

Improving Cervical Precancerous Image Segmentation using Pix2Pix

Marsella Vindriani¹, Muhammad Naufal Rachmatullah², Akhiar Wista Arum³

Department of Computer Engineering, Faculty of Computer Science, Universitas Sriwijaya

**marsellavindriani@gmail.com*

ABSTRACT

Cervical cancer remains a major health concern worldwide, particularly in developing countries where access to early screening and diagnosis is limited. Visual Inspection with Acetic Acid (VIA) is widely used for cervical cancer screening; however, its interpretation may be influenced by medical expertise and image quality. Deep learning-based segmentation methods have shown potential for medical image analysis, but their performance may be constrained by limited training data. This study evaluates a Pix2Pix-based synthetic image augmentation approach for cervical precancerous image segmentation using YOLOv11-seg. A total of 282 cervical precancerous images obtained from General Hospital Muhammad Hoesin, Indonesia and the International Agency for Research on Cancer (IARC) were annotated into three classes: Cervical Area, Columnar Area (CA), and Lesion. Two Pix2Pix configurations, 256×256 and 640×640 pixels, were evaluated using Fréchet Inception Distance (FID). The 640×640 configuration obtained a lower FID score of 25.9 compared with 266.0 for the 256×256 configuration under the implemented FID procedure. Synthetic images generated using the 640×640 configuration were subsequently used as additional training data for YOLOv11-seg. Five YOLOv11-seg variants (n, s, m, l, and x) were evaluated at 640×640 and 1024×1024 input resolutions using baseline and augmented datasets. YOLOv11l at 640×640 was selected as a representative configuration for detailed analysis. On the 66-image validation subset, the representative configuration showed increases in Lesion AP50, IoU, and Dice coefficient of 16.4%, 12.0%, and 14.7%, respectively. These results suggest that Pix2Pix-based augmentation can improve Lesion segmentation performance under the evaluated configuration. However, the findings are based on single-run experiments and a validation subset without an independent test set; therefore, they should not be interpreted as evidence of generalization to an independent patient population.

Keywords: Cervical Precancerous Image, Data Augmentation, Pix2Pix, Generative AI, Segmentation, YOLOv11-seg.

1. INTRODUCTION

Cervical cancer remains a major health concern worldwide, particularly in developing countries where access to healthcare and early screening may be limited[1]. Early detection and appropriate screening are important components of cervical cancer prevention and management[2]. Visual Inspection with Acetic Acid (VIA) is widely used in resource-limited settings because of its relatively simple

implementation and accessibility. However, interpretation of VIA images may be influenced by examiner expertise, image quality, illumination, cervical morphology, and the visual characteristics of abnormal regions[3], [4]. These challenges have motivated the development of artificial intelligence (AI)-based approaches for cervical image analysis.

Deep learning-based segmentation models have demonstrated potential for identifying anatomical structures and abnormal regions in medical images. YOLO-based segmentation models provide an integrated framework for object detection and segmentation and have been applied to various medical image analysis tasks[5]. However, the performance of deep learning models is influenced by the quantity and diversity of training data. Medical image datasets are often limited because of privacy considerations, acquisition costs, annotation requirements, and the need for expert validation [6-8].

Conventional data augmentation techniques, such as flipping, rotation, scaling, and cropping, can increase the variation of training images. However, these operations primarily modify existing images and may not introduce new visual patterns corresponding to different anatomical or pathological appearances[9]. Generative Adversarial Networks (GANs) provide an alternative approach by generating synthetic images that can potentially increase the visual diversity of a training dataset [7], [8].

Pix2Pix is a conditional Generative Adversarial Network (cGAN) designed for paired image-to-image translation[10], [11]. In medical imaging, Pix2Pix can learn the relationship between paired structural representations and corresponding medical images. In this study, segmentation masks are used as conditional inputs and cervical images as target images. This enables the generation of synthetic cervical images corresponding to the spatial structures represented by the segmentation masks.

Previous studies have investigated deep learning-based cervical image segmentation and GAN-based augmentation for medical image analysis. However, the use of paired Pix2Pix augmentation specifically for cervical precancerous image segmentation and its subsequent effect on YOLOv11-seg performance remains insufficiently examined in the dataset considered in this study. In addition, the effect of different Pix2Pix configurations on the generated image feature distribution and subsequent segmentation performance has not been systematically evaluated.

Therefore, this study evaluates two Pix2Pix configurations, 256×256 and 640×640 pixels, and assesses their generated images using FID. Synthetic images generated using the selected configuration are subsequently incorporated into YOLOv11-seg training. Five YOLOv11-seg variants, namely n, s, m, l, and x, are evaluated at 640×640 and 1024×1024 input resolutions using baseline and augmented datasets. The contribution of this study is primarily empirical, namely evaluating whether Pix2Pix-generated synthetic images can provide additional training data that improve cervical precancerous image segmentation, particularly for the Lesion class.

2. MATERIAL AND METHODS

The proposed methodology consists of dataset preparation, data preprocessing and annotation, Pix2Pix-based synthetic image generation, YOLOv11-seg training, and performance evaluation. The overall workflow is illustrated in Figure 1.

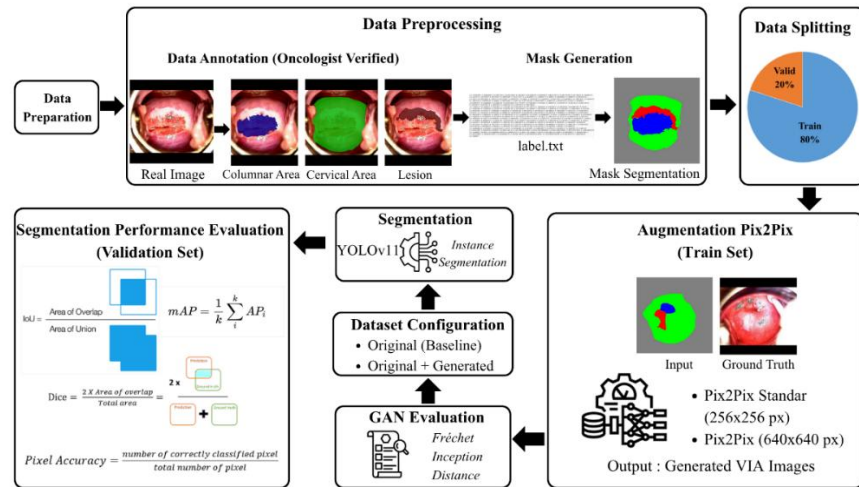


FIGURE 1. Workflow

The processed dataset is then divided into training and validation subsets. Subsequently, Pix2Pix is employed to generate synthetic cervical images to enhance dataset diversity. The generated images are evaluated using Fréchet Inception Distance (FID), and the selected augmentation results are incorporated into the segmentation training process using YOLOv11-seg. Finally, the segmentation performance is assessed using IoU, Dice coefficient, Pixel Accuracy, and AP50.

2.1 DATASET

The research began with the dataset preparation stage using cervical precancerous image datasets obtained from General Hospital Muhammad Hoesin, Indonesia and the International Agency for Research on Cancer (IARC). The dataset consisted of abnormal cervical precancerous images acquired after Visual Inspection with Acetic Acid (VIA) examinations[12].

The dataset comprised 151 images from General Hospital Muhammad Hoesin, Indonesia and 131 images from IARC. A total of 216 images were allocated to the training subset and 66 images to the validation subset. The split was performed at the image level based on the available dataset. No independent test set was used. Because sufficient patient-level metadata were not available, patient-level separation could not be independently verified. Therefore, the possibility that images from the same patient were distributed across different subsets cannot be completely excluded. The source-wise distribution is presented in Table 1.

TABLE 1.
Dataset Distribution

Source	Training Data	Validation Data	Total
General Hospital Muhammad Hoesin, Indonesia	131	20	151
IARC	85	46	131
Total	216	66	282

The 216 training images were used for Pix2Pix training and YOLOv11-seg training. The 66 validation images were reserved for model evaluation and were not used for Pix2Pix training or synthetic image generation.

2.2 DATA PREPROCESSING

The annotation process identified three classes: Cervical Area, Columnar Area (CA), and Lesion. Object annotation was performed using LabelMe, an open-source annotation tool capable of generating annotation coordinates in JSON format. The annotations were converted from JSON format into TXT format required for YOLO-based segmentation. The converted annotations were subsequently transformed into multi-class segmentation masks representing the Cervical Area, CA and Lesion classes at the pixel level.

The IARC dataset had previously been validated by IARC medical experts, while the General Hospital Muhammad Hoesin, Indonesia dataset was re-verified by oncology specialists from the hospital. Examples of the annotation results are shown in Figure 2, while examples of the real image and corresponding segmentation mask are presented in Figure 3.

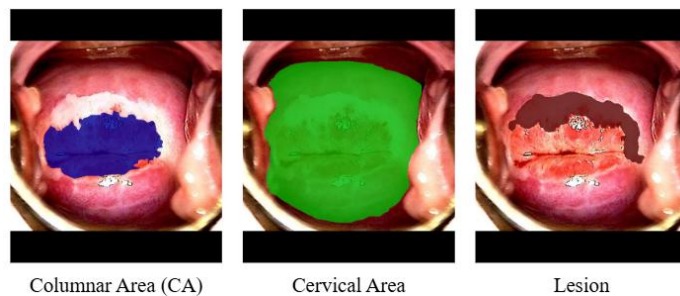


FIGURE 2. Image Annotation

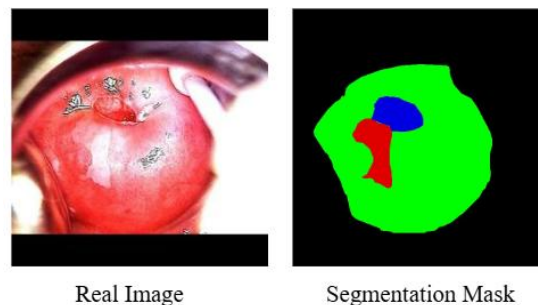


FIGURE 3. Real Image and Corresponding Segmentation Mask

After annotation and mask generation, the complete dataset consisted of 282 images, with 216 images used for training and 66 images reserved for validation. The validation subset was excluded from Pix2Pix training and synthetic image generation.

2.3 PIX2PIX AUGMENTATION MODEL

Pix2Pix is a conditional Generative Adversarial Network designed for paired image-to-image translation[13], [14]. In this study, the segmentation mask was used

as the conditional input and the original cervical image was used as the target image. Therefore, the translation direction was defined as illustrated in Figure 4.

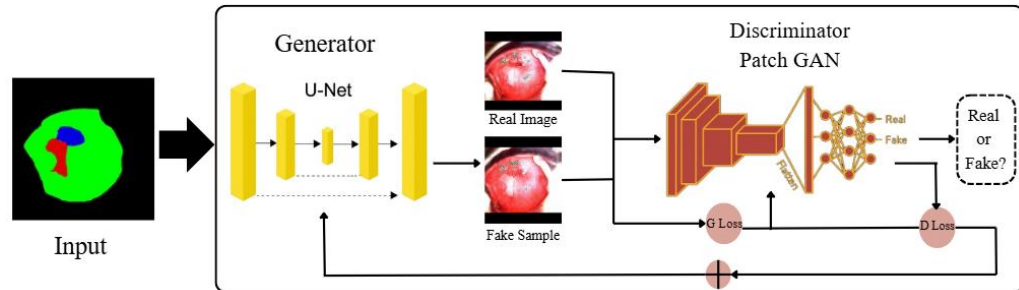


FIGURE 4. Pix2Pix Architecture

The model consisted of a U-Net-based generator and a PatchGAN discriminator. The generator was trained to produce synthetic cervical images from the segmentation masks, while the discriminator distinguished real image-mask pairs from generated image-mask pairs.

The Pix2Pix model was implemented using TensorFlow. The Adam optimizer was used with a learning rate of 2×10^{-4} and $\beta_1 = 0.5$. The batch size was 1, and the model was trained for 60,000 training steps. Input and target images were normalized to the range $[-1, 1]$. The generator loss consisted of adversarial binary cross-entropy loss and L1 reconstruction loss with an L1 loss weight of 100. The discriminator loss was calculated using binary cross-entropy. Model checkpoints were saved every 5,000 training steps.

Only the 216 training images and their corresponding segmentation masks were used for Pix2Pix training. The 66 validation images were excluded from Pix2Pix training and synthetic image generation. A fixed random seed was not explicitly specified in the original Pix2Pix implementation. Therefore, the Pix2Pix results represent a single stochastic training run without deterministic control of all stochastic components. The augmentation configurations are presented in Table 2.

TABLE 2.
Augmentation Configurations

No.	Scenario	Resolution	Preprocessing	Total Synthetic Images
1.	Standard Pix2Pix	256×256	Resize (256→286), Random Crop (286→256)	216
2.	Proposed Pix2Pix	640×640	No additional preprocessing	216

Based on Table 2, the first configuration used a 256×256 resolution with resizing from 256×256 to 286×286 followed by random cropping to 256×256. Random horizontal flipping was also applied during training. The second configuration used a 640×640 resolution without additional geometric preprocessing. Consequently, the comparison between the two configurations represents a comparison of their complete experimental settings rather than an isolated evaluation of image resolution.

2.4 SEGMENTATION MODEL

Five YOLOv11-seg variants, namely Nano (n), Small (s), Medium (m), Large (l), and Extra Large (x), were evaluated at two input resolutions: 640×640 and 1024×1024 . Each backbone variant was trained using both the original dataset and the Pix2Pix-augmented dataset. Thus, a total of 20 segmentation configurations were evaluated, consisting of five backbone variants, two input resolutions, and two dataset conditions.

The 640×640 synthetic images were used directly in both YOLOv11-seg input configurations. No separate 1024×1024 synthetic images were generated. The same augmented dataset containing 216 original and 216 synthetic images was used for both experiments, while the YOLOv11-seg input size was set to 640×640 or 1024×1024 through the `imgsz` training parameter. Accordingly, for the 1024×1024 configuration, the 640×640 synthetic images were automatically processed by the YOLOv11-seg data pipeline to match the specified input size. The segmentation training dataset configurations are presented in Table 3, and the training parameters used in this study are presented in Table 4.

TABLE 3.
Segmentation Training Dataset Configurations

No.	Scenario	Total Data	YOLOv11-seg Input Resolution
1.	Original dataset without augmentation (Baseline)	216	640×640 , 1024×1024
2.	Original dataset combined with synthetic images generated using Pix2Pix	216 original + 216 synthetic	640×640 , 1024×1024

TABLE 4.
Training Parameters

Parameters	640×640	1024×1024
Epochs	100	100
Batch Size	6	1
Optimizer	Adam	Adam
Learning Rate	$1e-3$	$1e-3$

The difference in batch size was caused by GPU memory limitations. A batch size of 6 was used at 640×640 , whereas a batch size of 1 was used at 1024×1024 to avoid GPU out-of-memory errors.

2.5 IMPLEMENTATION AND REPRODUCIBILITY

The Pix2Pix experiments were implemented using TensorFlow 2.20.0, while YOLOv11-seg was implemented using Ultralytics 8.3.228. The YOLOv11-seg experiments used Python 3.10.19 and PyTorch 2.5.1+cu121 on an NVIDIA GeForce RTX 2070 SUPER GPU with 8 GB VRAM. The Ultralytics implementation recorded a random seed of 0. However, each YOLOv11-seg configuration was evaluated using a single training run. Therefore, the study does not report variability across repeated runs, standard deviations, confidence intervals, or statistical significance testing. The

Pix2Pix implementation did not explicitly specify a fixed random seed. Consequently, the Pix2Pix experiments also represent single-run results.

2.6 EVALUATION METRICS

IoU, Dice coefficient, and Pixel Accuracy were calculated separately for each segmentation class. For each class, the metric was calculated for each validation image and then averaged across the validation images. Thus, the reported values represent class-specific mean performance rather than macro- or micro-averaged performance across classes

2.6.1 Fréchet Inception Distance

FID was used to compare the feature distributions of real and generated cervical images. Before feature extraction, all images were converted to RGB and resized to 299×299 pixels, corresponding to the standard spatial input size of InceptionV3. Pixel values were normalized to the range $[-1, 1]$. Features were extracted from the ImageNet-pretrained InceptionV3 network using the 2048-dimensional activations obtained after global average pooling[15]. In this study, FID is interpreted as a measure of feature-distribution similarity between real and generated images. It is not considered a direct measure of clinical realism, diagnostic validity, or clinical usefulness

2.6.2 Intersection over Union

Intersection over Union (IoU) is an evaluation metric used to measure the similarity between segmentation prediction results and ground truth masks. IoU is calculated as the ratio between the overlapping area of the prediction and the ground truth to their total union area[16].

$$IoU = \frac{A \cap B}{A \cup B} \quad (2)$$

In Equation (2), A represents the predicted segmentation area and B represents the ground truth area. The numerator ($A \cap B$) denotes the overlapping region between prediction and ground truth, whereas ($A \cup B$) represents the combined area of both regions.

2.6.3 Dice coefficient

Dice coefficient is a commonly used segmentation evaluation metric that measures the overlap similarity between prediction results and ground truth masks. The Dice coefficient ranges from 0 to 1, where values closer to 1 indicate better segmentation performance[17].

$$Dice\ Coefficient = \frac{2|A \cap B|}{|A| + |B|} \quad (3)$$

In Equation (3), $|A|$ and $|B|$ denote the number of positive pixels in the predicted mask and ground truth mask, respectively, while $2|A \cap B|$ represents twice the number of overlapping pixels between the two masks.

2.6.4 Pixel Accuracy

Pixel Accuracy is an evaluation metric used to measure the proportion of correctly classified pixels in segmentation tasks[18].

$$\text{Pixel Accuracy} = \frac{\text{number of correctly classified pixel}}{\text{total number of pixel}} \quad (4)$$

Although Pixel Accuracy provides a general overview of segmentation performance, this metric may become less representative in imbalanced datasets where majority classes dominate prediction results.

2.6.5 AP50

For AP50, the Average Precision (AP) was calculated separately for each segmentation class at an IoU threshold of 0.50 using the Ultralytics YOLOv11-seg evaluation results. The reported class-specific AP50 values correspond to Cervical Area, CA, and Lesion. The overall mAP50 is calculated as the mean AP50 across the three classes[19], [20].

3. RESULTS AND DISCUSSION

3.1 Pix2Pix Augmentation Results

Both Pix2Pix configurations were trained using the 216-image training subset. The 66 validation images were not used for Pix2Pix training or synthetic image generation. Each configuration generated 216 synthetic images.

3.1.1 Standard Pix2Pix 256×256

The 256×256 configuration used resizing, random cropping, and horizontal flipping during Pix2Pix training. The generated samples preserved the general cervical structures, although some samples showed reduced fine structural and texture information. Figure 5 shows examples of synthetic cervical precancerous images generated using the standard Pix2Pix 256×256 model.

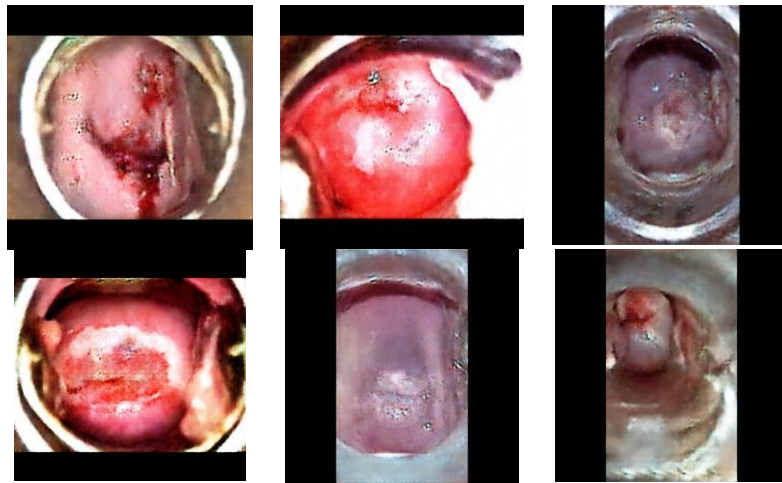


FIGURE 5. Examples of Synthetic Cervical Precancerous Images Generated Using Pix2Pix 256×256

However, several limitations were observed in the generated images. Due to image resizing, fine Lesion details and texture information became less visible. Some Lesion boundaries appeared blurry, and minor texture inconsistencies were observed in several generated samples. Although the generated images remained visually realistic, the lower image resolution reduced the ability of Pix2Pix to preserve detailed Lesion morphology.

3.1.2 Pix2Pix 640×640

The 640×640 configuration was trained without additional geometric preprocessing. The generated samples retained the spatial representation of the corresponding segmentation masks under the evaluated configuration, as shown in Figure 6.

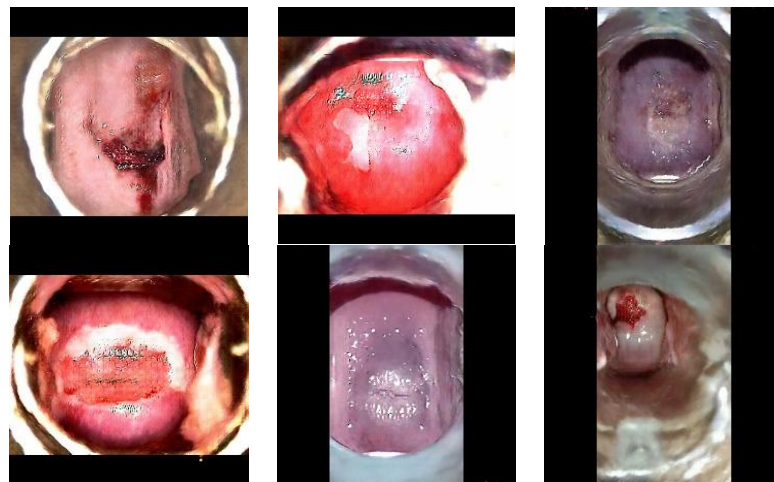


FIGURE 6. Examples of Synthetic Cervical Precancerous Images Generated Using Pix2Pix 640×640

Visual inspection provides qualitative information regarding the generated samples; however, it does not establish clinical validity. Therefore, FID was used as a quantitative measure of feature-distribution similarity.

3.1.3 Comparison of Pix2Pix Configurations

The two configurations produced different FID values. However, they differed in both image resolution and preprocessing. Therefore, the observed difference cannot be attributed exclusively to image resolution. The FID scores for the Pix2Pix-generated images are shown in Table 5.

TABLE 5.
FID Evaluation Results

Augmentation Model	FID Score
Standard Pix2Pix 256x256	266.0
Pix2Pix 640x640	25.9

Table 5 presents the FID evaluation results. The 640×640 configuration produced a lower FID value than the 256×256 configuration. Under the implemented FID procedure, this indicates greater similarity between the feature distributions of the generated and real images. However, FID does not directly establish clinical realism, diagnostic validity, or segmentation usefulness. Furthermore, because the two Pix2Pix configurations used different preprocessing procedures, the result should be interpreted as a comparison of the complete configurations rather than as evidence of an isolated resolution effect.

3.2 Segmentation Performance Evaluation

3.2.1 Segmentation Performance Across YOLOv11-seg Configurations

A total of 20 YOLOv11-seg configurations were evaluated, comprising five backbone variants (n, s, m, l, and x), two input resolutions (640×640 and 1024×1024), and two dataset conditions (baseline and Pix2Pix-augmented). The complete results are presented in Table 6.

TABLE 6.
Segmentation Performance Results

Backbone	Image size	Data	Metric	Cervical Area	CA	Lesion
Nano	640×640	Baseline	AP50	0.711	0.325	0.222
			IoU	0.852	0.394	0.351
			Pixel Accuracy	0.943	0.973	0.930
			Dice Coefficient	0.914	0.505	0.466
		Augmented	AP50	0.665	0.403	0.314
			IoU	0.864	0.416	0.378
			Pixel Accuracy	0.946	0.974	0.934
			Dice Coefficient	0.923	0.532	0.495

Small	1024×1024	Baseline	AP50	0.662	0.209	0.198
			IoU	0.771	0.245	0.130
			Pixel Accuracy	0.907	0.965	0.914
			Dice Coefficient	0.854	0.316	0.178
		Augmented	AP50	0.706	0.216	0.280
			IoU	0.810	0.328	0.254
			Pixel Accuracy	0.923	0.973	0.924
			Dice Coefficient	0.888	0.403	0.326
	640×640	Baseline	AP50	0.729	0.366	0.292
			IoU	0.877	0.429	0.331
			Pixel Accuracy	0.954	0.977	0.934
			Dice Coefficient	0.929	0.551	0.428
		Augmented	AP50	0.823	0.418	0.285
			IoU	0.857	0.399	0.379
			Pixel Accuracy	0.944	0.974	0.937
			Dice Coefficient	0.915	0.498	0.487
	1024×1024	Baseline	AP50	0.748	0.186	0.174
			IoU	0.791	0.315	0.232
			Pixel Accuracy	0.919	0.974	0.918
			Dice Coefficient	0.872	0.395	0.297
		Augmented	AP50	0.769	0.251	0.199
			IoU	0.795	0.308	0.312
			Pixel Accuracy	0.919	0.969	0.920
			Dice Coefficient	0.871	0.387	0.414
	640×640	Baseline	AP50	0.810	0.295	0.297
			IoU	0.826	0.378	0.349
			Pixel Accuracy	0.932	0.975	0.928
			Dice Coefficient	0.894	0.486	0.451
		Augmented	AP50	0.762	0.360	0.270
			IoU	0.878	0.448	0.360
			Pixel Accuracy	0.953	0.977	0.933
			Dice Coefficient	0.931	0.565	0.458
	1024×1024	Baseline	AP50	0.699	0.238	0.150
			IoU	0.798	0.319	0.234
			Pixel Accuracy	0.918	0.970	0.926
			Dice Coefficient	0.880	0.398	0.306
		Augmented	AP50	0.733	0.320	0.203
			IoU	0.836	0.320	0.343
			Pixel Accuracy	0.934	0.977	0.926
			Dice Coefficient	0.905	0.403	0.436
Large	640×640	Baseline	AP50	0.785	0.391	0.299
			IoU	0.808	0.417	0.325
			Pixel Accuracy	0.930	0.974	0.930
			Dice Coefficient	0.881	0.526	0.409
		Augmented	AP50	0.760	0.409	0.348
			IoU	0.868	0.396	0.364
			Pixel Accuracy	0.950	0.975	0.933
			Dice Coefficient	0.924	0.509	0.469

Marsella Vindriani, Muhammad Naufal Rachmatullah, Akhiar Wista Arum
Improving Cervical Precancerous Image Segmentation using Pix2pix

Extra Large	1024×1024	Baseline	AP50	0.577	0.164	0.193
			IoU	0.816	0.344	0.221
			Pixel Accuracy	0.926	0.971	0.915
			Dice Coefficient	0.894	0.438	0.297
		Augmented	AP50	0.669	0.426	0.295
			IoU	0.732	0.368	0.416
			Pixel Accuracy	0.894	0.976	0.931
			Dice Coefficient	0.827	0.454	0.525
	640×640	Baseline	AP50	0.795	0.340	0.325
			IoU	0.767	0.411	0.421
			Pixel Accuracy	0.908	0.972	0.934
			Dice Coefficient	0.853	0.515	0.533
		Augmented	AP50	0.773	0.436	0.364
			IoU	0.845	0.421	0.345
			Pixel Accuracy	0.940	0.976	0.931
			Dice Coefficient	0.908	0.538	0.439
		1024×1024	Baseline	AP50	0.606	0.157
				IoU	0.837	0.214
				Pixel Accuracy	0.936	0.970
				Dice Coefficient	0.906	0.272
			Augmented	AP50	0.656	0.319
				IoU	0.690	0.336
				Pixel Accuracy	0.873	0.975
				Dice Coefficient	0.796	0.421

The results demonstrate that segmentation performance varied across backbone variants, input resolutions, and dataset conditions. No single YOLOv11-seg variant consistently achieved the highest performance across all classes and evaluation metrics. The effect of Pix2Pix augmentation was also configuration-dependent. Several configurations showed improvements in Lesion-related metrics, while other configurations showed decreases in particular classes or metrics. Therefore, the results do not indicate a uniform improvement across all YOLOv11-seg configurations at 640×640 resolution.

3.2.2 Detailed Analysis of YOLOv11l at 640×640

For a more detailed baseline-versus-augmented analysis, the YOLOv11l configuration with 640×640 input was used as a representative configuration. This configuration showed consistent improvements across the three primary metrics after augmentation, namely AP50, IoU, and Dice coefficient. The selection of this configuration was intended to provide a detailed case analysis of the augmentation effect and should not be interpreted as indicating that YOLOv11l was the best-performing configuration across all metrics, classes, and input resolutions as shown in Table 7.

TABLE 7.
Detailed Segmentation Performance of YOLOv11l At 640×640

Metric	Class	Baseline	Augmented
AP50	Cervical Area	0.785	0.760
	CA	0.391	0.409

IoU	Lesion	0.299	0.348
	Cervical Area	0.808	0.868
	CA	0.417	0.396
Pixel Accuracy	Lesion	0.325	0.364
	Cervical Area	0.930	0.950
	CA	0.974	0.975
Dice coefficient	Lesion	0.930	0.933
	Cervical Area	0.881	0.924
	CA	0.526	0.509
	Lesion	0.409	0.469

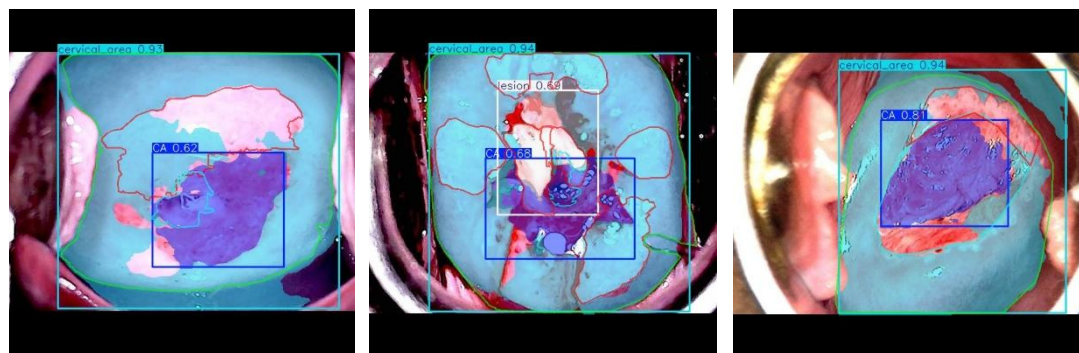
Based on Table 7, the Lesion class showed increases across all four evaluated metrics. Lesion AP50 increased from 0.299 to 0.348, representing a relative increase of 16.4%. IoU increased from 0.325 to 0.364 (+12.0%), Dice coefficient increased from 0.409 to 0.469 (+14.7%), and Pixel Accuracy increased from 0.930 to 0.933 (+0.3%).

The effect on the other classes was mixed. For the Cervical Area class, AP50 decreased from 0.785 to 0.760, while IoU, Pixel Accuracy, and Dice coefficient increased. For the CA class, AP50 increased from 0.391 to 0.409, whereas IoU decreased from 0.417 to 0.396 and Dice coefficient decreased from 0.526 to 0.509. These results indicate that the effect of synthetic augmentation was class-dependent rather than uniformly beneficial across all classes.

The comparison between 640×640 and 1024×1024 configurations did not demonstrate a consistent performance advantage for a single input resolution. Performance varied across backbone variants, classes, and evaluation metrics. Therefore, the higher input resolution cannot be assumed to consistently improve segmentation performance in the present dataset.

The computational requirements also differed between the two resolutions. The 640×640 experiments used a batch size of 6, whereas the 1024×1024 experiments used a batch size of 1 because of GPU memory limitations. This indicates a higher memory requirement for the larger input resolution. Therefore, the choice of input resolution should consider both segmentation performance and computational requirements.

3.2.3 Qualitative Segmentation Results



(a)

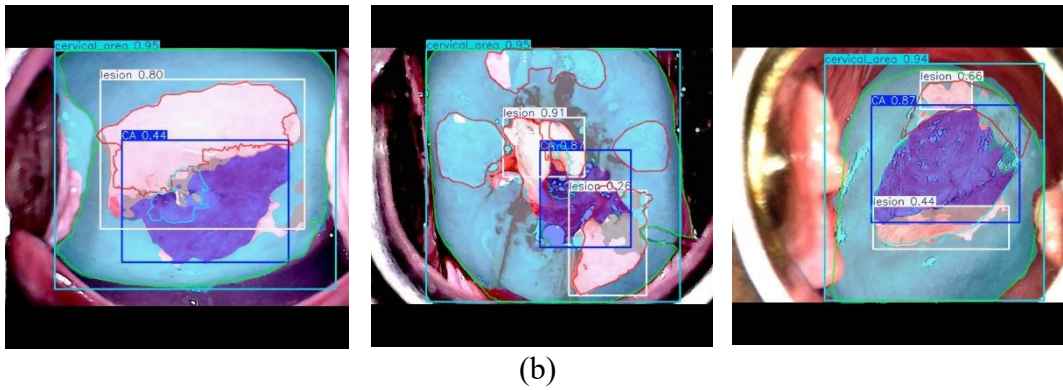


FIGURE 7. Comparison of YOLOv11-seg segmentation results using the (a) baseline dataset and (b) the dataset augmented with Pix2Pix 640-generated synthetic images. The examples illustrate differences in Lesion localization and boundary delineation between the baseline and augmented configurations.

Figure 7 presents qualitative examples of segmentation results obtained using the baseline and Pix2Pix-augmented datasets. In the displayed examples, the augmented configuration shows differences in Lesion localization and boundary delineation compared with the baseline configuration. These observations are consistent with the quantitative improvement observed for the Lesion class in the YOLOv11l 640×640 configuration. However, qualitative examples are illustrative and should not be interpreted as independent evidence of general performance improvement.

3.3 Discussion

The experimental results indicate that Pix2Pix-based synthetic augmentation was associated with improved Lesion segmentation performance in the representative YOLOv11l 640×640 configuration. Lesion AP50, IoU, and Dice increased by 16.4%, 12.0%, and 14.7%, respectively. However, the effect was class-dependent. For CA, AP50 increased from 0.391 to 0.409, while IoU and Dice decreased from 0.417 to 0.396 and 0.526 to 0.509, respectively. This suggests that synthetic augmentation did not improve all segmentation targets uniformly, and the possible differences in class characteristics should be investigated further.

These findings are in line with previous studies exploring deep learning and GAN-based approaches for cervical image analysis and medical image augmentation [5-8]. This study extends these approaches by using paired Pix2Pix augmentation with segmentation masks as conditional inputs and evaluating the generated images as additional training data for YOLOv11-seg. The observed improvement in Lesion segmentation indicates the potential of synthetic augmentation to provide useful visual variation, although its effectiveness may vary across segmentation classes.

The 640×640 Pix2Pix configuration also produced a lower FID (25.9) than the 256×256 configuration (266.0), indicating greater feature-distribution similarity under the implemented FID procedure. However, this difference cannot be attributed solely to resolution because the two configurations used different preprocessing procedures. FID also does not directly establish clinical realism or diagnostic validity.

Across the YOLOv11-seg configurations, augmentation did not consistently improve all models, resolutions, or classes. YOLOv11l at 640×640 was therefore used as a representative configuration for detailed analysis rather than as the overall best-performing model. The 1024×1024 experiments also required a smaller batch size due to GPU memory limitations, indicating higher computational requirements.

3.4 Limitations and Future Work

This study has several limitations. First, no independent test set was used, and patient-level separation could not be independently verified; therefore, the findings do not establish generalization to unseen patients or external datasets. Second, each configuration was evaluated in a single run without repeated experiments or statistical analysis. Third, the study did not compare Pix2Pix with other generative approaches such as CycleGAN, StyleGAN, or diffusion-based models. Finally, the 256×256 and 640×640 Pix2Pix configurations used different preprocessing procedures, limiting the ability to isolate the effect of resolution. Future studies should use larger patient-level separated datasets, repeated experiments, consistent preprocessing, and comparisons with alternative generative approaches.

4. CONCLUSION

The results indicate that Pix2Pix-based synthetic augmentation can improve Lesion segmentation performance in the evaluated validation setting, particularly in the YOLOv11l 640×640 configuration. However, the effect was class-dependent, and the findings are limited by the lack of an independent test set, single-run evaluation, differences in preprocessing, and the absence of comparisons with other generative methods. Further validation is needed to assess the robustness and generalizability of the approach.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

All methods were conducted according to relevant guidelines and regulations. All experimental protocols were approved by General Hospital Muhammad Hoesin, Indonesia. Informed consent was obtained from all subjects and/or their legal guardians. The clinical images were de-identified before analysis. The IARC images were obtained from the IARC Cervical Cancer Image Bank in accordance with its data-use conditions.

ACKNOWLEDGEMENTS

The authors would like to thank the Department of Computer System, Faculty of Computer Science, Universitas Sriwijaya, AIMed CoE Laboratory, and General Hospital Muhammad Hoesin, Indonesia, for supporting this research. The cervical clinical images used in this study were obtained from General Hospital Muhammad Hoesin, Indonesia, and the International Agency for Research on Cancer (IARC).

REFERENCES

- [1] S. Wang *et al.*, “Cervical Cancer Screening via Visual Inspection with Acetic Acid and Lugol Iodine for Triage of HPV-Positive Women,” *JAMA Netw. Open*, vol. 7, no. 3, p. e244090, 2024, doi: 10.1001/jamanetworkopen.2024.4090.
- [2] R. A. Adams and M. H. Botha, “Cervical cancer prevention in Southern Africa: A review of national cervical cancer screening guidelines in the Southern African development community,” *J. Cancer Policy*, vol. 40, p. 100477, 2024, doi: 10.1016/j.jcpo.2024.100477.

- [3] A. Lohiya *et al.*, “Effectiveness of Visual Inspection with Acetic Acid (VIA) Screening on Cervical Cancer Mortality and Incidence, A Systematic Review and Meta-Analysis,” *Asian Pacific J. Cancer Prev.*, vol. 23, no. 2, pp. 399–407, 2022, doi: 10.31557/APJCP.2022.23.2.399.
- [4] K. C. Carr and J. W. Sellors, “Cervical Cancer Screening in Low Resource Settings using Visual Inspection With Acetic Acid,” *J. Midwifery Women’s Heal.*, vol. 49, pp. 329–337, 2004, doi: 10.1016/j.jmwh.2004.02.015.
- [5] S. Nurmaini *et al.*, “Robust assessment of cervical precancerous lesions from pre- and post-acetic acid cervicography by combining deep learning and medical guidelines,” *Informatics Med. Unlocked*, vol. 52, p. 101609, 2025, doi: 10.1016/j.imu.2024.101609.
- [6] S. Nurmaini *et al.*, “Real time mobile AI-assisted cervicography interpretation system,” *Informatics Med. Unlocked*, vol. 42, p. 101360, 2023, doi: 10.1016/j.imu.2023.101360.
- [7] Y. Chen *et al.*, “Generative Adversarial Networks in Medical Image augmentation: A review,” *Comput. Biol. Med.*, vol. 144, p. 105382, 2022, doi: 10.1016/j.compbiomed.2022.105382.
- [8] H. C. Park, I. P. Hong, S. Poudel, and C. Choi, “Data Augmentation Based on Generative Adversarial Networks for Endoscopic Image Classification,” *IEEE Access*, vol. 11, pp. 49216–49225, 2023.
- [9] W. Zhang and S. Ji, “Rehabilitation Driven Optimized YOLOv11 Model for Medical X-Ray Fracture Detection,” *Sensors*, vol. 25, p. 5793, 2025, doi: 10.3390/s25185793.
- [10] J. Henry, T. Natalie, and D. Madsen, “Pix2pix GAN for Image-to-Image Translation,” vol. 10, p. 13140, 2025, doi: 10.13140/RG.2.2.32286.66887.
- [11] A. Van Der Schot, E. Sikkil, M. Spaanderman, and G. De Jong, “Placental Vessel Segmentation Using Pix2pix Compared to U-Net,” *J. Imaging*, vol. 9, no. 10, p. 226, 2023, doi: 10.3390/jimaging9100226.
- [12] International Agency for Research on Cancer, “IARC Cervical Cancer Image Bank,” International Agency for Research on Cancer (IARC). Accessed: Aug. 10, 2025. [Online]. Available: <https://screening.iarc.fr/cervicalimagebank.php>
- [13] S. Chen, D. Gao, L. Wang, and Y. Zhang, “Cervical Cancer Single Cell Image Data Augmentation Using Residual Condition Generative Adversarial Networks,” *2020 3rd Int. Conf. Artif. Intell. Big Data, ICAIBD 2020*, pp. 237–241, 2020.
- [14] E. Rasywir and S. Nurmaini, “Robust SegFormer-Based Cervical Neoplastic Lesion Segmentation Using Lesion-Specific Augmentation and Anatomical Label Optimization,” *IEEE Access*, vol. 13, pp. 205915–205931, 2025.
- [15] M. J. Chong and D. Forsyth, “Effectively unbiased FID and inception score and where to find them,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 6069–6078, 2020.
- [16] W. Zhong *et al.*, “A Classification Method for the Severity of Aloe Anthracnose Based on the Improved YOLOv11-seg,” *Agronomy*, vol. 15, p. 1896, 2025, doi: 10.3390/agronomy15081896.
- [17] K. K. Brar *et al.*, “Image segmentation review : Theoretical background and recent advances,” *Inf. Fusion*, vol. 114, p. 102608, 2025, doi: 10.1016/j.inffus.2024.102608.
- [18] Y. Jiang and Y. Deng, “Self-Attention Driven Image-to-image Translation with Conditional Adversarial Networks,” in *CSAI ’23: Proceedings of the 2023 7th International Conference on Computer Science and Artificial Intelligence*, 2024,

pp. 239–245.

- [19] A. Šidlauskas, A. Kriščiūnas, and D. Čalnerytė, “Continuous Satellite Image Generation from Standard Layer Maps Using Conditional Generative Adversarial Networks,” *ISPRS Int. J. Geo-Information*, vol. 13, no. 12, p. 448, 2024.
- [20] S. Nurmaini *et al.*, “Accurate detection of septal defects with fetal ultrasonography images using deep learning-based multiclass instance segmentation,” *IEEE Access*, vol. 8, pp. 196160–196174, 2020.