Object detection in HDR images

Muhammad Saad
School of Science and Engineering
Lahore University of Management Sciences
Lahore, Pakistan
24100012@lums.edu.pk

Rizwan Ahmad Bhatti
School of Science and Engineering
Lahore University of Management Sciences
Lahore, Pakistan
21030005@lums.edu.pk

***Abstract*—This work explores the potential of High Dynamic Range (HDR) images in improving object detection accuracy. Unlike existing studies that primarily focus on Low Dynamic Range (LDR) images, we provide a proof-of-concept for training YOLOv7 object detector directly with HDR images. Since annotated HDR image databases are scarce, we use pseudo HDR images resulting from range expansion of 1000 images from MS COCO dataset to train the model. Both models are trained for 100 epochs and batch size of 16. Trained models are used to detect objects in actual HDR dataset based on the HDR4RTT repository. True HDR test images are first tone mapped before being fed to the LDR trained model, whereas HDR trained model does not need this preprocessing step. We then compare the detection results of both models. Results show that the HDR trained model is able to detect objects in out-of-distribution HDR images. In comparison to LDR trained model, HDR trained model depicts reduced mean-average-precision. We recommend to repeat the experiment on a larger dataset of single object category and at least 300 epochs. This would required a GPU enabled machine with at least 12 GB GPU memory.**

***Index Terms*—image processing, hdr, object detection, yolov7, computer vision, deep learning, range expansion, tone mapping, image processing**

I. Introduction

Object detection stands at the forefront of contemporary research in computer vision, driven by advancements in Deep Learning that have revolutionized the landscape of visual perception tasks. Traditionally, Low Dynamic Range (LDR) images have served as the primary input for training object detection models, capturing scenes with limited contrast between bright and dark regions. However, the inherent limitations of LDR imagery in preserving fine details and nuanced information in high-contrast scenes have motivated researchers to explore alternative solutions.

In this work we delve into the exploration of High Dynamic Range (HDR) images as a novel avenue for enhancing object detection accuracy. HDR imaging technology captures a broader range of luminosity levels, offering a more comprehensive representation of scenes with varying lighting conditions. Unlike LDR images, which may struggle to capture details in extremely bright or dark areas, HDR images have the potential to provide a more nuanced and information-rich input to object detection models.

In pursuit of this objective, our research presents a comprehensive pipeline designed to harness the benefits of HDR imagery for training a YOLOv7 model [6], a state-of-the-art object detection framework known for its efficiency and accuracy.

Our work marks a departure from previous practices, as YOLOv7 has not been trained on HDR images before. Our experiment setup is based on official YOLOv7 repository, true HDR test dataset from HDR4RTT database [5], pre-trained Expand-Net model [4] and 1000 images belonging to 'person' class from MS COCO 2017 [9] repository.

II. Related work

A large amount of research has already been conducted on both object detection and HDR imaging, Range Expansion and what the state of the art is in Computer Vision, hence we won't delve too deep into these topics but will instead provide a brief overview.

A. HDR Imaging

HDR imagery encompasses almost the entirety of the visible light spectrum, spanning 13 orders of magnitude with information being commonly stored in high-precision floating-point formats (.hdr/.exr) [1]. In contrast Low Dynamic Range (LDR) imagery compresses scene information to three orders of magnitude, storing it in lower precision integer (.jpg/.png) formats[1]. The enhanced precision of HDR allows for the retention of more information in both brighter and darker regions of a scene.

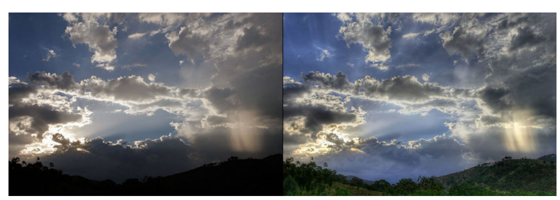

Fig. 1. HDR vs LDR Comparison [12]

1) Tone Mapping Operators: Tone mapping serves as an irreversible and lossy High Dynamic Range (HDR) processing technique, converting HDR content to Low Dynamic Range (LDR) content. Color space transformations, linear and non linear transfer functions are used to perform tone mapping. This paper utilizes the Reinhard TMO since it is easily available and shown to have consistently good results [11].

2) Range Expansion: Range Expansion involves predicting High Dynamic Range (HDR) images from single-exposure Low Dynamic Range (LDR) images. Traditionally, heuristic basis expansion operators have been central to the process, however with the improvement and development of convolutional neural networks, NN based expansion operators are becoming more commonly used and are able to perform as well as or even better than the heuristic based expansion operators. For this paper, we have decided to go with ExpandNet which is a Deep Convolutional Neural Network for High Dynamic Range Expansion from Low Dynamic Range Content. ExpandNet accepts LDR images as input and generates images with an expanded range, outperforming most expansion operators [4].

B. Object detection

Object detection is a computer vision task that involves identifying and locating objects of interest within an image or a video frame. The goal is to not only recognize the presence of objects but also to precisely determine their positions in the visual field. This task is fundamental for machines to comprehend and interpret the visual world, enabling applications ranging from autonomous vehicles and robotics to image and video analysis.

1) Applications of Object Detection:

- Autonomous Vehicles: Object detection is crucial for autonomous vehicles to identify and track pedestrians, vehicles, and obstacles in their surroundings. This enables the vehicle to make informed decisions for navigation and collision avoidance.
- Surveillance and Security: Object detection is widely used in surveillance systems to monitor and identify suspicious activities or objects in public spaces, airports, and critical infrastructure. It helps enhance security by automatically detecting and alerting authorities to potential threats.
- Environmental Monitoring: In fields like agriculture and ecology, object detection is used to monitor and track animals, assess plant health, and gather data for environmental studies.

In essence, object detection is a foundational technology with diverse applications that span numerous industries, contributing to the development of more intelligent and capable systems. Object detectors trained on High Dynamic Range images may be more robust to detection under different lighting conditions and are thus the next logical step forward.

2) YOLO detector: YOLO, which stands for ”You Only Look Once,” is a popular object detection algorithm that has gained prominence in the field of computer vision due to its speed and accuracy. The key characteristic of YOLO is its ability to perform object detection in real-time with a single pass through the neural network, making it significantly faster than many other object detection algorithms. In this work we are utilizing the YOLOv7 model of the YOLO detector family. The YOLOv7 model was released to the public in 2022 and is essentially the state of the art in object detection.

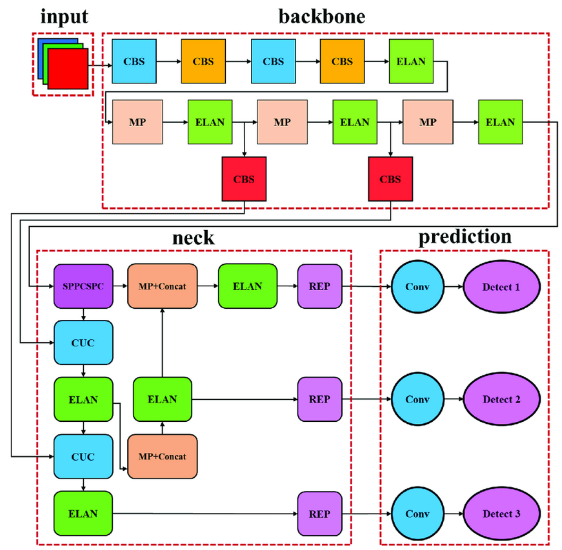

Fig. 2. YOLOv7 Architecture [13]

The YOLOv7 object detection model builds upon previous iterations through changes in the architecture of the network as well as model scaling. The architecture is built in order to achieve maximal efficiency while maintaining the performance.The overall structure is show in Figure 2. Extended Efficient Layer Extended Networks (E-ELAN) constitute the central addition to YOLOv7. E-ELAN uses merge, shuffle and expand cardinality to improve the models learning rate without slowing convergence. Computational blocks are divided into groups and each group can learn a different and diverse feature set, further improving learning [6]. The EELAN architecture is shown in Figure 3.

The model scaling is done by keeping in mind previous scaling techniques that utilize depth scaling which also results in an increase in width of the output computational block, the input width for the following transistion layer then also has to be increased. The YOLOv7 model instead proposes that only the depth of the computation block needs to be scaled and the transistion layer is subsequently width scaled. This allows the model to have an optimal structure [6].

Newer versions like the YOLOv8 detector have been released but as of yet they are not stable enough to be used. Additionally, YOLOv7 is miles ahead of the existing competition as show by Figure 4 all the while being stable and not experimented on too deeply. The YOLOv7 object detector has been trained on LDR images and our work tries to bring a shift in the paradigm of object detector training by utilizing High Dynamic Range images.

III. Methodology

In this work we propose an end-to-end pipeline for training and testing of YOLOv7 model on HDR dataset. True HDR images are used for test where the LDR trained model requires a tone-mapping preprocessing step. Detection results from LDR trained model is compared with those from HDR trained

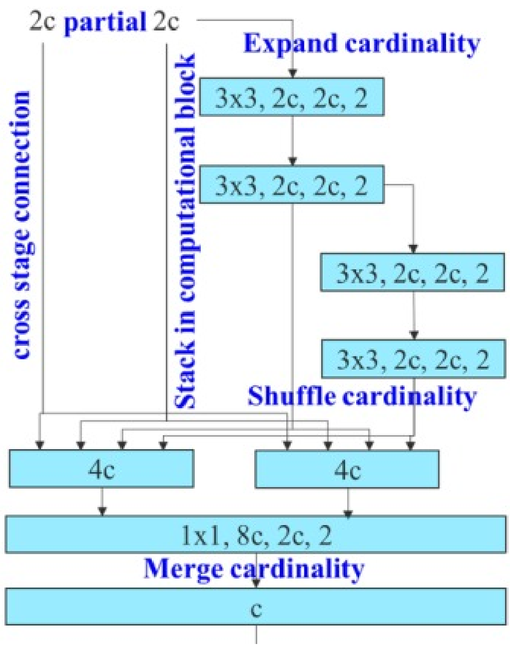

Fig. 3. EELAN Architecture [6]

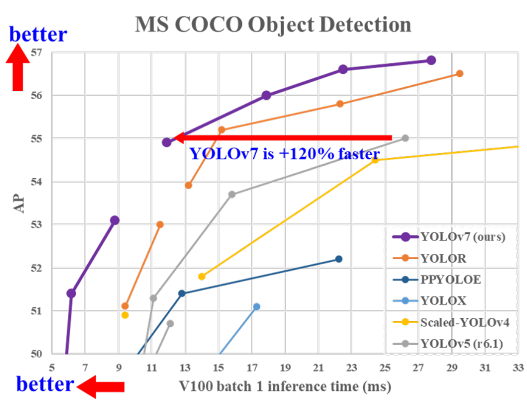

Fig. 4. Yolov7 vs other object detectors [6]

model. In the following sections we discuss the methodology in detail.

A. LDR Dataset

Various databases for annotated datasets are available for application of object detection [3]. These databases consist of LDR images. In this work we have chosen MS COCO [9] database of annotated images which consists of over 100,000 training images and 80 object categories. Due to limited resources, we acquire 1000 images of 'person' category only. This forms the yolo-coco-person dataset which is used to train the LDR YOLOv7 model depicted in Fig. 5.

B. Pseudo HDR Dataset

To train YOLOv7 on HDR images, we need an annotated database of HDR images. Most of the work available is for

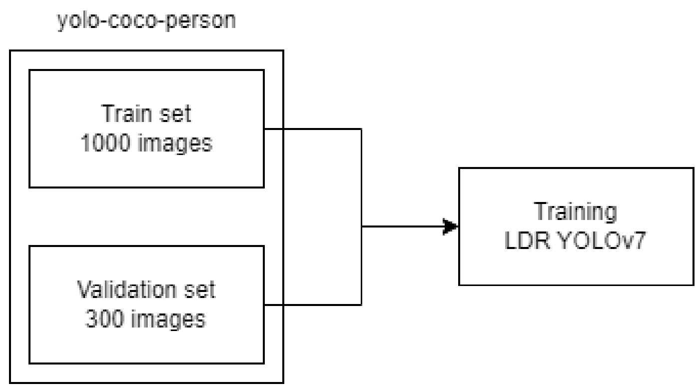

Fig. 5. LDR dataset involves 1000 training images and 300 validation images

LDR images therefore we propose Range Expansion [4]. LDR images can be expanded into HDR images. These would be pseudo HDR images since they have been synthetically generated. For range expansion an inverse tone mapping operator is needed. We have chosen Expand-Net [4] for converting LDR images into HDR images because of its better performance in comparison to other conventional range expansion methods. 1000 MS COCO images are range expanded to produce pseudo HDR images which will then be used to train the HDR YOLOv7 model. This forms the hdr-yolo-coco-person dataset which is used to train the HDR YOLOv7 model as depicted in Fig. 7.

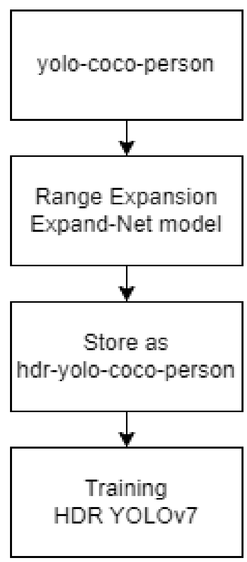

Fig. 6. HDR dataset is based on range expansion of yolo-coco-person images, stored as hdr-yolo-coco-person which is used for training of the HDR model.

C. HDR Dataset

True HDR images is required for testing of the YOLOv7 model trained on pseudo HDR dataset and also the LDR YOLOv7 model using the extra tone mapping preprocessing step. This allows evaluation on actually HDR dataset instead of range expanded images. HDR4RTT is a database of over 5000

HDR images with bounding box and class annotations [5] and we have chosen this database in this work to create yolo-hdr4rtt and tmo-yolo-hdr44tt datasets as depicted in Fig. **??**

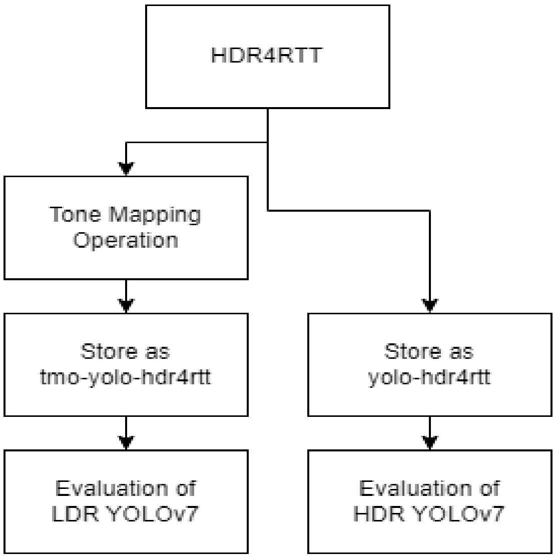

Fig. 7. LDR YOLOv7 model is evaluated against the tone mapped HDR4RTT images whereas HDR model is evaluated directly on true HDR images. Both sets are stored as tmo-yolo-hdr4rtt and yolo-hdr4rtt, respectively.

D. Object Detection with YOLOv7

YOLOv7 is a state-of-the-art detector surpassing other detectors in terms of accuracy and inference time [6]. Python implementation is opensource and a simple interface is provided for training and evaluation of customized YOLOv7 models [7]. The interface is based on YAML [7] to define the model, its hyper parameters and dataset. It has over 37 million parameters and requires CUDA GPU [8] support for training.

1) LDR pipeline: Fig. 8 shows the LDR training and evaluation pipeline. Results of the LDR images trained model would be useful to compare with those of the pseudo HDR images trained model.

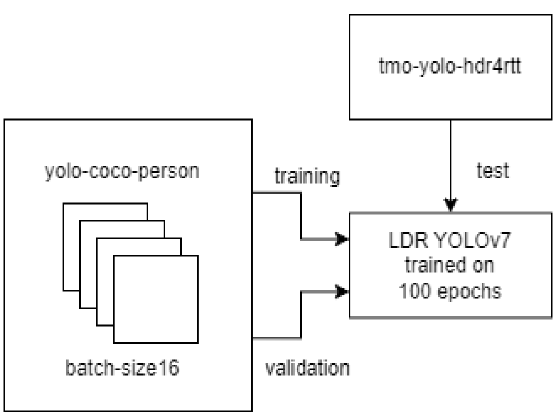

Fig. 8. LDR YOLOv7 is trained for 100 epochs on yolo-coco-person dataset. Model is continuously validated with validation images present in yolo-coco-person. Once all training is done, tmo-yolo-hdr4rtt dataset is used to test the model on out-of-distribution data.

2) HDR pipeline: Fig. 9 shows the HDR training and evaluation pipeline. Since the same LDR database is used, an extra Range Expansion stage is involved using the Expand-Net model. The output from this stage are the pseudo HDR images to be used for training of HDR-YOLOv7 model. These are stored while preparing the hdr-yolo-coco-person dataset.

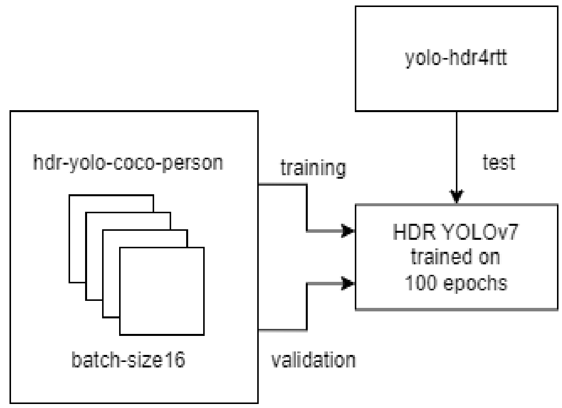

Fig. 9. HDR YOLOv7 is trained for 100 epochs on hdr-yolo-coco-person dataset. Model is continuously validated with validation images present in hdr-yolo-coco-person. Once all training is done, yolo-hdr4rtt dataset is used to test the model on out-of-distribution data.

E. Evaluation

Validation scores summary is available during training. Sample outputs after training also presented here.

RESULTS

TABLE I
MAP SCORES FOR LDR AND HDR MODELSR

Model	**mAP @ 0.5 IoU**
LDR YOLOv7	0.0175
HDR YOLOv7	0.0128

Table. I shows significantly lower mAP score because all 80 classes where being predicted but dataset was for a single class only. Further experiments are recommended to train for only a single class. Still the HDR model results are not far behind the LDR model. Fig. 12 and Fig. 13 show the variation in mAP score as training progressed for LDR and HDR models respectively.

Fig. 13 and Fig. 12 shows the detection results from HDR and LDR models respectively. It seems that tone mapped images when passed to LDR model show better detection as compared to when HDR image is passed directly to the HDR model. This observation can be attributed to low number of epochs. Moreover, we propose HDR model epochs to be higher than LDR model epochs to allow more convergence since HDR images have a higher amount of information as compared to LDR images.

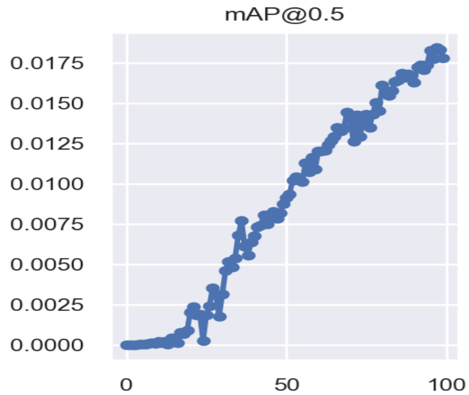

Fig. 10. Variation of mAP score as LDR training continued for 100 epochs. The increasing trend shows that more epochs would significantly improve the mAP scores.

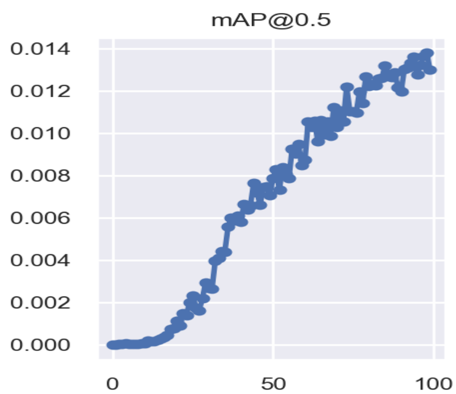

Fig. 11. Variation of mAP score as HDR training continued for 100 epochs. The increasing trend shows that more epochs would significantly improve the mAP scores.

Fig. 12. Object detection results on one image from tmo-yolo-hdr4rtt, showing three person detections.

Fig. 13. Object detection results on one image from yolo-hdr4rtt showing one person detection.

Conclusion

1000 images from MS COCO database belonging to person class were used to train the LDR YOLOv7 model. The same set of images were range expanded to train the HDR YOLOv7 model. 100 epochs were allowed for both models. We recommend to increase epochs to 300 and to update annotations to include only a single class label. Training should also be configured for a single class only. This will allow better mAP score comparison and faster convergence. The experiment repository is avialable on github [10] for further development and experimentation.

We also recommend to have a true HDR image database created for object detection purpose. This will require a custom software tool for creating annotations. True HDR images will allow better training of the HDR object detection models. An LDR image version of the HDR image can also be captured at the same time. These images can be used to train the LDR YOLOv7 model for results comparison.

References

[1] A. Artusi, F. Banterle, K. Debattista, and A. Chalmers, Advanced High Dynamic Range Imaging: Theory and Practice. Boca Raton, FL, USA: CRC Press, 2011.

[2] Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes (VOC) Challenge. IJCV (2010)

[3] Aria Salari, Abtin Djavadifar, Xiangrui Liu, Homayoun Najjaran, Object recognition datasets and challenges: A review

[4] ExpandNet: A Deep Convolutional Neural Network for High Dynamic Range Expansion from Low Dynamic Range Content

[5] https://drive.inesctec.pt/s/dPwxsHWpxj8xynk

[6] Wang, C.-Y., Bochkovskiy, A. and Liao, H.-Y.M. (2022) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. ArXiv: 2207.02696

[7] https://github.com/WongKinYiu/yolov7

[8] R. S. Dehal, C. Munjal, A. A. Ansari and A. S. Kushwaha, "GPU Computing Revolution: CUDA," 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, 2018, pp. 197-201, doi: 10.1109/ICACCCN.2018.8748495

[9] Lin et al. in Microsoft COCO: Common Objects in Context

[10] https://github.com/rizahmad/HDR_YOLOv7

[11] Cerdá-Company, X., Párraga, C. A., Otazu, X. (2018). Which tone-mapping operator is the best? A comparative study of perceptual quality. Journal of the Optical Society of America, 35(4), 626. https://doi.org/10.1364/josaa.35.000626

[12] Mullerova, Vera. (2015). HDR Image Artifact Copensation.
[13] Chang, B. R., Tsai, H.Hsieh, C. (2023). Accelerating the response of Self-Driving control by using rapid object detection and steering angle prediction. Electronics, 12(10), 2161. https://doi.org/10.3390/electronics12102161