

TECHNIQUES FOR IOU NON-MAX SUPPRESSION TO IMPROVE ROAD SURFACE DAMAGE DETECTION ACCURACY

R. Palani

Research Scholar, Department of Computer and Information Science,
Annamalai University, Chidambaram, Tamil Nadu, India, palanicdm@gmail.com

Dr. N. Puviarasan

Professor and Head, Department of Computer and Information Science,
Annamalai University, Chidambaram, Tamil Nadu, India, npuvi2410@yahoo.in

Dr. A. Rama Prasath

Associate Professor, School of Computing Sciences,
Hindustan University, Chennai, Tamil Nadu, India , rprasath@hindustanuniv.ac.in

Abstract: The aim of this research is to be implementing Intersection over Union (IOU) Non-Max Suppression (NMS) in Road Surface Object Detection helps reduce duplicate detections, improve localization and classification accuracy, reduce computational overhead, and enhance the overall reliability of the system. These benefits contribute to more accurate and robust road surface object detection, which is vital for various applications in the automotive industry, traffic management, and driver assistance systems.

Keywords: Computer Vision, IoU Non max suppression, Custom Object Detection, Road Surface Damage Detection.

1. Introduction

The introduction of non-maximum suppression (NMS) in object detection revolutionized the field of computer vision by addressing the problem of redundant and overlapping bounding box predictions. Object detection algorithms often generate multiple bounding box hypotheses for each object in an image, resulting in duplicate detections and decreased precision.

NMS was introduced as a post-processing step to filter out redundant detections and select the most accurate bounding boxes. It works by considering the confidence scores of the bounding box predictions and calculating the intersection over union (IoU) between them. By discarding bounding boxes with high IoU values, NMS ensures that only the most confident and non-overlapping detections remain.

The primary goal of NMS is to refine the output of object detection algorithms, improving object localization accuracy and reducing false positives. It has become a fundamental step in modern object detection pipelines, allowing for more precise and reliable object localization.

TECHNIQUES FOR IOU NON-MAX SUPPRESSION TO IMPROVE ROAD SURFACE DAMAGE DETECTION ACCURACY

NMS has seen widespread adoption and has become a standard technique in the computer vision community. Over the years, researchers have proposed variations and enhancements to NMS, such as soft-NMS and adaptive NMS, to address its limitations and further improve object detection performance.

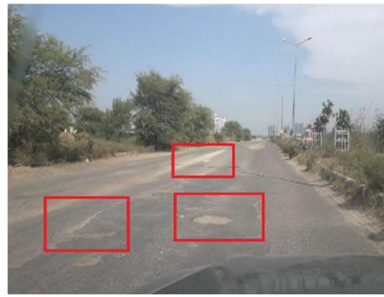
1.1. Object Detection

One of the subfields of computer vision, object detection, is extensively used in business. Two tasks are involved in object detection:

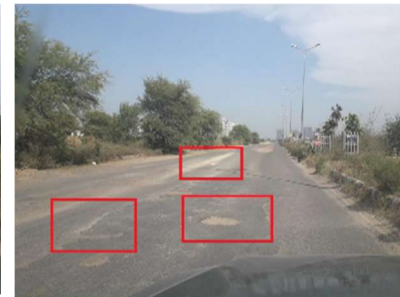
- Locating the object in the image
- Classifying the object in the image



Classifying the object in the
image



Locating the object in the
image



Classify and locate the
object, It is object detection
problem

Figure 1. Road Damage Classification and Detection.

1.2. How does Non-Max Supression work

Non-maximum suppression (NMS) is employed to choose the most suitable bounding box for an object while discarding or suppressing all other bounding boxes. NMS takes two factors into consideration:

- The objectiveness score assigned by the model.
- The overlap or Intersection over Union (IOU) of the bounding boxes.

In the provided image, observe the bounding boxes along with their respective objectiveness scores. This score indicates the model's level of confidence that the desired object is present within that specific bounding box.

Although all the bounding boxes appear to contain the object, only the green bounding box is deemed the best choice for detecting the object. How do we eliminate the redundant bounding boxes? During non-max suppression, the bounding box with the highest objectiveness score is initially selected. Then, all other boxes with significant overlap are removed. In Figure 2, The green bounding box for the pothole is chosen (as it possesses the highest objectiveness score of 98%).

The blue and red boxes for the pothole are discarded due to their substantial overlap with the green box. This process is repeated iteratively until no further reduction of boxes occurs. Ultimately, the remaining result will be as follows.

TECHNIQUES FOR IOU NON-MAX SUPPRESSION TO IMPROVE ROAD SURFACE DAMAGE DETECTION ACCURACY

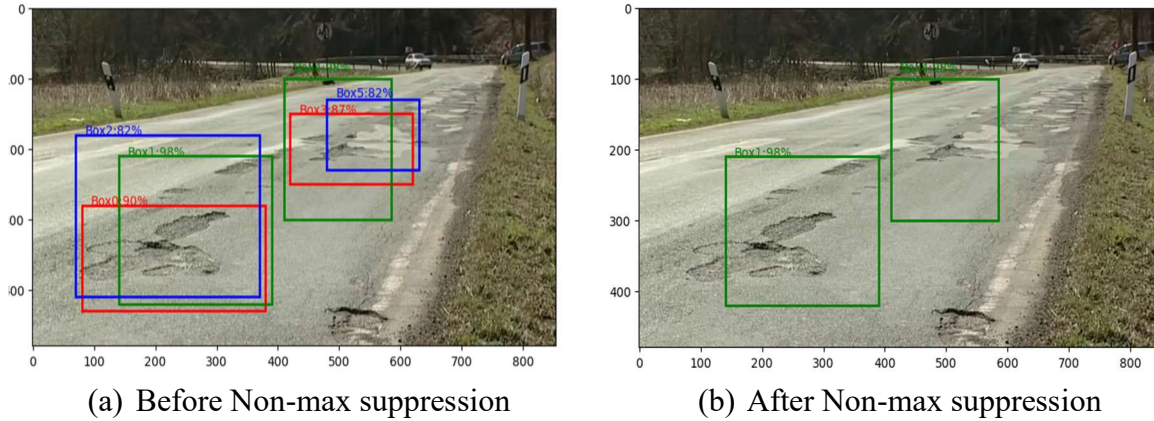


Figure 2. Non-max Suppression

1.3. Performance of NMS

Non-maximum suppression (NMS) is a post-processing algorithm commonly used in object detection tasks to eliminate redundant or overlapping bounding box predictions. Its primary purpose is to select the most accurate and non-overlapping bounding boxes that represent individual objects within an image. NMS is a critical step in achieving precise object localization and reducing duplicate detections. While the exact implementation details may vary, I'll provide an overview of NMS and its performance in object detection. NMS operates on a set of bounding box predictions generated by an object detection algorithm. These predictions typically include the coordinates of the bounding boxes, confidence scores indicating the likelihood of an object being present, and class labels if the detection is performed for multiple object classes.

The steps involved in NMS are as follows:

1. Sort the bounding box predictions based on their confidence scores in descending order.
2. Select the bounding box with the highest confidence score and consider it as a detection.
3. Calculate the intersection over union (IoU) between this selected bounding box and the remaining bounding boxes.
4. Remove all bounding boxes with IoU values above a certain threshold (e.g., 0.5 is a commonly used value) to avoid redundant detections of the same object.

Repeat steps 2-4 until all bounding boxes have been processed.

NMS ensures that only the most confident and non-overlapping bounding boxes remain after the elimination process. This process helps improve the precision and accuracy of the object detection system by reducing duplicate or overlapping detections.

In terms of performance, NMS has been widely adopted in the field of computer vision and has proven to be effective in enhancing object detection results. It helps remove redundant bounding boxes and refine the final set of detections, which can lead to improved object localization and reduced false positives.

However, it's worth noting that NMS is a heuristic approach and has some limitations. In scenarios where objects are densely packed or have significant overlap, NMS may struggle to eliminate all duplicate detections, leading to some false positives or missed detections. Researchers have proposed various

extensions and improvements to NMS to address these challenges, such as soft-NMS and adaptive NMS, which aim to handle overlapping instances more effectively.

Overall, NMS remains a fundamental and widely used technique in object detection pipelines, and its performance can significantly contribute to accurate and reliable object localization.

1.4. Various Types of NMS

The key differences between Intersection over Union (IoU) NMS, traditional NMS, and soft NMS lie in how they handle overlapping bounding boxes during the post-processing stage in object detection:

IoU NMS (used in YOLO):

- IoU NMS calculates the IoU (Intersection over Union) between bounding boxes to measure their overlap.
- It sets a threshold value (commonly 0.5) to determine when two boxes are considered overlapping.
- During the NMS process, boxes with IoU values higher than the threshold are suppressed or discarded.
- IoU NMS is a simple and widely used method that removes duplicate detections based on a fixed IoU threshold.

Traditional NMS:

- Traditional NMS is a broader term that encompasses various techniques used in object detection algorithms.
- It typically involves sorting bounding boxes based on confidence scores and eliminating redundant detections.
- The process usually includes selecting the box with the highest confidence score as a reference and comparing its IoU with other boxes.
- Boxes with IoU values above a threshold are discarded, and the process continues until all boxes have been evaluated.
- Traditional NMS does not necessarily specify the exact method for calculating IoU or suppressing boxes.

Soft NMS:

- Soft NMS is an alternative to traditional NMS that addresses the limitations of discarding lower-confidence overlapping boxes.
- It introduces a scoring mechanism that reduces the confidence scores of overlapping boxes instead of outright discarding them.
- Instead of a fixed threshold, Soft NMS uses a decay function (often a Gaussian decay) to reduce the scores of overlapping boxes based on their IoU.
- This allows boxes with lower confidence scores but high overlapping areas to still contribute to the final detections.
- Soft NMS provides a more flexible approach that aims to retain potentially valid detections with lower scores.

1.5 Research aim

This research aims of IOU NMS in object detection using YOLO is to advance the state-of-the-art in bounding box post-processing techniques. It aims to enhance the accuracy, efficiency, and robustness of the object detection pipeline, enabling more accurate localization and classification of objects in various real-world scenarios.

1. Related Works

Clustering detections: The de facto standard algorithm, GreedyNMS, has survived several generations of detectors, from Viola&Jones [32], over the deformable parts model (DPM) [7], to the current state-of-the-art R-CNN family [10, 9, 21]. Several other clustering algorithms have been explored for the task of NMS without showing consistent gains: mean-shift clustering [6, 35], agglomerative clustering [2], affinity propagation clustering [17], and heuristic variants [25]. Principled clustering formulations with globally optimal solutions have been proposed in [27, 23], although they have yet to surpass the performance of GreedyNMS.

Linking detections to pixels: Hough voting establishes correspondences between detections and the image evidence supporting them, which can avoid overusing image content for several detections [15, 1, 14, 34]. Overall performance of hough voting detectors remains comparatively low. [37, 5] combine detections with semantic labelling, while [36] rephrase detection as a labelling problem. Explaining detections in terms of image content is a sound formulation but these works rely on image segmentation and labelling, while our system operates purely on detections without additional sources of information. Co-occurrence. One line of work proposes to detect pairs of objects instead of each individual objects in order to handle strong occlusion [24, 29, 19]. It faces an even more complex NMS problem, since single and double detections need to be handled. [22] bases suppression decisions on estimated crowd density. Our method does neither use image information nor is it hand-crafted to specifically detect pairs of objects.

Co-occurrence: One line of work proposes to detect pairs of objects instead of each individual objects in order to handle strong occlusion [24, 29, 19]. It faces an even more complex NMS problem, since single and double detections need to be handled. [22] bases suppression decisions on estimated crowd density. Our method does neither use image information nor is it hand-crafted to specifically detect pairs of objects.

Auto-context: Some methods improve object detection by jointly rescored detections locally [30, 4] or globally [31] using image information. These approaches tend to produce fewer spread-out double detections and improve overall detection quality, but still require NMS. We also approach the problem of NMS as a rescored task, but we completely eliminate any post-processing.

Neural networks on graphs: A set of detections can be seen as a graph where overlapping windows are represented as edges in a graph of detections. [18] operates on graphs, but requires a pre-processing that defines a node ordering, which is ill-defined in our case.

End-to-end learning for detectors: Few works have explored true end-to-end learning that includes NMS. One idea is to include GreedyNMS at training time [33, 12], making the classifier aware of the NMS procedure at test time. This is conceptually more satisfying but does not make the NMS learnable. Another interesting idea is to directly generate a sparse set of detections, so NMS is unnecessary, which is done in [26] by training an LSTM that generates detections on overlapping patches of the image. At the boundaries of neighbouring patches, objects might be predicted from both patches, so post-processing is still required. [13] design a convnet that combines decisions of GreedyNMS with different overlap thresholds, allowing the network to choose the GreedyNMS operating point locally. None of these works actually completely remove GreedyNMS from the final decision process that outputs a sparse set of detections. Our network is capable of performing NMS without being given a set of suppression alternatives to choose from and without having another final suppression step.

3. Dataset

3.1 Data collection

The video footage was recorded on the roads located along the National Highways (NH) and Tamil Nadu State Highways (TNSH) in the districts of Cuddalore and Chengalpattu in Tamil Nadu, India. The video was shot using a Redmi5 and an iPhone12 mobile phone. To avoid any playback errors, the duration of the video file is limited to ten minutes. Additionally, some of the images used in the study were obtained from Google Images and other road datasets [43], and ethical considerations regarding filming the video were considered. Both the video and images are stored in Google Drive.

The images were originally in JPG format with varying resolutions. They were resized to 640 by 640 pixels to create square images and maintain consistency within the dataset. A total of 1640 images were captured under different weather conditions, including sunny, rainy, and winter. The images were also taken from various angles. Out of these, only 1129 images were deemed suitable for the study, as the remaining images either had poor color quality or did not clearly depict road damage.



Figure 3. State Highway Road Damage Dataset Sample.

3.2 Data Pre-Processing

TECHNIQUES FOR IOU NON-MAX SUPPRESSION TO IMPROVE ROAD SURFACE DAMAGE DETECTION ACCURACY

In this research, some of the key steps involved in data pre-processing for custom object detection:

- **Gathering and labelling data:** The first step is to gather a sufficient amount of data for training the model. This data should be labelled with the appropriate annotations that define the location and type of objects in the images.
- **Data cleaning:** Once the data has been gathered, it is important to perform data cleaning to remove any images that are of poor quality or contain incorrect annotations. This helps to ensure that the model is trained on high-quality data that is representative of the real-world scenarios.
- **Data augmentation:** Data augmentation is the process of artificially increasing the size of the training dataset by generating additional images from the original dataset. This helps to prevent overfitting and improves the generalization capabilities of the model.
- **Resizing and normalization:** The images in the dataset should be resized to a consistent size and normalized to ensure that the model receives consistent input. This helps to improve the training speed and accuracy of the model.
- **Data splitting:** The dataset should be split into training, validation, and testing sets. The training set is used to train the model, the validation set is used to tune the model hyperparameters, and the testing set is used to evaluate the final performance of the model.
- **Converting data to model input format:** The data should be converted to a format that can be input into the object detection model. This usually involves converting the images and annotations to a format that the model can read, such as YOLO format.

Here, sources images are various dimension which are mentioned below. Image size refers to the physical dimensions of an image. Scaling refers to the process of changing the size of an image while preserving its aspect ratio. Scaling can be done by either increasing or decreasing the number of pixels in the image. This paper includes a dataset of images of various sizes, as shown in Table 1.

Image sources	Dimension
RSDD2023	1280 x 870
Google Images	194 x 259
GRDD2020	600 x 600
Pothel_Dataset_V3	1280 x 780

Table 1. Dataset video file / image size

3.3 Damage Classification

The international study report [44] classifies various types of road damage. However, this research article specifically focuses on four categories of damages, as indicated in table 1, that are commonly found on Indian highways and considered significant. It should be noted that certain types of damage, including longitudinal construction joint part (D01), lateral construction joint part (D11), Cross walk blur (D43), and White line blur (D43), are not addressed in this study. The damages are identified and represented by image annotation labels, which are discussed below.

Damage Type			Detail	Class Name
Crack		Longitudinal	Wheel mark part	D00

TECHNIQUES FOR IOU NON-MAX SUPPRESSION TO IMPROVE ROAD SURFACE DAMAGE DETECTION ACCURACY

	Linear Crack			
		Lateral	Equal interval	D10
	Alligator		Partial Pavement, All Pavement'	D20
Other Corruption			Rutting, bump, pothole, Separation	D40

Table 2: Classification of Road surface Damages

3.4 Data Annotation

Data annotation involves the task of augmenting a dataset with metadata or labels to enhance its comprehensibility and usefulness for machine learning algorithms. Before the annotation process could commence, the dataset underwent a thorough examination to ensure the validity of all images, eliminate errors, and establish proper naming conventions, which consumed a substantial amount of time. The images were then annotated, with the exception of ambiguous ones, and assigned to their respective categories, as outlined in Table 1, which enumerates four distinct damage categories. To improve accessibility and usability, the images were systematically organized into folders and stored on Google Drive.

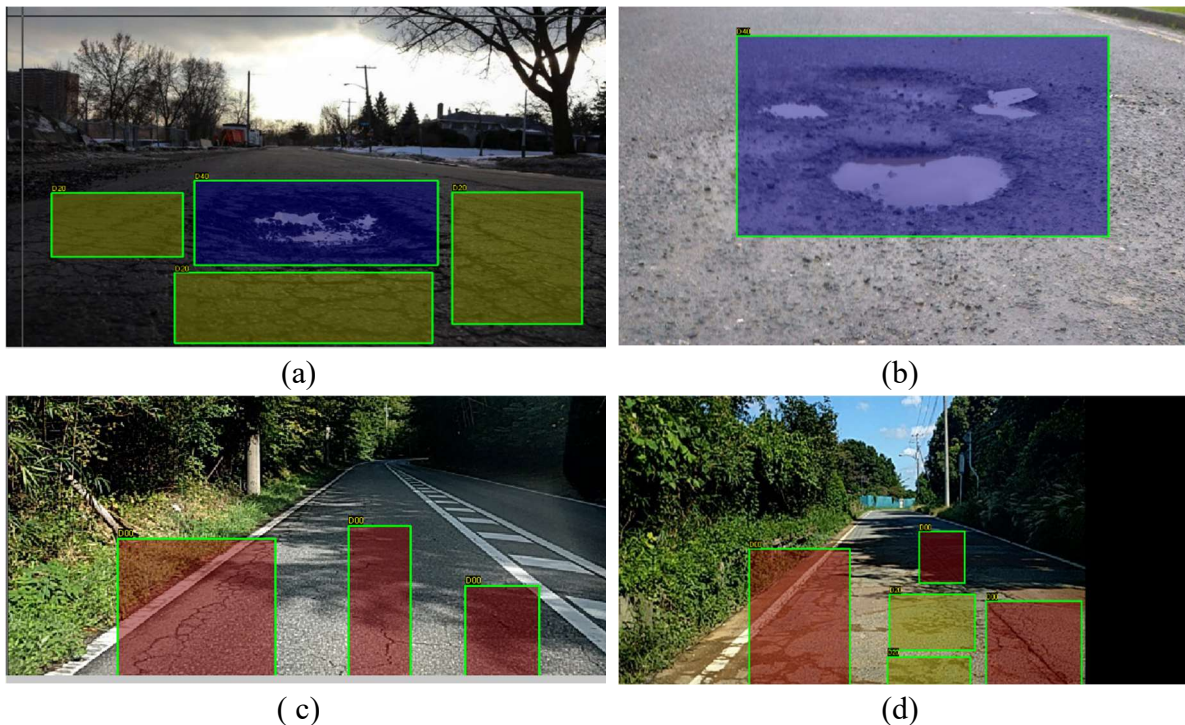


Figure 4. Damage class are marked with different colours.

3.5 Data statistics

The statistics for the Tamil Nadu Road datasets are shown in Figure 5. It should be noticed that datasets have an uneven distribution of occurrences for different damage classes. image augmentation techniques were used to create a balanced representation that can be used to train deep learning models.

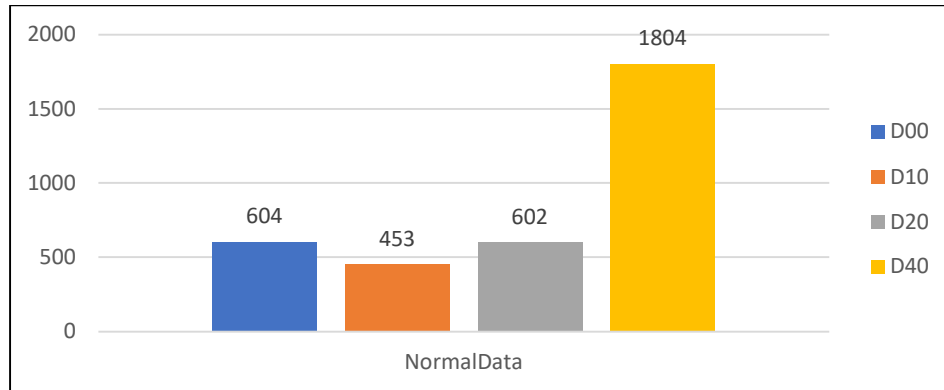


Figure 5. Statistics about the total of damage cases inside the underlying datasets.

4. Methodology

The proposed classification model was divided into five primary stages, as illustrated in Figure 3: data collection, data pre-processing and augmentation, implementation of several CNN models, experiment, and assessment. The damages were classified according to their types, pre-training models, and the proposed self-designed CNN model. Finally, this work evaluates and analyses the obtained results using the state-of-the-art performance metrics; the results are presented in Section 6.

4.1 Data Pre-processing

The sources images are various dimension which are mentioned below. Image size refers to the physical dimensions of an image. Scaling refers to the process of changing the size of an image while preserving its aspect ratio. Scaling can be done by either increasing or decreasing the number of pixels in the image. This paper includes a dataset of images of various sizes, as shown in Table 3 .

Image sources	Sizes
Redmi5 - mobile	854 x 480
Iphone12 -Video file	1280 x 780
Google images	268 X 188, 262 x 192, 271 x 186
GRDD2022 - Images	600 x 600

Table 3. Dataset video file / image size

4.2 Data Augmentation

Image data augmentation is a technique commonly used in computer vision tasks, such as image classification, object detection, and semantic segmentation. It involves applying a set of predefined transformations to existing images to create new variations of the original dataset. The purpose of data augmentation is to increase the size and diversity of the training data, which can lead to better model generalization and improved performance.

Here are some commonly used image data augmentation techniques are Horizontal/Vertical Flipping, Rotation, Scaling and Cropping, Translation, Shearing, Zooming, Adding Noise and Color Jittering. These techniques, along with others, can be applied individually or in combination to generate augmented images. It's important to strike a balance between introducing enough diversity to the dataset without distorting the original content too much, as excessively transformed images might confuse the model during training.

This research has challenges during the augmentation, for classes D00 & D10, some augmentations, such as flip rotation and left-to-right rotation, are invalid. The results are distorted when class D00 is turned to become class D10. Class D20 and D40 are exempt from image augmentation, nevertheless. All classes can benefit from contrast augmentation, which is a component of image augmentation. By enhancing 1129 base images, 2159 augmented images with 1791 bounding boxes are generated. As illustrated in Figure 4, these techniques were applied to each image to obtain a new training and testing sample. The training and testing portions of this dataset had a ratio of 80:20.

Classes	# Before Augment	# After Augment
D00	726	1316
D10	526	1013
D20	707	1450
D40	1908	3382
No. of Total Images	1129	2159
No. Rejected Images	538	711

Table 4. Statistical of Normal and Augmented Images

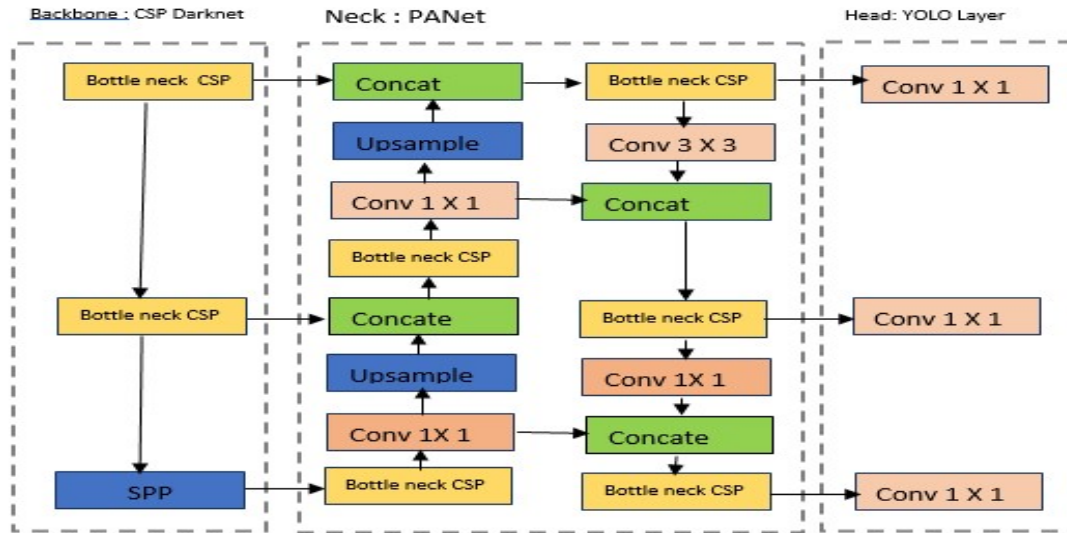
4.3 Architecture

In the YOLO object detection architecture, three key components—the backbone, neck, and head—collaborate to detect objects within an image[41]. The backbone assumes the role of the primary feature extractor. Typically, it is a deep convolutional neural network (CNN) employed to extract relevant features from the input image. This network is often pretrained on a large dataset like ImageNet, and subsequently fine-tuned specifically for the object detection task.

The neck establishes a connection between the backbone and the head. It generally comprises several layers of convolutional neural networks, serving to merge the features obtained from the backbone and prepare them for object detection. The neck is responsible for enhancing the feature maps derived from the backbone, employing techniques such as skip connections, pooling, and up-sampling.

TECHNIQUES FOR IOU NON-MAX SUPPRESSION TO IMPROVE ROAD SURFACE DAMAGE DETECTION ACCURACY

The head represents the final segment of the YOLO architecture and is accountable for object detection in the image. Typically, the head consists of several convolutional layers employed to predict bounding boxes and class probabilities for each identified object in the image. The head takes the features extracted by the neck and utilizes a set of fully connected layers to generate the ultimate predictions[41].



CSP: Cross Stage Partial Network Conv: Convolutional layer SPP: Spatial Pyramid
Pooling Concat: Concatenate function

Figure 6. Network architecture of YOLO

4.4 IoU NMS Algorithm

A typical object detection pipeline has one part that produces classification suggestion. The candidate regions for the object of interest are what constitute proposals. The majority of methods assign foreground/background scores based on the features computed in a sliding window over the feature map. The neighbouring windows are taken into consideration as potential regions because they have somewhat comparable scores. However, it takes a long time for the categorization network to process all these ideas. This results in a method called Non-maximum Suppression, which filters the ideas according to specific criteria.

NMS Input: A list of Proposal boxes B , corresponding confidence scores S and overlap threshold N .

Output: A list of filtered proposals D .

Algorithm

1. Choose the proposal with the highest confidence score, take it out of list B , and add it to list D with all the proposals. D is initially empty.
2. Next, compare this proposal to all the other ideas by calculating the IOU (Intersection over Union) between each one. Remove the proposal from B if the IOU exceeds the N -point cut-off.
3. Again, choose the proposal in B that you have the most faith in, take it out of B , and add it to D .

4. Calculate the IOU of this proposal once more using all the proposals in B, and then cross off any boxes with IOU levels above the threshold.
5. This procedure is repeated until B is empty of suggestions.

IOU calculation is used to measure the overlap between two proposals.

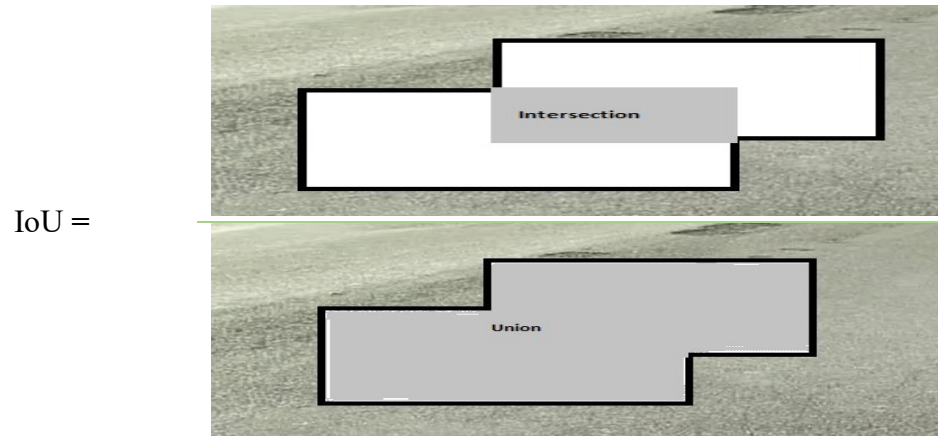


Figure 7 Intersection over union

IoU of 1 means your predicted bounding box perfectly matches the ground truth box. IoU of 0 means no part of your predicted bounding box overlaps with the ground truth box.

Pseudo code (IoU NMS)

- 1: *procedure* NMS (B,c)
- 2: $B_{nms} \leftarrow \emptyset$ Initialize empty set
- 3: *for* $b_i \in B$ *do* \Rightarrow Iterate over all the boxes
- 4: discard \leftarrow False Take Boolean variable and set it as false. This variable indicates whether $b(i)$
- 5: *for* $b_j \in B$ *do* should be kept or discarded
- 6: *if* $\text{same}(b_i, b_j) > \lambda_{nms}$ *then* if both boxes having same IOU
- 7: *if* $\text{score}(c, b_j) > \text{score}(c, b_i)$ *then*
- 8: discard \leftarrow True Compare the scores. If score of $b(i)$ is less than that of $b(j)$, $b(i)$ should be discarded, so set the flag to True.
- 9: *if not* discard *then* Once $b(i)$ is compared with all other boxes and still the discarded flag is False, then $b(i)$ should be considered. So add it to the final list.
- 10: $B_{nms} \leftarrow B_{nms} \cup b_i$ Do the same procedure for remaining and return the final list
- 11: *return* B_{nms}

TECHNIQUES FOR IOU NON-MAX SUPPRESSION TO IMPROVE ROAD SURFACE DAMAGE DETECTION ACCURACY

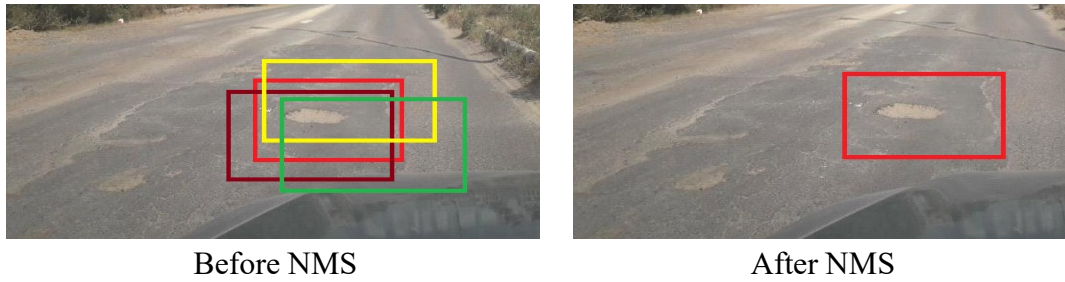


Figure 8 NMS implementation

the whole bounding box filtering process depends on single threshold value. So, selection of threshold value is key for performance of the model.

4.4 Objects Training

YOLO is an end-to-end architecture that can detect objects in an image and simultaneously classify them into predefined categories, all in a single pass. YOLO divides an input image into a grid of cells (8 X 8) and predicts bounding boxes 'B' and class probabilities for each cell. Each bounding box prediction includes the coordinates of the object's centre, width(B_w), and height(B_h). The class probabilities(P_c) indicate the likelihood of the object belonging to each of the predefined classes(C_1 , C_2 , C_3 and C_4).

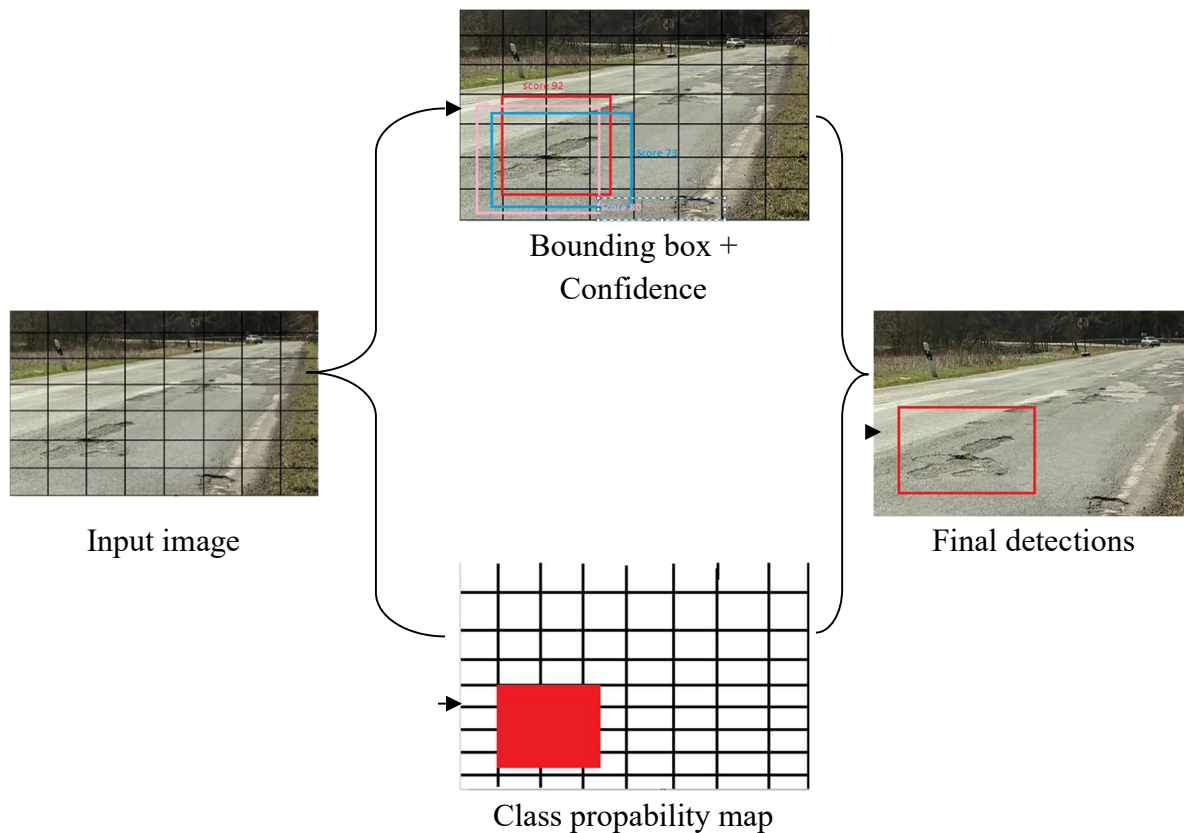


Figure 9. Steps of Object Detection

TECHNIQUES FOR IOU NON-MAX SUPPRESSION TO IMPROVE ROAD SURFACE DAMAGE DETECTION ACCURACY

YOLO uses a single convolutional neural network (CNN) to process the entire image, which makes it very fast compared to other object detection architectures.

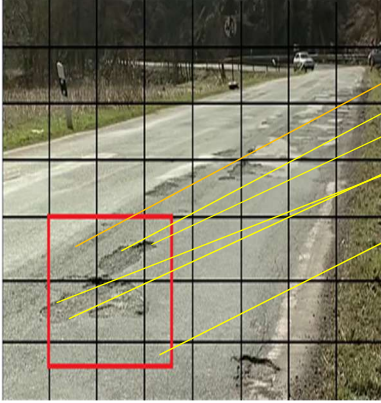
	P	0	1	1	1	1	0	0	0	0	Object Indicator
	b	?	30	35	25	25	?	?	?	?	Bounding box
	x		0	0	0	0					
	b	?	32	32	22	18	?	?	?	?	
	y		0	0	0	0					
	b	?	45	20	20	31	?	?	?	?	Class Label
	w		0	0	0	0					
	b	?	25	20	60	18	?	?	?	?	
	h		0			0					
	C	?	0	0	0	0	?	?	?	?	Class Label
	1										
	C	?	0	0	0	0	?	?	?	?	
	2										
	C	?	0	0	0	0	?	?	?	?	Class Label
	3										
	C	?	1	1	1	1	?	?	?	?	Class Label
	4										

Figure 10. Objects annotated by annotation tool.

4.5 Training model

The model was constructed using a dataset, and subsequently, it was evaluated using a set of testing images. The training model's logs and metrics were stored in a repository for reference. The focus of this paper is to investigate the intricacies of IoU NMS and assess its effectiveness. However, it should be noted that there is a lack of comprehensive information regarding the preparation of the training model.

4.6 Real-time Object detection by Trained model

The authenticity of real-time objects is confirmed through the utilization of a trained model. The object detection process is then assessed by comparing the results with a ground truth table. Further information regarding this evaluation is provided below.

True Positive and False Positive

Once the final predictions are established, the forecasted bounding boxes can be evaluated against the actual bounding boxes to calculate mean Average Precision (mAP) and assess the performance of the object detector. This evaluation involves identifying the number of true positives. A predicted bounding box is considered a true positive if it overlaps a ground truth bounding box by an Intersection over Union (IOU) threshold of 0.5, indicating a successful detection. Conversely, if a predicted bounding box has an overlap

TECHNIQUES FOR IOU NON-MAX SUPPRESSION TO IMPROVE ROAD SURFACE DAMAGE DETECTION ACCURACY

with a ground truth bounding box below the threshold, it is classified as a false positive and deemed an unsuccessful detection. The precision and recall metrics can then be computed based on the true positives and false positives, using the following formulas:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} = \frac{Count(True\ Positive)}{Count(True\ Positive + False\ Positive)}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} = \frac{Count(True\ Positive)}{Count(True\ Positive + False\ Negative)}$$

False Negative and True Negative

For each predicted bounding box within the image:

If the predicted class of the bounding box does not match any of the target classes in the image, label the bounding box as a false positive. If the predicted class matches one of the target classes, Compare the predicted bounding box with all the target boxes in the image. Determine the highest overlap between the predicted bounding box and a target box using the IOU threshold. If the highest overlap exceeds the IOU threshold, consider the target box as Successfully detected and record the predicted bounding box as a true positive. If the highest overlap is below the IOU threshold, mark the bounding box as a false positive. Store the successfully detected target box and proceed to the next predicted bounding box. Return the objectness score, predicted class, and a flag indicating if each prediction is a true positive.

F1 Score

The F1 score determines the optimal confidence score threshold that maximizes the balance between precision and recall. By calculating the F1 score, we can assess the trade-off between precision and recall. A higher F1 score indicates high precision and recall, while a lower F1 score suggests the opposite.

$$F1\ Score = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)}$$

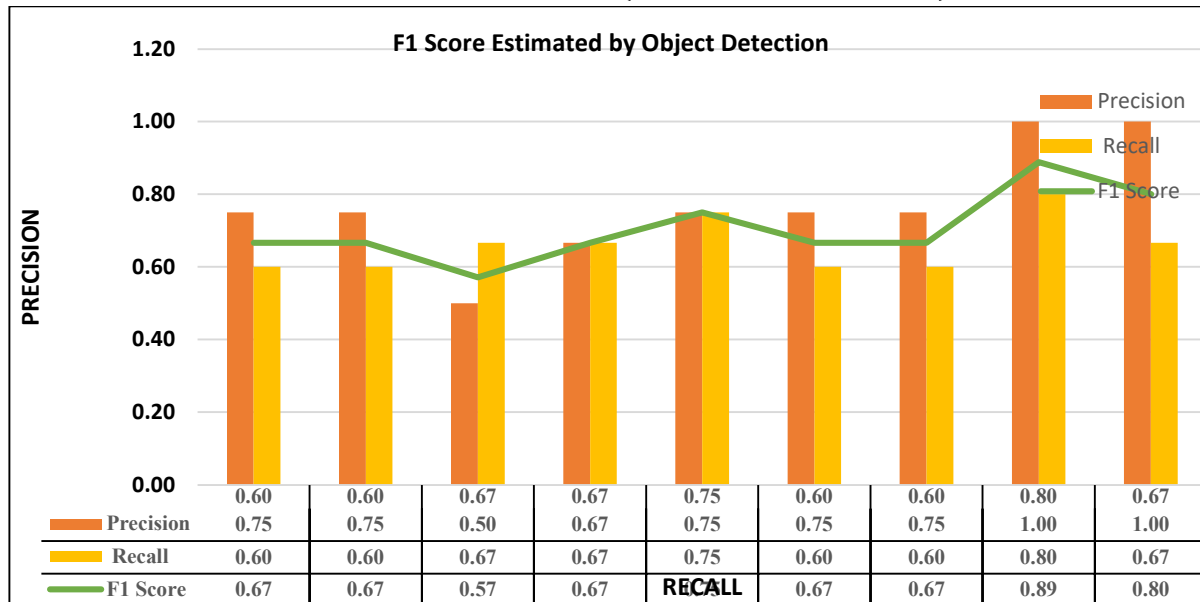


Figure 11. F1 Score estimated by Object detection.

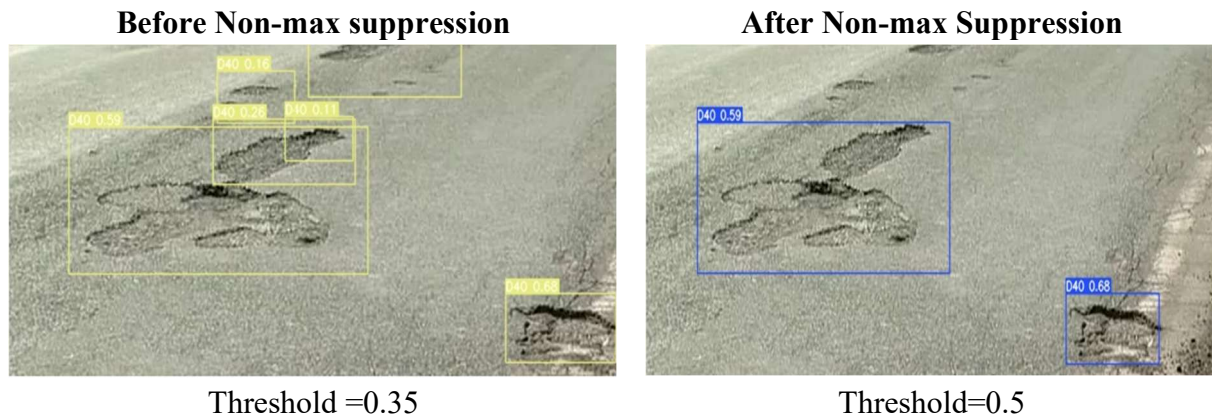


Figure 12. Bounding box eliminated by IoU NMS parameter.

5. Experimental setup

This research's data was analysed using the GPU environment of the NVIDIA Jetson nano hardware device. Since the CPU was unable to process the image data, the GPU environment was used. Most of the time training model prepared on Colab Research environment, which is another environment that is also available. Depending on availability, Google Colab gives users access to one of several different NVIDIA GPUs. This study's processing was done on a Tesla T4 processor running Torch 1.10.2+cu111 CUDA:0 with 12 GB of RAM. used Google Drive, which has a default 15 GB storage restriction, for storing files.

6. Result

The result of IOU NMS is a set of bounding boxes that have been filtered to remove redundant detections. The specific outcome of IOU NMS depends on the implementation and the parameters used. Typically, the process involves the following steps:

- Sorting: The initial bounding box detections are sorted based on their confidence scores or probabilities, usually in descending order.
- Selection: The bounding box with the highest confidence score (or probability) is selected as the starting point.
- Overlap calculation: The IOU scores are computed between the selected bounding box and all remaining boxes.
- Thresholding: Bounding boxes with IOU scores above a certain threshold (e.g., 0.5 or 0.7) are considered redundant and suppressed.
- Iteration: The process is repeated, selecting the next highest-scoring bounding box from the remaining detections and suppressing any overlapping boxes.
- Final selection: The resulting set of non-overlapping bounding boxes is considered the output of IOU NMS.

The results of the detected outputs are recorded in a git repository and the object detection in this study is evaluated by changes in IOU NMS threshold values with score confidence. URL: https://github.com/PalaniRamu/IoU_Non_Max_Suppression.git

7. Conclusion

IoU Non-Max Suppression (IoU NMS) is an essential post-processing step in road surface object detection that helps improve accuracy by removing redundant detections and retaining only the most reliable ones based on confidence scores and overlap thresholds.

References

- [1] O. Barinova, V. Lempitsky, and P. Kholi. On detection of multiple object instances using hough transforms. PAMI, 2012.
- [2] L. Bourdev, S. Maji, T. Brox, and J. Malik. Detecting people using mutually consistent poselet activations. In ECCV, 2010.
- [3] G. Burel and D. Carel. Detection and localization of faces on digital images. Pattern Recognition Letters, 1994.
- [4] G. Chen, Y. Ding, J. Xiao, and T. X. Han. Detection evolution with multi-order contextual co-occurrence. In CVPR, 2013.
- [5] J. Dai, K. He, and J. Sun. Convolutional feature masking for joint object and stuff segmentation. In CVPR, 2015.
- [6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, 2005.
- [7] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. PAMI, 2010.
- [8] J. Ferryman and A. Ellis. Pets2010: Dataset and challenge. In AVSS, 2010. 6
- [9] R. Girshick. Fast R-CNN. In ICCV, 2015.
- [10] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR, 2014.
- [11] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. In ECCV, 2016. 5
- [12] P. Henderson and V. Ferrari. End-to-end training of object class detectors for mean average precision. In ACCV, 2016.
- [13] J. Hosang, R. Benenson, and B. Schiele. A convnet for non-maximum suppression. In GCPR, 2016. 2, 3,4, 6
- [14] P. Kotschieder, S. Rota Bulò, M. Donoser, M. Pelillo, and H. Bischof. Evolutionary hough games for coherent object detection. CVIU, 2012.
- [15] B. Leibe, A. Leonardis, and B. Schiele. Robust object detection with interleaved categorization and segmentation. IJCV, 2008.
- [16] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, and S. Reed. Ssd: Single shot multibox detector. In ECCV, 2016. 1
- [17] D. Mrowca, M. Rohrbach, J. Hoffman, R. Hu, K. Saenko, and T. Darrell. Spatial semantic regularisation for large scale object detection. In ICCV, 2015.

- [18] M. Niepert, M. Ahmed, and K. Kutzkov. Learning convolutional neural networks for graphs. In ICML, 2016.
- [19] W. Ouyang and X. Wang. Single-pedestrian detection aided by multi-pedestrian detection. In CVPR, 2013.
- [20] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In CVPR, 2016.
- [21] S. Ren, K. He, R. Girshick, and J. Sun. Faster RCNN: Towards real-time object detection with region proposal networks. In NIPS, 2015.
- [22] M. Rodriguez, I. Laptev, J. Sivic, and J.-Y. Audibert. Density-aware person detection and tracking in crowds. In ICCV, 2011.
- [23] R. Rothe, M. Guillaumin, and L. Van Gool. Nonmaximum suppression for object detection by passing messages between windows. In ACCV, 2014.
- [24] M. A. Sadeghi and A. Farhadi. Recognition using visual phrases. In CVPR, 2011.
- [25] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. In ICLR, 2014.
- [26] R. Stewart and M. Andriluka. End-to-end people detection in crowded scenes. In CVPR, 2016.
- [27] S. Tang, B. Andres, M. Andriluka, and B. Schiele. Subgraph decomposition for multi-target tracking. In CVPR, 2015.
- [28] S. Tang, M. Andriluka, A. Milan, K. Schindler, S. Roth, and B. Schiele. Learning people detectors for tracking in crowded scenes. In ICCV, 2013.
- [29] S. Tang, M. Andriluka, and B. Schiele. Detection and tracking of occluded people. In BMVC, 2012.
- [30] Z. Tu and X. Bai. Auto-context and its application to high-level vision tasks and 3d brain image segmentation. PAMI, 2010.
- [31] A. Vezhnevets and V. Ferrari. Object localization in imagenet by looking out of the window. In BMVC, 2015.
- [32] P. Viola and M. Jones. Robust real-time face detection. In IJCV, 2004.
- [33] L. Wan, D. Eigen, and R. Fergus. End-to-end integration of a convolutional network, deformable parts model and non-maximum suppression. In CVPR, 2015.
- [34] P. Wohlhart, M. Donoser, P. M. Roth, and H. Bischof. Detecting partially occluded objects with an implicit shape model random field. In ACCV, 2012.
- [35] C. Wojek, G. Dorkó, A. Schulz, and B. Schiele. Sliding-windows for rapid object class localization: A parallel technique. In DAGM, 2008.
- [36] J. Yan, Y. Yu, X. Zhu, Z. Lei, and S. Z. Li. Object detection by labeling superpixels. In CVPR, 2015.
- [37] J. Yao, S. Fidler, and R. Urtasun. Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation. In CVPR, 2012.
- [38] Jan Hosang¹, Rodrigo², Benenson³, Max Planck Institut für Informatik Saarbrücken, Germany, arXiv:1705.02950v2 [cs.CV] 9 May 2017.
- [39] R. Palani¹, Dr. N. Puviarasan² and Dr. A. Rama Prasath³, Literature Review of Road Damage Detection with Repairing Cost Estimation, International Journal of Mechanical Engineering, ISSN: 0974-58232, Vol. 7 No. 2 February 2022

- [40] R. Palani¹ , Dr. N. Puviarasan² and Dr. A. Rama Prasath³ , Collection of distinct image frames using Structured similarity Index measured, Advanced Engineering Science, ISSN: 2096-3246 | Jan 2023
- [41] R. Palani¹ , Dr. N. Puviarasan² and Dr. A. Rama Prasath³, IMPROVE CUSTOM OBJECT DETECTION OF ROAD SURFACES BY SMOOTHING OF LABELS AND IMAGE ENHANCEMENT, Journal of Data Acquisition and Processing, ISSN: 1004-9037| Vol. 38 (1) 2023
- [42] R. Palani¹ , Dr. N. Puviarasan² and Dr. A. Rama Prasath³, Road Surface Damage Detection With Ensemble Techniques, Scandinavian Journal of Information Systems | ISSN: 0905-0167|1901-0990| 2023 35(1)
- [43] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” arXiv preprint arXiv:2207.02696, 2022.
- [44] Y. Du, N. Pan, Z. Xu, F. Deng, Y. Shen, and H. Kang, “Pavement distress detection and classification based on yolo network,” International Journal of Pavement Engineering, pp. 1–14, 2020.