Towards Better AI Generalization for Night Scenarios

Aman Mishra
Department of Electronics Engineering
Indian Institute of Technology (BHU)
Varanasi-221005
aman.mishra.ece19@itbhu.ac.in

Vaibhav Goyal
Department of Electronics Engineering
Indian Institute of Technology (BHU)
Varanasi-221005
vaibhav.goyal.ece19@itbhu.ac.in

Dharmendra Kumar
Department of Electronics Engineering
Indian Institute of Technology (BHU)
Varanasi-221005
dharmendra.kumar.ece19@itbhu.ac.in

Sandesh Kumar Sharma
Department of Electronics Engineering
Indian Institute of Technology (BHU)
Varanasi-221005
sandesh.krsharma.ece19@itbhu.ac.in

Dr. Sonam Jain
Assistant Professor, Supervisor
Department of Electronics Engineering
Indian Institute of Technology (BHU)
Varanasi-221005
sonam.ece@itbhu.ac.in

Abstract

Deep learning have seen a tremendous growth over the past few years, it lead to very efficient and robust models for tackling different tasks like Object detection, Tracking, Face Recognition, depth estimation, localization and many more. But most of the solutions present, are heavily based around the data on which it is trained, there are a loss of generalization somewhere. Similar case happens when we try to use model trained on day scenarios in night time situations, with a good amount of darkness and less light, model performances impact drastically. In this Proposal, we are developing or tuning models which will able to perform quite well both on day and night images. We will consider firstly Person Detection and Tracking in this proposal.

***K*eywords** Deep learning · Generalization · Person Detection · Tracking

1 Introduction

Developing methods for Night Vision is a very active area of research in the recent times. There were different solutions based on the intensity of lighting present in night images. Some methods directly using thermal cameras for dealing with night images as it provides a thermal image based on the heat radiated in different parts of scene but it creates a need for thermal camera to be available to each user. Other methods trying to process night images to reduce darkness and then use dataset specific methods which can lead to loss in generalization.Some other research focusses on translation the night images to day images using another deep learning technique known as GANs and then use that translated image to work further. This method is showing promising results for autonomous navigation of vehicles but we believe that translation night time face images to day is a very difficult problem to formulate as we need to have higher pixel level accuracies to maintain the correct face features across image, otherwise the face recognition will not work.

2 Our Approach

Our approach mainly consists of two parts, firstly, working on the image processing side, and secondly, on the model side. For the image processing side, we tried different algorithms for intensity and contrast management like, histogram equalization and Adaptive histogram equalization, and on the model side. we dropped the idea od training networks from scratch due to need of a bigger dataset and massive compute. Instead we created a small night dataset in our IIT BHU Campus and combined it with existing data and fine tuned the models which balanced accuracy and compute tradeoff. Figure 1 represents the pipeline of our Approach.

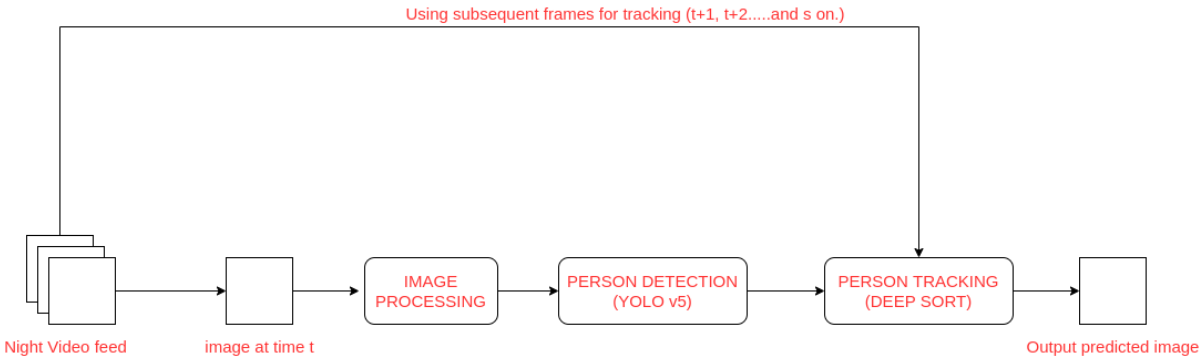

Figure 1: Proposed Pipeline for our Approach.

3 Object Detection

Object detection is a field of computer vision in which we have to classify the object as well as create a bounding box around the object to localize the object in the given image. Several models were used for Object detection which consist of One Stage detector or tow stage detector out of which YOLO series is one of the prominent solutions for one stage detector and RCNN Series IS from two stage detectors. But due to the speed and accuracy tradeoff, YOLO is preferred as it able to provide real time inference of more than 25 FPS.

3.1 Person Detection

It is a subclass of object detection whose main goal is to detect only the person class. WE had used YOLOv5 architecture for person detection and used the medium level of it which agrees between speed and accuracy and it able to provide FPS of over 25 for a video of size 640X480. Below is the architecture of YOLO model.

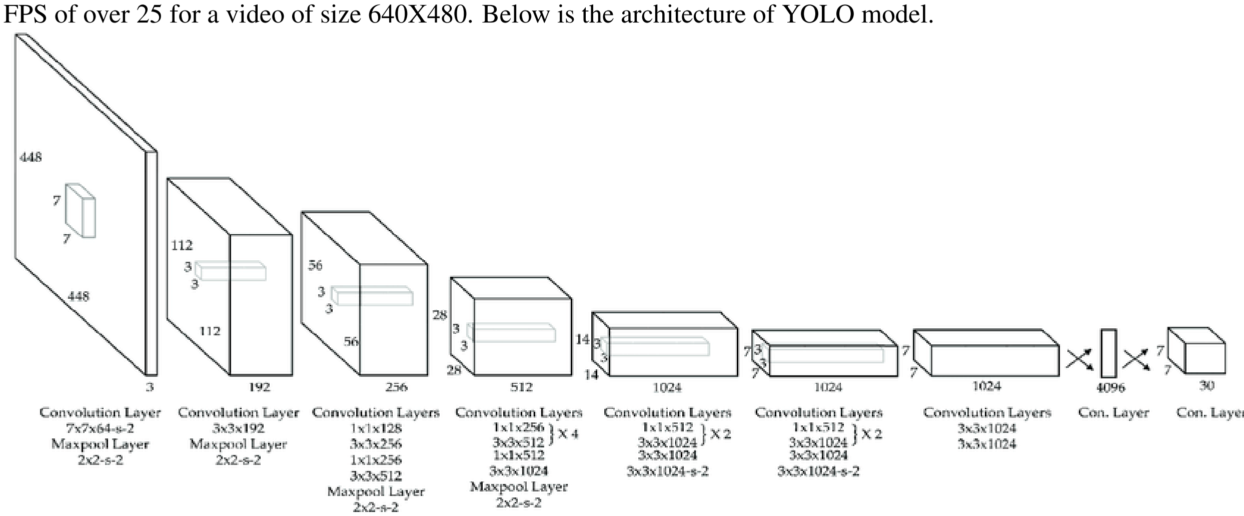

Figure 2: YOLOv5 Architecture.

4 Multi Object Tracking (MOT)

Multi Object Tracking is one of the most important problem in the field of computer vision and has tremendous application in fields where we have to preserve the Identity of ID of the detected person across the frames. It can be used in areas like surveillance, defence etc.
There were several methods for tackling this problem, some based on classical approaches, some on Deep learning. Out of various methods some are accurate but not real time, some are fast but not accurate. Based on the above tradeoff, two methods were considered good, SORT and DeepSORT. We used DeepSORT method in our case.

4.1 DeepSORT

DeepSORT is a multi tracking algorithm which works on three core functionalities.

4.1.1 Kalman Filter

A Kalman Filter is an algorithm that takes data inputs from multiple sources and estimates unknown variables, despite a potentially high level of signal noise. Often used in navigation and control technology, the Kalman Filter has the advantage of being able to predict unknown values more accurately than if individual predictions are made using singular methods of measurement.
Kalman Filters use a two-step process for estimating unknown variables. The algorithm works by first estimating the current state variables, and measures their uncertainties. Then, the algorithm updates the estimates using a weighted average, wherein more weight is attributed to estimates with higher levels of uncertainty. Because the filter takes in information from multiple sources, both current state and predicted state, the filter is able to provide a level of accuracy higher than if estimates were made given only one of the multiple sources.

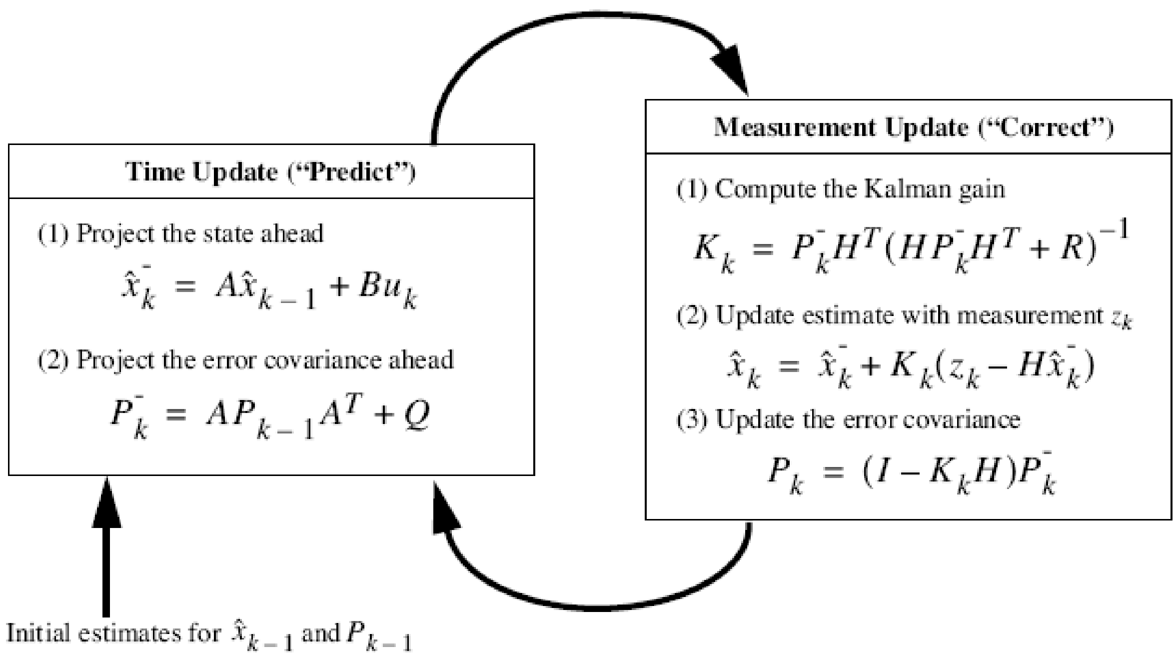

Figure 3: Kalman Filter Equations.

4.1.2 Hungarian Algorithm

The Hungarian algorithm, aka Munkres assignment algorithm, utilizes the following theorem for polynomial runtime complexity (worst case O(n3)) and guaranteed optimality: If a number is added to or subtracted from all of the entries of any one row or column of a cost matrix, then an optimal assignment for the resulting cost matrix is also an optimal assignment for the original cost matrix. We reduce our original weight matrix to contain zeros, by using the above theorem. We try to assign tasks to agents such that each agent is doing only one task and the penalty incurred in each case is zero.

4.1.3 Deep Association Metric

By using simple nearest neighbor queries without additional metric learning, successful application of our method requires a well-discriminating feature embedding to be trained offline, before the actual online tracking application. To this end, we employ a CNN that has been trained on a large-scale person re-identification dataset that contains over 1,100,000 images of 1,261 pedestrians, making it well suited for deep metric learning in a people tracking context.

5 Image Preprocessing

As night images are mostly dark and glary and has unequal contrast and intensity distribution, we used several preprocessing techniques based on histogram and frequency analysis. We used two methods from OpenCV library i.e. histogram equalization and Adaptive histogram equalization and correspondingly tested the processed image. An example of enhancement from them can be viewed from below image.

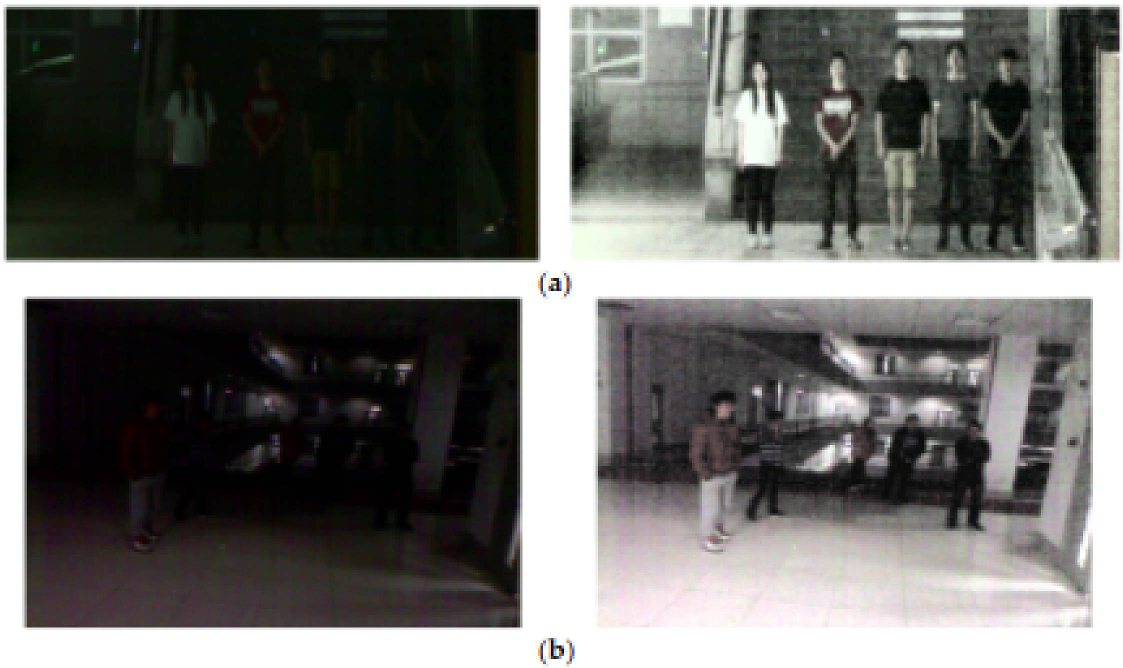

Figure 4: Preprocessed Results.

6 Dataset and Fine tuning

We cannot think with a high probability to train above models from scratch to fulfill our cases, as it requires both heavy computation and a large labelled dataset and both of them is difficult to get easily, especiallly the later one. SO we decided to fine tune the models by creating a small dataset. So, we collected 20-22 videos of several minutes of our campus's 4 major roads at different times covering almost equal distribution of day and night scenarios. Then we annotated this data and used that to fine tune the models and it helped in the overall performance. A snapshot of our data is shown below

Figure 5: One Snapshot from our collected data.

7 Results

We had provided the corresponding video results showing our performance in the presentation itself.

8 Website platform

For providing a user friendly solution, we worked on creating a website as well to show the real time feed of that person in the images and also show the results of our ML pipeline in that website to make it easier for tracking the results for a user. We had created version 1 of our website. The details of which is also included in the presentation.

9 Conclusion and Future Works

In this Proposal we worked on fine tuning the existing model by collecting and annotating our own dataset to increase generalization, In this proposal we added few novel preprocessing and fine tuning setups for Person Detection and Tracking. In our future Works, we will focus on face detection and recognition of those tracked persons within the video. Additionally, we will explore methods which translates night images to day images

10 References

[1] DEEP SORT PAPER - https://arxiv.org/pdf/1703.07402.pdf

[2] YOLO PAPER - https://arxiv.org/abs/1506.02640