Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 16(3), 1000-1013, 2026
Journal of the Institute of Science and Technology, 16(3), 1000-1013, 2026

ISSN: 2146-0574, eISSN: 2536-4618

Electrical Electronic Engineering **DOI: 10.21597/jist.1926599**

Research Article

Received: 09.04.2026 **Accepted: 04.05.2026**

To Cite: Turay, T. (2026). Analyzing the Impact of Photometric Data Augmentation on Medical Instance Segmentation Performance. *Journal of the Institute of Science and Technology*, 16(3), 1000-1013.

Analyzing the Impact of Photometric Data Augmentation on Medical Instance Segmentation Performance

Tolga TURAY[1*]

Highlights:

- The effect of photometric data augmentation on instance segmentation is analyzed.
- Color-based augmentations yield more consistent results.
- Augmentation depth may influence performance.

Keywords:

- Deep Learning
- Computer Vision
- Data Augmentation
- Medical Instance Segmentation

ABSTRACT:

This study analyzes the effect of photometric data augmentation strategies on medical instance segmentation using the Kvasir-SEG dataset. Mask R-CNN X101-FPN and YOLOv8l-seg are adopted as representative two-stage and one-stage segmentation models. Although data augmentation is commonly employed to enhance robustness in image classification, its systematic evaluation in instance segmentation remains limited due to the need to preserve pixel-level mask integrity during transformations. In this work, eight photometric augmentation techniques, including Hue, Saturation, Grayscale, Brightness, Contrast, Noise, Blur and Cutout, are applied both individually and through structured multi-level combinations. Each augmentation is evaluated within single, double, triple, and full pipelines. Segmentation performance is measured using mean Average Precision (mAP) based on the COCO evaluation protocol. The experimental results show that color-based augmentations provide more reliable accuracy improvements than distortion-based methods in polyp segmentation tasks, while excessive augmentation depth may slow convergence and restrict performance gains. This study presents a systematic analysis of augmentation depth and diversity and offers practical guidance for designing effective augmentation pipelines in medical instance segmentation.

[1*]Tolga TURAY (**Orcid ID:** 0000-0002-1445-296X), Atatürk University, Faculty of Engineering, Department of Electrical and Electronics Engineering, Erzurum, Türkiye.

***Corresponding Author:** Tolga TURAY, e-mail: tolgaturay@atauni.edu.tr

Ethics Committee Approval: Ethics committee approval is not required.

INTRODUCTION

The performance of instance segmentation models is primarily influenced by two key components: the characteristics of the segmentation architecture and the quality of the training data (Lin et al., 2014). From a model-oriented perspective, factors such as network design, feature extraction strategy, and optimization configuration strongly affect segmentation accuracy (He et al., 2017). From a data-oriented perspective, image resolution, visual variability, class imbalance, and the reliability of pixel-level annotations play a crucial role in achieving robust and generalizable segmentation outcomes (Everingham et al., 2010; Mumuni & Mumuni, 2022).

Among data-driven improvement strategies, data augmentation has become one of the most effective tools for enhancing model robustness and mitigating overfitting. Data augmentation increases the diversity of training samples by applying controlled transformations, including spatial modifications and appearance-based variations (Mumuni & Mumuni, 2022). While augmentation techniques have been extensively explored in domains such as natural language processing (Pellicer et al., 2023) and time-series analysis (Iglesias et al., 2023), data augmentation plays a particularly critical role in computer vision because visual models rely heavily on appearance information. Variations in illumination, texture, and contrast between training and test data can directly influence model behavior and lead to performance degradation if not properly addressed (Mumuni & Mumuni, 2022).

Data augmentation methods can be broadly divided into two conceptually distinct categories: data transformation and data synthesis (as illustrated in Figure 1). Data transformation approaches, (Nanni et al., 2021; Mumuni & Mumuni, 2022; Alomar et al., 2023a, 2023b; Kumar et al., 2024b) modify existing images through explicit operations applied at the image or pixel level. These methods include photometric transformations, such as color and intensity adjustments; geometric transformations, which alter spatial structure; and kernel or filter-based operations that emphasize specific visual patterns. In contrast, data synthesis–based augmentation methods (Mumuni et al., 2024) rely on deep learning models to generate new training samples. Data synthesis typically involves generative models that learn the underlying data distribution and produce novel samples rather than altering existing ones (Makhlouf et al., 2023).

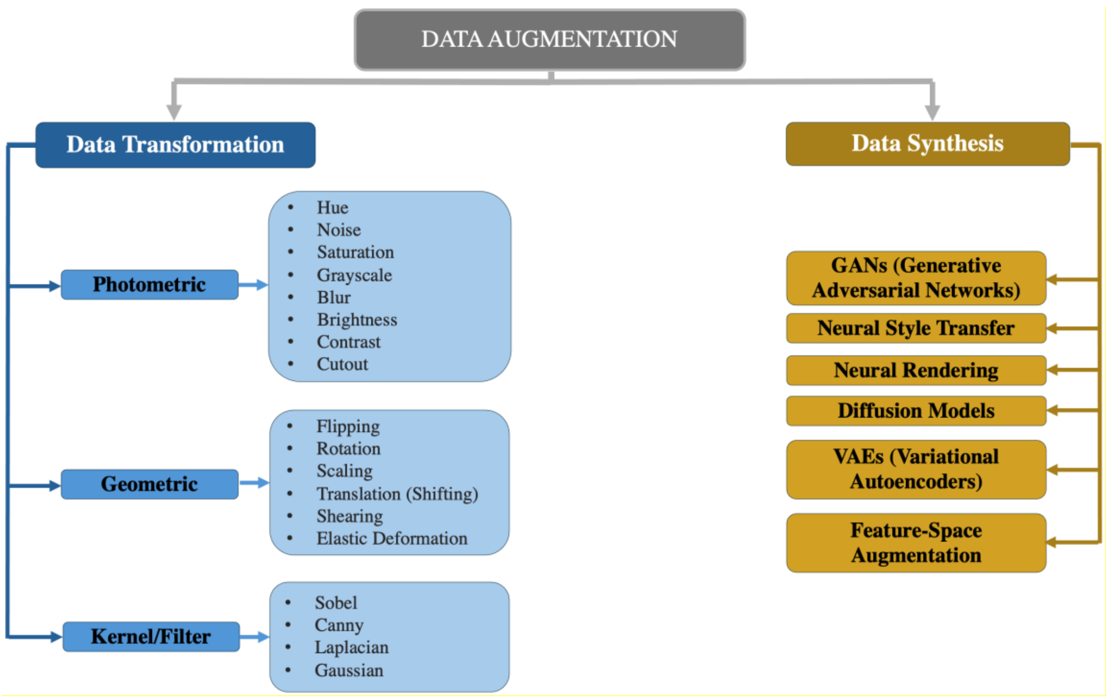


Figure 1. A detailed taxonomy of data augmentation approaches

Despite the widespread adoption of augmentation in computer vision, systematic investigations have predominantly focused on image classification (Alin et al., 2023; Cheung & Yeung, 2024; Gao et al., 2024) rather than instance segmentation. This imbalance stems from the additional complexity of segmentation tasks, where transformations must preserve exact pixel-level alignment between images and their corresponding annotation masks, increasing implementation difficulty and limiting their practical applicability. Consequently, many prior studies (Nanni et al., 2021; Paing et al., 2022; Goceri, 2023; Hwang et al., 2025) evaluate augmentation techniques individually, reporting the impact of isolated transformations without considering their combined effects, while their application in dense prediction tasks remains more challenging due to the need for consistent and accurate updates of pixel-level annotations. In light of these challenges and limitations, a closer examination of existing augmentation strategies in both image classification and instance segmentation domains is necessary to better understand their effectiveness and current research gaps.

In the context of image classification, a comprehensive study in (Nanni et al., 2021) evaluated more than ten augmentation methods, covering both standard transformations and signal-processing-based techniques such as Wavelet and Gabor filtering. The findings showed that applying augmentations individually can effectively reduce overfitting and improve generalization across both medical and natural image datasets. Similarly, another study (Goceri, 2023) explored multiple augmentation strategies on medical images obtained from different organs and imaging modalities, demonstrating that augmentation effectiveness is highly dependent on image characteristics. Both studies primarily focused on individual augmentation methods evaluated in isolation. In contrast, the work presented in (Korzhebin & Egorov, 2021) investigated combinations of augmentation techniques together with various transfer learning strategies across multiple architectures. Their results indicated that the strongest individual augmentations do not necessarily lead to optimal performance when combined, underscoring the importance of systematic analyses that examine interactions among augmentation methods.

In the context of instance segmentation, augmentation has been investigated in several studies; however, the majority of existing work remains limited to the evaluation of individual techniques in isolation. Three distinct examples from the literature are as follows:
(i) A recent instance segmentation study (Hwang et al., 2025) on building dimension estimation from street view images evaluated seven different augmentation techniques, including brightness, contrast, rotation, scaling, shearing, translation, and a single "all-together" configuration. The results showed that certain single augmentations, particularly contrast adjustment, consistently improved performance, whereas applying all augmentations simultaneously rarely yielded further gains and, in some cases, degraded accuracy. Notably, the study did not investigate intermediate augmentation combinations, such as pairwise or triple-method configurations, focusing instead on isolated techniques and a full aggregation strategy. (ii) In another work, (Paing et al., 2022) proposed an instance segmentation framework for multiple myeloma cell detection using Mask R-CNN (He et al., 2017), where data augmentation was incorporated through a deep learning–based strategy referred to as deep-wise augmentation. The study primarily relied on contrast-enhanced images to improve segmentation performance and reported high accuracy metrics. However, the augmentation design was limited in scope, as it focused on a single photometric transformation without exploring a broader set of augmentation techniques or their possible combinations. (iii) Finally, similar to (Paing et al., 2022), (Cerqueira et al., 2024) also applied only contrast-based augmentation to improve segmentation accuracy. Despite providing valuable results, the study employed a single data augmentation method without examining alternative strategies. In summary, the limitations of these studies arise from the lack

of a systematic evaluation of commonly used photometric and geometric augmentation methods, as well as the absence of an analysis of structured combinations of these techniques.

The studies discussed above predominantly rely on offline augmentation strategies or employ limited forms of online augmentation during training (Wagner et al., 2023; Cerqueira et al., 2024). Offline augmentation generates fixed transformed samples prior to training, which increases dataset size but constrains variability since augmented samples remain unchanged throughout the learning process. In contrast, online augmentation applies transformations dynamically and probabilistically during training, introducing greater diversity without expanding dataset size. Prior studies report that offline augmentation is commonly preferred for small datasets, whereas online augmentation tends to be more effective in large-scale training scenarios where dynamic variation improves robustness (Cerqueira et al., 2024; Wagner et al., 2023). However, systematic investigations of online augmentation strategies for medium-sized datasets remain limited, particularly in instance segmentation context.

In the present study, an online augmentation strategy implemented using PyTorch *transform* functions (PyTorch, 2025) is adopted to introduce dynamic and probabilistic variation during training on the Kvasir-SEG dataset (Jha et al., 2019). This setup enables an analysis of augmentation behavior in a medium-scale medical instance segmentation scenario, addressing an open gap in the existing offline/online-oriented literature.

In summary, to the best of our knowledge, there is no comprehensive study that systematically examines both the individual and combined effects of augmentation techniques on instance segmentation performance. Applying all augmentation techniques simultaneously, whether offline or online, is neither practical nor consistently effective. Instead, identifying well-balanced subsets of augmentations is essential for achieving meaningful performance gains. To address these gaps, the present work evaluates eight photometric augmentation methods (Hue, Saturation, Grayscale, Brightness, Contrast, Noise, Blur and Cutout) (Kumar et al., 2024b) applied individually and in multiple combinations. The goal is to provide deeper insight into how online augmentation depth and diversity influence segmentation accuracy and to determine effective augmentation configurations that improve mAP performance.

The remainder of this paper is structured as follows: the experimental protocol is presented in the Materials and Methods section, followed by a detailed description of findings in the Results and Discussion section. The paper concludes with the Conclusion section.

MATERIALS AND METHODS

This study follows a controlled experimental protocol to examine how photometric data augmentation strategies affect instance segmentation performance in a medical imaging setting. An online augmentation framework is adopted to vary augmentation diversity in a systematic manner while keeping the augmentation intensity fixed across all experiments. Eight photometric transformations are evaluated both individually and within predefined multi-augmentation combinations, enabling a structured analysis of augmentation depth under consistent training conditions. All experiments are conducted on the Kvasir-SEG (Jha et al., 2019) dataset using two instance segmentation architectures based on different learning paradigms, ensuring that observed performance differences mainly reflect the influence of augmentation strategies rather than model-specific or optimization-related factors.

Custom Dataset

All experiments are implemented on the Kvasir-SEG dataset (Jha et al., 2019), a publicly available benchmark comprising 1,000 colonoscopy images paired with pixel-level polyp segmentation masks. The annotations are clinically validated and were generated by expert gastroenterologists, making the dataset a standard reference for research on polyp segmentation, detection, and localization. Image resolutions vary between 332×487 and 1920×1072 pixels, and each image is matched with its corresponding mask using an identical file-naming scheme. A representative image–mask pair is shown in Figure 2. For reliable performance evaluation, the dataset is divided into training (80%) and test (20%) sets in accordance with common practice in medical image analysis. The provided masks are used as ground truth throughout all training and evaluation stages in this study.

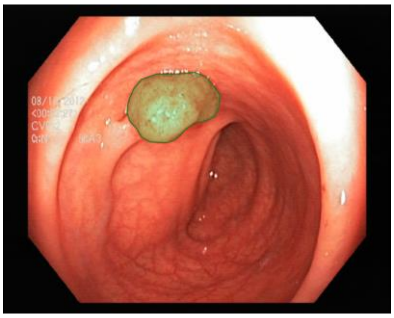


Figure 2. A representative image and corresponding segmentation mask from the Kvasir-SEG (Jha et al., 2019) dataset

Photometric Data Augmentation Techniques and Combination Strategy

Eight photometric data augmentation techniques were selected to systematically investigate their individual and combined impact on medical instance segmentation performance. These transformations were organized into two primary categories, namely color-based and distortion-based augmentations, as summarized in Table 1. Visual examples illustrating the application of each augmentation to a representative Kvasir-SEG image are provided in Figure 3.

Table 1. Overview of the single and multi-augmentation configurations evaluated in this study, grouped according to color-based and distortion-based augmentation categories

ID	Augmentation Combinations	Description
1	**Baseline (B)**	No augmentation
2	Hue (H)	Single color-based augmentation
3	Saturation (S)	Single color-based augmentation
4	Grayscale (GS)	Single color-based augmentation
5	Brightness (BR)	Single color-based augmentation
6	Contrast (C)	Single color-based augmentation
7	Noise (N)	Single distortion-based augmentation
8	Blur (BL)	Single distortion-based augmentation
9	Cutout (CO)	Single distortion-based augmentation
10	Brightness+Contrast (BRC)	Double color-based combination
11	Noise + Blur (NBL)	Double distortion-based combination
12	Brightness+Cutout (BRCO)	Double cross-family combination
13	Noise+Blur+Cutout (NBLCO)	Triple full distortion-based
14	Hue+Saturation+Grayscale+ Brightness+Contrast (HSGSBRC)	Quintuple full color-based
15	All Augmentations Combined (ALL)	Complete set of eight augmentations

Color-based augmentations modify the photometric characteristics of an image while maintaining its spatial structure. This group includes five transformations. Hue (H) alters the hue channel within a constrained range to reflect variability in illumination color and endoscopic lighting. Saturation (S) adjusts color intensity to emulate differences in tissue appearance. Grayscale (GS) converts images to single-channel representations, encouraging the model to focus on structural and textural information rather than chromatic cues. Brightness (BR) modifies overall pixel intensity to simulate underexposed or overexposed acquisition conditions commonly encountered during clinical procedures. Contrast (C) amplifies or attenuates local intensity variations, directly influencing the delineation of polyp boundaries and surrounding mucosal regions.

Distortion-based augmentations introduce stochastic degradations or partial occlusions to improve robustness against real-world imaging artifacts. Noise (N) injects random pixel-level perturbations to mimic sensor noise or compression effects. Blur (BL) applies Gaussian smoothing to replicate motion-induced blur or defocus frequently observed in endoscopic video streams. Cutout (CO) removes rectangular regions of the image, compelling the model to leverage contextual information and remain resilient to incomplete visual evidence.

In addition to single-augmentation experiments, structured multi-augmentation combinations were designed to examine whether interacting transformations yield complementary or adverse effects. As detailed in Table 1, the evaluated configurations span isolated augmentations as well as multiple combinations involving two, three, five, and all eight methods. This hierarchical design enables a controlled analysis of how augmentation diversity and depth influence instance segmentation performance within a unified online augmentation framework.

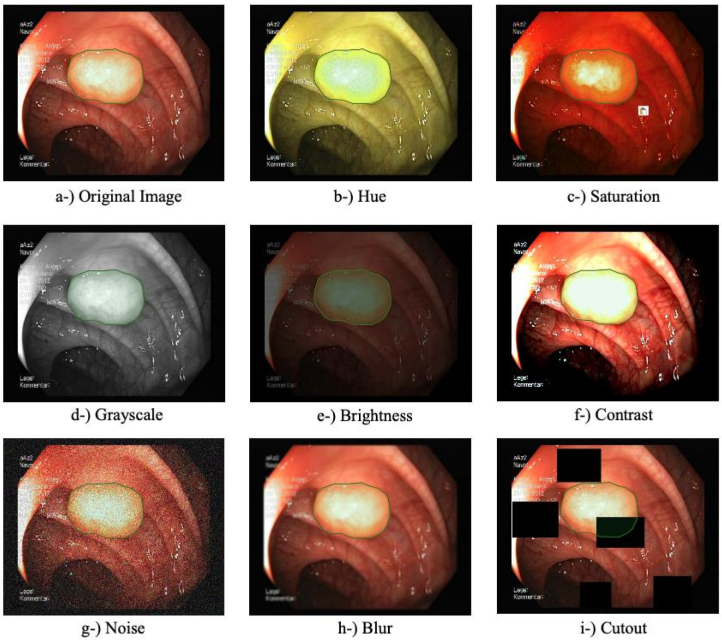


Figure 3. Illustration of the eight photometric augmentation techniques applied to a sample image from the Kvasir-SEG dataset (Jha et al., 2019): (a) original image, followed by (b) hue adjustment, (c) noise injection, (d) saturation modification, (e) grayscale conversion, (f) blurring, (g) brightness adjustment, (h) contrast enhancement, and (i) cutout

Methodological Framework

In contrast to offline augmentation approaches where transformed samples are generated once and stored, this study adopts an online augmentation framework in which transformations are applied dynamically during each training iteration. As a result, every image can be presented in multiple stochastic variants across epochs, increasing data diversity without expanding the dataset size.

Online augmentation is particularly suitable for medium-scale medical datasets such as Kvasir-SEG (Jha et al., 2019), since it continuously introduces new photometric variations and reduces the risk of memorizing fixed augmented samples. For this reason, all eight augmentation methods defined above were integrated into a unified online pipeline, enabling controlled comparisons between single-augmentation and multi-augmentation settings under identical training conditions.

Probability assignment and combination logic

To analyze the influence of augmentation depth and diversity on instance segmentation accuracy, three augmentation probability configurations were defined: (i) In the single-augmentation setting, each transformation was applied independently with a probability of 0.50 ($p=0.50$); (ii) In multi-augmentation configurations including double, triple, and quintuple combinations, the total augmentation intensity was fixed at 0.50 and distributed evenly among the selected methods. For example, in a double combination, each method was applied with a probability of 0.25 ($p = 0.25$), while in a triple combination the probability was approximately 0.17 ($p \approx 0.17$) per method; (iii) In the full eight-method configuration labelled ALL, the same 0.50 total intensity was shared across all transformations, resulting in an individual application probability of approximately 0.06 ($p = \approx 0.06\ (0.50 / 8)$) for each method. All augmentations followed independent Bernoulli (Ross, 2014) trials, meaning that each transformation was activated independently based on its assigned probability.

Although individual methods differed across configurations, the expected number of transformations applied per image remained close to p=0.5. This design ensured fair and balanced comparisons across different augmentation depths. When the batch size was at least sixteen (≥ 16), each training batch statistically preserved the intended augmentation rate (0.50 or 50%) within the online framework.

PyTorch-Based Implementation Framework

The online augmentation pipeline was implemented using the *torchvision.transforms* (PyTorch, 2025) framework. Each augmentation operation was embedded within a *RandomApply* wrapper, allowing an explicit probability to be assigned to each transformation. In practice, an augmentation step follows the general PyTorch structure *RandomApply([AugmentationFunction(parameters)], p=prob)*, enabling precise control over augmentation activation during training.

Under the online augmentation paradigm, every transformation is evaluated independently for each input image. When the assigned probability condition is met, the corresponding operation is applied to the image prior to being fed into the instance segmentation model. This design preserves statistical independence among transformations and ensures a transparent and reproducible augmentation process suitable for controlled experimental analysis.

In this setting, each augmentation is assigned an activation probability of $p = 0.06$, reflecting an equal distribution of the total augmentation intensity across all eight transformations. In contrast, when a single augmentation method is applied in isolation, its activation probability is set to $p = 0.50$ in order to maintain an equivalent overall augmentation strength.

Across all experiments, fixed parameter values were used for each augmentation technique to ensure consistency between single and multi-augmentation scenarios. Hue (H) transformation was implemented using *ColorJitter(hue=0.2)*, introducing controlled shifts in the hue channel while avoiding unrealistic color distortion. Saturation (S) adjustment employed *ColorJitter(saturation=0.3)*, enabling moderate modulation of color intensity without altering spatial content. Grayscale (GS) conversion was realized through *Grayscale(num_output_channels=3)*, preserving the expected three-channel input format while removing chromatic information at the pixel level. Brightness (BR) variation was applied using *ColorJitter(brightness=0.3)* to introduce global intensity changes, whereas contrast (C) adjustment relied on *ColorJitter(contrast=0.3*) to modulate local intensity relationships within the image. These operations were selected to remain within realistic photometric bounds commonly observed in endoscopic imagery.

Distortion-based augmentations were implemented to introduce controlled degradation effects. Gaussian blurring (BL) was applied via *GaussianBlur(kernel_size=5)*, producing mild spatial smoothing without excessive loss of edge information. Additive noise (N) was introduced using a custom *GaussianNoise(std=0.05)* module, which injects low-amplitude perturbations directly in the tensor domain. The cutout (CO) operation was implemented through a custom *RandomErasing(scale=(0.02, 0.15))* module, masking limited regions of the image while constraining the erased area to prevent excessive information removal.

Throughout all augmentation scenarios, these parameter settings were kept constant. Only the activation probabilities varied according to the experimental design, ensuring that differences in performance could be attributed to augmentation composition and depth rather than changes in transformation strength.

Training Configuration and Model Selection

Two state-of-the-art instance segmentation architectures, representing one of the most widely studied tasks in computer vision, were employed. These were Mask R-CNN X101-FPN implemented using (Detectron2, 2025) and YOLOv8l-seg implemented within the Ultralytics framework (YOLOv8, 2025). The selection intentionally includes both a two-stage and a one-stage architecture to ensure a representative methodological spectrum. Both models were trained using an identical training schedule to ensure fair comparisons across augmentation scenarios.

All experiments were conducted for 50 epochs with a batch size of 32. This batch size provided stable gradient estimates, sufficient stochastic variation for online augmentation, and compatibility with the available GPU memory. The initial learning rate was set to 0.001 and decayed by a factor of 0.1 after 30 epochs using a step-based learning rate schedule, which improved convergence stability for both architectures. Weight decay value with 10^{-4} was applied consistently to limit overfitting, and optimization was performed using stochastic gradient descent with a momentum coefficient of 0.9.

All experiments were executed on the High Performance Computing Center of Atatürk University using an 8xNVIDIA A40 GPU cluster with 48 GB VRAM per device. This infrastructure enabled efficient training across all augmentation scenarios and ensured full experimental reproducibility.

RESULTS AND DISCUSSION

During training, model optimization was analyzed through several loss components, including classification loss, bounding box regression loss, mask prediction loss, region proposal network classification loss, and RPN localization loss. The aggregation of these components constitutes the total loss, which was monitored across epochs to assess convergence stability under different augmentation strategies. While total loss provides insight into the optimization dynamics, it does not fully capture segmentation quality on unseen samples. For this reason, segmentation performance was primarily evaluated using mean Average Precision (mAP) metrics.

In line with standard instance segmentation benchmarks, performance is reported using mAP@[0.5:0.95], mAP@0.5, and mAP@0.75 following the COCO evaluation protocol (Lin et al., 2014). These metrics jointly reflect classification accuracy, localization precision, and mask quality, making them more representative of real-world segmentation performance than loss values alone. In contrast to validation loss, which is sensitive to short-term optimization behavior and augmentation-induced regularization effects, mAP offers a more stable and task-relevant indicator of generalization capability. Recent studies have similarly highlighted that reductions in training loss do not necessarily translate into improved segmentation accuracy (Alomar et al., 2023). To ensure consistency and fair comparison across all augmentation scenarios, all evaluations were conducted using the COCO Evaluator framework (Lin et al., 2014; Detectron2, 2025).

Table 2. Segmentation performance of Mask R-CNN X101-FPN and YOLOv8l-seg across single and combined augmentation scenarios

ID	Augmentation Combinations	Mask R-CNN X101-FPN mAP* / mAP@.5 / mAP@.75	YOLOv8l-seg mAP / mAP@.5 / mAP@.75
1	Baseline (B)	0.548 / 0.794 / 0.462	0.562 / 0.773 / 0.521
2	Hue (H)	0.572 / 0.811 / 0.487	0.588 / 0.796 / 0.543
3	Saturation (S)	0.578 / 0.817 / 0.495	0.595 / 0.801 / 0.551
4	Grayscale (GS)	0.565 / 0.805 / 0.476	0.583 / 0.789 / 0.538
5	Brightness (BR)	0.587 / 0.822 / 0.498	0.601 / 0.808 / 0.557
6	Contrast (C)	**0.592** / 0.834 / 0.503	**0.607** / 0.811 / 0.562
7	Noise (N)	0.553 / 0.791 / 0.458	0.568 / 0.779 / 0.516
8	Blur (BL)	0.544 / 0.787 / 0.451	0.561 / 0.782 / 0.509
9	Cutout (CO)	0.559 / 0.796 / 0.468	0.574 / 0.781 / 0.524
10	Brightness+Contrast (BRC)	**0.604** / 0.837 / 0.516	**0.618** / 0.822 / 0.574
11	Noise + Blur (NBL)	0.539 / 0.778 / 0.442	0.552 / 0.765 / 0.497
12	Brightness+Cutout (BRCO)	0.598 / 0.832 / 0.511	0.611 / 0.817 / 0.569
13	Noise+Blur+Cutout (NBLCO)	0.532 / 0.772 / 0.437	0.545 / 0.759 / 0.488
14	Hue+Saturation+Grayscale+ Brightness+Contrast (HSGSBRC)	**0.612** / 0.843 / 0.525	**0.625** / 0.829 / 0.581
15	All Augmentations Combined (ALL)	0.545 / 0.788 / 0.455	0.558 / 0.771 / 0.513

*mAP@[.5, .95]

Impact of Color-Based and Distortion-Based Augmentations on mAP

A structured comparison between color-based and distortion-based augmentation strategies indicates consistent yet moderate trends in how appearance-level and structure-level transformations influence instance segmentation performance on the Kvasir-SEG dataset. As shown in Table 2, appearance-level, color-based augmentations generally provide more stable and consistent improvements for both Mask R-CNN X101-FPN (He et al., 2017) and YOLOv8l-seg (YOLOv8, 2025), whereas structure-level, distortion-based transformations tend to exhibit higher sensitivity and less predictable effects, particularly at stricter IoU thresholds, suggesting a stronger impact on boundary precision and mask integrity.

The primary analysis focuses on mAP (mAP@[.5:.95]), as it provides the most comprehensive and discriminative assessment of instance segmentation performance. Additional results for mAP@.5 and mAP@.75 are reported to support the main findings. While these single-threshold metrics are commonly presented in related studies, they capture performance under less restrictive IoU conditions and thus offer a narrower perspective. In this study, the performance trends identified using mAP@[.5:.95] are consistently mirrored by mAP@.5 and mAP@.75, indicating coherent behavior across evaluation criteria, with single-threshold scores serving as complementary rather than decisive measures.

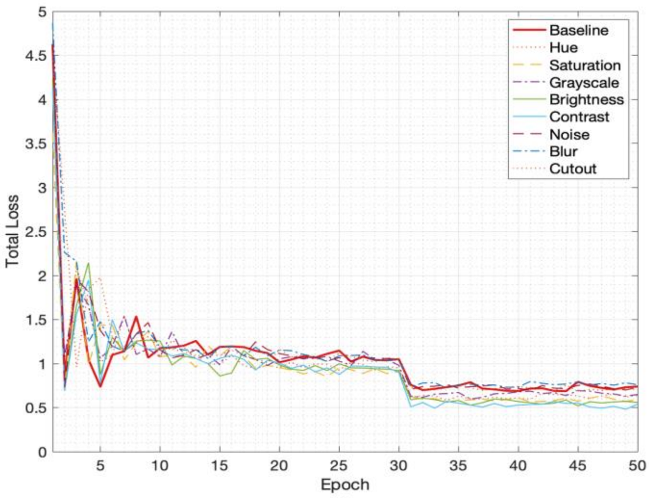
Among the eight individual augmentation techniques, Contrast (C) and Brightness (BR) yield the most pronounced performance gains. For Mask R-CNN, Contrast increases mAP from 0.548 to 0.592, while YOLOv8l-seg shows a comparable improvement from 0.562 to 0.618. These improvements suggest that controlled illumination variations enhance the model's ability to cope with the heterogeneous lighting conditions commonly observed in endoscopic imagery. Saturation (S) and Hue (H) also lead to performance gains, although their impact remains more modest. Overall, these findings indicate that expanding color variability can support generalization, provided that the structural characteristics of the polyps are preserved.

In contrast, distortion-based augmentations demonstrate less consistent behavior and may even degrade segmentation accuracy. Noise (N) and Blur (BL) tend to reduce performance, particularly at stricter IoU thresholds. For example, Blur decreases Mask R-CNN mAP from 0.548 to 0.544 and YOLOv8l-seg from 0.562 to 0.561, with a more noticeable decline at mAP@0.75. Cutout (CO) performs comparatively better within this group, yielding a slight improvement for Mask R-CNN (mAP = 0.559). This outcome may indicate that limited occlusion encourages reliance on contextual information, although excessive masking risks removing clinically relevant texture details and adversely affecting boundary localization.

The effects of augmentation become more evident when multiple transformations are combined. The Brightness+Contrast (BRC) configuration achieves one of the strongest performances among all evaluated settings (mAP = 0.604 for Mask R-CNN and 0.618 for YOLOv8l-seg), surpassing all single-augmentation results. This observation suggests that complementary illumination-related transformations can yield cumulative benefits. In contrast, the Noise+Blur (NBL) combination leads to reduced performance for both segmentors, implying that simultaneous structural distortions may amplify visibility challenges around polyp boundaries. Similarly, the triple combination NBLCO produces the lowest scores, reinforcing the notion that aggressive distortion undermines fine-grained texture information critical for accurate segmentation.

More extensive color-based combinations deliver the most favorable outcomes. The five-method configuration HSGSBRC achieves the highest mAP values for both models (0.612 for Mask R-CNN and 0.625 for YOLOv8l-seg). These results indicate that increased color diversity can substantially improve robustness when structural integrity is maintained throughout training.

Although the ALL configuration applies all eight augmentation methods simultaneously, its performance reverts toward baseline levels. This behavior likely results from the concurrent presence of beneficial color perturbations and disruptive structural distortions, where the latter counteract the advantages of the former. Collectively, these findings reveal that increasing augmentation diversity alone does not guarantee performance gains. Instead, careful selection and balancing of augmentation strategies are necessary to preserve clinically meaningful visual cues.



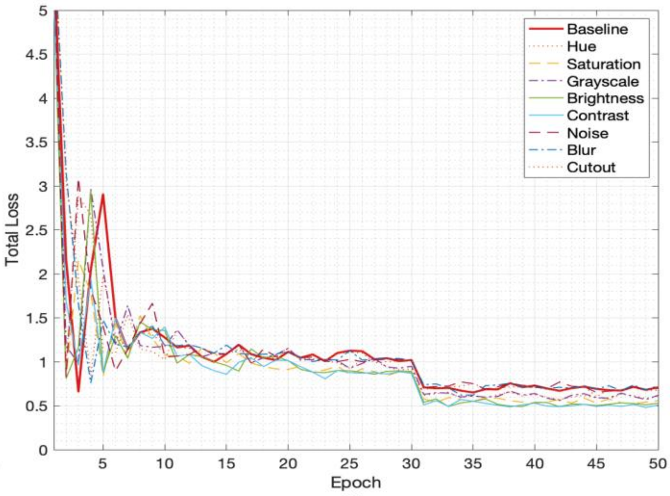


Figure 4. Mask R-CNN: total loss vs. epochs (baseline and single augmentation scenarios) (a) and YOLOv8l-seg: total loss vs. epochs (baseline and single augmentation scenarios) (b)

Convergence Trendss

Training convergence behavior is illustrated through four figures. Figures 4 (a) and (b) present total loss curves for Mask R-CNN and YOLOv8l-seg under baseline and single augmentation conditions, while Figures 5 (a) and (b) extend this analysis to combined augmentation scenarios summarized in Table 1. Together, these visualizations provide insight into how different augmentation strategies influence optimization stability, convergence dynamics, and the resulting performance margins between models.

Training behavior with single augmentation scenarios

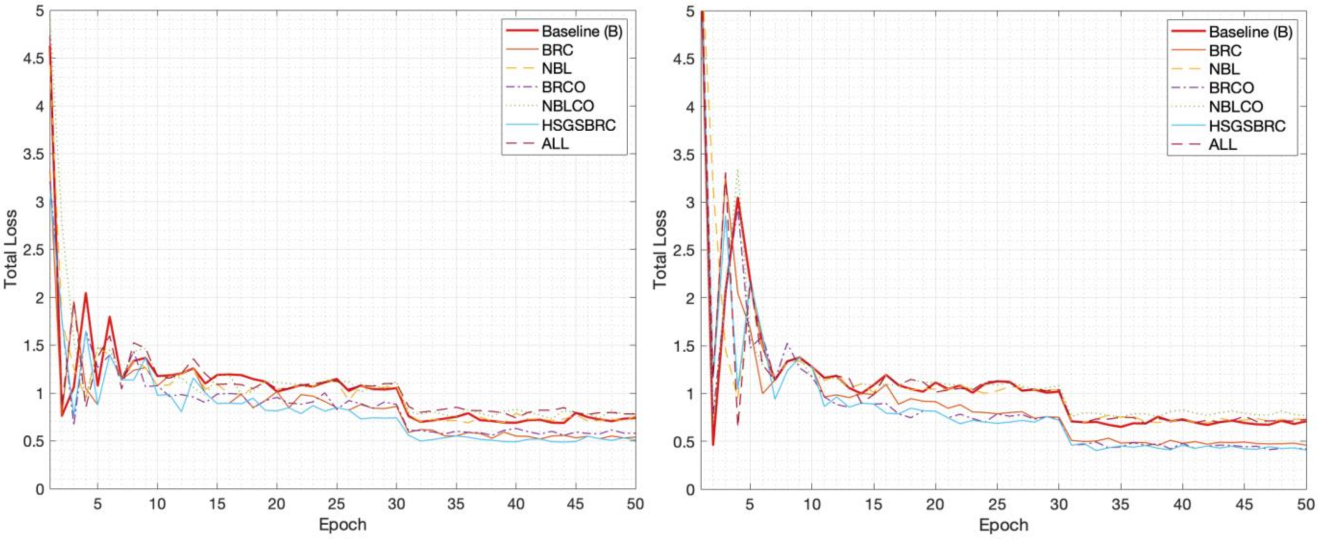
Under single-augmentation conditions, both models exhibit stable and interpretable convergence patterns. As shown in Figures 4 (a) and (b), most augmentation-enhanced training curves converge below the baseline trajectory, which is highlighted by a thicker red line. This observation aligns with the quantitative results in Table 2, where baseline and blur configurations correspond to the lowest mAP values.

YOLOv8l-seg displays more pronounced fluctuations during the initial 10 epochs compared with Mask R-CNN. This behavior is consistent with its one-stage design, where classification and localization are optimized jointly, increasing early sensitivity to augmentation-induced variability. In contrast, Mask R-CNN's two-stage architecture appears to mitigate this effect, producing smoother early convergence.

Augmentations such as Brightness and Contrast achieve the lowest final loss values, which is consistent with their superior mAP performance. This agreement between convergence behavior and segmentation accuracy suggests that certain appearance-level perturbations enhance generalization without compromising structural coherence. Overall, while both models benefit from augmentation, YOLOv8l-seg exhibits greater early volatility, whereas Mask R-CNN demonstrates more stable optimization dynamics.

Training behavior with combined augmentation scenarios

Combined augmentation experiments further clarify the interaction between combined transformations and optimization behavior. In Figure 5 (a), Mask R-CNN shows stable convergence for color-dominant combinations, particularly HSGSBRC, which consistently follows the lowest loss trajectory and corresponds to its strong mAP performance. This trend indicates that structured color-based combinations can introduce beneficial diversity without disrupting essential anatomical cues.



(a) (b)

Figure 5. Mask R-CNN: Total Loss vs. Epochs (Baseline and Combined Augmentation Scenarios) (a) and YOLOv8l-seg: Total Loss vs. Epochs (Baseline and Combined Augmentation Scenarios) (b)

By contrast, distortion-heavy combinations such as NBL, BRCO, and NBLCO converge to higher loss values and yield lower mAP scores, as reported in Table 2. The ALL configuration is especially notable, as its loss curve remains above the baseline throughout training, mirroring its limited accuracy gains. This behavior suggests that indiscriminate application of heterogeneous augmentations may impose excessive regularization and hinder effective optimization.

YOLOv8l-seg exhibits similar trends in Figure 5 (b), though with amplified early-stage volatility, particularly for distortion-dominated combinations such as NBLCO and ALL. While convergence stabilizes in later epochs, final loss values remain higher than those observed for color-based configurations, indicating a less favorable optimization trajectory. This sensitivity is consistent with the model's architectural characteristics and its heightened response to increased input variability, especially under aggressive structural perturbations.

Taken together, these results indicate that moderate online multi-augmentation strategies (especially those focused on color-based transformations) can support improved convergence and segmentation accuracy. In contrast, overly aggressive combinations spanning multiple transformation families tend to yield diminishing or negative returns. Future work may extend this analysis to larger-scale datasets such as MS COCO (Lin et al., 2014) or Pascal VOC (Everingham et al., 2010) to further assess the generality of these findings.

CONCLUSION

This study presents a systematic investigation of how single and combined augmentation strategies affect training dynamics and instance segmentation performance. The results demonstrate that color-based augmentations consistently enhance both convergence behavior and segmentation accuracy, whereas distortion-based transformations must be applied with caution due to their potential to introduce disruptive artifacts. Moreover, extensive augmentation that combines incompatible transformations can impose excessive regularization, limiting convergence and degrading performance.

By evaluating augmentation strategies both individually and in structured combinations, this work highlights the importance of balanced augmentation design. Carefully selected color-based transformations provide the most reliable improvements, while indiscriminate mixing, as in the ALL configuration, may negate potential gains. These findings offer practical guidance for designing augmentation pipelines in medical instance segmentation tasks, where preserving structural fidelity is critical.

Conflict of Interest

The article author declares that there is no conflict of interest between them.

Author's Contributions

The author declares that they have contributed equally to the article.

REFERENCES

Alin, A. Y., Kusrini, & Yuana, K. A. (2023). Data Augmentation Method on Drone Object Detection with YOLOv5 Algorithm. *2023 8th International Conference on Informatics and Computing, ICIC 2023*. https://doi.org/10.1109/ICIC60109.2023.10382123

Alomar, K., Aysel, H. I., & Cai, X. (2023a). Data Augmentation in Classification and Segmentation: A Survey and New Strategies. *Journal of Imaging 2023, Vol. 9, Page 46*, *9*(2), 46. https://doi.org/10.3390/JIMAGING9020046

Alomar, K., Aysel, H. I., & Cai, X. (2023b). Data Augmentation in Classification and Segmentation: A Survey and New Strategies. *Journal of Imaging 2023, Vol. 9, Page 46*, *9*(2), 46. https://doi.org/10.3390/JIMAGING9020046

Cerqueira, V., Santos, M., Roque, L., Baghoussi, Y., & Soares, C. (2024). Online Data Augmentation for Forecasting with Deep Learning. *Lecture Notes in Computer Science*, *16121 LNAI*, 217–229. https://doi.org/10.1007/978-3-032-05176-9_17

Cheung, T. H., & Yeung, D. Y. (2024). A Survey of Automated Data Augmentation for Image Classification: Learning to Compose, Mix, and Generate. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(10), 13185–13205. https://doi.org/10.1109/TNNLS.2023.3282258

Detectron2. (2025). *model_zoo*. https://github.com/facebookresearch/detectron2/blob/main/MODEL_ZOO.md

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, *88*(2), 303–338. https://doi.org/10.1007/S11263-009-0275-4/METRICS

Gao, X., Xiao, Z., & Deng, Z. (2024). High accuracy food image classification via vision transformer with data augmentation and feature augmentation. *Journal of Food Engineering*, *365*, 111833. https://doi.org/10.1016/J.JFOODENG.2023.111833

Goceri, E. (2023). Medical image data augmentation: techniques, comparisons and interpretations. *Artificial Intelligence Review*, *56*(11), 12561–12605. https://doi.org/10.1007/S10462-023-10453-Z/TABLES/10

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, *2017-October*, 2980–2988. https://doi.org/10.1109/ICCV.2017.322

Hwang, D., Kim, J. J., Moon, S., & Wang, S. (2025). Image Augmentation Approaches for Building Dimension Estimation in Street View Images Using Object Detection and Instance Segmentation Based on Deep Learning. *Applied Sciences 2025, Vol. 15, Page 2525*, *15*(5), 2525. https://doi.org/10.3390/APP15052525

Iglesias, G., Talavera, E., González-Prieto, Á., Mozo, A., & Gómez-Canaval, S. (2023). Data Augmentation techniques in time series domain: a survey and taxonomy. *Neural Computing and Applications*, *35*(14), 10123–10145. https://doi.org/10.1007/S00521-023-08459-3/FIGURES/9

Jha, D., Smedsrud, P. H., Riegler, M. A., Halvorsen, P., de Lange, T., Johansen, D., & Johansen, H. D. (2019). Kvasir-SEG: A Segmented Polyp Dataset. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11962 LNCS*, 451–462. https://doi.org/10.1007/978-3-030-37734-2_37

Korzhebin, T. A., & Egorov, A. D. (2021). Comparison of Combinations of Data Augmentation Methods and Transfer Learning Strategies in Image Classification Used in Convolution Deep Neural Networks. *Proceedings of the 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2021*, 479–482. https://doi.org/10.1109/ELCONRUS51938.2021.9396724

Kumar, T., Brennan, R., Mileo, A., & Bendechache, M. (2024a). Image Data Augmentation Approaches: A Comprehensive Survey and Future Directions. *IEEE Access*, *12*, 187536–187571. https://doi.org/10.1109/ACCESS.2024.3470122

Kumar, T., Brennan, R., Mileo, A., & Bendechache, M. (2024b). Image Data Augmentation Approaches: A Comprehensive Survey and Future Directions. *IEEE Access*, *12*, 187536–187571. https://doi.org/10.1109/ACCESS.2024.3470122

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8693 LNCS*(PART 5), 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

Makhlouf, A., Maayah, M., Abughanam, N., & Catal, C. (2023). The use of generative adversarial networks in medical image augmentation. *Neural Computing and Applications*, *35*(34), 24055–24068. https://doi.org/10.1007/S00521-023-09100-Z/FIGURES/7

Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, *16*, 100258. https://doi.org/10.1016/J.ARRAY.2022.100258

Mumuni, A., Mumuni, F., & Gerrar, N. K. (2024). A survey of synthetic data augmentation methods in computer vision. *Machine Intelligence Research*, *21*(5), 831–869. https://doi.org/10.1007/s11633-022-1411-7

Nanni, L., Paci, M., Brahnam, S., & Lumini, A. (2021). Comparison of Different Image Data Augmentation Approaches. *Journal of Imaging 2021, Vol. 7, Page 254*, *7*(12), 254. https://doi.org/10.3390/JIMAGING7120254

Paing, M. P., Sento, A., Bui, T. H., & Pintavirooj, C. (2022). Instance Segmentation of Multiple Myeloma Cells Using Deep-Wise Data Augmentation and Mask R-CNN. *Entropy 2022, Vol. 24, Page 134*, *24*(1), 134. https://doi.org/10.3390/E24010134

Pellicer, L. F. A. O., Ferreira, T. M., & Costa, A. H. R. (2023). Data augmentation techniques in natural language processing. *Applied Soft Computing*, *132*, 109803. https://doi.org/10.1016/J.ASOC.2022.109803

PyTorch. (2025). *torchvision.transforms — Torchvision 0.8.1 documentation*. https://docs.pytorch.org/vision/0.8/transforms.html

Ross, S. M. (2014). Introduction to Probability Models. *Introduction to Probability Models: Eleventh Edition*, 1–767. https://doi.org/10.1016/C2012-0-03564-8

Wagner, F., Eltner, A., & Maas, H. G. (2023). River water segmentation in surveillance camera images: A comparative study of offline and online augmentation using 32 CNNs. *International Journal of Applied Earth Observation and Geoinformation*, *119*, 103305. https://doi.org/10.1016/J.JAG.2023.103305

YOLOv8. (2025). Retrieved December 31, 2025, from https://docs.ultralytics.com/models/yolov8/