Pig Detection Using Computer Vision Models Trained on Synthetic Datasets

Alvaro V. Martins[1], **Thiago M. Paixão**[1], **Francisco A. Boldt**[1], **Isabella C. Condotta**[2]

[1] Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo (IFES)
Serra – ES – Brazil

alvarovinicius7@gmail.com, {franciscoa, thiago.paixao}@ifes.edu.br

[2]University of Illinois Urbana-Champaign
Urbana-Champaign – Illinois – United States of America

icfsc@illinois.edu

Abstract. *Pork is a major commodity in the global food industry, yet monitoring pigs in their living spaces remains a challenge due to the scarcity of annotated datasets for AI models. In this study, an existing method for generating training datasets by utilizing a limited number of annotated images is adapted for pig detection, with the goal of enabling accurate pig counting in future applications. Using the COCO dataset and the Multi-Camera Pig Dataset, we applied Digital Image Processing and Meta's Segment Anything Model 2 (SAM2) to augment data. A YOLO model trained on the generated datasets demonstrated that manual annotation can be reduced by 40% with minimal performance loss, significantly decreasing dataset production effort.*

1. Introduction

Pork is one of the most important types of meat in the world. Approximately 34% of the meat consumed globally is pork, with its global consumption having increased by 77% from 63.5 million tons in 1990 to 113 million tons in 2022 [Kim et al. 2024]. From 1998 to 2018, developing countries accounted for around 85% of the rise in global swine meat consumption [Whitnall and Pitts 2019]. In Brazil, pork production increased by 326% over the past 30 years, with the swine meat industry in the country expecting growth of 16% over the next decade [Roppa et al. 2024].

Consequently, the increase in supply and demand for pork brings to light some challenges encountered in pig production facilities, the pig farms. Despite being bred only to be harvested for meat, pigs are living beings as well, and thus are deserving of some level of respect and dignity. Many countries have modeled their animal welfare legislations based on the Five Freedoms concept, including the members of the European Union [EUR-Lex 2024], Japan [MAFF 2023] and Brazil [SEBRAE 2016]. The welfare requirements are varied, such as the methodologies that can be implemented to fulfill them. Despite this, many methods are challenging to maintain, in special animal monitoring and supervision. Visual monitoring of the animals is important for assuring good conditions of the pen, as well as early detection of abnormalities in the enclosures or pigs themselves. This task becomes harder if the pen is large or if there are many pigs to monitor at the same time, making the process very inefficient if attempted by common farm workers.

Many studies in computer vision applied to pig monitoring utilize primarily locally obtained pig footage, which is then manually annotated to create training datasets [Van der Zande et al. 2021, Guo et al. 2023, Jaoukaew et al. 2024]. All these works used the same methodology of dataset collection and training. This methodology is effective but can become extremely labor-intensive depending on the implementation domain, mainly due to manual effort to annotate data, which takes the most amount of time in the whole process. This does not even take into account the time required to obtain the initial footage of the pigs in their pens.

The most notable work relevant to our context is the development of a novel method for training an object detection model in the traffic signs domain using artificially generated images with traffic sign templates [Tabelini et al. 2022]. This method demonstrated that a model can be effectively trained without manually annotated data and potentially improve performance when artificially generated images are used alongside manually annotated ones. However, this method was only tested on rigid objects (traffic signs) and not on more deformable, plastic objects. The form of an object significantly influences the model's detection capability, which in turn could affect the method's performance.

The primary objective of this work is to adapt the existing methodology for generating synthetic, pre-processed, and pre-annotated image datasets to enhance the efficiency and effectiveness of training AI computer vision models in diverse scenarios, including those with deformable objects. The proposed method extracts sections of the object of interest from a limited set of manually annotated images, segments them using Meta's Segment Anything Model 2 (SAM2), and integrates them into randomly selected natural images as foreground elements. This adaptation proved competitive in relation to model performance with the original method, and our results indicate a 40% reduction in the manual annotation effort required for model training.

2. Related Works

Computer vision is defined as the ability of a computer to process, analyze, and understand information from images or videos [Szeliski 2022]. Prior to the advent of Machine Learning (ML) algorithms, computer vision tasks were executed using traditional image processing methods, such as edge detection and mathematical transforms, where the rules for handling visual information had to be predefined [Forsyth and Ponce 2002]. Deep learning has enabled these rules and patterns to be learned dynamically, rather than manually adjusted, making computer vision tasks more flexible, generalized, and robust. Convolutional neural networks (CNNs) have made complex image pattern recognition not only feasible but also accessible for commercial applications [Khan et al. 2021].

With the growing demand for automated livestock monitoring, CV models are increasingly being developed to track and analyze animals in their living spaces. Notable works in this field include vision-based models for monitoring chicken behavior and health in poultry farms [Bhuiyan and Wree 2023], cow detection in a free stall barn [Tassinari et al. 2021], and multi-camera tracking systems for pigs [Shirke et al. 2021]. All of these works used the You Only Look Once (YOLO) algorithm trained using transfer learning methodology, which involves leveraging pre-trained neural networks on large-scale, generic domain datasets and fine-tuning them on domain-specific data

[Tian et al. 2025a]. This approach allows models to retain valuable low-level features, such as edges and textures, while adapting higher-level features to the specific task, reducing training time and improving performance, especially when limited annotated data is available [Zhuang et al. 2020].

Despite significant advancements in the training efficiency of CNNs in recent years, their performance remains highly dependent on large-scale, high-quality annotated datasets [Sager et al. 2021]. The manual annotation process is not only labor-intensive and time-consuming but also prone to inconsistencies and errors, which can negatively impact model accuracy and generalization [Zhao et al. 2024]. As a result, reducing the reliance on extensive manual annotation while maintaining or improving model performance has become a critical challenge in the field of computer vision.

In the domain of synthetic datasets, several studies have explored generative approaches to address challenges in image dataset collection and annotation for object detection. Some methods enhance training data quality through 3D renderings, others rely solely on 2D imaging, and some combine both techniques [Paulin and Ivasic-Kos 2023, Man and Chahl 2022]. The primary objective of these approaches is to introduce greater variability in training data, normalize datasets, and incorporate edge cases that may be underrepresented in real-world datasets.

A particularly effective technique for generating training datasets involves using object templates and merging them with real or synthetic backgrounds to create pre-annotated, high-quality artificial datasets. This method has been successfully applied in traffic sign detection, especially for autonomous vehicle training [Tabelini et al. 2022].

However, existing applications of these techniques have largely focused on rigid objects, such as traffic signs [de Mello et al. 2021] or industrial components [Abou Akar et al. 2022]. There has been limited exploration of their effectiveness in domains involving deformable objects, such as pig detection, where shape variations and slight deformations occur. The success of these methods in previous studies suggests that similar approaches could be both feasible and effective for animal monitoring applications.

3. Materials & Methods

3.1. Proposed Method

The proposed method, illustrated in Figure 1, follows a multi-step pipeline. First, pig footage is captured and used to extract individual pig templates based on existing annotations. Second, two background groups were created: one consisting solely of generated noise images and the other containing randomly selected images from the COCO dataset. Third, templates undergo random transformations, including rotation, flipping, and, if necessary, rescaling to match the target training resolution. Fourth, transformed templates are randomly placed on the backgrounds with varied positions and orientations, and annotations are automatically generated using the known dimensions of the templates, backgrounds, and placement coordinates. Finally, additional image processing techniques were applied. The resulting synthetic images form the training dataset for the object detection model. Further details of the method are provided in the subsequent sections.

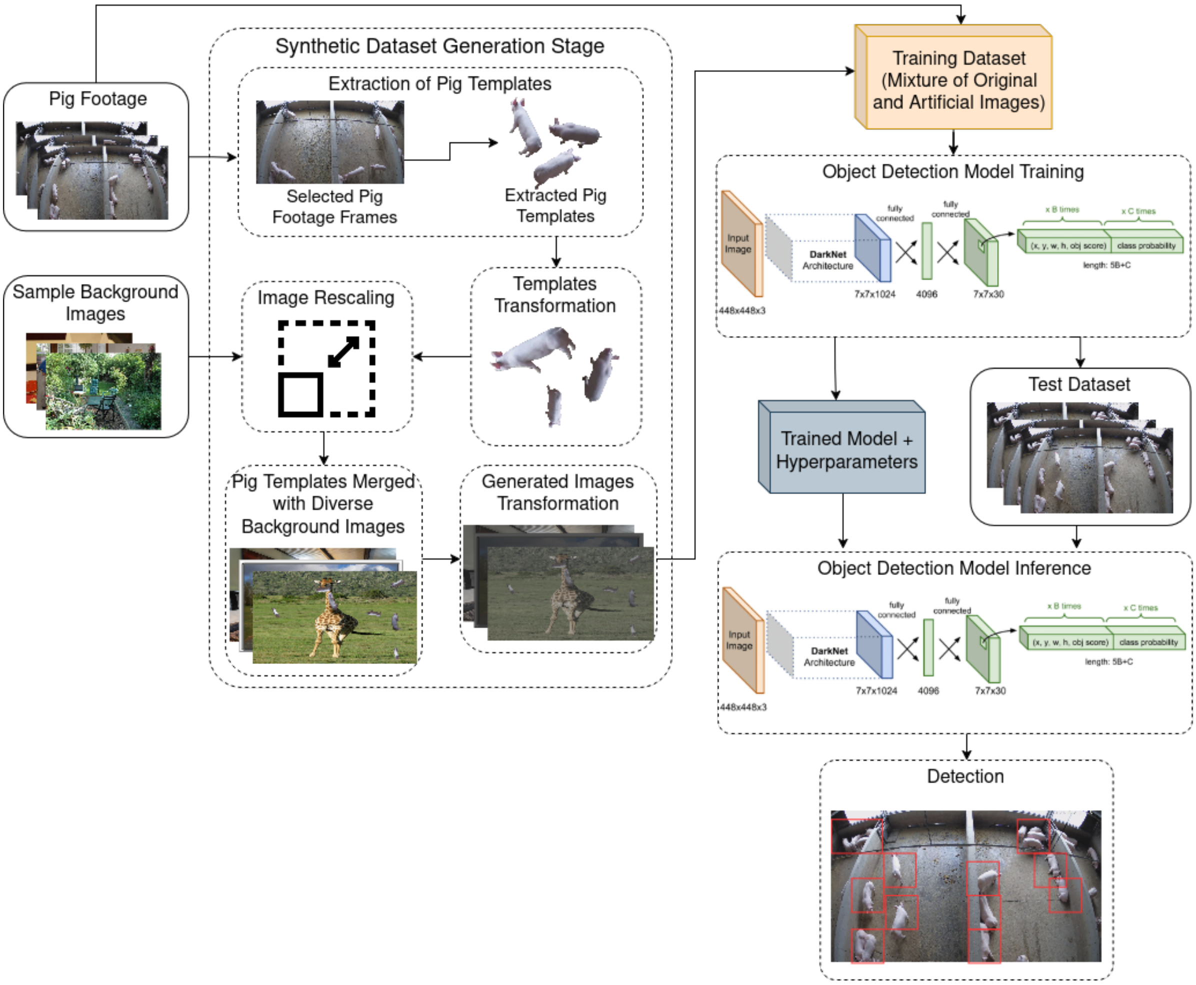

Figure 1. Overview of the Proposed Method.

3.2. Pig Footage

The pig footage used in this study comes from the Multi-Camera Pig Dataset, provided by the AIFARMS project at the University of Illinois Urbana-Champaign. The dataset comprises three distinct 8-minute, 4K-resolution videos capturing pigs in their living spaces. Each video was extracted into a set of 7,200 images, which were all manually annotated. The frames of the first video, and its extracted templates, were used exclusively for training. The second and third were similarly used for validation, and testing, respectively.

3.3. Background Images

Two types of background images were used: randomly generated noise images and natural images. The noise backgrounds were created at a fixed resolution of 640×360—commonly used in computer vision training—with randomly assigned pixel values, illustrated in Figure 2. For natural backgrounds, we utilized 5,000 images from the Common Objects in Context (COCO) dataset [Lin et al. 2014], which features diverse scenes in varying standard-definition (SD) resolutions, as shown in fig 3.

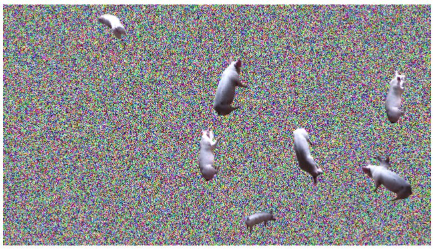

Figure 2. Image generated with random noise background

Figure 3. Image generated with a random natural image as background.

3.4. Template Extraction

Pig templates were extracted using traditional digital image processing with the Pillow Python library [Clark 2015] and Meta's Segment Anything Model 2 (SAM2) [Ravi et al. 2024]. Initially, pig-containing regions were cropped from a limited number of frames based on annotation data, ensuring that overlapping or intersecting pigs were excluded to maintain training quality, as shown in Figure 4. The images were then passed to SAM2, which performs zero-shot segmentation based on prompts such as bounding boxes, points, or center coordinates. This process enabled precise segmentation of pigs from their surroundings, yielding high-quality templates for dataset generation, illustrated in Figure 5. This approach differs from Tabelini's work, which used traffic sign graphics instead of zero-shot segmentation to extract templates from more varied context images.

Figure 4. Image section containing pig cropped from the frame.

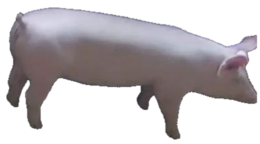

Figure 5. Pig template obtained by segmenting the pig image crop using SAM2.

3.5. Training Generation

To generate pre-annotated synthetic images, pig templates were randomly positioned over background images, with bounding box annotations automatically calculated based on the known dimensions and placement of each template.

A random number of pig templates (between 1 and 10) was selected and merged onto a background, randomly sampled from the COCO dataset. For comparison purposes in the first experiment, we also created background images composed of random pixel noise (referred to as *noise images*), allowing us to assess the impact of background realism on model training.

To ensure visual coherence and avoid unrealistic edges—a known issue when overlaying segmented objects onto new backgrounds—we implemented a multi-step blending strategy. Initially, Gaussian blur was applied to soften template edges. However, this alone proved insufficient. We therefore introduced a transparency gradient by stacking four eroded versions of each template, each with decreasing opacity in steps of 0.25.

The most eroded layer had full opacity, while the least eroded had higher transparency. This compositing method produced smooth, feathered borders that visually blended the pig templates into their backgrounds, minimizing abrupt contours and enhancing realism.

Following blending, all images were rescaled to 640×360 pixels—a common resolution in object detection benchmarks. Background images were resized directly, while template images (originally from 4K video frames) were scaled with aspect ratio preservation, using a scaling factor defined as $SF = \frac{OD}{TD}$ where OD and TD represent the original and target dimensions, respectively.

To further increase variability, templates were randomly rotated (±90°), flipped (25% chance horizontally or vertically), and slightly rescaled using a secondary factor between -0.05 and 0.05. After placing each template at a random location on the background, annotations were computed for YOLO based on the template's position and size. Finally, full-image augmentations—including adjustments to gamma, sharpness, contrast, and color—were applied within a ±0.75 range to simulate photometric variability found in real-world scenarios.

Figures 6 and 7 illustrate examples of synthetic images before and after the blending and augmentation pipeline, highlighting the visual similarity to naturally acquired samples.

Figure 6. Example of generated image before transformation.

Figure 7. Example of generated image after transformation.

3.6. Model Training

The model we trained to evaluate the efficacy of the method was the YOLO model [Redmon et al. 2016], version 12 [Tian et al. 2025b]. As of this writing, YOLOv12 is the most recent iteration of the YOLO family and represents the state of the art among non-transformer-based computer vision models. YOLO is primarily used for object detection tasks, but it also supports functionalities such as object tracking, pose estimation, and object classification. Unlike other detectors like R-CNN [Girshick 2015], which separate Regions of Interest (RoIs) before classification and detection, YOLO employs a single-stage approach, processing the entire image in a single pass. This enables near real-time object detection. We chose YOLO due to its ease of use, simplified transfer learning process, our prior experience with the tool, and its widespread adoption in related works.

3.7. Evaluation Metrics

To evaluate the models and, consequently, the proposed method, we adopted five complementary metrics: Precision, Recall, F1-Score, mean Average Precision at 50% IoU (mAP50), and mean Average Precision across 50–95% IoU (mAP@50-95). Precision measures the proportion of correctly identified positive detections relative to all positive predictions, defined as $P = \frac{TP}{TP+FP}$, where *TP* (True Positives) are the correctly identified positive instances and *FP* (False Positives) are the incorrect positive predictions. Recall quantifies the model's sensitivity, representing the proportion of correctly identified positive instances among all actual positives, given by $R = \frac{TP}{TP+FN}$, where *FN* (False Negatives) are the actual positive instances that were incorrectly predicted as negative. A well-optimized model balances Precision and Recall, which is captured by the F1-Score, the harmonic mean of the two, computed as $F1 = \frac{2\times(P\times R)}{(P+R)}$.

To assess object detection performance, we employed mAP50 and mAP@50-95. mAP50 considers a detection correct if the Intersection over Union (IoU) between the predicted and ground truth bounding boxes is at least 50%, providing a more lenient evaluation. In contrast, mAP@50-95 calculates the mean average precision across multiple IoU thresholds, ranging from 50% to 95% in increments of 5%. This approach offers a more rigorous and comprehensive assessment of model performance by evaluating the precision at various levels of detection difficulty, thus providing a more nuanced understanding of the model's accuracy and robustness.

3.8. Experiments

Random vs. COCO Background Experiment The first experiment aims to compare the use of random images versus natural images from the COCO dataset as background settings for the proposed method. We also evaluated the default training setup using the original manually annotated data to establish an empirical upper bound for model performance. To this end, we extracted a total of 494 pig templates from 100 randomly selected video frames—50 frames from the training video recording and 50 from the validation recording. Among these templates, 323 were sourced from frames designated for training, while 171 came from validation frames. Using these templates, we generated a dataset consisting of 14,200 synthetic images, with an 80/20 split for training and validation. We then trained the YOLO model on these datasets, performed inference on the test set, and finally compared their performance metrics. The model hyperparameters, detailed in Table 1, largely followed YOLO's standard configuration, except for image and batch sizes, which were reduced due to hardware constraints.

Model	Epochs	Image Size	Batch Size	Workers	Learning Rate
YOLO12m	100	416x416	4	8	0.01

Table 1. Hyperparameters used for the experiment.

Artificial Training Ratio Experiment This experiment aims to determine the optimal balance between traditional and artificial training data to maximize model performance while evaluating whether a higher proportion of artificial data degrades results. To this end, we created four additional datasets, each containing a fixed set of 7,200 traditional images—a quantity previously shown to yield reliable training outcomes—combined with increasing amounts of artificial images (1,800, 4,800, 7,200, 10,800, and 28,800). This resulted in datasets with varying traditional-to-artificial image ratios: 0% (only traditional images), 20%, 40%, 50%, 60%, 80% and 100% (only artificial images). Model hyperparameters remained consistent with those used in the first experiment.

4. Results & Discussion

This section presents the findings and insights derived from two key experiments conducted to evaluate the effectiveness of the proposed method for generating artificial training datasets in the context of pig detection. The first experiment explores the impact of different background types on model performance, while the second investigates the optimal ratio of artificially generated images to manually annotated images in the training dataset. Together, these experiments provide valuable insights into the feasibility, limitations, and potential of the proposed approach, offering a foundation for future improvements and applications in object detection tasks.

4.1. Random vs. COCO Background Experiment

This experiment aimed to evaluate the impact of different artificially generated backgrounds on training effectiveness, as well as to verify the feasibility of the proposed method by comparing it with training on the original data.

The results, presented in Figure 8, show that datasets using random noisy backgrounds were ineffective for training, achieving an F1-Score of 58.5 and a mAP@50-95 score of only 24. In contrast, datasets that employed natural images from the COCO dataset as backgrounds performed significantly better, with an F1-Score of 92.5 and a mAP@50-95 of 61.6—values that are highly competitive with those obtained by the model trained with regular, manually annotated images.

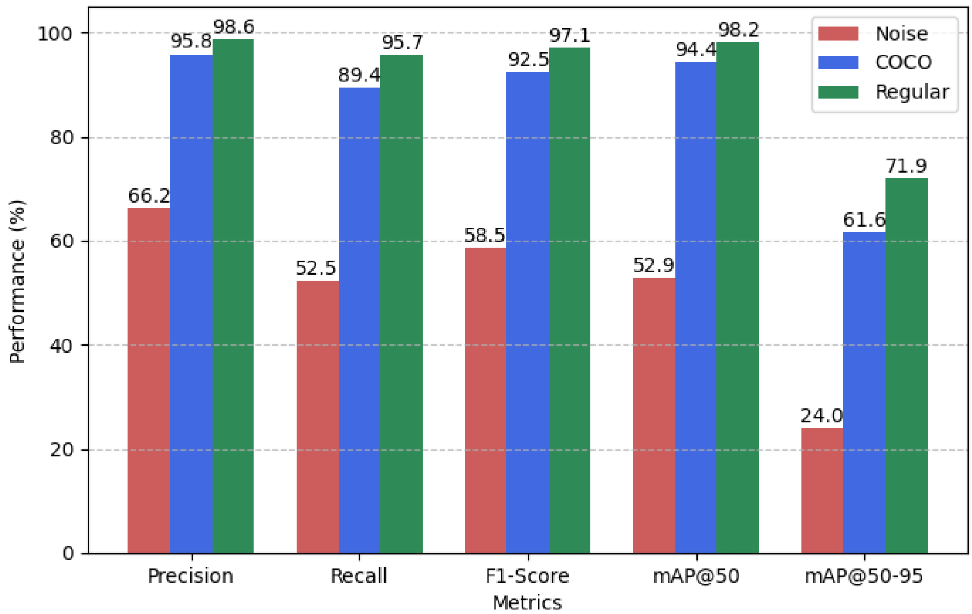

Figure 8. Experiment 1: Comparing random images and natural images from the COCO dataset as background settings. The original dataset, with manual annotations, was used to establish an empirical upper bound for performance.

The obtained results can be attributed to the nature of the background and the limited number of object templates. CNNs learn by identifying spatial patterns and object-specific features.

In conventional training with original data, background features can sometimes aid object detection by providing additional contextual cues. However, in artificially generated datasets with a limited set of templates, the model's ability to generalize relies more on background diversity and transformation techniques (e.g., sharpness, color, and brightness adjustments). Insufficient background variation may cause the model to overfit to the small template set, reducing its robustness to new scenarios. Conversely, artificial images with diverse random backgrounds, combined with transformation, improved model generalization. Notably, the results suggest that, despite being outside the original application domain (pig monitoring), natural images from the COCO dataset introduce greater variability, reducing the model's reliance on templates and encouraging more generalizable feature learning. This further supports their suitability for integration with traditionally annotated datasets in subsequent experiments.

Another key observation is that while the differences in F1-Score and mAP@50 between the COCO-based and original datasets remained within 5 percentage points—with the COCO-based dataset achieving scores above 92%— the performance gap widened to over 10 percentage points for mAP@50-95. This suggests that the proposed dataset generation method is more sensitive to stricter Intersection over Union (IoU) thresholds, leading to lower-confidence predictions under higher localization accuracy requirements. As a result, while the method exhibits strong performance under more lenient evaluation criteria, its effectiveness declines when precise object localization is essential.

4.2. Artificial Training Ratio Experiment

The second experiment aimed to identify the optimal proportion of artificially generated images, using COCO dataset backgrounds, in relation to manually annotated images within the training dataset. A higher proportion of pre-annotated synthetic images, combined with a reduced number of manually labeled samples, could significantly alleviate the annotation burden. Additionally, we sought to assess whether an excessive reliance on artificial images would negatively impact detection performance.

As shown in Figure 9, and more detailed in Table 2, the results demonstrate that integrating a balanced mix of original and artificially generated images can enhance model performance. Specifically, the findings suggest that an optimal training dataset composition consists of approximately 40% to 50% artificially generated images. Beyond this threshold, the benefits plateau or decline, indicating the importance of maintaining a sufficient proportion of manually annotated images to ensure robust model generalization.

Artificial Ratio	Precision	Recall	F1-Score	mAP@50	mAP@50-95
20%	+0,26%	+0,99%	+0,64%	+0,16%	+2,42%
40%	+0,53%	+1,45%	+1,00%	+0,53%	+2,30%
50%	+0,42%	+1,67%	+1,06%	+0,54%	+2,55%

Table 2. Performance improvements across different artificial data ratios in the training dataset, measured as relative gains (%) over the baseline (0% artificial data). The metrics include Precision, Recall, F1-Score, mAP@50, and mAP@50-95.

When analyzing the stricter metric, mAP@50-95, the same trend is observed, with performance peaking near the 50% mark. This can be attributed to the well-preserved realistic context provided by original images, complemented by the increased variability

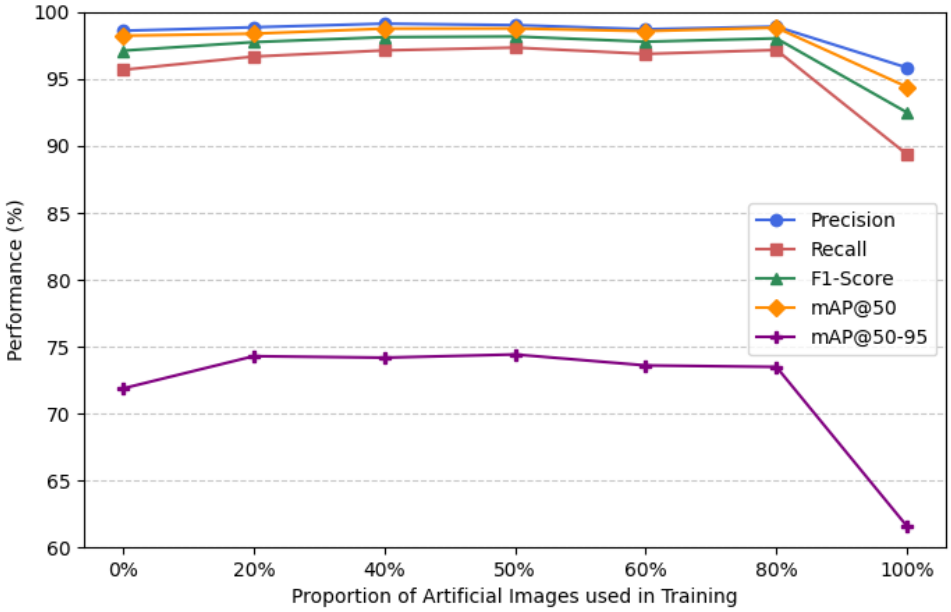

Figure 9. Experiment 2: Evaluating the effect of mixing artificial and original training images on model performance.

introduced by artificial images. These findings suggest that a sufficient number of traditionally annotated images must be retained in the dataset for the benefits of artificial transformation to take effect. Conversely, using the proposed method with a low number of original images could lead to a slight performance degradation.

5. Conclusion

This study investigated the use of synthetic image generation to reduce the annotation effort required for training object detection models, with pig monitoring selected as the application domain due to the deformable nature of the target objects. The proposed methodology involved extracting pig templates from a limited set of manually annotated images, segmenting them using SAM2, and compositing them onto random natural images from the COCO dataset. Annotations were automatically generated based on image dimensions and template placement, significantly streamlining dataset creation.

Experimental results demonstrated that synthetic datasets using COCO backgrounds achieved strong performance (F1-Score: 92.5, mAP@50-95: 61.6), comparable to traditional datasets. In contrast, noise-based backgrounds were ineffective. Although the COCO-based dataset showed some sensitivity to stricter IoU thresholds, the overall results confirmed the viability of the proposed method. Furthermore, training performance was optimized when artificial images comprised 40% to 50% of the dataset, while excessive reliance on original data led to slight performance degradation due to limited contextual diversity.

In conclusion, the proposed method proved effective in generating synthetic datasets that support high-performing object detection models with reduced manual annotation effort. Future work will extend this approach to other livestock species and explore the use of bounding boxes and masks as input prompts for SAM models, potentially further minimizing human intervention in dataset creation.

Acknowledgements. The authors would like to thank FAPES/UnAC (No. FAPES 1228/2022 P 2022-CD0RQ, No. SIAFEM 2022-CD0RQ) for the financial support provided through the UniversidaES System.

References

Abou Akar, C., Tekli, J., Jess, D., Khoury, M., Kamradt, M., and Guthe, M. (2022). Synthetic object recognition dataset for industries. In 2022 35th SIBGRAPI conference on graphics, patterns and images (SIBGRAPI), volume 1, pages 150–155. IEEE.

Bhuiyan, M. R. and Wree, P. (2023). Animal behavior for chicken identification and monitoring the health condition using computer vision: A systematic review. IEEE access, 11:126601–126610.

Clark, A. (2015). Pillow (pil fork) documentation.

de Mello, J. P. V., Tabelini, L., Berriel, R. F., Paixao, T. M., De Souza, A. F., Badue, C., Sebe, N., and Oliveira-Santos, T. (2021). Deep traffic light detection by overlaying synthetic context on arbitrary natural images. Computers & Graphics, 94:76–86.

EUR-Lex (2024). Animal welfare.

Forsyth, D. A. and Ponce, J. (2002). Computer vision: a modern approach. prentice hall professional technical reference.

Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision, pages 1440–1448.

Guo, Q., Sun, Y., Orsini, C., Bolhuis, J. E., de Vlieg, J., Bijma, P., and de With, P. H. (2023). Enhanced camera-based individual pig detection and tracking for smart pig farms. Computers and Electronics in Agriculture, 211:108009.

Jaoukaew, A., Suwansantisuk, W., and Kumhom, P. (2024). Robust individual pig tracking. International Journal of Electrical and Computer Engineering (IJECE), 14(1):279–293.

Khan, A. A., Laghari, A. A., and Awan, S. A. (2021). Machine learning in computer vision: A review. EAI Endorsed Transactions on Scalable Information Systems, 8(32).

Kim, S. W., Gormley, A., Jang, K. B., and Duarte, M. E. (2024). Current status of global pig production: an overview and research trends. Animal Bioscience, 37(4):719.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In Computer vision–ECCV 2014: 13th European conference, zurich, Switzerland, September 6-12, 2014, proceedings, part v 13, pages 740–755. Springer.

MAFF (2023). Farm animal management in consideration of animal welfare.

Man, K. and Chahl, J. (2022). A review of synthetic image data and its use in computer vision. Journal of Imaging, 8(11):310.

Paulin, G. and Ivasic-Kos, M. (2023). Review and analysis of synthetic dataset generation methods and techniques for application in computer vision. Artificial intelligence review, 56(9):9221–9265.

Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K. V., Carion, N., Wu, C.-Y., Girshick, R., Dollár, P., and Feichtenhofer, C. (2024). Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 779–788.

Roppa, L., Duarte, M. E., and Kim, S. W. (2024). Pig production in latin america. Animal Bioscience, 37(4):786.

Sager, C., Janiesch, C., and Zschech, P. (2021). A survey of image labelling for computer vision applications. Journal of Business Analytics, 4(2):91–110.

SEBRAE (2016). Bem-estar animal na produção de suínos.

Shirke, A., Saifuddin, A., Luthra, A., Li, J., Williams, T., Hu, X., Kotnana, A., Kocabalkanli, O., Ahuja, N., Green-Miller, A., et al. (2021). Tracking grow-finish pigs across large pens using multiple cameras. arXiv preprint arXiv:2111.10971.

Szeliski, R. (2022). Computer vision: algorithms and applications. Springer Nature.

Tabelini, L., Berriel, R., Paixão, T. M., De Souza, A. F., Badue, C., Sebe, N., and Oliveira-Santos, T. (2022). Deep traffic sign detection and recognition without target domain real images. Machine Vision and Applications, 33(3):50.

Tassinari, P., Bovo, M., Benni, S., Franzoni, S., Poggi, M., Mammi, L. M. E., Mattoccia, S., Di Stefano, L., Bonora, F., Barbaresi, A., et al. (2021). A computer vision approach based on deep learning for the detection of dairy cows in free stall barn. Computers and Electronics in Agriculture, 182:106030.

Tian, Y., Ye, Q., and Doermann, D. (2025a). Yolov12: Attention-centric real-time object detectors.

Tian, Y., Ye, Q., and Doermann, D. (2025b). Yolov12: Attention-centric real-time object detectors. arXiv preprint arXiv:2502.12524.

Van der Zande, L. E., Guzhva, O., and Rodenburg, T. B. (2021). Individual detection and tracking of group housed pigs in their home pen using computer vision. Frontiers in animal science, 2:669312.

Whitnall, T. and Pitts, N. (2019). Global trends in meat consumption. Agricultural Commodities, 9(1):96–99.

Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., and Parmar, M. (2024). A review of convolutional neural networks in computer vision. Artificial Intelligence Review, 57(4):99.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., and He, Q. (2020). A comprehensive survey on transfer learning. Proceedings of the IEEE, 109(1):43–76.