep learning models

Pre-trained Models	**Augmented Convolution Layers**	**Augmented last Fully connected layer**	**Output layer**
Alex Net	not adding convolution layer	4096x1	1x1
VGG-16 Net	CONV2D(512,1024,1x1) CONV2D(1024,128,5x5) CONV2D(128,16,1x1)	96x1	1x1
Res-101 Net	CONV2D(2048,1024,1x1) CONV2D(1024,128,5x5) CONV2D(128,8,1x1)	96x1	1x1
Res-152 Net	CONV2D(2048,1024,1x1) CONV2D(1024,128,5x5) CONV2D(128,8,1x1)	96x1	1x1
Inception Net	**Main:**CONV2D(2048,1024,1x1) CONV2D(1024,512,2x2) CONV2D(512,128,3x3) **Aux:**CONV2D(768,128,4x4) CONV2D(128,32,2x2)	**Main:**256x1 **Aux:**640x1	**Main:**1x1 **Aux:**1x1
Dense-201 Net	CONV2D(1920,1024,1x1) CONV2D(1024,128,5x5) CONV2D(128,16,1x1)	96x1	1x1
Dense-161 Net	CONV2D(2208,1024,1x1) CONV2D(1024,128,5x5) CONV2D(128,16,1x1)	96x1	1x1

3.7 Experiment for finding angle between bisector and horizontal axis of plane

The angle between bisector and horizontal axis of plane help us to find position of drone in corridor. We use Mean Absolute loss as functions for train our network .With the help of these function compute the loss of every batches and back propagate into to the network to update the the weights of network.

- We use 35000 images for training and 600 images for test
- We use three loss functions for training

3.7.1 Mean Absolute Loss

Mean Absolute Loss is a measure of difference between two continuous values

$$Mean\ Absolute\ Loss(\hat{y}, y) = \frac{1}{n}\sum_{t=1}^{n} |\hat{y_t} - y_t|$$

Where ,

- $\hat{y}$ are predicted value vector of batch
- y are ground-truth value vector of batch
- $\hat{y}_t$ is the prediction made on t^{th} training example.
- y_t is the ground-truth value on t^{th} training example.
- n number of training examples

3.7.2 Convergence Graphs of Deep Learning models using Mean Absolute Loss

Convergence graph of different model trained in our experiment is shown below,where x-axis describe number of epochs and y-axis describe the training loss.

- **Convergence Graph of training on Alex Net**

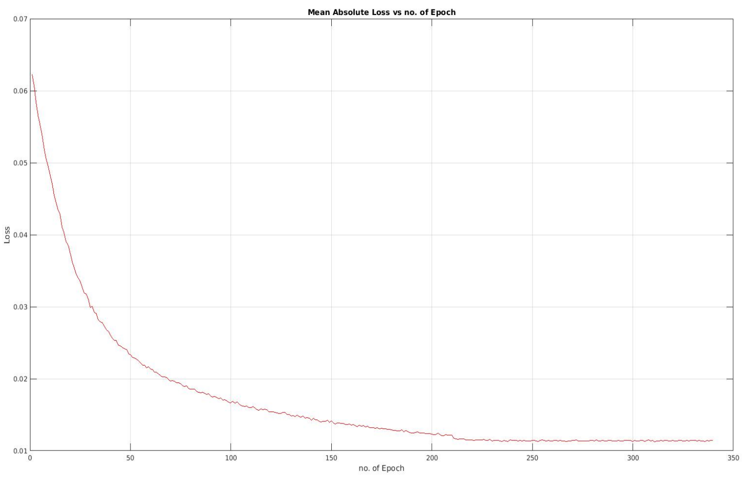


Figure 3.27: Convergence Graph of Alex Net

- **Convergence Graph of training on VGG-16 Net**

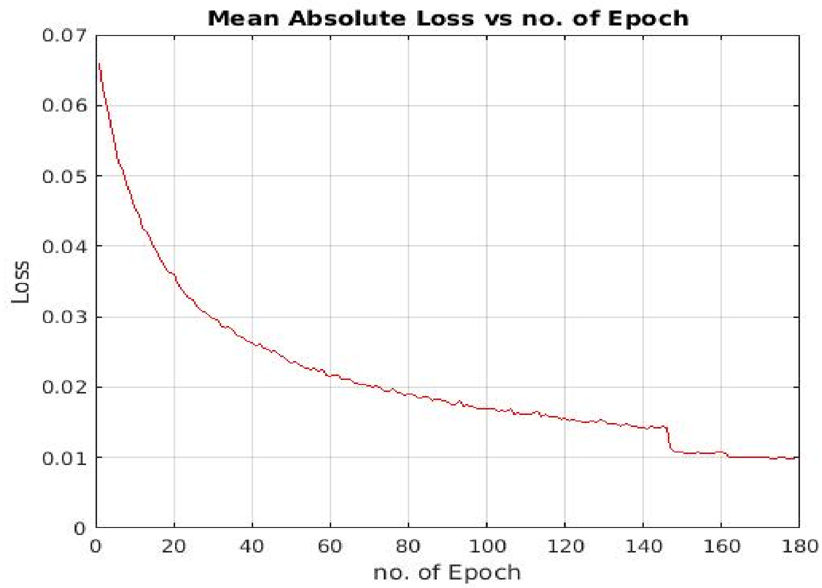


Figure 3.28: Convergence Graph of VGG-16 Net

- **Convergence Graph of training on Res-101 Net**

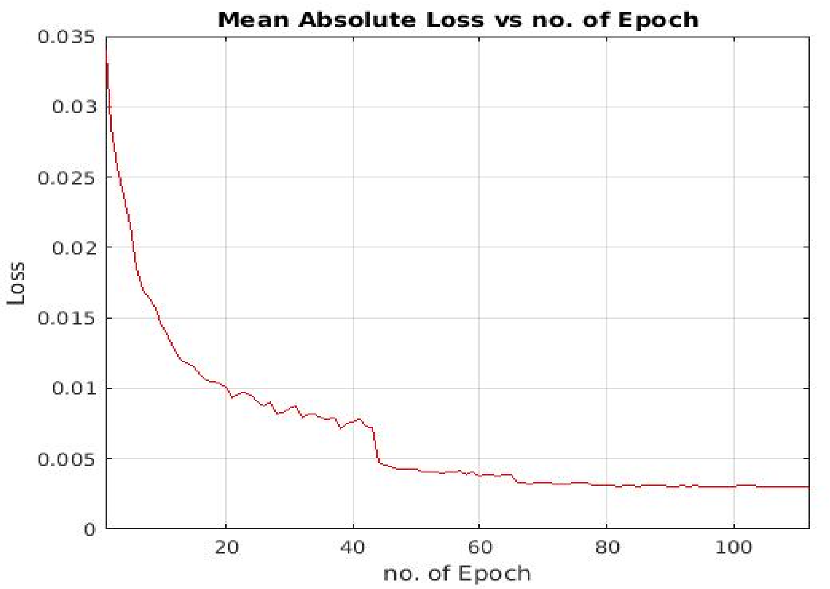


Figure 3.29: Convergence Graph of Res-101 Net

- **Convergence Graph of training on res-152 Net**

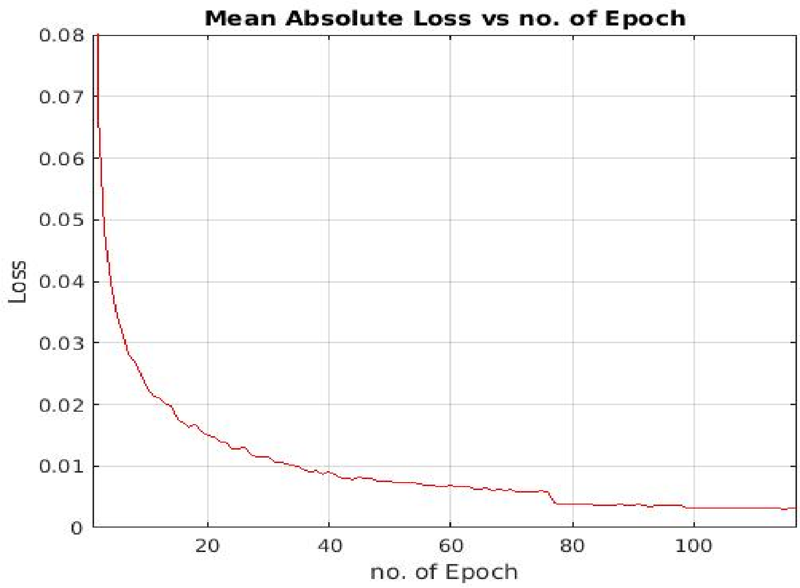


Figure 3.30: Convergence Graph of res-152 Net

- **Convergence Graph of training on Inception Net**

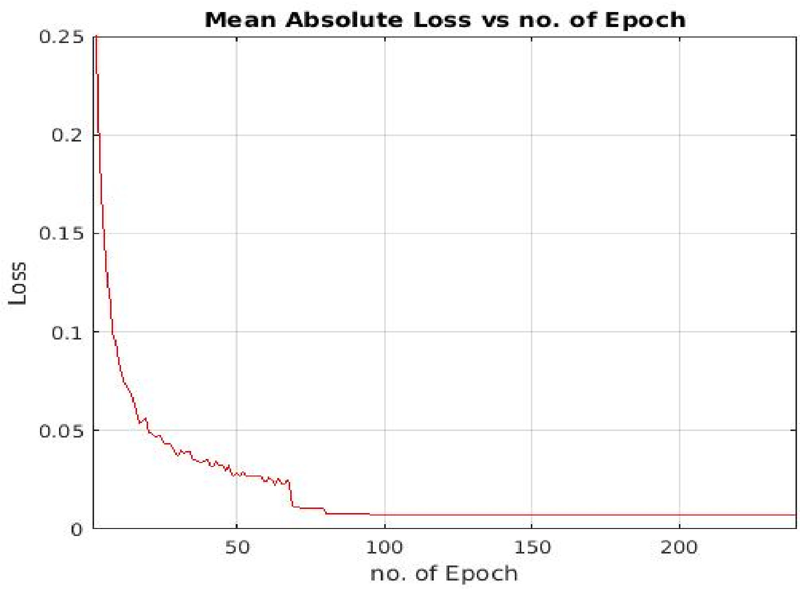


Figure 3.31: Convergence Graph of Inception Net

- **Convergence Graph of training on Dense-201 Net**

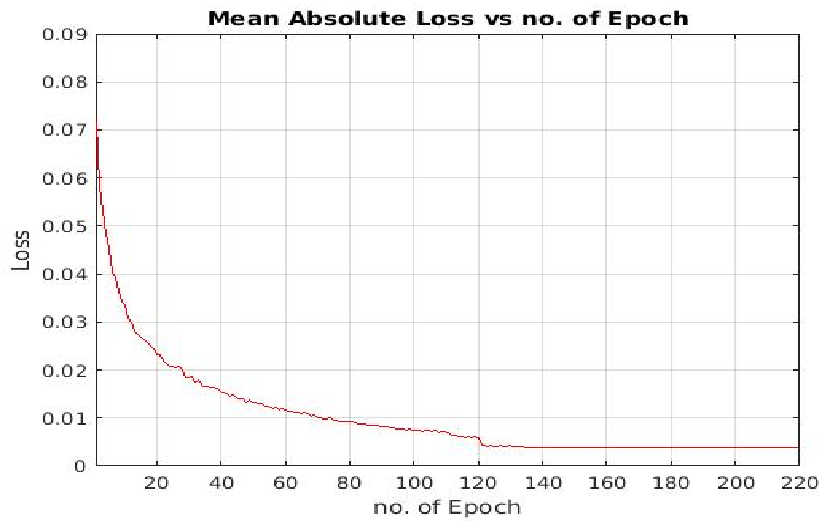


Figure 3.32: Convergence Graph of Dense-201 Net

- **Convergence Graph of training on Dense-161 Net**

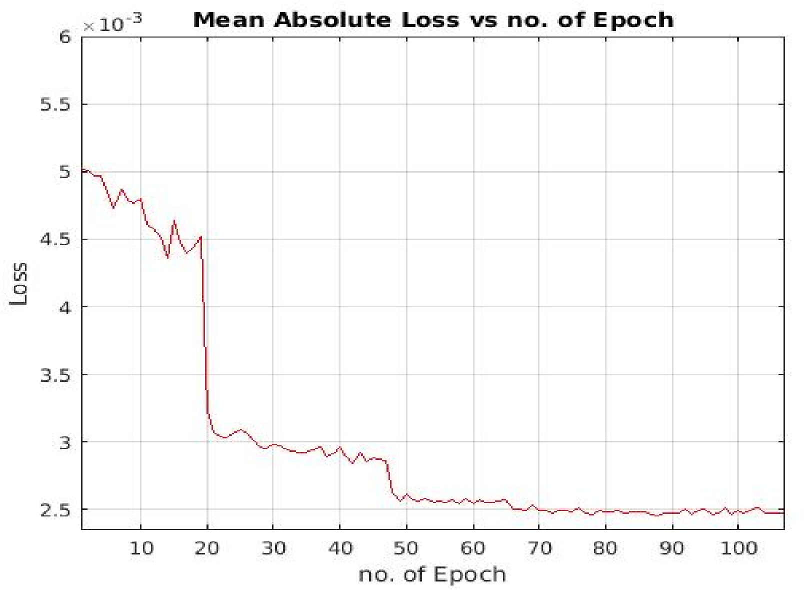


Figure 3.33: Convergence Graph of Dense-161 Net

3.7.3 Test Results

We use three error metric for measure the accuracy of model on test images.The error metrics are Mean Square Error,Mean Absolute Error and Mean Relative Error.

$$Mean\ Squared\ Error(\hat{y}, y) = \frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2$$

$$Mean\ Absolute\ Error(\hat{y}, y) = \frac{1}{n}\sum_{t=1}^{n} |\hat{y}_t - y_t|$$

$$Mean\ Relative\ Error(\hat{y}, y) = \frac{1}{n}\sum_{t=1}^{n} \frac{|\hat{y}_t - y_t|}{y_t}$$

Where ,

- $\hat{y}$ are predicted value vector of batch
- y are ground-truth value vector of batch
- $\hat{y}_t$ is the prediction made on t^{th} test example.
- y_t is the ground-truth value on t^{th} test example.
- n number of test examples

Mean Square Error,Mean Absolute Error and Mean Relative Error for testset of all used deep learning model shown in table3.3 Comparison between

Table 3.3: Error in test set angle prediction

Result for angle prediction in Degree			
Pre-trained Model	**Mean Square Error**	**Mean Absolute Error**	**Mean Relative Error**
Alex Net	0.21997	1.72831	1.39280
VGG-16 Net	0.47318	2.84597	2.41795
Res-101 Net	0.11103	1.46875	1.17946
Res-152 Net	0.11032	1.41558	1.16529
Inception Net	0.11929	1.44906	1.20959
Dense-201 Net	0.12383	1.79144	1.42709
Dense-161 Net	**0.05791**	**1.32693**	**1.08712**

models shown in bar graph3.34 on basis of Mean Absolute Error.
Result of some selected test images shown in table3.4

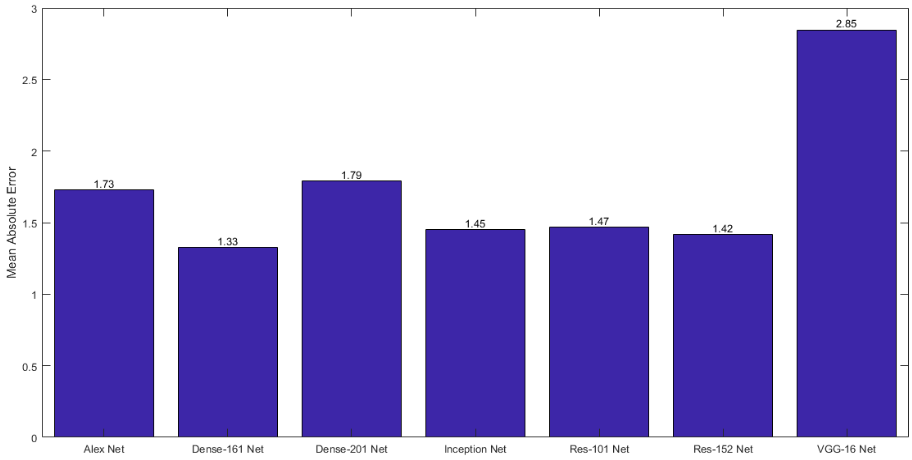
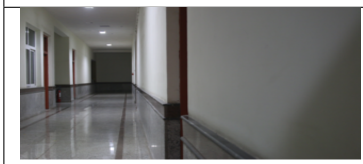
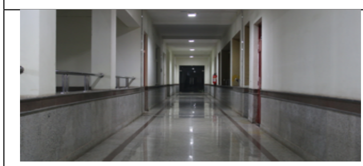
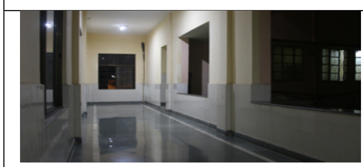


Figure 3.34: Comparison between models on basis of Mean Absolute Error on test images

Table 3.4: Results of angle prediction some selected images in degree

Images	**Location of corridor**	**Actual angle**	**Predicted angle**
	**TIIR Building**	**132.677**	**132.293**
	**Physics Department**	**47.055**	**46.707**
	**Life Science Department**	**90.000**	**91.142**
	**New Building CSE Department**	**136.507**	**135.253**
	**TIIR Building**	**90.720**	**90.303**
	**New Building CSE Department**	**48.366**	**48.413**

3.8 Time Comparison And Parameter of Model Used

Time taken by one image in the models in seconds shown in table 3.5

Table 3.5: Time taken by different model on same parameter on CPU and GPU

Time Comparison of Models in seconds			
Model Name	**CPU TIME (sec)**	**GPU TIME (sec)**	**Parameters**
Alex Net	0.09164	0.00264	57007937
VGG-16 Net	0.87111	0.00417	138357544
Res-101 Net	0.51802	0.02673	44549160
Res-152 Net	1.31313	0.03809	60192808
Inception Net	0.46774	0.02731	30720450
Dense-201 Net	0.49752	0.05514	23345297
Dense-161 Net	0.58392	0.04647	32019857

3.9 Summary

In this chapter we discussed localization algorithm, dataset creation and sea used modified deep learning models architecture.In experiment part we saw training of deep learning models using transfer learning and result on test set.By seeing testing result on test set we conclude Dense-161 Network architecture give minimum error in all among networks.

Chapter 4

Conclusion

This thesis focuses on drone localization in GPS-denied indoor corridor environment. The research includes design of localization algorithm for drones.This research also includes the design of a fully convolutional deep architecture for estimating the angle of bisector of corridor in image plane which help us to detect the position of drone. The deep learning architecture facilitates end-to-end training and improves the overall performance and reduce overhead of feature extraction which use in the conventional machine learning.

In Chapter 1, a brief introduction of Unmanned Aerial Vehicles (UAV), challenges to autonomous UAV navigation, research motivation with goals and objective are discussed.

In Chapter 2, we discussed about deep learning basics, component of deep learning architecture, transfer learning and recent deep learning models which are use in our research to estimation of position of drone in corridors.

In Chapter 3, we discussed about our proposed work. The first phase of my research work is designed localization algorithm which helps us to find position of drone in flat indoor corridor environment. In our algorithm we take help of bisector of corridor floor to find position of drone. The angle between bisector and horizontal axis of image defined the position of drone.

The second phase of our research work is creation of dataset. There is no benchmark data set of our project work so we need to create dataset for our project. For creation of dataset we capture images of all possible position of drone in corridor. Our dataset covers around 100 corridor of NIT Rourkela and it consist 35000 images for training.

The third phase of my work is designing of deep learning models. We use help of transfer learning that is take deep learning network already trained with Imagenet challenge dataset and further trained with our dataset. In result we can see Dense-161 Network give best result in comparison with other deep learning network .

Scope for Further Research

The research findings out of this thesis have shown multiple promising directions for further investigations.

- We work on drone localization in flat corridor it can be enhance for curve corridor.
- Our algorithm can be use as a autonomous navigation algorithm with boundary condition. Boundary is end of corridor and what happen if corridor turn.
- We can also enhance the proposed algorithm with obstacle detection in corridor.

References

[1] Krizhevsky, A., Sutskever, I., and Hinton, G. E., 2012. "Imagenet classification with deep convolutional neural networks". In Advances in neural information processing systems, pp. 1097–1105.

[2] Zeiler, M. D., and Fergus, R., 2014. "Visualizing and understanding convolutional networks". In European conference on computer vision, Springer, pp. 818–833.

[3] He, K., Zhang, X., Ren, S., and Sun, J., 2016. "Deep residual learning for image recognition". In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

[4] Huang, G., Liu, Z., Weinberger, K. Q., and van der Maaten, L., 2017. "Densely connected convolutional networks". In Proceedings of the IEEE conference on computer vision and pattern recognition, Vol. 1, p. 3.

[5] Simonyan, K., and Zisserman, A., 2014. "Very deep convolutional networks for large-scale image recognition". *arXiv preprint arXiv:1409.1556*.

[6] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., et al., 2015. "Going deeper with convolutions". Cvpr.

[7] Bristeau, P.-J., Callou, F., Vissiere, D., and Petit, N., 2011. "The navigation and control technology inside the ar. drone micro uav". *IFAC Proceedings Volumes,* **44**(1), pp. 1477–1484.

[8] de Babo Martins, F., 2015. "Visual-inertial based autonomous navigation of an unmanned aerial vehicle in gps-denied environments".

[9] Surmann, H., Lingemann, K., Nüchter, A., and Hertzberg, J., 2001. "A 3d laser range finder for autonomous mobile robots". In Proceedings of the 32nd ISR (International Symposium on Robotics), Vol. 19, pp. 153–158.

[10] Engel, J., Sturm, J., and Cremers, D., 2012. "Camera-based navigation of a low-cost quadrocopter". In Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on, IEEE, pp. 2815–2821.

[11] Doncieux, S., Mouret, J.-B., Muratet, L., and Meyer, J.-A., 2004. "The robur project: towards an autonomous flapping-wing animat". *Proceedings of the Journes MicroDrones, Toulouse*.

[12] Michelson, R. C., and Reece, S., 1998. "Update on flapping wing micro air vehicle research-ongoing work to develop a flapping wing, crawling entomopter". In 13th Bristol International RPV/UAV Systems Conference Proceedings, Bristol England, Vol. 30, pp. 30–1.

[13] Cai, G., Lum, K.-Y., Chen, B. M., and Lee, T. H., 2010. "A brief overview on miniature fixed-wing unmanned aerial vehicles". In Control and Automation (ICCA), 2010 8th IEEE International Conference on, IEEE, pp. 285–290.

[14] Ferreira, F. O. R. T., 2014. "Video analysis in indoor soccer with a quadcopter".

[15] Bry, A., Bachrach, A., and Roy, N., 2012. "State estimation for aggressive flight in gps-denied environments using onboard sensing". In Robotics and Automation (ICRA), 2012 IEEE International Conference on, IEEE, pp. 1–8.

[16] Roberts, J. F., Stirling, T., Zufferey, J.-C., and Floreano, D., 2007. “Quadrotor using minimal sensing for autonomous indoor flight”. In European Micro Air Vehicle Conference and Flight Competition (EMAV2007), no. LIS-CONF-2007-006.

[17] Checchin, P., Gérossier, F., Blanc, C., Chapuis, R., and Trassoudaine, L., 2010. “Radar scan matching slam using the fourier-mellin transform”. In Field and Service Robotics, Springer, pp. 151–161.

[18] Engel, J., Schöps, T., and Cremers, D., 2014. “Lsd-slam: Large-scale direct monocular slam”. In European Conference on Computer Vision, Springer, pp. 834–849.

[19] Mei, C., Sibley, G., Cummins, M., Newman, P., and Reid, I., 2011. “Rslam: A system for large-scale mapping in constant-time using stereo”. *International journal of computer vision,* **94**(2), pp. 198–214.

[20] Çelik, K., and Somani, A. K., 2013. “Monocular vision slam for indoor aerial vehicles”. *Journal of electrical and computer engineering,* **2013**, pp. 4–1573.

[21] Bachrach, A., He, R., and Roy, N., 2009. “Autonomous flight in unknown indoor environments”. *International Journal of Micro Air Vehicles,* **1**(4), pp. 217–228.

[22] Fraundorfer, F., Heng, L., Honegger, D., Lee, G. H., Meier, L., Tanskanen, P., and Pollefeys, M., 2012. “Vision-based autonomous mapping and exploration using a quadrotor mav”. In Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on, IEEE, pp. 4557–4564.

[23] Achtelik, M., Bachrach, A., He, R., Prentice, S., and Roy, N., 2009. “Stereo vision and laser odometry for autonomous helicopters in gps-denied indoor environments”. In Unmanned Systems Technology XI, Vol. 7332, International Society for Optics and Photonics, p. 733219.

[24] Bills, C., Chen, J., and Saxena, A., 2011. “Autonomous mav flight in indoor environments using single image perspective cues”. In Robotics and automation (ICRA), 2011 IEEE international conference on, IEEE, pp. 5776–5783.

[25] Pomerleau, D. A., 1989. “Alvinn: An autonomous land vehicle in a neural network”. In Advances in neural information processing systems, pp. 305–313.

[26] Ross, S., Melik-Barkhudarov, N., Shankar, K. S., Wendel, A., Dey, D., Bagnell, J. A., and Hebert, M., 2013. “Learning monocular reactive uav control in cluttered natural environments”. In Robotics and Automation (ICRA), 2013 IEEE International Conference on, IEEE, pp. 1765–1772.

[27] Kim, D. K., and Chen, T., 2015. “Deep neural network for real-time autonomous indoor navigation”. *arXiv preprint arXiv:1511.04668*.

[28] LeCun, Y., et al., 2015. “Lenet-5, convolutional neural networks”. *URL: http://yann. lecun. com/exdb/lenet*, p. 20.

[29] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al., 2015. “Imagenet large scale visual recognition challenge”. *International Journal of Computer Vision,* **115**(3), pp. 211–252.