**International Journal of Intelligent Information Technologies**
Volume 18 • Issue 1

# Deep Learning-Based Object Detection in Diverse Weather Conditions

Ravinder M., Indira Gandhi Delhi Technical University for Women, India
Arunima Jaiswal, Indira Gandhi Delhi Technical University for Women, India
Shivani Gulati, Indira Gandhi Delhi Technical University for Women, India

## ABSTRACT

The number of different types of composite images has grown very rapidly in current years, making object detection an extremely critical task that requires a deeper understanding of various deep learning strategies that help to detect objects with higher accuracy in less time. A brief description of object detection strategies under various weather conditions is discussed in this paper with their advantages and disadvantages. So, to overcome this, transfer learning has been used and implementation has been done with two pretrained models (i.e., YOLO and Resnet50) with denoising which detects the object under different weather conditions like sunny, snowy, rainy, and hazy. And comparison has been made between the two models. The objects are detected from the images under different conditions where Resnet50 is identified to be the best model.

## KEYWORDS

COCO Dataset, Deep Learning, Object Detection, Resnet50, Transfer Learning, YOLO

## INTRODUCTION

Computers are like machines which unlike humans can understand only numbers (on a high level). But the advancements in technology require computers to deal with and understand an image or video data. Thus, to create a bridge between the two, Computer Vision (CV) and Image Processing (IP) played an important part. The extremely vital field in Artificial Intelligence (AI) helps to detect objects in different weather conditions. The use of Object Detection (OD) is everywhere around like it is used for Content-based image retrieval, Surveillance Industry, Traffic tracking systems, Activity Recognition, Sentiment Analysis, Human-Computer Interaction, Document Summarization, Robot Vision, Consumer Electronics, Security Systems, Autonomous Driving, and Augmented Reality. OD is the most critical problem in CV. Also, this is the main task and is at a high level in the field of Artificial Intelligence that coexists with us into our lives. There are several OD methods (Saritha, R. R. et al., 2019) that exist. The goal of OD is to build rectangular bounding boxes around the objects and figure out what class they belong to. Scenes in images can be analyzed or understood with the help of recognition of image that comes under computer vision which is linked to OD. Elements in scenes can be understood at pixel level which is created by image segmentation whereas only a class

DOI: 10.4018/IJIIT.296236

label could be obtained via IR. The motivation behind this topic is that it is the strategy that will overdo all the physical tasks like Counting in Crowd, Cars: Self Driving, Surveillance System, and Detection of a Face in any type of weather.

OD is used in several applications and this can be customized up to various extents based on the real-time requirement. There are different ways to detect an Object like Point-based Detection and Optical Flow, out of which, the most frequent ones are Stage Detectors that can detect an Object in a Single Shot. The detection can be traditional or deep learning-based and it can be a two-stage or one stage. In one-stage detector methods, objects can be detected in a single go whereas in two-stage detectors the objects cannot be detected in a single go instead they can be detected in two phases.

Deep Learning has a great deal of potential in the Object Detection arena, and it can help us construct a complete solution. These algorithms can also be strengthened to provide predictions that are as near to the original bounding box as possible. As a result, the system will produce tighter and sharper bounding box estimations.

The Transfer Learning approach has been used in an effective way to detect an object. Like here it is not required to build and train the model from the start but to make faster progress the pre-downloaded weights can be used which are trained by someone else on the network architecture. So, this is a perfect fit that can be used as a pre-trained model to transfer the knowledge for a task like detecting an object in adverse weather conditions like snowy, hazy, rainy, foggy, etc., The computer vision research society has done a fine job of publishing a lot of data sets on the web, such as image net or coco or pascal kinds of data sets. These are all titles of different data sets that people have been posting online and that a lot of computer scholars have skilled their algorithms on. Often this training takes several weeks and may require many GPUs. A few open-source implementations of neural networks are available for download, including not only the code but also the weights, and there are many networks to choose from that have been trained on, such as the image net data set, which has thousands of different classes, so the network might have a softmax unit that outputs one of a thousand multiple categories. What else can be done is to delete the existing softmax layer and create a new softmax unit, or the first few layers can be frozen, and post-processing can be accomplished based on the structure. For example, depending on the framework, stuff like trainable parameter equals zero can be set for these earlier layers, and then a shallow softmax model can be trained from this feature vector. Let say, there is one Pretrained Model which is already trained on huge datasets, then it can be used with a small dataset for Post Processing of layers and it can be used for a customized purpose.

Object Detection can serve in almost every area around us like it is useful in some of the areas like Content-based image retrieval, Surveillance industry for tracking of behavior, activities, or facts amassing, influencing, managing, or directing. This can encompass remarks from a distance through the digital gadget, consisting of closed-circuit television (CCTV), or interception of electronically transmitted facts, consisting of Internet site visitors. It also can encompass easy technical strategies, consisting of human intelligence amassing and postal interception (Shidik, G. F.et al., 2019), Traffic tracking system where Road site visitors is a complicated phenomenon, in which diverse entities (pedestrians, motors, vehicles, busses, tramps, bicycles, etc.) engage one every different, while the usage of not unusual place infrastructure. The site visitors' control and manage, because of infrastructure constraints and growing variety of cars, is a complicated challenge and calls for the utility of devoted algorithms collectively with unique site visitors' facts (each historic and current) The facts approximately variety of cars and their kinds is useful in decreasing journey instances and emissions. Precise site visitors facts lets in us now no longer handiest to boom effectiveness of site visitors manage, however additionally to confirm control coverage to converting situations and expect infrastructure bottlenecks (Li, C., Dobler, G., Feng, X., & Wang, Y., 2019), Activity Recognition where in Activity popularity ambitions to apprehend the moves and dreams of 1 or extra dealers from a chain of observations at the dealers' moves and the environmental situations (Ehatisham-Ul-Haq. et al., 2019), Human-Computer Interaction here Human–laptop interplay research the layout and use of laptop generation, centered at the interfaces among human beings and computer systems.

Researchers withinside the area of HCI look at the methods (Vuletic, T. et al., 2019), Document Summarization Task in Object detection used for file summarizing wherein the precis of a file might also additionally rent phrases or terms which do now no longer seem withinside the authentic file. Multi-file summarization is also necessitated because of the fast boom in online facts. Thus, many researcher's awareness in this challenge the usage of item detection to extract vital capabilities out of a file (Liu, L. et al., 2020), Robot Vision entails the usage of an aggregate of digital digicam hardware and laptop algorithms to permit robots to technique visible facts from the international. For instance, your machine may want to have a 2D digital digicam that detects an item for the robotic to select out up. An extra complicated instance is probably to apply a 3-D stereo digital digicam to manual a robotic to mount wheels onto a transferring car. Without Robot Vision, your robotic is blind. This isn't always a hassle for lots of robot obligations, however for a few packages Robot Vision is beneficial or maybe vital (Burnett, K. et al., 2019), Consumer electronics or domestic electronics are digital gadgets meant for regular use, commonly in non-public homes. Consumer electronics encompass gadgets used for entertainment, communications, and recreation, Autonomous Driving here comes Self-using cars are motors or vehicles wherein human drivers are in no way required to take manage to soundly function the car. Also referred to as self-sufficient or "driverless" motors, they integrate sensors and software programs to manage, navigate, and pressure the car (Wang, Y. et al., 2019), and Augmented reality (AR) generation integrates virtual facts with the bodily environment, stay and in actual time. Through the addition of graphics, sounds, haptic remarks, and/or odor to the herbal international because it exists, AR can integrate actual existence with a super-imposed photo or animation the usage of the digital digicam on a cell tool or a unique headgear.

Different aspects that are presumed to identify the Object are recognized on basis of literature analysis conducted, and how multiple networks react to object detection tasks in various weather conditions and basic knowledge of how to identify and analyze are detailed clarified in the background of this article.

## BACKGROUND

In this real-world object detection is used to detect various surrounding instances like dogs, cars, laptops, various vegetables and fruits, and humans. That gives a way to recognize objects, to localize and detect objects with numerous techniques such as Point Detection, Traditional, Deep Learning based detectors, Optical Flow-based. Deep study has been done in Deep learning detectors like RCNN (Girshick, R. et al., 2014), Faster RCNN (Girshick, R. 2015), Mask RCNN (He, K. et al., 2017) are examples of one-stage detectors, while YOLO (Redmon, J. et al., 2016), Single Shot Multi Box (Liu, W. et al., 2016), and Retina Net (Lin, T. Y., et al.2017) are examples of two-stage detectors.

Supervised pre-training for supplementary tasks and segmentation of objects and to bottom-up region proposals high-capacity convolutional neural networks have been applied, which is a combination of these above-mentioned approaches and this is the technique named RCNN. Here in this technique high-performance boost has been observed by domain-specific fine-tuning. As it merges regional proposals with convolutional neural networks hence it was named as Regions with CNN features by the authors. The three stages are depicted in the diagram above, which together form the RCNN architecture. To begin with, Regional Proposal Generation: RCNN employs selective seeking to produce 2k concepts for each photo. Second, Feature Extraction: Every area concept is cut into the constant decision, and a preferable CNN variant is employed to extract a 4096-dimensional characteristic as a final representation. For all area's ideas, this feature has a high level of semantic and strong illustration. Finally, classification and localization: The retrieved features are fed into a support vector machine, which uses them to classify the item. This model integrates two strategies: supervised pre-training for auxiliary tasks and applying high-capacity convolutional neural networks to bottom-up region recommendations to localize and segment objects. For accurate object detection and semantic segmentation, rich feature hierarchies are used.

Fast RCNN It has higher mean average precision, unmarried degree schooling, schooling which makes all community layer updated, and disk garage is not always needed for characteristic capturing. The structure of Fast R-CNN takes a photo as input. It then approaches photos with a convolutional and max-pooling layer for providing a convolutional characteristic map. A characteristic vector in a constant layer is then fetched from every characteristic map employing an area of hobby pooling layer for every area notion. Then these are then fed to completely linked layers. Output layers are then classified by them.

The Faster R-CNN model is made up of two modules: a deep convolutional community that features areas and a detector that uses those areas. The RPN feeds a photo as entering and produces the output of square item proposals. Every rectangle has an objectless rating. To resolve this hassle, (Shih, K. H. et al., 2019) delivered a further (RPN), that serves in an almost value-unfastened manner through sharing complete-photo convolution capabilities with detection networks i.e in preference to the usage of selective seek set of rules at the characteristic map which discovers area proposals, a distinct community which is required to expect the area proposals. RPN is completed with a convolutional community, that may expect item bounds and ratings at every function simultaneously. An arbitrary-length photo is fed to RPN which produces a fixed of square item proposals. In other words, a small community is slid over a convolutional characteristic map, which would be the output of the final convolutional layer, to develop "proposals" for the region in which the item is located.

Convolution of K anchor packing containers has been conducted in RPN along with each sliding window, resulting in continuous vectors that may be extracted using the magnificence layer and regression layer to obtain the desired output. The community moves across the characteristic map, blending completely with a n* n spatial frame. After then, in every sliding window, a vector of low dimensions is acquired and passed into FC layers. This technique is presented in (Ren, S., He, K., Girshick, R., & Sun, J., 2015). In good illumination and normal weather patterns, the features retrieved from an image can aid in the identification of an object. Images captured by camera systems in minimal light situations such as night, dusk, and dawn, as well as in poor weather conditions such as rain and snow, contain partially lit objects, poor contrast, and poorly informed content. Object detection algorithms have a hard time with these images. Also, the dataset used was KITTI (Geiger, A., et al., 2013), FLIR (FLIR, 2018), and Indian Driving Dataset (Varma, G., et al., 2019).

In SPP Net it follows RCNN, because of the life of FC layers, CNN calls for a set length enter, and because of this RCNN crop every area notion into the identical length. It might also additionally occur that items might also additionally in part seem withinside the wrapped area and undesirable geometric distortion can be produced because of the wrapping task. The material losses would lessen popularity in terms of accuracy, in particular, while the items vary. To address the problem, researchers developed the SPP-Net model, which is based on the notion of spatial pyramid matching (SPM). SPPNet (Purkait, P., et al., 2017) received the first runner-up place in the Object Detection Task, second runner-up place in Image Classification, and fifth position in the Localization Task.

Review Net has a unique design that allows it to learn based entirely on a dehazing solution – the use of color space, appearance structuring, the one-of-a-kind loss weights ratio for the major and secondary presentations, and the use of bottleneck parallel spatial cleaning. It is vital to recognize that the multi-appearance structure utilized by Review Net does now no longer upload directly to the variety of parameters due to the fact the second one appearance or by skip isn't always introduced after a static first appearance. Instead, retaining the variety of parameters and length of the structure constant, the whole community is cut up into seems – making the whole shape extra green for photo enhancement and mild sufficient to be included in self-sufficient cars. The proposed Review Net encompasses a set of encoder-decoder modules. (Mehra, A. et al., 2020) Used five datasets in this work – RESIDE-preferred indoor (thirteen,990 photo pairs), RESIDEβ outdoor (72, a hundred thirty-five photo pairs), RESIDE HSTS (10 photo pairs) [41, HazeRD (seventy-five image pairs) and D-Hazy (1,499 photo pairs) (Zhang, Y., Ding, L., & Sharma, G., 2017). RESIDE dataset (Li, B., et al., 2017) became selected for the schooling due to the fact, aside from being one in every of the

biggest publicly to be had datasets for dehazing, it benchmarks 9 consultant countries of the artwork dehazing strategies through supplying complete reference assessment metrics required for artificial goal checking outset.

ReViewNet (Mehra, A., et al., 2020), for Photo dehazing for actual time packages in self-sufficient using. The ReViewNet structure seems two times on a hazy photo, and the community is optimized. One of the kind losses is carried out to decide the foremost ratio. Outcomes had been confirmed over 32 kinds of scenarios. The paintings provide to the country-of-the-artwork employing demonstrating conclusive evidence that the usage of context particular additives and capabilities of a deep gaining knowledge of community can bring and extra correct photo modules for enhancement. In addition, it propagates to higher preprocessing and better overall performance.

All the preceding item techniques have been discussed like CNN Family and ReviewNet and SPP Net (Purkait, P., et al., 2017) where the area is required to locate the item in the image. The community now no longer examines the entire Photo, instead seems at elements of the PHOTO that have excessive chances. Yolo or You Look Only Once, proposed is a unique object detection set of rules a great deal one kind from the area primarily based algorithms visible. It consists of several convolutional layers with different grid sizes.

It comes with the great advantage of fastness but since object detection takes place in one step only, shows lesser accuracy.

SSD is likewise part of the own circle of relatives of networks which expect the bounding packing containers of items in every Photo. It is easy to stop the unmarried community, eliminating many steps concerned in different networks which attempts to obtain identical challenge, on the time of its publishing. It works higher than the county of Faster RCNN in instances of better dimensional pictures. The essential idea of SSD is frequently primarily based totally on the feed ahead convolutional community. It then generates ratings forth the presence of every item magnificence in every default field and produces modifications to higher in the shape of the item. The SSD version is constituted of in particular structures: Base Community and Auxiliary Community. The base community is the early part of the version that's primarily based on the preferred structure used for the excessive-high satisfactory photo category. The Auxiliary community has capabilities Object Detection the usage of Single Shot Detector for a Self-Driving Car in particular centered for items with one kind scales or factor ratios. SSD has additives in its shape. The first issue known as the base community is used for a category of pictures. The second issue has 3 beneficial capabilities together with multi-scale capabilities maps for detection. The requirements of item localization and categorization are accomplished ahead of time by skipping the community in the Single Shot technique. The call of a way for bounding field regression is the Multibox community.

A Progressive Recurrent Network (PReNet) for recursively removing rain from a single image. It is possible to remove some rain at each iteration, and the remaining rain may be removed progressively in subsequent iterations. The result is a rain-free image after a certain number of iterations. Besides some residual ResNet blocks, PReNet (Wang, J., et al., 2017) contains a CNN layer that works on both the original rainy image and the current output image. A second CNN layer provides the current output image. The convolutional LSTM is further combined with a recurrent layer to exploit deep feature dependencies across iterations is presented by (Renet at., 2019).

A modified Track-Oriented Multiple Hypothesis Tracking (MHT) algorithm is used to track the surrounding vehicles using the data from these sensors. A modification was introduced in the well-known MHT algorithm to prevent the exponential growth of possible hypotheses and thereby reduce computation time without sacrificing critical information. A real-time implementation of the system has been implemented using a dataset collected at Texas A&M University (Bhadoriya et al., 2021).

A study to analyze the performance of object detection methods trained and tested using images captured in clear and rainy weather conditions and also examines the effectiveness and limitations of leading de-raining approaches, deep-learning-based domain adaptation, and image translation frameworks that have been considered for addressing the problem of object detection under rainy

conditions. As part of this tutorial, experimental results are presented for a variety of surveyed techniques (MazinHnewa et al., 2020).

## MATERIALS AND METHODS

### Dataset and Attribute Information

The dataset used in this model is trained under the COCO model set. MS-COCO dataset consists of 3,30,000 pictures including 2,50,000 categorized pictures; 150 are the item images, 80 are item classes, 91 are stuff classes and 5 are captions according to the pictures. This dataset is well-known because it shows capabilities containing context popularity, multi-item supportability according to the images, and item segmentation. COCO dataset can predict as well as to detect 80 different objects in the surroundings, i.e., it can find 80 different classes from various objects. While using the transfer learning concept, the pre-trained model is required at the end and for the last few layers, data set with very few images are required for performing de-noising. Eg: Yolo pre-trained model consists of 112 layers and the Resnet50 pre-trained model consists of 147 layers, out of which a layer present, in the end, is responsible for de-noising. De-noising helps enhance the features of an image so that the model can detect objects with great accuracy and draw a bounding box around the object. The dataset being used has 99 snowy images, 96 foggy images, 94 frizzy, and 101 rainy images.

**Number of classes:** 80
**Area:** Every basic object has been covered by COCO DATASET

In adverse weather conditions like foggy, frizzy, rainy, snowy, etc., the model can detect all the objects mentioned above in Table 1.

### Deep Learning

In the world of object detection, Deep learning plays a vital role as it builds an end-to-end approach for detecting objects in every weather condition (Huang, S. C., et al., 2020). Behind the scenes, the original image is passed in the neural networks rather than just the patches of original images which lead to the reduction of the dimensions. A neural network helps in suggesting selective patches. It's a branch of machine learning in which the layers of a neural network go deeper. In every discipline, this technique has yielded incredible outcomes. The input layer, the hidden layer where the weights are established, and the output layer make up a primary neural network. It is referred to be a neural network when more layers are included. Multiple levels of representation are present in deep learning models (Jiao, L., et al., 2019). Nonlinear and simple modules are used to create this representation. These modules raise the presentation level of the representation from level one to higher. Deep learning models understand more complex functions as a result of this transition. Computer Vision, which includes the recurrent convolutional network and the fast recurrent convolutional network, has been in use for a long time and has gained popularity in recent years as the advancement of hardware has equipped machines with more processing power. These neural networks have their own set of benefits and drawbacks, which have already been discussed in the preceding section. One or more convolutional layers make up a convolutional neural network (Liu, L., et al., 2020). The architecture of a convolutional neural network is meant to take the benefit of a 2D image, which is accomplished through local connections and pooling layers to obtain features extraction. In general, there are two types of convolution: one that is fully connected with layers to the hidden layers, and the other that is not. This strategy is not only practical, but it also performs better.

Transfer learning (Torrey, L., et al., 2010) is one of the most important domains under Deep Learning. It has a lot of potential in the object detection space that is why it came into the picture wherein, the basic idea is to transfer the learning of a model to another. After the learning is transferred,

**Table 1. Dataset description**

| Different Classes of Objects | |
|---|---|
| Person | Wine Glass |
| Bicycle | Cup |
| Car | Fork |
| Motorcycle | Knife |
| Airplane | Spoon |
| Bus | Bowl |
| Train | Banana |
| Truck | Apple |
| Boat | Sandwich |
| Traffic Light | Orange |
| Fire Hydrant | Broccoli |
| Stop Sign | Carrot |
| Parking Meter | Hotdog |
| Bench | Pizza |
| Bird | Donut |
| Cat | Cake |
| Dog | Chair |
| Horse | Couch |
| Sheep | Potted plant |
| Cow | Bed |
| Elephant | Dining Table |
| Bear | Toilet |
| Zebra | TV |
| Giraffe | Laptop |
| Backpack | Mouse |
| Umbrella | Remote |
| Handbag | Keyboard |
| Tie | Cellphone |
| Suitcase | Microwave |
| Frisbee | Oven |
| Skis | Toaster |
| Snowboard | Sink |
| Sports ball | Refrigerator |
| Kite | Book |
| Baseball Bat | Clock |
| Baseball Glove | Vase |
| Skateboard | Scissors |
| Surfboard | Teddy Bear |
| Tennis Racket | Hair Dryer |
| Bottle | Toothbrush |

a lot of efforts were reduced. The only thing it makes sense is when there is a Task 1 which is readily available on the internet and for Task 2 we have some fewer options like for example Car features will remain the same irrespective of weather or any other real-world object.

A pre-trained model is required which has already been trained on the best dataset available over the internet and can be used based on the requirement in the real world. Like several methods are available for Detection as well as Segmentation and one can make use of this methodology so that the time and cost of the Advance Model is saved that can be used for some requirement-based customize purpose.

Tiny YoloV3 (Redmon, J., et al., 2018) has been implemented just to check on gather some hands-on on how pre-trained model works behind the scenes with different use cases like rainy, snowy, foggy and hazy weather has been covered and Output has been saved in local and accuracy has come up as variation with several objects like the person is 99% and the car is 97% and truck is 87% and so on that will be covering result section.

## PROPOSED MODEL

The most feasible and efficient method to detect an object in adverse weather conditions is to use the trending concept of transfer learning. Transfer learning is a perfect fit for the same as in the real world, the object will remain the same irrespective of weather conditions. Enough dataset is available on the internet for different variety of objects. Also, pre-trained models are available, that are trained on the MSCOCO dataset (Lin, T. Y., el al., 2014).

YOLO, and RESENET 50 (He, K., et al., 2016) pretrained models are efficient enough to detect an object as they have already been trained on the network architecture with downloaded weights.

A pre-trained model is already a very well-trained model. That is why it plays a very important role in a proposed model. While working with the pre-trained model, the efforts required are less, because it doesn't require the complete model to be prepared from the scratch, which is why it becomes easy to implement the pre-trained model with the de-noising approach for preparing the proposed model. This pre-trained model has been trained previously with huge datasets, so there is no need to repeatedly work with huge data sets. Therefore, the pre-trained model is the best one to work with, when it comes to the proposed model. The suggested model consists mostly of incorporating a de-noising method into a pre-trained model. The layered view of the de-noising approach. The pre-trained model is aware of the features of different objects which are to be detected. The properties and features of the objects remain the same. Hence, for the pre-trained model, it doesn't matter, under what weather conditions; the objects are to be detected. For different weather conditions, the de-noising approach comes into play. The image used is enhanced with the help of this de-noising approach. As soon as the image objects are enhanced using this approach, it becomes easy and convenient for the pre-trained model to detect the objects present in the image. Thus, the bounding box is prepared easily around the object after detection. A rectangular boundary is made, as much as closer and accurate as it can. The rectangular boundary to the object tells that how accurately the object is detected using the proposed model.

### Pre-Requisites

**Step 1:** The required dependencies must be resolved like Keras, NumPy, TensorFlow, Matplotlib should be imported to have that environment setup ready on which further implementation can be performed.

**Step 2:** Directory structure should be there for all the assets and dependency tree resolved like an input data set is made ready for use as an object must be detected from those images present in the input data set comes under assets and then the main file and other required files and Output Folder.

**Step 3:** A framework has been designed and developed to process the input in the model and receive the output from the model.

As soon as the pre-trained model is downloaded, weights are required. The pre-trained model is efficient and convenient to use and is doing the tasks exactly as needed. The softmax layer units are created, due to which the previous layers are frozen. At the time of post-processing, the smaller data sets are used, which are completely dependent on the provided framework. The training parameters are set to zero for the layers at the beginning, to compute the feature vector. A shallow softmax model is used for the post-processing which will include all the steps and processes coming under the de-noising approach. The effort has been done to showcase the same in Figure 1.

## RESULTS AND DISCUSSIONS

### Results

YOLO and RESNET50 with denoising model have been implemented and both were trained on the COCO dataset for classification and detection of different objects in snowy, rainy, and foggy weather conditions. These models are implemented on an i5 processor in Python IDLE. Yolo is implemented by importing the required libraries and by feeding the image into the model of 112 layers and storing the output in the destination folder. It shows and detects an object with different accuracy and draws a bounding box around the object.

The proposed Model of the pre-trained Model of RESNET 50 with Denoising has been implemented and provides results for that images also where YOLO was unable to detect a single object. It shows better accuracy than YOLO as it consists of a denoising approach which enhances the feature of objects that helps in drawing a bounding box around the objects.

The comparisons between variants of YOLO and RESNET50 have shown below in different weather conditions:

## DISCUSSION

As per the observation given in Table 2 and Table 3, this has been found that in some cases for snowy weather YOLO has performed not well whereas Resnet50 performed better. As shown in Figure 2 in

**Figure 1. Layered view of the proposed model**

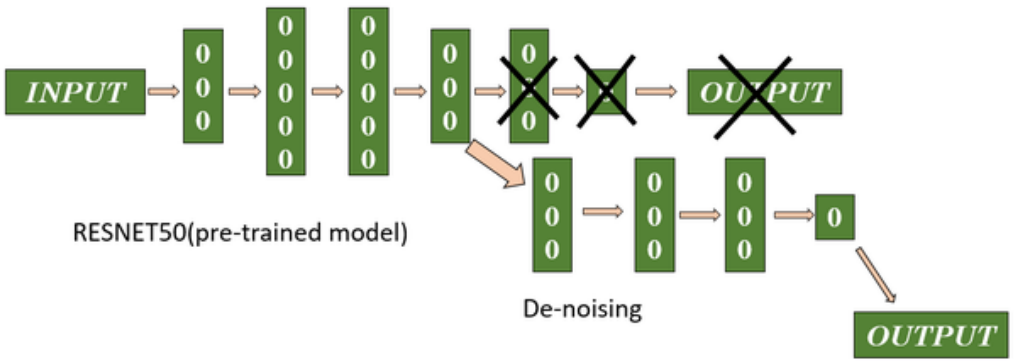


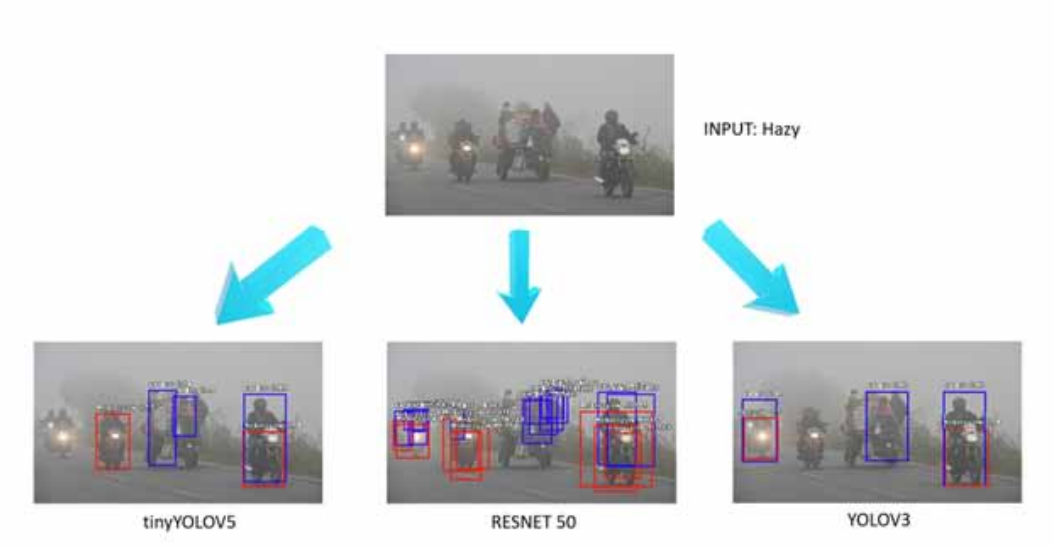
**Table 2. Performance percentage probability of all three pre-trained model in all the 5 kinds of weather**

| Pretrained Model | Weather == FOGGY | Weather == SNOWY | Weather == Rainy(MILD) | Weather == Rainy(HIGH) | Weather == Sunny |
|---|---|---|---|---|---|
| YOLOV3 | 67% | 82% | 63% | 79% | 97% |
| tinyYOLOV5 | 87% | 57% | 61% | 43% | 98% |
| RESNET50 | 99% | 98% | 92% | 82% | 99% |

**Table 3. Detection time of all the three pre-trained models in all 5 kinds of weather**

| Pretrained Model | Weather == FOGGY | Weather == SNOWY | Weather == Rainy(MILD) | Weather == Rainy(HIGH) | Weather == Sunny |
|---|---|---|---|---|---|
| YOLOV3 | 39.542678830270 | 1.90836763381958 | 0.9579235645654 | 0.98579456123235 | 0.97845612356645 |
| tinyYOLOV5 | 45.548793145445 | 3.84545445445474 | 2.5242415242221 | 3.58525784447575 | 1.23546894763214 |
| RESNET50 | 33.027085542678 | 0.90836763381958 | 0.9079473018646 | 0.90048003196716 | 0.89579820632934 |

**Figure 2. Output from tinyYOLOV5, YOLOV3, and Resnet50 in hazy weather conditions**




hazy weather season Resnet50 performs better, whereas YOLO performs better in hazy weather, in Figure 3 under rainy season accuracy of YOLO is less than that of Resnet50 as it can detect 2 persons along with that vehicle, in Figure 4 under little rainy season YOLO was unable to detect rest of the objects other than the car that too with very less accuracy on the other hand Resnet covers every object with greater accuracy, and in Figure 5 in snowy weather, YOLO didn't detect any object and Resnet was able to detect the objects in the image.

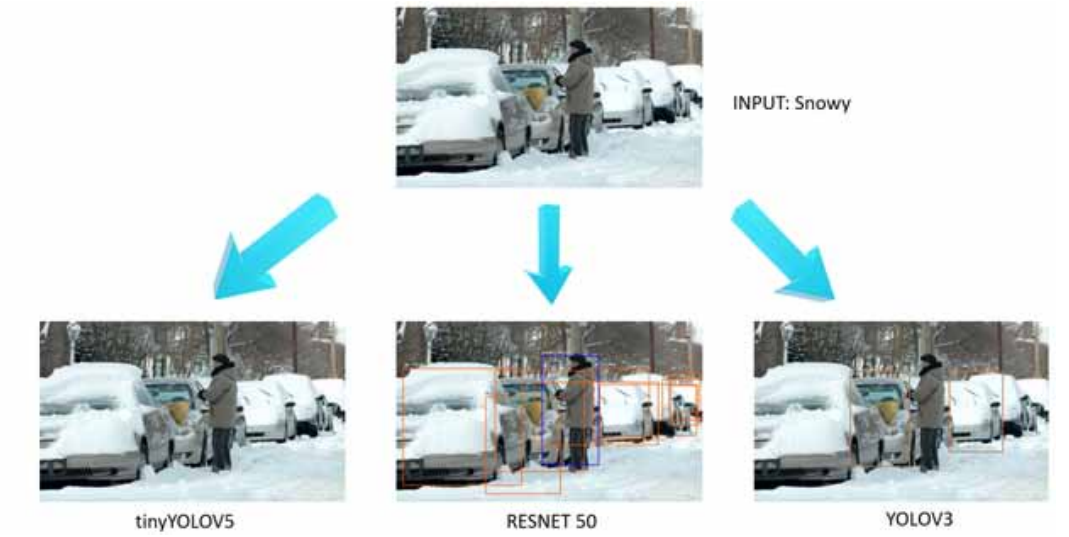
Figure 3. Output from tinyYOLOV5, YOLOV3, and Resnet50 in rainy weather conditions


Figure 4. Output from tinyYOLOV5, YOLOV3, and Resnet50 in little rainy weather conditions


## CONCLUSION

Analysis of existing methods of object detection has been done. And it was observed that using the concept of transfer learning fits the best. This research work concerning object detection in adverse weather conditions involved the literature review implementation of models and algorithms involving Deep Learning Approaches such as Transfer Learning

**Figure 5. Output from tinyYOLOV5, YOLOV3, and Resnet50 in snowy weather conditions**




analysis have been made from two pre-trained models like YOLO and Resnet50 with Denoising concludes that YOLO performs better in hazy weather conditions whereas in Rainy and Snowy weather Resnet50 performs the best. As it detected each small object with greater accuracy.

## REFERENCES

Bhadoriya, A. S., Vegamoor, V. K., & Rathinam, S. (2021). *Object Detection and Tracking for Autonomous Vehicles in Adverse Weather Conditions* (No. 2021-01-0079). SAE Technical Paper.

Burnett, K., Samavi, S., Waslander, S., Barfoot, T., & Schoellig, A. (2019, May). autotrack: A lightweight object detection and tracking system for the saeautodrive challenge. In *2019 16th Conference on Computer and Robot Vision (CRV)* (pp. 209-216). IEEE.

Ehatisham-Ul-Haq, M., Javed, A., Azam, M. A., Malik, H. M., Irtaza, A., Lee, I. H., & Mahmood, M. T. (2019). Robust human activity recognition using multimodal feature-level fusion. *IEEE Access: Practical Innovations, Open Solutions*, *7*, 60736–60751.

FLIR. (2018). *Free FLIR thermal dataset for algorithm training*. https://www.flir.com/oem/adas/adas-dataset-form/

Geiger, A., Lenz, P., Stiller, C., & Urtasun, R. (2013). Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, *32*(11), 1231–1237.

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448). IEEE.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587). IEEE.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969). IEEE.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778). IEEE.

Hnewa, M., & Radha, H. (2020). Object Detection Under Rainy Conditions for Autonomous Vehicles: A Review of State-of-the-Art and Emerging Techniques. *IEEE Signal Processing Magazine*, *38*(1), 53–67.

Huang, S. C., Le, T. H., & Jaw, D. W. (2020). DSNet: Joint semantic learning for object detection in inclement weather conditions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Jiao, L., Zhang, F., Liu, F., Yang, S., Li, L., Feng, Z., & Qu, R. (2019). A survey of deep learning-based object detection. *IEEE Access: Practical Innovations, Open Solutions*, *7*, 128837–128868.

Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., & Wang, Z. (2017). *Reside: A benchmark for single image dehazing.* arXiv preprint arXiv:1712.04143, 1.

Li, C., Dobler, G., Feng, X., & Wang, Y. (2019). *Tracknet: Simultaneous object detection and tracking and its application in traffic video analysis.* arXiv preprint arXiv:1902.01466.

Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980-2988). IEEE.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., . . . Zitnick, C. L. (2014, September). Microsoft coco: Common objects in context. In *European conference on computer vision* (pp. 740-755). Springer.

Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., & Pietikäinen, M. (2020). Deep learning for generic object detection: A survey. *International Journal of Computer Vision*, *128*(2), 261–318.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21-37). Springer.

Purkait, P., Zhao, C., & Zach, C. (2017). *SPP-Net: Deep absolute pose regression with synthetic views*. arXiv preprint arXiv:1712.03452.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788). IEEE.

Redmon, J., & Farhadi, A. (2018). *Yolov3: An incremental improvement.* arXiv preprint arXiv:1804.02767.

Ren, D., Zuo, W., Hu, Q., Zhu, P., & Meng, D. (2019). Progressive image deraining networks: A better and simpler baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3937-3946). IEEE.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, *28*, 91–99.

Saritha, R. R., Paul, V., & Kumar, P. G. (2019). Content based image retrieval using deep learning process. *Cluster Computing*, *22*(2), 4187–4200. doi:10.1007/s10586-018-1731-0

Shidik, G. F., Noersasongko, E., Nugraha, A., Andono, P. N., Jumanto, J., & Kusuma, E. J. (2019). A systematic review of intelligence video surveillance: Trends, techniques, frameworks, and datasets. *IEEE Access: Practical Innovations, Open Solutions*, *7*, 170457–170473. doi:10.1109/ACCESS.2019.2955387

Shih, K. H., Chiu, C. T., Lin, J. A., & Bu, Y. Y. (2019). Real-time object detection with reduced region proposal network via multi-feature concatenation. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(6), 2164–2173.

Torrey, L., & Shavlik, J. (2010). Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques* (pp. 242–264). IGI Global.

Varma, G., Subramanian, A., Namboodiri, A., Chandraker, M., & Jawahar, C. V. (2019, January). IDD: A dataset for exploring problems of autonomous navigation in unconstrained environments. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1743-1751). IEEE.

Vuletic, T., Duffy, A., Hay, L., McTeague, C., Campbell, G., & Grealy, M. (2019). Systematic literature review of hand gestures used in human computer interaction interfaces. *International Journal of Human-Computer Studies*, *129*, 74–94.

Wang, J., Wang, W., Wang, L., & Tan, T. (2017, October). PreNet: Parallel Recurrent Neural Networks for Image Classification. In *CCF Chinese Conference on Computer Vision* (pp. 461-473). Springer.

Wang, Y., Chao, W. L., Garg, D., Hariharan, B., Campbell, M., & Weinberger, K. Q. (2019). Pseudo-lidar from visual depth estimation: Bridging the gap in 3d object detection for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8445-8453). IEEE.

Zhang, Y., Ding, L., & Sharma, G. (2017, September). Hazerd: an outdoor scene dataset and benchmark for single image dehazing. In *2017 IEEE international conference on image processing (ICIP)* (pp. 3205-3209). IEEE.

*Ravinder M. received his B.Tech (Computer Science & Engineering) and M.Tech (Computer Science) degrees from B V Raju Institute of Technology (BVRIT), and School of Information Technology-Jawaharlal Nehru Technological University Hyderabad, Telangana, India in 2003 and 2009 respectively, and his Ph.D (Computer Science & Engineering) degree from the Jawaharlal Nehru Technological University Kakinada (JNTUK), Andhra Pradesh, India in 2017. He is working as Assistant Professor at the Department of Computer Science & Engineering, Indira Gandhi Delhi Technical University for Women (IGDTUW), Delhi, India since March 2018 He has more than eight years of teaching experience. Prior to joining IGDTUW, he has worked in Koneru Lakshmaiah Deemed to be University (KL Deemed to be University), Guntur, Andhra Pradesh, and Geethanjali College of Engineering and Technology (GCET), Hyderabad, Telangana, India etc. His areas of research interest are Image Processing, Video Retrieval, Machine Learning, and Deep Learning.*

*Arunima Jaiswal is a Ph.D in Computer Science & Engineering from Delhi Technological University, Delhi, India. She has received her M.Tech (Master of Technology) degree from Delhi Technological University, Delhi, India and B.Tech (Bachelor of Technology) in Information Technology from University School of Information Technology, Guru Gobind Singh Indraprastha University, Delhi, India. Currently, she is working as a University Assistant Professor in Dept. of Computer Science & Engineering at the Indira Gandhi Delhi Technical University for Women, Delhi, India. She has many publications to her credit in various Journals with high impact factor and International Conferences. Her research interests are in the area of Sentiment Analysis, Social Media Analytics, Soft Computing, Machine Learning, Deep Learning, Social & Semantic Web.*