Multiclass Geospatial Object Detection using Machine Learning-Aviation Case Study

Durga Prasad Dhulipudi(PhD)
LSI
IIIT-H
Hyderabad, India
durgaprasad.d@research.iiit.ac.in

Dr Rajan KS PhD,Tokyo University
LSI
IIIT-H
Hyderabad, India
rajan@iiit.ac.in

***Abstract*— There is growing interest to explore the autonomous taxiing that can sense its environment and maneuver safely with little or no human input. This technology is like the one developed for driver less cars that synthesize information from multiple sensors, which sense surrounding environment to detect road surface, lanes, obstacles and signage. This paper presents application of computer vision and machine learning to autonomous method for the surface movement of an air vehicle. We present a system and method that uses pattern recognition which aids unmanned aircraft system (UAS) and enhance the manned air vehicle landing and taxiing. Encouraged with our previous results [1], we extend upon our research to include multiple object relevant to taxiing. The objective of the current project is to build training dataset of annotated objects acquired from overhead perspective. It is useful for training a deep neural network to learn to detect, count specific airport objects in a video or image. This paper details the procedure and parameters used to create training dataset for running convolutional neural networks (CNNs) on a set of aerial images for efficient and automated object recognition. In this method, multiple airport surface signage dataset from satellite images are subjected to training for pattern recognition. This trained system learns and then identifies and locates important visual references from imaging sensors and could help in decision making during taxiing phase.**

Keywords—auto-land, air vehicle, UAV, UAS, Machine Learning, Computer Vision, CNN, Deep Learning, landing, taxi, Runways

I. Introduction

Compared to other phases ,landing is the most demanding flight phase for both manned and unmanned aerial vehicles .It is also the most accident-prone stage for manned, remotely piloted and autonomous aircrafts. Auto-landing systems were introduced to reduce the workload on pilot during this critical phase. These systems based on the Instrument Landing System (ILS) have already proven their importance through decades. In an IFR auto-land systems were designed to make landing possible in visibility too poor to permit any form of visual landing, although they can be used at any level of visibility. They are usually used when visibility is less than 600 meters' runway visual range and/or in adverse weather conditions, although limitations do apply for most aircraft. The auto-land systems work in conjunction with radio altimeter, Inertial Navigation System (INS), ILS, MLS or Global Navigation Satellite System (GNSS). Closer to the runway, both under VFR or IFR, pilots are expected to rely on the visual references for landing. Modern systems like HUD or Combined Vision System (CVS) allows a trained pilot to manually fly the aircraft using guidance cues from the flight guidance system. This significantly reduces the cost of operating in very low visibility, and allows aircraft that are not equipped for automatic landings to make a manual landing safely at lower levels of look ahead visibility or runway visual range (RVR).

Notwithstanding the type of landing and instruments used, typically Pilots are expected to have the runway threshold markings, aiming point, displacement arrows , touch down point, helipad markings/lights in sight before Minimum Decision altitude (MDA).

Modern imaging sensors are integrated with displays to provide the pilots with better images than unaided human vision. With benefits provided by such systems like Enhanced Flight Vision Systems (EFVS) and Combined Vision System (CVS), there has been a growing interest in making use of vision as important apparatus for autonomous landing of UAS. Weiwei Kong et al [5] presents the main research groups involved in the development of vision-based autonomous landing systems. In these problems, different types of visual features were considered including geometric model of the target, points, corners of the runway, binormalized Plücker coordinates, the three parallel lines of the runway (at left, center and right sides) and the two parallel lines of the runway along with the horizon line and the vanishing point. Rather than points, lines or other features, researchers from IST/TU Lisbon [6] demonstrated the dense information could also be applied into vision-based landing system. Cesetti [7] et al. proposed a vision-based guide system for UAV navigation and safe landing using natural landmarks. This method might not be effective when the real-time image was much different from the reference image and its performance depends on matching algorithm.

II. OBJECTIVE

Object detection in machine learning makes use of Deep Learning for object classification, detection and instance segmentation. Deep learning is a subset of machine learning in Artificial Intelligence (AI) that has networks which are capable of learning by training on annotated data.. In our previous study [1], our objective was to create a dataset of single object in perspective view and apply the Machine Learning to detect the object in the perspective view. The objective of the current study is to build training dataset of annotated objects acquired from overhead perspective for multiple surface markings. We picked up those markings that are used by pilots for decision making in important landing and taxiing phases.

This paper details the procedure and parameters used to create training dataset for running convolutional neural networks (CNNs) on a set of aerial images for efficient and automated object recognition.

This part of activity is precursor to the overall research objective of pattern detection for autonomous decision-making. The deep learning model created then identifies and locates important visual references from imaging sensors and could help in decision making during landing and taxiing.

III. SCOPE

This study limits the scope to provide an input for landing and taxiing system. Specifically, the study is restricted to below mentioned aspects:

- Creation of dataset from satellite images
- Apply *transfer learning* technique of Machine Learning on these images and create object detection inference model
- Verify if this model is suitable to detect airport objects from images acquired from imaging sensors

This study does not intend to cover:

- System of autonomous landing
- Aero dynamics of landing and mechanism of autonomous taxiing

IV. DATA USED

QGIS-Open Layers plugin is used to access high resolution satellite and aerial imagery base map. Polygon boundaries of the objects of interest are generated from the vector data available at FAA website .Negative training set is created from the declassified New York state GIS data clearing house link .The dataset has the following attributes:

1) Data from overhead at about 1 meter per pixel resolution at ground.
2) Data from 1025 distinct locations in United States and Europe.
3) We collected about 800 images comprising of 4 specific patterns that are most important in landing and taxiing phase.

V. METHODOLOGY

Transfer learning or inductive transfer is a research problem in machine learning that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem. For example, knowledge gained while learning to recognize cars could apply when trying to recognize trucks.

While we continued with Faster RCNN [2] as in our previous study, we moved to PyTorch Deep learning framework and Detectron2 [3] instead of Tensorflow Object detection framework . We found this new ML framework more contemporary and has greater ease of use and implementation in addition to providing us a consistent tool for mask recognition and our future research. Detectron2 is Facebook AI Research's next generation software system that implements state-of-the-art object detection algorithms. It is a ground-up rewrite of the previous version, Detectron, and it originates from Mask R-CNN [4]-benchmark.

We continued the use ***transfer learning*** method on to create training dataset for multiple airport surface marking detection.

VI. STUDY AREA

We originally focus on the aviation domain, namely airport feature detection and classification. We view this problem as a compelling use case that also provides insights into semi-automatic feature extraction in sub meter resolution satellite imagery. Therefore, data from 1025 distinct locations in United States and Europe from over 400 airports is used in this study. This is like the area that we used for our previous study.

VII. PREPARING GEOSPATIAL DATA FOR MACHINE LEARNING

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. It focuses on the development of computer programs that can access data and use it learn for tasks like object classification, clustering, pattern recognition, localization, object detection and segmentation. The process of learning starts from providing relevant annotated data to ML algorithms, which find relationships, develop understanding, make decisions, and evaluate their confidence from the training data they're given. Such trained models have ability to generalize to previously unseen data. It is critical that we feed them the right data for the problem we want to solve. The success of any ML algorithm depends on the quality and quantity of training data. The process for providing data ready for object detection task using ML algorithms can be summarized as below sequence of steps:

A. Data Collection

This step is concerned with selecting the available data that we will be working with. For the purposes of this project, we could not get the dataset from public sources directly in image format. QGIS-Open Layers plugin is used to access high resolution satellite and aerial imagery base map. The data is free for academic purpose and research. Polygon boundaries of the specific objects of interest are generated from the vector data available at *Federal Aviation Administration - AIS Open Data* website. Additionally, several locations are annually digitized to achieve desired variability expected by ML algorithms.

For threshold classification purpose, we created a repository of 800 images- 200 for each of the 4 categories of interest at various zoom levels scales [Figure 1 **Error! Reference source not found.**].

Figure 1 Snapshot of the training data for 4 categories (QGIS Open Layers)

B. Data Preprocessing

This preprocessing step involves data cleaning, sampling, labeling, and splitting into training and test sets. If the data is collected from diverse source and in various formats, it is required to convert them into standard image format like JPEG or PNG depending on the ML library being used. In this case, we used RGB images in JPEG format. It is very likely that the machine learning tools we use on the data will influence the preprocessing we are required to perform. For example, In our previous study ,we usedTensorflow Object Detection accepts labels for classification in PASCAL VOC XML In this study ,we moved to COCO JSON format suitable for PyTorch mask detection.

Data cleaning requires removing some images that are not representative of selected pattern in terms of visual appearance though their geographical location is correct.

Sampling is the process of selecting representative subset of the available dataset. It involved visually inspecting each sample and weeding out objects that are too similar. The exercise is essential to avoid overfit of the training dataset. Our sampling process followed data augmentation guidelines prescribed to prevent overfitting . Thus, images that have good mixture of varied backgrounds, rotations, lighting, colors, textures, occlusions etc. are part of the sampled data [Figure 2].

Figure 2 Snapshot of sample representativeness among 4 categories

Labeling is the process of annotating objects of interest for classification training. Since Object Detection not only classifies the object but detects the location of an object in the image, we need to pass it a bounding box identifying where the object is in the image and the label associated with that bounding box. In our previous study, we just added a bounding box around the object. In the current study, we used masks or contours that represent the shape of the object.

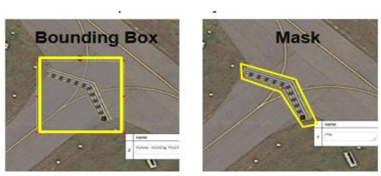

Figure 3 Bounding box and Mask

We used VGG Image Annotator labeling tool, a simple and standalone manual annotation software for image, audio and video. VIA runs in a web browser and does not require any installation or setup. The bounding boxes and masks are represented by a series of vertices saved in COCO json format which is de-facto for storing annotation data.

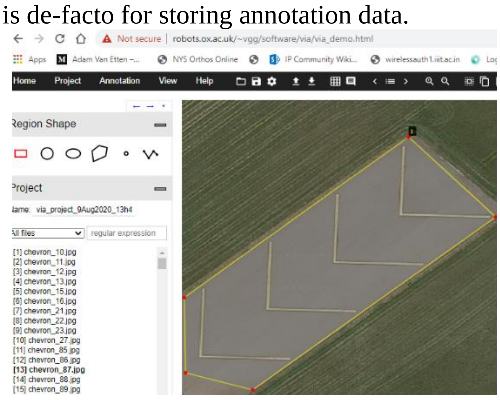

Figure 4 Annotation with LabelImg

To ensure robustness data is split into training and testing sets. Training data is used to train an algorithm. Generally, training data is a certain percentage of an overall dataset along with testing set. As a rule, the better the training data, the better the algorithm or classifier performs. Once a model is trained on a training set, it's usually evaluated on a test set to check if there is overtraining or model is generalized well.

To explore the transfer learning and ascertain the confidence in implementing the method for the airport object detection, we took a smaller representative sample of the selected data. Larger datasets often result in longer running times for algorithms, longer computational and memory requirements.

C. Data Formatting

In machine learning, we typically obtain the data and ensure that it is formatted before starting the training process. PyTorch Object Detection API is used for this study, which expect the images and annotation in COCO json format. All the images and annotations are converted into this format. More images are created dynamically by the process of data augmentation via training configuration file.

D. Prvious Study with Single Object

In phase1 and phase 2 of our previous study, we could reliably identify the bounding boxes at confidence levels exceeding 94% for overhead images even in case of in case of highly unclear ambiguous images **Error! Reference source not found.**

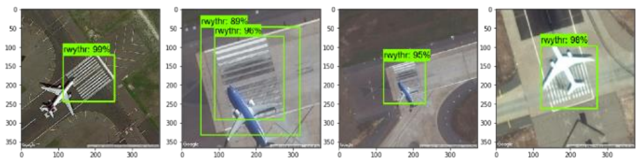

Figure 5 Object detection in occlusion

Similar level of confidence levels was noticed in case of highly unclear ambiguous images -Figure 6.

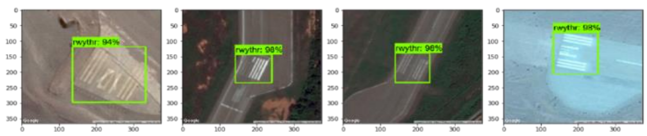

Figure 6 Object detection in ambiguous scenarios

Our hypothesis of training the model with overhead images and detecting for perspective images Figure *7* is validated.

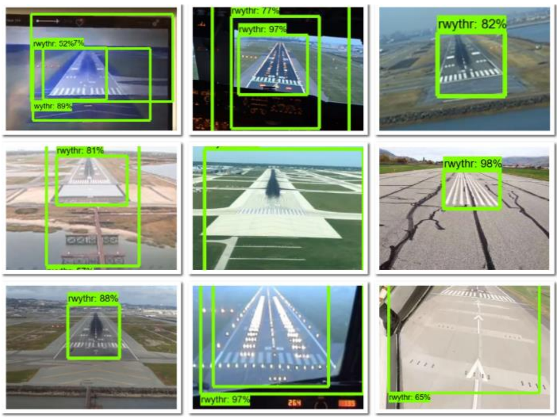

Figure 7 Object detection from aircraft landing/taxiing videos (YouTube)

Encouraged with the results obtained, our research continued to expand the study to include multiple object detection; Improve the model to obtain tight boundaries in the perspective images; Create a dataset of airport objects that comprises of images, annotations and masks.

VIII. Object Detection , classification AND SEGEMENTATION

This study included multiple airport surface markings as continuity to previous study. Equipped with the data and format that is suitable for training using PyTorch, the experiment for object detection and classification and segmentation is carried out on Google Colab Machine Learning environment with two GPUs.

A. Current Study with multiple airport surface markings

With the results obtained in our previous study, our research expanded to include:

- Create a dataset of airport objects that comprises of images, annotations and masks.
- Multiple object detection for airport surface marking recognition

While only Runway thresholds location sign pattern was used in our previous study, we extended our data collection to include 4 patterns -Runway Threshold, Chevron, ILS Holding Position and Runway Holding Position Markings.

Table VIII-1 shows the distribution of instances among all 4 categories. For object detection, classification and segmentation purpose, we created a repository of images of 1:500 and 1:1200 scales.

Table VIII-1 DISTRIBUTION OF INSTANCES AMONG ALL 4 CATEGORIES

Category	#Instances
Runway Threshold	200
Chevron	206
ILS Holding Position	200
Runway Holding Position	199

B. System Configuration and Machine Learning environment

Table VIII-2 summarizes the results obtained in Phase1 with hardware and software configuration details.

Table VIII-2 SYSTEM CONFIGURATION AND TRAINING DATA

Category	Property
Methodology	Transfer Learning
Image Segmentation Algorithm	Detectron2 with Mask RCNN 101 FPN 3x configuration
Number of Images	800 in toal , 200 for each class
Data Augmentation	Yes
Training Epochs	3000
Training Time	25 minutes
Average inference time	0.36 seconds and fps:2.77

C. Detecting objects captured overhead perpective

The Faster-RCNN model that is retrained for this custom data is now applied on airport images acquired from overhead perspective for inference. Our results show that for almost all test images, the bounding boxes have a confidence level of more than 98% for all 4 surface pattern categories. Figure *8* illustrates the output of the inferences made by the model developed to detect these 4 surface marking patterns.

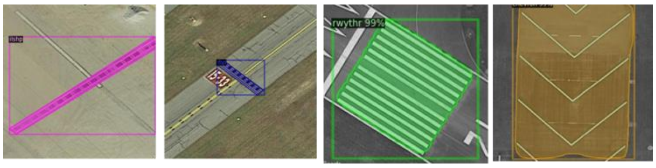

Figure 8 Bounding box and Mask detection of all 4 patterns

To test the model inference, we selected perspective images that have high variability in terms of pavement type, surroundings, distance of the pattern, visibility and wear and tear. **Error! Reference source not found.** illustrates that bounding boxes and masks are accurate for images that have good mixture of varied backgrounds, rotations, lighting, colors, textures, occlusions etc.

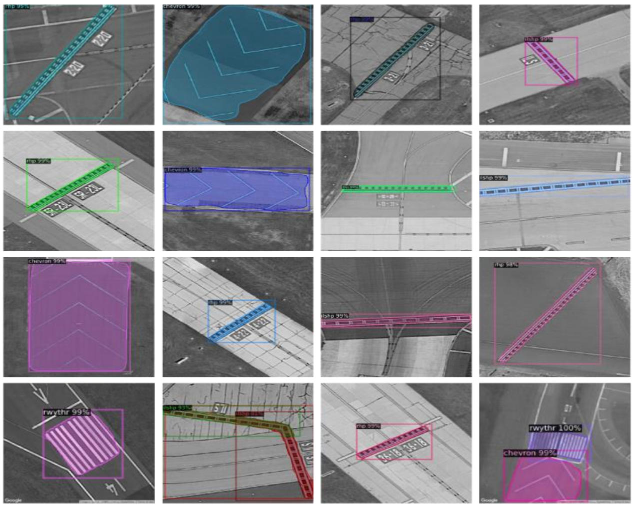

Figure 9 Detection in disparate images

D. *Detecting objects captured from aviation imaging sensors*

The Faster-RCNN model that is retrained for this custom data is now applied on airport images acquired from imaging sensors for inference. Notice that these images are snapshots from publicly available aircraft landing and taxiing videos, typically that are shot from onboard imaging sensors.

We selected several samples from landing and taxiing videos that has the 4 selected patterns visible. **Error! Reference source not found.** illustrates the output of the inferences made by the model developed to detect these 4 surface marking patterns. To test the model inference, we selected perspective images that have high variability in terms of pavement type, surroundings, distance of the pattern, visibility and wear and tear.

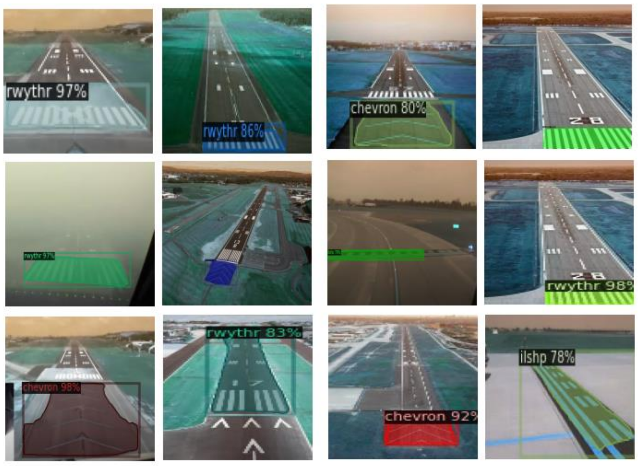

Figure 10 Object detection and Mask detection from aircraft landing/taxiing videos (YouTube)

E. *Model Evaluation Metrics*

We used metrics chosen for popular competitions like COCO Object Detection Challenge. The evaluation is based on a measure called Intersection over Union (IoU). Also, referred to as the Jaccard Index, IoU is an evaluation metric that quantifies the similarity between the ground truth bounding box and the predicted bounding box to evaluate how good the predicted box is. The IoU score ranges from 0 to 1, the closer the two boxes, the higher the IoU score. Formally, the IoU measures the overlap between the ground truth box and the predicted box over their union.

The surface pattern ground truth and predicted bounding per IoU metric are per Figure *11* .

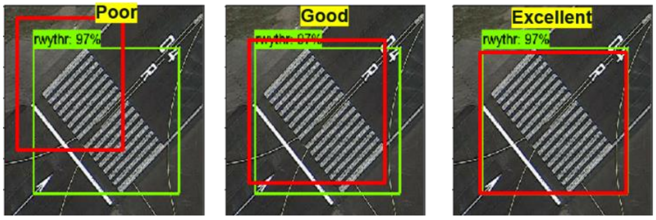

Figure 11 Intersection Over Union (IoU)

The COCO Object Detection Challenge evaluates detection using 12 metrics where: mAP (interchangeably referred to in the competition by AP) is the principal metric for evaluation in the competition, where AP is averaged over all 10 thresholds and all 80 COCO dataset categories. This denoted by AP@ [.5: .95] or AP@ [.50: .05: .95] incrementing with .05. Hence, a higher AP score per the COCO evaluation protocol indicates that detected objects are perfectly localized by the bounding boxes. Additionally, COCO individually evaluates on AP at 0.5 and 0.75 IoU thresholds, this is denoted AP@.50 or AP^{IOU=0.50}
IOU=0.50 and AP@.75 or AP^{IOU=0.75}
IOU=0.75 respectively. Table VIII-3 has the evaluation results for bounding boxes detected by the model trained to detect the 4 categories under consideration.

Table VIII-3Evaluation results for bounding box

Metric	Value
AP	85.0441
AP-Chevron	85.87187
AP-ILS Holding Position	83.51583
AP-Runway Holding Position	79.71215
AP-Runway Threshold	91.07654
AP50	99.10995
AP75	98.87533
APl	85.03289
APm	80.58168

Table VIII-4 summarizes the evaluation results for bounding boxes detected by the model trained to detect the 4 categories under consideration.

Table VIII-4 Evaluation metrics for segmentation boundary

Metric	Value
AP	79.30977
AP-Chevron	80.82801
AP-ILS Holding Position	74.35408

Metric	Value
AP-Runway Holding Position	71.09814
AP-Runway Threshold	90.95885
AP50	99.22977
AP75	96.00683
APl	79.7537
APm	53.99485

FUTURE WORK

Our previous study proved to detect the trained pattern from perspective images, our hypothesis is validated for single object. The current study extended to study the same for multiple surface markings. Encouraged with the results obtained, our future research intends to expand the study to include multiple object detection in videos; Improve the model to obtain tight boundaries in the perspective images; Experiment on automatic decision making choices or opportunities with this proof.

REFERENCES

[1] Durga Prasad Dhulipudi& Dr Rajan KS ,Geospatial Object Detection using Machine Learning-Aviation Case Study, DASC 2019-20

[2] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

[3] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo and Ross Girshick, Detectron2. https://github.com/facebookresearch/detectron2, 2019.

[4] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 2980-2988, doi: 10.1109/ICCV.2017.322.

[5] Weiwei Kong , Dianle Zhou , Daibing Zhang and Jianwei Zhang, "Vision-based autonomous landing system for unmanned aerial vehicle: A survey",2014 International Conference on Multisensor Fusion and Information Integration for Intelligent Systems (MFI).

[6] T. Goncalves, J. Azinheira, and P. Rives, "Vision-Based Automatic Approach and Landing of Fixed-Wing Aircraft Using a Dense Visual Tracking," , Informatics in Control Automation and Robotics, pp. 269-282..

[7] Cesetti, A., Frontoni, E., Mancini, A. et al. J Intell Robot Syst (2010) 57: 233. https://doi.org/10.1007/s10846-009-9373-3

[8] O. M. Parkhi, A. Vedaldi, A. Zisserman, C. V. Jawahar ;Cats and Dogs; IEEE Conference on Computer Vision and Pattern Recognition, 2012