DIRT AND WATER DETECTION FOR FLOOR CLEANING ROBOTS

Manohar Akula

***Abstract*—Modern floor-cleaning robots are becoming more and more complex, and when equipped with digital cameras and a powerful vision system, they can navigate and assess their surroundings with more independence. The YOLOv7 and YOLOv7 tiny frameworks are used to propose a vision system that can find water and stains on floors. The vision system's goal is to reduce wasteful energy use and material consumption by activating the robot's cleaning system only when a filthy region is spotted and then using just as much cleaning supplies as necessary to get the job done. I'll be making use of data from Flobot's FLOBOT Perception Dataset. Using their dataset as a starting point, I developed a model to identify water and dirt on the floor, and then propose a vision system based on CV frameworks. On a real dataset (ACIN), the best model achieved a mean average precision map@0.5 attained 0.6775 and the map@:0.5:0.95 obtained 0.2541.**

***Index Terms*—Floor cleaning, Image Classification, Machine Learning, Neural Networks, and Deep Learning.**

I. Introduction

Robotic floor cleaners equipped with digital cameras to identify uncleaned or dirt areas are an emerging technology. More recently, camera-based mapping has been investigated in floor-cleaning robots as a supplementary navigation sensor [1]. It makes sense to put these camera-equipped floor-cleaning robots to use for anything besides navigation. The majority of these additional duties include searching for stains. The literature suggests that researchers are aiming to detect filthy places to save robot resources [2] to differentiate between solid and liquid dirt, and to identify dirt items from valuable goods [3].

For my research, I chose to use the FLOBOT perception dataset, which includes 2174 photos taken at a pilot supermarket site and including dirt and wetness on the floor. There are 1571 pictures showing dirty or wet floors, whereas just 603 show clean ones. The term ”dirt” is used to describe a wide variety of materials found on the floor, including but not limited to: coffee grounds, chip crumbs, paper scraps, curds, milk, soda, and other liquids. Water, it characterizes the puddles on the ground. The resolution of a screen specifies the maximum number of horizontal and vertical pixels that may be shown at once. I have utilized YOLO in my project. The convolutional neural network (CNN) aids the YOLO object identification technique. The inspiration for this architecture came from algorithms like the Region Based Convolutional Neural Network (RCNN) [4], which suggest areas of interest in the picture to be examined and then feed those regions into a CNN to identify objects. The YOLO framework provided a solution to this problem by partitioning the input picture into a grid, with the idea that for each block, a bounding box and its probability of belonging to a given class could be created. This would save on processing costs. The picture just has to be processed by the CNN once, making this method substantially quicker than alternatives like RCNN, which need extra stages like region proposals. The YOLO framework not only outperforms its competitors in terms of speed, but also boasts state-of-the-art outcomes as assessed by the mAP in a number of benchmarks like COCO [5]. In light of this, the study proposing the use of YOLOv7 to identify dirt was able to achieve state-of-the-art results on the use of floor-cleaning robots by approaching the issue as an object detection problem. The following were the most common problems discussed in relation to robot floor cleaners: Wide ranges in luminosity. Intricate floor plans. Not many soiled-situation captioned pictures exist. Image blurring caused by the robot's rapid motion. Class distinction between dirty and clean.

For the sake of overcoming the obstacles I was able to manually annotate the whole dataset into two classes, such as ”Water” and ”Dirt,” using an annotation tool (https://www.makesense.ai/). Each picture is being annotated with many border boxes.

What's so special about YOLOv7? To respond to this inquiry The YOLOv7 team aimed to improve upon previous work in object identification by developing a network architecture that could properly forecast bounding boxes at a faster pace than competing methods. YOLOv7 advances the state of the art in object detection with quicker and more accurate inferences compared to prior versions (i.e. YOLOv5). Comparisons with other real-time object identification models indicate that YOLOv7 models infer at a quicker rate (x-axis) and with higher accuracy (y-axis).

A. Related Work

In recent years, a variety of industrial applications, including automated monitoring, control, management, and maintenance, have been created and implemented.

The authors of [6] discuss recent developments in the Automatic System for Visual Detection of Dirt Buildup on Conveyor Belts Using Convolutional Neural Networks field. A convolutional neural network is trained using RGB pictures in the approach they suggested. It has shown to be quite successful to retrain consolidated networks for image classification using our gathered photos using the transfer learning approach. They came to the conclusion that identifying dirt buildup using ML can help with the inspection process of BC buildings by reducing uptime and assisting decision-making in maintenance sectors.

The researchers behind A Deep Learning-Based Dirt Detection Computer Vision System for Floor-Cleaning Robots with

Improved Data Collection evaluate the state-of-the-art in [7]. With a mAP of 0.488 and 0.733, respectively, they came to the conclusion that the YOLOv5m6 achieved the greatest results on both multi-class and binary classification.

I was unable to locate any similar works to identify transparent things like water on the floor or a conveyor belt, other from the pertinent results reported in the literature. Photoelectric sensors find it challenging to obtain steady sensing conditions when dealing with transparent and translucent things. Using an annotation tool, https://www.makesense.ai/, I manually annotated the entire dataset into 2 classes, such as ”Water” and ”Dirt,” to identify the various classes on the floor. It took me about 3 days to name them all. Each image has many border boxes added during annotation. The lowest and maximum x, y coordinate values for the whole picture are provided by these boundary boxes, which are made of either dirt or water. Later, I translated each image’s border box values into an XML file. The data original is where these xml files were kept. Later, to identify the raw data and preserve it inside the preprocessed, xml files were employed. The best model on a real dataset (ACIN) produced a mean average accuracy of 0.6775 for map@0.5 and 0.2541 for map@:0.5:0.95.

II. EXPLORATORY DATA ANALYSIS

The dataset was obtained from the European business FLOBOT (Floor Cleaning Robot) at http://lcas.github.io/FLOBOT. The collection includes 2174 examples of interior supermarket photos. The dataset was also limited to simply the pictures of the wet and dirty floors. Between 1 and 1572, the floor is covered with either water or water and dirt. To be more specific, the photographs with dirt had chips, coffee, ice cream, and soda all over the floor. Additionally, the water in the photographs is merely water that has spilled onto the ground. The last 600 pictures have a tidy floor. To determine the substance of the floor, I manually assigned soil and water classes to each photograph from 1 and 1572.

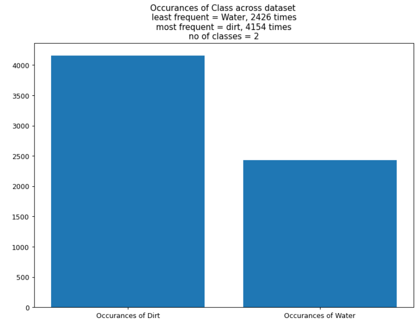
Initial detection of classes like Water and Dirt in the images was made using the annotation box values, as seen in the figures below.

Fig.1: Separation of Dirt and Water classes from the Image

A Qualitative Analysis was performed on the extracted classes and the results are shown in the below figure.

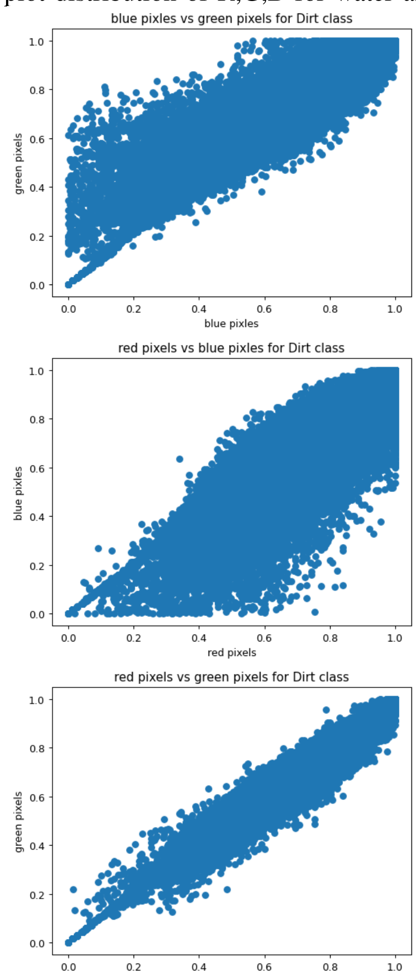
- Class across dataset :least frequent = Water, 2426 times
- most frequent = dirt, 4154 times

Fig.2: Class Distribution of Water and Dirt classes



Since my study involves image classification, the quantitative analysis I used to extract some characteristics from the image dataset had no bearing on the outcomes. Finally, I can determine the minimum, maximum, and median pixel values for dirt and water.

Fig.3: Scatter plot distribution of R,G,B for water and Dirt classes



Quantitative Analysis: For red pixels:

- Min Value = 0.0
- Max Value=1.0
- Median=0.5529412031173706

For blue pixels:

- Min Value = 0.0

- Max Value=1.0
- Median=0.545098066329956

For green pixels:

- Min Value = 0.0
- Max Value=1.0
- Median=0.5529412031173706

III. Methodology

The concept of segmenting one image into smaller images is the foundation of YOLO. The picture has been divided into an SS square grid. The cell in charge of detecting an item is the one that contains the object's center. B-boxes with a confidence rating for each box. For this design, the model predicts two bounding boxes by default. The classification score will range from "0.0" to "1.0," with "0.0" denoting the lowest confidence level and "1.0" denoting the greatest. If there is no item in the cell, the confidence score should be "0.0," and the score should be "1.0" if the model is 100 percent confident in its prediction.

For effective inference speed, the convolutional layers in the backbone of the YOLO network must be efficient. Cross Stage Partial Networks were WongKinYiu's first step toward maximizing layer efficiency.

The authors of YOLOv7 [8] extend on prior research in this area, taking into account both the memory requirements for maintaining layers in memory and the length of time it takes for a gradient to back-propagate across layers. Their network will be able to learn more effectively the smaller the gradient. They settle on E-ELAN, an extended version of the ELAN computational block, as their final layer aggregate.

To train, test, and forecast the output for YOLOv7 and YOLOv7 small models, the following procedures must be taken.

- Step-1: Creating an environment
- Step-2: Installation required libraries and software's
- Step-3: Model modification or tuning hyperparameters
- Step-4:Training:
- Step-5: Predicition:

IV. Results and Discussion

A. Model-1: YOLOv7 Pretrained weights

Construction: Changing the hyperparameters of the training model with batch-size 10, epochs 100, img size 640*640, and pretrained weights Yolov7.pt. The following results were attained with the Model 1. Here the whole dataset was used and achieved highest mAP for model prediction. After looking at the results the dirt and water was predicted accurately with high confidence. Please refer to the video in evaluations/sample.png

The mAP@0.5 started at 0.0002746 and map@0.5:0.95 started at 4.554e-05 intially after training the 100 epochs the map@0.5 achieved 0.6775 and the map@:0.5:0.95 achieved 0.2541. After testing this model with a raw video input the model is able to detect the dirt and water accurately. Please refer to the video in the evaluations folder.

For the model-1 initial values for the mAP@0.5 and map@0.5:0.95 in the model one are 0.0002746 and 4.554e-05, respectively. After training for 100 epochs, the map@0.5

Fig.4: Prediction of Dirt and Water classes on test data with YOLOv7

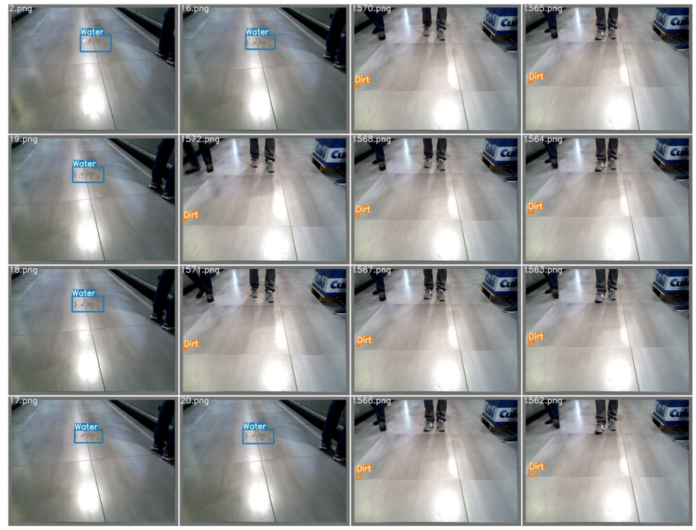


Fig.5: Results of YOLOv7

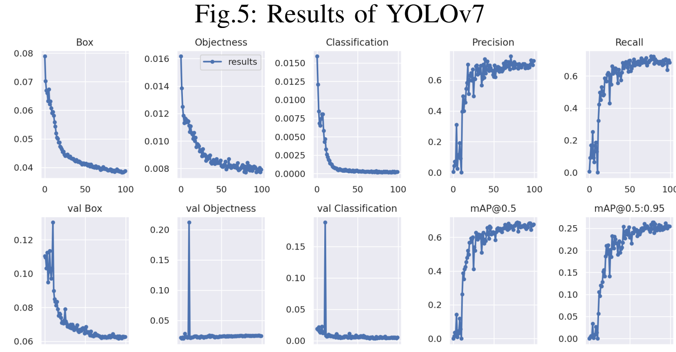
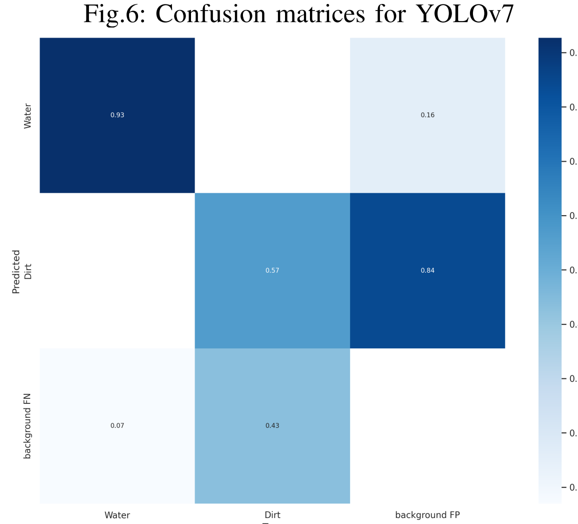


Fig.6: Confusion matrices for YOLOv7



attained 0.6775 and the map@:0.5:0.95 obtained 0.2541. This model can detect the wetness and dirt with accuracy after being tested with a raw video input. Refer to the evaluations folder's video for further information.

B. Model-2: YOLOv7 tiny Pretrained weights

Construction: Altering the hyperparameters of the Yolov7 tiny.pt training model with batch-size 8, epochs 100, and 640x640 image size. The Model 2 produced the following outcomes.

Here the model was trained with the 50 percent of data with the yolov7 pretrained weights and achieved highest mAP as compared to the previous model. But while predicting the dirt and water on the floor the model detected them moderatly with a less confidence.

The mAP@0.5 started at 0.0002746 and map@0.5:0.95 started at 4.554e-05 intially after training the 100 epochs the map@0.5 achieved 0.6775 and the map@:0.5:0.95 achieved 0.2541. After testing this model with a raw video input the model is able to detect the dirt and water accurately. Please refer to the video in the evaluations folder.

Fig.7: Prediction of classes on test data with YOLOv7-tiny

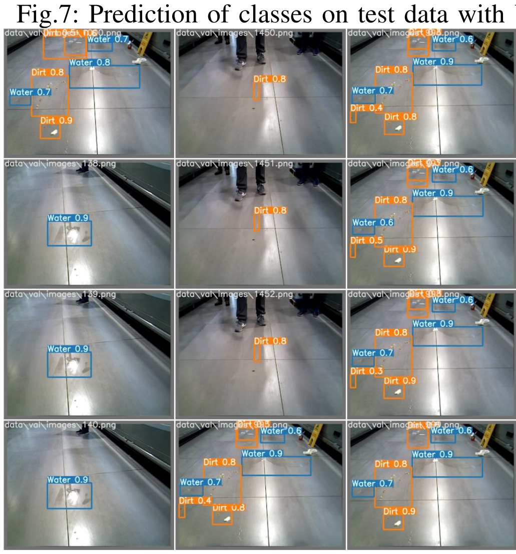
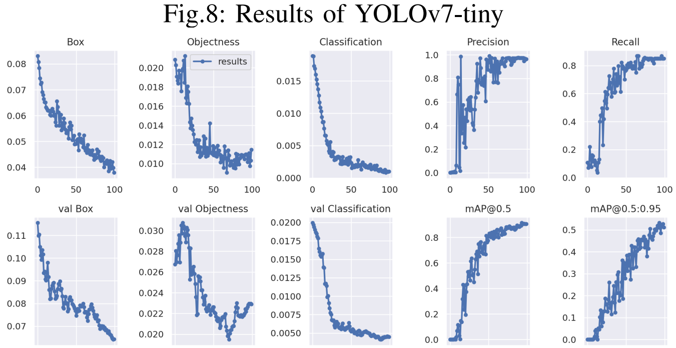
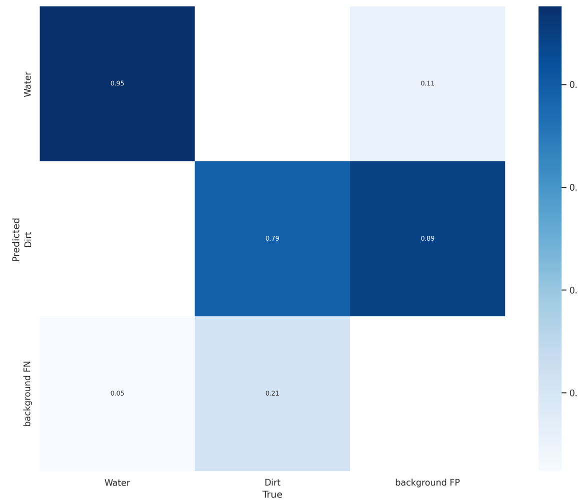


Fig.8: Results of YOLOv7-tiny



The mAP@0.5 started at 0.0002 and map@0.5:0.95 started at 5.324e-05 intially after training the 100 epochs the map@0.5 achieved 0.7827 and the map@:0.5:0.95 achieved 0.3958 . After testing this model with a raw video input the model is not able to detect the dirt and water accurately

Fig.9: Confusion matrices for YOLOv7 tiny



REFERENCES

[1] Kang, M.C.; Kim, K.S.; Noh, D.K.; Han, J.W.; Ko, S.J. A robust obstacle detection method for robotic vacuum cleaners.IEEE Trans. Consum. Electron. 2014, 60, 587–595.

[2] Grünauer, A.; Halmetschlager-Funek, G.; Prankl, J.; Vincze, M. The power of GMMs: Unsupervised dirt spot detection for industrial floor cleaning robots. In Towards Autonomous Robotic Systems; Gao, Y., Fallah, S., Jin, Y., Lekakou, C., Eds.; Springer: Cham, Switzerland; Guildford, UK, 2017; pp. 436–449

[3] Connolly, C.; Fleiss, T. A study of efficiency and accuracy in the transformation from RGB to CIELAB color space. IEEE Trans. Image Process. 1997, 6, 1046–1048. ACM SIGMOD International Conference on Management of Data, SIGMOD '15, pages 1383– 1394, New York, NY, USA. ACM.

[4] Riese, F.M.; Keller, S. Soil Texture Classification with 1D Convolutional Neural Networks based on Hyperspectral Data. arXiv 2019, arXiv:1901.04846.

[5] BBhatia, Y.; Rai, R.; Gupta, V.; Aggarwal, N.; Akula, A. Convolutional neural networks based potholes detection using thermal imaging. J. King Saud Univ. Comput. Inf. Sci. 2019.

[6] André A. Santos, Filipe A. S. Rocha , Agnaldo J. da R. Reis and Frederico G. Guimarães Automatic System for Visual Detection of Dirt Buildup on Conveyor Belts Using Convolutional Neural Networks //https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7601099/pdf/sensors-20-05762.pdf

[7] Canedo D, Fonseca P, Georgieva P, Neves AJR. A Deep Learning-Based Dirt Detection Computer Vision System for Floor-Cleaning Robots with Improved Data Collection. Technologies. 2021; 9(4):94. https://doi.org/10.3390/technologies9040094

[8] Chien-Yao Wang1 , Alexey Bochkovskiy, and Hong-Yuan Mark Liao1 1 Institute of Information Science, Academia Sinica, Taiwan. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors //https://doi.org/10.48550/arXiv.2207.02696