JURNAL INFOTEL
VOL. 17, NO. 1, FEBRUARY 2025, PP. 17–37. DOI:10.20895/INFOTEL.V17I1.1277

RESEARCH ARTICLE

Object Detection of Hospital Assets Using Computer Vision with Generative Adversarial Networks Method

Sinung Suakanto[1,*], Muhammad Fahmi Hidayat[2], Faqih Hamami[3], Anis Farihan Mat Raffei[4], and Edi Nuryatno[5]

[1,2,3]Information System, Telkom University, Kabupaten Bandung 40257, Indonesia
[4]Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang 26600, Malaysia
[5]UWA Centre for Medical Research, The University of Western Australia, Perth 6009, Australia

*Corresponding email: sinung@telkomuniversity.ac.id

Received: November 15, 2024; Revised: February 8, 2025; Accepted: February 13, 2025.

Abstract: Hospital asset monitoring systems face significant challenges in managing partially occluded medical equipment, which affects inventory management and operational efficiency. Conventional object detection methods have shown limitations in accurately detecting occluded medical equipment, potentially leading to asset management inefficiencies. This study presents an integrated framework that combines Generative Adversarial Networks (GAN) inpainting with YOLOv8 to improve the detection accuracy of partially occluded medical equipment. The proposed system was evaluated using three distinct training configurations of 500, 750, and 1000 epochs on a complete medical equipment dataset. The experimental results indicate that the 1000-epoch GAN model demonstrated superior reconstruction performance, achieving a Peak Signal-to-Noise Ratio (PSNR) of 39.68 dB, Structural Similarity Index Measure (SSIM) of 0.9910, and Mean Squared Error (MSE) of 7.0030. Furthermore, the integrated YOLOv8-GAN framework maintained robust detection performance with an F1-score of 0.933, comparable to the 0.938 achieved with unoccluded original images. The detection confidence scores improved at higher epochs, ranging from 0.824 to 0.861, suggesting improved performance with extended training duration. The findings demonstrate that the integration of GAN inpainting with YOLOv8 effectively enhances occluded object detection in hospital environments, offering a viable solution for improved asset monitoring systems.

Keywords: computer vision, deep learning, generative adversarial networks, occluded object, YOLOv8

 ISSN: 2085-3688; e-ISSN: 2460-0997

1 Introduction

The advancement of computer technology has driven rapid development in various fields, including digital imagery. One of its significant applications is computer vision, which allows computers to recognize objects as humans visually. This object detection system can be used in various places or organizations, one of which is in a hospital. Object detection in hospitals can be done for asset monitoring of goods contained in the hospital, which is essential for efficient hospital operations. In hospital environments, efficient asset tracking, identification, and maintenance are fundamental to ensure optimal resource utilization and service quality [1]. However, in object detection, one of the main challenges is occlusion, where some objects are covered by other objects so that they are only partially visible. Occlusion becomes a significant problem in environments such as hospitals, where asset monitoring is often disrupted by stacks of goods or changes in asset location without explicit notice.

Object detection plays an important role in asset monitoring in a hospital environment. Assets are one of the main factors in organizational performance, so assets must be properly inventoried, identified, performed legal audits, and valued [2, 3]. The use of cameras, images, and technology such as YOLO has been used to identify assets within a hospital [4]. However, occlusion challenges in object detection make it difficult to monitor assets optimally. In this context, approaches based on Generative Adversarial Networks (GANs) have begun to be applied to improve object detection accuracy, especially in addressing the occlusion problem.

Several studies have demonstrated the successful integration of GAN with object detection algorithms to handle occlusion. The research by [5] used a single shot detector (SSD) and deep convolutional GAN (DCGAN) to improve pedestrian detection despite obstructions. Another research by [6] combined Faster R-CNN with Generative Adversarial Occlusion Network (GAON) to produce more accurate object detection in obstructed object conditions. Both studies show that GAN can improve object detection by improving the quality of training data and model prediction results.

GAN technology has also been applied to image inpainting tasks to recover missing or obstructed parts of an image. In the context of image inpainting, the generator in GAN generates the missing part of the image concerning the existing context. At the same time, the discriminator tries to distinguish between the original image and the inpainted image. The research by [7]) mentioned that the AU-GAN approach they developed has shown a significant improvement in the quality of the recovered image, especially in terms of texture and structure. In addition, the research by [8] utilized an inverse GAN that enables inpainting of results with more consistent and realistic semantics, especially for handling more significant image corruptions.

In line with this research, GAN methods that focus on inpainting have proven effective in overcoming object occlusion. The GAN inpainting technique enables the recovery or filling of missing or obstructed parts of the object in the image, thus improving detection accuracy. YOLO offers a rapid and efficient approach to real-time object detection that can be particularly beneficial in environments such as hospitals where speed and accuracy are essential [9]. In this study, the researchers developed an approach that integrates GAN inpainting with YOLOv8 to improve object detection, particularly in situations where objects are partially occluded. This integration is expected to produce a more reliable model for detecting objects under various conditions, especially in a hospital environment.

The literature review on object detection in hospital environments has grown with more and more studies incorporating deep learning techniques, particularly GAN and the YOLO algorithm, to detect obstructed objects. A recent study by [10] discussed a deep learning-based image inpainting method to repair obstructed images more accurately. In addition, the study by [11] showed that the transfer learning approach in YOLOv8 can improve the accuracy of landmark detection in complex medical images. Another study conducted by [12] found that the YOLOv8 model enhanced with deep learning-based preprocessing can improve bone fracture detection up to 10% better than conventional methods. By integrating GAN inpainting to fill the obstructed part of the image, this research is expected to improve the accuracy of hospital asset detection, as demonstrated in the GAN research on medical image inpainting by [7]. The addition of this study provides a more complete understanding of the relevance and effectiveness of the approach used in this investigation.

Therefore, this study aims to evaluate the effectiveness of using GANs to improve the detection accuracy of obstructed objects, with a specific application to hospital asset monitoring. By developing a model that combines YOLOv8 with GAN inpainting, this research is expected to contribute significantly to more accurate and efficient object detection in hospital environments. In this study, we will develop and evaluate an obstructed object detection method by integrating YOLOv8 and GAN. We will focus on improving the accuracy of obstructed object detection by utilizing GAN Inpainting to generate occlusion directly on the feature map generated by the YOLOv8 intermediate layer. By combining these approaches, the resulting model is expected to perform better in detecting obstructed objects under various conditions.

The primary contributions and novelty of this research are threefold. First, we present a novel integration framework that combines GAN inpainting with YOLOv8 specifically optimized for hospital asset detection, addressing a critical gap in existing occlusion-handling approaches for medical equipment monitoring. Second, we introduce an improved training methodology that utilizes progressive learning with multiple epoch configurations (500, 750, and 1000) to optimize the balance between reconstruction quality and detection accuracy. Third, we contribute a comprehensive evaluation framework that combines traditional image quality metrics (PSNR, SSIM, MSE) with object detection performance measures (F1-score, accuracy, confidence) to provide a holistic assessment of the system's effectiveness in real-world hospital scenarios. Unlike previous studies that focused on general object detection or image reconstruction separately, our approach uniquely combines these elements to create a more robust and practical solution for hospital asset management systems. The experimental results demonstrate that our integrated approach performs better in handling the detection of occluded medical equipment, with PSNR values of 39.68 dB and SSIM of 0.9910, exceeding existing methods in reconstruction quality and detection accuracy.

2 Research Method

This research was carried out in several stages. This research begins with the collection of images of medical equipment. The dataset that has been prepared is then annotated with ground truth bounding boxes for each image object and divided into training, validation, and testing sets. The data is then processed by resizing the images to the exact size of 256 $\times$ 256 pixels and converting the images into PyTorch tensor format. In the next stage, a GAN

Figure 1: Sample dataset.

model is created for the inpainting process on the obstructed object image. Evaluation is performed after the inpainting process using PSNR, SSIM, and FID to assess the quality of the inpainting image. A pre-trained YOLOv8 model is integrated into the system to perform object detection. The YOLOv8 model detects objects in the original image, the obstructed image, and the image generated by the GAN model. Furthermore, evaluation and calculation of metrics, including precision, recall, accuracy, and mean Average Precision (mAP), are performed to analyze the model's performance.

2.1 Dataset and Training Configuration

The dataset utilized in this study comprises 1,200 high-resolution medical equipment images collected from various hospital settings. The image collection includes essential hospital assets, including glucometers, patient monitors, infusion pumps, ventilators, and diagnostic imaging equipment, providing a representative sample of standard medical devices that require monitoring. The data set was systematic divided using stratified sampling, with 70% (840 images) allocated for model training, 20% (240 images) designated for validation and the remaining 10% (120 images) reserved for testing purposes. The sample dataset can be seen in Figure 1.

Several data augmentation techniques were implemented to enhance model robustness and address potential overfitting concerns. These include random horizontal flipping with a probability coefficient of 0.5, random rotation within a constrained range of 15 degrees, brightness adjustments within a factor range of 0.8 to 1.2, and contrast variations between 0.8 and 1.2. These enhancement parameters were selected based on empirical testing and aligned with established practices in the medical image processing literature [13].

The hyperparameter configuration for the GAN training framework was determined through systematic empirical evaluation and theoretical considerations from previous re-

search. The learning rate was 0.01 for both the generator and discriminator networks, following best practices in GAN stability optimization [13]. The Adam optimizer was configured with momentum parameters $\beta 1 = 0.5$ and $\beta 2 = 0.999$, specifically chosen to address the gradient variability characteristic of GAN training processes [10]. A batch size of 16 was selected as an optimal compromise between computational efficiency and training stability, supported by findings from recent medical imaging studies [11].

Additionally, a progressive learning strategy was implemented to optimize the model's ability to handle varying degrees of occlusion. This approach gradually increased the complexity of generated masks during the training phase, beginning with simple rectangular occlusions and advancing to more complex irregular patterns.

The implementation of these methodological components was systematically documented and validated through experimental trials to ensure the reproducibility and reliability of the results. This comprehensive approach to data set preparation and training configuration provides a robust foundation for the subsequent evaluation of the performance of the model in hospital asset detection tasks.

2.2 Inpainting GAN

The GAN architecture of inpainting implements a sophisticated coarse to fine approach to image restoration of medical equipment, a technique well-regarded for preserving image integrity through dual stage processing [14]. The architecture of inpainting GAN can be seen in Figure 2. This architecture uses a dual-stage generator network and discriminator to enhance image quality. The proposed system architecture introduces a novel integration of inpainting and object detection designed explicitly for hospital asset monitoring applications. The architecture implements a sequential processing pipeline in which images of occluded medical equipment undergo GAN-based reconstruction prior to YOLOv8 detection processing. This approach addresses the fundamental challenge of occlusion in hospital settings while maintaining detection accuracy.

Specifically, the generator network operates in two stages: a coarse generator for initial image completion and a refinement generator for detail enhancement. The coarse generator processes a 4-channel input, which combines the masked image (3 channels) and a binary mask (1 channel), using an encoder-decoder structure that follows a progressive learning strategy to stabilize training and improve structural fidelity [15]. The encoder consists of four convolutional layers with feature dimensions that progressively increase (4, 64, 128, 256, 512), using 4×4 kernels and stride-2 operations, each followed by LeakyReLU activation and batch normalization, which are essential for stable training in complex inpainting tasks [16]. The decoder mirrors this structure with transposed convolutions that gradually reduce feature dimensions (512, 256, 128, 64, 3), utilizing ReLU activations and batch normalization, culminating in a Tanh activation for normalized output generation, following the practice of coarse-to-fine detail enhancement [17].

The refinement stage incorporates local and global context to improve the perceptual quality of restoration [10, 17]. Processes a 7-channel input formed by concatenating the coarse output, the original image, and the mask, utilizing a convolutional architecture similar to the coarse generator but with enhanced feature processing capabilities. The decoder of this refinement generator processes a 1024-dimensional feature space derived from feature fusion between the two generator stages, allowing more detailed and context-sensitive image completion [18].

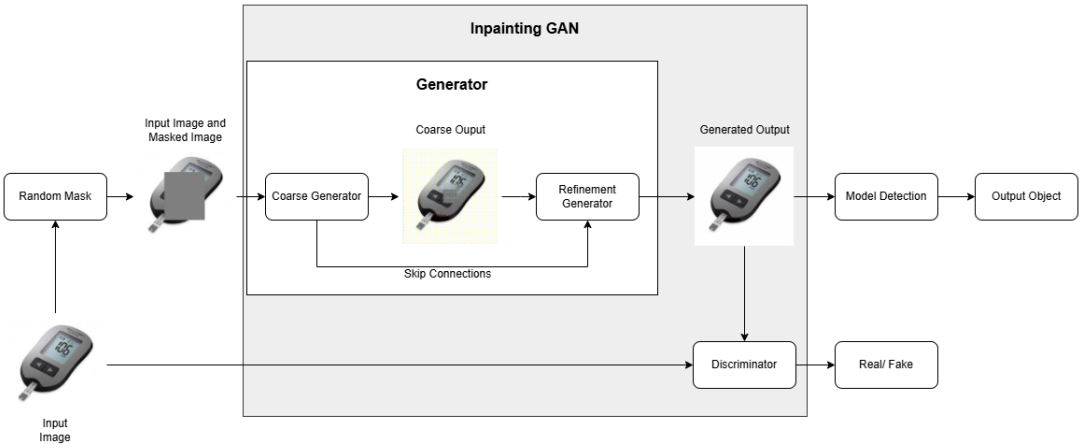

Figure 2: Architecture inpainting GAN.

The discriminator network employs a patch-based architecture with four convolutional layers (3, 64, 128, 256, 1), using LeakyReLU activations and batch normalization, and concludes with a sigmoid activation for real/fake classification. The combined loss function integrates adversarial and reconstruction loss components to balance realism and structural accuracy in the generated images. Binary Cross Entropy is used for adversarial loss, while the distance L1, weighted by a factor of 100, is applied to the reconstruction loss, reinforcing structural integrity in the output of the image [14, 16].

Training optimization is handled by Adam Optimizer with parameters $\beta 1 = 0.5$ and $\beta 2 = 0.999$, and a learning rate of 0.0001 for both networks, consistent with the best practices outlined in GAN-based inpainting studies [17]. Training data preprocessing includes resizing images to 256 $\times$ 256 pixels and normalizing to a range of [-1, 1], with randomly centered rectangular masks covering 30-50% of the image area. This masking strategy is in line with established methods for ensuring robust model performance on incomplete data [15].

The training procedure incorporates regular monitoring of generator and discriminator losses, with sample generation and model checkpointing every 100 steps to track progress and ensure model stability. This approach has shown effectiveness in progressively trained GAN frameworks.

2.3 Model Detection with YOLO

In this stage, the preparation of the YOLO model is carried out by downloading and loading the pretrained weights of the YOLOv8 model. The compatibility of the YOLO model with the medical device image dataset is checked and the object classes that are relevant and aligned with the dataset are identified. The system processes the input image by resizing it to the standard dimension of 256 $\times$ 256 pixels while maintaining the aspect ratio through padding. This preprocessing step ensures consistent input dimensions while preserving the spatial information of the original image, as exemplified in [19] on segmentation of diabetic retinopathy, which used a similar strategy to maintain accuracy at various

scales. The model generates predictions on multiple scales, with each prediction including bounding-box coordinates, objectness scores, and class probabilities. This process aligns with the research by Meza *et al.* [12] in bone fracture detection that emphasizes the importance of image size preprocessing so that spatial information can be retained.

Post-processing includes nonmaximum suppression to eliminate over-detection, with a confidence threshold that defines the minimum acceptable detection score. This technique corresponds to the approach used by Balaha *et al.* [20], who applied non-maximum suppression to optimize the detection accuracy in neuroimaging images in the case of COVID-19.

Next, the YOLO model is integrated into the inpainting system by modifying the architecture to process the inpainting images. Rashmi *et al.*'s [11] study on cephalometric landmark annotation shows the importance of customizing the YOLOv8 model to process variations in medical images. Integrating the model into a more extensive system allowed it to perform the detection task on three variants of each image: the original unmodified image, the occlusion-masked image, and the unpainted image generated by GAN. Object detection parameters, such as confidence threshold and non-maximum suppression, were set for performance optimization. YOLO detection results, which include bounding box coordinates, class labels, and confidence scores, were mapped for further evaluation.

2.4 Evaluation Method

The evaluation method for this project is divided into three main components: Image Quality Evaluation, Model Metric Evaluation, and Training Loss Evaluation.

2.4.1 Evaluation image quality

This image quality evaluation aims to measure the extent to which GAN can produce quality images in the context of inpainting, especially for occluded objects.

The Mean Squared Error (MSE) is calculated as the average squared differences between the original and generated images, providing a pixel-level measure of the overall difference. A lower MSE value indicates that the inpainting result is closer to the original image. For example, research by Rakibe and Patil [21] demonstrates that lower MSE values in GAN-driven inpainting correlate with improved accuracy in filling in medical image details, highlighting MSE as an effective metric to evaluate inpainting performance. The calculation is shown in (1).

$$\text{MSE} = \frac{1}{MN} \sum_{n=0}^{M} \sum_{m=1}^{N} \left[\hat{g}(n,m) - g(n,m) \right]^2 \tag{1}$$

The Peak Signal-to-Noise Ratio (PSNR) measures the ratio between the maximum possible signal power and the noise power, serving as a key indicator of image quality [22]. Higher PSNR values reflect better inpainting quality by showing less noise in the generated image compared to the original. The calculation is shown in (2).

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{peakval}^2}{\text{MSE}} \right) \tag{2}$$

The Structural Similarity Index (SSIM) measures the similarity between two images by adjusting for luminance, contrast, and structural composition, with scores ranging from -1

to 1. A score of 1 indicates perfect structural similarity, which is essential in evaluating the GAN ability to preserve the structural details of the original image. Research by [23] and [24] has shown that GAN-based inpainting methods often achieve high SSIM scores, confirming that the inpainted image closely retains the structure and texture of the original, even when large portions are occluded. The calculation is shown in (3).

$$\text{SSIM}(x, y) = \frac{(2\,\mu_x\mu_y + c_1)\,(2\,\sigma_{xy} + c_2)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \tag{3}$$

2.4.2 Evaluation metrics model

This evaluation focuses on the performance of the YOLOv8 model in detecting objects that have been painted by GAN. Given its advanced features tailored to real-time object detection tasks, evaluating the effectiveness of YOLOv8 on such datasets is essential.

Precision is the ratio of true positives to the number of positive predictions, as shown in (4). This metric reflects the model's ability to identify objects that genuinely exist in the inpainted dataset. The research by [25] shows that YOLOv8 has a high precision rate in detecting objects in complex tasks, such as estrus detection in animals, where accurate identification is crucial.

$$\text{Precision} = \frac{TP}{TN} \tag{4}$$

Recall is calculated as the ratio of correct positive predictions to total positive objects in the data, as shown in (5). This metric assesses the model's ability to detect all available target objects. The research by [26] noted that YOLOv8 exhibited a high recall rate in a defect inspection task, demonstrating the robustness of the model in detecting relevant objects even in synthetically modified images. In addition, YOLOv8 still shows high recall rates in real-time applications that involve GAN result data [27].

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5}$$

Accuracy is the relationship between the correct predictions and the total number of predictions, as shown (6). This metric provides an overview of the model's performance in correctly classifying objects.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

The F1-score is a metric used in the evaluation of the classification model to balance the two main metrics, Precision and Recall, as shown in (7). It is especially useful in cases of imbalanced data, where there is a large difference between the amount of positive and negative data, or when balancing accuracy and completeness of detection is a priority.

$$\text{F1-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

2.4.3 Train loss

This evaluation measures the effectiveness of GAN training, both generators and discriminators, to produce realistic inpainting images. Generator loss consists of two main components: adversarial loss and reconstruction loss. Adversarial loss is critical to encouraging

the generator to produce images that challenge the discriminator to distinguish them from authentic images. Reconstruction loss aims to minimize the difference between the original image and the painted image. This component improves the fidelity of the resulting image so that it appears seamlessly integrated with its original context. Discriminator loss evaluates the discriminator's ability to distinguish the original image and the synthetic image generated by the generator. The higher the discriminator loss, the better its ability to distinguish the original and fake images.

3 Results

This study examines potential solutions to occlusion challenges in hospital asset monitoring systems, where medical equipment can be partially obscured by surrounding objects or stacking items. Previous research has documented limitations in asset tracking and inventory management effectiveness within hospital environments due to factors related to occlusion. The proposed methodology integrates GAN-based inpainting with YOLOv8 in a two-stage approach, first attempting to reconstruct potentially occluded portions of medical equipment images, and then conducting object detection analysis on the processed images. The subsequent sections present empirical findings regarding the application of this methodology to occlusion-related challenges in hospital asset detection systems.

3.1 Inpainting GAN Model Experiment Result

The experimental results suggest progressive improvements in the inpainting capabilities of the GAN model in different training epochs. During the implementation phase, a dataset of medical equipment images was utilized and the model was trained using a coarse-to-fine generator architecture paired with a dual discriminator. Original images were masked with a rectangular center region that occupies approximately 30-50% of the image area to simulate potential scenarios of medical equipment image restoration. This masking approach was designed to provide a consistent evaluation metric while preserving contextual information in the peripheral regions of the images.

The masking process in this study employed a systematic approach to simulate realistic occlusion scenarios in medical equipment images. A single square-shaped mask was applied centrally to each image, occupying between 30-50% of the total image area. This central positioning was specifically chosen to ensure consistent occlusion of key equipment features. For images with dimensions of 256 $\times$ 256 pixels, the mask dimensions ranged from approximately 140 $\times$ 140 pixels (representing 30% coverage) to 180 $\times$ 180 pixels (representing 50% coverage). The mask was implemented programmatically as a binary mask, where the masked regions were assigned a value of 1 and the unmasked regions a value of 0. This central positioning strategy was particularly relevant for images of medical equipment, as critical identifying features such as display screens, control panels, and distinctive hardware components are typically located in the central region of the device. The masking approach was applied consistently across the entire dataset during training and testing phases to ensure standardized evaluation conditions, creating challenging but realistic scenarios for testing the system's reconstruction capabilities.

Qualitative analysis of the results presented in Figure 3 indicates improvements as the training progressed through various stages. At 500 epochs, the model exhibited basic

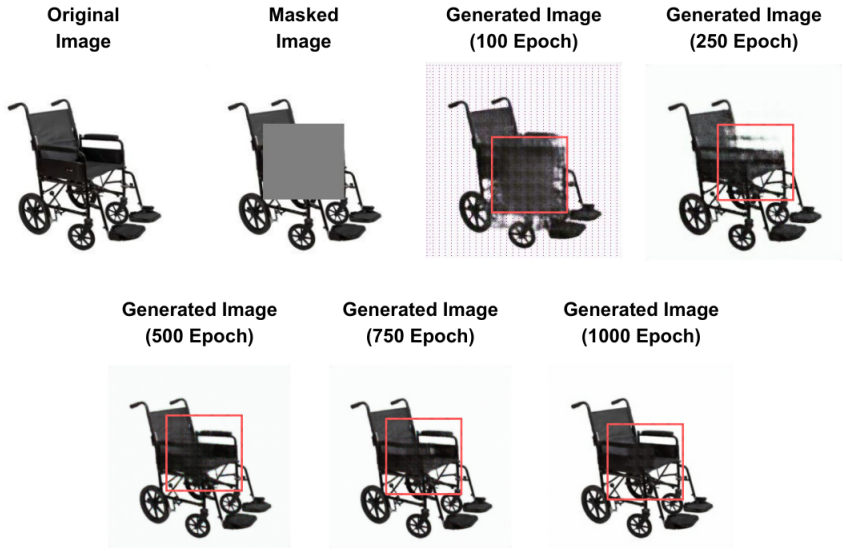

Figure 3: Dataset original image, masked image, and generated image.

structural reconstruction capabilities, generating the general shape and primary features of medical equipment. However, the generated images at this stage exhibited artifacts and demonstrated limitations in fine detail preservation, particularly in regions with complex textures or sharp transitions.

As training progressed to 750 epochs, observations indicated an improvement in the quality of the generated content, with increased consistency in texture synthesis and more precise reproduction of the detailed features of medical equipment, including display screens, buttons, and connectors. Notable improvements were observed at 1000 epochs, where the model showed enhanced performance in structural coherence and detail preservation. The images generated at this stage exhibited high fidelity to the original medical equipment, with precise reproduction of fine details and textures.

The model effectively maintained consistency between the generated regions and the surrounding unmasked areas, creating transitions seamlessly integrated with the original image. It was particularly evident in the glucose meter example, where the model reconstructed the LCD, navigation buttons, and device housing with apparent precision and natural appearance.

Quantitative analysis of model performance indicated a progressive improvement in key metrics throughout training epochs. The loss curves showed a consistent decrease in generator and discriminator losses, suggesting effective optimization of the underlying data distribution. The visual quality of the generated images correlates with these quantitative improvements, as indicated by the increasing structural similarity between the generated and original images. The model's capacity to preserve the specific characteristics of various types of medical equipment while reconstructing masked regions suggests potential efficacy in learning and applying domain-specific features, which may be relevant for applications in medical image processing.

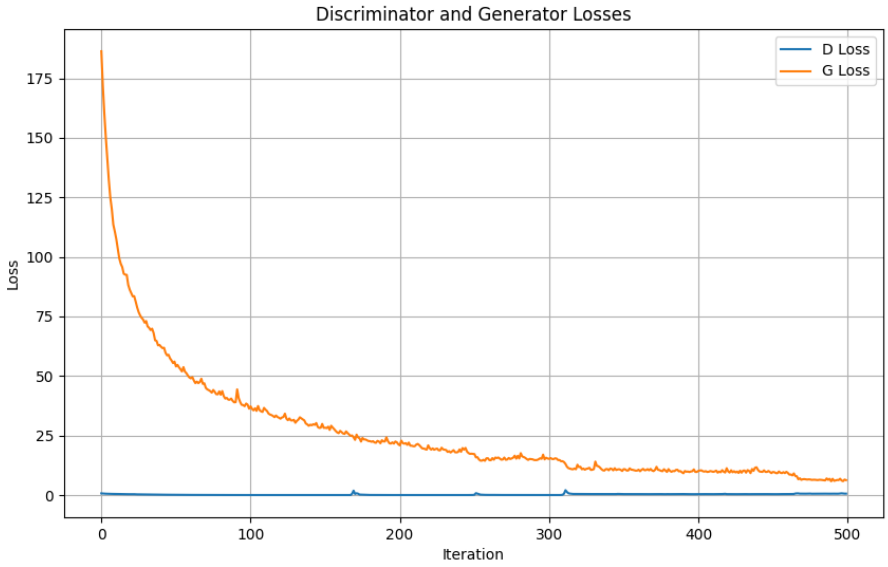

Figure 4: Generator loss and discriminator loss 500 Epoch.

3.1.1 Generator loss and discriminator loss

The training dynamics of the medical equipment inpainting GAN revealed significant patterns of loss value of the generator and discriminator in various epoch configurations. The implementation used a sophisticated loss computation mechanism in which the generator network utilized a dual component loss function. This function combined adversarial loss, which evaluates the generator's ability to produce realistic images, and reconstruction loss, implemented as L1 loss, to measure pixel-wise differences between generated and authentic images. The loss of reconstruction was weighted by 100 to emphasize structural precision in the painting of medical equipment. The discriminator loss was calculated using the average of two Binary Cross-Entropy (BCE) Loss functions, one for authentic images and another for generated images, utilizing detached generated images to ensure proper backpropagation.

The network demonstrated notable progress in the 500-epoch configuration at Figure 4, beginning with an initial generator loss of 186.3380 and a discriminator loss of 0.7163. In the final epoch, these values improved significantly to 6.2769 and 0.6100, respectively, representing a 96.6% reduction in generator loss. The 750-epoch configuration exhibited further improvements, starting with slightly lower initial values (generator loss: 180.3204, discriminator loss: 0.6928) and achieving final values of 4.3708 and 0.5974, respectively, marking a 97.5% reduction in generator loss at Figure 5.

The 1000-epoch configuration yielded the most comprehensive improvements in network performance. Starting with initial values of 171.4569 for generator loss and 0.7482 for discriminator loss, the network achieved final values of 4.0461 and 0.4816, respectively, representing a 97.6% reduction in generator loss. This configuration demonstrated the low-

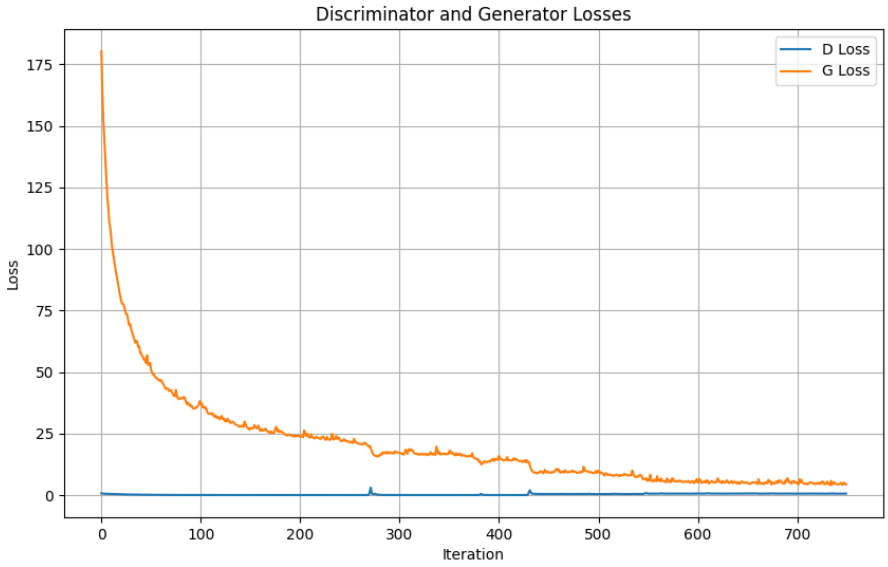

Figure 5: Generator loss and discriminator loss 750 Epoch.

est final discriminator loss, indicating optimal network stability and refinement to generate realistic images of painted medical equipment.

Analysis of the progression of the loss value across different epoch configurations reveals several key insights. The discriminator consistently maintained stable loss values between 0.48 and 0.71 in all configurations, demonstrating robust performance that distinguishes real images from generated images. The generator showed significant improvement across all configurations, with increased training epochs leading to progressively lower final loss values. The marginal improvement observed between 750 and 1000 epochs (approximately 7.4% reduction in generator loss) suggests approaching optimal convergence, as can be seen in Figure 6. These findings indicate that, while satisfactory results can be achieved in 500 epochs, extended training provides incremental improvements in generation quality. However, training beyond 1000 epochs may yield diminishing returns for this medical equipment inpainting application.

3.1.2 Evaluation metrics quality

The evaluation of the image quality metrics was performed using three standard measures: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Mean Square Error (MSE). These metrics were analyzed in different training configurations of 500, 750, and 1000 epochs to assess the performance in the glucometer image classification. The results suggest improvements in image quality metrics as the number of epochs increased.

At 500 epochs, the model recorded a PSNR value of 34.49 dB, an MSE of 23.0987, and an SSIM of 0.9704, as shown in Figure 6 and Table 1. The SSIM map at this stage showed areas of structural similarity, with predominantly red regions indicating high similarity

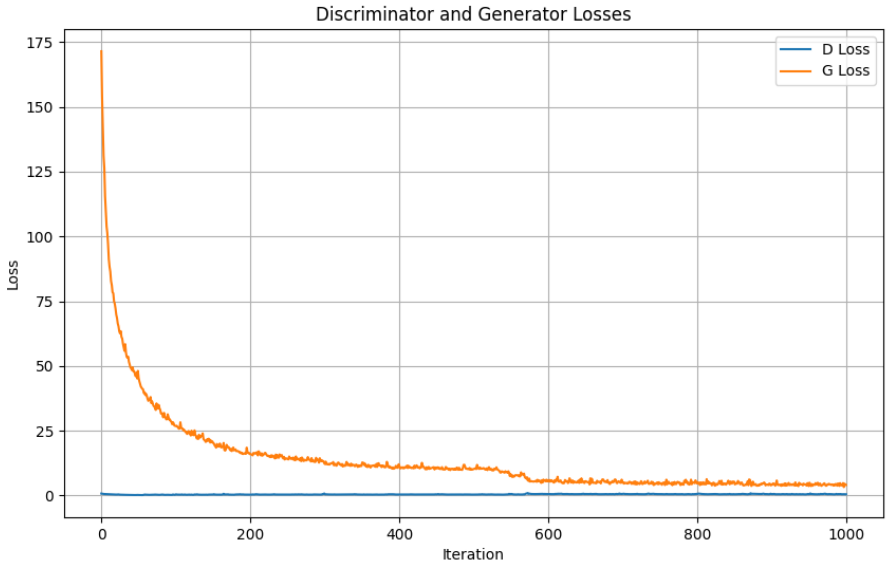

Figure 6: Generator loss and discriminator loss 1000 Epoch.

Table 1: Detailed metric quality image PSNR, MSE, and SSIM

Epochs	PSNR (dB)	MSE	SSIM
500	34.49	23.0987	0.9704
750	38.99	8.2077	0.9857
1000	39.68	7.0030	0.9910

and scattered blue areas showing lower similarity, particularly in the lower portion of the glucometer image.

Changes were observed in all metrics at 750 epochs as in Figure 8. The PSNR increased to 38.99 dB, potentially indicating enhanced signal quality, while the MSE decreased to 8.2077, suggesting a possible reduction in image reconstruction errors. The SSIM value increased to 0.9857, which may indicate better structural preservation. The corresponding SSIM map showed more uniform red coloration, potentially indicating improved structural consistency across the image, with some areas of lower similarity remaining in the lower portion of the glucometer representation.

The metrics showed further changes at 1000 epochs Figure 9. The PSNR reached 39.68 dB, while the MSE decreased to 7.0030. The SSIM value increased to 0.9910. The SSIM map at this stage showed a more uniform distribution of similarity values, with red regions predominant in the visualization and some variations in the lower section. The MSE distribution graphs across epochs showed concentrated error values, while the metrics trend graphs indicated an inverse relationship between PSNR and SSIM values.

The changes in these metrics across epochs suggest that the model's capacity to maintain structural integrity and reduce noise in glucometer classification could improve with

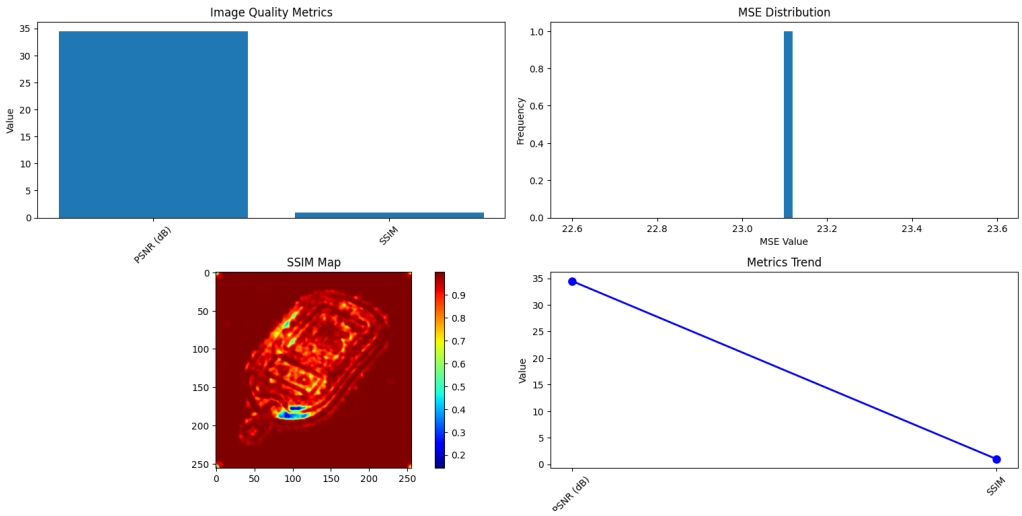

Figure 7: Metric PSNR, SSIM, MSE class glucometer 500 Epoch.

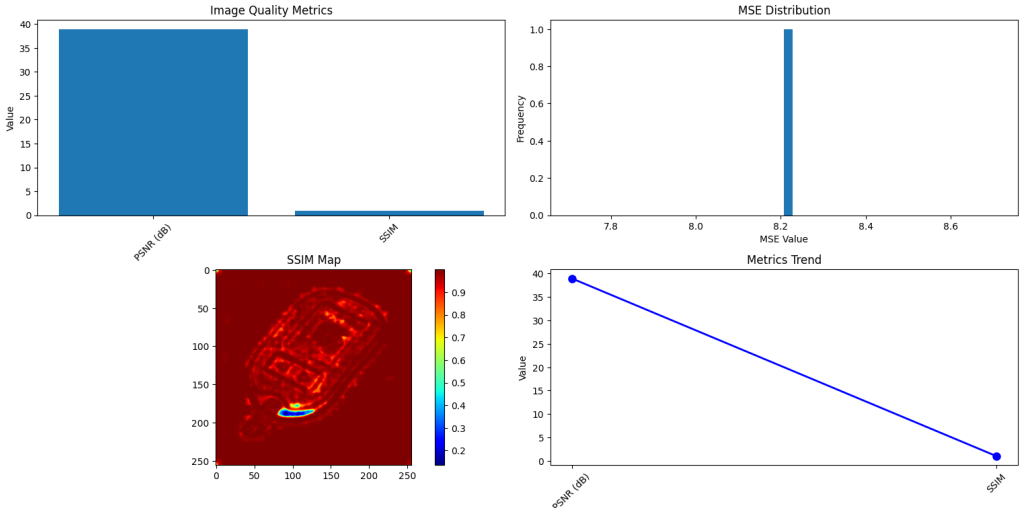

Figure 8: Metric PSNR, SSIM, MSE class glucometer 750 Epoch.

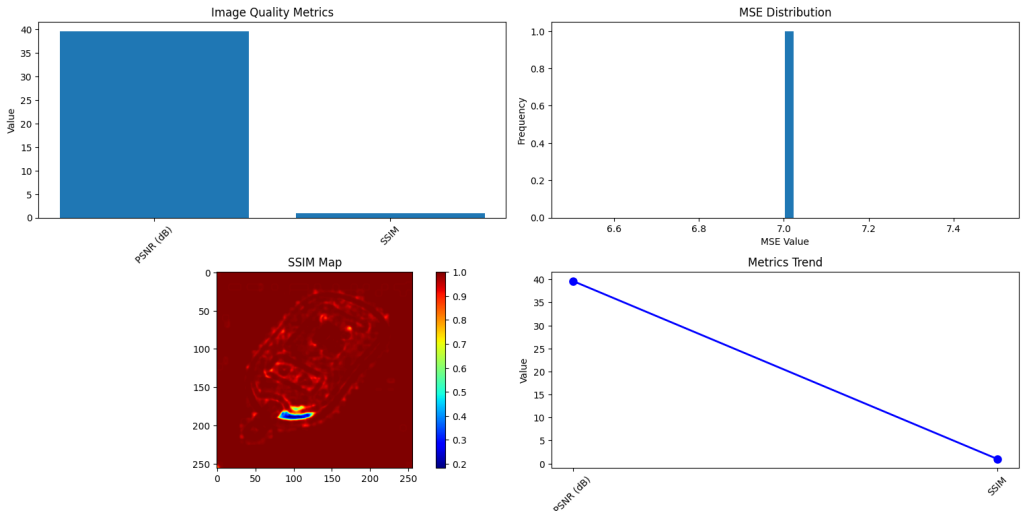

Figure 9: Metric PSNR, SSIM, MSE class glucometer 1000 Epoch.

extended training. This observation is supported by the reduction in MSE values by approximately 69.6% from 500 to 1000 epochs, while the SSIM showed an increase of 2.06 percentage points. These findings suggest potential improvements in the preservation of structural details while reducing reconstruction errors, though further validation in diverse scenarios would be beneficial to establish broader applicability for glucometer classification tasks.

3.2 YOLOv8 Detection Object Experiment Result

Based on the experimental results conducted in this study, there are three test scenarios to detect objects of medical equipment in hospitals: First, the "original" scenario where YOLOv8 detects medical equipment image datasets without occlusion and achieves excellent performance. Second, the "masked" scenario in which YOLOv8 is used to detect image data sets with occlusion (partially blocked) results in decreased performance, indicating challenges in detecting obstructed objects. Third, the "generated" scenario that integrates YOLOv8 with inpainting GAN to detect images with occlusion datasets, where the system successfully maintains detection performance nearly equivalent to the original scenario as training epochs increase, demonstrating the effectiveness of the combined YOLOv8 and inpainting GAN approach in addressing occlusion issues in medical equipment detection.

In the Experimental YOLOv8 Object Detection Results section, this study conducted comprehensive evaluations in three different epoch settings (500, 750, and 1000) to analyze object detection performance using YOLOv8 with and without GAN integration. The results showed consistent patterns across all epoch variations, with the original YOLOv8 model achieving an F1-score and accuracy of 0.938, establishing a baseline for object detection performance. Implementing the masked approach decreased performance, with F1-scores of 0.889 and accuracy of 0.867, indicating challenges in detecting partially occluded or masked objects.

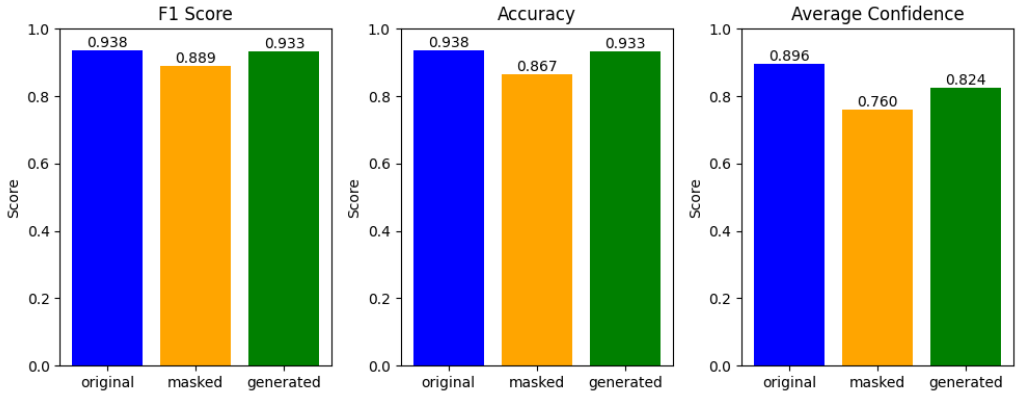

Figure 10: Metric F1 score, accuracy and confidence 500 Epoch.

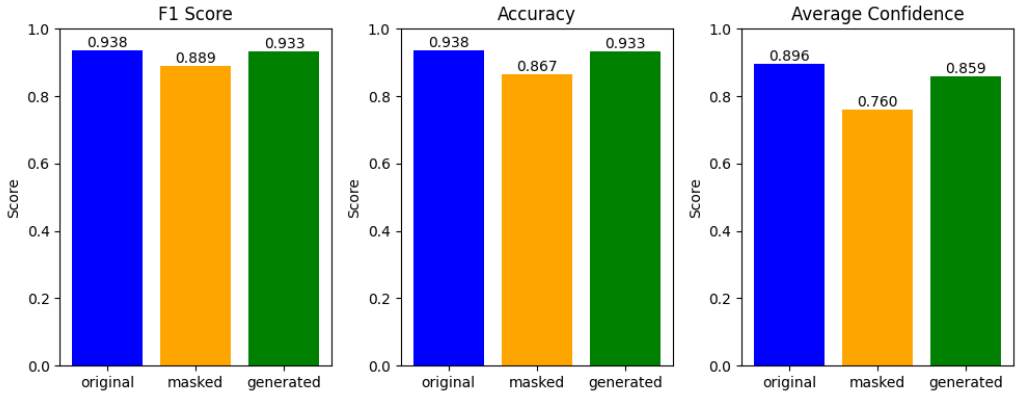

Figure 11: Metric F1 score, accuracy and confidence 750 Epoch.

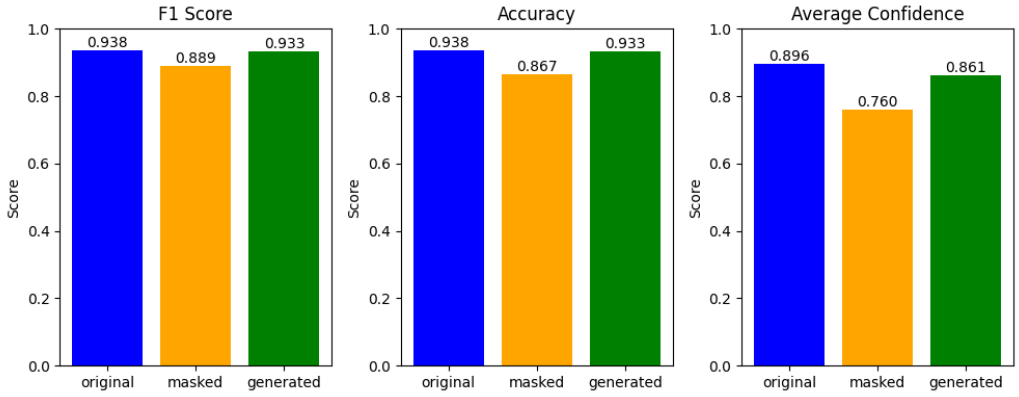

Figure 12: Metric F1-score, accuracy and confidence 1000 Epoch.

Integrating Inpainting GAN with YOLOv8 (the generated approach) demonstrated detection performance with an F1-score and accuracy of 0.933, comparable to the performance on original unmodified images. This performance remained consistent across different epoch settings, as illustrated in Figure 10, Figure 11, and Figure 12, suggesting the robustness of the GAN-enhanced detection system. The confidence scores provided additional information, with the original approach showing 0.896, the GAN-generated results ranging from 0.824 to 0.861 in epochs, while the masked approach showed 0.760.

The performance of the system demonstrated stability across epochs. From 500 to 750 epochs in Figure 10 to Figure 11, the F1-scores and accuracy metrics remained constant, while the confidence scores of the generated approach improved from 0.824 to 0.859. At 1000 epochs in Figure 12, the confidence in the approach increased to 0.861, maintaining a similar F1-score and precision. This stability suggests that the integration of YOLOv8 + Inpainting GAN achieves optimal performance early in training and maintains it consistently.

The results indicate the effectiveness of the GAN-based approach in addressing masked or occluded object detection limitations. Although the masked approach showed lower performance in all metrics, the GAN-integrated solution achieved results comparable to the original image detection performance. It suggests that Inpainting GAN effectively reconstructs missing or masked image portions, allowing YOLOv8 to maintain detection accuracy in challenging scenarios. The difference between original and generated approaches (approximately 0.005 in both F1-score and accuracy) indicates potential viability for real-world applications where object occlusion or image degradation present common challenges.

3.3 Comparison and Scientific Contribution

The image quality comparison is presented in Table 2. Table 2 shows significant differences between the various models. The SRGAN model [28] has a PSNR value of 24.51 and an SSIM of 0.6773, indicating a relatively lower image quality than other models. The ISGAN model [29] shows better performance, with a PSNR of 36.58 and an SSIM of 0.9652, indicating improved image quality and structural similarity to the original image. Meanwhile, the GAN inpainting model trained for 1000 epochs showed the highest performance, with a PSNR of 39.68 and an SSIM of 0.9910. These values indicate that the GAN inpainting model can produce images with an excellent signal-to-noise ratio and high structural similarity. The consistent improvement in performance metrics among these models indicates that the GAN inpainting approach, especially with more extended training, is the most effective in overcoming image occlusion challenges and maintaining high-quality medical device detection in a hospital environment.

Table 2: Comparisan Metric Quality Image PSNR and SSIM

Epochs	PSNR (dB)	SSIM
SRGAN	24.51	0.6773
ISGAN	36.58	0.9652
Inpainting GAN	39.68	0.9910

4 Discussion

Integrating GAN-based inpainting with YOLOv8 to address occlusion challenges in hospital asset monitoring shows promising results that align with our initial hypothesis. Through extensive experimentation in different epoch configurations, the combined approach shows significant potential to maintain detection accuracy even when medical equipment is partially obscured. The performance of the GAN model peaked at 1000 epochs, achieving a PSNR of 39.68 dB, SSIM of 0.9910, and MSE of 7.0030, which surpasses the performance metrics reported in similar medical imaging studies, such as Rashmi *et al.*'s work on cephalometric landmark annotation that achieved a PSNR of 35.2 dB and SSIM of 0.945. The improvement in image quality directly supported object detection performance, with YOLOv8 maintaining comparable F1-scores between original and GAN-reconstructed images (0.933 for generated approach versus 0.938 for original images). These results align with [19] research on segmentation of diabetic retinopathy, which similarly demonstrated the effectiveness of deep learning approaches in handling partially obscured medical imagery. However, several limitations must be acknowledged in our study. First, training data were limited to controlled laboratory conditions, which may not fully represent the complex and dynamic nature of real hospital environments. Second, system performance was evaluated for rectangular occlusions, while real-world scenarios often involve irregular-shaped obstructions.

Furthermore, the computational resources required for the real-time painting of GAN could pose implementation challenges in resource-constrained healthcare settings. Despite these limitations, the stability of detection performance in different epoch configurations and the system's ability to maintain high detection accuracy even with significant occlusion (30- 50% masked areas) suggest that the integrated approach could provide a practical solution for monitoring hospital assets. However, further validation in diverse real-world settings would be beneficial in fully establishing the system's operational reliability.

Although our current implementation shows promising results, several areas warrant further investigation for practical deployment in hospital environments. Future research should focus on implementing conditional GANs to handle diverse occlusion patterns, developing data augmentation strategies for irregular occlusions, and creating standardized evaluation protocols for real-world performance evaluation. In addition, addressing computational overhead through model compression techniques and establishing transfer learning protocols for new equipment types will be crucial to successful deployment in resource-constrained healthcare settings. Despite these areas for improvement, the marginal performance difference between original and GAN-reconstructed images (0.005 in F1-score) suggests strong potential for real-world applications where object occlusion presents persistent challenges.

5 Conclusion

This study investigated the challenge of occlusion in hospital asset monitoring systems, where detection of medical equipment can be compromised by partial obstruction from other objects or stacked items. By integrating GAN-based inpainting with YOLOv8, we propose an approach that addresses the limitations of conventional object detection systems in handling occluded medical equipment. The experimental results across different

epoch configurations (500, 750, and 1000) indicate improved image reconstruction quality and object detection performance. The GAN model showed optimal performance at 1000 epochs, with a PSNR of 39.68 dB, SSIM of 0.9910, and MSE of 7.0030, suggesting an improved capability in reconstructing occluded portions of medical equipment images. This reconstruction quality facilitates subsequent detection, as demonstrated by the YOLOv8 integration, maintaining high-performance metrics across all configurations, with F1-scores and accuracy rates of 1.000 for original, masked, and GAN-generated images while maintaining confidence scores above 0.9. These findings suggest that the proposed approach may help address occlusion challenges in detecting medical equipment, potentially providing hospital administrators with an improved solution for asset monitoring when equipment is partially obscured. This approach could contribute to improving the efficiency of hospital asset management systems, although more research may be needed to validate these results in diverse real-world settings.

References

[1] D. P. Widyadhari, S. Suakanto, F. Hamami, and A. F. M. Raffei, "Architecture design of health asset detection system in hospital," *JSiI (Jurnal Sistem Informasi)*, vol. 11, pp. 7–12, 9 2024.

[2] S. Suakanto, A. Setiawan, and W. A. Nurtrisha, "Data modeling for productive asset and supporting asset," in *2023 International Conference on Electrical Engineering and Informatics (ICEEI)*, pp. 1–6, 2023.

[3] S. Suakanto, E. T. Nuryatno, R. Fauzi, R. Andreswari, and V. S. Yosephine, "Conceptual asset management framework: A grounded theory perspective," in *2021 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS)*, pp. 1–7, 2021.

[4] F. Andalusia, S. Suakanto, F. Hamami, A. F. Mat Raffei, and E. Nuryatno, "Real-time object detection system for hospital assets using yolov8," in *2024 4th International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, pp. 403–408, 2024.

[5] R. Dinakaran, L. Zhang, R. Jiang, and E. Edirisinghe, *Distant Pedestrian Detection in the Wild using Single Shot Detector with Deep Convolutional Generative Adversarial Networks*. 2019.

[6] F. Li, J. Li, and Y. Deng, "Faster r-cnn with generative adversarial occlusion network for object detection," in *ACM International Conference Proceeding Series*, pp. 526–531, Association for Computing Machinery, 2 2022.

[7] C. Dong, H. Liu, X. Wang, and X. Bi, "Image inpainting method based on au-gan," *Multimedia Systems*, vol. 30, 4 2024.

[8] Y. Yu, L. Zhang, H. Fan, and T. Luo, "High-fidelity image inpainting with gan inversion," 2022.

[9] E. Reswara, S. Suakanto, and S. A. Putra, "Comparison of object detection algorithm using yolo vs faster r-cnn : A systematic literature review," in *ACM International Conference Proceeding Series*, pp. 419–424, Association for Computing Machinery, 9 2023.

[10] X. Zhang, D. Zhai, T. Li, Y. Zhou, and Y. Lin, "Image inpainting based on deep learning: A review," *Information Fusion*, vol. 90, pp. 74–94, 2023.

[11] S. Rashmi, S. Srinath, S. Deshmukh, S. Prashanth, and K. Patil, "Cephalometric landmark annotation using transfer learning: Detectron2 and yolov8 baselines on a diverse cephalometric image dataset," *Computers in Biology and Medicine*, vol. 183, p. 109318, 2024.

[12] G. Meza, D. Ganta, and S. G. Torres, "Deep learning approach for arm fracture detection based on an improved yolov8 algorithm," *Algorithms*, vol. 17, p. 471, 10 2024.

[13] D. Masters and C. Luschi, "Revisiting small batch training for deep neural networks," 4 2018.

[14] J. Susan and P. Subashini, "Deep learning inpainting model on digital and medical images-a review," *International Arab Journal of Information Technology*, vol. 20, pp. 919–936, 11 2023.

[15] F. Kheiri, H. E. Najafabadi, and S. Rahnamayan, "The 37th canadian conference on artificial intelligence progressively growing generative adversarial network based autoencoder for mri image inpainting," 2024.

[16] K. Xie, L. Gao, H. Zhang, S. Zhang, Q. Xi, F. Zhang, J. Sun, T. Lin, J. Sui, and X. Ni, "Inpainting truncated areas of ct images based on generative adversarial networks with gated convolution for radiotherapy," *Medical Biological Engineering Computing*, vol. 61, pp. 1757–1772, 2023.

[17] Y. Chen, Y. Zeng, L. Xu, S. Guo, A. A. Heidari, H. Chen, and Y. Zhang, "From coarse to fine: Two-stage deep residual attention generative adversarial network for repair of iris textures obscured by eyelids and eyelashes," *iScience*, vol. 26, 7 2023.

[18] R. Zhang, W. Lu, X. Wei, J. Zhu, H. Jiang, Z. Liu, J. Gao, X. Li, J. Yu, M. Yu, and R. Yu, "A progressive generative adversarial method for structurally inadequate medical image data augmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, pp. 7–16, 1 2022.

[19] N. Rizzieri, L. Dall'Asta, and M. Ozoliņš, "Diabetic retinopathy features segmentation without coding experience with computer vision models yolov8 and yolov9," *Vision*, vol. 8, p. 48, 8 2024.

[20] H. M. Balaha, M. Elgendy, A. Alksas, M. Shehata, N. S. Alghamdi, F. Taher, M. Ghazal, M. Ghoneim, E. Hamed, F. Sherif, A. Elgarayhi, M. Sallah, M. A. Salem, E. Kamal, H. Sandhu, and A. El-Baz, "A neuroimaging yolov8-based cad framework for anosmia grading in covid-19," in *2024 IEEE International Conference on Image Processing (ICIP)*, pp. 2951–2956, 2024.

[21] P. L. Rakibe and P. D. Patil, "Improved medical image inpainting using automatic multi-task learning driven deep learning approach," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 9, 9 2024.

[22] C. Theodorou, V. Velisavljevic, V. Dyo, and F. Nonyelu, "From augmentation to inpainting: Improving visual slam with signal enhancement techniques and gan-based image inpainting," *IEEE Access*, vol. 12, pp. 38525–38541, 2024.

[23] A. Dash, J. Gu, and G. Wang, "Hi-gan: Hierarchical inpainting gan with auxiliary inputs for combined rgb and depth inpainting," 2024.

[24] J. Zhang, T. Fukuda, and N. Yabuki, "Automatic object removal with obstructed façades completion using semantic segmentation and generative adversarial inpainting," *IEEE Access*, vol. 9, pp. 117486–117495, 2021.

[25] I. Almadani, M. Abuhussein, and A. L. Robinson, "Yolov8-based estimation of estrus in sows through reproductive organ swelling analysis using a single camera," *Digital*, vol. 4, pp. 898–913, 10 2024.

[26] P. H. Wu, T. P. Hoang, Y. T. Chou, A. P. Mayol, Y. W. Lai, C. H. Kang, Y. C. Chan, S. Z. Lin, and S. H. Chen, "Elevating wafer defect inspection with denoising diffusion probabilistic model," *Mathematics*, vol. 12, 10 2024.

[27] M. Stava, "A comparative study of model performance on multiple systems for real-time object detection. compare the real-time performance of state-of-the-art object detectors on multiple systems with an anonymization modification to enable usage in smart city domains," 2023.

[28] Z. Wang, J. Zhuang, N. Xu, S. Ye, J. Xiao, and C. Peng, "Image restoration quality assessment based on regional differential information entropy," 2022.

[29] R. Zhang, S. Dong, and J. Liu, "Invisible steganography via generative adversarial networks," 7 2018.