

Synthetic Data Augmentation for Enhanced Chicken Carcass Instance Segmentation

Yihong Feng, Chaitanya Pallerla, Xiaomin Lin, Pouya Sohrabipour Sr, Philip Crandall, Wan Shou, Yu She, Dongyi Wang

Abstract—The poultry industry has been driven primarily by broiler chicken production and has grown into the world’s largest animal protein sector. Automated detection of chicken carcasses on processing lines is vital for quality control, food safety, and operational efficiency in slaughterhouses and poultry processing plants. However, developing robust deep learning models for tasks like instance segmentation in these fast-paced industrial environments is often hampered by the need for laborious acquisition and annotation of large-scale real-world image datasets.

To this end, we present the first pipeline generating photo-realistic, automatically labeled synthetic images of chicken carcasses under varying poses. We also introduce a new benchmark dataset containing 300 annotated real-world images, curated specifically for poultry segmentation research. Using these datasets, this study investigates the efficacy of synthetic data and automatic data annotation to enhance the instance segmentation of chicken carcasses, particularly when real annotated data from the processing line is scarce. A small real dataset (60 images of chicken carcasses) with varying proportions of synthetic images were evaluated in prominent instance segmentation models: YOLOv11-seg, Mask R-CNN (with R50 and R101 backbones), and Mask2Former. Results show that synthetic data significantly boosts segmentation performance for chicken carcasses across all models. YOLOv11-seg consistently achieved the highest accuracy. Notably, models with greater capacity (R101) and transformer-based architectures (Mask2Former) derived greater benefits, particularly with larger volumes of synthetic data. Furthermore, model-specific optimal ratios of synthetic-to-real data were observed. This research underscores the value of synthetic data augmentation as a viable and effective strategy to mitigate data scarcity, reduce manual annotation efforts, and advance the development of robust AI-driven automated detection systems for chicken carcasses in the poultry processing industry.

Index Terms—Synthetic Data, Chicken, Data Augmentation, Instance Segmentation, Blender, Mask-RCNN, YOLOv11.

I. INTRODUCTION

OVER the past two decades, driven by increased demand in established and emerging markets, poultry has become

Yihong Feng, Pouya Sohrabipour Sr, Dongyi Wang are with the Department of Biological and Agricultural Engineering, University of Arkansas, Fayetteville, AR 72701 USA (e-mail: yihongf@uark.edu, pouyas@uark.edu, dongyw@uark.edu).

Chaitanya Pallerla, Philip Crandall are with the Department of Food Science, University of Arkansas, Fayetteville, AR 72701 USA (e-mail: pallerla@uark.edu, crandal@uark.edu).

Xiaomin Lin is with Department of Electrical Engineering, University of South Florida, Tampa, FL 33620 USA (e-mail: xlin2@usf.edu).

Wan Shou is with Department of Mechanical Engineering, University of Arkansas, Fayetteville, AR 72701 USA (e-mail: wshou@uark.edu).

Yu She is with Department of Industrial Engineering, Purdue University, West Lafayette, IN 47907, USA (e-mail: yushe@purdue.edu). This paper was produced by the IEEE Publication Technology Group. They are in Piscataway, NJ.

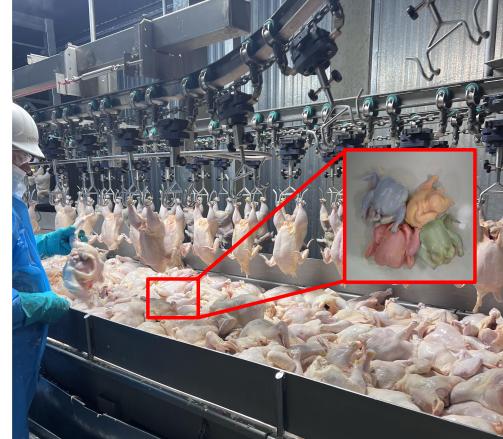


Fig. 1. Chicken carcasses on a processing line in a poultry slaughter plant. The inset displays examples from our dataset with instances of overlapping carcasses overlaid with segmentation masks, where each color represents an individual chicken carcass.

the most widely consumed animal protein worldwide [1]. Moreover, the poultry industry is projected to remain the world’s largest meat exporting sector over the next decade [1]. From 2001 to 2021, the rapid expansion of poultry production and rising consumer demand resulted in record-high global poultry imports [2]. The poultry industry of the United States takes the lead domestically and internationally, supported by advanced production structures, modern poultry genetics, abundant domestic feed resources, and strong consumer demand [2]. Global poultry sales reached \$70.2 billion in 2024, up 4% from \$67.4 billion in 2023.

Despite its economic success, the poultry industry faces significant challenges during the processing phase, particularly within slaughter facilities. The working conditions in chicken processing plants are harsh because of low environmental temperatures around 4 °C and high humidity [3]. In addition, workers are often required to work at rates dictated by 140 birds per minute processing line speeds, manually hanging chilled chicken carcasses on shackles for subsequent automated deboning and packaging tasks [4]. These repetitive and manual tasks not only result in a high incidence of workplace injuries but have also led to a shortage of skilled labor [5], which has forced the industry to seek inventive solutions that can mitigate worker safety concerns, maintain production efficiency, and ensure product quality.

In recent years, the integration of automation technologies has offered great potential to improve poultry processing practices by reducing human exposure to hazardous conditions

and improving production efficiency and food safety. In the poultry production phase, automation and robotics have made great progress in precision animal management [6], such as monitoring environmental conditions, health management [7], and egg picking [4]. However, in comparison, research on automation in poultry slaughter plants, especially for carcass handling is still limited [8] because of unique challenges in image understanding [9] and robotic manipulation [10]. Challenges to image understanding, include (a) the variability in posture and shape of chicken carcasses, piling carcasses on conveyor belts where they exhibit diverse postures such as extended wings or bowed legs; (b) visual feature similarity, since all the carcasses are often stacked, and share similar colors and textures that makes visual differentiation of individual carcasses difficult; (c) occlusion and overlap, where carcasses are partially or fully obstructing one another on the conveyor; (d) the variations of lighting conditions, including shadows and reflections degrade image quality; Without intelligent visual understanding, robotic systems would struggle to manipulate chicken carcasses with the level of precision and accuracy necessary to meet industry standards. [11]. This study focuses on the image understanding component and aims to improve image recognition accuracy and system robustness in these complex environments.

II. RELATED WORK

The integration of machine vision technologies has significantly advanced automation systems by providing intelligent input capabilities. Image instance segmentation and object detection are two main tasks in vision based industrial automation.

For object detection, YOLO(you only look once)-based models have gained popularity for real-time object detection due to their speed and robustness in throughout the agri-food industry. In viticulture, YOLOv5s with a tracking algorithm enabled real-time grape cluster counting at 50.4 FPS with 84.9% accuracy [12]. For robotic weeding, a Multimodule-YOLOv7-L model achieved 97.1% mAP at 37.3 FPS for lettuce identification and weed severity classification [13]. YOLO models have also been effectively applied to detect apples, citrus, and other fruits under diverse conditions [14]. In poultry farming, YOLOv8x-DB supports dust bathing behavior detection in cage-free hens [15], and the lightweight YOLO-Claw network achieved 97.1% mAP for limb-health assessment [16], enabling precise welfare monitoring and health assessment for growing broilers.

While object detection models like YOLO based models are good at rapidly identifying and locating objects, many agricultural and food processing applications demand a more granular understanding, requiring not only detection but also precise pixel-level delineation. This need for detailed spatial information brings instance segmentation models to the forefront.

Instance segmentation models, such as Mask R-CNN and its variations, have been widely used in agriculture for tasks such as identifying and counting fruit, monitoring plant health, and detecting pests. These tasks require not only the detection

and location of objects but also being able to differentiate among individual objects of the same class. This pixel-level classification is crucial in complex environments involving occlusions, variable shapes, and overlapping objects. Apples and strawberries are common benchmark crops in research due to their visual complexity and commercial value. For instance, a customized Mask R-CNN with deformable convolution and attention modules achieved a segmentation for apples under occlusion and varying lighting conditions [17]. Beyond the field, tasks often require objects handling and quality control where simple detection falls short. For instance, Mask R-CNN has been effectively applied to food item identification under diverse lighting conditions, showcasing its adaptability to visual challenges inherent in processing environments [18] [19]. More broadly, instance segmentation supports automated quality control, such as identifying defects in table olives [20], and facilitates robotic bin-picking of soft, deformable food items like chicken fillets [21] by providing precise location and shape information, even when items are overlapping. These applications show how instance segmentation's ability to deliver detailed, pixel-level understanding is critical for advancing automation and precision in food processing.

Despite the effectiveness of deep learning-based visual models, their deployment and performance depends largely on large-scale annotated datasets, which are particularly difficult to obtain pixel-level annotation in poultry processing due to the complex visual features of carcasses, such as irregular shapes, overlapping limbs, soft, deformable tissue, and varying postures. Furthermore, variable lighting, reflections from wet surfaces, and contamination from feathers or blood exacerbate the difficulty of consistent object recognition [22]. The Fig. 1 illustrates a representative example of these challenges which shows a typical overhead image from a poultry processing line, where multiple carcasses are partially overlapping, with irregular poses and shadow effects. Even for human annotators, precisely outlining each body part for segmentation is labor-intensive and error-prone. These issues highlight why robust instance segmentation is difficult to achieve without advanced data strategies.

To overcome the challenges of data collection and annotation, there are different image augmentation strategies that can help improve model robustness as well as resolve overlapping target objects by exposing individual objects to a variety of orientations and scales. Traditional image augmentation methods, such as flipping, rotating, cropping, and scaling, effectively expand the size of the training dataset [23]. However, these affine transformation techniques have limited capabilities to introduce new, diverse instances of objects and are not able to simulate significant changes in object appearance, such as changes in texture, occlusion, or complex deformations [24], [25]. These limitations are especially evident in poultry processing, where real-world unpredictability is notable, and traditional augmentations lack the ability to capture complex carcass handling conditions.

Synthetic data generation provides a new and more advanced method to expand the dataset for neural network training. There are several approaches to generating the synthetic data. 3D modeling and rendering: the methods rely

on physically-based rendering (PBR) engines to create synthetic scenes from scratch, simulating lighting, textures, and camera perspectives [26], [27]. Recent advances have shifted toward neural rendering methods such as Neural Radiance Fields (NeRF), which model volumetric scenes from multiple views to synthesize realistic novel perspectives [28], and 3D Gaussian Splatting (3DGS), which represents surfaces using anisotropic Gaussian primitives for efficient real-time rendering [29]. Additional applications encompass strawberry plant disease detection systems, where synthetic imagery facilitates model training for identifying various pathological stages under diverse environmental conditions [30]. These neural methods enable a more faithful replication of object-level details, dynamic geometry, and realistic occlusion. By leveraging these methods, researchers can generate complex synthetic datasets that more accurately reflect real-world variability and lighting conditions.

Building upon these advanced visual rendering techniques, physics-based simulation imitates real-world dynamics [31], which offers another pathway for synthetic data generation, focusing on reproducing the dynamic behaviors and interactions among objects and environments beyond just their visual appearance [32]. This approach generates data by emulating real-world physical laws, ensuring that the synthetic contents are not only visually plausible but also that their behavior, such as object motion, collisions, deformations, fluid dynamics, and environmental changes all adhere to physical logic. Consequently, researchers can create large datasets encompassing diverse complex scenarios, edge cases, or interactions that would be difficult to replicate in the actual processing plant [33] [34], thereby enhancing the robustness and generalization capabilities of machine learning models.

The combination of these two approaches is particularly powerful: physics-based simulation can generate 3D scenes and events with complex dynamic behaviors and physical interactions, while neural rendering techniques like 3DGS can efficiently render these dynamic, physically plausible scenes from arbitrary new viewpoints with high fidelity. This synergy allows the generated synthetic data to achieve high standards in both visual realism and behavioral authenticity, providing invaluable data resources for training more robust vision models.

In addition, style transfer and image synthesis using generative models can produce new data variations for training purposes [35]. Generative Adversarial Network (GAN) [36] is one of the generative frameworks designed to synthesize new data with the same characteristics that visually resemble the data in the training set [37], [38]. In agriculture community, GANs enhance dataset have been used for tasks like weed segmentation [39], apple detection [40], and oyster recognition [41]. GANs improve model accuracy and robustness, addressing challenges like data scarcity and environmental variability. However, GANs have limitations, including mode collapse, where the model generates limited diversity in outputs, and training instability due to the adversarial process. Additionally, GAN-generated data may introduce biases, potentially affecting model accuracy in real-world applications [42];

Another prominent generative framework is the diffusion

model [43] [44]. These models operate through a two-stage process: systematically adding noise to data and then training a neural network to remove the noise, iteratively by transforming the added noise into coherent data. Diffusion models have gained acceptance for their ability to generate high-fidelity and diverse images [45], often surpassing other models in sample quality and training stability. This makes diffusion models promising for specialized applications, such as generating synthetic oysters for aquaculture [46], [47] or augmenting datasets for rare plant diseases or produce defects. However, a primary limitation of these models is their slow, iterative sampling process, which can be computationally intensive compared to single-pass generation methods [48]. Continuing to address these sampling speed challenges and explore their utility in various scientific domains.

Although most synthetic data generation methods significantly increase the quantity and visual realism of training datasets, they still do not address one of the most persistent difficulties in deep learning: the time consuming and labor intensive task of manual annotation. Even when synthetic images are available, many applications still require manual masks or bounding boxes annotation to ensure model accuracy. This makes it difficult to scale up datasets efficiently.

In this paper, we propose a practical framework that generates automatically labeled synthetic data and combines it with real datasets under various configurations to enhance instance segmentation performance using an initial example from the poultry processing industry. We evaluated three leading deep learning models: Mask R-CNN, Mask2Former, and YOLOv11 across five dataset settings, including a real chicken carcass dataset baseline and four augmented datasets with up to 1000 Blender-generated synthetic images. Blender, a 3D rendering platform allows for precise control of the scene parameters and automatic generation of accurate annotations, offers greater structural fidelity and annotation reliability compared to other generative models like GANs or diffusion methods, which are particularly important in tasks involving non-rigid, anatomically varied, and piled objects. Our study focuses on the rehanging phase of poultry processing, where intact broiler carcasses are typically piled on a take-away belt when exiting the cold-water chiller and must be rehung on an overhead, moving schackle line. It is critical that reliable real-time segmentation will be available for the next phase, using robots to perform this precise identification of the orientation of the piled poultry carcass, pick-up an individual carcass and hang it by the hocks on the moving schackle line. This study presents a simulation-to-reality (Sim2Real) framework for generating synthetic training data targeting the segmentation of occluded chicken carcasses in dense configurations. This work supports broader efforts to improve the worker safety, efficiency, consistency, and scalability of poultry production systems through data efficient solutions.

III. METHODOLOGY

This study elaborates on the collection of real-world and synthetic chicken carcass segmentation datasets, subsequently proposing a dataset augmentation method for improving poultry instance segmentation performances. In this study, a deep

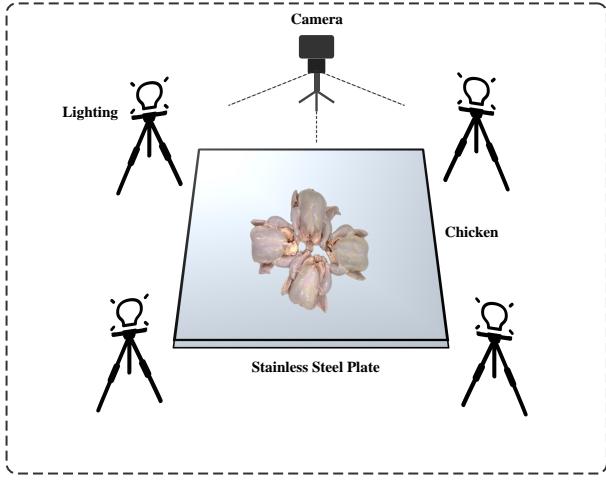


Fig. 2. The real chicken data acquisition system

neural network was trained on a combination of a small-scale annotated real-world data and large-scale synthetic data rendered from Blender (a computer graphics software). In this experiment, the real-world data was collected using a custom hardware system, as shown in Fig. 2, however, obtaining such datasets with annotation requires significant time and resources. To address this challenge, this study adopted a “Sim-to-Real” strategy, which depends on a task-specific and data-efficient solution. By using Blender to simulate anatomically plausible carcass identification and positions and generate high-fidelity ground-truth masks, our framework directly addresses key obstacles in this specific dataset: annotation bottlenecks and visual complexity. A comparison analysis of competition deep learning models provide strong support for this method’s ability to improve segmentation accuracy and enhance practical feasibility.

A. Real-world dataset collection

We built a real-world poultry dataset using a custom image acquisition system designed to capture diverse scenarios. To ensure the sample consistency, all poultry carcasses were obtained from local grocery stores affiliated with the same supplier network that follows standardized packaging and cold-chain procedures, which helped reduce variation in carcass appearance. To simulate various real-world conditions, poultry carcasses were randomly arranged in different positional combinations, including stacked and closely placed configurations. The image acquisition system was equipped with controlled lighting to maintain consistent illumination throughout the image capture process, as shown in Figure 2. The background of the poultry carcasses was a stainless-steel plate, mimicking the stainless steel food contact surfaces used in poultry production. In total, 300 real world images were collected, which were composed of images with different carcass overlays and image capturing angles. These images form the baseline of our dataset, providing a robust performance evaluation of our segmentation model under varying conditions. Each image was manually annotated with pixel-level instance

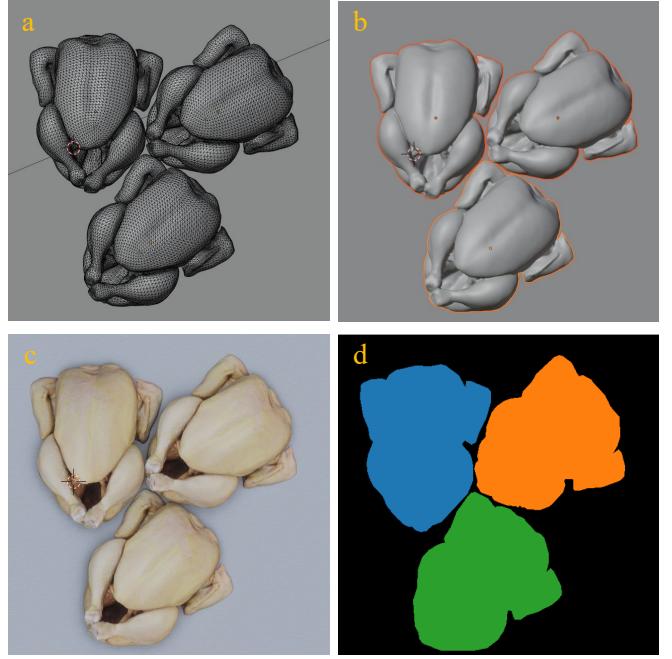


Fig. 3. An overview of our synthetic data generation workflow. (a) Mesh Preparation: High quality 3D chicken models are imported into Blender and arranged in a non-overlapping layout. Mesh resolution and topology are inspected in wireframe mode. (b) Geometry Refinement: Surface smoothness, orientation, and model boundaries are adjusted under solid viewport shading to ensure clean geometry before rendering. (c) RGB Rendering: Realistic lighting, textures, and materials are applied to generate photorealistic synthetic images using the Cycles engine. (d) Mask Generation: Each instance is assigned a unique ID to generate the corresponding segmentation masks. These automated generated masks are used as ground truth for training instance segmentation models.

segmentation masks using LabelMe [49]. This collection of 300 annotated real-world images was then divided into distinct training, validation, and testing sets to develop and evaluate our segmentation model. Additional details are discussed in Section C.

B. Synthetic Dataset Generation

Blender is an open-source 3D creation software, has extensive modeling, rendering, and animation features [50]. By using tools like Blender, researchers can generate synthetic datasets that closely mimic real-world scenarios [26] [41] [27]. Its Python API can be used to automatically generate synthetic datasets by altering the camera positions, lighting conditions, and object properties. Ground truth labels are automatically generated during the rendering process, which allows the researchers to quickly build large-scale synthetic datasets with accurate annotations. The software’s physically-based rendering engine can simulate realistic lighting, shadows, and material properties, which making the synthetic images more representative of real-world conditions.

Blender has demonstrated considerable efficiency across diverse agricultural applications for synthetic data generation. Barth et al. utilized Blender to generate a comprehensive dataset comprising 10,500 synthetic images of Capsicum annuum through systematic randomization of 3D plant mod-

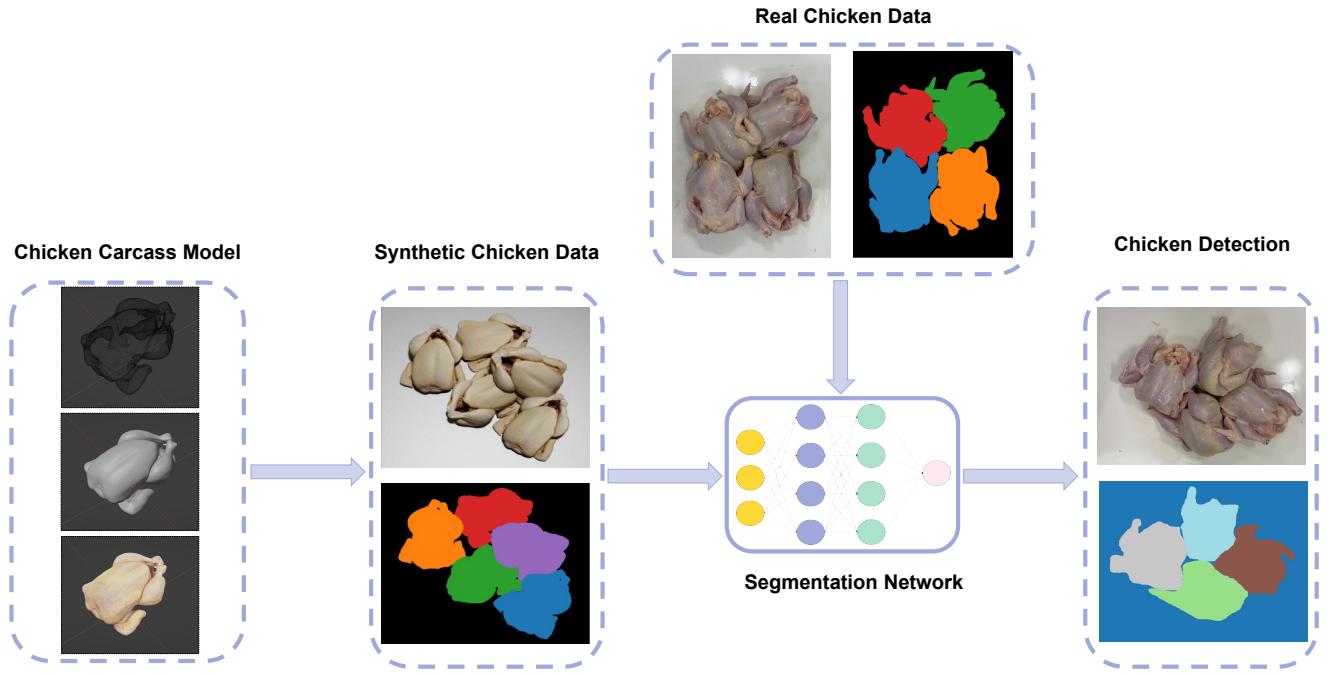


Fig. 4. An overview for Synthetic data generation and model training. The process begins with a high fidelity 3D chicken carcass model, which is used to generate a large-scale synthetic chicken dataset. This synthetic data is then combined in varying quantities with our real-world training data to form a series of distinct hybrid datasets. Each datasets are subsequently used to train and evaluate the deep learning Segmentation Networks (including Mask R-CNN with ResNet-50/101 backbones, Mask2Former, and YOLOv11-seg). This comprehensive approach allows for a systematic investigation of how synthetic data impacts performance across different architecture. The trained network can then perform robust instance segmentation (Chicken Detection) on new, unseen real-world images, accurately identifying and delineating each individual carcass even in cluttered scenes.

els based on empirical morphological measurements. This dataset was specifically designed to replicate visual conditions prevalent in greenhouse environments and facilitate instance segmentation tasks in precision agriculture [51]. The ability of Blender-generated synthetic data to substantially enhance model performance in food processing applications, by effectively capturing complex real-world variations, has been highlighted by several studies. For instance, Ummadisingu et al. demonstrated this enhancement for food item segmentation in meal-assisting robotic systems [52], and Jonker et al. showcased its benefits for chicken fillet handling in automated robotic systems [21].

While previous applications have demonstrated Blender's capabilities in agricultural contexts, poultry processing presents unique challenges that have been largely unaddressed. Unlike the aforementioned applications involving static plant models or food items, poultry carcass handling has complex anatomical variations from each carcass. Our approach leverages Blender's capabilities to develop comprehensive training datasets that capture the intricate variations inherent in industrial poultry processing, including realistic simulation of diverse carcass orientations and anatomical overlap typically encountered during rehanging operations [9]. This represents a significant advancement beyond existing applications by addressing the dynamic and morphologically complex nature of biological products typically encountered in food processing environments.

In poultry processing plants, chicken carcasses are typically spread or stacked on flat surfaces during inspection and

handling procedures. Our study focuses on addressing the challenges specific to poultry carcass instance segmentation, such as overlapping and close positioning, which make it difficult for visual models to distinguish among individual carcasses. These complex spatial arrangements also present significant challenges for human annotators, making manual labeling both time consuming and error prone, which highlights the advantages of the Blender approach.

In the virtual simulation, carcasses were placed in configurations ranging from isolated to tightly clustered or overlapped. As shown in Fig. 3, each carcass was treated as an individual input in the mask annotations, while the background remained uniform to avoid unnecessary complexity. The top-view perspective was chosen to reflect practical application scenarios of how the human operator sees the carcasses in production environments. Using this approach, a total of 1000 synthetic images were generated, each containing RGB images and the corresponding instance masks with COCO format. The synthetic dataset provided a robust training data, ensuring the model learns to handle overlapping and closely positioned carcasses effectively.

C. Deep Neural Network and training details

This paper implemented Mask R-CNN [53], Mask2Former [54], and YOLOv11-seg [55] as the baseline instance segmentation models for our analysis. Each model was systematically evaluated to determine optimal performance characteristics for industrial poultry processing applications. Following are some implementation details:

For Mask R-CNN, the model was deployed with two different backbone configurations (ResNet-50-FPN and ResNet-101-FPN) to evaluate the impact of network depth on segmentation performance. Both backbones integrated Feature Pyramid Network (FPN) to hierarchically combine multi-scale features, enhancing semantic representation across resolution levels. The Region Proposal Network (RPN) generated candidate object regions through sliding window analysis, proposing anchors that underwent refinement to localize potential carcass instances. These refined proposals subsequently facilitated the prediction head's triple task of classification, bounding box regression, and mask generation, yielding precise per-instance segmentation masks for poultry carcasses. For Mask2Former, the model was implemented with a Swin Transformer Tiny (Swin-T) backbone, which offers a hierarchical representation with shifted windows for efficient self-attention computation. This model replaced the traditional convolutional backbone with a transformer architecture, providing enhanced capability to capture global context while maintaining computational efficiency. For the YOLO model, the latest YOLO architecture (YOLOv11) was utilized with its segmentation capability, which offered a one-stage detection approach with improved speed-accuracy trade-offs compared to traditional two-stage detectors.

TABLE I
CHICKEN DETECTION MODELS TRAINED WITH REAL AND SYNTHETIC DATASETS.

Setting	Real Train	Real Val	Real Test	Synthetic Train
Real_Baseline	60	60	180	0
Rea+Syn-250	60	60	180	250
Rea+Syn-500	60	60	180	500
Rea+Syn-750	60	60	180	750
Rea+Syn-1000	60	60	180	1000

All models were trained with consistent training hyperparameters to ensure fair comparisons: 200 epochs for all models, batch sizes of 8, image resolution of 640×640 pixels, utilizing NVIDIA A100 GPU (40GB memory) provided by the Arkansas High Performance Computing Center (AHPCC), and data augmentation including random flipping (probability=0.5), random cropping (512×512), and normalization. For Mask R-CNN and Mask2Former, a step-wise learning rate schedule with linear warmup over 500 iterations and learning rate decay at epochs 30, 60, and 100. Mask R-CNN was optimized using SGD with momentum (0.9) and weight decay (0.0001), while Mask2Former utilized AdamW optimizer with parameter-wise learning rates to account for the different components in the transformer architecture [56] [54]. For YOLOv11, the default optimizer configuration is with an initial learning rate of 0.01, momentum of 0.937, and weight decay of 0.0005.

The experimental design incorporated a systematic approach to evaluate the impact of synthetic data augmentation on model performance. The 300 real-world images were randomly partitioned into training (20%), validation (20%), and test (60%) sets, with a fixed random seed to ensure reproducibility. We

allocated less data in training set and more data in the test set to prove the effectiveness of using the synthetic data for model trianing. As shown in Table I, to evaluate the contribution of synthetic data, five training combinations of the real images and synthetic data. The baseline models were trained solely on 60 real-world images, and four additional settings will include the same number of real-world data supplemented by 250, 500, 750, and 1000 synthetic images, respectively. The validation (60 images) and test (180 images) datasets remain unchanged which comprised exclusively of real-world poultry processing imagery, ensuring fair assessment of how synthetic data augmentation affects model generalization to actual industrial conditions.

D. Evaluation Metrics

To evaluate the performance of the poultry carcass segmentation model, the COCO dataset evaluation plugin provided in PyTorch, which is a standard for instance segmentation tasks. Each model processed RGB imagery and generated instance predictions comprising segmentation masks, confidence values, and localization boundaries for individual poultry carcasses. Performance evaluation centered on the instance segmentation quality using mask overlap analysis. We utilized the spatial coincidence ratio between predictions and ground-truth annotations, calculated as:

$$\text{IoU} = \frac{\text{Area of overlap}}{\text{Area of union}}. \quad (1)$$

The primary metric used for assessment was mean Average Precision (mAP), which provides a comprehensive measure of segmentation accuracy. The analysis framework reported three principal metric variants: AP₅₀ (average precision at IoU threshold 0.5), AP₇₅ (at IoU threshold 0.75), and AP (mean average precision across all IoU thresholds from 0.5 to 0.95). These metrics were calculated separately for both instance masks and bounding boxes to evaluate segmentation and localization performance independently.

IV. RESULTS

Table II summarises the quantitative performance of four instance-segmentation frameworks, MaskR-CNN with ResNet-50 (R50) and ResNet-101 (R101) backbones, Mask2Former, and YOLOv11-Seg trained on five different dataset settings. The real-only configuration is denoted **Real_Baseline**, whereas **Real+Syn-N** ($N \in \{250, 500, 750, 1000\}$) indicates synthetic images were added to the same real set. All models were evaluated on same 180-real world images in the test set using the COCO-style mAP, which averaged precision over IoU thresholds from 0.50 to 0.95. AP₅₀ and AP₇₅ were reported to capture low and high IoU performance.

A. Overall trends

For all models, adding synthetic images consistently improved both detection and segmentation metrics. However, the extent of this improvement and the optimal amount of

TABLE II
PERFORMANCE OF INSTANCE SEGMENTATION MODELS TRAINED WITH VARYING AMOUNTS OF SYNTHETIC DATA.

Model	Setting	bbox_mAP	bbox_mAP_50	bbox_mAP_75	segm_mAP	segm_mAP_50	segm_mAP_75
Mask R-CNN (R50)	Real_Baseline	0.740	0.963	0.866	0.746	0.965	0.863
	Real+Syn-250	0.761	0.962	0.861	0.736	0.954	0.838
	Real+Syn-500	0.784	0.959	0.883	0.755	0.959	0.865
	Real+Syn-750	0.795	0.962	0.896	0.765	0.961	0.877
	Real+Syn-1000	0.793	0.960	0.886	0.759	0.950	0.871
Mask R-CNN (R101)	Real_Baseline	0.750	0.960	0.869	0.733	0.956	0.836
	Real+Syn-250	0.795	0.962	0.883	0.751	0.952	0.856
	Real+Syn-500	0.773	0.961	0.884	0.751	0.951	0.857
	Real+Syn-750	0.776	0.964	0.874	0.753	0.963	0.860
	Real+Syn-1000	0.805	0.962	0.881	0.766	0.953	0.875
Mask2Former	Real_Baseline	0.434	0.712	0.437	0.523	0.795	0.597
	Real+Syn-250	0.530	0.776	0.548	0.669	0.899	0.739
	Real+Syn-500	0.652	0.896	0.694	0.737	0.933	0.835
	Real+Syn-750	0.672	0.896	0.715	0.726	0.925	0.809
	Real+Syn-1000	0.661	0.880	0.694	0.737	0.927	0.816
YOLOv11-seg	Real_Baseline	0.835	0.969	0.890	0.784	0.959	0.846
	Real+Syn-250	0.845	0.965	0.891	0.817	0.958	0.874
	Real+Syn-500	0.842	0.966	0.888	0.817	0.964	0.883
	Real+Syn-750	0.857	0.966	0.906	0.828	0.965	0.891
	Real+Syn-1000	0.852	0.972	0.896	0.833	0.971	0.898

synthetic data depended on the model architecture and the specific evaluation metric. For example, the Mask2Former model, which had a lower baseline, showed a substantial relative increase with synthetic data. More robust baseline models like YOLOv11-seg also benefited, enhancing the high performance more, particularly with larger volumes of synthetic images. There was not always a linear improvement with the addition of more synthetic data; for some models, a diminishing return or even a slight decrease in performance was observed after a certain amount of synthetic data was added. For instance, Mask R-CNN R50 beyond 750 synthetic images.

B. Effect of the synthetic to real ratio

The real training dataset was set to 60 images. The experiments explored synthetic-to-real data ratios by adding 0, 250, 500, 750, or 1000 synthetic images, corresponding to ratios of 0:60 (baseline), approximately 4.17:1 (Syn-250), 8.33:1 (Syn-500), 12.5:1 (Syn-750), and 16.67:1 (Syn-1000). For Mask R-CNN (R50), a ratio around 12.5:1 appeared optimal for both *bbox_mAP* (0.795) and *segm_mAP* (0.765). For Mask R-CNN (R101), the highest *bbox_mAP* (0.805) and *segm_mAP* (0.766) were achieved at the largest ratio of approximately 16.67:1, suggesting it could leverage more synthetic data effectively than its R50 model. Mask2Former showed impressive gains up to a ratio of 8.33:1, achieving *bbox_mAP* of 0.652 and *segm_mAP* of 0.737. Further increases to 12.5:1 yielded the peak *bbox_mAP* (0.672), while *segm_mAP* was similar or peaked at 16.67:1 with 0.737. YOLOv11-Seg consistently benefited from increasing synthetic-to-real ratios. The best *bbox_mAP* (0.8570) was at a 12.5:1 ratio, while the highest *segm_mAP* (0.8325) and several other key metrics (*bbox_mAP_50*, *segm_mAP_50*, *segm_mAP_75*) peaked at the 16.67:1 ratio.

Those results indicate that different model capacities and architectures respond differently to the proportion of syn-

thetic data. While lighter models like R50 stop improving significantly as more synthetic data is added, but deeper or transformer-based architectures continue to benefit from larger synthetic datasets.

C. Model-Specific Performance Analysis

1) Mask R-CNN (R50 and R101) and Backbone Comparison:

- Mask R-CNN (R50): The baseline *bbox_mAP* was 0.740 and *segm_mAP* was 0.746. Adding synthetic data generally improved performance, with the “Real+Syn-750” setting yielding the best results (*bbox_mAP* 0.795, *segm_mAP* 0.765). Performance slightly decreased with 1000 synthetic images.
- Mask R-CNN (R101): The baseline *bbox_mAP* (0.750) was slightly higher than R50, but *segm_mAP* (0.733) was lower. The R101 backbone generally benefited from more synthetic data than R50, achieving its peak *bbox_mAP* (0.805) and *segm_mAP* (0.766) with 1000 synthetic images.
- Backbone Comparison (R50 vs. R101): The deeper R101 backbone, achieved a slightly higher peak *bbox_mAP* and *segm_mAP* (0.805 and 0.766 respectively with 1000 synthetic images) compared to the R50 backbone’s peak (0.795 *bbox_mAP* and 0.765 *segm_mAP* with 750 synthetic images). The R101 also seemed to better utilize larger amounts of synthetic data (peaking at 1000 images) compared to R50 (peaking at 750 images). However, the baseline performance of R101 for segmentation was only slightly lower than R50’s. At their respective optimal synthetic data settings, R101 slightly outperformed R50.

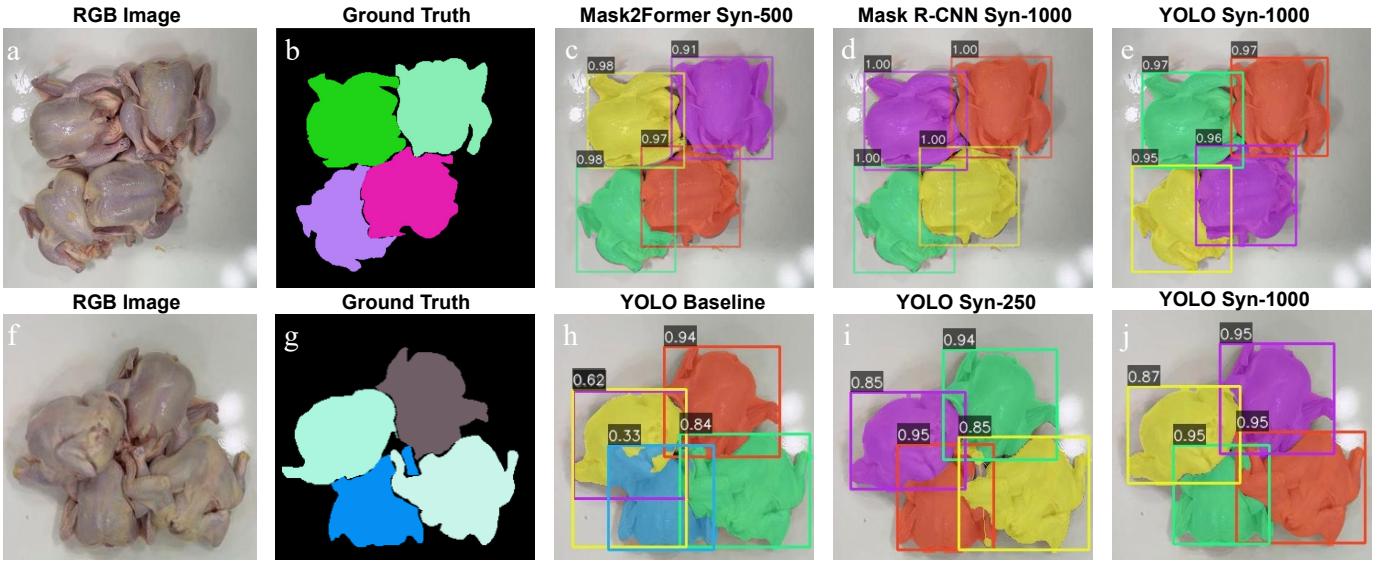


Fig. 5. Comparison of bounding box and instance segmentation results obtained with the best-performing training ratios for each model (Mask-RCNN with ResNet-101 backbones, Mask2Former, and YOLOv11-seg) and with progressively larger synthetic sets for YOLOv11-seg. Example predictions on two real chicken-carcass images. (a, f) Input RGB images. (b, g) Ground-truth instance masks. (c–e) Predictions for the raw image (a) using the three baseline architectures, each shown with the training configuration that yielded its highest COCO mAP_{0.50 : 0.95}: (c) Mask2Former trained with Real + Syn-500, (d) Mask R-CNN (ResNet-101) trained with Real + Syn-1000, and (e) YOLOv11-seg trained with Real + Syn-1000. (h–j) Predictions for image (f) from YOLOv11-seg models trained with increasing synthetic-to-real ratios: (h) Real-only baseline, (i) Real + Syn-250, and (j) Real + Syn-1000. Masks are colour-coded per instance and overlaid with the associated bounding box and confidence score.

D. Bounding Box Detection and Instance Segmentation Quality

The qualitative impact of these metric improvements is visualized in Figure. 5. The top row directly compares the best configurations of the three architectures. While all models perform well, notable differences in detail are apparent. For instance, focusing on the bottom-left carcass, the mask contour from YOLOv11-seg (e) provides the tightest fit to the ground-truth edges, whereas the masks from Mask R-CNN (R101) (d) and Mask2Former (c) exhibit a degree of under-segmentation in this area, failing to capture the full details of the object.

The bottom row of Figure. 5 provides a clear qualitative illustration of the value of synthetic data. While the baseline model (h) detected all instances, its segmentation quality and confidence scores were notably deficient, particularly for the two lower, touching carcasses. This performance is substantially improved with the addition of synthetic data. When increasing to 1000 synthetic images (j), the mask delineations for these two lower carcasses become significantly more precise, and their confidence scores dramatically increase from 0.33 and 0.84 in (h) to 0.95 for both in (j). This visually demonstrates that synthetic data not only boosts the model’s detection confidence but also effectively refines segmentation boundaries in complex, overlapping scenes.

Generally, bounding box detection (*bbox_mAP*) were similar with instance segmentation (*segm_mAP*), but with some difference:

- For Mask R-CNN (R50), the optimal amount of synthetic data (Syn-750) was the same for both *bbox_mAP* and *segm_mAP*.

- For Mask R-CNN (R101), 1000 synthetic images yielded the best results for both *bbox_mAP* and *segm_mAP*.
- For Mask2Former, *bbox_mAP* peaked with 750 images, while *segm_mAP* peaked with 500 and 1000 synthetic images, indicating a slightly different optimal point or a plateau for segmentation.
- For YOLOv11-seg, *bbox_mAP* peaked with 750 synthetic images, while *segm_mAP* (and related *segm_mAP₇₅*) continued to improve up to 1000 synthetic images, suggesting that segmentation quality, which requires more precise localization, might benefit more from larger synthetic datasets with this model.

The *mAP₅₀* scores (for both *bbox* and *segm*) were generally very high across all models and settings with synthetic data, often above 0.95 for Mask R-CNN and YOLOv11-seg. This indicated that all models became proficient at detecting and segmenting the easier instances (IoU threshold of 0.5). These improvements from synthetic data were often more pronounced in the stricter *mAP₇₅* metric and the overall *mAP* (0.50:0.95), highlighting that synthetic data helped improve the precise localization capabilities of the models. For example, the *segm_mAP₇₅* of YOLOv11-seg increased 5.72% from 0.8461 (baseline) to 0.8975 (with 1000 synthetic images).

Summary of Results

The addition of synthetic data was broadly beneficial for all instance segmentation models tested, improving both bounding box detection and instance segmentation performance. YOLOv11-seg demonstrated the highest overall performance, effectively leveraging up to 1000 synthetic images to achieve top scores, particularly in segmentation metrics. This quanti-

tative superiority was also reflected in the qualitative analysis, where YOLOv11-seg consistently produced more precise segmentation masks compared to the other architectures. Mask R-CNN models also showed consistent improvements, with the R101 backbone benefiting from a larger volume of synthetic data than the R50. Mask2Former, while starting from a lower baseline, exhibited significant relative gains, underscoring the value of synthetic data for this architecture. Notably, for several models, an optimal balance of synthetic to real data was identified, beyond which performance gains plateaued or, in some isolated cases, slightly decreased. The ideal amount of synthetic data and the resulting performance gains depended on the models, and these improvements were typically more pronounced at higher IoU thresholds. The visual evidence strongly supported this, qualitatively demonstrating how synthetic data transformed low-confidence, imprecise detections into high-confidence, accurate segmentations, especially in challenging cases of object overlap.

V. DISCUSSION

In this paper, the observed performance improvements across various instance segmentation models when augmenting real training data with synthetic images can be attributed to several factors, which align with, and in some cases nuance, findings from previously published literature. Primarily, the introduction of synthetic data likely increased the diversity of training samples. This exposed the models to a wider range of chicken carcass orientations, occlusions, and background variations than were available in our limited real dataset (60 images). Such enhanced diversity generally helps models generalize better to unseen data and reduces overfitting to the specific characteristics of a small real dataset. For instance, synthetic data can provide numerous examples of challenging scenarios, such as chicken carcasses in dense clusters or under unusual lighting, which might be rare or absent in limited real-world imagery. This principle is well-supported by other studies. For example, early work by Richter et al. demonstrated the enhanced performance of using synthetic data from game engines for semantic segmentation [57], and Tremblay et al. showed benefits for object detection using domain randomization [58]. More recently, Vanherle et al. developed a Gaussian Splatting based pipeline specifically for generating high-quality, context-aware, and varied synthetic data for instance segmentation, emphasizing its role in creating diverse training instances and reducing the domain gap [59]. Furthermore, studies by Eli et al. on face parsing [60] and Jordan et al. in zero-shot classification highlight that the diversity of content within synthetic datasets can be even more crucial for model generalization than achieving perfect photorealism, a notion that resonates with our findings where varied synthetic chicken carcass data proved beneficial [61].

The superior performance of the YOLOv11-seg model in our experiments, consistently achieving the highest accuracy metrics, can be attributed to its advanced architecture [62], which effectively combines efficient object detection with segmentation capabilities. Its pre-training step might also contribute to its robust feature extraction. This observation

aligns with the general trend in the YOLO family, where newer iterations consistently push performance boundaries in detection and segmentation tasks, as documented in various reviews and benchmark studies on YOLO architecture evolution [63].

Furthermore, the ability of the Mask R-CNN model with an R101 backbone effectively leverage larger proportions of synthetic data compared to its R50 counterpart suggests that backbone capacity plays a significant role. A deeper network like R101 has a higher model capacity, allowing it to learn more complex features and patterns from a larger and more diverse dataset, such as that augmented with synthetic images, without saturating as quickly. This resonates with fundamental concepts in deep learning [64]. More recent research by Yu et al. in the context of domain adaptation for remote sensing image classification also found that ResNet-101 achieved higher average accuracy gains (5.22%) compared to ResNet-50 (4.99%) when adapting to new domains, suggesting deeper architectures can indeed better utilize information from varied data distributions [65].

The substantial relative performance gains exhibited by Mask2Former when trained with augmented datasets also warrant specific discussion. This finding is likely linked to its transformer-based architecture. Transformer models are often characterized as being “data-hungry”, and their attention mechanisms may particularly benefit from the increased variability and quantity of training instances offered by synthetic data. This allows them to learn more robust and generalizable representations. This observation is echoed by broader research into transformer architectures in computer vision, such as the Vision Transformer (ViT) proposed by Dosovitskiy et al., which demonstrated that transformers can achieve remarkable performance, especially when pre-trained on large-scale datasets and subsequently fine-tuned on diverse data [66].

For the several models tested, an optimal balance of synthetic to real data was identified, beyond which the segmentation performance plateaued or, in some isolated cases, slightly decreased. This observation suggests a critical trade-off. While synthetic data provides valuable diversity, an overreliance on it can be counterproductive, especially when a noticeable “domain gap” exists between the synthetic and real images. In such cases, two distinct problems can arise. First, the model may begin to overfit the synthetic data, learning to recognize rendering artifacts or other patterns that are unique to the artificial images but are not present in real-world scenarios. Second, the sheer volume of synthetic examples can dilute the influence of the limited real data during training, causing the model to overlook or “forget” subtle but crucial features that are only present in the authentic images. This nuanced behavior, where our specific models like YOLOv11-seg and Mask2Former showed sustained benefits up to a certain tested point, contrasts with some studies that might report more rapidly diminishing returns from synthetic data. Recent research supports this observation. For instance, Regina et al. in their study on drone detection, found that fine-tuning models pre-trained on synthetic data with small shares (5-10%) of real data significantly boosts performance, highlighting the importance of real data as an “anchor” [67]. Chang et al. found that for multi-object tracking, 60-80% of real data could

be substituted with synthetic data without performance loss, emphasizing the role of flexible data generators in narrowing domain gaps [68]. Conversely, relying too heavily on synthetic data, especially if it is not perfectly aligned with the target domain or lacks sufficient diversity, can be detrimental for the model performance. Li et al. explicitly stated that directly using synthetic data from diffusion models can degrade performance due to feature distribution discrepancies [69], and Tremblay et al. mentioned that low-quality synthetic samples can impede learning [58]. Research into language models by Ilia et al. also warns of “model collapse” when recursively training on purely synthetic data, suggesting a maximal proportion of synthetic data is advisable when mixed with real data to avoid degradation [70]. This underscores that the optimal data blend and the tolerance for synthetic data can be highly model-specific and task-dependent, influenced by factors such as model architecture, capacity, the quality and diversity of the synthetic data, and the nature of the synthetic-to-real domain gap.

These model-specific insights into leveraging synthetic data are particularly pertinent given the well-documented challenges in data acquisition and annotation within the field of poultry. Our findings offer a potential mitigation strategy for these challenges, aligning with the successful application of synthetic data in related agricultural domains, such as precision agriculture and food processing. By demonstrating the effective use of synthetic data for chicken instance segmentation, our work contributes to the growing body of evidence supporting its role in advancing AI applications in agriculture where data scarcity is a common bottleneck.

VI. LIMITATION AND FUTURE WORK

Despite the promising results, this study has several limitations that warrant acknowledgment. First, the size of our real dataset (60 images for training) is quite small. While this highlights the utility of synthetic data in data-scarce scenarios, it might also exaggerate the perceived impact of synthetic augmentation, as the baseline performance with only real data is likely to be sub-optimal. Second, the realism and specific characteristics of our synthetic data, while designed to be diverse, are still an approximation of real-world variability. There might be inherent biases in the generation process, or it may lack certain subtle differences present in real images, potentially creating a domain gap that could limit performance on truly unseen real-world data. Third, our investigation was limited to a specific set of instance segmentation models. Other architectures, including different YOLO variants or emerging segmentation models, might respond differently to synthetic data augmentation. Finally, our evaluation was conducted on a specific test set derived from a similar environment as the real training data. The generalization capability of the trained models to completely different plant environments remains to be thoroughly tested.

Several avenues for future research from our findings and limitations: 1) Advanced Synthetic Data Generation: investigate more sophisticated synthetic data generation techniques, such as Generative Adversarial Networks (GANs), diffusion

models or physics-based simulation, which could produce even more realistic and diverse training samples, potentially reducing the domain gap. 2) Real-World Robustness Testing: Conduct extensive testing of the trained models in diverse, real-world chicken processing line environments to assess their generalization capabilities and practical deployment readiness. 3) Exploring Synthetic-to-Real Ratios: Perform a more granular analysis of the impact of different synthetic-to-real data ratios, potentially developing adaptive strategies for determining the optimal mix for specific model architectures and dataset characteristics. 4) Cost-Benefit Analysis: Undertake a formal cost-benefit analysis comparing the resources required for generating and curating high-quality synthetic data versus the cost and effort of acquiring and annotating more real data. 5) Domain Adaptation Techniques: Explore unsupervised or semi-supervised domain adaptation techniques to further bridge the gap between the synthetic and real data domains, aiming to improve model performance when real labeled data is extremely scarce.

ACKNOWLEDGEMENTS

This work was supported by awards no. 2023-67021-39072, 2023-67022-39074, and 2023-67022-39075 from the U.S. Department of Agriculture (USDA)’s National Institute of Food and Agriculture (NIFA) in collaboration with the National Science Foundation (NSF) through the National Robotics Initiative (NRI) 3.0. This research is also supported by the Arkansas High Performance Computing Center, which is funded through multiple National Science Foundation grants and the Arkansas Economic Development Commission.

REFERENCES

- [1] “USDA ERS - Poultry Expected To Continue Leading Global Meat Imports as Demand Rises,” <https://www.ers.usda.gov/amber-waves/2022/august/poultry-expected-to-continue-leading-global-meat-imports-as-demand-rises/>.
- [2] “Poultry & Eggs - Sector at a Glance | Economic Research Service,” <https://www.ers.usda.gov/topics/animal-products/poultry-eggs/sector-at-a-glance>.
- [3] I. De Medeiros Esper, P. J. From, and A. Mason, “Robotisation and intelligent systems in abattoirs,” *Trends in Food Science & Technology*, vol. 108, pp. 214–222, Feb. 2021.
- [4] G. Ren, T. Lin, Y. Ying, G. Chowdhary, and K. Ting, “Agricultural robotics research applicable to poultry production: A review,” *Computers and Electronics in Agriculture*, vol. 169, p. 105216, Feb. 2020.
- [5] N. F. Dias, A. S. Tirloni, D. Cunha dos Reis, and A. R. P. Moro, “The effect of different work-rest schedules on ergonomic risk in poultry slaughterhouse workers,” *Work*, vol. 69, no. 1, pp. 215–223, 2021.
- [6] W. F. Pereira, L. da Silva Fonseca, F. F. Putti, B. C. Góes, and L. de Paula Naves, “Environmental monitoring in a poultry farm using an instrument developed with the internet of things concept,” *Computers and electronics in agriculture*, vol. 170, p. 105257, 2020.
- [7] Q. Tong, J. Wang, W. Yang, S. Wu, W. Zhang, C. Sun, and K. Xu, “Edge ai-enabled chicken health detection based on enhanced fcos-lite and knowledge distillation,” *Computers and Electronics in Agriculture*, vol. 226, p. 109432, 2024.
- [8] D. Yang, D. Cui, and Y. Ying, “Development and trends of chicken farming robots in chicken farming tasks: A review,” vol. 221, no. C. [Online]. Available: <https://doi.org/10.1016/j.compag.2024.108916>
- [9] P. Sohrabipour, C. K. R. Pallerla, A. Davar, S. Mahmoudi, P. Crandall, W. Shou, Y. She, and D. Wang, “Cost-effective active laser scanning system for depth-aware deep-learning-based instance segmentation in poultry processing,” *AgriEngineering*, vol. 7, no. 3, p. 77, 2025.

- [10] A. Davar, Z. Xu, S. Mahmoudi, P. Sohrabipour, C. Pallerla, Y. She, W. Shou, P. Crandall, and D. Wang, "Chicgrasp: Imitation-learning based customized dual-jaw gripper control for delicate, irregular bio-products manipulation," *arXiv preprint arXiv:2505.08986*, 2025.
- [11] E. U. Chowdhury and A. Morey, "Application of optical technologies in the US poultry slaughter facilities for the detection of poultry carcass condemnation," *British Poultry Science*, vol. 61, no. 6, pp. 646–652, Nov. 2020.
- [12] L. Shen, J. Su, R. He, L. Song, R. Huang, Y. Fang, Y. Song, and B. Su, "Real-time tracking and counting of grape clusters in the field based on channel pruning with yolov5s," *Computers and Electronics in Agriculture*, vol. 206, p. 107662, 2023.
- [13] R. Hu, W.-H. Su, J.-L. Li, and Y. Peng, "Real-time lettuce-weed localization and weed severity classification based on lightweight yolo convolutional neural networks for intelligent intra-row weed control," *Computers and Electronics in Agriculture*, vol. 226, p. 109404, 2024.
- [14] Z. Liu, R. R. D. Abeyrathna, R. M. Sampurno, V. M. Nakaguchi, and T. Ahamed, "Faster-yolo-ap: A lightweight apple detection algorithm based on improved yolov8 with a new efficient pdwconv in orchard," *Computers and Electronics in Agriculture*, vol. 223, p. 109118, 2024.
- [15] B. Paneru, R. Bist, X. Yang, and L. Chai, "Tracking dustbathing behavior of cage-free laying hens with machine vision technologies," *Poultry Science*, vol. 103, no. 12, p. 104289, 2024.
- [16] D. Wu, Y. Ying, M. Zhou, J. Pan, and D. Cui, "Yolo-claw: A fast and accurate method for chicken claw detection," *Engineering Applications of Artificial Intelligence*, vol. 136, p. 108919, 2024.
- [17] D. Wang and D. He, "Fusion of Mask RCNN and attention mechanism for instance segmentation of apples under complex background," *Computers and Electronics in Agriculture*, vol. 196, p. 106864, 2022.
- [18] Y. Li, X. Xu, and C. Yuan, "Enhanced mask r-cnn for chinese food image detection," *Mathematical Problems in Engineering*, vol. 2020, no. 1, p. 6253827, 2020.
- [19] S. P. Mohanty, G. Singhal, E. A. Scuccimarra, D. Kebaili, H. Héritier, V. Boulanger, and M. Salathé, "The food recognition benchmark: Using deep learning to recognize food in images," *Frontiers in Nutrition*, vol. 9, p. 875143, 2022.
- [20] M. Macías-Macías, H. Sánchez-Santamaría, C. J. García Orellana, H. M. González-Velasco, R. Gallardo-Caballero, and A. García-Manso, "Mask r-cnn for quality control of table olives," *Multimedia Tools and Applications*, vol. 82, no. 14, pp. 21 657–21 671, 2023.
- [21] M. Jonker, "Robotic Bin-Picking Pipeline for Chicken Fillets with Deep Learning-Based Instance Segmentation using Synthetic Data," 2023.
- [22] I. Attri, L. K. Awasthi, T. P. Sharma, and P. Rathee, "A review of deep learning techniques used in agriculture," *Ecological Informatics*, p. 102217, 2023.
- [23] D. Lewy and J. Mańdziuk, "An overview of mixing augmentation methods and augmentation strategies," *Artificial Intelligence Review*, vol. 56, no. 3, pp. 2111–2169, 2023.
- [24] B. Hüttnerrauch, "Limitations of data augmentation and outlook," in *Targeting Using Augmented Data in Database Marketing: Decision Factors for Evaluating External Sources*. Springer, 2016, pp. 279–290.
- [25] B. Zoph, E. D. Cubuk, G. Ghiasi, T.-Y. Lin, J. Shlens, and Q. V. Le, "Learning data augmentation strategies for object detection," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16*. Springer, 2020, pp. 566–583.
- [26] S. I. Nikolenko, *Synthetic Data for Deep Learning*. Springer, 2021, vol. 174.
- [27] A. Gaur, C. Liu, X. Lin, N. Karapetyan, and Y. Aloimonos, "Whale detection enhancement through synthetic satellite images," in *OCEANS 2023-MTS/IEEE US Gulf Coast*. IEEE, 2023, pp. 1–7.
- [28] B. Mildenhall, P. Srinivasan, M. Tancik, J. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," in *ACM Transactions on Graphics (TOG)*, vol. 40, no. 4, 2021, pp. 1–12.
- [29] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering," in *ACM Transactions on Graphics (TOG)*, vol. 42, no. 4, 2023, pp. 1–14.
- [30] K. Aghamohammadesmaeilketabforoosh, "Enhancing Strawberry Disease and Quality Detection: Integrating Vision Transformers with Blender-Enhanced Synthetic Data and SwinUNet Segmentation Techniques," 2024.
- [31] A. Kar, A. Prakash, M.-Y. Liu, E. Cameracci, J. Yuan, M. Rusiniak, D. Acuna, A. Torralba, and S. Fidler, "Meta-sim: Learning to generate synthetic datasets," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 4551–4560.
- [32] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *2012 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2012, pp. 5026–5033.
- [33] M. Lin, X. Wang, Y. Wang, S. Wang, F. Dai, P. Ding, C. Wang, Z. Zuo, N. Sang, S. Huang *et al.*, "Exploring the evolution of physics cognition in video generation: A survey," *arXiv preprint arXiv:2503.21765*, 2025.
- [34] C. Gan, J. Schwartz, S. Alter, D. Mrowca, M. Schrimpf, J. Traer, J. De Freitas, J. Kubilius, A. Bhandwaldar, N. Haber *et al.*, "Threedworld: A platform for interactive multi-modal physical simulation," *arXiv preprint arXiv:2007.04954*, 2020.
- [35] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8798–8807.
- [36] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [37] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.
- [38] Y. Karabatis, X. Lin, N. J. Sanket, M. G. Lagoudakis, and Y. Aloimonos, "Detecting olives with synthetic or real data? olive the above," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 4242–4249.
- [39] Y. Tian, G. Yang, Z. Wang, E. Li, and Z. Liang, "Detection of apple lesions in orchards based on deep learning methods of CycleGAN and YOLOV3-dense," *Journal of Sensors*, vol. 2019, no. 1, p. 7630926, 2019.
- [40] M. Fawakherji, V. Suriani, D. Nardi, and D. D. Bloisi, "Shape and style GAN-based multispectral data augmentation for crop/weed segmentation in precision farming," *Crop Protection*, vol. 184, p. 106848, 2024.
- [41] X. Lin, N. J. Sanket, N. Karapetyan, and Y. Aloimonos, "Oysternet: Enhanced oyster detection using simulation," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 5170–5176.
- [42] C. Wang and Z. Xiao, "Lychee surface defect detection based on deep convolutional neural networks with gan-based data augmentation," *Agronomy*, vol. 11, no. 8, p. 1500, 2021.
- [43] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, and R. Rombach, "Sdxl: Improving latent diffusion models for high-resolution image synthesis," *arXiv preprint arXiv:2307.01952*, 2023.
- [44] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.
- [45] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in neural information processing systems*, vol. 34, pp. 8780–8794, 2021.
- [46] X. Lin, V. Mange, A. Suresh, B. Neuberger, A. Palnitkar, B. Campbell, A. Williams, K. Baxevani, J. Mallette, A. Vera *et al.*, "Odyssee: Oyster detection yielded by sensor systems on edge electronics," *arXiv preprint arXiv:2409.07003*, 2024.
- [47] B. Campbell, A. Williams, K. Baxevani, A. Campbell, R. Dhone, R. E. Hudock, X. Lin, V. Mange, B. Neuberger, A. Suresh *et al.*, "Is ai currently capable of identifying wild oysters? a comparison of human annotators against the ai model, odyssee," *Frontiers in Robotics and AI*, vol. 12, p. 1587033, 2025.
- [48] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," *arXiv preprint arXiv:2010.02502*, 2020.
- [49] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "Labelme: a database and web-based tool for image annotation," *International Journal of Computer Vision*, vol. 77, no. 1, pp. 157–173, 2008.
- [50] M. Denninger, M. Sundermeyer, D. Winkelbauer, D. Olefir, T. Hodan, Y. Zidan, M. Elbadrawy, M. Knauer, H. Katam, and A. Lodhi, "Blenderproc: Reducing the reality gap with photorealistic rendering," in *16th Robotics: Science and Systems, RSS 2020, Workshops*, 2020.
- [51] R. Barth, J. IJsselmuizen, J. Hemming, and E. J. Van Henent, "Data synthesis methods for semantic segmentation in agriculture: A Capsicum annum dataset," *Computers and electronics in agriculture*, vol. 144, pp. 284–296, 2018.
- [52] A. Ummadisingu, K. Takahashi, and N. Fukaya, "Cluttered Food Grasping with Adaptive Fingers and Synthetic-Data Trained Object Detection," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 8290–8297.
- [53] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

- [54] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar, “Masked-attention mask transformer for universal image segmentation,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 1290–1299.
- [55] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “Yolov11: Robust, fast and efficient object detection,” *arXiv preprint arXiv:2308.11562*, 2023.
- [56] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” *arXiv preprint arXiv:1711.05101*, 2017.
- [57] S. R. Richter, V. Vineet, S. Roth, and V. Koltun, “Playing for data: Ground truth from computer games,” in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14*. Springer, 2016, pp. 102–118.
- [58] J. Tremblay, A. Prakash, D. Acuna, M. Brophy, V. Jampani, C. Anil, T. To, E. Cameracci, S. Boochoon, and S. Birchfield, “Training deep networks with synthetic data: Bridging the reality gap by domain randomization,” in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 969–977.
- [59] B. Vanherle, B. Zoomers, J. Put, F. Van Reeth, and N. Michiels, “Cut-and-splat: Leveraging gaussian splatting for synthetic data generation,” *arXiv preprint arXiv:2504.08473*, 2025.
- [60] E. Friedman, A. Lehr, A. Gruzdev, V. Loginov, M. Kogan, M. Rubin, and O. Zvitia, “Knowing the distance: Understanding the gap between synthetic and real data for face parsing,” *arXiv preprint arXiv:2303.15219*, 2023.
- [61] J. Shipard, A. Willem, K. N. Thanh, W. Xiang, and C. Fookes, “Boosting zero-shot classification with synthetic data diversity via stable diffusion,” *arXiv preprint arXiv:2302.03298*, vol. 3, no. 5, 2023.
- [62] R. Khanam and M. Hussain, “Yolov11: An overview of the key architectural enhancements,” *arXiv preprint arXiv:2410.17725*, 2024.
- [63] R. Sapkota and M. Karkee, “Comparing yolov11 and yolov8 for instance segmentation of occluded and non-occluded immature green fruits in complex orchard environment,” *arXiv preprint arXiv:2410.19869*, 2024.
- [64] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [65] Y. Liang, S. Cao, J. Zheng, X. Zhang, J. Huang, and H. Fu, “Low saturation confidence distribution-based test-time adaptation for cross-domain remote sensing image classification,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 139, p. 104463, 2025.
- [66] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2020.
- [67] T. R. Dieter, A. Weinmann, S. Jäger, and E. Brucherseifer, “Quantifying the simulation–reality gap for deep learning-based drone detection,” *Electronics*, vol. 12, no. 10, p. 2197, 2023.
- [68] C.-J. Chang, D. Li, S. Moon, and M. Kapadia, “On the equivalency, substitutability, and flexibility of synthetic data,” *arXiv preprint arXiv:2403.16244*, 2024.
- [69] X. Li, X. Tan, Z. Chen, Z. Zhang, R. Zhang, R. Guo, G. Jiang, Y. Chen, Y. Qu, L. Ma *et al.*, “One-for-more: Continual diffusion model for anomaly detection,” *arXiv preprint arXiv:2502.19848*, 2025.
- [70] I. Shumailov, Z. Shumailov, Y. Zhao, N. Papernot, R. Anderson, and Y. Gal, “Ai models collapse when trained on recursively generated data,” *Nature*, vol. 631, no. 8022, pp. 755–759, 2024.