Journal of Theoretical and Applied Information Technology
31st August 2025. Vol.103. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

THE INSPECTION OF VARIOUS INPUT TRANSFORMATIONS TOWARDS HUMAN DETECTION AND TRACKING

TANVEER KADER[1], AHMAD FAKHRI AB. NASIR [2*], ANWAR P.P. ABDUL MAJEED[3], M. ZULFAHMI TOH[4], NOORLIN MOHD ALI[5]

[1,2,4,5]Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia
[2]Centre for Artificial Intelligence & Data Science, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Gambang, Kuantan, Pahang, Malaysia
[3]School of Engineering and Technology, Sunway University, Bandar Sunway, 47500, Selangor, Malaysia
E-mail: [1]tanveerkaderedu@gmail.com, [2]afakhri@umpsa.edu.my, [3]anwarm@sunway.edu.my, [4]zulfahmi@umpsa.edu.my, [5]noorlin@umpsa.edu.my
** corresponding author*

ABSTRACT

Considering the role of color space in enhancing object detection this paper finds the impact of detection models trained on different color spaces in human tracking system within the Tracking by detection (TBD) framework. A customized dataset over 8k frames including indoor and outdoor human movement videos was developed following the MOT15 structure. YOLO12s was fine-tuned separately on six color spaces RGB, Grayscale, HSV, HSI, YCbCr and YES followed by SORT and DeepSORT for tracking. YOLO12s trained on RGB provides the best MOTA with 31.5% and 50% for SORT and DeepSORT respectively. Competitive results observed from the Grayscale color space.

Keywords: *Human tracking, tracking-by-detection, color spaces, SORT and DeepSORT*

1. INTRODUCTION

Multi object tracking (MOT) is a fundamental task in computer vision and plays an important role in the development of human tracking systems. Unlike single object tracking which focuses on one target, the MOT system detects all individuals in a video frame and consistently tracks their identities across time. Tracking by detection (TBD) is a commonly used framework in modern MOT systems where tracking process is followed by the detection of objects generated in each video frame. The initial studies on human detection can be found in references [1]–[3]. In this process, it must deal with certain challenges like occlusion, motion blur, rapid movement etc.

Color spaces such as Grayscale, YCbCr, HSI, HSV, YES etc. can sometimes play an effective role in improving the detection quality of detection models depending on the detection target like greenery, certain color of vegetables, skin color etc. Each of these color modes offer special features such as RGB preserves real world visual similarity in colors, Grayscale can simplify image texture, HSV separates intensity from color, HSI is almost similar to HSV but more convenient to human vision, YCbCr separates luminance from the chrominance components that helps to handle brightness and color independently. As discussed in Section 2 many detection systems improved their accuracy using several color methods [5]–[18]. Still, in tracking system most of the work focuses on model architectures or optimizing performance on standard RGB videos [19]–[25], the impact of other color spaces left unexplored which led the research to find answers to the followings:

(1) How do various color spaces influence the tracking systems?

(2) Which is the best color space providing the highest tracking accuracy and why?

In tracking system while most of the work focuses on model architectures or optimizing performance on standard RGB videos, the impact of other color spaces left unexplored. This study aims to systematically investigate how the different color models affect the overall performance of tracking systems for humans. A dataset was developed in

RGB color space following the MOT15 dataset structure and 6 different versions of it generated by transforming the color spaces into Grayscale, YCbCr, HSI, HSV and YES. The YOLO12s model was fine-tuned on different color spaces for each dataset. Then SORT and DeepSORT trackers are used to associate the detected humans over the frames with unique ids. The output results include the frames, videos and text file that contains bounding boxes and with the assigned ids. The evaluation was done by the motmetrics python library for tracking metrics like accuracy (MOTA), precision (MOTP), FP (False Positive), FN (False Negative), IDSW (ID switches) etc.

This paper contributes the following:

(1) A systematic investigation of six color spaces influences tracking performance using YOLO12s with SORT and DeepSORT trackers.

(2) A custom dataset with proper experimental setup to ensure fair comparison between RGB, Grayscale, YCbCr, HSV, HIS and YES color spaces.

(3) Detailed analysis using MOT metrics to highlight the impact of color in accuracy, precision, and id switches.

The rest of the paper is organized as follows, section 2 summarizes the work done before and finds the importance of this work, dataset development process is shown in section 3, section 4 describes the working methodology in detail, the outcomes and the conclusions are discussed in section 5 and 6 respectively.

2. RELATED WORKS

Color space plays an important role in detection systems. In many cases using other color spaces rather than RGB significantly improves the detection accuracy. Water surface detection in multi-temporal and multi-spectral image analysis method uses image transformation from RGB color space to HSV to get the dynamic information at the continental scale [4]. Hyperspectral image data used to detect tomato spotted wilt virus infection on tobacco and found it effective while detecting in an early stage [5]. It's also used to detect powdery mildews: a threat to the wheat production. HSI of leaf and canopy were taken and then segmented the images based on HSV color space and proceed to further steps [6]. RGB, Lab, HIS, HSV color spaces are seen to be used in feature selection to detect oil palm leaf disease [7] and rare earth molten salt temperature detection [8]. Abdulwahab Ismail Durojaiye et al. [9] emphasizes the importance of HSI in food and agricultural products and suggests the necessary requirements for the ease of HSI replication all over the world. HSV color space is also used in brain MR image segmentation [10] and satellite image [11]. Grayscale images can reduce computational complexity as well as tackle color distortion and blurring in detecting objects in deep water [12]. Peng wang et al. [13] proposed a deep learning-based method to perform wielding defect detection on 8-bit high grayscale x-ray films of solid rocket shells. It's also used in surface temperature mapping of magnesium alloys [14]. Khamar Basha Shaik et al. [15] performs skin color detection and segmentation in HSV and YCbCr color space. Their study shows the efficiency of these color spaces compared to RGB color space. YCbCr color space is also seen in improving identification of fire as well as lowering the computational complexity [16]. It's also used in shadow removal form images that can improve segmentation, tracking, or recognition algorithms [17].

In TBD framework YOLO with SORT and DeepSORT is a widely used method because of its simplicity and high performance. But existing tracking studies are mostly performed on publicly available RGB datasets. YOLOv3, YOLOv3 tiny, YOLOv3 custom datasets are used and filtered for the person class only to detect human followed by DeepSORT tracker that shows dataset quality can improve tracking [18]. Another combination of YOLOv3 and DeepSORT was used in tracking from a top view perspective [19]. YOLOv3 and YOLOv3-TINY were trained on OpenImagesv5 dataset for real time gender detection and tracking [20]. Fang Yang et al. [21] shows comparative results between YOLOv5(s, m, l) and YOLOv7 with DeepSORT using the MOT15 challenge dataset. Modification of trackers can also be seen for pedestrians and vehicles tracking by training the detection models on MSCOCO and UA-DETRAC datasets [22]. Pedestrian detection and tracking performed on MOT16 and VERI-Wild datasets using YOLOv5-Lite and modified DeepSORT [23]. Another modified DeepSORT on top of YOLOv8 used MOT15 and 16 datasets for multi pedestrian tracking [24].

Most of the tracking systems rely on publicly available datasets which are in RGB color space. Despite improvement in detection system using several color spaces no tracking system explored

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the impact of different color transformation input in the tracking by detection framework.

3. DATASET DEVELOPMENT

The development of the dataset went through several steps like recording videos, extracting frames from the videos, annotating humans from the frames and finally organizing the frames and labels following the MOT15 [25] dataset structure. Human movement videos were captured in the lobby area, entrance of the lobby and parking area. This provides different lighting conditions and variety in video scenes. A total of 7 video scenes were chosen from resolution 2688x1512 and 5.02 frames per second (fps) except one video that is 4.95 fps.

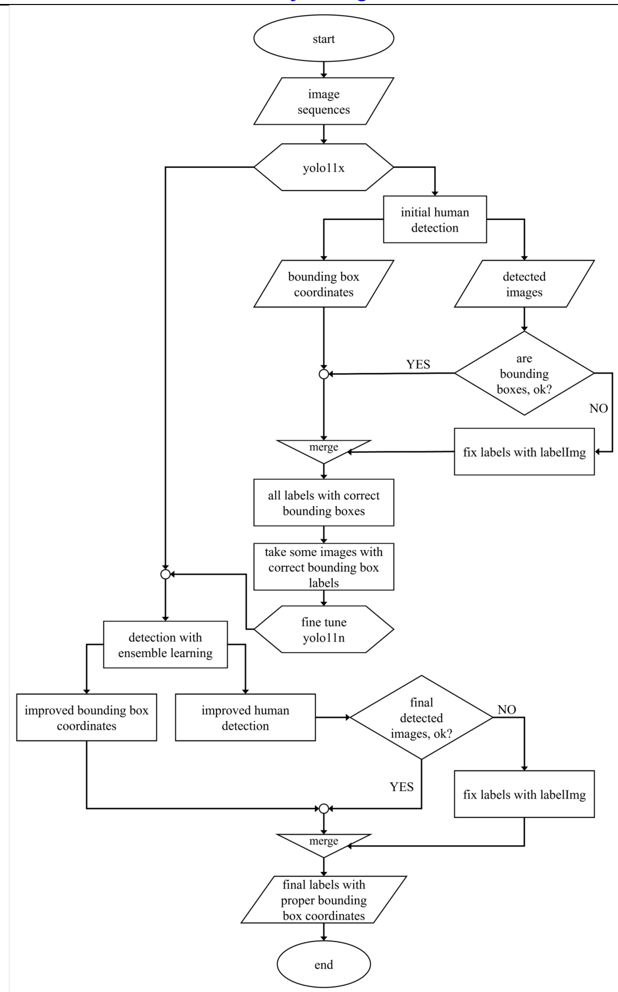
Frames extracted from each video using cv2 library are kept in distinct folders by the sequence name and the frames are named serially like frame_0001, frame_0002 etc. Figure 1 (b) shows it has a lowest of 902 and a highest of 1801 frames form the shortest and the longest videos respectively. To speed up the tiring bounding box annotation process the leverage of pretrained yolo models were taken. Figure 1 (a) shows that YOLO11x, the largest YOLO11 model for detection was chosen to directly predict humans and save the bounding boxes coordinates in labels directory for each of the frames with corresponding frame numbers. Along with these bounding boxes the detected frames were also saved in another folder so that the frames can be checked for any missing or wrong predictions. If any wrong or miss detection occurred, the frame was then annotated manually for the bounding boxes using the popular labelImg tool. Then the labels are merged with the right ones. After annotating some frames, a small YOLO11n model was fine-tuned with some correct labels. Then the fine-tuned YOLO11n and default YOLO11x are jointly used as ensemble models to predict the bounding boxes again. This time the detection becomes more accurate.

After another checking and fixing the wrong ones, all these were merged with the correct labels. Figure 1 (b) shows the total number of bounding boxes annotated in a video. The 5-minute video contains the highest bounding boxes that contain 10033 and one 3-minute video has the lowest 660 bounding boxes which represent the highest and least human appearances in the videos respectively. It sums up to 40544 bounding boxes which annotate humans in the frames. The standard MOT dataset format was followed, specifically the MOT15 dataset with a little bit of changes for the ease of the work. Unlike MOT15 it contains separate trains, validation and test set. In MOT15 each line in detection file contains 10 values that are frame number, id number, bounding box (bb) left, bb top, bb width, bb height, detection confidence and followed by three -1 value which indicates the x, y, z parameters that are not useful in MOT15 because it is a 2D dataset. The id is not assigned to -1 in detection files and in the ground truth file it contains unique id for each object. The customized dataset is also the same for the ground truth files except the detection files because it follows the YOLO detection format which is object class, x and y coordinates of objects center, width and height of the object.

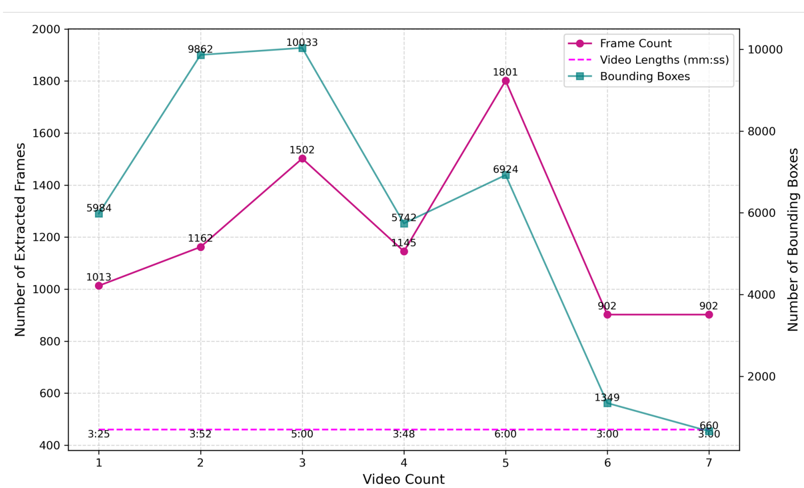
Figure 2 shows that indoor lobby 1, 4, 3 and outdoor parking sequences are chosen for training contain 1013, 1145, 1502 and 902 images that sum up to 4562 images. The validation set contains 1162 and 902 images from indoor lobby2 and outdoor entrance respectively. The longest video that contains 1801 image sequence is chosen for the test set to examine the tracking outcome for longer scenes. It also shows the data split ratio which is roughly 54:24:21 for train, validation and test set. This ratio differs from the ideal dataset split ratio which is 70:20:10. The reasons are to maintain variation of indoor and outdoor scenarios videos in train and validation set. The frame numbers populated each set. Frames from a video kept together otherwise it will break the sequence hence provided the unusual split ratio. Besides, a larger validation set will prevent overfitting issues, and the longest video was chosen for test set for a fair testing. All the videos were captured in RGB the default dataset is the RGB one. The other 5 datasets were prepared by transforming the RGB images to the respective color space which are Grayscale, YCbCr, HIS, HSV and YES.

Journal of Theoretical and Applied Information Technology
31st August 2025. Vol.103. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



(a)



(b)

Figure 1: (a) Annotation process (b) Extracted frames and number of bounding boxes

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

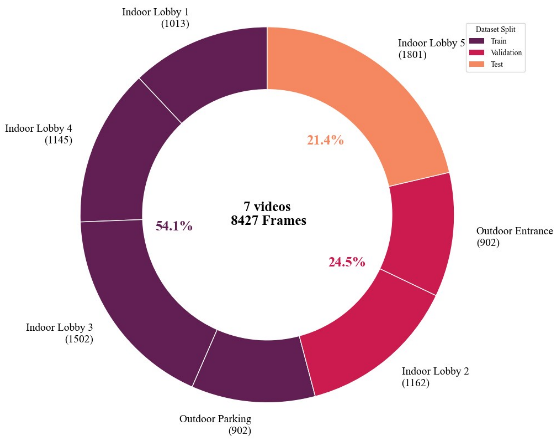


Figure 2: Dataset Analysis

From all the color spaces mentioned above only three of them have the library support from opencv-python which is the popular computer vision library. The supported color spaces are Grayscale, YCbCr and HSV. Using the cv2 module it is easily transformable from RGB to these color spaces. For the remaining two-color spaces, which are HSI and YES the values were manually calculated to transform from RGB to these color spaces.

4. METHODOLOGY

4.1 YOLO12 Detection Framework

YOLO12 is the latest object detector available till date. It surpasses all the other versions released previously. During the iterations of previous versions, it can be observed that most of the improvements are focused on CNN architecture to gain faster processing. Unlike the previous one this version YOLO12 incorporates attention centric mechanism inspired by the proven track of modeling capability of attention mechanism and successfully maintained the speed as well. It improved the block level residual design and feature aggregation model by introducing residual efficient layer aggregation networks (R-ELAN). All these new techniques helped to improve the detection model in terms of accuracy and speed.

Like previous YOLO models YOLO12 comes with five different sizes namely nano (N), small (S), medium (M), large (L) and extra-large (X). The size of the model is generally chosen depending on the dataset. The larger models tend to provide better detection, being computationally expensive and slower than the smaller ones. Here for all the color space datasets total 4562 and 2064 image frames are available for training. The small model is a good choice for this amount of data which will provide a good balance between speed and accuracy.

4.2 Tracking Algorithms

SORT: SORT is a motion-based tracking system. Coming up from the tracking by detection genre, it leverages the CNN detection models in multi object tracking by incorporating the Faser region CNN (FrCNN) which is a two-stage end to end framework. Feature extraction and proposal of regions are passed to the second stage where the classification of the objects is done from the proposed regions. Besides relying on detection models, it also utilized tracking components by combining the Kalman filter and Hungarian algorithm where these components solve the velocity models and id assignments respectively. The threshold value from intersection over union (IoU) eliminates the redundant assignments. It assigns id to a detected object after a minimum number of consecutive frame appearance.

DeepSORT: DeepSORT is a remarkable improvement to the previous SORT by integrating appearance information. To maintain the simplicity of the SORT tracker it adds on a re-identification model which is trained on large scale person reidentification dataset that helps the tracker to handle most of the computational complexity into this offline pretraining stage. This deep neural network model generates feature vectors which gives better tracking results after combining it with IoU matching. To find out the similarity between two objects it calculates the similarity matric and the maximum distance value which ensures the objects within maximum allowed distance are linked across the frames. Appearance feature association matric really helped the tracker to perform better on occlusions and reduced id switches while keeping the simplicity of the SORT algorithm.

4.3 YOLO12s Tuning

The train set is not very high or too low. YOLO12s is a good choice because of its good balance of speed and accuracy. Large sized models can lead to false detections which will badly impact the tracking process. As multiclass detectors focus

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

only on humans, the detection class was set only to persons and ignoring all other objects. To have an efficient and good tuning of this YOLO12s model for human detection we incorporated a bunch of hyperparameters which will help the model learn and avoid overfitting. The model was trained using adamW optimizer, batch size of 48 and set for 140 epochs with a learning rate of 0.002 which will give it a smooth training. L2 regularization was used to tackle the overfitting issue. Besides, the training utilizes the augmentation parameters by mixing multiple images using mixup and mosaic parameters, flipping, rotating and scaling images to bring variation in data so that the model learns better. The post processing thresholds which are confidence threshold and intersection over union (IoU) and are set to 0.22 and 0.5 to filter detections for validation after each epoch which will help the model tune properly. Early stopping was enabled to reduce unnecessary computational cost.

4.4 Tracking Pipeline

Setup: The setup for the tracking process is done in a way so that the same setup can be reused for each dataset. It enhanced the architecture for the experiments in a generalized manner to minimize the redundancy as well as increasing the flexibility of the reusable coding. A python script named run_tracker.py was written which imported the functions from the tracking algorithms and is modified with proper parameters and set under a run tracker function which can be called to start the tracking process. A customized jupyter notebook script named as main.ipynb was written to handle all the inputs and outputs properly. This script is also used for initiating the tracking process. Both trackers used this generalized initiator to run experiments on different datasets. This script worked as a bridge between datasets, tracking algorithms and the output directories. All the fine-tuned models from all those color spaces are kept in a directory so that it can be used easily. Separate directories are created for detecting image frames, output video and tracking results. Parameters like IoU threshold and confidence threshold can be easily tunable from this script. With this setup all the experiments were done separately for each fine-tuned model and saved the outputs to observe the effect of the detection models of various color spaces on both SORT and DeepSORT trackers.

Human detection with fine-tuned model: After activating the tracking function, it first initiates the detection models which are stored in one place with proper naming after the color model it was trained. The detection models take each frame from the test sequence and start detecting humans from the frames. The detections are generated in 'xyxy' format where it indicated two corners of the bounding boxes top-left (x1, y1) and bottom-right (x2, y2). For a bounding box coordinates like [200, 300, 250, 400] the top-left corner is (200, 250) and the bottom-right corner is (300, 400). The weak and wrong detections were eliminated by filtering them with the confidence score and IoU threshold with value of 0.5 and 0.2 respectively. After generating the detections those were converted to 'xywh' format to make the detections compatible for the SORT and DeepSORT trackers where (x, y) identifies the top-left corner, w and h are the width and height of the bounding box. These values are calculated from the difference of the x and y coordinates meaning w = x2 – x1 and h = y2 – y1. For the example given above the coordinates after converting to 'xywh' format will be [200, 300, 50, 100]. Finally, all these detections are fed into the trackers.

YOLO12s SORT: SORT takes the detection results and predicts locations of the objects in the next frame. It uses Kalman filtering to perform that prediction. It then performs a comparison between predicted bounding box and the current YOLO detection using IoU. Based on this IoU cost the Hungarian algorithms match the detections to the trackers meaning the assignment of the object id. If the detection matches a tracker, it updates it on the other hand if it does not create new tracker and deletes the old tracker if not matched for N frames. In the experiment the IoU threshold is 0.2 which eliminates the bounding boxes with IoU score less than that, max_age is 15 which deletes the unmatched tracks after 15 frames and min_hits is 3 to establish a new track after 3 consecutive detections.

YOLO12s DeepSORT: DeepSORT enhances the power of SORT by incorporating appearance features to the bounding boxes. The detections pass through a CNN based feature extractor to generate feature vectors. In this case MobilenetV2 is chosen as a pretrained model on ImageNet dataset. The standard input size for this feature extractor model is 224 x 224. Before generating feature vectors, the bounding box regions form the detected frames were cropped. These cropped images were resized to 224 x 244 dimensions which will help to run the feature extractor to perform faster. Cosine similarity is responsible to calculate the distances between vectors which will identify the similar

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

objects and combining it with Kalman filtering and IoU it improves the id assignment of the Hungarian algorithm making the tracker more reliable. The max_distance is set to 0.5 which will discard the detections that exceed this cut-off value while calculating cosine similarity.

The results were saved in MOT format for both trackers on top of each detection model that are trained on their respective color space datasets.

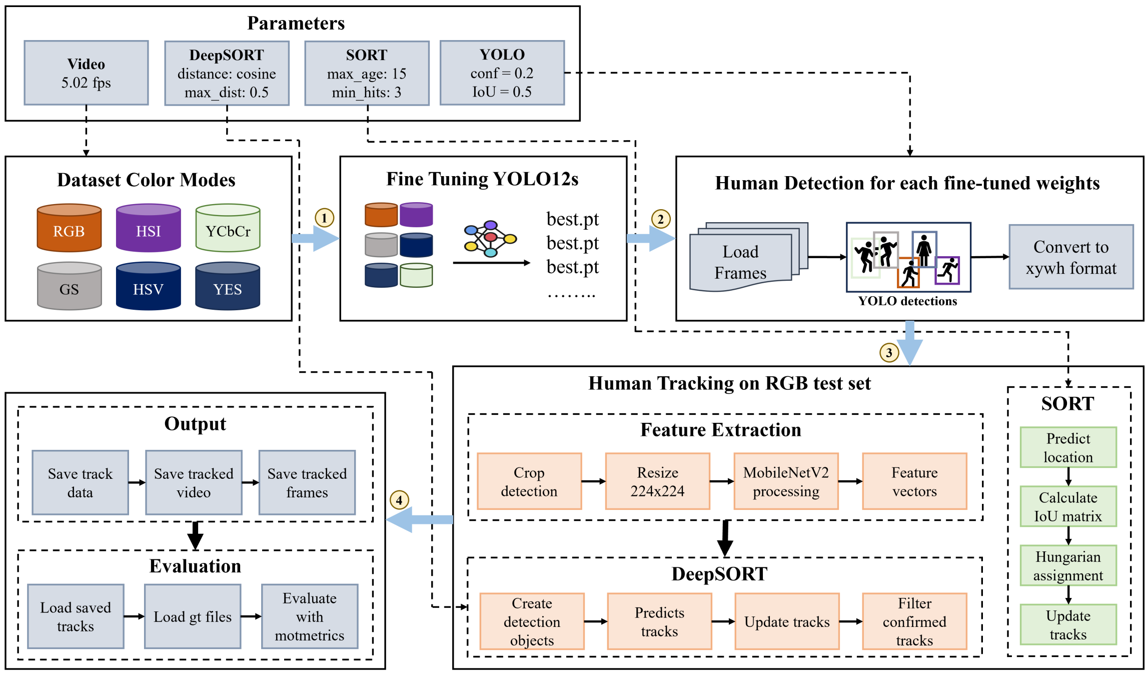


Figure 3: Human Tracking by Detection Pipeline

5. RESULTS AND DISCUSSIONS

This evaluation is done using the motmetrics python library. Multi object tracking accuracy (MOTA): This is the primary metric to measure the performance of trackers. It depends on False Positive (FP): tracks that don't exist in ground truth, False Negative (FN): missed tracks that are available in ground truth and ID switches: the number of times id is incorrectly changed. Multi Object tracking precision (MOTP): It determines how precise bounding boxes are. It is the average IoU between the predicted bounding boxes and actual ground truth boxes for the correct detections. ID F1 score (IDF1): Determines how well the trackers maintain identity of objects across frames. Higher IDF1 means less ID switches. Mostly Tracked (MT): It is the ground truth that are tracked successfully by more than 80% of their lifespan. Mostly Lost (ML): The objects that were tracked by less than 20% of their lifespan. Partially Tracked (PT): The objects which tracking percentage is between 20% to 80% are considered as partially tracked objects. These metrics were considered for the evaluation process for this study.

Each fine-tuned YOLO12s detection model was used for the detection phase of the tracking pipeline for both SORT and DeepSORT tracker. Table 1 shows all the tracking results collected from the experiments on various color spaces. YOLO12s fine-tuned on RGB obtain highest MOTA of 50.0% and 31.5% for both DeepSORT and SORT respectively. Almost near MOTA is observed from grayscale trained model for DeepSORT (45.7%) and SORT (29.1%). Grayscale always provided fewer FP for DeepSORT (376) and SORT (73) indicating better prevention of ghost tracking. Even grayscale trainset for DeepSORT provided the highest IDF1 (43.8%) which led to fewer IDSW (7). But it contains more FN (3370) meaning missed a lot of detections. In case of SORT the performance of other color spaces based on MOTA YCbCr (16.6%), HSV (13.8%) and HIS (11.3%). On the other hand, the DeepSORT tracker obtained 22.0%, 21.3% and 17.6% for HSV, YCbCr and HSI color space respectively. The remaining YES color space provided the worst outcome with negative MOTA score indicating it totally failed to track the humans on test videos for both trackers.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

To test the intuition that the results might get improved, some tests conducted using SDG optimizer in detection model tuning or running the tracker on same color space test set. For example, here for DeepSORT the detection model trained using AdamW optimizer provided only 43.0% MOTA. Also, a test tracking performed on Grayscsale test set which gained 45.4% MOTA indicating no improvements.

Table 1: Tracking Results

Color Modes	MOTA ↑	MOTP ↑	IDF1 ↓	IDSW ↓	FP ↓	FN ↓	MT ↑	PT	ML ↓
YOLO12s DeepSORT									
RGB	50.0	23.7	43.3	15	483	2955	47	46	184
Grayscale	45.7	21.2	43.8	7	376	3370	42	38	197
HSV	22.0	29.3	38.7	5	684	3540	18	21	179
HSI	17.6	23.5	28.0	25	1347	4292	25	51	280
YCbCr	21.3	26.4	34.8	1	1116	4098	21	20	216
YES	Negative	-	-	-	-	-	-	-	-
RGB-SDG optimizer[1]	43.0	24.6	42.5	27	591	3320	41	55	181
GS-GS test set[2]	45.4	21.7	41.9	6	385	3378	45	38	194
YOLO12s SORT									
RGB	31.5	23.6	34.6	8	119	2452	18	63	159
Grayscale	29.1	20.7	33.1	11	73	2480	13	53	154
HSV	13.8	29.2	27.6	2	217	2089	7	22	133
HSI	11.3	22.6	20.6	8	367	3023	10	50	161
YCbCr	16.6	22.6	28.5	1	280	2305	6	20	173
YES	Negative	-	-	-	-	-	-	-	-

using SDG optimizer[1] tracking on Grayscale test set[2]

6. CONCLUSIONS

This study investigates the influence of different color spaces for detection models in tracking systems. The detection model that uses the RGB color space always provides better tracking results compared to the other color spaces. Though it can be seen that other color spaces noticeably improve detection performance in the tracking system the answer can be found in the mechanism of the tracking algorithms. The DeepSORT tracker uses a feature extraction model which incorporates appearance features with bounding box positions. This appearance similarity calculation becomes more accurate while using RGB images because it offers better distinguishable features for different objects. On the other hand, the remaining color spaces don't offer this variety of colors which makes tracking difficult. The SORT only relies on bounding box positions so the accuracy difference with the other color spaces is not very high. Though it does not care about appearance features, the YOLO detection model misses some humans, that's why it cannot surpass the one with the RGB. In tracking algorithms color input is an essential factor as it identifies an object and tries to continuously keep the track for upcoming frames.

Full manual annotation of the frames will give better bounding boxes which may lead to better training. The experiments were conducted using only six color spaces while there are other color spaces available such as CMYK (Cyan, Magenta, Yellow, Black) and LAB (Lightness, A, B). Besides, fine tuning for the detection models were done using limited hyperparameters. The possibility remains to obtain better results with more fine-tuned models or better values for tracker parameters or using the other color spaces left in this study.

Finally, from the detailed evaluation of the results and the dependency on features of the trackers it is clear that the impact of the colors in tracking shows significant fluctuations in tracking performance. Tracking algorithms that use appearance features tend to keep the margin higher on RGB color space compared to the others. Moreover, the color choice gives a nice trade of between most detection and fewer id switches depending on the need of work. If the goal is to track most of the humans RGB is best. On the other hand, Grayscale could be a good choice, especially for DeepSORT if maintaining identity over time is more important even if some detections are being ignored.

ACKNOWLEDGEMENT

The authors would like to thank the Ministry of Higher Education for providing financial support under Fundamental Research Grant Scheme

(FRGS) No. FRGS/1/2023/ICT02/UMP/02/3 (University reference RDU230117).

REFERENCES:

[1] T. J. Cheng, A. F. Ab. Nasir, A. P.P. A. Majeed, M. A. M. Razman, and T. L. Lim, "Vision-based Human Detection by Fine-Tuned SSD Models," International Journal of Advanced Computer Science and Applications, vol. 13, no. 11, pp. 386–390, Jan. 2022, doi: 10.14569/IJACSA.2022.0131143.

[2] T. J. Cheng, A. F. Ab. Nasir, A. P.P. A. Majeed, T. L. Lim, and I. M. Khairuddin, "CenterNet: A Transfer Learning Approach for Human Presence Detection," In: Tan, A., et al. Advances in Intelligent Manufacturing and Robotics. ICIMR 2023. Lecture Notes in Networks and Systems, vol. 845, pp. 41–51, Feb. 2024, doi: 10.1007/978-981-99-8498-5_4.

[3] T. J. Cheng, A. F. Ab. Nasir, A. P.P. A. Majeed, T. L. Lim, M. A. M. Razman, and I. M. Khairuddin, "Fine-tuned RetinaNet models for Vision-based Human Presence Detection," MEKATRONIKA, vol. 4, no. 2, pp. 16–23, July 2022, doi: 10.15282/mekatronika.v4i2.8850.

[4] J.-F. Pekel et al., "A near real-time water surface detection method based on HSV transformation of MODIS multi-spectral time series data," Remote Sens. Environ., vol. 140, pp. 704–716, Jan. 2014, doi: 10.1016/j.rse.2013.10.008.

[5] Q. Gu et al., "Early detection of tomato spotted wilt virus infection in tobacco using the hyperspectral imaging technique and machine learning algorithms," Comput. Electron. Agric., vol. 167, p. 105066, Dec. 2019, doi: 10.1016/j.compag.2019.105066.

[6] L. An et al., "Classification of wheat powdery mildew based on hyperspectral: From leaves to canopy," Crop Prot., vol. 177, p. 106559, Mar. 2024, doi: 10.1016/j.cropro.2023.106559.

[7] H. Hamdani, A. Septiarini, A. Sunyoto, S. Suyanto, and F. Utaminingrum, "Detection of oil palm leaf disease based on color histogram and supervised classifier," Optik, vol. 245, p. 167753, Nov. 2021, doi: 10.1016/j.ijleo.2021.167753.

[8] Z. Zhang, M. Xu, F. Liu, X. Wu, Y. Zhang, and Y. Shen, "Multi-color space features and data regression neural network-based method for molten salt temperature detection in rare earth electrolytic processes," Int. Commun. Heat Mass Transf., vol. 159, p. 108069, Dec. 2024, doi: 10.1016/j.icheatmasstransfer.2024.108069.

[9] A. I. Durojaiye, S. T. Olorunsogo, B. A. Adejumo, A. Babawuya, and I. I. Muhamad, "Deep learning techniques for the exploration of hyperspectral imagery potentials in food and agricultural products," Food Humanity, vol. 3, p. 100365, Dec. 2024, doi: 10.1016/j.foohum.2024.100365.

[10] P. Singh, "A neutrosophic-entropy based clustering algorithm (NEBCA) with HSV color system: A special application in segmentation of Parkinson's disease (PD) MR images," Comput. Methods Programs Biomed., vol. 189, p. 105317, Jun. 2020, doi: 10.1016/j.cmpb.2020.105317.

[11] P. Ganesan and V. Rajini, "Value based semi automatic segmentation of satellite images using HSV color space, histogram equalization and modified FCM clustering algorithm," in 2013 International Conference on Green Computing, Communication and Conservation of Energy (ICGCE), Dec. 2013, pp. 77–82. doi: 10.1109/ICGCE.2013.6823403.

[12] M. Zhou, B. Li, J. Wang, and K. Fu, "A lightweight object detection framework for underwater imagery with joint image restoration and color transformation," J. King Saud Univ. - Comput. Inf. Sci., vol. 35, no. 9, p. 101749, Oct. 2023, doi: 10.1016/j.jksuci.2023.101749.

[13] P. Wang, L. Li, X. Li, L. Duan, Z. Lü, and R. Di, "An automatic welding defect detection method based on deep learning for super 8-bit high grayscale X-ray films of solid rocket motor shells," NDT E Int., vol. 151, p. 103306, Apr. 2025, doi: 10.1016/j.ndteint.2024.103306.

[14] X. Xu, W. Jia, S. Xu, F. Ning, and J. Lei, "Relationship between high-level color features and temperature mapping of magnesium alloy surface images based on the K-nearest neighbor algorithm," Appl. Therm. Eng., vol. 259, p. 124940, Jan. 2025, doi: 10.1016/j.applthermaleng.2024.124940.

[15] K. B. Shaik, P. Ganesan, V. Kalist, B. S. Sathish, and J. M. M. Jenitha, "Comparative Study of Skin Color Detection and Segmentation in HSV and YCbCr Color Space," Procedia Comput. Sci., vol. 57, pp. 41–48, 2015, doi: 10.1016/j.procs.2015.07.362.

[16] X. Chen, Q. An, and K. Yu, "Fire identification based on improved multi feature fusion of YCbCr and regional growth," Expert Syst.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Appl., vol. 241, p. 122661, May 2024, doi: 10.1016/j.eswa.2023.122661.

[17] X. Zhu, R. Chen, H. Xia, and P. Zhang, “Shadow removal based on YCbCr color space,” Neurocomputing, vol. 151, pp. 252–258, Mar. 2015, doi: 10.1016/j.neucom.2014.09.045.

[18] M. I. H. Azhar, F. H. K. Zaman, N. Md. Tahir, and H. Hashim, “People Tracking System Using DeepSORT,” in 2020 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), Aug. 2020, pp. 137–141. doi: 10.1109/ICCSCE50387.2020.9204956.

[19] I. Ahmed, M. Ahmad, A. Ahmad, and G. Jeon, “Top view multiple people tracking by detection using deep SORT and YOLOv3 with transfer learning: within 5G infrastructure,” Int. J. Mach. Learn. Cybern., vol. 12, no. 11, pp. 3053–3067, Nov. 2021, doi: 10.1007/s13042-020-01220-5.

[20] Z. M. Peerun and R. K. Moloo, “Real-time gender and people tracking using YOLO,” in 2024 Sixth International Conference on Computational Intelligence and Communication Technologies (CCICT), Apr. 2024, pp. 448–454. doi: 10.1109/CCICT62777.2024.00079.

[21] F. Yang, X. Zhang, and B. Liu, “Video object tracking based on YOLOv7 and DeepSORT,” Jul. 25, 2022, arXiv: arXiv:2207.12202. doi: 10.48550/arXiv.2207.12202.

[22] D. Meimetis, I. Daramouskas, I. Perikos, and I. Hatzilygeroudis, “Real-time multiple object tracking using deep learning methods,” Neural Comput. Appl., vol. 35, no. 1, pp. 89–118, Jan. 2023, doi: 10.1007/s00521-021-06391-y.

[23] W. Luo, P. Sun, F. Zhong, W. Liu, T. Zhang, and Y. Wang, “End-to-end Active Object Tracking via Reinforcement Learning,” in Proceedings of the 35th International Conference on Machine Learning, PMLR, Jul. 2018, pp. 3286–3295. Accessed: Oct. 28, 2024. [Online]. Available: https://proceedings.mlr.press/v80/luo18a.html

[24] W. Sheng et al., “Multi-objective pedestrian tracking method based on YOLOv8 and improved DeepSORT,” Math. Biosci. Eng., vol. 21, no. 2, Art. no. mbe-21-02-077, 2024, doi: 10.3934/mbe.2024077.

[25] L. Leal-Taixé, A. Milan, I. Reid, S. Roth, and K. Schindler, “MOTChallenge 2015: Towards a Benchmark for Multi-Target Tracking,” Apr. 08, 2015, arXiv: arXiv:1504.01942. doi: 10.48550/arXiv.1504.01942.