# Enhancing Depth Perception in Single-Lens Cameras through Barrier Detection and Artificial Intelligence

**Rakesh Kumar Yadav[1], Karthikeyan M P[2], Dr. Naresh Kaushik[3], Akash Kumar Bhagat[4], Prakriti Kapoor[5], Shikhar Gupta[6]**

*[1]Associate Professor, Maharishi School of Engineering & Technology, Maharishi University of Information Technology, Uttar Pradesh, India, Email Id- rkymuit@gmail.com, Orcid Id- 0000-0002-0151-4981*

*[2]Assistant Professor, Department of Computer Science and Information Technology, JAIN (Deemed-to-be University), Bangalore, karnataka, India, Email Id- karthikeyan.mp@jainuniversity.ac.in, Orcid Id- 0000-0002-2346-0283*

*[3]Assistant Professor, Department of uGDX, ATLAS SkillTech University, Mumbai, Maharashta, India, Email Id- naresh.kaushik@atlasuniversity.edu.in, Orcid Id- 0000-0002-9896-4662*

*[4]Assistant Professor, Department of Computer Science & IT, ARKA JAIN University, Jamshedpur, Jharkhand, India, Email Id- akash.b@arkajainuniversity.ac.in, Orcid Id- 0000-0001-8717-764X*

*[5]Centre of Research Impact and Outcome, Chitkara University, Rajpura- 140417, Punjab, India prakriti.kapoor.orp@chitkara.edu.in https://orcid.org/0009-0002-0877-3545*

*[6]Chitkara Centre for Research and Development, Chitkara University, Himachal Pradesh-174103 India shikhar.gupta.orp@chitkara.edu.in https://orcid.org/0009-0004-0138-3987*

Single-lens cameras are popular but lack depth perception. Augmented reality, autonomous navigation and 3D reconstruction are limited by scene depth information. For autonomous navigation and object recognition, single-lens camera depth perception can be improved. Single-lens camera systems need to improve their depth assessment to be feasible. This study tackles this limitation by combining barrier detection with AI techniques to improve depth perception in single-lens cameras. Better depth perception is the goal of jpg image dataset, a standard for autonomous driving. In preparation, the dataset is rotated, flipped and scaled to improve training. The Single Shot Detector (SSD) has a barrier detection module that detects and evaluates scene depth cues. The Enriched Recurrent Neural Network (ERNN) improves depth perception. ERNN uses temporal relationships in sequential data to improve scene comprehension and depth estimate. The suggested method is compared to current methods using Error and Accuracy Parameters. Our experiments show that single-lens cameras can perceive depth better than earlier methods. This study shows how to overcome current limits and improve depth perception in single-lens cameras. SSD barrier detection and ERNN depth perception speed things up. Single-lens cameras will get more accurate depth estimates in real life with this strategy.

**Keywords:** Artificial Intelligence, Single Shot Detector (SSD), Single-Lens Cameras, Depth Perception, Enriched Recurrent Neural Network (ERNN).

## 1. Introduction

Depth prediction is challenging for artificial intelligence (AI), which impacts visual perception. Although synchronization and picture generation are complicated, dedicated hardware, multi-view geometry and machine learning (ML) approaches are employed. Accurate depth estimation is essential for visual perception and poorly limited picture generation might produce inaccurate outcomes [1]. Depth retrieval is necessary for distinguishing individuals in observation. This can be achieved by capturing several pictures from various angles using many cameras. The main challenge is dealing with conventional photography that captures two-dimensional projections of 3D scenes [2]. The emerging field of AI uses segmentation and object identification to help persons with psychological and physical problems. This gadget helps vision-impaired people overcome perilous outdoor obstacles. AI can help them by broadening their senses and delivering precise descriptions [3]. The capacity to distinguish between short depth intervals using stereo acuity has been demonstrated to be sensitive to blur caused by monocular vision. Numerous studies indicate lower stereo acuity with mono-vision correction in presbyopes, consistent with this sensitivity [4]. Visual perception, graphics, recognition and robotics depend on depth estimation in AI.

However, dedicated hardware, multi-view geometry and ML are expensive and power-hungry are limited to specialized situations [5]. Developing bezel-less smart phones requires mass manufacture of camera lens modules before assembly. A camera lens module typically found smart phones consists of stacked spherical and aspheric lenses in a customized barrel [6]. Fast photographic cameras enable periodic field screening of huge field plots, breeder individuals and genomic vaults. A phenomenotypic record that is accurate, unbiased and comparative can be produced with little to no human intervention through effective data automation manufacturing. Disposable sensors, for example, are quick, easy to use and deliver convenience and a benefit-cost. [7]. Interest in fisheye cameras has increased due to their wide field of view, enabling applications like augmented or virtual reality (AR/VR) content and robot vision. They need improvement due to geometric distortions and object appearance changes [8].

## 2. Literature Review

According to the author of, [9] improved real-time object identification in industrial automation, construction, healthcare and autonomous cars with You Only Look Once (YOLO) and Faster Region Convolutional Neural Network (Faster R-CNN) algorithms despite data privacy and computational complexity. Study [10] analyzed deep learning (DL) systems to discriminate malignant and benign tumors in various visual modalities. These technologies are in early clinical trials despite claims of improved accuracy. This review covers digital image-based AI advances and problems. The proposed a glasses-free 3D auto-stereoscopic display system that uses an eye monitoring technique for presenting computed tomography (CT) [11]

cardiac images. The system employs various machine-learning approaches to generate pictures of excellent quality with minimum crosstalk while making use of an illumination unit and slit barrier. To developed Double Star, a long-range assault that worked on commercialized stereo imaging devices and validated on actual devices [12]. Double star injected artificial obstacle depth employing pure light from two sources in autonomous machines.

The study examine a house-wide fall alert system that distinguishes between real crashes and routine activities using machine learning and an inexpensive single-board computer [13]. Every fall in high-risk areas guarantees security in assisted living facilities, rehabilitation institutes and elderly communities. The study analyzes the crucial visual component of near-eye displays that lower computational load and lessen visual discomfort, including foveation and defocus [14]. There is a need for more research to be done on the association between lower discrimination thresholds and visual eccentricity, as evidenced by two psychophysical tests. To employ the technological operation of radar detectors, Light Detection and Ranging (LiDAR) [15] along with vision cameras in self-driving cars are examined in this study with an emphasis on their usefulness and possibilities. It looks at multi-sensor fusion methods and freely available calibrated tools for self-driving automobile devices' identifying objects. Author [16] autonomous driving relies on detection for path planning, motion prediction and collision avoidance. Challenges include semantic alignments, representation training and perception repair. The survey covers detectors, performance indicators and techniques. Studies [17] present a procedure for automatic recognizing collisions that focuses on obstacle lengths from single-camera images or movies. Even while cameras are used for collision detection, existing techniques must be safer for automated driving.

Research gap is need for the ability of current single-lens camera systems to reliably assess depth and restricts their usefulness in practical situations. This study overcomes this restriction by presenting a unique strategy for improving depth perception in single-lens cameras by combining barrier detection with artificial intelligence (AI) algorithms.

Significant of the study is single-lens cameras with barrier detection and AI increase depth perception, focusing, portrait mode, 3D scene reconstruction, AR, object recognition and real-time depth mapping. Improvements boost autofocus, portrait mode and 3D scene reconstruction. Tracking and segmenting objects improves low-light performance and dynamic scene interpretation. These developments produce better imaging.

Objective of the study is optimizing the depth vision in single-Lens cameras involves a framework using AI-based Barrier Detector-Depth Perception (BD-DP) architecture. The Single Shot Detector (SSD) is used for barrier detection, while Enriched Recurrent Neural Networks (ERNN) is used for depth perception. This approach improves barrier distance estimation accuracy and efficiency, enhancing the performance of single-lens cameras.

The following are the next section of the study; Section 2: Study Methods, Section 3: Result and Discussion and Section 4: Conclusion is discussed.

## 3. Method

A single-lens camera using SSD for boundary detection and ERNN for depth perception combines the strengths of both architectures. A more detailed system flow follows (Figure 1).

The SSD and ERNN are used in BD-DP in a single-lens camera to combine information from both tasks. This fusion technique involves combining features from both modules, training a classifier and evaluating the performance on diverse datasets. Regular testing along with validation ensures the robustness and generalization of the fused model in various scenarios.

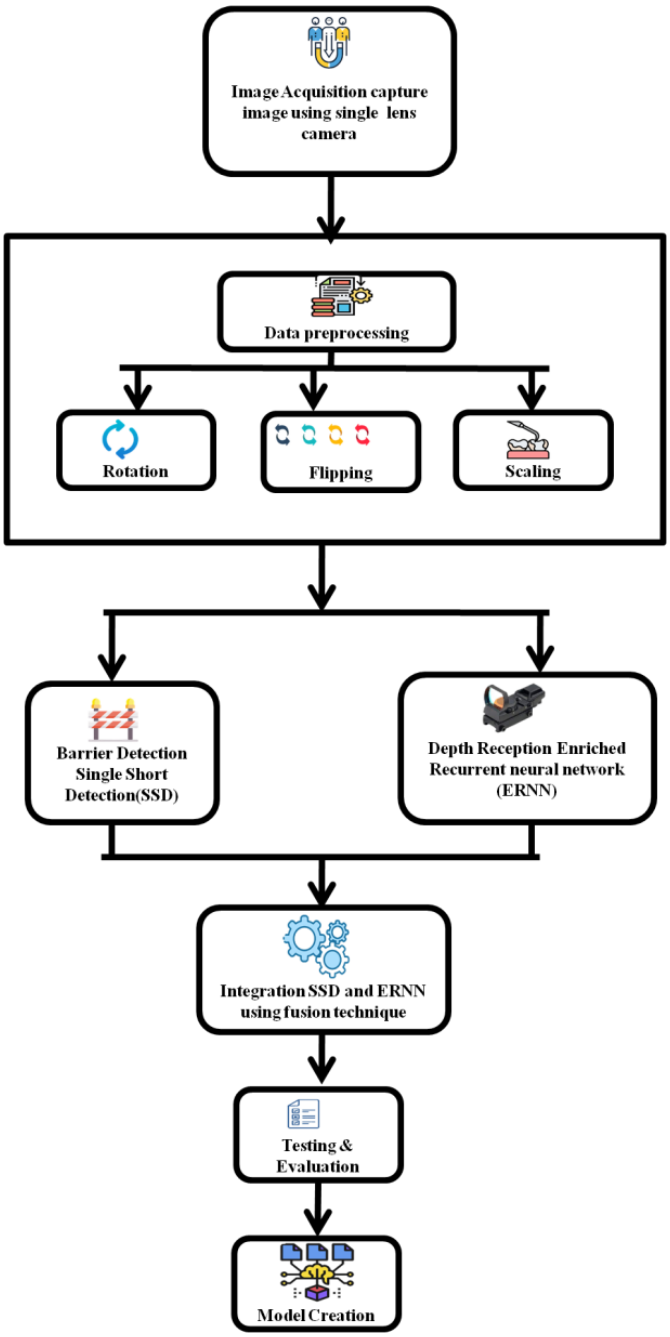


Figure 1: Work flow Model [Source: Author]

### 3.1 Image Acquisition

The research and training a single lens camera captured image and BD-DP algorithms dataset consisting of RGB photos (280 x 960 pixels) charged with a single lens camera, stored in JPG format and includes object class, bounding box coordinates, camera angle and object distance as in figure 2.

Figure 2: Input Image Capture by Single Lens Camera (source: https://vancouverpublicspace.ca/dev/wordpress/wp-content/uploads/2017/03/Sergio-Ruiz-Flickr-Mint-Plaza-406x272.jpg )

### 3.2 Data Preprocessing

The dataset was preprocessed by replacing bounding box coordinates with object detector positions using the suggested framework. Normalize the captured images to ensure consistency and resize the pictures if necessary. Model performance was compared using the original and identified bounding boxes. Resize images to a standard size. Normalize pixel values. Augment data to increase the diversity of the dataset.

- Rotation,
- Flipping,
- Scaling.

The intersection over the union function found a border percentage integrating. Learning, verification and evaluation datasets were randomly selected from the 26,927-object modified dataset.

### 3.3 Barrier Detection Using Single Shot Detector

The dataset trains a Single Shot Detector (SSD) model for barrier detection, which is adjusted for specific barrier types, bounding boxes and class scores are extracted for each object. SSD is Convolution Neural Network (CNN) architecture for object localization, generating feature maps at different scales. It predicts object class scores and bounding box offsets using multiple convolutional layers. SSD performs feature fusion across various scales to improve detection accuracy. The model is trained using a composite loss function, combining localization and confidence loss. It is trained on a dataset with annotated bounding boxes and class labels. SSD is suitable for real-time object detection in applications like autonomous vehicles, surveillance,

robotics, ensuring efficient and effective object detection. Convolutional networks for feature extraction, such as VGG, ResNet, Inception, Inception Resnet-v2 and MobileNets, are the main subjects in the study.

To detect more minor things, the first layer creates a large-scale feature map; to catch more prominent objects, the second layer creates multi-scale feature maps (See Figure 3). Using a tiny kernel, the final convolutional layers forecast bounding box positions and confidences for several categories. To generate 8952 boxes per class, the SSD network consists of the MobileNets basic model, additional feature layers and classifier convolutional layers. The dimensions of the input photos are 300 by 300 (SSD300). MobileNets is used instead of VGG-16 or RESNET to increase accuracy and detection speed. Mobile networks are perfect for real-time object detection because they use less computational resources and shorter processing time. They are smaller and less computationally demanding. Each item needs an input image and ground truth boxes to train it. Several dimension feature maps examine a limited collection of standard shapes with varying perspective percentages. For each object category, including wildflowers and cats, these boxes forecast shape variances and their trust. These default boxes are first matched to the ground truth boxes during training; the huge barrier box for 8 x 8 scales and the tiny barrier box for 4 × 4 scales are considered positives and the remaining packages are considered negatives. Localization loss and confidence loss are combined to form the weighted model loss. SSD generates several feature maps that align with objects and provide the thing's name. Localization and classification loss comprise the loss function used to assess the SSD model.

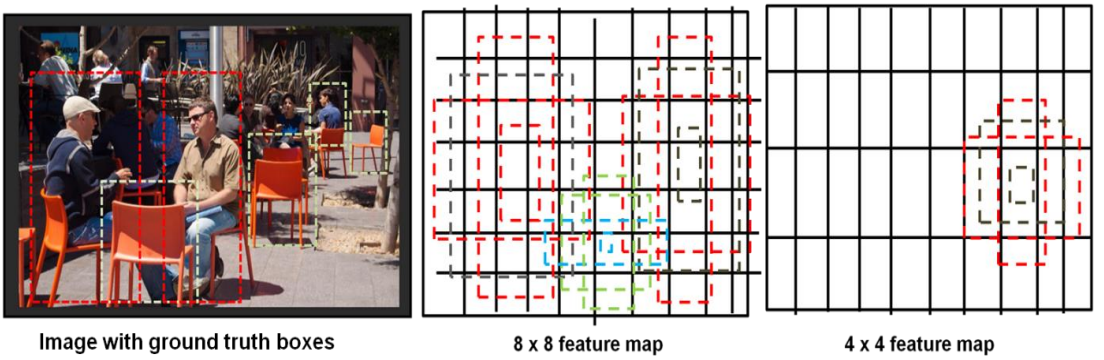


Figure 3: SSD Frame Work [Source: Author]

The loss can be calculated in this way in Equation (1).

$$\mathcal{L} = \frac{1}{M}\left(L_{con\,e} + L_{loc}\right) \quad (1)$$

Here, $L_{loc}$ stands for localization loss, $L_{conf}$ for confidence loss and M for the number of positive matches in Equation (2).

$$\mathcal{L}_{loc}(w, l, h) = \sum_{j\epsilon\,Pos}^{M} \sum_{n\epsilon dw,dz,x,g} w_{j.i}^{l}\, smooth_K 1(k_j^n - h_i^n) \quad (2)$$

$$h_i^{dw} = (h_i^{dw} - c_j^{dw})/c_j^x h_i^{dz} = (h_i^{dz} - c_j^{dz})/c_j^z \quad (3)$$

$$h_i^x = \log \frac{h_i^x}{c_i^x} \quad (4)$$

$$h_i^g = \log \frac{h_i^g}{c_i^g} \quad (5)$$

Between the 'Predicted-Box' (k) and the 'Ground-Truth-Box' (h) parameters, the localization loss is a $\text{smooth}_K$loss. The Default-Bounding-Boxes (w) middle, dimension (x) and length (g) are $(dw, wz)$ the softmax loss over several classes, or confidence (c), represents the confidence loss and it is displayed as follows in Equation (6):

$$\mathcal{L}_{loc}(w, z) = -\sum_{j\epsilon Pos}^{M} w_{ji}^{o} \log\left(c_j^o\right) - \sum_{j\epsilon Neg}^{M} \log\left(c_j^o\right)\left(c_j^o\right) \quad (6)$$

Here $c_j^o = \frac{\exp c_j^o}{\sum_o \exp c_j^o}$.

### 3.4 Enriched Recurrent Neural Network (ERNN) -Based Depth Perception

Develop a depth perception module to determine the depth of the barriers that have been identified, with the help of an ERNN. Provide depth charts or values related to the obstacles that were found. Artificial intelligence (AI) has advanced due to recent developments in deep learning, making it feasible to obtain fundamental characteristics from raw sensory information. Gated Recurrent Unit (GRU) cells are a crucial choice for recurrent neural networks (RNNs) in tasks involving BD-DP in a single-lens camera. These cells are designed to work with sequential data, making them suitable for time-series information studies. The goal is to produce accurate class predictions based on the temporal sequence of time-of-flight measurements, understanding patterns and relationships. Figure 4 tells the depth map for the input image. GRU cells can retain long-term material information through internal gates and optimizable parameters. These gates control the flow of information and allow the model to better capture patterns as well as dependencies in the time-of-flight measurements.

Figure 4: Estimated Depth Map [Source: Author]

#### 3.4.1 Gated-Recurrent Unit

We present GRUs following the depth perception in an object. Set $z \in D$ be the ground truth class label and let $w = (w_1, \ldots., w_s, \ldots, w_S), w_s \epsilon \mathbb{R}^m$be an ordered collection of S events. A GRU cell gets $w_s$ at each successive cycle and produces a stimulus $g_s \epsilon \mathbb{R}^n$answer in Equation

(7).

$$g_s^i = \left(1 - y_s^i\right)g_{s-1}^i + y_s^i g_s^i \quad (7)$$

Through merging a potential stimulation via the present period $\acute{g}_s^i$ with inspiration via the preceding time $g_{s-1}^i$ . The modification gate or trade-off factor $y_s^i$, is computed as in Equation (8),

$$y_s^i = \sigma(X_y w_s + V_y g_{s-1})^i \quad (8)$$

Where, σ is an exponential function that produces values in the interval (0, 1) and $x_y \in \mathbb{R}^{n\times m}$ and $V_y \in \mathbb{R}^{n\times m}$ is optimizable parameters identical throughout each s. It computes the potential deactivation in Equation (9).

$$\acute{g}_s^i = \tanh\left(X_{ws} + V(w_s \odot g_{s-1})^i\right) \quad (9)$$

Where $w_s$ frequently implied to as the initialization gate and $\odot$ are the element-wise products of the two vectors. Keep in mind that $X \in \mathbb{R}^{n\times m}$ and $V \in \mathbb{R}^{n\times m}$ is distinct parameter sets from Equation (10) $X_y$ and $V_y$.

The gate that is reset is comparable to the updated gate $w_s$.

$$q_s^i = \sigma(X_q w_s + V_q g_{s-1})^i \quad (10)$$

The final phase injects the last bit of GRU stimulation at time S into a dense layer with an activation function similar to soft-max in Equation (11-12).

The logarithm for $y^j$is calculated from the dense layer using,

$$y^j = \sum_i x_t^{ji} g_S^i, \quad (11)$$

Where the soft-max layer weights are represented by $x_t = (x_t^{ji})$, next, the sequence of segmentation labeling can be acquired by implementing the function for softmax activation.

$$\acute{y}^j = \frac{f^{y^j}}{\sum_j f^{yj}} \quad (12)$$

Moreover, stacking GRU layers can create a more profound GRU architecture. The sequence of observations x is sent into the first GRU layer and the activation outputs of each succeeding layer are fed into them. At last, we transmit the stimulation of the highest stacking layer to the soft-max dense layer. The framework is shown in Figure 5. Circles represent the concatenation procedure. Grey rectangles represent bidirectional GRU layers and arrays represent information flow.

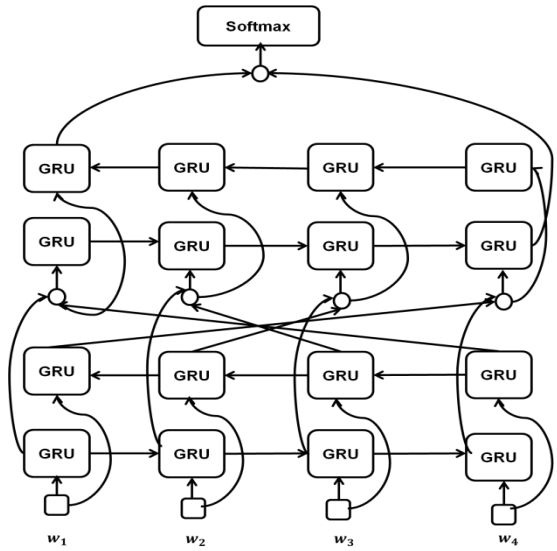


Figure 5: Fundamental Structure of GRU layer

[Source: https://www.mdpi.com/1099-4300/21/4/414/pdf]

3.5 Integrating of SSD and ERNN

Utilizing a multi-model fusion technique entails the identification of barriers in the input frame by an object detection module such as SSD and estimating the depth of the discovered walls by a depth perception module such as ERNN GRU-based Depth Estimation. The fusion technique combines features from both modules at a higher level to produce a fused feature representation.

## 4. Result and Discussion

The model was implemented using PyTorch 1.9.1 on specialized hardware like a Field-Programmable Gate Array (FPGA) for BD-DP on computers or phones. The study evaluated various size barrier prediction models; with ERNN performing better in error metric (see Table 1 and Figure 6) and accuracy metric (see Table 2 and Figure 7). The SSD model was implemented for barrier detection in distance prediction frameworks for mini or massive-scale objects. ERNN achieved depth perception for single-lens cameras.

4.1 Evaluation matrix

To assess how well the barrier-detecting models performed using the data set is tested, this experiment used the Mean-Absolute-Error (MAE). Equation (13) establishes the MAE for the anticipated barrier distance,

$$\mathrm{MAE} = \frac{1}{\mathrm{M}}\sum_{\mathrm{j}=1}^{\mathrm{M}}\left|\mathrm{c_j} - \acute{\mathrm{c}}_\mathrm{j}\right| \quad (13)$$

Where, $\acute{\mathrm{c}}_\mathrm{j}$ is the anticipated barrier distance, $\mathrm{c_j}$ is the actual barrier distance and M is the total number of items as shown in Equation (14-18). This study employed an additional five

variables to compare the performance of the suggested structure with different approaches. The five metrics were the following: threshold accuracy (Threshold), root mean squared error (RMSE), squared relative difference (SquaRel), relative absolute mistake (AbsRel) and RMSE log.

$$\text{Abs Rel} = \frac{1}{M}\sum_{j=1}^{M}\frac{|c_j - \acute{c}_j|}{d_j} \tag{14}$$

$$\text{SqualRel} = \frac{1}{M}\sum_{j=1}^{M}\frac{\left||c_j - \acute{c}_j|\right|^2}{d_j} \tag{15}$$

$$\text{RMSE} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left|c_j - \acute{c}_j\right|^2} \tag{16}$$

$$\text{RMSE log} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left|\log c_j - \log\acute{c}_j\right|^2} \tag{17}$$

$$\text{Threshold} = \%\text{ of } c_j \text{ s.t.} \max\left(\frac{\acute{c}_j}{c_j}, \frac{\acute{c}_j}{c_j}\right) = \delta < \text{threshold} \tag{18}$$

The three values that the threshold value requires are "$\delta < 1.25$, $\delta < 1.25^2$ and $\delta < 1.25^3$". It calculates the merged image's contrast. A high standard deviation (δ) would accompany a highly colorful visual and obtain higher accuracy.

Table 1: Outcomes of Error and Accuracy Matrix for SSD-ERNN in a Single Lens Camera

| Method | Error Measurement | | | |
|---|---|---|---|---|
| | AbsRel | SqRel | RMSE | RMSElog |
| XGboost-LSTM [18] | 0.184 | 1.596 | 4.873 | 0.276 |
| Encoder-Decoder Architecture[19] | 0.048 | 0.146 | 2.097 | 0.278 |
| YOLO-Faster R-CNN [20] | 0.487 | 0.463 | 6.809 | 0.570 |
| SSD-ERNN (Proposed) | 0.048 | 0.047 | 0.094 | 0.083 |

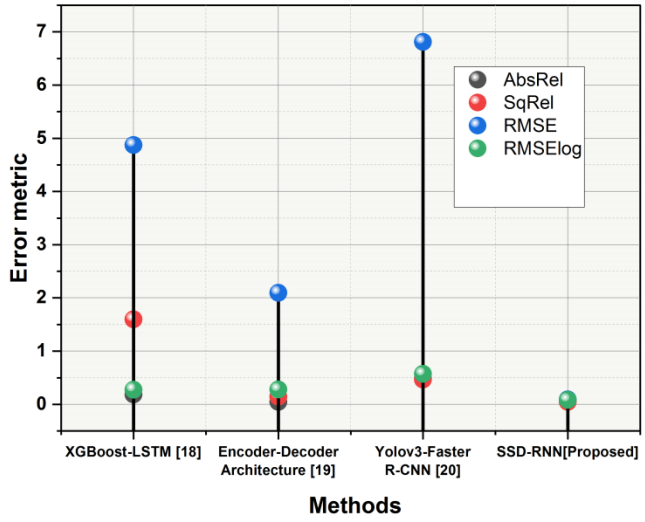


Figure 6: Conclusions of the SSD-ERNN Error Matrix in a Single Lens Camera

Table 2: Outcomes of Accuracy Matrix for SSD-ERNN in a Single Lens Camera

| Method | Accuracy (%) | | |
|---|---|---|---|
| | $\delta_1$ <1.25 | $\delta_2$<1.25$^2$ | $\delta_3$<1.25$^3$ |
| XGboost-LSTM [18] | 0.84 | 0.902 | 0.703 |
| Encoder-Decoder Architecture[19] | 0.89 | 0.92 | 0.847 |
| YOLO-Faster R-CNN [20] | 0.74 | 0.83 | 0.67 |
| SSD-ERNN (Proposed) | 0.989 | 0.997 | 0.982 |

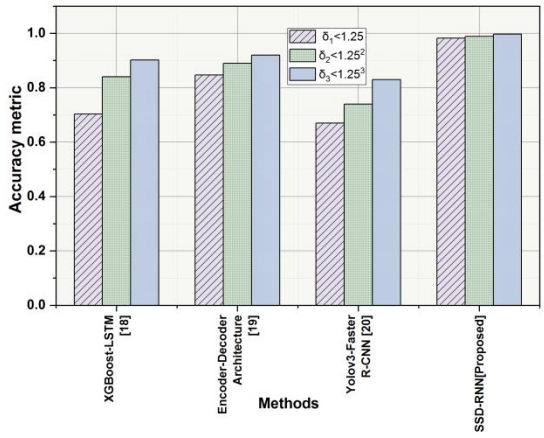


Figure 7: Consequences of the Single Lens Camera SSD-ERNN Accuracy Matrix

## 5. Discussion

BD-DP is a crucial task in AI, but Extreme Gradient Boosting (XGBoost) [18], a popular ML algorithm, is limited to tabular data and it is feature engineering dependent that can need help with generalization to unseen data. XGBoost is expensive, unsuitable for image data and requires careful hyper parameter tuning. Sequential LSTM [18] networks could not capture spatial dependencies well, rendering them unfit for these purposes. LSTM is more complex, computationally expensive and time-consuming to train. Long-term dependencies, lighting conditions and data efficiency can challenge LSTMs. They escape the vanishing gradient issue since they are not built for depth perception. Encoder-decoder architectures are effective for barrier detection and depth perception. They suffer from loss of information during encoding, limited receptive field, scale variation issues, training data dependency, dynamic scenes, computational complexity, hyper parameter sensitivity and occlusions. Despite these limitations, architectural changes, attention mechanisms and advanced training procedures help construct encoder-decoder models [19]. Popular object identification method YOLO-Faster R-CNN [20] detects barriers in single-lens cameras. Its shortcomings are its low localization accuracy, trouble recognizing small items, limited knowledge of object relationships, difficulty managing obstructed objects and difficulty seeing variable-sized things. Its performance depends on training data quality and diversity, yet its speed vs. accuracy trade-off could prove difficult. Faster R-CNN is utilized for depth perception and

barrier identification in single-lens cameras. Its drawbacks include computational complexity, slowness, training time, hyper parameter sensitivity, limited generalization to unforeseen circumstances, single-lens camera restrictions, occlusion handling problems and colossal model size. Computer vision research can improve these limits and applicability for specific applications [20].

## 6. Conclusion

The suggested model calculates the separations across large and small objects. The proposed framework's object detector finds the objects' bounding boxes and classes. The suggested framework's depth estimator calculates the acquired image's depth map. The Depth-Map is integrated in the bounding bundles to obtain depth-sensitive attributes for every item. Based on the object's depth characteristics and bounding box, the ERNN methodology is applied to estimate the depth perception between the small and big things. The accuracy of the suggested framework for calculating depth perception ranged from 98.9 to 99.7% depending on the barrier distance level and object size.

## References

1. Zhang, Linlin, Xiang Yu, Yaw Adu-Gyamfi, and Carlos Sun. "Spatio-temporal fusion of LiDAR and camera data for omnidirectional depth perception." Transportation Research Record (2023): 03611981231184187. https://doi.org/10.1088/1755-1315/542/1/012068
2. Goud, B. Karthik, SDVS Jagannadha Raju, and K. Divakar Rao. "Novel defocus hologram aberration compensation method in digital holography integrated pathology microscope for label-free 3-D imaging." Optics and Lasers in Engineering 140 (2021): 106514. https://doi.org/10.1016/j.optlaseng.2020.106514
3. Bauer, Zuria, Alejandro Dominguez, Edmanuel Cruz, Francisco Gomez-Donoso, Sergio Orts-Escolano, and Miguel Cazorla. "Enhancing perception for the visually impaired with deep learning techniques and low-cost wearable sensors." Pattern Recognition Letters 137 (2020): 27-36. https://doi.org/10.1016/j.patrec.2019.03.008
4. Smith, Carrie E., Robert S. Allison, Frances Wilkinson, and Laurie M. Wilcox. "Monovision: Consequences for depth perception from large disparities." Experimental Eye Research 183 (2019): 62-67. https://doi.org/10.1016/j.exer.2018.09.005
5. Garg, Rahul, Neal Wadhwa, Sameer Ansari, and Jonathan T. Barron. "Learning single camera depth estimation using dual-pixels." In Proceedings of the IEEE/CVF international conference on computer vision, pp. 7628-7637. 2019. http://openaccess.thecvf.com/content_ICCV_2019/html/Garg_Learning_Single_Camera_Depth_Estimation_Using_Dual-Pixels_ICCV_2019_paper.html
6. Kim, Sung Wook, Young Gon Lee, Bayu Adhi Tama, and Seungchul Lee. "Reliability-enhanced camera lens module classification using the semi-supervised regression method." Applied Sciences 10, no. 11 (2020): 3832. https://www.mdpi.com/2076-3417/10/11/3832#
7. Rudolph, R., Katja Herzog, Reinhard Töpfer, and V. Steinhage. "Efficient identification, localization and quantification of grapevine inflorescences and flowers in unprepared field images using Fully Convolutional Networks." Vitis 58, no. 3 (2019): 95-104. https://doi.org https://doi.org/10.5073/vitis.2019.58.95-104
8. Li, Tangwei, Guanjun Tong, Hongying Tang, Baoqing Li, and Bo Chen. "Fisheyedet: A self-

study and contour-based object detector in fisheye images." IEEE Access 8 (2020): 71739-71751. https://doi.org/10.1109/ACCESS.2020.2987868

9. Shianto, Kevin Adiputra, Kartika Gunadi, and Endang Setyati. "Deteksi jenis mobil menggunakan metode yolo dan faster r-cnn." Jurnal Infra 7, no. 1 (2019): 157-163. https://publication.petra.ac.id/index.php/teknik-informatika/article/download/8065/7275
10. Goyal, Manu, Thomas Knackstedt, Shaofeng Yan, and Saeed Hassanpour. "Artificial intelligence-based image classification methods for skin cancer diagnosis: Challenges and opportunities." Computers in biology and medicine 127 (2020): 104065. https://doi.org/10.1016/j.compbiomed.2020.104065
11. Kang, Dongwoo, Jin-Ho Choi, and Hyoseok Hwang. "Autostereoscopic 3d display system for 3d medical images." Applied Sciences 12, no. 9 (2022): 4288. https://doi.org/10.3390/app12094288
12. Zhou, Ce, Qiben Yan, Yan Shi, and Lichao Sun. "{Double Star} :{ Long-Range} Attack towards Depth Estimation based Obstacle Avoidance in Autonomous Systems." In 31st USENIX Security Symposium (USENIX Security 22), pp. 1885-1902. 2022. sec22-zhou-ce.pdf (usenix.org)
13. Shu, Francy, and Jeff Shu. "An eight-camera fall detection system using human fall pattern recognition via machine learning by a low-cost android box," Scientific Reports 11, no. 1 (2021): 2471. https://doi.org/10.1038/s41598-021-81115-9
14. Sun, Qi, Fu-Chung Huang, Li-Yi Wei, David Luebke, Arie Kaufman, and Joohwan Kim. "Eccentricity effects on blur and depth perception." Optics expresses 28, no. 5 (2020): 6734-6739. https://doi.org/10.1364/OE.28.006734
15. Yeong, De Jong, Gustavo Velasco-Hernandez, John Barry, and Joseph Walsh. "Sensor and sensor fusion technology in autonomous vehicles: A review." Sensors 21, no. 6 (2021): 2140. https://www.mdpi.com/1424-8220/21/6/2140#
16. Qian, Rui, Xin Lai, and Xirong Li. "3D object detection for autonomous driving: A survey." Pattern Recognition 130 (2022): 108796. https://doi.org/10.1016/j.patcog.2022.108796
17. Liu, Yang, Evan Gunnell, Yu Sun, and Hao Zheng. "An Object-Driven Collision Detection with 2D Cameras using Artificial Intelligence and Computer Vision." In CS & IT Conference Proceedings, vol. 12, no. 6. CS & IT Conference Proceedings, 2022. DOI: https://doi.org/10.5121/csit.2022.120626
18. Lee, Seungyoo, Kyujin Han, Seonyeong Park, and Xiaopeng Yang. "Vehicle Distance Estimation from a Monocular Camera for Advanced Driver Assistance Systems." Symmetry 14, no. 12 (2022): 2657. https://www.mdpi.com/2073-8994/14/12/2657#
19. Khan, Faisal, Shahid Hussain, Shubhajit Basak, Joseph Lemley, and Peter Corcoran. "An efficient encoder–decoder model for portrait depth estimation from single images trained on pixel-accurate synthetic data." Neural Networks 142 (2021): 479-491. https://doi.org/10.1016/j.neunet.2021.07.007
20. Rane, Nitin. "YOLO and Faster R-CNN object detection for smart Industry 4.0 and Industry 5.0: applications, challenges, and opportunities." Available at SSRN 4624206 (2023). https://dx.doi.org/10.2139/ssrn.4624206