

Article

PigFRIS: A Three-Stage Pipeline for Fence Occlusion Segmentation, GAN-Based Pig Face Inpainting, and Efficient Pig Face Recognition

Ruihan Ma ^{1,†} , Seyeon Chung ^{2,†} , Sangcheol Kim ^{2,*}  and Hyongsuk Kim ^{1,2,*} ¹ Department of Electronics and Information Engineering, Jeonbuk National University, Jeonju 54896, Republic of Korea² Core Research Institute of Intelligent Robots, Jeonbuk National University, Jeonju 54896, Republic of Korea

* Correspondence: sckim7777@gmail.com (S.K.); hskim@jbnu.ac.kr (H.K.)

† These authors contributed equally to this work.

Simple Summary: Accurate pig identification is essential in smart farming, yet fences and other obstructions often block critical facial features, significantly reducing recognition accuracy—especially given the high similarity among individual pigs. To address these challenges, we propose a three-stage Pig Face Recognition and Inpainting System (PigFRIS), which integrates a Fence Occlusion Segmentation module (YOLOv11L), a GAN-based inpainting model (AOT-GAN) to restore occluded regions, and an EfficientNet-B2 recognition component for robust pig classification. By precisely detecting and correcting missing features, PigFRIS shifts the focus from incidental cues to the pig's genuine facial attributes, delivering reliable recognition even in highly obstructed farm environments.

Abstract: Accurate animal face recognition is essential for effective health monitoring, behavior analysis, and productivity management in smart farming. However, environmental obstructions and animal behaviors complicate identification tasks. In pig farming, fences and frequent movements often occlude essential facial features, while high inter-class similarity makes distinguishing individuals even more challenging. To address these issues, we introduce the Pig Face Recognition and Inpainting System (PigFRIS). This integrated framework enhances recognition accuracy by removing occlusions and restoring missing facial features. PigFRIS employs state-of-the-art occlusion detection with the YOLOv11 segmentation model, a GAN-based inpainting reconstruction module using AOT-GAN, and a lightweight recognition module tailored for pig face classification. In doing so, our system detects occlusions, reconstructs obscured regions, and emphasizes key facial features, thereby improving overall performance. Experimental results validate the effectiveness of PigFRIS. For instance, YOLOv11 achieves a recall of 94.92% and a AP_{50} of 96.28% for occlusion detection, AOTGAN records a FID of 51.48 and an SSIM of 91.50% for image restoration, and EfficientNet-B2 attains an accuracy of 91.62% with an F1 Score of 91.44% in classification. Additionally, heatmap analysis reveals that the system successfully focuses on relevant facial features rather than irrelevant occlusions, enhancing classification reliability. This work offers a novel and practical solution for animal face recognition in smart farming. It overcomes the limitations of existing methods and contributes to more effective livestock management and advancements in agricultural technology.

Keywords: deep learning; smart pig farm; pig face recognition; smart farm; face inpainting; fence occlusion segmentation



Academic Editor: Danilo Florentino Pereira

Received: 23 January 2025

Revised: 19 February 2025

Accepted: 20 February 2025

Published: 28 March 2025

Citation: Ma, R.; Chung, S.; Kim, S.; Kim, H. PigFRIS: A Three-Stage Pipeline for Fence Occlusion Segmentation, GAN-Based Pig Face Inpainting, and Efficient Pig Face Recognition. *Animals* **2025**, *15*, 978. <https://doi.org/10.3390/ani15070978>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Occlusions have emerged as a key obstacle to reliable pig face recognition, particularly under real-world farm conditions where metal fences, bars, or gating structures frequently obscure the most discriminative facial features [1,2]. Eyes, snouts, and ears—essential for distinguishing one pig from another—are often partially or fully blocked, thereby causing substantial drops in identification accuracy. In commercial barns, this issue is further compounded by the inherently high facial similarity among pigs, which already makes them challenging to differentiate. As a result, fence-induced occlusions have become one of the most pressing hurdles to the practical application of non-invasive pig face recognition systems, preventing otherwise promising deep learning models from achieving consistent performance.

While traditional livestock identification methods such as ear tags, tattoos, or RFID may circumvent the occlusion problem by relying on physical markers, they often entail significant labor costs, can be intrusive, and may fail when devices are lost or damaged [3,4]. Moreover, such methods do not leverage the benefits of automated, camera-based monitoring—an approach that has gained traction in smart farming for reducing manual intervention and improving animal welfare [5]. Motivated by these limitations, non-invasive approaches have gained increasing traction. Biometric systems using facial recognition offer a promising alternative that can improve both animal welfare and operational efficiency. Deep learning methods have been successfully applied to facial recognition in various species, including cattle [6,7], sheep [8], dogs [9], and birds [10]. Pig face recognition research has likewise advanced, with Hansen et al. [11] achieving around 83% accuracy using Convolutional Neural Networks, and Marsot et al. [12] noting that focusing on features like the eyes and snout can boost accuracy to about 91%. Recent works, including those of Li [13] and Shi [14], further demonstrated the potential of deep learning techniques in achieving rapid and robust pig face detection under practical farm conditions. Moreover, lightweight models also show promise for real-time farm applications [15]. Yet, these models often underperform when confronted with large-scale, systematic occlusions from fences or other barn structures. In one notable study, Liu et al. found that accuracies dropped from 94% to 81% when fences obstructed key areas of the pig's face [16].

Various methods have been explored to address the occlusion problem in pig face recognition, loosely divided into two categories. The first involves occlusion-robust feature extraction, where models are either trained with augmented datasets that include occlusions or are designed with specialized attention mechanisms. These techniques can offer some resilience but often assume that the occluded regions are random or small in area, which may not hold when metal bars systematically block large portions of the face. The second category is image inpainting, a field that has advanced considerably since the introduction of Generative Adversarial Networks (GANs) [17]. GAN-based inpainting has proven adept at restoring missing or damaged pixels in various contexts [18,19], yet generic solutions trained on diverse datasets can struggle to reproduce the subtle textures and geometries of pig faces, especially when large segments of the face are fenced off. The shape of the bars themselves—often linear, repetitive, and covering critical features—further complicates generic inpainting pipelines, which may fill in occlusions with visually plausible yet identity-irrelevant content.

A direct response to these challenges requires a system that explicitly detects fences, recovers occluded areas with pig-specific details, and then leverages the restored images for recognition. Motivated by this need, we propose an integrated three-stage framework called the Pig Face Recognition and Inpainting System (PigFRIS). Rather than relying on partial solutions that either attempt to learn occlusion-robust features or apply generic inpainting, PigFRIS unifies fence detection, targeted GAN-based restoration, and a lightweight

classifier into a single pipeline. By localizing the fence occlusions first, the method can precisely identify which image regions need to be reconstructed, allowing a specialized inpainting algorithm to focus its generative capacity on reproducing core features vital for identification. Subsequently, a resource-friendly classification model processes these refined images, achieving substantially improved accuracy under realistic farm conditions compared to using raw, occlusion-riddled inputs.

In PigFRIS, the first stage of this pipeline is the Fence Occlusion Segmentation module, which isolates metal bars and other structural obstructions that block critical facial features. We employ a supervised segmentation approach based on a customized YOLOv11L model [20] and a dataset of pig face images that are naturally occluded by fences in various orientations and lighting conditions. Human annotators use pixel-level labeling tools to precisely outline the fence regions in each image, thereby teaching the model to differentiate barred occlusions from the pig's facial areas. Once trained, the model automatically generates accurate occlusion masks for newly acquired images, providing essential guidance for the inpainting process that follows.

After the occlusion masks are extracted, the second stage—the Pig Face Inpainting module—relies on an Aggregated Contextual Transformations GAN (AOT-GAN) [21] that is fine-tuned to the texture and geometry of pig faces. This component similarly requires a supervised framework, meaning it learns to recover the missing facial regions based on pairs of “clean” and “masked” images. To build such training data, we again utilize non-occluded pig faces as ground truth and systematically overlay simulated fence masks to generate the source images. During training, the GAN sees both the masked input (with fence-induced occlusions) and the corresponding unmasked face, thereby acquiring the ability to fill in realistic details for eyes, snouts, and other salient markings. This arrangement ensures that the model, once trained, can seamlessly replace occluded areas with plausible pig facial features. The inpainting process is especially critical in preserving subtle distinctions among pigs, given that even minor inaccuracies around the eyes or snout can degrade recognition performance.

In the final stage, the Pig Face Recognition module utilizes an EfficientNet-B2 classifier [22] to identify individual pigs from both inpainted and original images. We opt for this lightweight model in anticipation of real farm environments, where computational resources may be constrained but rapid inference is desired. The improved clarity delivered by the occlusion segmentation and GAN-based inpainting steps markedly enhances recognition accuracy compared to using raw, occlusion-riddled images alone. Empirical observations show that when fence occlusions remain uncorrected, recognition accuracy drops significantly, underscoring how severely obstructed faces impair model performance. By contrast, once PigFRIS has automatically removed fence regions and reconstructed the missing facial features, the identification rates increase notably. The system therefore offers a practical solution to the long-standing limitation of occluded pig faces, promising non-invasive animal identification in large-scale operations and contributing to the broader adoption of smart farming practices.

In practical on-farm deployment, the Pig Face Recognition and Inpainting System (PigFRIS) is designed to process incoming video streams or image feeds in near real-time, addressing fence occlusions on the fly. As the camera captures pig faces, the Fence Occlusion Segmentation module first detects the bars or other obstructing structures using the trained YOLOv11L-based segmentation network. Whenever an occlusion mask is identified, the system immediately feeds the masked pig face into the Pig Face Inpainting module, which employs the fine-tuned AOT-GAN to reconstruct the concealed facial regions. Once the missing areas are realistically restored, the resulting inpainted image is passed to the EfficientNet-B2 classifier for final identification. By chaining these three stages together

in a continuous loop, PigFRIS ensures that obstructed pig faces are accurately processed and recognized as they appear in real-time, offering a practical solution that aligns with the rapid pace of modern farm operations.

- We propose an integrated three-stage system (PigFRIS) that systematically addresses fence occlusions in pig face recognition by unifying segmentation, GAN-based restoration, and lightweight classification rather than focusing solely on recognition or inpainting.
- We employ a customized YOLOv11L segmentation approach trained on pig faces with synthetic fence masks, enabling precise detection of real-farm obstructions such as metal bars that obscure critical facial cues.
- We apply inpainting technology specifically to pig face recognition for restoring occluded facial features. By targeting the challenge of fence-induced obstructions, our approach significantly enhances identification accuracy compared to baseline methods that ignore or inadequately handle such occlusions.
- We adopt EfficientNet-B2 as a resource-friendly recognition module, achieving strong identification accuracy under computational constraints. Empirical evaluations demonstrate a notable accuracy boost when occluded faces are repaired by the GAN before recognition.
- We present a newly collected and annotated dataset of pig faces frequently occluded by farm structures such as fences. This dataset captures realistic variations in lighting, pose, and environmental conditions, filling a gap in existing resources and enabling more accurate evaluations of occlusion-handling techniques in livestock identification.
- Contribution to Smart Farming and Animal Welfare. Our work provides a practical and scalable solution to the occlusion problem in pig face recognition, promoting non-invasive identification methods and supporting more efficient and ethical livestock management practices.

By addressing the “invisible barrier” of occlusions in pig face recognition, our research advances the field of smart farming technology. It demonstrates how integrating state-of-the-art computer vision techniques can overcome practical challenges in agriculture, leading to improved animal welfare and farm efficiency.

The remainder of this paper is organized as follows. Section 2 introduces the proposed PigFRIS framework, detailing methodologies for occlusion detection, image inpainting, and pig face recognition. Section 3 describes the evaluation metrics used to assess the effectiveness of our approach. Section 4 presents the experimental setup, results, and analysis, including comparisons with existing methods and ablation studies. Finally, Sections 5 and 6 conclude the paper and outline potential directions for future research.

2. Materials and Methods

2.1. Overview of the Proposed PigFRIS System

In this section, we detail the methodology behind our Pig Face Recognition and Inpainting System (PigFRIS) as shown in Figure 1. The system is designed to detect and restore occluded pig faces caused by metal fences, thereby improving recognition accuracy in realistic farm settings. To achieve this, PigFRIS combines three interconnected modules: (1) a supervised Fence Occlusion Segmentation model, (2) a GAN-based inpainting mechanism, and (3) a lightweight pig face recognition network. The following subsections describe the architectural overview of the system, followed by in-depth explanations of each module.

In Stage 1, to handle the commonly observed fence occlusion in farm environments, an encoder–decoder segmentation model is employed to identify and locate the occluded

regions. We utilize Structural Similarity (SSIM) and manual annotations to prepare and label the pig face images, thus obtaining ground-truth masks that guide the segmentation network. Afterward, the labeled images are fed into the segmentation network during training to predict the occlusion regions. In Stage 2, once the occlusion regions are detected, the system applies a Generative Adversarial Network (GAN)-based inpainting model to fill the occluded areas. The input to the GAN consists of the original image and the occlusion masks. The generator reconstructs realistic textures for the occluded regions with the help of the discriminator, ensuring the quality of inpainting. The training process incorporates various constraints, such as perceptual loss, adversarial loss, and reconstruction loss based on the original image (ground truth), to enhance the authenticity and consistency of the inpainting results. In Stage 3, after the pig face image has been inpainted, clearer and more distinguishable facial features can be extracted for improved recognition. A classification network, whether conventional or deep learning-based, is used to extract features from the restored image, ultimately classifying the pig's identity (e.g., "ID 2"). By combining the results of the previous two stages, Stage 3 achieves accurate recognition even in complex farm environments. By integrating these three stages, PigFRIS effectively addresses fence occlusion, improves the clarity of key facial features, and facilitates accurate and stable pig identification in real-world farm environments.

Taken as a whole, PigFRIS operates in a sequential yet tightly coupled workflow. Occlusion masks produced by the first module feed directly into the second, guiding the GAN toward precise restoration of missing features. The inpainted outputs then enter the recognition module, where even small gains in accuracy can have significant operational impact in large-scale pig farming. By addressing the systematic problem of fence-induced occlusions, our approach not only improves recognition performance but also lessens the need for invasive methods such as physical tagging or RFID implants, ultimately supporting both animal welfare and farm efficiency.

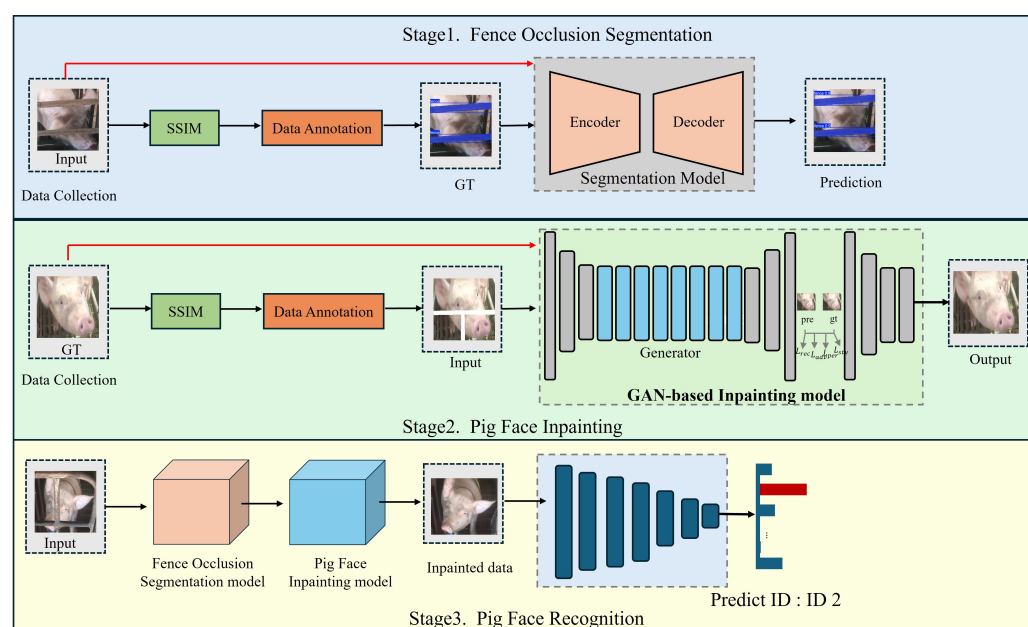


Figure 1. Illustration of the our proposed three-stage System PigFRIS. The top portion (Stage 1) shows how occluded pig faces are collected, annotated, and fed into the Fence Occlusion Segmentation model. The middle portion (Stage 2) depicts the GAN-based inpainting pipeline, where ground-truth (GT) images are used to train the model to restore missing facial regions. The bottom portion (Stage 3) demonstrates how inpainted pig faces are ultimately passed to a recognition network that predicts the correct pig ID, thereby overcoming the challenges posed by fence obstructions.

2.2. Dataset

This study utilizes three distinct datasets, each corresponding to a specific task within the proposed pig face recognition pipeline. All images were collected under typical indoor farm conditions, capturing various lighting scenarios, pig poses, and occlusion patterns. Figure 2. illustrates representative samples from our three proposed datasets and Table 1. shows some details from our proposed three datasets. The first dataset targets Fence Occlusion Segmentation by pairing real occluded images with pixel-level fence annotations; the second focuses on pig face inpainting, consisting of artificially masked inputs and their unoccluded ground-truth images; and the third comprises pig face recognition data, enabling both raw occlusion and inpainted comparisons to assess performance gains in identifying individual pigs.

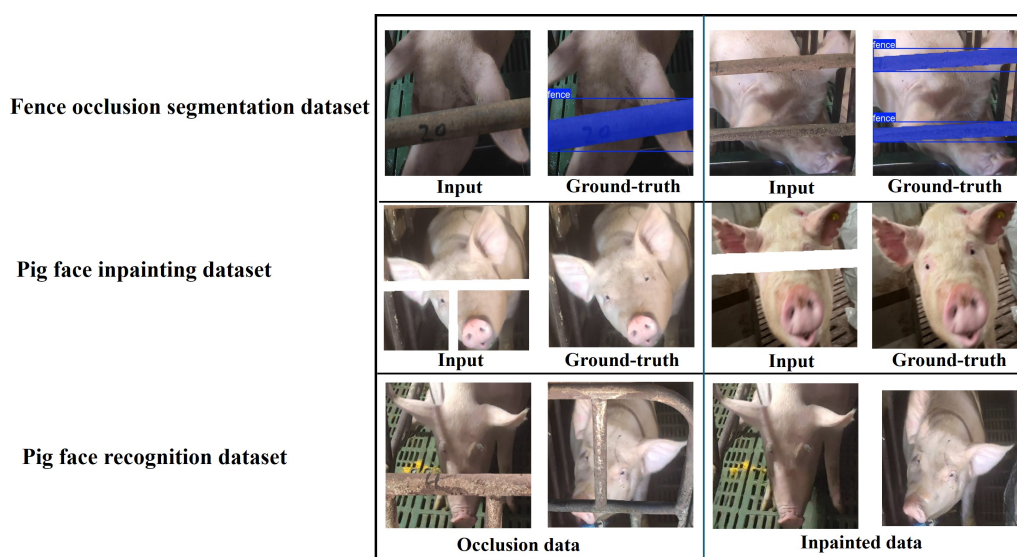


Figure 2. Examples of visualization from the three custom datasets used in our study. The top row showcases the Fence Occlusion Segmentation dataset, highlighting how fences partially obscure the pig’s face (input) and the corresponding ground-truth annotations. The middle row illustrates the pig face inpainting dataset, where artificially masked images (input) are paired with unoccluded ground-truth images. The bottom row shows the pig face recognition dataset, which contains both raw occlusion data and the resulting images after inpainting, enabling performance comparisons between obstructed and restored faces in the final recognition stage.

Table 1. Summary of dataset augmentation, resulting sizes, and usage.

Dataset	Augmentation Strategy	Original Images	Augmented Dataset Size	Usage
Occlusion segmentation dataset	zoom + cut-out	186	558	Fence Segmentation
Pig Face Inpainting Dataset	zoom + horizontal flipping	500	1000	Occlusion Removal
Pig Face Recognition Dataset	zoom horizontal, dropout	1000	2000	Pig Face Recognition

2.2.1. Fence Occlusion Segmentation Dataset

Fence occlusions represent a significant hurdle for automated pig face recognition in indoor farming. To train our segmentation model in a supervised manner, we collected 186 images of pig faces that were actually occluded by fence bars in real farm settings. These images range from minor obstructions—where a single metal bar crosses part of the pig’s snout—to more substantial occlusions—where large portions of the face are blocked.

Each occluded image was then meticulously annotated using LabelMe. Skilled annotators outlined every visible fence segment intersecting the pig’s face at a pixel level,

effectively producing a ground-truth (GT) mask for each image. This mask highlights the exact regions where the fence obscures the pig's face, thereby instructing the model on which areas are fence-related occlusions and which belong to the pig's facial features.

Following the annotation, the dataset was partitioned into training (60%), validation (20%), and testing (20%) subsets. Because real occlusion instances can be limited in variability, especially regarding fence orientation and lighting conditions, we augmented the training set with targeted transformations such as rotation, zoom-in, and random cropping. By diversifying how the fences appear, we increased the total size of the training set while ensuring the model is exposed to a broader spectrum of occlusion patterns.

This rigorously annotated dataset enables the YOLOv11L-based segmentation module to learn robust, fine-grained distinctions between pig facial regions and obstructing fence bars. Accurate segmentation masks are essential to the downstream inpainting step; if the fence detection is unreliable, the inpainting module risks distorting unoccluded facial features or leaving remnants of metal bars in the reconstructed images. Hence, establishing a high-quality occlusion segmentation dataset with real-world farm imagery is vital to the overall success of PigFRIS.

2.2.2. Pig Face Inpainting Dataset

To address the frequent loss of critical facial details when pig faces are partially obscured by fences, we curated a dedicated inpainting dataset designed to train and evaluate supervised GAN-based inpainting methods. In a real farm environment, it is exceptionally difficult to collect perfectly matched “occluded” and “unoccluded” images of the same pig at the same pose. Therefore, our solution involves selecting pig faces with no visual obstructions as ground truth (GT) images, then artificially simulating fence-induced masks on these images to serve as the “occluded” inputs. First, a pool of 500 high-quality pig face images was collected under indoor farm conditions. Each image depicted a pig face that was not covered by any fence. These images were carefully inspected to ensure they captured varied angles, lighting conditions, and individual pig characteristics.

Next, labeling software was employed to draw fence-like shapes—replicating the typical geometry and placement of metal bars—on the otherwise unoccluded pig faces. These simulated occlusion masks formed the input data, while the original, unobstructed images were preserved as the GT. By maintaining a one-to-one pairing between artificially masked images and their unmasked counterparts, the dataset supports supervised learning, allowing the inpainting model to learn how to reconstruct occluded regions based on the GT's intact facial features.

Following this mask generation process, the dataset was split into training (60%), validation (20%), and testing (20%) subsets. To further enrich the model's exposure to diverse occlusion patterns, the training set was augmented with horizontal flipping and zoom-in transformations, producing an additional 1000 inpainted samples. This augmentation step ensures that the model encounters a wide spectrum of fence placements, pig facial poses, and partial occlusions—factors that mirror real-world challenges in pig farming. By training on this combination of artificially created occlusions and carefully curated GT images, the inpainting model learns to restore missing facial regions in a realistic and coherent manner, ultimately preserving essential identifying traits such as the eyes, snout, and ears. This approach balances practicality and scientific rigor. On the one hand, it respects the logistical hurdles of collecting perfectly paired real occlusion data in a farm environment; on the other, it provides a controlled, high-fidelity training set that reliably teaches the inpainting model to handle visually challenging fence obstructions. The result is a systematic, replicable methodology for generating supervised datasets that improve the completeness and accuracy of pig face images in preparation for downstream recognition tasks.

2.2.3. Pig Face Recognition Dataset

Two versions of the pig face recognition dataset were curated to demonstrate performance under authentic farm conditions and to assess the efficacy of inpainting in addressing fence-induced occlusions. The first version preserves real-world occlusions captured directly from an indoor pig farm, where partial fence obstructions naturally occur. The second version applies our inpainting method to the same occluded images, reconstructing missing facial features and offering a direct comparison of how restored data influences recognition accuracy. This dataset features 20 individually tagged pigs, each assigned a unique ID and contributing approximately 50 images—resulting in around 1000 images overall. These photographs were collected over multiple sessions to capture variations in pose, lighting, and typical farm activities, reflecting everyday challenges in commercial piggyery. All images were then split into training (60%), validation (20%), and testing (20%) subsets, ensuring that each pig identity was proportionally represented across the three splits.

To enhance the robustness of the training process, a range of data augmentation techniques was applied to the training set. These augmentations included horizontal flipping and zoom transformations, effectively increasing data diversity. By exposing the network to multiple angles, partial occlusions, and subtle differences in facial appearance, the model learns to generalize across the spectrum of conditions present in working farm environments. The resulting paired (occluded vs. inpainted) dataset thus enables a thorough evaluation of the pipeline's recognition performance, both with and without restoration of fence-obscured pig faces.

2.3. Architecture of the Proposed PigFRIS System

2.3.1. Occlusion Detection Module

Accurate detection of occlusions is paramount for the effectiveness of the PigFRIS framework. In indoor pig farming environments, metal fences frequently obstruct pigs' faces, obscuring critical facial features essential for reliable identification. To address this, the Occlusion Detection module employs the YOLOv11 segmentation model, a state-of-the-art object detection architecture renowned for its high accuracy and real-time performance capabilities. We chose YOLOv11 because of its improved multi-scale feature extraction, balanced speed–accuracy trade-off, and enhanced segmentation performance compared to earlier YOLO versions. In particular, these characteristics allow for robust occlusion detection in challenging farm environments where high inference speed and precise segmentation are critical for subsequent inpainting and recognition tasks.

The architecture of YOLOv11 comprises several key components. At the core is the backbone network, an enhanced version of CSPDarknet, which efficiently extracts multi-scale features from input images. This backbone utilizes Cross Stage Partial (CSP) connections to facilitate better gradient flow and reduce computational redundancy, thereby enhancing feature extraction capabilities. Following the backbone, YOLOv11 incorporates a Path Aggregation Network (PANet) as the neck, which is responsible for feature fusion across different scales. PANet enhances the receptive field and improves the model's ability to detect objects of varying sizes by effectively aggregating features from multiple layers. This multi-scale feature fusion is crucial for accurately identifying occluded regions that may vary in size and shape. The detection head of YOLOv11 is bifurcated into two primary components: the bounding box prediction head and the segmentation head. The bounding box prediction head localizes objects within the image, while the segmentation head generates pixel-wise masks for identified occlusions. The segmentation head employs lightweight convolutional layers to ensure real-time processing without compromising segmentation accuracy. Throughout the network, Leaky ReLU activation

functions [22] and Batch Normalization layers [23] are utilized to accelerate training and enhance model stability.

The YOLOv11 model optimizes a composite loss function designed to balance the objectives of object detection [24] and segmentation [25]. This loss function comprises several components, including the Complete Intersection over Union (CIoU) loss [26] for bounding box optimization, defined as:

$$\mathcal{L}_{bbox} = 1 - \text{CIoU}(b_{pred}, b_{true}) \quad (1)$$

where b_{pred} and b_{true} represent the predicted and ground truth bounding boxes, respectively. To enhance classification accuracy, we utilize the cross-entropy loss:

$$\mathcal{L}_{cls} = - \sum_{c=1}^C y_c \log(p_c) \quad (2)$$

Here, C denotes the number of classes, y_c is the ground truth label (one-hot encoded), and p_c is the predicted probability for class c . For the segmentation task, Binary Cross-Entropy (BCE) loss is employed:

$$\mathcal{L}_{seg} = - \frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (3)$$

where N is the number of pixels, y_i is the ground truth mask pixel value, and p_i is the predicted mask pixel value. The total loss function is a weighted sum of these components:

$$\mathcal{L}_{total} = \lambda_{bbox} \mathcal{L}_{bbox} + \lambda_{cls} \mathcal{L}_{cls} + \lambda_{seg} \mathcal{L}_{seg} \quad (4)$$

where λ_{bbox} , λ_{cls} , λ_{seg} are weight coefficients that balance the contributions of each loss component.

In the PigFRIS system, the YOLOv11 segmentation model is integrated as the core component of the Occlusion Detection module. The process begins with the input of raw pig face images captured in indoor farming environments, which are preprocessed by resizing and normalizing to meet YOLOv11's input requirements. The model then processes these images, extracting features through the CSPDarknet backbone and fusing multi-scale features via the PANet neck. The segmentation head generates precise pixel-wise masks delineating the occluded regions caused by metal fences. These occlusion masks are subsequently utilized by the Image Inpainting module, where the Aggregated Contextual Transformations Generative Adversarial Network (AOT-GAN) reconstructs the obscured facial features. By providing accurate and detailed masks, YOLOv11 ensures that the inpainting process targets only the relevant regions, thereby preserving the integrity of the original image and facilitating high-quality restoration of missing facial features.

The choice of YOLOv11 within PigFRIS is justified by its superior performance in both accuracy and speed compared to previous versions and other object detection models. YOLOv11's ability to perform real-time segmentation makes it highly suitable for large-scale farming operations where timely identification of occlusions is essential. Furthermore, its advanced segmentation capabilities ensure that occluded regions are precisely identified, which is critical for the subsequent image inpainting process. This integration of YOLOv11 not only enhances the accuracy of occlusion detection but also contributes to the overall efficiency and reliability of the PigFRIS system.

2.3.2. Pig Face Inpainting Module

Once occlusions are detected and segmented, the next critical step is to restore the obscured facial features to facilitate accurate recognition. The Image Inpainting module utilizes the Aggregated Contextual Transformations Generative Adversarial Network (AOT-GAN), a powerful model designed for high-fidelity image reconstruction. AOT-GAN was chosen for its ability to capture both local and global contextual information through aggregated contextual transformations, resulting in more coherent and realistic inpainted regions than traditional GAN-based methods. This ensures that the reconstructed facial features seamlessly blend with the surrounding areas, preserving the integrity of the original facial structure—an essential requirement for subsequent recognition tasks.

The architecture of AOT-GAN consists of a generator and a discriminator network. The generator employs an encoder that extracts features from the masked input image, followed by an Aggregated Contextual Transformer (ACT) [27] that aggregates contextual information to fill in the missing regions. The decoder then reconstructs the inpainted image from these aggregated features. The discriminator network [28] is a convolutional neural network trained to distinguish between real and inpainted images, thereby enforcing realism in the generator's outputs.

The AOT-GAN model is optimized using a combination of loss functions to balance reconstruction accuracy and visual realism. The primary components of the loss function include:

$$\mathcal{L}_{\text{rec}} = \|\hat{I} - I_{\text{gt}}\|_1 \quad (5)$$

$$\mathcal{L}_{\text{adv}} = \mathbb{E}_{I_{\text{gt}}} [\log D(I_{\text{gt}})] + \mathbb{E}_{\hat{I}} [\log (1 - D(\hat{I}))] \quad (6)$$

$$\mathcal{L}_{\text{perc}} = \sum_i \|\phi_i(\hat{I}) - \phi_i(I_{\text{gt}})\|_1 \quad (7)$$

$$\mathcal{L}_{\text{style}} = \sum_i \|G_i(\hat{I}) - G_i(I_{\text{gt}})\|_1 \quad (8)$$

where $\hat{I}$ is the inpainted image, I_{gt} is the ground truth image, D is the discriminator, ϕ_i represents features extracted from the i -th layer of a pre-trained network, and G_i is the Gram matrix of features from layer i . The total loss function for the generator is a weighted sum of these components:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{rec}} \mathcal{L}_{\text{rec}} + \lambda_{\text{adv}} \mathcal{L}_{\text{adv}} + \lambda_{\text{perc}} \mathcal{L}_{\text{perc}} + \lambda_{\text{style}} \mathcal{L}_{\text{style}} \quad (9)$$

where λ_{rec} , λ_{adv} , λ_{perc} , λ_{style} are hyperparameters that balance the contributions of each loss component.

In the PigFRIS framework, the AOT-GAN model serves as the core component of the Image Inpainting module. The integration process begins with the application of the occlusion masks generated by the YOLOv11 module to the raw pig face images, creating masked images where occluded regions are highlighted. These masked images, along with their corresponding masks, are then input into the AOT-GAN generator to reconstruct the obscured facial features. The generator leverages the aggregated contextual transformations to ensure that the inpainted regions are both visually realistic and structurally consistent with the non-occluded parts of the image.

The high-quality restoration provided by AOT-GAN is crucial for maintaining the integrity of the facial features, which are essential for accurate pig identification. By effectively reconstructing the occluded regions, the Image Inpainting module ensures that the Recognition module receives complete and detailed facial information, thereby enhancing the overall classification accuracy of the PigFRIS system.

2.3.3. Pig Face Recognition Module

In the pig face recognition model, achieving an optimal balance between computational efficiency and classification accuracy is paramount. To this end, we employ EfficientNet-b2 as the primary recognition model within the PigFRIS framework. EfficientNet-b2 was selected due to its optimal balance between accuracy and parameter count, making it particularly suitable for real-time deployments in farm environments. By maintaining high predictive performance while minimizing computational requirements, it facilitates efficient inference under practical on-site conditions. As a member of the EfficientNet [29] family, EfficientNet-b2 leverages a compound scaling strategy that proportionally expands the network's depth, width, and resolution. This unified approach enhances accuracy while restricting the total number of parameters relative to conventional convolutional neural networks. Consequently, EfficientNet-b2 can effectively operate on resource-limited hardware, reducing inference latency and enabling real-time implementation in pig farming scenarios.

The architecture of EfficientNet-b2 consists of a series of convolutional blocks that include mobile inverted bottleneck convolution (MBConv) layers, each augmented with squeeze-and-excitation (SE) modules to enhance feature representation by recalibrating channel-wise feature responses. The structure of EfficientNet-b2 begins with a stem block that processes the input image through a standard convolution layer, followed by a sequence of MBConv blocks that extract hierarchical features at multiple scales. Each MBConv block integrates SE modules, which adaptively recalibrate channel-wise feature maps by explicitly modeling interdependencies between channels. This mechanism allows the network to focus on the most informative features, thereby improving its discriminative capability without incurring significant computational overhead.

For the classification task, EfficientNet-b2 concludes with a fully connected layer that maps the extracted features to the number of pig identities, followed by a softmax activation function that generates probability distributions over the classes. To optimize the Recognition module, we employ the categorical cross-entropy loss function, which measures the discrepancy between the predicted probability distributions and the ground truth labels, as described in Formula (2).

Within the PigFRIS system, EfficientNet-b2 is integrated as the core component of the Recognition module. After the Occlusion Detection and Image Inpainting modules have processed the raw pig face images, the resulting inpainted images are fed into EfficientNet-b2 for classification. The model leverages its efficient architecture to rapidly process each image, extracting and emphasizing the most salient facial features necessary for distinguishing individual pigs. The use of SE modules within EfficientNet-b2 ensures that the model dynamically focuses on the most informative channels, thereby enhancing its ability to discriminate between pigs with high facial similarity. The deployment of EfficientNet-b2 in PigFRIS offers several advantages. Its compact architecture reduces the computational load, enabling real-time classification without sacrificing accuracy. This efficiency is critical for large-scale farming operations where rapid and accurate identification of individual pigs is required to monitor health, behavior, and productivity. Furthermore, the attention mechanisms embedded within EfficientNet-b2 improve the model's robustness to variations in pose, lighting, and occlusions, ensuring consistent performance across diverse farm environments.

This multi-model approach not only improves recognition accuracy but also ensures the system's scalability and efficiency, making PigFRIS suitable for large-scale farming operations. By leveraging the strengths of each model, PigFRIS achieves a robust and reliable pig face recognition system that enhances operational efficiency and promotes better animal welfare through non-invasive identification methods.

3. Evaluation Metrics

To quantitatively evaluate the performance of our classification model, we utilize classification performance metrics such as Accuracy [30], Precision [31], Recall [32], and F1-Score [33]. These metrics allow us to assess the effectiveness of our classification model comprehensively.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} \quad (10)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (11)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (12)$$

$$\text{F1-Score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (13)$$

Accuracy provides an intuitive measure of how accurate the model is, allowing us to understand its overall outlier detection performance. The metrics are widely used in performance evaluation, with Precision indicating fewer FP (False Positives) as it approaches 1, and Recall indicating fewer FN (False Negatives) as it approaches 1. The F1-Score represents the harmonic mean between Precision and Recall, which are inversely related. A value closer to 1 indicates that the model's performance is balanced.

We further employ FID (Frechet Inception Distance) [34] as evaluation metrics for the GAN model. These metrics are widely used in assessing the performance of generative models. FID measures the statistical difference between generated images and real images. It utilizes the Inception model to extract features from images and computes the Frechet distance (or Wasserstein-2 distance) assuming that the distributions of these features follow Gaussian distributions. The calculation formula for FID is as follows:

$$\text{FID} = \|\mu_r - \mu_g\|^2 + \text{Tr}(\sum_r + \sum_g - 2(\sum_r \sum_g)^{1/2}) \quad (14)$$

Here, μ_r and μ_g represent the mean feature vectors of real and generated data, respectively, while $\sum_r$ and $\sum_g$ denote their respective covariance matrices. Tr denotes the trace of a matrix, and this formula represents the difference between the feature vector distributions of real and generated data. A lower FID value indicates that the quality of generated images is similar to that of real images.

SSIM(x, y) [35] is an index that measures the structural similarity between two images x and y , defined by the following formula:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (15)$$

where μ_x and μ_y are the mean values of images x and y , respectively, σ_x^2 and σ_y^2 are the variance values of images x and y , respectively, and σ_{xy} is the covariance value between images x and y . C_1 and C_2 are small constants for stabilization. The SSIM value ranges from 0 to 1, with a value closer to 1 indicating higher structural similarity between the two images.

PSNR (peak signal-to-noise ratio) [36] is another metric for measuring the similarity between two images, defined by the following formula:

$$\text{PSNR}(x, y) = 10 \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}(x, y)} \right) \quad (16)$$

where MAX is the maximum pixel value of the image, and $MSE(x, y)$ [37] is the mean squared error between images x and y . MSE is defined as follows:

$$MSE(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2 \quad (17)$$

where N is the total number of pixels in the image. A higher PSNR value indicates greater similarity between the two images.

MAE (Mean Absolute Error) [38] is an index that measures the absolute difference between two images, defined by the following formula:

$$MAE(x, y) = \frac{1}{N} \sum_{i=1}^N |x_i - y_i| \quad (18)$$

where N is the total number of pixels in the image. A lower MAE value indicates a smaller difference between the two images.

Thus, FID, SSIM, PSNR, and MAE are important metrics for evaluating the similarity between the generated image and the actual image. Higher SSIM and PSNR values, and lower FID and MAE values, indicate that the quality of the generated image is similar to the actual image.

4. Experiments

4.1. Fence Occlusion Segmentation

In this part of our experiment, we evaluate the effectiveness of the Fence Occlusion Segmentation module in accurately detecting and localizing occluded regions in the Fence Occlusion Segmentation dataset.

4.1.1. Experimental Setup

The Fence Occlusion Segmentation experiment was conducted using the YOLOv11 Large segmentation model on a dataset specifically curated for pig face occlusions caused by fences. The input image size was consistently fixed at 640×640 pixels, and the dataset was split into an 80% training set and a 20% validation set. The model was trained for 200 epochs with a batch size of 16, and early stopping was applied with patience of 100 epochs to mitigate overfitting. The optimizer was automatically selected (Adam) under the Auto setting, initializing the learning rate at 0.1 and gradually reducing it by a final learning rate factor of 0.01. The momentum was set to 0.937, with a weight decay of 0.0005 to address potential overfitting. A warmup phase was employed for the first 3 epochs, starting with a warmup momentum of 0.8. In terms of loss configuration, the box loss weight was set to 7.5, the class loss weight to 0.5, and the DFL (Distribution Focal Loss) weight to 1.5. Evaluation metrics focused on segmentation accuracy, including measures such as Precision, Recall, and mAP at different IoU thresholds, ensuring a comprehensive assessment of how effectively the model identified and segmented occluded fence regions in pig face images.

4.1.2. Experimental Results

In this experiment, multiple YOLO-based architectures were evaluated on a challenging occlusion segmentation dataset, where fences and other farm structures frequently obscure pig faces. The compared models include YOLO8n, YOLO8s, YOLO8m, YOLO8l, YOLO9c, YOLO11n, YOLO11s, YOLO11m, and YOLO11l. Among these, YOLO11 variants emerged as particularly effective, with YOLO11l delivering the strongest performance.

Table 2 summarizes the segmentation outcomes. Notably, YOLO11l achieved an AP_{50} of 96.28%, AP_{75} of 91.9%, and AP_{50-95} of 89.48%, surpassing all other YOLO models. Additionally, it reached a Recall of 94.92%, indicating a high capacity to accurately localize occluded regions. Meanwhile, YOLO11n demonstrated a balanced performance, with a Precision of 89.49% and a Recall of 91.53%, suggesting the reliable coverage of occlusions and minimal false positives. Figure 3. presents the fence occlusion segmentation results using our customized YOLOv11L model on the testing dataset. Each column pair displays the original occluded image (top) alongside its corresponding segmentation result (bottom). The model effectively identifies and highlights fence bars (in blue) across diverse real-world occluded pig face images. Overall, these results underscore the robustness of YOLOv11l in handling complex occlusion patterns common in indoor pig farming environments. By achieving higher mAP values across diverse thresholds, YOLO11—particularly its large variant—produces more precise segmentation masks critical for subsequent inpainting. This high-quality occlusion detection not only reduces the likelihood of missed occlusions but also ensures cleaner boundaries for a more accurate restoration of hidden facial features in the ensuing stages of the PigFRIS pipeline.

Table 2. Experimental results of YOLO models on test dataset of Fence Occlusion Segmentation dataset.

Model	Precision	Recall	AP_{50}	AP_{75}	AP_{50-95}
YOLO8n [39]	94.97 $\pm$ 0.7	89.83 $\pm$ 1.2	95.48 $\pm$ 0.4	90.17 $\pm$ 0.1	88.62 $\pm$ 0.5
YOLO8s	87.46 $\pm$ 0.4	94.58 $\pm$ 0.2	94.87 $\pm$ 1.1	90.60 $\pm$ 0.6	84.97 $\pm$ 0.8
YOLO8m	88.71 $\pm$ 0.1	93.22 $\pm$ 0.4	96.22 $\pm$ 0.1	92.35 $\pm$ 0.9	86.86 $\pm$ 0.5
YOLO8l	90.66 $\pm$ 0.2	86.44 $\pm$ 0.4	95.78 $\pm$ 1.6	93.29 $\pm$ 0.6	84.88 $\pm$ 1.7
YOLO9c [40]	90.15 $\pm$ 1.1	93.05 $\pm$ 0.5	91.98 $\pm$ 0.1	88.82 $\pm$ 1.2	85.43 $\pm$ 0.6
YOLO11n [20]	89.49 $\pm$ 0.4	91.53 $\pm$ 1.2	93.68 $\pm$ 0.8	88.28 $\pm$ 1.3	84.21 $\pm$ 0.4
YOLO11s	88.66 $\pm$ 0.6	88.14 $\pm$ 0.2	92.41 $\pm$ 1.7	89.01 $\pm$ 0.6	85.37 $\pm$ 0.8
YOLO11m	92.04 $\pm$ 1.1	89.83 $\pm$ 0.3	92.12 $\pm$ 0.6	89.47 $\pm$ 0.4	85.01 $\pm$ 1.6
YOLO11l	87.05 $\pm$ 0.4	94.92 $\pm$ 0.7	96.28 $\pm$ 1.4	91.90 $\pm$ 1.1	89.48 $\pm$ 1.1

Note: The values that recorded the highest performance for each metric are highlighted in bold.

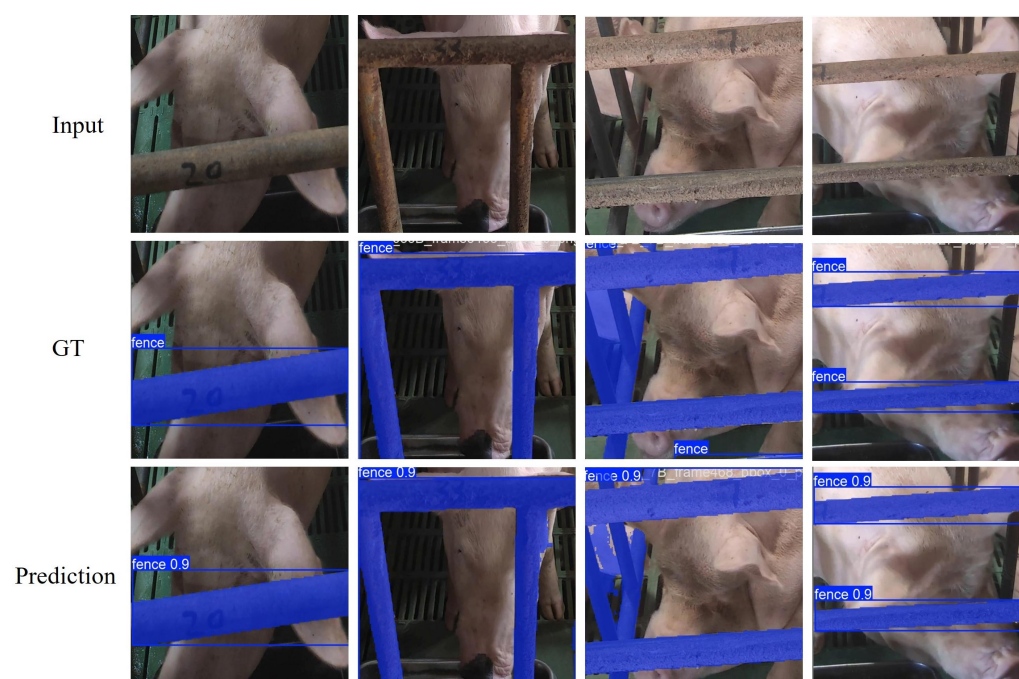


Figure 3. Fence Occlusion Segmentation Results using YOLOV11L model on the Testing Dataset.

4.2. Pig Fence Inpainting

In this phase, we compare the performance of several inpainting models to evaluate how effectively they restore occluded facial regions in the pig face inpainting dataset. Detailed descriptions of our model configurations, baseline comparisons, and evaluation metrics can be found in the subsequent subsection.

4.2.1. Experimental Setup

The inpainting experiments were performed on pig face images where occluded regions were identified and masked prior to restoration. Each image was resized to 256×256 pixels to ensure consistent input dimensions. A comprehensive evaluation was conducted using four advanced inpainting models—DeepFillv2, AOTGAN, RFR, and TFill—to determine the most suitable model for the given application. All models were trained for 15,000 iterations, starting from pre-trained weights on the CelebA dataset as provided by the original authors. The batch size was set to 8, and the Adam optimizer was employed with a learning rate of 1×10^{-4} for both the generator and the discriminator, using $\text{Beta1} = 0.5$ and $\text{Beta2} = 0.999$ to ensure stable and efficient convergence. Multiple loss terms guided the model updates. An L1 Loss, weighted as 1, encouraged fidelity to the original image structure. A Style Loss, weighted as 250, preserved texture and stylistic details, while a Perceptual Loss, weighted as 0.1, ensured the restored regions aligned with high-level semantic features. To further enhance realism, an adversarial loss was incorporated at a weight of 0.01, prompting the generator to produce outputs that closely resembled authentic images.

Evaluation metrics included the Fréchet Inception Distance (FID) to assess distributional similarity, the Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR) to quantify structural and pixel-level accuracy, and the Mean Absolute Error (MAE) to measure overall deviation from the ground truth. This combination of metrics offered a comprehensive and robust evaluation of each model's inpainting quality. By comparing DeepFillv2, AOTGAN, RFR, and TFill under these standardized conditions, we ensured a fair assessment of performance and a well-informed decision on which inpainting model best meets the requirements of our application.

4.2.2. Experimental Results

The inpainting outcomes underscore the advantages of AOTGAN in reconstructing occluded pig face images. As presented in Table 3, AOTGAN attains an FID of 51.48, outperforming other methods and suggesting a closer match to the statistical distribution of real, unobstructed images. Moreover, AOTGAN achieves an SSIM of 91.5, indicating the effective preservation of structural details and perceptual consistency in facial features. Although TFill yields marginally higher PSNR and a slightly lower MAE, these gains in pixel-level accuracy do not translate into producing globally realistic textures as effectively as AOTGAN. DeepFillv2, while offering competitive SSIM and PSNR values, exhibits a higher FID, implying less alignment with the overall realism of genuine pig faces. RFR, with substantially lower SSIM and PSNR, struggles to generate coherent restorations, revealing limitations in both fine-grained detail and global structural consistency.

Overall, AOTGAN's balanced performance stands out. Its superior FID and SSIM, in conjunction with competitive PSNR and MAE, affirm its ability to faithfully capture intricate textures. Consequently, AOTGAN emerges as the primary inpainting choice, ensuring that restored images closely resemble high-quality, authentic data—an essential requirement for accurate pig face reconstruction.

Table 3. Performance evaluation results assessed using image restoration performance metrics.

Model	FID ↓	SSIM ↑	PSNR ↑	MAE ↓
Deepfillv2 [41]	72.5 ± 0.6	90.2 ± 0.2	29.3 ± 0.2	7.2 ± 0.4
RFR [39]	80.87 ± 0.4	15.14 ± 0.9	14.21 ± 0.3	76.5 ± 1.1
TFill [19]	53.98 ± 0.4	90.9 ± 0.5	30.34 ± 0.7	6.3 ± 0.2
AOTGAN[21]	51.48 ± 0.8	91.5 ± 0.1	30.25 ± 0.5	6.6 ± 0.1

Note: The values that recorded the highest performance for each metric are highlighted in bold.

Figure 4 presents examples of failure cases observed during the inpainting experiments, illustrating that while the model generally performs effectively, certain challenging scenarios remain problematic. In (A), the model appears overfitted, producing visually inconsistent or unrealistic patterns in the restored region. In (B), a key facial feature—the pig’s eye—has been inadvertently removed during restoration, indicating difficulty in maintaining essential anatomical details. In (C), the originally open eye was restored as closed, suggesting that subtle variations in facial features are not always accurately captured. These cases highlight the limitations of the model’s current configuration and underscore the need for improved feature discrimination and context-aware reconstruction strategies. By acknowledging these shortcomings, future work can focus on refining the model to better handle complex occlusions, intricate textures, and subtle facial expressions.

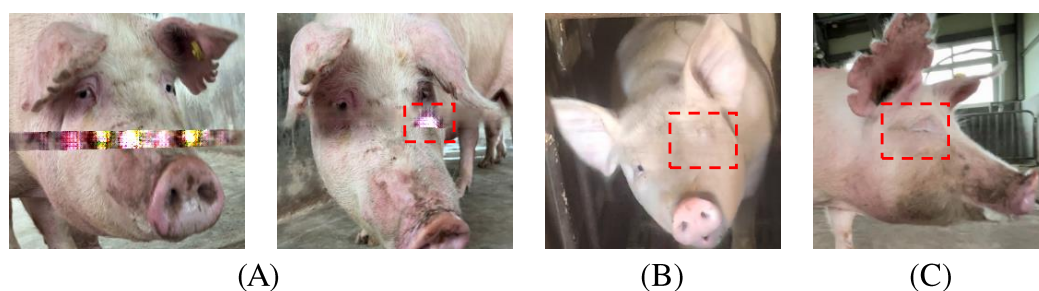


Figure 4. Erroneously restored images from the image inpainting experiment. (A) shows an overfitted inpainting result, (B) shows an inpainting result where one eye was removed, and (C) shows an inpainting result where the original open eye was restored in a closed state.

Figure 5 displays the reconstructed pig face images produced by different inpainting models, including AOT-GAN, DeepFillv2, RFR, and TFill, alongside the original and masked inputs. While the masked images reveal substantial missing information, certain models demonstrate remarkable skill in recovering realistic facial features. AOT-GAN preserves subtle details such as the texture and contour of the pig’s ears and snout, resulting in inpainted images that closely resemble the original. DeepFillv2 also produces visually plausible restorations, though occasional artifacts or inconsistencies may appear in more complex occlusion regions. In contrast, RFR tends to struggle in maintaining coherent structure, often introducing distortions or blurred patches that reduce overall realism. TFill achieves strong pixel-level accuracy, offering slightly sharper local detail in some instances but sometimes lacks the global perceptual alignment that AOT-GAN achieves. Viewed together, these results highlight the trade-offs between different methods and emphasize that AOT-GAN consistently provides a superior balance of perceptual realism and structural integrity when restoring occluded pig faces.

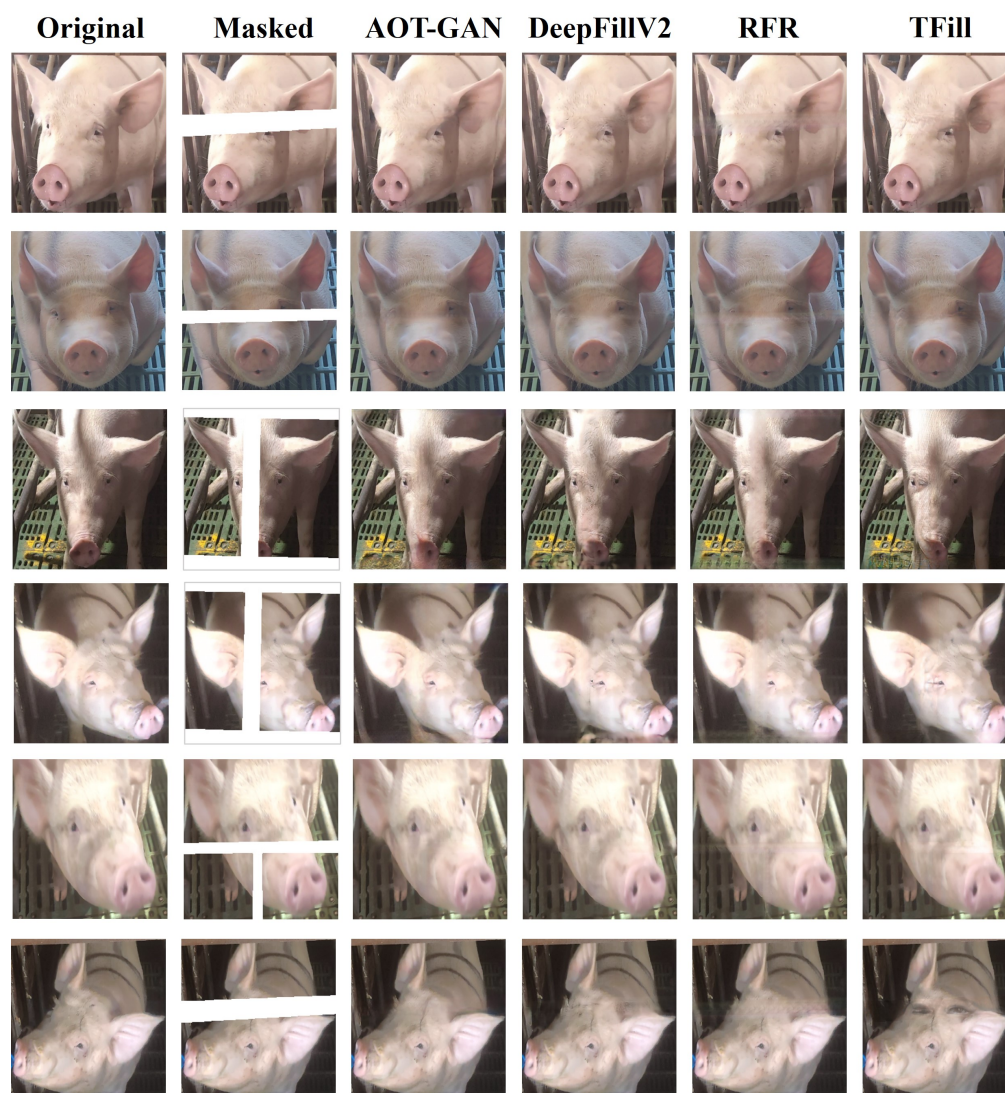


Figure 5. Image inpainting results of the GAN models used in the experiment.

4.3. Pig Face Classification

In this section, we compare multiple classification networks—including EfficientNet-b2—on the Pig Face Recognition dataset to accurately distinguish individual pigs using inpainted facial features. By measuring classification accuracy, inference time, and parameter efficiency, we confirm the model’s practicality in real farm scenarios. Key implementation details and evaluation metrics are provided in the following subsection.

4.3.1. Experimental Setup

The pig face recognition experiments were conducted using a model initialized with ImageNet pre-trained weights. Each pig face image was resized to 256×256 pixels before being fed into the model. Training proceeded with a batch size of 32, and a global average pooling layer was employed to extract informative features from the convolutional backbone. To prevent overfitting, a dropout rate of 0.3 was applied, and a fully connected layer comprising 512 units and ReLU activation was introduced prior to the final classification layer. The output layer used a softmax activation function to produce class probabilities for N distinct pig identities. An Adam optimizer was employed with the learning rate set to 0.0004, and training was performed for 100 epochs. This provided sufficient time for the model to converge without incurring excessive computational expense. To comprehensively evaluate the recognition performance, Accuracy, Precision, Recall, and

F1-Score were used as metrics. Accuracy measured overall correctness, Precision, and Recall assessed the model's ability to handle class imbalances effectively, and the F1-Score provided a harmonic mean that balanced both Precision and Recall. This combination of metrics ensured a thorough understanding of the model's classification capabilities and its practical utility in distinguishing individual pigs.

4.3.2. Experimental Results

Table 4 provides a comprehensive comparison of several deep learning models on our pig face recognition dataset, showcasing the consistent and significant performance gains enabled by removing fence occlusions through the PigFRIS pipeline. Each model is evaluated in two distinct modes: training on raw, occluded images (without inpainting) versus training on images that have passed through PigFRIS's first two stages. In these stages, a YOLOv11L-based Fence Occlusion Segmentation module initially detects and masks the fence bars covering the pig's face, and an AOT-GAN inpainting module subsequently reconstructs the missing regions. This yields inpainted images (with inpainting), effectively restoring critical facial details around the eyes, snout, and ears for subsequent recognition.

Table 4. Classification performance of multiple deep learning models on the test set of our pig face recognition dataset: raw (occluded) vs. PigFRIS-inpainted images.

Model	Accuracy	Precision	Recall	F1 Score
MobileNet-V2 (w/ inpainting)	87.71 $\pm$ 0.8	87.66 $\pm$ 1.1	87.71 $\pm$ 0.7	86.27 $\pm$ 0.9
MobileNet-V2 (w/o inpainting)	78.51 $\pm$ 1.0	76.96 $\pm$ 0.5	75.51 $\pm$ 1.2	73.01 $\pm$ 0.8
MobileNet-V3 (w/ inpainting)	89.94 $\pm$ 0.6	92.21 $\pm$ 1.3	89.94 $\pm$ 0.8	89.02 $\pm$ 1.2
MobileNet-V3 (w/o inpainting)	72.45 $\pm$ 0.8	83.06 $\pm$ 1.0	72.45 $\pm$ 0.9	72.86 $\pm$ 1.3
EfficientNet-B0 (w/ inpainting)	89.39 $\pm$ 1.0	86.52 $\pm$ 0.9	89.39 $\pm$ 1.2	87.11 $\pm$ 0.7
EfficientNet-B0 (w/o inpainting)	80.61 $\pm$ 1.3	86.27 $\pm$ 0.8	80.61 $\pm$ 0.6	79.80 $\pm$ 1.4
EfficientNet-B1 (w/ inpainting)	74.30 $\pm$ 1.5	78.90 $\pm$ 0.9	78.90 $\pm$ 1.1	69.03 $\pm$ 1.0
EfficientNet-B1 (w/o inpainting)	71.43 $\pm$ 1.2	81.68 $\pm$ 0.7	71.43 $\pm$ 0.8	68.38 $\pm$ 1.2
EfficientNet-B2 (w/ inpainting)	91.62 $\pm$ 0.9	93.22 $\pm$ 0.6	91.62 $\pm$ 0.8	91.44 $\pm$ 1.1
EfficientNet-B2 (w/o inpainting)	86.22 $\pm$ 1.3	87.93 $\pm$ 1.2	86.22 $\pm$ 1.0	85.88 $\pm$ 0.7
ResNet50 (w/ inpainting)	89.39 $\pm$ 0.8	90.83 $\pm$ 1.3	89.39 $\pm$ 1.2	89.05 $\pm$ 0.9
ResNet50 (w/o inpainting)	78.06 $\pm$ 1.1	83.81 $\pm$ 1.0	78.06 $\pm$ 0.7	78.03 $\pm$ 1.4
ResNet101 (w/ inpainting)	87.15 $\pm$ 0.9	89.52 $\pm$ 1.1	87.15 $\pm$ 0.6	86.14 $\pm$ 1.0
ResNet101 (w/o inpainting)	77.55 $\pm$ 1.2	80.05 $\pm$ 0.9	77.55 $\pm$ 1.3	76.46 $\pm$ 1.0

The "(w/ inpainting)" designation in the table signifies models trained on the PigFRIS-inpainting-enhanced dataset, while the "(w/o inpainting)" versions are trained on the original, occluded pig face images. Bold values indicate the best performance within the same model when comparing results with and without inpainting.

A clear trend emerges across MobileNet-V2, MobileNet-V3, EfficientNet-B0/B1/B2, and ResNet50/101: models trained on inpainted images demonstrate higher Accuracy, Precision, Recall, and F1-Scores. This improvement underscores the benefit of accurately isolating fence bars via YOLOv11L segmentation and then reconstructing occluded areas through AOT-GAN, ensuring that classifiers receive a clearer view of key pig facial cues. Notably, both MobileNet-V2 and MobileNet-V3 exhibit marked performance jumps after occlusion removal, suggesting that even lighter-weight architectures profit from training with reconstructed data. Meanwhile, deeper ResNet networks (ResNet50 and ResNet101) also display metric boosts, indicating that enhanced facial features benefit from a broad range of model complexities. Among all tested configurations, EfficientNet-B2—employed in the third stage of PigFRIS—achieves the highest overall metrics, with Accuracy increasing from 86.22% without inpainting to 91.62% with inpainting. This strong performance reflects the synergy of EfficientNet-B2's balanced scaling strategy with the fence segmentation and GAN-based restoration steps.

In summary, the results in Table 4 illustrate how removing fence bars and reconstructing missing features significantly elevates pig face recognition performance, irrespective of a model's size or complexity. By integrating Fence Occlusion Segmentation (Stage 1), AOT-GAN inpainting (Stage 2), and a robust classifier (Stage 3), PigFRIS ensures that each network can focus on discriminative pig facial details. This pipeline approach validates our proposed system, demonstrating that a dedicated occlusion-removal process can substantially enhance model accuracy—particularly in architectures like EfficientNet-B2, which capitalize on richer visual information once obstructions are eliminated.

Figure 6 provides a detailed heatmap visualization that contrasts occluded pig faces (A) with their post-inpainting counterparts (B). In (A), the heatmaps clearly highlight the metal bars as regions of high intensity—typically represented in warmer colors such as red or yellow—which slice across the pig's facial region. These high-intensity areas indicate that the segmentation model is capturing strong gradients or activation in those obstructed regions, reflecting the difficulty of extracting continuous facial features when fence occlusions are present. Conversely, (B) displays the same images after processing through our AOT-GAN inpainting module. In these visualizations, the previously occluded areas now exhibit a more homogeneous intensity distribution, with cooler colors replacing the disruptive high-intensity patterns. This shift in the heatmap signals that the inpainting process has effectively removed the metal bars and reconstructed the occluded regions, allowing the network's attention to refocus on intrinsic facial cues such as the snout, eyes, and ears. The refined intensity gradients in (B) suggest a clearer delineation of morphological details, which ultimately contributes to improved downstream recognition performance. Overall, these heatmaps not only demonstrate the successful removal of occlusions but also provide visual evidence of how inpainting enhances feature continuity and clarity, thereby facilitating more robust pig face recognition.

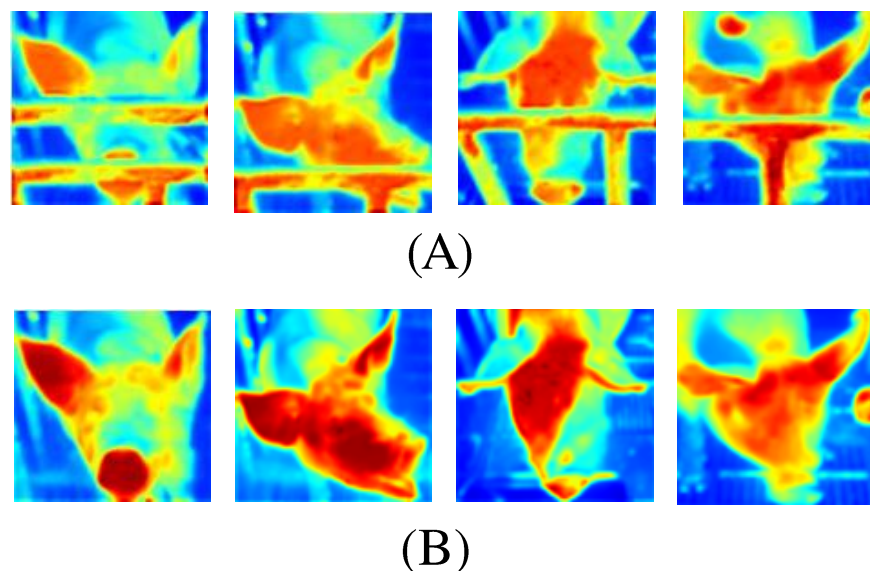


Figure 6. Heatmap Visualization of Occluded vs. Inpainted Pig Faces (A) shows heatmaps of pig faces with fence bars prominently visible as high-intensity obstructions, highlighting how occlusions disrupt the model's ability to capture key facial cues. (B) displays the same images after AOT-GAN removes and reconstructs the missing regions, revealing more continuous heat distributions across the pig's head. This inpainting step allows the network to focus on salient identifiers—such as the eyes, ears, and snout—thereby improving overall face recognition performance.

Figure 7A displays the two-dimensional t-SNE projection of feature embeddings extracted by EfficientNet-B2 from pig faces that remain partially obscured by fence bars. Due to these occlusions, many points cluster less distinctly, indicating that key identifying

cues—such as the eyes and snout—are partially missing or distorted. Figure 7B presents the same images after passing through the first two stages of PigFRIS, where YOLOv11L segments the fence regions and AOT-GAN inpaints the missing facial areas. With the obstructing bars removed and the facial features more fully restored, the embeddings form tighter, more separable clusters. This clear improvement demonstrates how accurately recovering occluded facial details before recognition empowers EfficientNet-B2 to differentiate individual pigs more effectively.

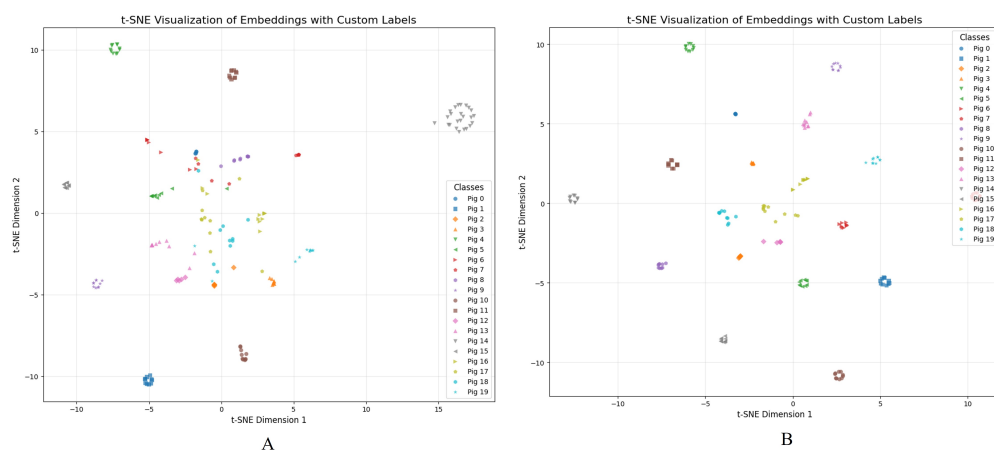


Figure 7. t-SNE Visualization of Occluded (A) vs. Inpainted (B) Pig Faces Using EfficientNet-B2 on Testing dataset of Pig Face Recognition Dataset.

Figure 8 compares two confusion matrices reflecting EfficientNet-B2's classification performance on (A) raw, fence-occluded pig faces and (B) faces restored via PigFRIS. In (A), a significant portion of the off-diagonal cells exhibit high values, indicating frequent misclassifications among individual pig IDs. These errors stem from missing or distorted facial cues—such as partially covered eyes or snouts—that obscure critical distinguishing traits. By contrast, (B) reveals a confusion matrix with a more pronounced diagonal and substantially lower off-diagonal counts. This improvement underscores how removing the fence and reconstructing the occluded regions enhances each pig's unique facial signature, allowing the model to more accurately differentiate among individuals. Ultimately, the clearer diagonal trend in the inpainted scenario confirms the transformative impact of our GAN-based restoration process on overall recognition accuracy.

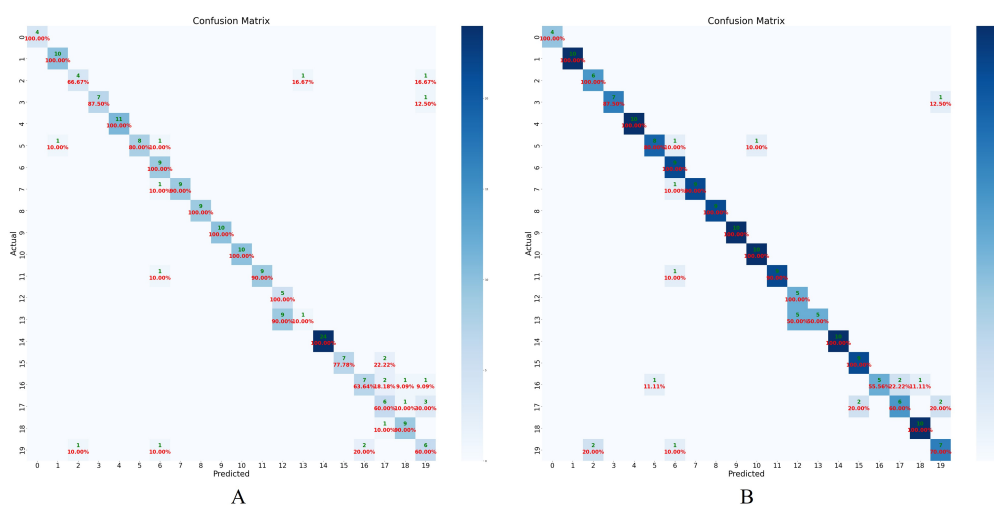


Figure 8. Confusion Matrix for Occluded (A) vs. Inpainted Pig Faces (B) Using EfficientNet-B2 on Testing dataset of Pig Face Recognition Dataset.

5. Discussion

The experimental findings illustrate that effectively handling occlusions is critical for improving pig face recognition performance in complex farm environments. Our occlusion segmentation module, based on the YOLOv11 model, provides highly accurate masks that are fundamental to the subsequent inpainting step. Without precise segmentation, the inpainting model would be forced to guess which regions need restoration, potentially leading to suboptimal reconstructions or the introduction of visually inconsistent artifacts. For example, in challenging lighting conditions or when fences are partially visible, the segmentation module sometimes fails to capture all occluded areas, which, in turn, causes the inpainting stage to generate blurred or distorted features. The reliable identification of occluded areas ensures that the generator focuses its efforts where it matters most—on restoring essential facial features rather than regions that are already visible.

By integrating this accurate occlusion segmentation with an advanced inpainting approach (AOT-GAN), our pipeline recovers hidden facial traits and guides the recognition model toward meaningful discriminative cues. The results confirm that when fences and other obstructions are removed at the pixel level, recognition models no longer resort to non-informative patterns for classification. Instead, they learn to rely on authentic pig facial features, ultimately increasing classification accuracy and robustness. Nevertheless, certain failure cases still persist. For instance, when the inpainting module encounters large occluded regions or complex textures—such as around the eyes or snout—it may produce reconstructions with subtle distortions or fail to preserve delicate expression nuances. These issues sometimes lead to recognition errors, particularly in instances where pigs exhibit very similar facial markings or when rapid head movements create motion blur. The presence of these failure cases highlights that even state-of-the-art segmentation and inpainting methods have limitations, suggesting that more pig-specific data and context-aware restoration algorithms are necessary to achieve consistently high-fidelity results. While leveraging pre-trained models on human-centric datasets like CelebA offers a shortcut to realistic texture generation, domain mismatches may occasionally distort the outcome. In future research, we plan to collect broader and more diverse pig face datasets, capturing a wider range of environmental conditions and facial variations. Additionally, exploring hybrid approaches that integrate other data modalities (e.g., thermal imaging or depth data) and incorporating more sophisticated generative frameworks could further improve restoration quality. Furthermore, these improvements in recognition underscore the potential impact of occlusion segmentation and inpainting on real-world livestock management systems. By ensuring that the model's focus shifts from reliance on accidental fence patterns to genuine facial characteristics, the system gains better generalizability. In increasingly automated and data-driven farming environments, this reliability stands to benefit animal welfare monitoring, health assessment, and behavior analysis. To this end, future research should also focus on integrating the pipeline with advanced hardware solutions. Our system is designed with practical deployment in mind, being compatible with edge computing platforms and resource-constrained hardware. We plan to deploy our system using JSON Nano hardware in pig farms in Korea, aiming for real-time monitoring and efficient processing in actual field conditions. Additionally, hardware optimizations such as model quantization, pruning, and energy-efficient inference methods are promising areas to explore.

In summary, the occlusion segmentation model plays a foundational role, enabling inpainting to function optimally and thereby enhancing pig face recognition accuracy. These results underscore the importance of treating the data pipeline holistically—accurately identifying and removing occlusions at their source is as critical as the subsequent reconstruction and classification processes. This discussion not only underscores the efficacy of

our proposed system but also outlines clear future research directions, particularly in improving the inpainting module to achieve more accurate pig face completion and ultimately elevate recognition performance. These avenues pave the way for more robust, adaptable, and data-efficient recognition frameworks, advancing the field of precision farming and intelligent animal husbandry.

6. Conclusions

In this study, we introduced an integrated pipeline that combines occlusion segmentation, inpainting, and classification to significantly enhance pig face recognition in challenging farm environments. Our approach transforms unreliable, fence-obstructed images into refined facial representations that drive more accurate and robust recognition, offering a practical solution for smart farming. Rather than relying on incidental, non-informative cues, PigFRIS focuses on authentic pig facial features, thereby improving overall recognition performance and enabling real-world applications such as livestock monitoring, health assessment, and behavior analysis. The system's design also supports deployment on edge computing platforms, making it suitable for resource-constrained farm environments and real-time operation. We acknowledge that some limitations persist. Variations in lighting conditions, rapid animal movements, and other environmental factors can occasionally affect the precision of occlusion detection and inpainting, imposing real-time constraints on the system. To address these challenges, future research will focus on expanding the diversity of pig-specific datasets, refining our inpainting model for higher-fidelity restoration, and exploring advanced hardware optimizations—such as model quantization and pruning—for smoother, real-time performance. Overall, this work not only advances the state of the art in pig face recognition but also paves the way for more efficient, welfare-oriented smart farming solutions. By bridging advanced computer vision techniques with practical farm applications, PigFRIS establishes a solid foundation for transformative improvements in precision livestock management.

Author Contributions: R.M. and S.C. developed the overall system framework, collected the data, and wrote the manuscript. S.C. and R.M. conducted all experiments. S.K. and H.K. supervised the entire implementation process and provided critical revisions to enhance the quality of the manuscript. All authors have read and approved the final version of the manuscript for publication.

Funding: This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS-2019-NR040079), and by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through the Agri-Bioindustry Technology Development Program, funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (RS-2025-02307882).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated and/or analyzed during the current study are not publicly available due to confidentiality agreements with participants but are available from the corresponding author upon reasonable request.

Acknowledgments: We extend our appreciation to the reviewers and editors for their constructive feedback, which has significantly enhanced the quality of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Xu, S.; He, Q.; Tao, S.; Chen, H.; Zhu, S.; Huang, Z. Pig Face Recognition Based on Trapezoid Normalized Pixel Difference Feature and Trimmed Mean Attention Mechanism. *IEEE Trans. Instrum. Meas.* **2022**, *72*, 1–13. [\[CrossRef\]](#)
- Jonna, S.; Nakka, K.K.; Sahay, R.R. Deep learning based fence segmentation and removal from an image using a video sequence. *arXiv* **2016**, arXiv:1609.07727.
- Maselyne, J.; Saeys, W.; De Ketelaere, B.; Mertens, K.; Vangeyte, J.; Hessel, E.F.; Millet, S.; Van Nuffel, A. Validation of a High Frequency Radio Frequency Identification (HF RFID) system for registering feeding patterns of growing-finishing pigs. *Comput. Electron. Agric.* **2014**, *102*, 10–18.
- Pereira, E.; Araújo, Í.; Silva, L.F.V.; Batista, M.; Júnior, S.; Barboza, E.; Santos, E.; Gomes, F.; Fraga, I.T.; Davanzo, R.; et al. RFID Technology for Animal Tracking: A Survey. *IEEE J. Radio Freq. Identif.* **2023**, *7*, 609–620. [\[CrossRef\]](#)
- Schindler, F.; Steinhage, V. Identification of animals and recognition of their actions in wildlife videos using deep learning techniques. *Ecol. Inform.* **2021**, *61*, 101215. [\[CrossRef\]](#)
- Xu, B.; Wang, W.; Guo, L.; Chen, G.; Li, Y.; Cao, Z.; Wu, S. CattleFaceNet: A cattle face identification approach based on RetinaFace and ArcFace loss. *Comput. Electron. Agric.* **2022**, *193*, 106675. [\[CrossRef\]](#)
- Li, G.; Sun, J.; Guan, M.; Sun, S.; Shi, G.; Zhu, C. A New Method for non-destructive identification and Tracking of multi-object behaviors in beef cattle based on deep learning. *Animals* **2024**, *14*, 2464. [\[CrossRef\]](#)
- Wan, Z.; Tian, F.; Zhang, C. Sheep face recognition model based on deep learning and bilinear feature fusion. *Animals* **2023**, *13*, 1957. [\[CrossRef\]](#)
- Mougeot, G.; Li, D.; Jia, S. A deep learning approach for dog face verification and recognition. In Proceedings of the PRICAI 2019: Trends in Artificial Intelligence: 16th Pacific Rim International Conference on Artificial Intelligence, Cuvu, Yanuca Island, Fiji, 26–30 August 2019; Proceedings, Part III 16; Springer: Cham, Switzerland, 2019; pp. 418–430.
- Wang, K.; Yang, F.; Chen, Z.; Chen, Y.; Zhang, Y. A fine-grained bird classification method based on attention and decoupled knowledge distillation. *Animals* **2023**, *13*, 264. [\[CrossRef\]](#)
- Marsot, M.; Mei, J.; Shan, X.; Ye, L.; Feng, P.; Yan, X.; Li, C.; Zhao, Y. An adaptive pig face recognition approach using Convolutional Neural Networks. *Comput. Electron. Agric.* **2020**, *173*, 105386.
- Wang, Z.; Liu, T. Two-stage method based on triplet margin loss for pig face recognition. *Comput. Electron. Agric.* **2022**, *194*, 106737.
- Li, G.; Jiao, J.; Shi, G.; Ma, H.; Gu, L.; Tao, L. Fast recognition of pig faces based on improved Yolov3. In *Proceedings of the Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2022; Volume 2171, p. 012005.
- Li, G.; Shi, G.; Jiao, J. YOLOv5-KCB: A new method for individual pig detection using optimized K-means, CA attention mechanism and a bi-directional feature pyramid network. *Sensors* **2023**, *23*, 5242. [\[CrossRef\]](#)
- Ma, R.; Ali, H.; Chung, S.; Kim, S.C.; Kim, H. A lightweight pig face recognition method based on automatic detection and knowledge distillation. *Appl. Sci.* **2023**, *14*, 259. [\[CrossRef\]](#)
- Shigang, W.; Jian, W.; Meimei, C.; Jinyang, W. A pig face recognition method for distinguishing features. In Proceedings of the 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, 14–16 April 2021; pp. 972–976.
- Shi, C.; Liu, W.; Meng, J.; Jia, X.; Liu, J. Self-prior guided generative adversarial network for image inpainting. *Vis. Comput.* **2024**, 1–13. [\[CrossRef\]](#)
- Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Generative image inpainting with contextual attention. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 5505–5514.
- Zheng, C.; Cham, T.J.; Cai, J. Tfill: Image completion via a transformer-based architecture. *arXiv* **2021**, arXiv:2104.00845.
- Khanam, R.; Hussain, M. Yolov11: An overview of the key architectural enhancements. *arXiv* **2024**, arXiv:2410.17725.
- Zeng, Y.; Fu, J.; Chao, H.; Guo, B. Aggregated contextual transformations for high-resolution image inpainting. *IEEE Trans. Vis. Comput. Graph.* **2022**, *29*, 3266–3280. [\[CrossRef\]](#)
- Dubey, S.R.; Singh, S.K.; Chaudhuri, B.B. Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing* **2022**, *503*, 92–108.
- Segu, M.; Tonioni, A.; Tombari, F. Batch normalization embeddings for deep domain generalization. *Pattern Recognit.* **2023**, *135*, 109115.
- Kaur, R.; Singh, S. A comprehensive review of object detection with deep learning. *Digit. Signal Process.* **2023**, *132*, 103812.
- Ma, R.; Fuentes, A.; Yoon, S.; Lee, W.Y.; Kim, S.C.; Kim, H.; Park, D.S. Local refinement mechanism for improved plant leaf segmentation in cluttered backgrounds. *Front. Plant Sci.* **2023**, *14*, 1211075.
- Wang, X.; Song, J. ICIOU: Improved loss based on complete intersection over union for bounding box regression. *IEEE Access* **2021**, *9*, 105686–105695.
- Shang, S.; Liu, J.; Yang, Y. Multi-layer transformer aggregation encoder for answer generation. *IEEE Access* **2020**, *8*, 90410–90419.

28. Feng, X.; Song, D.; Chen, Y.; Chen, Z.; Ni, J.; Chen, H. Convolutional transformer based dual discriminator generative adversarial networks for video anomaly detection. In Proceedings of the 29th ACM International Conference on Multimedia, Chengdu, China, 20–24 October 2021; pp. 5546–5554.
29. Koonce, B.; Koonce, B. EfficientNet. *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*; Apress: Berkeley, CA, USA, 2021; pp. 109–123.
30. Naidu, G.; Zuva, T.; Sibanda, E.M. A review of evaluation metrics in machine learning algorithms. In *Proceedings of the Computer Science On-Line Conference*; Springer: Cham, Switzerland, 2023; pp. 15–25.
31. Terven, J.; Cordova-Esparza, D.M.; Ramirez-Pedraza, A.; Chavez-Urbiola, E.A.; Romero-Gonzalez, J.A. Loss functions and metrics in deep learning. *arXiv* **2023**, arXiv:2307.02694.
32. Kynkäänniemi, T.; Karras, T.; Laine, S.; Lehtinen, J.; Aila, T. Improved precision and recall metric for assessing generative models. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 3929–3938.
33. Yacoubby, R.; Axman, D. Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models. In Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems, Online, 20 November 2020; pp. 79–91.
34. Soloveitchik, M.; Diskin, T.; Morin, E.; Wiesel, A. Conditional frechet inception distance. *arXiv* **2021**, arXiv:2103.11521.
35. Snell, J.; Ridgeway, K.; Liao, R.; Roads, B.D.; Mozer, M.C.; Zemel, R.S. Learning to generate images with perceptual similarity metrics. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 4277–4281.
36. Helmrich, C.R.; Bosse, S.; Schwarz, H.; Marpe, D.; Wiegand, T. A study of the extended perceptually weighted peak signal-to-noise ratio (XPSNR) for video compression with different resolutions and bit depths. *ITU J. ICT Discov.* **2020**, *3*, 65–72.
37. Boursalie, O.; Samavi, R.; Doyle, T.E. Evaluation metrics for deep learning imputation models. In *Proceedings of the International Workshop on Health Intelligence*; Springer: Cham, Switzerland, 2021; pp. 309–322.
38. Hodson, T.O. Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geosci. Model Dev. Discuss.* **2022**, *2022*, 1–10.
39. Jocher, G.; Chaurasia, A.; Qiu, J. YOLO by Ultralytics. 2023. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 19 February 2025).
40. Wang, C.Y.; Yeh, I.H.; Mark Liao, H.Y. Yolov9: Learning what you want to learn using programmable gradient information. In *Proceedings of the European Conference on Computer Vision*; Springer: Cham, Switzerland, 2025; pp. 1–21.
41. Patwardhan, K.A.; Sapiro, G.; Bertalmio, M. Video inpainting of occluding and occluded objects. In Proceedings of the IEEE International Conference on Image Processing 2005, Genova, Italy, 14 September 2005; Volume 2, p. II-69.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.