http://www.jourmag.org/index.php/jma

Journal of Magnetism and Its Applications

Received August 17th 2026 | Revised October 6th 2026 | Accepted November 18th 2026

Research article

Automated Detection of Nocturnal Blue Tears Using YOLO12n and CutMix Data Augmentation

Shahbaz Nazeer[1], Abdul Qadeer[1], Rabia Saleem ,Saher Asghar , Faiq Hassan[1]

[1] Department of Information Technology, Government College University Faisalabad, Faisalabad 38000, Pakistan

shahbaznazeer@gcuf.edu.pk,qadeeryaseen1058@gmail.com,rabiasaleem@gcuf.edu.pksaherasgharnagra@gmail.com, i.faiqhassan1@gmail.com

Article info

Keywords:
Blue Tears; bioluminescence; YOLO12n; CutMix; object detection; BT3.8k; coastal monitoring

Abstract

Blue Tears are transient coastal bioluminescent events that are challenging to detect in low-light images. This study evaluates YOLO12n with CutMix augmentation on BT3.8k, a benchmark of 3,827 annotated images extracted from 241 user-uploaded videos. Training used 260 epochs, 640 × 640 inputs and CutMix probability 0.5. The reported evaluation yielded 95.09% mAP@50, 72.03% mAP@50–95, 92.45% precision, 89.89% recall and a 91.15% F1-score. Compared with YOLO12n without CutMix, mAP@50 improved by 7.86 percentage points. The detector contains 2.56 million parameters versus 22.39 million reported for BT-YOLO, although this cross-study comparison did not use a unified retraining protocol. These findings support augmentation-assisted detection on BT3.8k, while clip-level separation, geographically independent evaluation and deployment latency still require validation.

1. Introduction

Blue Tears, or sea sparkle, refers to visible blue bioluminescence produced in coastal water, frequently associated with the dinoflagellate Noctiluca scintillans. The light is most apparent at night and may be triggered by wave action, movement of vessels or other mechanical disturbance. Although these events have environmental and tourism significance, their occurrence can be transient and spatially irregular [1], [2]. Conventional field observation is difficult to sustain continuously, while many established satellite water-quality techniques primarily measure daytime optical signals rather than faint nocturnal light emission [3], [4].

Recent one-stage object detectors provide an alternative way to localize visual events in images collected by cameras and mobile devices. YOLO models jointly predict object locations and confidence values in a single detection pipeline, making them candidates for automated observation systems [5], [6]. Nevertheless, Blue Tears imagery is challenging: naturally luminous wave fronts may be narrow or partially obscured, shoreline lights may resemble the target, and social-media photographs can vary considerably in resolution, framing and exposure. Training such detectors on limited video-derived datasets therefore raises questions about robustness and generalization.

Data augmentation can increase the diversity of visual contexts available during learning without requiring new annotations. CutMix forms composite samples by inserting image regions from one example into another, with corresponding label adjustments [7], [8]. Unlike approaches that primarily alter the detector architecture, augmentation seeks to improve representation learning from the available training set. Whether this improves nocturnal Blue Tears localization is the question addressed by the present experiments.

This paper condenses the experimental study reported in the source MS thesis. Its contributions are (i) an experimental pipeline for one-class Blue Tears detection using YOLO12n, (ii) a within-model CutMix ablation, (iii) evaluation with precision, recall, F1-score and mAP measures, and (iv) an accuracy–parameter comparison against BT-YOLO values reported in the thesis. The study concerns coastal video-derived images rather than direct multispectral observations of small inland lakes. Its claims are therefore restricted to image-based detection of visible nocturnal blooms.

2. Experimental Methods

2.1 Dataset and evaluation split

The BT3.8k dataset contains 3,827 images extracted from 534 short clips derived from 241 user-uploaded videos. It includes content from WeChat, YouTube, TikTok, Bilibili and REDnote, acquired under different viewing angles, lighting conditions and wave patterns. The dataset includes close, distant and some aerial views; approximately 80% of clips were recorded on mobile phones, 17% on mixed camera types and 3% using drones, according to the source thesis. Each annotated Blue Tears region was represented by a bounding box in YOLO format.

*Corresponding author Faiq Hassan Email: i.faiqhassan1@gmail.com

Table 1. Video sources and extracted images in BT3.8k.

Platform	Raw videos	Clips (≤10 s)	Annotated images
WeChat	75	122	1,220
YouTube	26	112	804
TikTok	25	38	272
Bilibili	11	92	317
REDnote	104	170	1,214
Total	241	534	3,827

Images were assigned to a 70:20:10 training/validation/test partition: 2,679 training images, 765 validation images and 383 test images. The training set was used for weight optimization, validation for monitoring and checkpoint selection, and testing for final evaluation. Because multiple still images originate from the same source clips, independence of these partitions depends on whether nearby frames were grouped before splitting. The thesis describes a fixed split but does not document clip-group identifiers or a leakage audit; this limitation is retained in the interpretation of results.

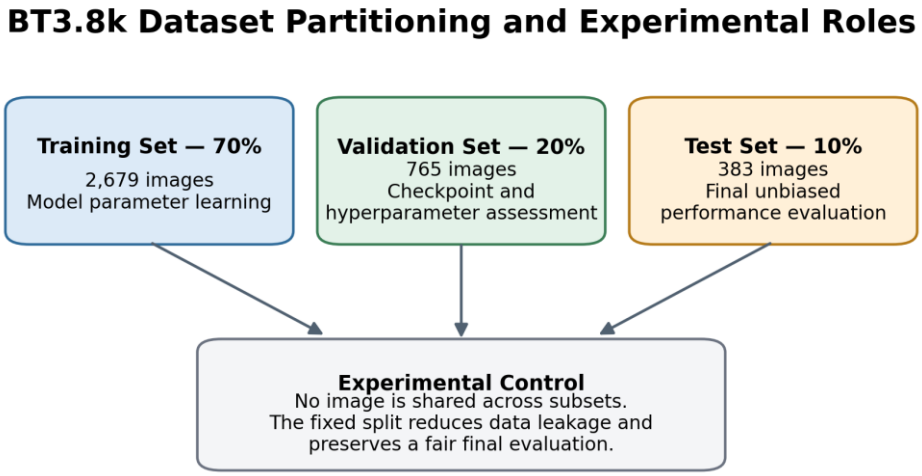


Fig. 1. Reported partition and experimental roles for the BT3.8k dataset (adapted from thesis Figure 3.2).

2.2 Preprocessing and augmentation

The preprocessing procedure checked image–label pairs, the validity of single-class identifiers, normalized bounding-box coordinates and the absence of malformed or zero-area boxes. Images were resized to 640 × 640 pixels for model input. The thesis reports correction of annotation errors where identified, but does not provide counts of corrected boxes or a detailed annotation agreement study.

During training, CutMix was activated with probability 0.5. A rectangular patch from a donor image was inserted into a second image, and detection labels were adjusted to account for donor boxes, obscured regions and image bounds. This exposes the detector to combinations of glowing wave regions, dark water and artificial background lights, potentially reducing dependence on particular scenes. The augmentation was applied during model fitting; the reported validation and test partitions were not described as augmented. Figure 2 illustrates the data-mixing principle. The original CutMix strategy is described in [8].

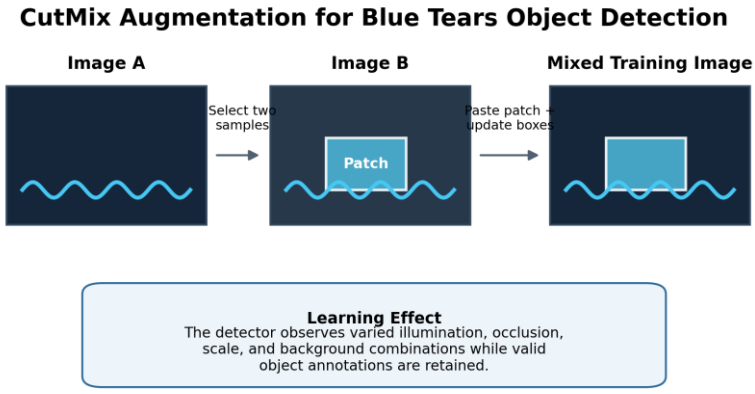


Fig. 2. Conceptual CutMix augmentation applied to Blue Tears imagery, including bounding-box updates (source thesis Figure 3.4).

*Corresponding author Faiq Hassan Email: i.faiqhassan1@gmail.com

2.3 Detector architecture and training settings

The experimental backbone was the nano-scale YOLO12n one-stage object detector, implemented with the Ultralytics framework. Its backbone extracts hierarchical image features, the neck combines multi-scale representations, and the detection head produces localization and confidence estimates. The smaller parameter budget was chosen to support possible future deployment on constrained hardware; runtime or frame-rate benchmarks were not supplied in the thesis. YOLO-family design and detection principles are described in [5], [6], [9].

Table 2. Training and software configuration reported in the thesis.

Parameter	Reported setting
Detection model	YOLO12n (2.56 million parameters)
Dataset and classes	BT3.8k; one class (Blue Tears)
Data split	2,679 train / 765 validation / 383 test
Epochs / batch size	260 / 16
Input size	640 × 640 pixels
CutMix probability	0.5
Optimizer	AdamW (reported)
Framework / environment	Ultralytics / Kaggle
GPU	NVIDIA Tesla T4 (reported)
Workers / caching	2 / enabled

The thesis does not provide a fixed random seed, exact dependency versions, model checkpoint identifier or training-source repository. Therefore, the settings above describe the reported experiment but do not alone constitute a fully reproducible computational record. In addition, the relative attribution of gains to CutMix depends on holding all remaining training conditions fixed between its on/off configurations.

2.4 Evaluation criteria and comparisons

A prediction was counted as a true positive when it matched a labeled region at the specified intersection-over-union (IoU) threshold; unmatched predictions and unmatched ground-truth targets contributed false positives and false negatives, respectively. Precision was computed as TP/(TP + FP), recall as TP/(TP + FN), and F1-score as 2PR/(P + R). The mAP@50 metric summarizes detection average precision at IoU 0.50; mAP@50–95 averages performance across progressively stricter IoU thresholds between 0.50 and 0.95 [10], [11].

Two comparisons are reported. First, an internal ablation contrasts YOLO12n trained with and without CutMix on BT3.8k. This is the more directly interpretable test of augmentation benefit. Second, the thesis compares its final mAP@50 and parameter count against reported BT-YOLO benchmark values. Because a common training/evaluation protocol for both architectures is not documented, this second comparison is descriptive and does not prove that all observed differences arise from architectural or augmentation choices.

3. Results and Discussion

3.1 Detection accuracy and training progression

At epoch 260, the source thesis reports 95.09% mAP@50, 72.03% mAP@50–95, 92.45% precision, 89.89% recall and 91.15% F1-score for YOLO12n with CutMix. The difference between mAP@50 and mAP@50–95 indicates that stricter spatial localization remains more difficult than detection at IoU 0.50. In particular, faint or fragmented luminous patterns may be recognized but difficult to bound precisely. These figures are reproduced from the source thesis and were not independently recomputed from raw detections.

Table 3. Metric progression during training, reproduced from the thesis.

Epoch	mAP@50 (%)	Precision (%)	Recall (%)	F1-score (%)
50	82.50	79.90	73.70	76.66
100	85.14	82.58	75.75	79.02
130	87.16	84.10	78.92	81.43
160	88.92	85.71	80.76	83.17
190	89.78	86.92	82.22	84.50
220	90.61	88.28	83.23	85.68
260	95.09	92.45	89.89	91.15

All four reported epoch-wise metrics increase from the early checkpoint at epoch 50 to epoch 260. mAP@50 rises by 12.59 percentage points, from 82.50% to 95.09%, while recall rises from 73.70% to 89.89%. The numerical trajectory is displayed in Figure 3. The table gives selected checkpoints rather than every epoch; therefore, it does not establish monotonicity at each individual update or exclude transient training instability.

*Corresponding author Faiq Hassan Email: i.faiqhassan1@gmail.com

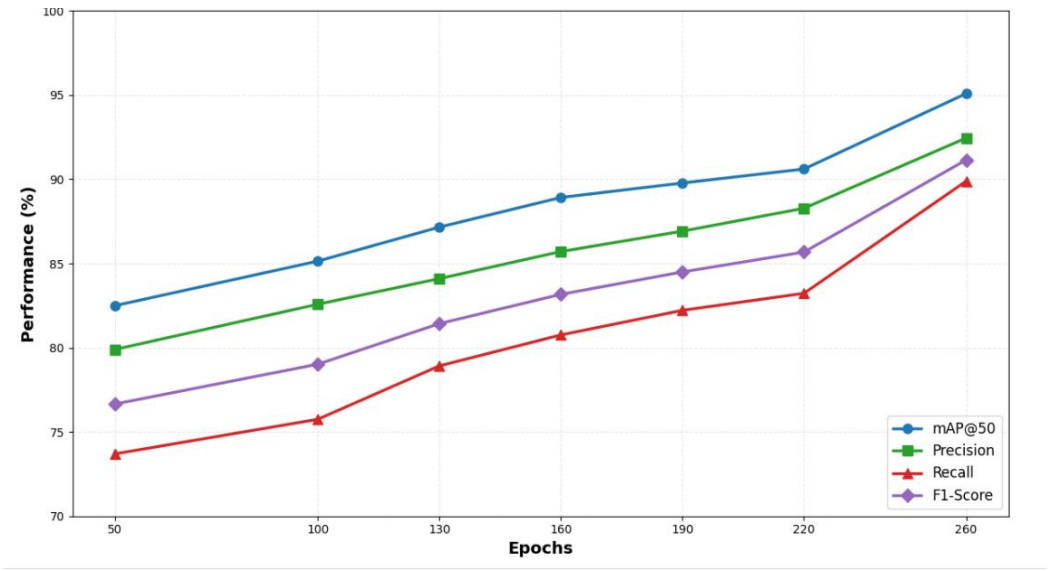


Fig. 3. Training progression of mAP@50, precision, recall and F1-score (source thesis Figure 4.6).

3.2 CutMix ablation

The most direct experimental evidence for the proposed strategy is the within-model CutMix ablation. The thesis reports 87.23% mAP@50 without CutMix and 95.09% with CutMix. Precision increases by 7.89 points, recall by 8.55 points and F1-score by 8.23 points. These improvements are consistent with the hypothesis that patch-based mixing increases robustness to visual variation in nocturnal coastal scenes. However, the thesis reports a single summarized run for each configuration; variability across seeds and confidence intervals were not available.

Table 4. YOLO12n ablation with and without CutMix. All values are percentages.

Configuration	mAP@50	Precision	Recall	F1-score
Without CutMix	87.23	84.56	81.34	82.92
With CutMix	95.09	92.45	89.89	91.15
Difference (pp)	+7.86	+7.89	+8.55	+8.23

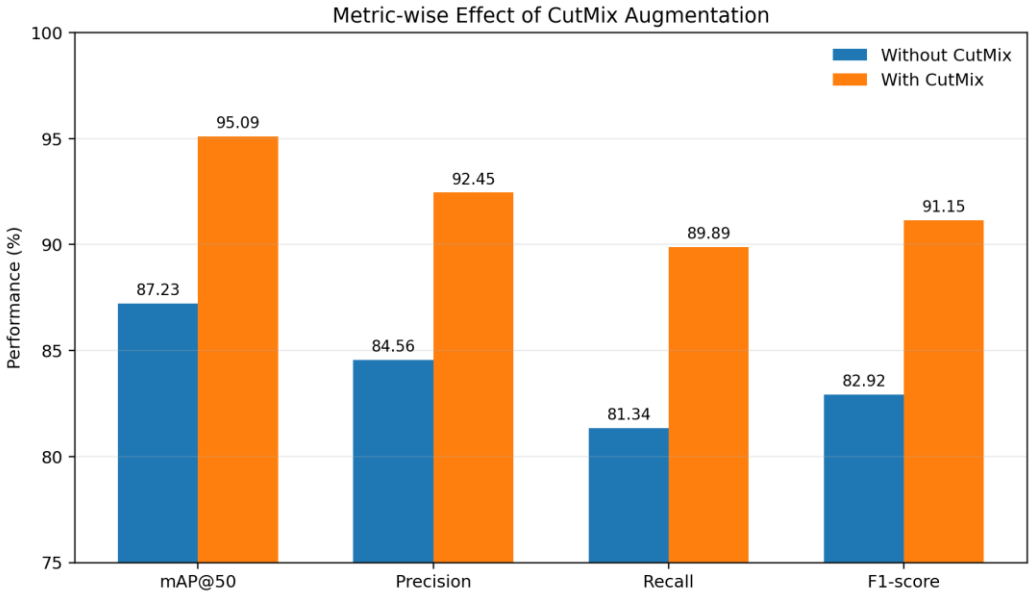


Fig. 4. Comparison of reported performance with and without CutMix augmentation (source thesis Figure 4.12).

Recall improvement is particularly relevant to low-light monitoring because missed detections may occur when the glow is weak, partly obscured or visually similar to reflections. Nonetheless, precision and recall depend on the confidence threshold and evaluation labels; they are not direct measures of real-world false alarm frequency per hour. The gain observed on the BT3.8k split should therefore be validated using independent footage before operational interpretation.

*Corresponding author Faiq Hassan Email: i.faiqhassan1@gmail.com

3.3 Accuracy–complexity comparison with BT-YOLO

The thesis cites 80.00% mAP@50 and 22.39 million model parameters for BT-YOLO, compared with 95.09% mAP@50 and 2.56 million parameters for YOLO12n with CutMix. Thus, the quoted mAP@50 values differ by 15.09 percentage points, and the proposed model has approximately 88.6% fewer parameters (2.56 versus 22.39 million). The model-size reduction does not by itself establish faster inference; measured latency, memory use and throughput on the same device are needed for that conclusion.

Table 5. Reported benchmark comparison (not a controlled head-to-head retraining).

Model	mAP@50 (%)	Parameters (million)
BT-YOLO (reported reference)	80.00	22.39
YOLO12n + CutMix (this thesis)	95.09	2.56

These benchmark numbers should be interpreted cautiously. The thesis does not establish that both models were re-evaluated under an identical clip-grouped data partition, threshold settings, preprocessing chain and hardware. In addition, the original BT-YOLO publication is discussed in the thesis but its complete bibliographic entry is absent from the supplied reference list. Verification of the original comparator and a harmonized retraining experiment are required before making a formal state-of-the-art claim.

3.4 Failure modes, practical interpretation and limitations

The thesis identifies non-bloom light sources, reflections, heterogeneous video quality and weak distant targets as practical error sources. Bright shoreline lamps and reflected light can be confused with bioluminescence, whereas faint or small glow regions can be missed. Potential countermeasures include targeted negative examples, improved exposure normalization, temporal consistency checks and more geographically diverse data. These are proposed mitigations, not experimentally verified improvements in the present study.

A potential monitoring workflow would sample frames from coastal cameras, apply the trained detector and require repeated detections before alerting an operator or updating a dashboard. The conceptual system in the thesis does not constitute an evaluated field deployment; model latency, energy consumption, calibration stability and false alerts over prolonged observation remain unmeasured.

The study also has dataset and inference limitations. First, the 3,827 images arise from only 241 videos, so neighboring frames may share visual content unless clip-level splitting is enforced. Second, platform and location imbalance may limit transfer to unseen coasts and camera hardware. Third, biological species identification and water-quality measurements were not undertaken, so the detector recognizes visible Blue Tears-like image patterns rather than confirming a species or toxin status. Fourth, neither significance testing across repeated seeds nor an independent external benchmark is reported. These limits frame the reported metrics as benchmark-specific rather than proof of universal or operational performance.

4. Conclusions

The source thesis demonstrates that a compact YOLO12n object detector combined with probabilistic CutMix augmentation can achieve strong reported detection scores on the BT3.8k Blue Tears image benchmark. The reported final values are 95.09% mAP@50, 72.03% mAP@50–95, 92.45% precision, 89.89% recall and 91.15% F1-score. Relative to the same architecture trained without CutMix, mAP@50 improved by 7.86 percentage points. The model contains 2.56 million parameters, supporting further investigation of resource-constrained coastal monitoring. Future work should document a video-grouped split, repeat experiments across random seeds, verify the BT-YOLO baseline in a unified protocol, and measure generalization and latency on new cameras and coastlines. These checks are essential before treating a promising benchmark detector as a validated early-warning system.

Acknowledgements

This article is derived from the MS Information Technology thesis of Ayesha Mushtaq at Government College University Faisalabad (2026). The thesis acknowledges academic supervision by Mr. Shahbaz Nazeer. Specific funding information was not supplied in the source document.

References

[1] D. M. Anderson, A. D. Cembella, and G. M. Hallegraeff, “Progress in understanding harmful algal blooms: Paradigm shifts and new technologies for research, monitoring, and management,” Annual Review of Marine Science, vol. 4, pp. 143–176, 2012. doi: 10.1146/annurev-marine-120308-081121.

[2] H. W. Paerl and T. G. Otten, “Harmful cyanobacterial blooms: Causes, consequences, and controls,” Microbial Ecology, vol. 65, no. 4, pp. 995–1010, 2013.

[3] C. Hu, “A novel ocean color index to detect floating algae in the global oceans,” Remote Sensing of Environment, vol. 113, no. 10, pp. 2118–2129, 2009.

[4] D. M. Anderson, “Approaches to monitoring, control and management of harmful algal blooms,” Ocean & Coastal Management, vol. 52, no. 7, pp. 342–347, 2009. doi: 10.1016/j.ocecoaman.2009.04.006.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, real-time object detection,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2016, pp. 779–788.

[6] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal speed and accuracy of object detection,” arXiv:2004.10934, 2020.

[7] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” Journal of Big Data, vol. 6, article 60, 2019.

[8] S. Yun et al., “CutMix: Regularization strategy to train strong classifiers with localizable features,” in Proc. IEEE/CVF Int. Conf. Computer Vision, 2019, pp. 6023–6032.

[9] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” arXiv:2207.02696, 2022.

[10] M. Everingham et al., “The Pascal Visual Object Classes (VOC) challenge,” International Journal of Computer Vision, vol. 88, no. 2, pp. 303–338, 2010.

[11] T.-Y. Lin et al., “Microsoft COCO: Common objects in context,” in Computer Vision – ECCV 2014, 2014, pp. 740–755.

*Corresponding author Faiq Hassan Email: i.faiqhassan1@gmail.com