2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) | 979-8-3503-3086-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAC3N60023.2023.10541599

The development of a new edge computing device based on deep learning YOLOv8 algorithm for diabetic retinopathy Detection

Devendra Singh
Department of Computer Science and Engineering
Graphic Era Deemed to be University, Dehradun, India
devendrasuno@gmail.com

Lalit Mohan
Department of Computer Science and Engineering
, Graphic Era Hill University, Bhimtal, India
lalitmohan@gehu.ac.in

Janmejay Pant
Department of Computer Science and Engineering
, Graphic Era Hill University, Bhimtal, India
geujay2010@gehu.ac.in

Amit Mittal
Department of Allied Science,
, Graphic Era Hill University, Bhimtal, India
amitforestry26@gmail.com

Dinesh C. Dobhal
Department of Computer Science Application,
, Graphic Era Deemed to be University, Dehradun, India
dineshdobhal@gmail.com

***Abstract*: Diabetic retinopathy is a prevalent eye disorder that occurs in individuals with diabetes. The condition is caused by damage to the retinal blood vessels caused by elevated blood sugar levels. The retina is the tissue behind the eye that is sensitive to light. Among the leading causes of blindness worldwide is diabetic retinopathy. However, early detection and prompt treatment can effectively prevent or minimize vision loss. Deep learning algorithms, such as "You Only Look Once" (YOLO) Convolutional Neural Networks (CNNs), have demonstrated state-of-the-art performance on benchmark datasets and computer vision competitions. YOLO is a single end-to-end model that enables real-time object identification with high accuracy. Our proposed YOLOv8-based model integrated with Jetson Nano Developer Kit with CSI camera module shows the highest top1 accuracy of 93.77 on the training dataset and validation dataset with all given classes its measured accuracy top1 accuracy is 72.6 with 2.6ms inference time.**

Keywords—Diabetic Retinopathy (DR), YOLO, YOLOv8, Machine learning (ML) algorithms, Jetson Nano Developer Kit, LabelMe.

I. INTRODUCTION

Diabetic Retinopathy (DR) is a condition caused by unstable blood glucose levels, which results in various disorders that cannot be treated. It is a prevalent disease that starts with mild symptoms, progresses to intermediate stages, and ultimately leads to complete loss of vision. According to estimates, by 2030, the number of people diagnosed with DR is expected to increase to 191 million, up from the current global diagnosis rate of 126.6 million [1]. Furthermore, Vision Threatening Diabetic Retinopathy (VTDR) cases are projected to increase from 37.3 million to 56.3 million if no interventions are implemented.

Extensive research has been conducted to find effective treatments for ailments such as diabetic retinopathy (DR). Early detection and effective management of diabetic retinopathy (DR), a potentially sight-threatening complication of diabetes, is crucial to prevent further damage to the retina. Accurate categorization of the disease stage plays a vital role in assisting ophthalmologists in determining the most appropriate treatment strategies to halt or slow down the progression of DR [2].

Object Detection is a process aimed at identifying all objects within an image. It involves training a dataset to enable the system to detect objects on its own. Existing methods have limitations such as long training times, unsuitability for real-time applications, and scalability issues. Differentiating between objects of the same type is a challenge for machines. Object Detection is crucial in various scenarios such as hospitals, traffic control, and self-driving cars. The goal is to create a real-time Object Detection system that combines image classification and object localization, involving assigning class labels and drawing bounding boxes around objects in an image.

Two types of problems can be solved by supervised machine learning, namely regression and classification. In the context of image processing, image classification refers to the task of identifying objects within images, and is akin to traditional classification methods. Following image classification, the next step often involves localizing the identified objects using bounding boxes, which are rectangular boxes that are determined around objects using deep neural networks. This process, known as object detection, can be implemented in either one or two stages, depending on the specific approach employed.

ISBN: 979-8-3503-3086-1/23/$31.00©2023 IEEE

An object detector with two stages is widely used in computer vision tasks for detecting objects. The detectors use a Region Proposal Network to generate Regions of Interest (RoI) in the first stage and then predict objects and bounding boxes for these regions in the second stage. In addition to RCNNs, Fast RCNNs, and Faster RCNNs are also examples of two-stage detectors. These detectors are favored for their ability to select highly probable region proposals and achieve high accuracy in object detection.

Single-stage object detectors are simpler in architecture and are intended to detect objects by taking into account all spatial dimensions at the same time. These detectors can identify objects and their respective probability of belonging to a certain class in one go, eliminating the need for a separate stage for region proposals. The output includes bounding boxes around the objects detected.

Over the past few years, there has been a significant rise in the attention given to solving the issue of object detection in a single step by utilizing deep neural networks. This is primarily due to the evolution of You Only Look Once (YOLO) and its subsequent versions. These algorithms formulate the localization problem as a regression task and have gained popularity for their efficiency and accuracy. While YOLO is one of the well-known algorithms in this category, it is worth noting that other recent algorithms such as SSD [3], DSSD [4], RetinaNet [5], M2Det [6], and RefineDet++ [7] also utilize the Single Shot Detector (SSD) approach for object detection.

Because YOLO is simple, low-complexity, and easy to implement, despite being more complex and powerful, it usually outperforms multi-stage detectors. YOLO has gained popularity as a common choice in production settings due to its high accuracy and fast inference time, making it a tough competitor not only to other two-stage detectors but also to previous single-stage detectors.

As the YOLO (You Only Look Once) framework continues to advance, several key trends and possibilities are expected to shape its future developments. Researchers and developers will incorporate state-of-the-art techniques in deep learning, data augmentation, and training methodologies into the YOLO architecture. This ongoing process of innovation is likely to result in improved performance, robustness, and efficiency of the model.

The way we test object detection models might change in the future. The current benchmark, COCO 2017, could be replaced. The new benchmark would be more advanced and difficult. This transition from less demanding benchmarks, such as VOC 2007 used in earlier versions of YOLO, reflects the need to keep pace with the increasing sophistication and accuracy of models.

There is anticipation that the number of YOLO models and their applications will continue to grow, as more models are introduced each year and their uses expand. The versatility and power of the YOLO framework are likely to enable its application in diverse domains, ranging from home appliances to autonomous cars.

Additionally, YOLO models possess the ability to expand their functionalities beyond just detecting and segmenting objects. They can also be utilized for tracking objects in videos and estimating 3D key points. As the development of these models progresses, they have the potential to serve as the basis for novel solutions that cater to a wider variety of computer vision assignments.

Furthermore, it is anticipated that YOLO models will be flexible enough to adjust to a wide range of hardware platforms, spanning from small IoT gadgets to high-performance computing clusters. Such flexibility will permit the utilization of YOLO models in a variety of different settings, depending on the specific needs and limitations of the application.
Customizing YOLO models to fit various hardware requirements can increase their availability and usefulness to diverse users and industries.

II. Literature Survey

The following research papers discuss various approaches to object detection in computer vision YOLO algorithm, as proposed by Joseph Redmon in "You Only Look Once: Unified, Real-Time Object Detection," which employs a regression algorithm to achieve real-time object detection with high accuracy [8]. "Understanding of Object Detection Based on CNN Family and YOLO" by Juan Du provides a general overview of several other object detection strategies, including CNN and R-CNN, and compares their efficiency, introducing the YOLO algorithm as a more efficient approach [9]. Lastly, "Learning to Localize Objects with Structured Output Regression" by Matthew B. Blaschko focuses on object localization, utilizing a bounding box method as a solution to the limitations of the sliding window method [10].

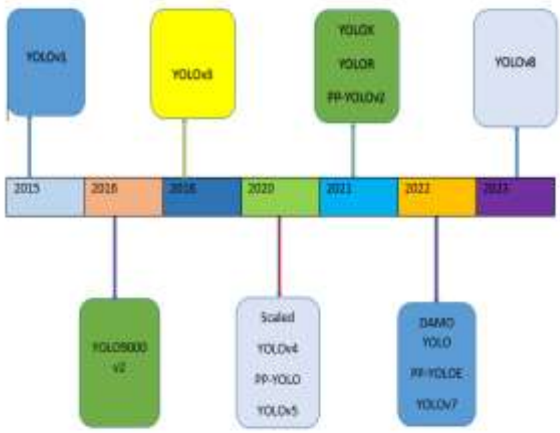

Fig. 1. All YOLO versions with their released date

Implementing object detection to solve real-world category vision applications poses significant challenges, such as resource allocation, system robustness, scalability, efficiency, and latency. Moreover, machine learning-based computer vision often requires Internet of Things (IoT) communication for data streaming, where images serve as input and produced detection results as output. To address these challenges, the concept of Edge AI has emerged, utilizing Edge Computing in combination with Machine Learning. Edge AI facilitates ML processing at the edge of the network, closer to the data source (e.g., camera), thereby forming distributed edge systems comprising multiple interconnected edge devices (such as Mobile Edge Computing or cloud-based edge).

Object detection, a key technology in computer vision, is typically performed on edge devices equipped with CPUs or GPUs, as well as specialized hardware such as neural processing units (NPUs) or vision accelerators. These NPUs have gained popularity for their efficiency in AI-based computer vision inference tasks. Examples of such NPUs include the Neural Compute Stick (NCS) by Intel, Jetson AI edge devices by NVidia and Neural Processing Engine by Qualcomm.

III. RESEARCH METHODOLOGY

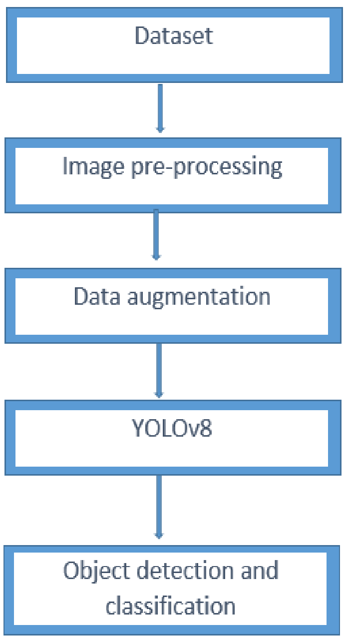

Fig. 2. Proposed author machine learning pipeline

3.1 Image pre-processing

The Indian Diabetic Retinopathy Image Dataset (IDRID) is a comprehensive open-access resource that serves as the first database of the Indian population for studying diabetic retinopathy (DR) [11]. The dataset comprises retinal fundus images, which are divided into two sections: the first half section contains images with DR and diabetic macular edema (DME) annotations, while the second section contains normal retinal images without any DR or DME annotations. In total, the DR segmentation dataset of IDRID consists of 122 images that showcase various DR anomalies such as microaneurysms (MA), hard exudates (EX), hemorrhage's (HE), and soft exudates (SE).

Table 3.1 Dataset details

Severity Level category	**Total number of images**
Microaneurysms (MA) category	81
Soft Exudates (SE) category	40
Hard Exudates (EX) category	81
Hemorrhage's (HE) category	80

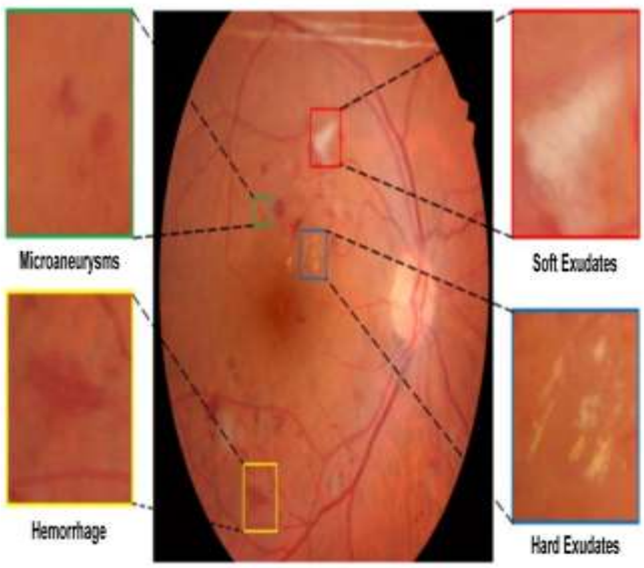

Fig. 3. Annotated fundus picture showing microaneurysms, haemorrhages, soft exudates, and hard exudates.

3.2 Data augmentation

Data augmentation techniques are useful to increase the size of a real dataset by generating multiple variations of the original data. These techniques are commonly utilized in computer vision and natural language processing (NLP) models to mitigate issues related to limited data availability and lack of diversity. Incorporating data augmentation strategies can significantly benefit machine learning models. Experimental results indicate that a deep learning model trained with various image augmentation techniques outperforms a model which is trained without augmentation in terms of training loss (i.e., penalization for incorrect predictions) and precision, as well as validation loss and accuracy, particularly in the context of image classification tasks.

3.3 YOLOv8

YOLOv8, developed by Ultralytics, is the latest generation of YOLO-based Object Detection models, known for their cutting-edge performance. With the previous versions of YOLO as a foundation, YOLOv8 has been developed to enhance both speed and accuracy. Furthermore, YOLOv8 presents a combined framework for training models to carry out Image Classification, Object Detection, and Instance Segmentation tasks. Although some features are still pending in the Ultralytics YOLOv8 repository, such as the complete set of export features for trained models, the team plans to release a paper on Arxiv that compares YOLOv8 with other state-of-the-art vision models, providing valuable insights into its capabilities .

In the YOLOv8 model series, there are five different models available for detection, segmentation, and classification tasks, each falling into a specific category. The YOLOv8 Nano model is the smallest and fastest, while the YOLOv8 Extra Large (YOLOv8x) model is the most accurate but slower in terms of performance. The models in the YOLOv8 series are named as follows: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x.

Furthermore, YOLOv8 is equipped with pre-existing models for object detection, image classification, and instance segmentation. These particular models were taught utilizing the well-known COCO dataset for identification and segmentation, with a resolution of 640, and the ImageNet dataset for categorization, with a resolution of 224.

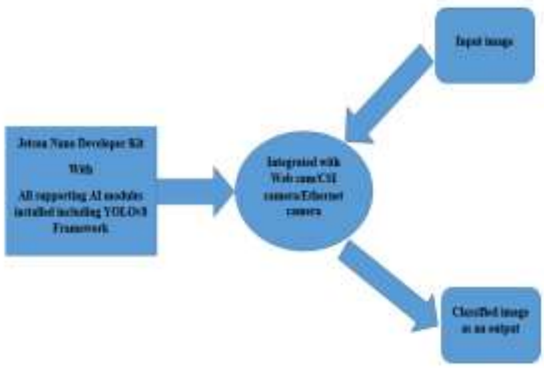

Fig.4. Proposed edge device diagram

3.4 Results

We have configured Jetson Nano developer with YOLOv8. We installed all the necessary libraries on the Jetson Nano developer to run YOLOv8. Our experiment utilized Ultralytics YOLOv8 as a tool. We clone the YOLOv8 GitHub repository on Jetson and follow all the steps to properly configure YOLOv8. For input initially, a webcam is used but users can also use other input devices like Camera Serial Interface (CSI) camera or Ethernet camera.

We have used pre-trained weight for YOLOv8 and then finally train that model on our custom dataset.

train/loss	mertics/accuracy_top1	metrics/accuracy_top5	val/loss
0.26657	0.7242	1	0.69499
0.17981	0.83452	1	0.41385
0.15695	0.86477	1	0.39399
0.14038	0.89146	1	0.32446
0.13061	0.89146	1	0.31535
0.11974	0.90391	1	0.22824
0.11335	0.90925	1	0.2659
0.10687	0.92349	1	0.23045
0.09837	0.93772	1	0.20348

Fig.5. The top1 and top5 training accuracy and loss

Classes	top1_acc	top5_acc
All	0.726	1
Speed:	0.4ms	
Inference:	2.6ms	

Fig.6. the top1 and top5 training accuracy and loss

The above figures show top1 accuracy of 93.77 on the training dataset and a top1 accuracy of 72.6 on all classes available on validation with 2.6ms inference time.

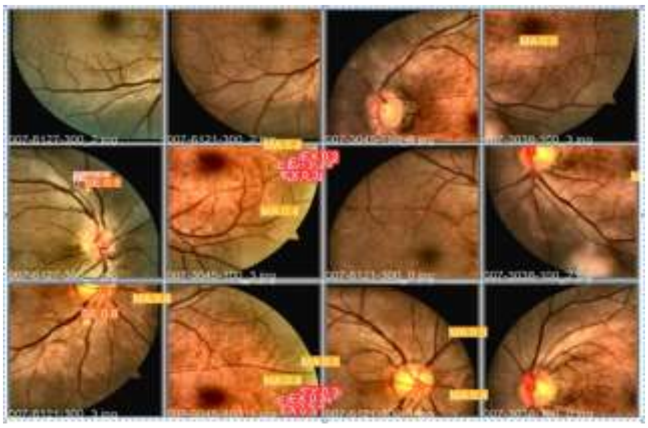

Fig.7. Result of a validation dataset

IV. CONCLUSION AND FUTURE SCOPE

In conclusion, our study highlights the potential of deep learning algorithms, specifically the YOLOv8-based model integrated with Jetson Nano Developer Kit with CSI camera
Module, for early identification and detection of diabetic retinopathy. The results obtained from our proposed model demonstrate high accuracy rates, with a top1 accuracy of 93.77 on the training dataset and a top1 accuracy of 72.6 on the validation dataset with all given classes, along with a fast inference time of 2.6ms. These findings suggest that our model has the potential to assist in the timely identification and treatment of diabetic retinopathy, which could help prevent or reduce vision loss in diabetic patients. Further research and validation on larger datasets and clinical trials are warranted to establish the clinical utility of our proposed model for diabetic retinopathy screening and management. Overall, our study contributes to the growing body of literature on the application of deep learning algorithms in medical imaging, specifically for diabetic retinopathy, and holds promise for improving early diagnosis and management of this sight-threatening condition.

REFERENCES

[1] Zheng Y, He M, Congdon N. The worldwide epidemic of diabetic retinopathy. Indian J Ophthalmol. 2012 Sep-Oct;60(5):428-31. doi: 10.4103/0301-4738.100542. PMID: 22944754; PMCID: PMC3491270.

[2] Sivaprasad S, Pearce E. The unmet need for better risk stratification of non-proliferative diabetic retinopathy. Diabet Med. 2019 Apr;36(4):424-433. doi: 10.1111/dme.13868. Epub 2018 Dec 7. PMID: 30474144; PMCID: PMC6587728.

[3] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) Ssd: single shot multibox detector. European Conf Comput Vis 2016:21–37. https://doi.org/10.48550/arXiv.1512.02325

[4] Fu CY, Liu W, Ranga A, Tyagi A, Berg AC (2017) Dssd: Deconvolutional single shot detector. arXiv preprint arXiv:1701.06659

[5] Lin TY, Goyal P, Girshick R, He K, Dollár P (2017) Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision, pp 2980–2988. https://doi.org/10.48550/arXiv.1708.02002

[6] Zhao Q, Sheng T, Wang Y, Tang Z, Chen Y, Cai L, Ling H (2019) M2det: a single-shot object detector based on multi-level feature pyramid network. Proceed AAAI Conf Artif Intell 33:9259–9266

[7] Zhang S, Wen L, Lei Z, Li SZ (2020) RefineDet++: single-shot refinement neural network for object detection. IEEE Trans Circuits Syst Video Technol 31(2):674–687. https://doi.org/10.1109/TCSVT.2020.2986402

[8] Joseph Redmon, Santosh Divvala, Ross Girshick, “You Only Look Once: Unified, Real-Time Object Detection”,The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.

[9] YOLO Juan Du1,”Understanding of Object Detection Based on CNN Family”,New Research, and Development Center of Hisense, Qingdao 266071, China.

[10] Matthew B. Blaschko Christoph H. Lampert, “Learning to Localize Objects with Structured Output Regression”, Published in Computer Vision – ECCV 2008 pp 2-15.

[11] Prasanna Porwal, Samiksha Pachade, Ravi Kamble, Manesh Kokare,
Girish Deshmukh, Vivek Sahasrabuddhe, and Fabrice Meriaudeau,”
Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research” Data, vol. 3, June 2018.