Transactions in GIS

WILEY

Transactions in GIS

RESEARCH ARTICLE OPEN ACCESS

A Novel Multi-Layer Attention Boosted YOLOv10 Network for Landslide Mapping Using Remote Sensing Data

Naveen Chandra[1] | Himadri Vaidya[2,3] | Neelima Satyam[4] | Xiaochuan Tang[5] | Saurabh Singh[6] | Sansar Raj Meena[7,8]

[1]Geomorphology, Environment, and Engineering Geology, Wadia Institute of Himalayan Geology, Dehradun, Uttarakhand, India | [2]Computer Science and Engineering, Graphic Era Hill University, Dehradun, Uttarakhand, India | [3]Computer Science and Engineering, Graphic Era Deemed University, Dehradun, Uttarakhand, India | [4]Department of Civil Engineering, Indian Institute of Technology, Indore, India | [5]State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, College of Computer Science and Cyber Security, Chengdu University of Technology, Chengdu, China | [6]Remote Sensing and GIS Lab, Department of Geology, Institute of Science, Banaras Hindu University, Varanasi, Uttar Pradesh, India | [7]Machine Intelligence and Slope Stability Laboratory, Department of Geosciences, University of Padova, Padova, Italy | [8]Center for Remote Sensing, Department of Earth and Environment, Boston University, Boston, Massachusetts, USA

Correspondence: Sansar Raj Meena (sansarraj.meena@unipd.it)

Received: 20 February 2025 | **Revised:** 20 February 2025 | **Accepted:** 21 February 2025

Funding: This work was supported by Department of Science and Technology, Science and Engineering Research Board, New Delhi, India, EEQ/2022/000812.

Keywords: attention mechanism | Hazard | landslides | remote sensing | YOLO models

ABSTRACT

Detecting landslides is a critical challenge within the remote sensing fraternity, especially given the need for timely and accurate hazard assessment. Traditional methods for identifying landslides from remote sensing data are often manual or partially automated; however, with the progress of computer vision technology, the automated methods based on deep learning algorithms have gained significant attention. Furthermore, attention mechanisms, inspired by human visual structure, have grown remarkably in various applications, including hazard studies. In this study, we leverage the capabilities of YOLO models, especially YOLOv10 and its variants, to automate the detection of landslides. We applied four prevailing attention mechanisms: CBAM, ECA, GAM, and SA. Models are trained using the Bijie landslide detection database. Moreover, the best results are unveiled based on the evaluation criteria, that is, precision, recall, f-score, and mAP. The YOLOv10m+CBAM showed the best performance with map@50-95 of 78.5%. Our results demonstrate a robust system capable of rapidly identifying and localizing landslide events with significant detection speed and accuracy improvements. This advancement augments the process of landslide detection and supports more effective disaster response and management.

1 | Introduction

Landslides are pervasive geo-hazards triggered by seismic, climatic, and human-induced factors, posing a noteworthy risk to society and causing substantial loss of life, property/infrastructure, and economy (Ma and Mei 2021a; Zhang and Wang 2024; Du et al. 2024). The severity and frequency of landslide cases are likely to surge due to existing weather scenarios, urban growth, and rising population, especially in hilly regions globally (Ma and Mei 2021b; Casagli et al. 2023). Accordingly,

Abbreviations: AI, artificial intelligence; AP, average precision; C2FCIB, cross-stage partial connections with bottleneck; CBAM, convolution block attention module; CIB, compact inverted block; CNN, convolution neural networks; CSPNet, cross stage partial network; DEM, digital elevation model; DL, deep learning; ECA, efficient channel attention; FPN, feature pyramid network; GAM, global attention mechanism; GAP, global average pooling; GFLOPs, giga floating point operations per second; GMP, global max pooling; GPU, graphics processing unit; IoU, intersection over union; mAP, mean average precision; ML, machine learning; MLP, multi-layer perception; NMS, non-maximum suppression; PAN, path aggregation network; PSA, partial spatial attention; SA, shuffle attention; SC-Down, spatial-channel decoupled down-sampling; SGD, stochastic gradient descent; UAVs, unmanned aerial vehicles; YOLO, you only look once.

effective mitigation policies and risk reduction methods are crucial for human safety and infrastructure protection. Landslide risk management relies on three procedures (Casagli et al. 2023): comprehensive monitoring, prediction, and detection. Constant monitoring is crucial for forecasting landslides, while accurate detection is essential for post-hazard recovery efforts and maintaining updated inventories. Additionally, it supports mapping vulnerable locations, which is vital for sustainable planning and development. Hence, increasing emphasis is on rapidly and precisely detecting landslide events to enhance preparedness and response strategies.

The traditional method of landslide extraction, which depended upon on-site visits, was labor-intensive, time-consuming, and less efficient in emergency response conditions (Liu, Wei, et al. 2021; Malamud et al. 2004; Guzzetti et al. 2012). Nevertheless, innovations in remote sensing technology have revolutionized and have become essential in landslide investigation. These technological improvements enhance efficiency and effectiveness, allowing for quicker and more accurate assessments of landslide-prone areas.

Currently, four methods (Han et al. 2023) specifically employ visual analysis, pixel-based techniques, an object-oriented approach, and AI to use remote sensing-based data for landslide recognition. Visual analysis-based methods depend on expert knowledge but are time-consuming and labor-intensive. They are often inadequate for rapid response situations and may not meet the urgent demands of landslide emergency responses. Pixel-based methods address the limitations of the visual approach by using a binary method (Liu, Wei, et al. 2021) to categorize pixels within remote sensing imagery, such as landslides or backgrounds. However, these methods can struggle to differentiate objects that have the same spectral characteristics/features as landslides accurately, making correct classification challenging. Object-oriented techniques use multiscale segmentation, which analyzes many image primitives like shape, texture, and spectrum (Lu et al. 2011). These methods need experimental threshold settings to differentiate between landslide and non-landslide locations. While effective, they can struggle when exposed high resolution remote sensing data of large topographical regions, considering prompt segmentation. Keyport et al. (2018) (Keyport et al. 2018) proposed a significant work to evaluate the strengths and weaknesses of both methods for landslide detection.

On the contrary, AI-based models have seen significant developments across various domains (Hong et al. 2024), including hazard analysis (Stephen et al. 2024). This progress has been particularly valuable for landslide investigation, where the integration of AI with remote sensing data (Zhang et al. 2021; Hong et al. 2024) has proven to be transformative. Further, the development of remote sensing-based big data (Ma et al. 2015; Hong et al. 2024) has provided substantial help to AI in addressing complex challenges in landslide-related studies with greater accuracy and efficiency by offering large volumes of high-quality, diverse datasets necessary for training and improving AI models. Moreover, ML has demonstrated substantial enhancements in landslide assessment studies (Ma and Mei 2021b). ML-based algorithms provide powerful tools for analyzing large datasets and identifying patterns that can predict landslide occurrences (Ma and Mei 2021a). The ML-driven models have substantially enhanced the accuracy of various landslide mapping methods.

With the advent of powerful computational tools, especially GPUs (Iqbal et al. 2024), DL (Shrestha and Mahmood 2025), a subdivision of ML, has excelled in numerous areas such as image segmentation (Minaee et al. 2022), object detection (Zhao et al. 2019), and scene classification (Zeng et al. 2021). DL has also gained remarkable attention in the studies of geohazard analysis, especially landslides (Ma and Mei 2021a). In the past, different DL algorithms, mainly CNNs, with different architectures have been suggested for landslide event detection (Mohan et al. 2021; Chandra and Vaidya 2024b). Notably, U-Net (Devara et al. 2024; Chandra et al. 2023; Ghorbanzadeh et al. 2021) and Mask-RCNN (Ullo et al. 2021; Liu, Shao, et al. 2021). In recent times, YOLO (You Only Look Once) has been very popular in object detection tasks due to its speed and accuracy (Diwan et al. 2023). Variants of YOLO models, particularly YOLOv3 (Han et al. 2023), YOLOv4 (Li and Li 2022), YOLOv5 (Mo et al. 2023), YOLOv6 Chandra and Vaidya (2024a), YOLOv7 (Liu et al. 2023), and YOLOv8 (Mao et al. 2024) have been proposed for landslide detection. Subsequently, attention modules (Niu et al. 2021), encouraged by the human visualization process, have been assimilated into CNN to boost detection accuracy. For example, YOLO-SA (Cheng et al. 2021), 3D-SCAM (Ji et al. 2020), SW-MSA (Jia et al. 2024), and LA-YOLO-LLL (Yang et al. 2024).

Despite remarkable efforts from researchers, investigating a suitable DL-driven algorithm for landslide event detection remains a major challenge due to the complexity of remote sensing data and the diverse environmental factors affecting landslide appearance. Moreover, while attention mechanisms designed to improve model performance by emphasizing important features are promising, their integration and evaluation within algorithms like YOLO are still in the preliminary stages. This gap highlights the need for comprehensive research to assess the effectiveness of attention models in landslide detection. To address these limitations, this research shifts its focus to the latest YOLOv10 model. Therefore, this article aims to develop a more efficient and robust landslide detection network based on YOLOv10, considering the specific issues of scale variation and environmental complexity in remote sensing imagery. Addressing this gap is crucial for developing more robust, accurate, and efficient models capable of near real-time landslide detection, thereby advancing the field and improving disaster response strategies. Therefore, we propose an attention-guided YOLOv10 network to infer its capabilities and effectiveness concerning landslide event detection from Earth observational images. The key contributions of this research work are as follows:

1. To develop an improved and optimized YOLOv10 network specifically designed for advanced landslide mapping.
2. To significantly enhance the network's capability to capture and represent critical features from input imagery by integrating attention modules specifically, CBAM, ECA, GAM, and SA discretely within the neck of the YOLOv10 architecture.
3. To conduct a thorough performance evaluation and assessment using a benchmark database to confirm that the

FIGURE 1 | Examples of the Bijie landslide detection data set.

results are comparable with other prevailing methods in studies of landslide detection.

The remaining sections of the paper are organized as follows: Section 2 describes the data set. Section 3 demonstrates the methodology. Section 4 presents the results, and Section 5 comprises a discussion. Lastly, Section 6 includes the conclusion.

2 | Data Set Description

For this research work, we employed the Bijie landslide detection database (Ji et al. 2020), a popular dataset developed to support the research on landslide detection using remote sensing images. It covers the entire Bijie town in Guizhou province (China). The total coverage area is 26,853 km^2. The images are acquired from the TripleSat (between May and August 2018). The dataset includes high-resolution imagery with a ground resolution of 0.8 m. It features 770 annotated landslide samples (rock slides, debris slides, and rock fall), and 2003 negative instances representing various backgrounds to balance training and evaluation. The size of landslides in each sample is different. Moreover, non-landslide samples comprise mountains, villages, rivers, roads, agricultural land, and forest. The dataset also provides references/ground truth images for accurate reference and DEMs with a 2-m accuracy, offering critical topographic data like slope, aspect, and elevation. The example of the Bijie landslide detection database is shown in Figure 1. The Bijie landslide detection dataset, covering an extensive area offers a diverse and challenging set of environmental conditions ideal for testing the robustness of our model. This large area includes varied topographies, and other complex landscapes, are crucial for landslide detection studies. The dataset's high spatial resolution ensures that even small-scale landslide events can be accurately detected, making it highly suitable for our research. The size of the study area allows the model to learn from a wide range of geomorphological features and real-world scenarios, which improves the model's generalizability. This dataset not only provides ample examples of landslides but also includes challenging negative samples, which are necessary for developing a model that can perform well in different regions. This broad geographic coverage ensures that the results from this dataset can be extended to other landslide-prone areas, including the Himalayan region, making it an appropriate and comprehensive dataset for this study.

3 | Methodology

This section illustrates the YOLOv10 (baseline) model and describes the proposed method. Moreover, details of the training setting and evaluation criteria are provided.

3.1 | Background of Original YOLOv10 Model

The YOLOv10 (Wang et al. 2024) network builds upon the robust and efficient architecture of its previous variants while incorporating several advanced features to improve performance in object detection tasks. The key variants of YOLOv10 include

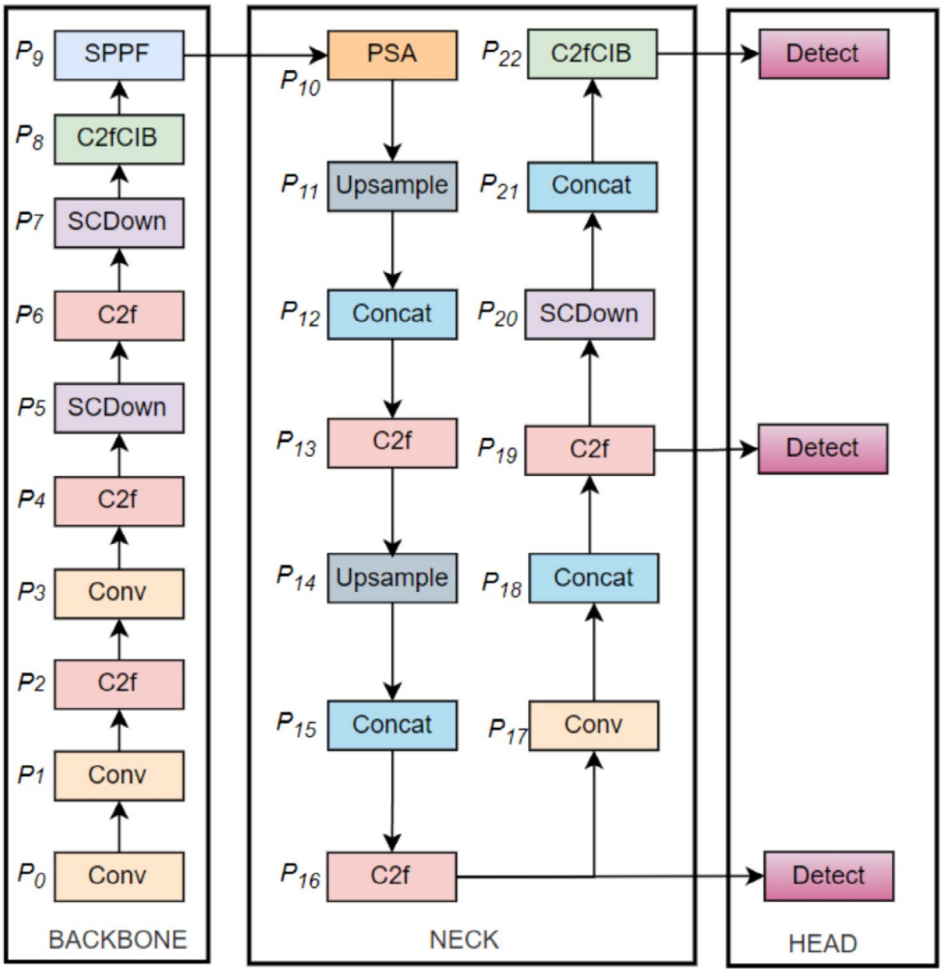

FIGURE 2 | Architecture of the original YOLOv10 model.

YOLOv10n (nano), YOLOv10s (small), YOLOv10m (medium), YOLOv10b (balanced), YOLOv10l (large), and YOLOv10xl (extra-large), each optimized for specific use cases and computational constraints. The architecture comprises the backbone, followed by the Neck, and the head module. Figure 2 represents the architecture of the YOLOv10 network. The backbone of YOLOv10 extracts the significant features from the input imagery. The backbone is a CNN that processes the image through a series of layers to generate feature maps of different resolutions. YOLOv10 includes a modified version of the CSPNet Wang, Mark Liao, et al. (2020) as its backbone to enhance gradient flow and reduce computational complexity. CSPNet splits the feature map into two fragments and combines them through a cross-stage structure, which improves learning and reduces the model's size. Multiple convolutional layers with varying kernel sizes, strides, and padding are used to capture spatial hierarchies and patterns within the image. Residual connections help to resolve the vanishing gradient problem and facilitate the training of deeper architectures.

The Neck module of the YOLOv10 network is created to combine features from different scales and improve the feature representation. The feature pyramid network (FPN) merges feature maps from different phases of the backbone to generate a more composite representation Lin et al. (2017). The path aggregation network (PAN) further enhances feature maps by providing improved information flow between different levels of the feature pyramid, refining the localization and classification of objects Liu et al. (2018). The FPN and PAN structures are important in YOLOv10, as they combine features from different layers to produce a multi-scale feature representation. This supports the detection of small and large objects, improving detection accuracy.

The Head of YOLOv10 makes the predictions. It takes the combined features from the neck and performs landslide detection, which comprises bounding box regression, object classification, and confidence scoring.

NMS-free training in the YOLOv10 algorithm is advantageous because it removes the need for NMS during inference, which traditionally filters out redundant predictions but adds computational overhead (Wang et al. 2024). YOLOv10 presents dual label assignments, combining one-to-many assignments for rich supervision during training and one-to-one matching for efficient inference, resulting in faster, end-to-end predictions without performance loss. This process enhances real-time object detection capabilities by reducing latency and maintaining accuracy, offering an improvement over traditional NMS-based techniques, which are slower due to their reliance on post-processing (Wang et al. 2024). Additionally, YOLOv10 includes large-kernel convolutions and PSA modules to expand performance by capturing broader

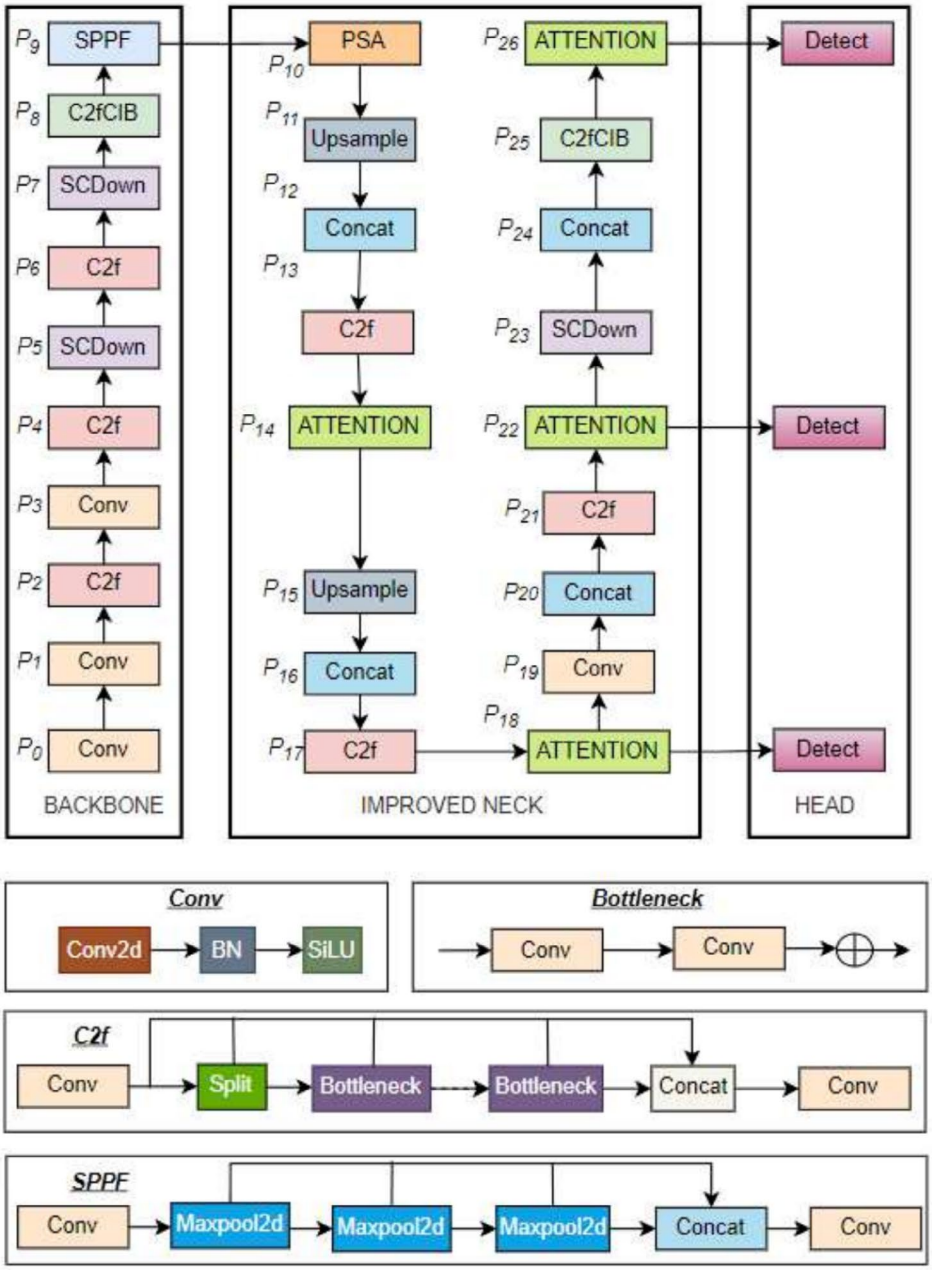

FIGURE 3 | Structure of the proposed model.

context and complex patterns within the data, all without increasing computational cost. These improvements make YOLOv10 an influential and competent model for real-time object detection in complex settings, such as landslide event extraction.

3.2 | Proposed Multi-Layer Attention Based YOLOv10 Network

The proposed YOLOv10 integrates attention modules discretely at several locations inside the neck of the network to augment feature representation and enhance detection accuracy. The architecture of the proposed model is shown in Figure 3. Explicitly, attention modules are introduced at four locations: First, between the C2f Module (P13th layer) and Up Sample (P15th layer): The C2f module is intended to capture features with different receptive fields. It sums multi-scale features utilizing a combination of convolutions and feature fusion. The Upsample process surges the spatial resolution of the feature map, making it appropriate for detecting smaller objects. Integrating an attention module between these two components allows the network to refine the feature map before it is upsampled. This ensures that the upsampling process is applied to more relevant features, leading to better preservation of critical information needed for detecting landslides.

Second, between the C2F module (P17th layer and the Convolution Layer) (P19th layer), by adding an attention module here, the network can concentrate on the most relevant features learned by the C2f module before further processing them through convolutions. This confirms that the subsequent convolutions operate on the most important features, boosting the detection accuracy for complex backgrounds.

Third, between the C2f Module (P21th layer and SC Down layer) (P23th layer), the C2f module continues to extract and integrate multi-scale features. Further, SC Down Layer reduces the spatial dimensions while attempting to maintain as much informative content as possible.

Inserting an attention module here helps in emphasizing critical features before the down-sampling process. This ensures that the reduction in spatial dimensions does not lead to the loss of important information necessary for landslide detection.

Fourth, after the C2FCIB Layer, at the P26th layer: It combines features from different stages and compresses them to enhance the representational capacity of the network. By adding an attention layer after the C2FCIB layer, the network gains the ability to further refine these combined features before the final detection stage. The C2FCIB layer merges features from different stages, which might include some irrelevant or redundant information. The architecture of CIB and PSA is shown in Figure 4. The CIB in YOLOv10 is proposed to improve feature extraction by enabling interactions between different spatial regions in the feature maps, while the PSA module augments pixel-level feature refinement by fine-tuning attention weights across the spatial domain. The attention layer helps in filtering out these unnecessary details and highlights the most critical aspects of the feature map. Just before the detection stage, the feature map must be as relevant and focused as possible. The attention mechanism ensures that the final features used for detection are precisely those that contribute most to identifying landslides. By refining the features at this late stage, the attention layer improves the network's ability to correctly identify and localize landslides, especially in complex and diverse environmental conditions typical of remote sensing images.

3.3 | Adddition of CBAM Layer

The CBAM (Woo et al. 2018) refines feature maps by applying both channel and spatial attention mechanisms. Let $F \in \mathbb{R}^{C\times H\times W}$ represent the input feature map, where C is the number of channels, and H and W are the spatial dimensions.

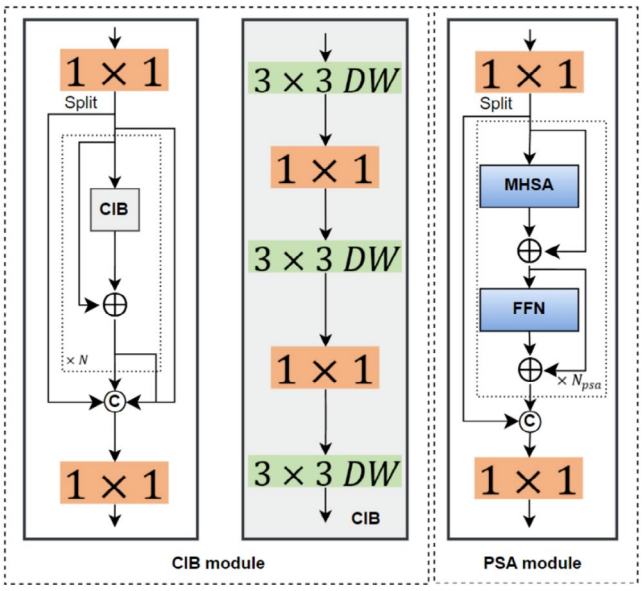

FIGURE 4 | Description of CIB and PSA module.

The CAM focuses on the importance of different channels. Given F, the channel attention M_c is computed through GAP and GMP operations.

$$F_{\text{avg}} = \text{GAP}(F) \quad \text{and} \quad F_{\max} = \text{GMP}(F) \tag{1}$$

where both F_{avg} and $F_{\max}$ are in $\mathbb{R}^{C\times 1\times 1}$. The pooled features are computed by a shared MLP:

$$MLP(F_{\text{avg}}) = W_1\sigma(W_0F_{\text{avg}}) \quad \text{and} \quad MLP(F_{\max}) = W_1\sigma(W_0F_{\max}) \tag{2}$$

where W_0 and W_1 are the weights of the MLP, and σ denotes the ReLU activation function.

Moreover, the integration and activation, followed by rescaling, result in the channel attention mechanism.

$$M_c = \sigma(\text{MLP}(F_{\text{avg}}) + \text{MLP}(F_{\max})) \tag{3}$$

$$F' = M_c \cdot F \tag{4}$$

The spatial attention mechanism focuses on the significant spatial regions within the feature map F'. The spatial attention M_s is processed by computing average as well as max pooling along the channel axis:

$$F^c_{\text{avg}} = \text{AvgPool}(F') \quad \text{and} \quad F^c_{\max} = \text{MaxPool}(F') \tag{5}$$

These are concatenated:

$$F_{\text{concat}} = \left[F^c_{\text{avg}}; F^c_{\max}\right] \tag{6}$$

further, we apply a 7×7 convolution to the concatenated feature map before rescaling

$$M_s = \sigma(\text{Conv}(F_{\text{concat}})) \tag{7}$$

$$F'' = M_s \cdot F' \tag{8}$$

3.4 | Addition of ECA Layer

ECA (Wang, Wu, et al. 2020) module improves channel-wise attention without increasing the computational cost. Below, we demonstrate the mathematical operations involved in ECA.

First, we apply global average pooling to the input feature map to produce a channel-wise descriptor:

$$F_{\text{avg}} = \text{GAP}(F) \tag{9}$$

where $F_{\text{avg}} \in \mathbb{R}^{C\times 1\times 1}$. Instead of employing a multi-layer perceptron (MLP), ECA utilizes a 1D convolution to capture channel-wise dependencies. Let k be the kernel size of the convolution. Apply a $1 \times 1 \times k$ convolution to F_{avg}:

$$E_{\text{ca}} = \text{Conv}_{1D}(F_{\text{avg}}) \tag{10}$$

where $E_{ca} \in \mathbb{R}^{C\times1\times1}$. The outcomes of the convolution is passed through a sigmoid activation function to produce the final channel attention map:

$$M_c = \sigma(E_{ca}) \tag{11}$$

where σ corresponds to the sigmoid function. Finally, rescale the original feature map F by the channel attention map M_c:

$$F' = M_c \cdot F \tag{12}$$

where $F' \in \mathbb{R}^{C\times H\times W}$

3.5 | Addition of GAM Layer

It computes both channel and spatial attention mechanisms to refine features effectively (Liu, Shao, et al. 2021). The overall process is as follows: initially, we perform the GAP to aggregate spatial information into a channel descriptor:

$$F_{avg} = \text{GAP}(F) \tag{13}$$

where $F_{avg} \in \mathbb{R}^{C\times1\times1}$. Further, we generate channel attention weights by applying a fully connected layer followed by a sigmoid activation:

$$M_c = \sigma(W_c F_{avg}) \tag{14}$$

where W_c is the weight matrix of the fully connected layer and σ denotes the sigmoid function. The result $M_c \in \mathbb{R}^{C\times1\times1}$ represents the channel attention map. Lastly, we rescale the input feature map F with the channel attention map M_c:

$$F' = M_c \cdot F \tag{15}$$

3.6 | Addition of SA Layer

The SA (Zhang and Yang 2021) mechanism consists of three primary steps: First, spatial attention: To aggregate spatial information, we perform the GAP operation along the channel dimension.

$$F_{avg} = \text{GAP}(F) \tag{16}$$

where $F_{avg} \in \mathbb{R}^{1\times H\times W}$. To aggregate spatial information we conduct GAP operation along the channel dimension

$$F_{max} = \text{GMP}(F) \tag{17}$$

where $F_{max} \in \mathbb{R}^{1\times H\times W}$. Now we concatenate the pooled feature maps

$$F_{concat} = \text{Concat}(F_{avg}, F_{max}) \tag{18}$$

where $F_{concat} \in \mathbb{R}^{2\times H\times W}$. Furthermore, we apply a convolutional layer followed by a sigmoid activation to compute spatial attention weights:

$$M_s = \sigma(\text{Conv}(F_{concat})) \tag{19}$$

where Conv denotes a convolutional layer with a 7×7 kernel, and $M_s \in \mathbb{R}^{1\times H\times W}$ is the spatial attention map. Rescale the input feature map F with the spatial attention map M_s:

$$F' = M_s \cdot F \tag{20}$$

Second, channel attention: We perform GAP along the spatial dimensions to aggregate information into a channel descriptor:

$$F^c_{avg} = \text{GAP}(F') \tag{21}$$

where $F^c_{avg} \in \mathbb{R}^{C\times1\times1}$.

Moreover, implementing GMP along the spatial dimensions to aggregate information into a channel descriptor:

$$F^c_{max} = \text{GMP}(F') \tag{22}$$

where $F^c_{max} \in \mathbb{R}^{C\times1\times1}$. Now concatenate and transform the feature map Concatenate F^c_{avg} and F^c_{max} and apply a shared 1×1 convolution followed by a sigmoid activation:

$$F^c_{cat} = \text{Concat}(F^c_{avg}, F^c_{max}) \tag{23}$$

$$M_c = \sigma(\text{Conv}(F^c_{cat})) \tag{24}$$

where σ implies the sigmoid function and $M_c \in \mathbb{R}^{C\times1\times1}$ is the channel attention map. Finally, rescale the spatially attended feature map F' with the channel attention map M_c:

$$F'' = M_c \cdot F' \tag{25}$$

Third, channel shuffling: Reshape the feature map F'' to group channels, then shuffle the grouped channels to enhance inter-channel information flow:

$$F_{shuffled} = \text{Shuffle}(F'') \tag{26}$$

where $F_{shuffled} \in \mathbb{R}^{C\times H\times W}$ is the outcome of the SA module. It is the refined feature map $F_{shuffled}$ that incorporates both spatial and channel attention mechanisms, with enhanced inter-channel information flow due to the shuffling operation.

The detection layers predict bounding-boxes, objectness, and class objectness scores, and class probabilities for each anchor box. The bounding-box coordinates (b_m, b_n) and dimensions (b_w, b_h) are predicted as follows:

$$b_m = \sigma(t_m) + c_m \tag{27}$$

$$b_n = \sigma(t_n) + c_n \tag{28}$$

$$b_w = p_w e^{t_w} \tag{29}$$

$$b_h = p_h e^{t_h} \tag{30}$$

where t_m, t_n, t_w, t_h denotes the predicted offsets; c_m, c_n signifies the center coordinates of the anchor box; p_w, p_h represents the prior width and height of the anchor box; and σ is the sigmoid

function. The objectness score $\mathrm{Pr}_{\mathrm{object}}$ indicates the confidence that an object is present in the bounding box:

$$\mathrm{Pr}_{\mathrm{object}} = \sigma\left(t_{\mathrm{object}}\right) \tag{31}$$

where t_{object} is the raw objectness score output by the network. The class probabilities $\mathrm{Pr}_{\mathrm{class}}$ for each class are predicted using the softmax function:

$$\mathrm{Pr}_{\mathrm{class}} = \mathrm{softmax}\left(t_{\mathrm{class}}\right) \tag{32}$$

where t_{class} is the raw class scores output by the network.

3.7 | Loss Function Estimation

First, the localization loss (bounding box regression) estimates the accuracy of the predicted bounding boxes against the ground truth boxes. It comprises coordinates loss and dimensions loss.

$$\mathscr{L}_{\mathrm{loc}} = \lambda_{\mathrm{coord}} \sum_{i=0}^{G^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\mathrm{obj}} \left[\left(m_i - \hat{m}_i\right)^2 + \left(n_i - \hat{n}_i\right)^2 + \left(w_i - \hat{w}_i\right)^2 + \left(h_i - \hat{h}_i\right)^2 \right] \tag{33}$$

In the above equation, λ_{coord} is a weight parameter; $\mathbb{1}_{ij}^{\mathrm{obj}}$ indicates if the j-th bounding box in cell i is responsible for the object; and G is the grid size. Second, the confidence loss evaluates the model's confidence concerning the existence of target in a predicted bounding box.

$$\mathscr{L}_{\mathrm{conf}} = \lambda_{\mathrm{obj}} \sum_{i=0}^{G^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\mathrm{obj}} \left(C_i - \hat{C}_i\right)^2 + \lambda_{\mathrm{noobj}} \sum_{i=0}^{G^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\mathrm{noobj}} \left(C_i - \hat{C}_i\right)^2 \tag{34}$$

where, λ_{obj} and λ_{noobj} are weight parameters for the object and no-object confidence loss, respectively; C_i indicates the predicted confidence score, and $\hat{C}_i$ shows the ground truth confidence score. Third, the classification loss measures the accuracy of the predicted class probabilities against the ground truth class labels for each cell that contains an object.

$$\mathscr{L}_{\mathrm{cls}} = \lambda_{\mathrm{class}} \sum_{i=0}^{G^2} \mathbb{1}_{i}^{\mathrm{obj}} \sum_{c \in \mathrm{classes}} \left(\mathrm{pr}_i(c) - \hat{\mathrm{p}}\mathrm{r}_i(c)\right)^2 \tag{35}$$

where λ_{class} represent a weight parameter, $\mathrm{pr}_i(c)$ signifies the predicted probability of class c for cell i, and $\hat{\mathrm{p}}\mathrm{r}_i(c)$ denotes the ground truth probability. The total loss function in YOLOv10 encompasses these three components:

$$\mathscr{L}_{\mathrm{total}} = \mathscr{L}_{\mathrm{loc}} + \mathscr{L}_{\mathrm{conf}} + \mathscr{L}_{\mathrm{cls}} \tag{36}$$

3.8 | Training Settings and Machine Specifications

The proposed model is trained with the following hyperparameters: image size = 640 × 640, batch size = 16, epochs = 500, initial momentum factor = 0.937, initial learning rate = 0.01, optimizer = SGD, and initial weight decay = 0.0005. The experiments are conducted on a machine with the following specifications: Python-3.10.12, torch-2.0.1+cu117, CUDA:0 (NVIDIA GeForce RTX 4090, 24207MiB), CUDA:1 (NVIDIA GeForce RTX 4090, 24210MiB).

3.9 | Evaluation Criteria

For the quantitative evaluation, four standard metrics are computed: precision, recall, F-score, and mAP (Cheng et al. 2021; Mao et al. 2024; Liu et al. 2023). These metrics are estimated as follows:

$$\mathrm{Precision} = \frac{TP}{\mathrm{TP} + \mathrm{FP}} \tag{37}$$

where TP corresponds to the number of true positives, and FP signifies the number of false positives.

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{38}$$

where FN denotes the number of false negatives.

$$\mathrm{F\text{-}score} = 2 \cdot \frac{\mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{39}$$

mAP evaluates the overall performance of the detection model across multiple classes and thresholds. It is computed by averaging the average precision (AP) for each class defined below.

$$\mathrm{mAP} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{AP}_i \tag{40}$$

where N is the number of classes, and AP_i is the average precision for the i-th class.

These metrics ensure that both the correctness and completeness of detections are considered.

4 | Results

This section contains the quantitative and qualitative findings of the proposed study.

4.1 | Quantitative Results

The quantitative results of the proposed YOLOv10+attention and baseline models are given in Table 1. Notably, the YOLOv10l+CBAM model shows a remarkable increase in mAP@50-95 by 10.3% compared to the original YOLOv10l network. Similarly, adding GAM to YOLOv10l improves performance by 9.4%, while the YOLOv10l+SA and YOLOv10l+ECA models show improvements of 8.2% and 7.9%, respectively. Furthermore, the inclusion of attention layers significantly boosts the mAP@50-95 of YOLOv10b, with YOLOv10b+GAM, YOLOv10b+SA, and YOLOv10b+CBAM achieving improvements of 4.3%, 4.2%, and 3.4%, respectively, and no improvements in YOLOv10b+ECA. For the YOLOv10x model, the

TABLE 1 | Experimental results of the proposed models.

Models	GFLOP	Params	P	R	mAP@50	mAP@50–95	Time	GPU
YOLOv10n	8.2	26.94	95.4	97.1	98.8	75.2	0.359	2.91
YOLOv10n+GAM	9.6	33.76	97.1	90.6	98.5	75.5	2.622	1.77
YOLOv10n+SA	8.2	26.95	96.4	97.2	98.9	75.8	0.421	1.64
YOLOv10n+ECA	8.2	26.95	99.0	97.2	99.2	76.1	0.443	2.90
YOLOv10n+CBAM	8.2	27.97	93.6	96.2	98.7	76.2	0.281	1.69
YOLOv10m	63.4	164.5	95.4	98.8	98.8	74.9	1.027	8.62
YOLOv10m+ECA	8.2	269.5	98.1	96.9	98.9	76.2	1.205	2.93
YOLOv10m+GAM	77.9	219.1	97.1	93.7	98.6	76.6	2.971	5.66
YOLOv10m+SA	80.3	175.1	96.3	97.8	99.3	76.9	0.537	2.94
YOLOv10m+CBAM	66.7	181.7	98.0	94.1	99.1	**78.5**	0.717	4.67
YOLOv10b	97.9	204.1	91.9	96.0	98.2	73.1	3.01	6.11
YOLOv10b+CBAM	97.9	212.66	94.4	95.0	98.6	76.5	0.778	6.18
YOLOv10b+GAM	127.7	297.3	92.8	97.8	98.4	77.4	3.482	7.62
YOLOv10b+ECA	97.9	204.1	94.0	88.8	97.6	72.9	0.806	11.9
YOLOv10b+SA	109.6	240.8	95.7	95.3	98.7	77.3	0.971	6.23
YOLOv10l	126.3	257.1	83.9	92.5	95.3	67.7	2.392	7.24
YOLOv10l+CBAM	143.8	320.7	92.3	97.2	98.6	78.0	1.030	6.80
YOLOv10l+GAM	161.9	368.7	96.8	94.3	98.8	77.1	3.228	8.68
YOLOv10l+ECA	143.8	312.2	99.0	98.1	99.4	75.6	1.393	12.7
YOLOv10l+SA	143.8	312.2	98.6	89.6	98.5	75.9	0.913	7.0
YOLOv10x	169.8	315.8	93.5	94.9	97.9	72.6	2.928	10.3
YOLOv10x+CBAM	197.2	415.1	94.5	97.3	99.0	77.0	1.775	9.94
YOLOv10x+GAM	225.4	490.1	94.0	89.1	97.7	74.0	3.664	11.04
YOLOv10x+ECA	197.2	401.8	95.3	95.2	98.3	73.7	2.286	18.1
YOLOv10x+SA	197.1	401.8	98.1	95.7	98.4	75.1	1.989	10.4

Note: Bold letter indicates this model performs best.

enhancements are 4.4% for YOLOv10x+CBAM, 1.4% for YOLOv10x+SA, 1.1% for YOLOv10x+GAM, and 2.5% for YOLOv10x+ECA. Similarly, the YOLOv10m model shows the best performance with CBAM, which improves accuracy by 3.6%. The performance of YOLOv10m+SA, YOLOv10m+GAM, and YOLOv10m+ECA increases by 2.0%, 1.7%, and 1.3%, respectively. Additionally, YOLOv10n+CBAM shows an improvement of 1.0%, with marginal gains noted for YOLOv10n+ECA (0.9%), YOLOv10n+SA (0.6%), and YOLOv10n+GAM (0.3%). The observed improvements in mAP@50-95 indicate that the attention mechanisms significantly upgrade the model's capability to accurately detect landslides, making them valuable additions to the YOLOv10 framework.

Furthermore, the highest mAP@50 computed is 99.4% (YOLOv10l+ECA), which represents an improvement of 4.1% over the baseline model. Following this, YOLOv10x+CBAM demonstrated progress of 1.1%, achieving an mAP@50 of 99.0%. Additionally, YOLOv10m+SA, YOLOv10n+ECA, and YOLOv10m+CBAM also estimated high mAP@50 scores of 99.3%, 99.2%, and 99.1%, respectively. The remaining variants of YOLOv10 combined with attention mechanisms (13 models in total) also achieved competitive mAP@50 scores of 98%. Even the YOLOv10x + GAM and YOLOv10b + ECA models, with mAP@50 scores of 97.7% and 97.6%, respectively, demonstrated nearly similar performance.

The mAP@50 metric at an IoU threshold of 50% is crucial for evaluating the accuracy of the proposed network. It indicates the model's ability to identify and localize landslides within the images correctly. The high mAP@50 values indicate that the models accurately detect objects at a 50% overlap. This level of precision is significant as it ensures that the detected objects closely match the actual objects, which is essential for applications that require high accuracy, such as landslide detection from satellite images. The observed improvements and high scores demonstrate the effectiveness of integrating attention mechanisms with the YOLOv10 models, enhancing their

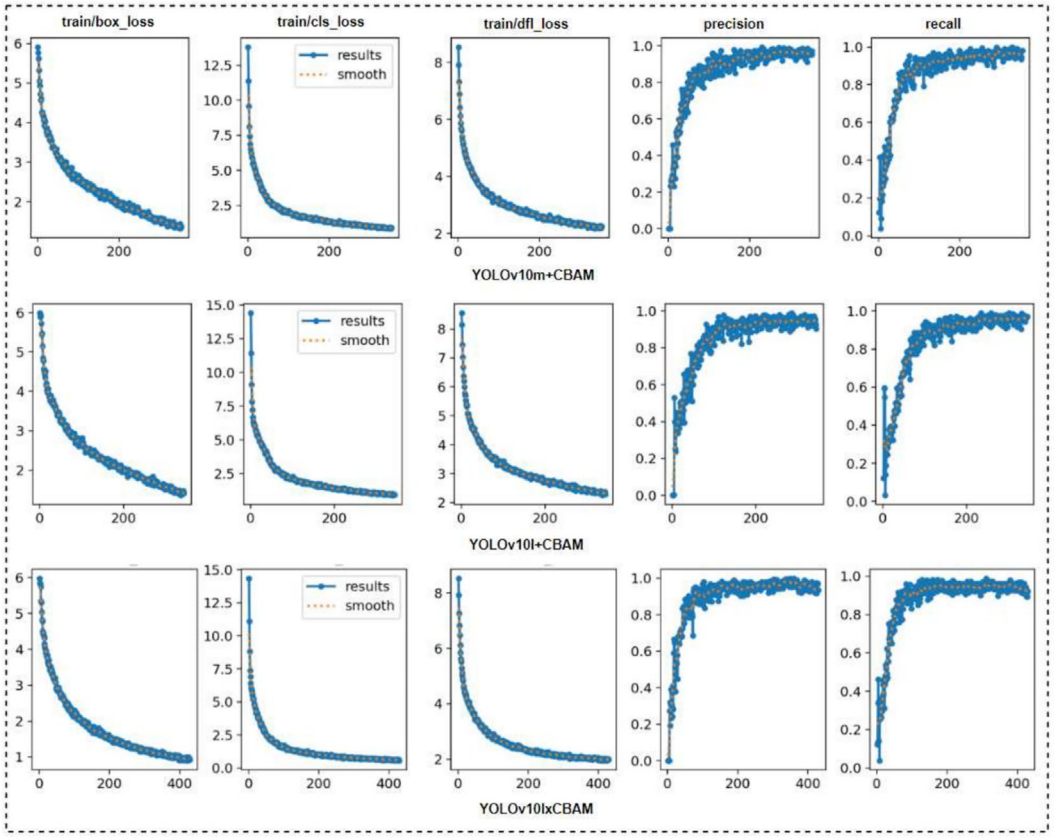

FIGURE 5 | Training loss function.

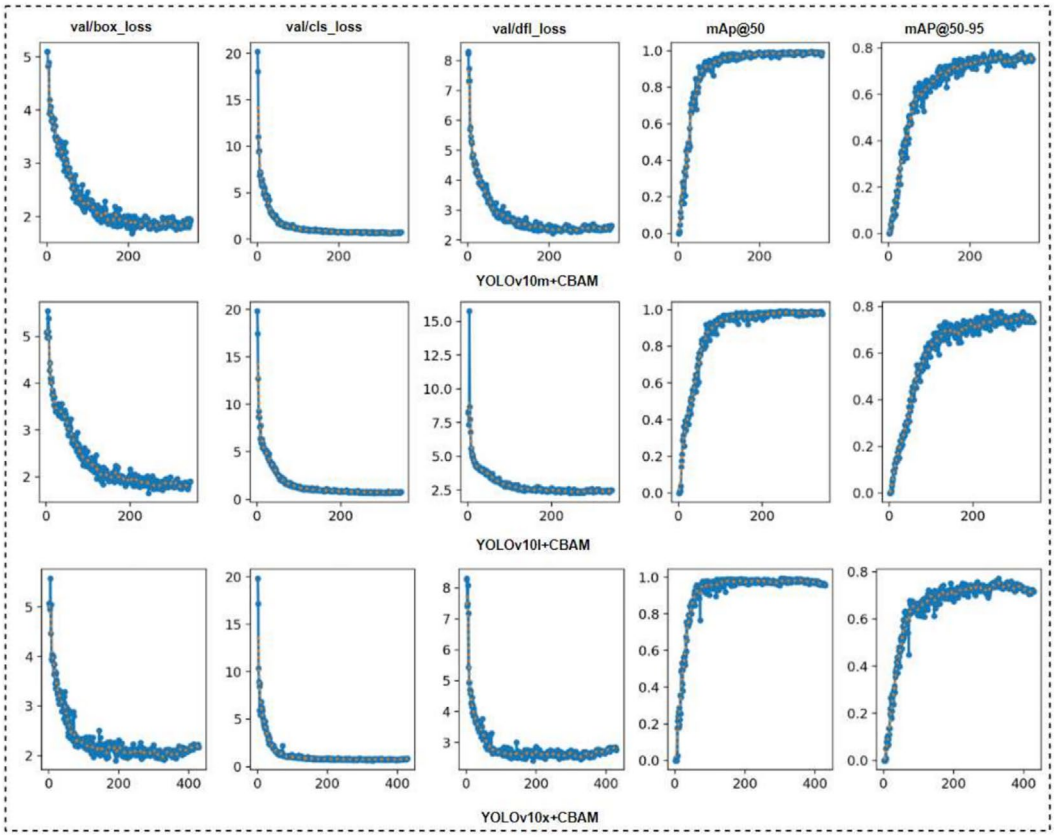

FIGURE 6 | Validation loss function.

detection capabilities and making them reliable for practical use in critical tasks.

Precision is a critical metric for evaluating the model performance, reflecting the proportion of true positive detection among all positive detection completed by the model. High precision shows that the model makes fewer false positive errors, which is essential in applications where the cost of false positives is high. In the context of landslide detection from satellite images, high precision ensures that detected landslides are actual landslides, minimizing the risk of false alarms. The precision reached 99.0% for both YOLOv10n+ECA and YOLOv10l+ECA models. The YOLOv10l+attention models exhibited the most significant improvement in precision, ranging from 8.7% to 15.1%. For YOLOv10x+attention models, precision improvements ranged from 0.5% to 4.6%. The precision of YOLOv10n+attention models increased by 1.0% to 3.6%. However, improvements for YOLOv10b and YOLOv10m models with attention layers ranged from 0.9% to 3.8% and 0.9% to 2.7%, respectively, with no improvement observed for YOLOv10n+CBAM. The improvements in precision with the addition of attention mechanisms highlight their significance in enhancing the model's reliability and accuracy, making them more effective for landslide detection. In the context of landslide detection from satellite images, high precision ensures that detected landslides are actual landslides, minimizing the risk of false alarms. This is predominantly valued in hazard assessment and disaster management, where false alarms could lead to inefficient resource allocation. The improvements in precision with the addition of attention mechanisms highlight their significance in enhancing the model's reliability and accuracy, making them more effective and operational for real-world applications, especially landslide detection. In addition, recall quantifies the model's ability to detect all important instances within the dataset. The maximum and minimum recall are computed for YOLOv10l+ECA (98.1%), and YOLOv10b+ECA (88.8%) respectively. The highest improvement (1.8%–5.6%) is observed for the YOLOv10l+attention model, demonstrating the effectiveness of incorporating attention mechanisms in enhancing the model's sensitivity.

Moreover, we estimated two factors to infer the efficiency and performance of the models. The GFLOP metric indicates the computational complexity of the model. It is estimated between

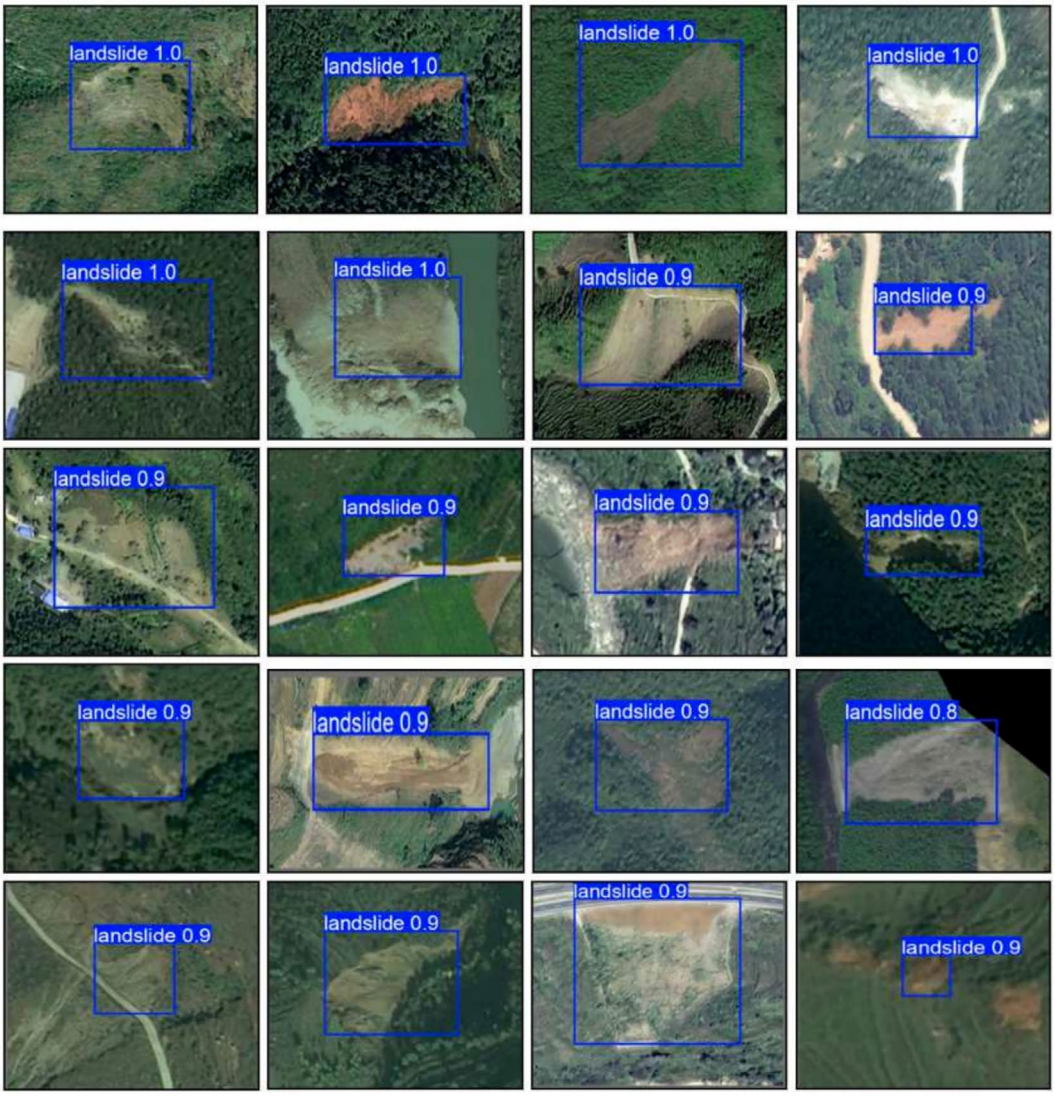

FIGURE 7 | Results of the detected landslides by the proposed network YOLOv10m+CBAM model.

8.2 (YOLO10n) and 225.4 (YOLO10x+GAM). A lower GFLOP count implies that the model is more efficient and can process data faster, which is vital for real-time applications like disaster response where timely detection is crucial. The number of parameters in a model reflects its size and complexity. Our experiments ranged between 26.94 (YOLOv10n) and 490.1 (YOLO10x+GAM). Models with fewer parameters require less memory and computational resources, making them more appropriate for deployment on systems with limited hardware capabilities. Additionally, to understand the resource utilization capability of the proposed model, we estimated the computational time of each model. This measures how long it takes to train and infer using the model (0.281 h (YOLOv10n+CBAM) to 3.664 h (YOLOv10x+GAM)). Faster models enable quicker iterations and experimentation during the development phase. Additionally, shorter inference times are essential for real-time monitoring and forecasting models. Moreover, GPU Utilization signifies the memory used by each model (1.64GBs (YOLOv10n+SA) to 11.04 GBs (YOLOv10x+GAM)), which can significantly affect training times and model performance. High-end GPUs can handle more complex models and larger datasets, reducing training time and enabling more sophisticated models to be used. The

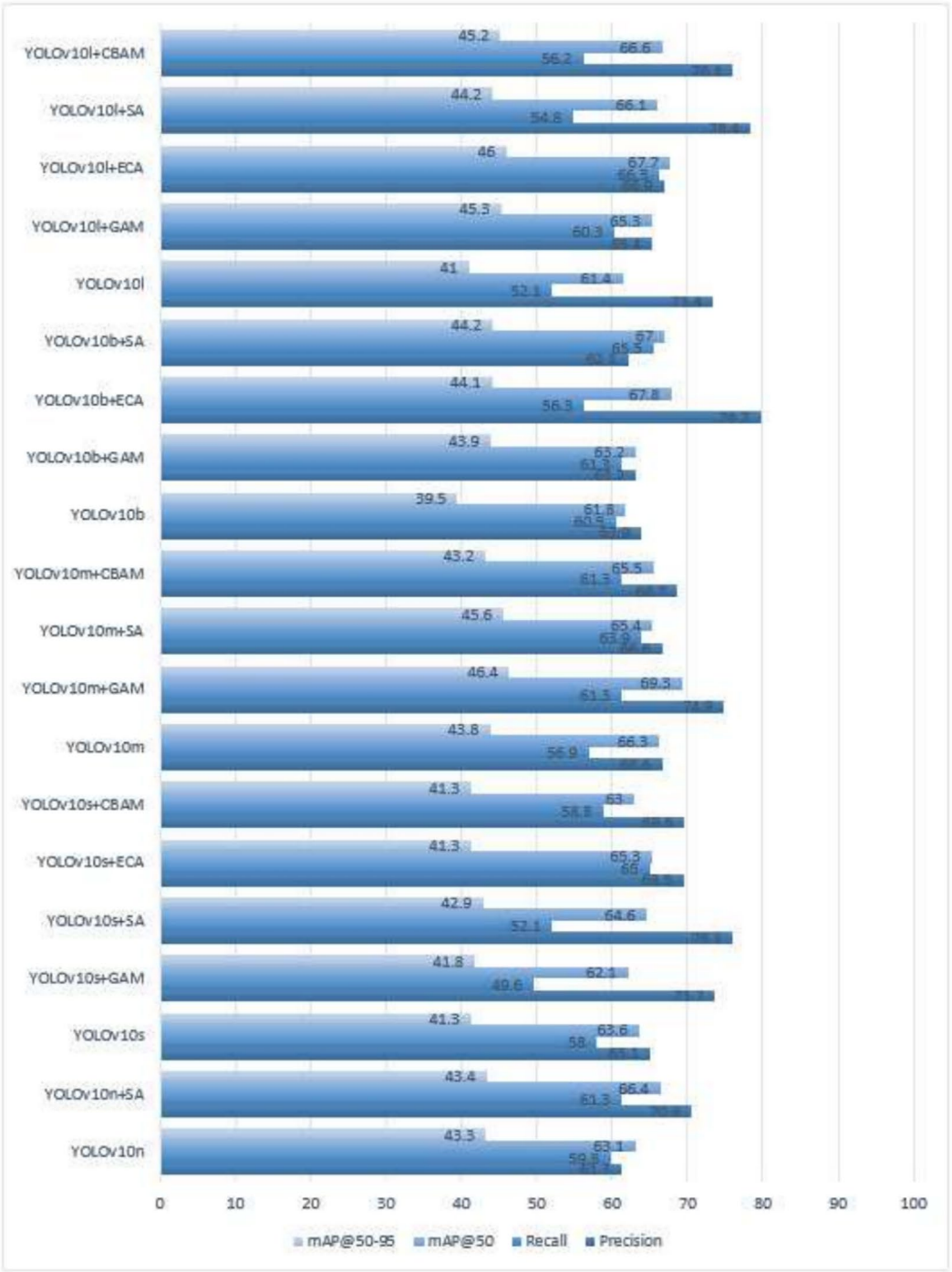

FIGURE 8 | Results of the Kodagu district, Western Ghats, India.

computational complexities of each model are shown in Table 1. The estimated training loss and the validation loss functions are represented in Figures 5 and 6, respectively.

4.2 | Qualitative Results

A qualitative assessment of the obtained results is required to identify the performance of the model beyond numerical estimation. It provides an understanding of the model's capability in real-world scenarios and highlights the possibilities for improvement. It represents a complete interpretation of model performance and ability, including visual and contextual accuracy. The qualitative estimation of extracted landslides encompasses a visual analysis of the model's output to measure the correctness and reliability of the forecasts. This procedure includes investigating the detected landslides in the context of the original input images and comparing the predicted bounding boxes with the real features in the images. We evaluate whether the bounding boxes generated by the model correctly enclose the entire landslide area without missing substantial features. Figure 7 represents the detected landslides by the YOLOv10m + CBAM model. The images involve complex environmental features like road networks, water bodies, and vegetation, though the model extracts the landslides effectively. These elements can obstruct the landslide detection process, as they may resemble landslide areas in certain visual characteristics. The qualitative evaluation supports an understanding of how the model deals with these complexities and whether it effectively distinguishes between landslide-prone areas and other landforms. It is seen that our model can detect both large and small landslides correctly. This detection ability of our model is a significant advantage, particularly in the context of landslide hazard assessment. In real-world scenarios, landslides vary greatly in size, and the consequences of failing to detect smaller landslides can be very critical.

Small landslides, though less immediately destructive, can often serve as precursors to larger events. Hence, detecting these smaller features allows for proactive measures before they intensify into major hazards. Visual inspection of the model's predictions represents that, despite varying landslide sizes, the model maintains consistent accuracy, identifying landslide areas in challenging environments. Moreover, if the bounding box does not encompass the entire affected area, it may lead to an incorrect assessment of the landslide's extent, which could affect critical decision-making in disaster response and management. In some of our results, the predicted bounding boxes

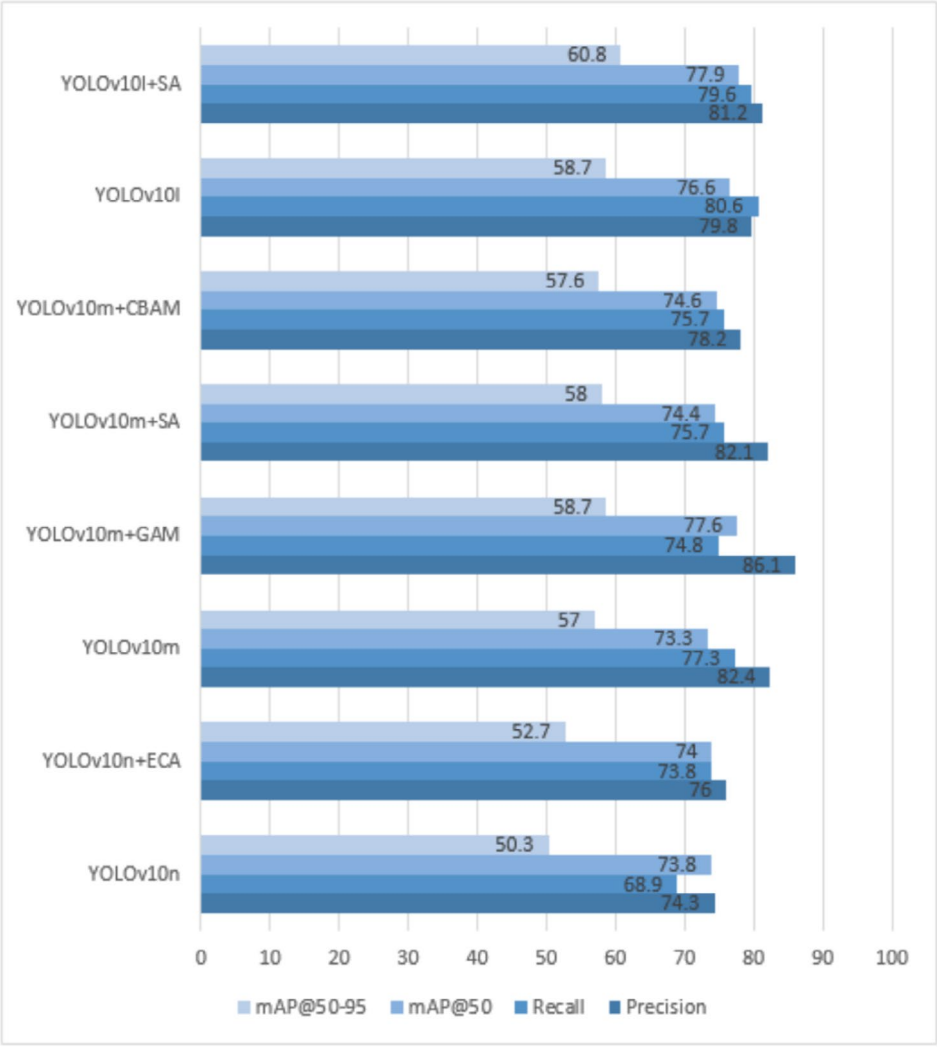

FIGURE 9 | Results of the Nepal Database.

struggled to enclose the complete landslide region, particularly for irregularly shaped landslides. The partial detections can have significant implications for hazard analysis, especially in terms of accurately estimating the size and impact of a landslide.

5 | Discussion

The proposed network effectively recognizes landslides, and to evaluate its appropriateness, we utilized an auxiliary dataset. Our results were compared with previous studies based on landslide detection. We highlighted the necessity of the patience factor in training the model. Lastly, we discussed the potential application of this research and the future directions of this research work.

5.1 | Performance Evaluation Using Benchmark Dataset

A comprehensive evaluation of the YOLOv10+attention model is also assessed using an additional benchmark dataset. The metrics mentioned in section 2.4 provide a comprehensive understanding of the model's accuracy, reliability, and overall performance in detecting landslides.

First, the prevalent benchmark dataset used for evaluation, HR-GLDD, comprises high-resolution PlanetScope images (Meena et al. 2023). We considered the case study of the Kodagu district, Western Ghats, India. Figure 8 shows the results of the performance metrics of various YOLOv10 models and their variants (integrating attention modules specifically SA, GAM, CBAM, and ECA) for landslide detection tasks. Each attention model enhances specific aspects of performance by enabling the network to focus on more relevant features during training, improving detection and localization. The addition of attention mechanisms significantly boosts precision across models. For instance, YOLOv10b + ECA shows the highest precision at 79.7% (an improvement of +15.80% from the baseline model), indicating that ECA effectively captures fine-grained channel dependencies, leading to more accurate detections. YOLOv10n (baseline) has a precision of 61.3%, which improves (+9.3%) to 70.6% with the SA module, demonstrating better feature refinement. Recall improvements (from the baseline models) are also estimated in different configurations. For example, YOLOv10l+ECA indicates an enhancement of 14.20%. Moreover, YOLOv10l+GAM showed an increase of 8.2%. An equal improvement (+7.0%) is calculated in the recall of YOLOv10l+ECA, and the progress of +5.0% is estimated for YOLOv10b+SA. Furthermore, adding attention mechanisms significantly improves mAP@50 across various YOLOv10 configurations. ECA achieves the highest improvements, with YOLOv10m+ECA and YOLOv10l+ECA recording impressive gains of +6.0% and +6.3% in mAP@50, respectively, highlighting ECA's strength in refining channel-wise features for better detection. CBAM demonstrates achieving a notable +5.2% increase in the YOLOv10s+CBAM model. Similarly, SA contributes consistently across scales, with YOLOv10n+SA, YOLOv10m+SA, and YOLOv10l+SA showing gains of +3.3%, +5.2%, and +4.7%, respectively, reflecting its ability to enhance spatial feature focus. GAM also delivers competitive improvements, particularly in larger models, with YOLOv10l+GAM achieving +3.9% followed by YOLOv10m+GAM (+3.0%). The attention models consistently improve the mAP@50-95 across configurations, suggesting they are particularly effective for more challenging IoU thresholds. Second, we considered the popular Nepal landslide detection database (Grzybowski et al. 2021). Figure 9 presents the overall outcomes of the baseline and attention-induced models. The most significant improvement in mAP@50 is observed in the YOLOv10m+GAM model, with an increase of +4.3%. Both YOLOv10m+CBAM and YOLOv10l+SA show a similar enhancement, each improving by +1.3%, while YOLOv10m+SA exhibits a+1.1% increase in mAP@50, and YOLOv10n+ECA shows a marginal improvement of +0.2%. When the threshold is increased to mAP@50-90, the YOLOv10n+ECA model demonstrates an improvement of +2.4%, highlighting its efficiency in correctly detecting and localizing landslides across variable surroundings. Similarly, the YOLOv10l+SA model also shows significant progress with a+2.1% increase, underscoring its capability in handling complex terrain and diverse landslide features.

TABLE 2 | Comparison of our results with previous work.

Reference	Model	*P* (%)	*R* (%)	*F* (%)
Our study	YOLOv10m+CBAM	98.0	94.1	96.0
Tanatipuknon et al. (2021)	R-CNN	85.0	64.0	71.0
	Improvement	+13.0	+30.1	+25.0
	Faster R-CNN	86.0	77.0	79.0
	Improvement	+12.0	+17.1	+17.0
Han et al. (2023)	Dynahead-YOLOv3	87.1	87.5	87.0
	Improvement	+10.1	+6.6	+9.0
Du et al. (2024)	LBE-YOLO	90.6	86.5	88.5
	Improvement	+7.4	+7.6	+7.5
Qin et al. (2023)	Faster R-CNN+VGG16	59.1	86.2	70.0
	Improvement	+38.9	+7.9	+20.0
	Faster R-CNN+ ResNet50	68.0	95.6	80.0
	Improvement	+30.0	−1.5	+16.0
	Faster R-CNN+ DarkNet53	69.5	93.8	80.0
	Improvement	+28.5	+0.3	+16.0
Yang et al. (2024)	YOLOv4+ECA	93.5	91.4	92.4
	Improvement	+4.5	+2.7	+3.6
	YOLOv4+CBAM	94.0	92.2	93.1
	Improvement	+4.0	+1.9	+2.9
	YOLOv4+VGG16	96.1	91.9	94.0
	Improvement	+1.9	+2.2	+2.0
	YOLOv4+LPFRF	95.5	94.2	94.9
	Improvement	+2.5	−0.1	+1.1

Additionally, the mAP@50-90 of YOLOv10m+GAM and YOLOv10m+SA improved by +1.7% and +1.0%, respectively, demonstrating their robustness in performance. Furthermore, while YOLOv10m+CBAM shows a smaller improvement of +0.6%, it still indicates that despite this relatively lower enhancement, it maintains competitiveness in precise landslide detection. These improvements in mAP@50-90 are crucial as they demonstrate the models' abilities to detect landslides and enhance the precision of localization, which is vital for practical applications in landslide hazard mapping.

5.2 | Comparative Analysis

First, we compared our outcomes with five previous research works that proposed 11 different methods for object detection models using the Bijie dataset. Table 2 lists these methods, showcasing the results. Our proposed methods demonstrated an improvement in calculated *F*-score, ranging from +1.1% to +25.0%. Precision also showed substantial progress, with improvements between +1.9% and +38.9%, indicating highly competitive results. Similarly, recall improved by +0.3% to +30.1%. An equivalent recall was noted in the case of YOLOv10m+CBAM (94.1%) and YOLOv4+LPFRF (94.2%), with no progress compared to Faster R-CNN+ResNet50 (95.6%). Nevertheless, the overall comparative results highlight the model's capability and effectiveness.

Second, we compared the outcomes of YOLOv10 with nine state-of-the-art algorithms specifically used for landslide detection: Faster R-CNN, SSD, YOLOv3 tiny, YOLOv5, YOLOv6n, YOLOv6t, YOLOv6s, YOLOv7, and YOLOv8. Figure 10 presents the overall comparison of these models based on the F-score. The YOLOv10m model demonstrated the maximum improvement, ranging from +9.17% to 18.57%. YOLOv10s showed an enhancement of +8.68% to 18.04%. Following closely, the YOLOv10n model also exhibited significant progress, with improvements between +8.34% and 17.74%. Similar advancements were observed for YOLOv10x (+6.3% to 15.7%) and YOLOv10l (+6.0% to 15.4%). However, the improvement for YOLOv10l was comparatively lower, ranging from +0.09% to 9.49%.

Comparing the performance of the proposed method with previous studies and state-of-the-art models is significant for several reasons. Firstly, it provides a benchmark to evaluate the effectiveness and efficiency of the new models. By demonstrating improvements over existing algorithms, we can validate the advancements in detection capabilities. Secondly, it highlights the robustness and adaptability of the YOLOv10 models, confirming their potential for real-world applications. Lastly, such

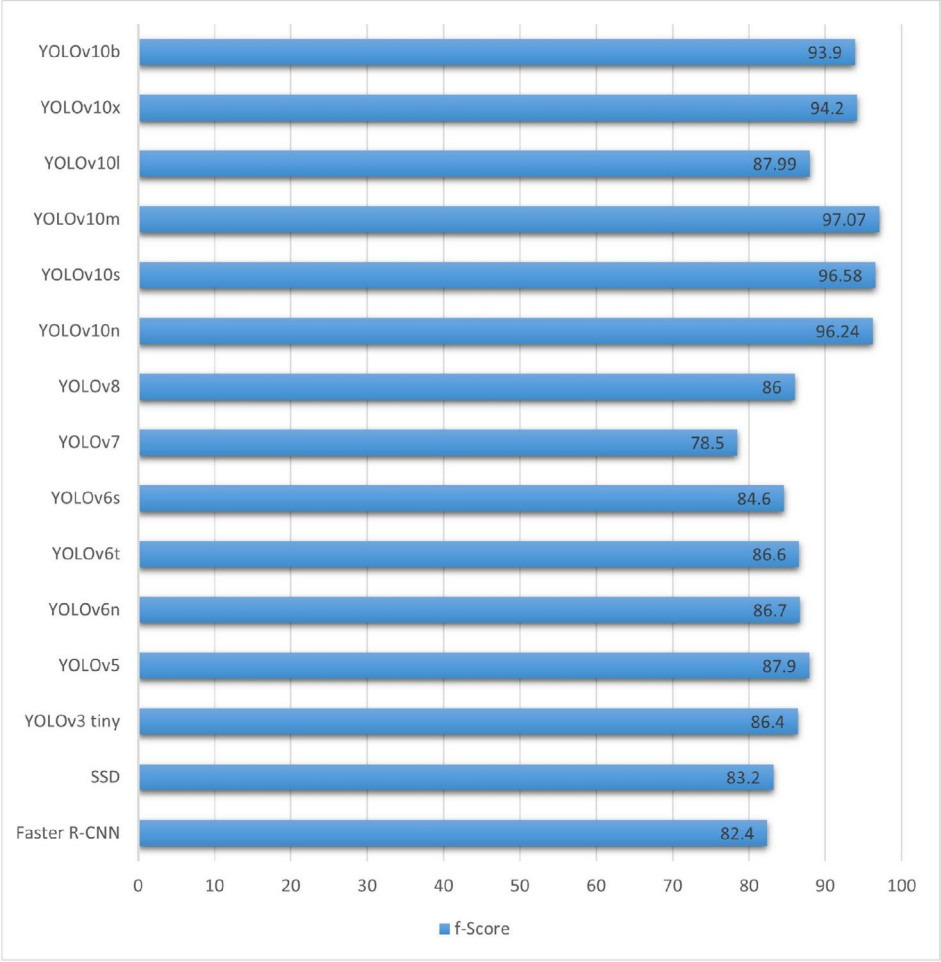

FIGURE 10 | Comparison of YOLOv10 variants with state-of-the-art models.

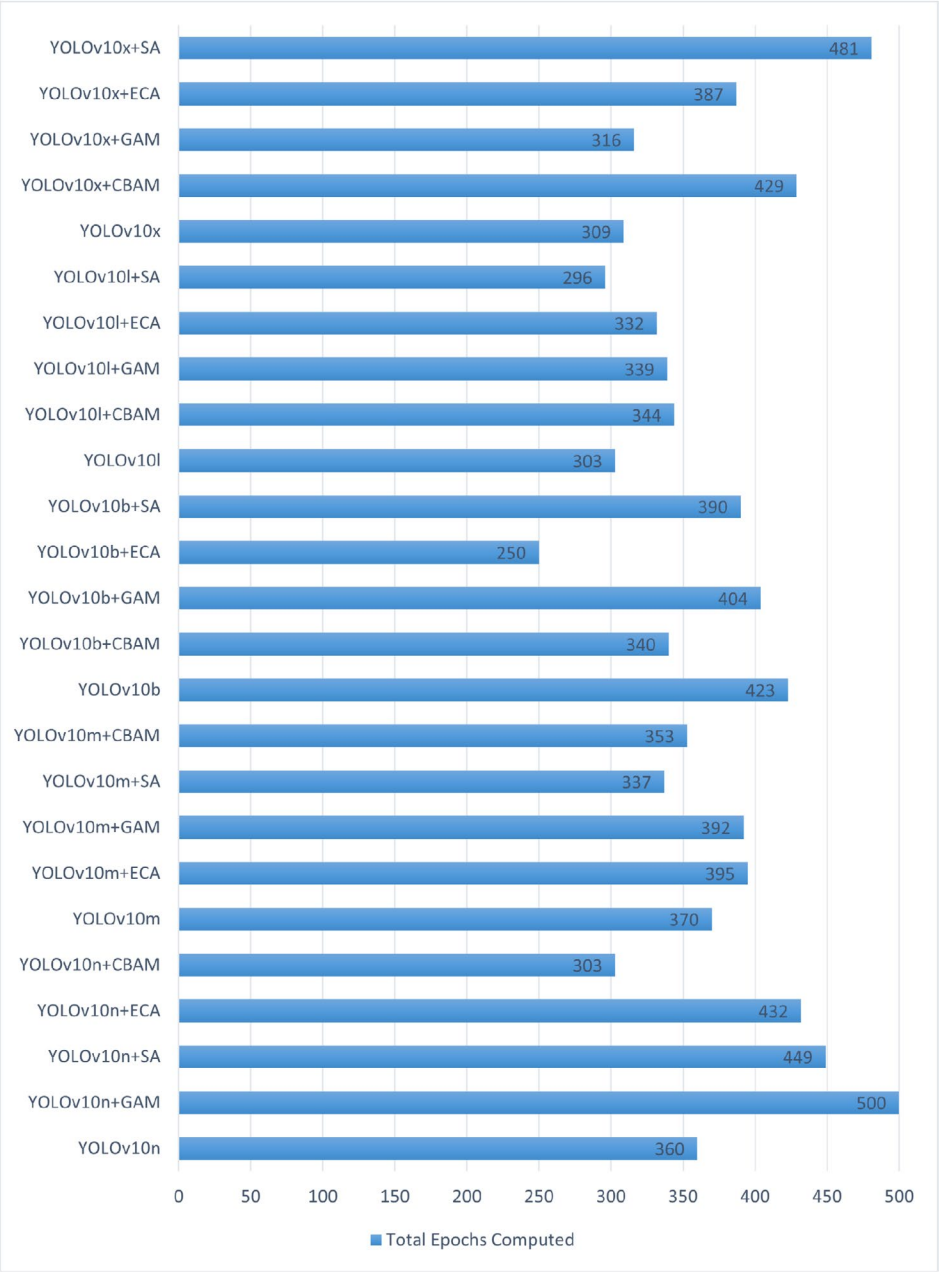

FIGURE 11 | Results of Early Stopping Criteria.

comparisons foster innovation and encourage the continuous development of more accurate and reliable models for critical tasks like landslide detection, ultimately contributing to better disaster management and mitigation strategies.

5.3 | Early Stopping Settings

Early stopping is a technique used in DL to prevent overfitting and enhance model generalization. It operates by monitoring the model's performance on a validation set during training and stopping the training process when the performance on this set starts to decrease, indicating that the model is beginning to overfit the training dataset. The early stopping parameter (termed patience) for our proposed YOLOv10+attention network is set to 100 (Kim et al. 2023; Dai and Fan 2022). This implies that during training, if the validation performance does not improve for 100 consecutive epochs, the training process will stop. The default patience value of 100 ensures an optimal approach for training YOLOv10 for our study. The total number of epochs computed for each model is shown in Figure 11. It is noted that, except for YOLOv10n+GAM, all the models stopped early. The patience parameter is flexible and can be easily adjusted according to the specific requirements of the training process. Setting patience to zero (0) effectively disables the early stopping mechanism. Disabling patience (patience = 0) can be useful in scenarios where the model needs to be trained for a fixed number of epochs regardless of the validation performance. When enabled, the patience parameter helps prevent overfitting by stopping the training once the model starts to overfit the training data, as indicated by a lack of improvement on the validation set. This is significant for the large volume of datasets, where the risk of overfitting is likely high.

5.4 | Potential Application and Future Directions

The proposed study has applications in real-world scenarios, particularly for mountain regions exposed to landslides, such as the northern part of India, especially Uttarakhand. The attention-integrated YOLO networks support the model in distinguishing landslide features from composite and mixed backgrounds. By refining the network's capacity to emphasize pertinent features, the model can detect small and large landslides, providing broad monitoring across the area. The suggested study holds significant potential for prompt and correct mapping of landslide scenes. This ability is vital for effective rescue and recovery operations during landslide events, where timely data can save lives and reduce damage. The results of our work can be important in time-restricted landslide mapping efforts, helping to generate up-to-date inventories of landslide occurrences. Correctly identifying and localizing the landslides is crucial for developing robust forecasting systems. This information can benefit in understanding the factors contributing to landslide occurrences and predicting future events, thus enhancing early warning systems. Exploiting UAVs equipped with high-resolution cameras and the attention-based YOLOv10 model can be deployed for real-time landslide monitoring. Furthermore, one of the main aims for future study is to enlarge the dataset size providing an additional comprehensive illustration of diverse landslides and their changing appearances, refining the model's accuracy and robustness. Besides, access to high-performance GPUs can suggestively accelerate the training method, allowing for more extensive experimentation with varying network arrangements and hyperparameters. Ablation experiments should explore and verify the effect of individual attention mechanisms in each position to evaluate the model's performance. Using distributed computing methods can further boost the training procedure, allowing the handling of big datasets and complex networks more efficiently.

6 | Conclusion

The addition of the attention layers within the YOLOv10 significantly improves the model's performance, particularly for landslide detection. By incorporating an attention mechanism, the model effectively captures relevant features and suppresses irrelevant ones, even in complex and diverse environmental conditions. The attention-augmented YOLOv10 outperforms traditional models, providing a robust solution for real-time and precise landslide hazard assessment. This approach demonstrates the potential for rapid and reliable mapping of landslide-prone regions, crucial for timely disaster response and mitigation efforts, particularly in vulnerable regions like the Himalayas. Future research can further refine these models and expand their applicability by leveraging larger datasets and advanced computational resources, ultimately enhancing the resilience of communities against landslide hazards.

Acknowledgments

The authors would like to thank the director of the Wadia Institute of Himalayan Geology (WIHG), Dehradun, for encouraging this research work. The contribution number of this research work is WIHG/0385.

Open access publishing facilitated by Universita degli Studi di Padova, as part of the Wiley - CRUI-CARE agreement.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Casagli, N., E. Intrieri, V. Tofani, G. Gigli, and F. Raspini. 2023. "Landslide Detection, Monitoring and Prediction With Remote-Sensing Techniques." *Nature Reviews Earth and Environment* 4, no. 1: 51–64. https://doi.org/10.1038/s43017-022-00373-x.

Chandra, N., S. Sawant, and H. Vaidya. 2023. "An Efficient u-Net Model for Improved Landslide Detection From Satellite Images." *PFG-Journal of Photogrammetry Remote Sensing and Geoinformation Science* 91, no. 1: 13–28. https://doi.org/10.1007/s41064-023-00232-4.

Chandra, N., and H. Vaidya. 2024a. "Automated Detection of Landslide Events From Multi-Source Remote Sensing Imagery: Performance Evaluation and Analysis of Yolo Algorithms." *Journal of Earth System Science* 133, no. 3: 127.

Chandra, N., and H. Vaidya. 2024b. "Deep Learning Approaches for Landslide Information Recognition: Current Scenario and Opportunities." *Journal of Earth System Science* 133, no. 2: 1–25.

Cheng, L., J. Li, P. Duan, and M. Wang. 2021. "A Small Attentional Yolo Model for Landslide Detection From Satellite Remote Sensing Images." *Landslides* 18: 2751–2765. https://doi.org/10.1007/s10346-021-01694-6.

Dai, G., and J. Fan. 2022. "An Industrial-Grade Solution for Crop Disease Image Detection Tasks." *Frontiers in Plant Science* 13: 921057. https://doi.org/10.3389/fpls.2022.921057.

Devara, M., V. Maurya, and R. Dwivedi. 2024. "Landslide Extraction Using a Novel Empirical Method and Binary Semantic Segmentation u-Net Framework Using Sentinel-2 Imagery." *Remote Sensing Letters* 15: 326–338. https://doi.org/10.1080/2150704X.2024.2320178.

Diwan, T., A. Ani, and J. Tembhurne. 2023. "Object Detection Using Yolo: Challenges, Architectural Successors, Datasets and Applications." *Multimedia Tools and Applications* 82, no. 6: 9243–9275. https://doi.org/10.1007/s11042-022-13644-y.

Du, Y., X. Xu, and X. He. 2024. "Optimizing Geo-Hazard Response: Lbe-yolo's Innovative Lightweight Framework for Enhanced Real-Time Landslide Detection and Risk Mitigation." *Remote Sensing* 16: 534. https://doi.org/10.3390/rs16030534.

Ghorbanzadeh, O., A. Crivellari, P. Ghamisi, H. Shahabi, and T. Blaschke. 2021. "A Comprehensive Transferability Evaluation of u-Net and Resu-Net for Landslide Detection From Sentinel-2 Data (Case Study Areas From Taiwan, China, and Japan)." *Scientific Reports* 11, no. 1: 14629. https://doi.org/10.1038/s41598-021-94190-9. https://api.semanticscholar.org/CorpusID:235999785.

Grzybowski, J., L. Bragagnolo, L. Rezende, and R. Silva. 2021. "Convolutional Neural Networks Applied to Semantic Segmentation of Landslide Scars." *Catena* 201: 1. https://doi.org/10.1016/j.catena.2021.105189.

Guzzetti, F., A. Mondini, M. Cardinali, F. Fiorucci, and M. Santangelo. 2012. "Landslide Inventory Maps: New Tools for an Old Problem." *Earth-Science Reviews* 112: 42–66. https://doi.org/10.1016/j.earscirev.2012.02.001.

Han, Z., Z. Fang, Y. Li, and B. Fu. 2023. "A Novel Dynahead-Yolo Neural Network for the Detection of Landslides With Variable Proportions

Using Remote Sensing Images." *Frontiers in Earth Science* 10: 1077153. https://doi.org/10.3389/feart.2022.1077153.

Hong, D., C. Li, B. Zhang, N. Yokoya, J. A. Benediktsson, and J. Chanussot. 2024. "Multimodal Artificial Intelligence Foundation Models: Unleashing the Power of Remote Sensing Big Data in Earth Observation." https://www.the-innovation.org/article/id/65e4f30cbf979918c9d8124d.

Iqbal, U., T. Davies, and P. Perez. 2024. "A Review of Recent Hardware and Software Advances in Gpu-Accelerated Edge-Computing Single-Board Computers (Sbcs) for Computer Vision." *Sensors (Basel, Switzerland)* 24, no. 15: 4830. https://doi.org/10.3390/s24154830. https://www.mdpi.com/1424-8220/24/15/4830.

Ji, S., Y. Dawen, C. Shen, W. l. Li, and Q. Xu. 2020. "Landslide Detection From an Open Satellite Imagery and Digital Elevation Model Dataset Using Attention Boosted Convolutional Neural Networks." *Landslides* 17, no. 6: 1337–1352. https://doi.org/10.1007/s10346-020-01353-2.

Jia, L., X. Leng, X. Wang, and M. Nie. 2024. "Recognizing Landslides in Remote Sensing Images Based on Enhancement of Information in Digital Elevation Models." *Remote Sensing Letters* 15: 224–232. https://doi.org/10.1080/2150704X.2024.2313611.

Keyport, R., T. Oommen, T. Martha, K. Sajinkumar, and J. Gierke. 2018. "A Comparative Analysis of Pixel- and Object-Based Detection of Landslides From Very High-Resolution Images." *International Journal of Applied Earth Observation and Geoinformation* 64: 1–11. https://doi.org/10.1016/j.jag.2017.08.015.

Kim, K., K. Kim, and S. Jeong. 2023. "Application of Yolo v5 and v8 for Recognition of Safety Risk Factors at Construction Sites." *Sustainability* 15: 15179. https://doi.org/10.3390/su152015179.

Li, B., and J. Li. 2022. "Methods for Landslide Detection Based on Lightweight yolov4 Convolutional Neural Network." *Earth Science Informatics* 15: 765–775. https://doi.org/10.1007/s12145-022-00764-0.

Lin, T. Y., P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie. 2017. "Feature Pyramid Networks for Object Detection." In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 936–944.

Liu, P., Y. Wei, Q. Wang, J. Xie, Y. Chen, and Z. Li. 2021. "A Research on Landslides Automatic Extraction Model Based on the Improved Mask r-Cnn." *ISPRS International Journal of Geo-Information* 10, no. 3: 168. https://doi.org/10.3390/ijgi10030168. https://www.mdpi.com/2220-9964/10/3/168.

Liu, Q., T. Wu, Y. Deng, and Z. Liu. 2023. "Se-yolov7 Landslide Detection Algorithm Based on Attention Mechanism and Improved Loss Function." *Landscape* 12: 1522. https://doi.org/10.3390/land12081522.

Liu, S., L. Qi, H. Qin, J. Shi, and J. Jia. 2018. "Path Aggregation Network for Instance Segmentation." In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8759–8768.

Liu, Y., Z. Shao, and N. Hoffmann. 2021. "Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions." arXiv Preprint arXiv: 2112.05561.

Lu, P., A. Stumpf, N. Kerle, and N. Casagli. 2011. "Object-Oriented Change Detection for Landslide Rapid Mapping." *Geoscience and Remote Sensing Letters, IEEE* 8: 701–705. https://doi.org/10.1109/LGRS.2010.2101045.

Ma, Y., H. Wu, L. Wang, et al. 2015. "Remote Sensing Big Data Computing: Challenges and Opportunities." *Future Generation Computer Systems* 51: 47–60. https://doi.org/10.1016/j.future.2014.10.029.

Ma, Z., and G. Mei. 2021a. "Deep Learning for Geological Hazards Analysis: Data, Models, Applications, and Opportunities." *Earth-Science Reviews* 223: 1–33. https://doi.org/10.1016/j.earscirev.2021.103858.

Ma, Z., and G. Mei. 2021b. "Machine Learning for Landslides Prevention: A Survey."

Malamud, B. D., D. L. Turcotte, F. Guzzetti, and P. Reichenbach. 2004. "Landslide Inventories and Their Statistical Properties." *Earth Surface Processes and Landforms* 29, no. 6: 687–711. https://doi.org/10.1002/esp.1064.

Mao, Y., R. Niu, B. Li, and J. Li. 2024. "Potential Lanslides Identification Based on Improved yolov8 and Insar Phase-Gradient Stacking." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17: 1–10. https://doi.org/10.1109/JSTARS.2024.3399788.

Meena, S., L. Nava, K. Bhuyan, et al. 2023. "Hr-Gldd: A Globally Distributed Dataset Using Generalized Deep Learning (Dl) for Rapid Landslide Mapping on High-Resolution (Hr) Satellite Imagery." *Earth System Science Data* 15, no. 7: 3283–3298. https://doi.org/10.5194/essd-15-3283-2023.

Minaee, S., Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos. 2022. "Image Segmentation Using Deep Learning: A Survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, no. 7: 3523–3542. https://doi.org/10.1109/TPAMI.2021.3059968.

Mo, P., D. Li, M. Liu, J. Jia, and X. Chen. 2023. "A Lightweight and Partitioned CNN Algorithm for Multi-Landslide Detection in Remote Sensing Images." *Applied Sciences* 13: 8583. https://doi.org/10.3390/app13158583.

Mohan, A., A. Singh, B. Kumar, and R. Dwivedi. 2021. "Review on Remote Sensing Methods for Landslide Detection Using Machine and Deep Learning." *Transactions on Emerging Telecommunications Technologies* 32, no. 7: e3998. https://doi.org/10.1002/ett.3998.

Niu, Z., G. Zhong, and H. Yu. 2021. "A Review on the Attention Mechanism of Deep Learning." *Neurocomputing* 452: 48–62. https://doi.org/10.1016/j.neucom.2021.03.091.

Qin, H., J. Wang, X. Mao, Z. Zhao, X. Gao, and W. Lu. 2023. "An Improved Faster r-Cnn Method for Landslide Detection in Remote Sensing Images." *Journal of Geovisualization and Spatial Analysis* 8, no. 1: 2. https://doi.org/10.1007/s41651-023-00163-z.

Shrestha, A., and A. Mahmood. 2025. "A Metabolomic Study Uncovering Key Amino Acids and Amines in Duroc Boar Semen as Biomarkers of Sexual Maturity." *Animal Reproduction Science* 275: 107800. https://doi.org/10.1109/ACCESS.2019.2912200.

Stephen, A., I. Gratchev, D. H. Kim, and S. Y. Ohn. 2024. "Application of Artificial Intelligence and Remote Sensing for Landslide Detection and Prediction: Systematic Review." *Remote Sensing* 16: 2947. https://doi.org/10.3390/rs16162947.

Tanatipuknon, A., P. Aimmanee, Y. Watanabe, et al. 2021. "Study on Combining Two Faster r-Cnn Models for Landslide Detection With a Classification Decision Tree to Improve the Detection Performance." *Journal of Disaster Research* 16, no. 4: 588–595. https://doi.org/10.20965/jdr.2021.p0588.

Ullo, S., A. Mohan, A. Sebastianelli, et al. 2021. "A New Mask R-CNN-Based Method for Improved Landslide Detection." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14: 1. https://doi.org/10.1109/JSTARS.2021.3064981.

Wang, A., H. Chen, L. Liu, et al. 2024. "Yolov10: Real-Time End-to-End Object Detection." arXiv Preprint arXiv:2405.14458.

Wang, C. Y., H. Y. Mark Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh. 2020. "Cspnet: A New Backbone That Can Enhance Learning Capability of Cnn." In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1571–1580.

Wang, Q., B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu. 2020. "Eca-Net: Efficient Channel Attention for Deep Convolutional Neural Networks." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11534–11542.

Woo, S., J. Park, J. Y. Lee, and I. S. Kweon. 2018. "Cbam: Convolutional Block Attention Module." In *Proceedings of the European Conference on Computer Vision (ECCV), 2018*, 3–19.

Yang, Y., Z. Miao, H. Zhang, B. Wang, and L. Wu. 2024. "Lightweight Attention-Guided Yolo With Level Set Layer for Landslide Detection From Optical Satellite Images." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17: 1–19. https://doi.org/10.1109/JSTARS.2024.3351277.

Zeng, D., M. Liao, M. Tavakolian, et al. 2021. "Deep Learning for Scene Classification: A Survey." ArXiv, abs/2101.10531. https://api.semanticscholar.org/CorpusID:231709709.

Zhang, Q., and T. Wang. 2024. "Deep Learning for Exploring Landslides With Remote Sensing and Geo-Environmental Data: Frameworks, Progress, Challenges, and Opportunities." *Remote Sensing* 16, no. 8: 1344. https://doi.org/10.3390/rs16081344. https://www.mdpi.com/2072-4292/16/8/1344.

Zhang, Q. L., and Y. B. Yang. 2021. "Sa-Net: Shuffle Attention for Deep Convolutional Neural Networks." In *ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2235–2239.

Zhang, X., Z. Yanan, and J. Luo. 2021. "Deep Learning for Processing and Analysis of Remote Sensing Big Data: A Technical Review." *Big Earth Data* 6: 1–34. https://doi.org/10.1080/20964471.2021.1964879.

Zhao, Z. Q., P. Zheng, S. T. Xu, and X. Wu. 2019. "Object Detection With Deep Learning: A Review." *IEEE Transactions on Neural Networks and Learning Systems* 30, no. 11: 1–21. https://doi.org/10.1109/TNNLS.2018.2876865.