JAST (Journal of Animal Science and Technology) TITLE PAGE

Upload this completed form to website with submission

ARTICLE INFORMATION	**Fill in information in each box below**
Article Type	Research Article
Article Title (within 20 words without abbreviations)	Comparison of estimating vegetation index for outdoor free-range pig production using convolutional neural networks
Running Title (within 10 words)	Comparison of estimating vegetation index using different CNNs
Author	Sang-Hyon OH [first_author] 1, Hee-Mun Park [first_author] 2, Jin-Hyun Park2
Affiliation	1 Division of Animal Science, College of Agriculture and Life Science, Gyeongsang National University, Jinju 52725, Korea, Republic of 2 School of Mechatronics Engineering, Engineering College of Convergence Technology, Gyeongsang National University, Jinju 52725, Korea, Republic of
ORCID (for more information, please visit https://orcid.org)	Sang-Hyon OH (https://orcid.org/0000-0002-9696-9638) Hee-Mun Park (https://orcid.org/0000-0001-5182-1739) Jin-Hyun Park (https://orcid.org/0000-0002-7966-0014)
Competing interests	No potential conflict of interest relevant to this article was reported.
Funding sources State funding sources (grants, funding sources, equipment, and supplies). Include name and number of grant if available.	
Acknowledgements	
Availability of data and material	
Authors' contributions Please specify the authors' role using this form.	Conceptualization: OH SH, Park JH Data curation: OH SH, Park HM, Park JH Formal analysis: Park HM, Park JH Methodology: OH SH, Park JH Software: Park HM, Park JH Validation: OH SH, Park HM, Park JH Investigation: OH SH, Park HM, Park JH Writing - original draft: OH SH, Park HM Writing - review & editing: OH SH, Park HM, Park JH
Ethics approval and consent to participate	The present experiment was reviewed and approved by the Institutional Animal Care and Use Committee of North Carolina A&T University (IACUC: 12-003.0).

CORRESPONDING AUTHOR CONTACT INFORMATION

For the corresponding author (responsible for correspondence, proofreading, and reprints)	**Fill in information in each box below**
First name, middle initial, last name	Jin-Hyun Park
Email address – this is where your proofs will be sent	uabut@gnu.ac.kr

Secondary Email address	
Address	
Cell phone number	
Office phone number	
Fax number	

6

7

Comparison of estimating vegetation index for outdoor free-range pig production using convolutional neural networks

Sang-Hyon OH[1†], Hee-Mun Park[2†] and Jin-Hyun Park[2*]

[1]Division of Animal Science, College of Agriculture and Life Science, Gyeongsang National University, Jinju 52725, South Korea

[2]School of Mechatronics Engineering, Engineering College of Convergence Technology, Gyeongsang National University, Jinju 52725, South Korea

[†] Both authors contributed equally to this manuscript.

***Corresponding author**: uabut@gnu.ac.kr

Running Title: Comparison of estimating vegetation index using different CNNs

ORCID

Sang-Hyon Oh	https://orcid.org/0000-0002-9696-9638
Hee-Mun Park	https://orcid.org/0000-0001-5182-1739
Jin-Hyun Park	https://orcid.org/0000-0002-7966-0014

Title of the manuscript: Comparison of estimating vegetation index for outdoor free-range pig production using convolutional neural networks

ABSTRACT

This study aims to predict the change in corn share according to the grazing of 20 gestational sows in a mature corn field by taking images with a camera-equipped UAV. Deep learning based on convolutional neural networks (CNNs) has been verified for its performance in various areas. It has also demonstrated high recognition accuracy and detection time in agricultural applications such as pest and disease diagnosis and prediction. A large amount of data is required to train CNNs effectively. Still, since UAVs capture only a limited number of images, we propose a data augmentation method that can effectively increase data. And most occupancy prediction predicts occupancy by designing a CNN-based object detector for an image and counting the number of recognized objects or calculating the number of pixels occupied by an object. These methods require complex occupancy rate calculations; the accuracy depends on whether the object features of interest are visible in the image. However, in this study, CNN is not approached as a corn object detection and classification problem but as a function approximation and regression problem so that the occupancy rate of corn objects in an image can be represented as the CNN output. The proposed method effectively estimates occupancy for a limited number of cornfield photos, shows excellent prediction accuracy, and confirms the potential and scalability of deep learning.

Keywords: outdoor, pig, vegetation index, image analysis, convolutional neural network

INTRODUCTION

Pasture-based pig production is a common practice adopted in various countries, providing an opportunity for small-scale farmers to generate additional value within the context of corporate-driven swine industries. Iberico pork in Spain is a prime example, which has successfully demonstrated the potential benefits of Pasture-based pig production [1]. However, the expansion of this practice may lead to land degradation issues, warranting careful assessment and mitigation strategies. We presented two previous publications that have addressed the need for Pasture-based pig production and its associated land degradation problem [2, 3]. In this study, we focus on the crucial method of the land degradation assessment process: defining a suitable approach for measuring the extent of degradation in affected areas.

Digital image recognition technology is an image processing technology from computer vision. It has been applied in various areas of modern life, including security, the military, transportation, agriculture, medicine, and daily life [4,5,6]. However, it was difficult to recognize object features affected by camera settings, brightness around the object, and shadows. Utilizing multilayer artificial neural network algorithms in image recognition can allow more accurate object recognition even when there are changes to object features. However, this approach could have been impractical due to its high computation requirements. With the recent developments in semiconductor technologies, devices capable of parallel computing have been developed by integrating thousands of processing units into a single device, making it easier to implement algorithms with large amounts of computation. As a result, image recognition based on deep convolutional neural networks has also become practical technology [5-7].

Most yield predictions involve designing a Convolutional Neural Network (CNN)-based object detector for an image and predicting the yield or the occupancy by counting the number of detected objects or calculating the number of pixels occupied by the objects. However, these methods require multiple computational steps in addition to the detector, and their accuracy depends heavily on whether the object features of interest are clearly visible in the image. For example, it is difficult to classify crops within images captured from high altitudes or wide areas. Moreover, deep learning-based image recognition in the agricultural domain requires a large amount of image data collected by experts in the field, and these images differ depending on the cultivation method, environment, and location [8-11]. Basic data augmentation involves applying various image processing techniques to preserve the characteristics of the original image while maintaining the diverse characteristics of the objects. These techniques vary from physically transforming images by randomly flipping, rotating, and cropping them to techniques that change the color or brightness of the images, such as inverting and channel mixing [8-13]. Two stages of processing are required to predict the yield or occupancy of a specific object. The first step is to classify specific objects in an image using a CNN, and the second step is to represent the occupancy rate of the classified objects based on the number of objects and the area they occupy. Calculating the degree of occupancy is a very cumbersome process. But another advantage of a CNN is that it also can be applied to function approximation and regression problems in addition to classifying objects [12,13]. Therefore, if a CNN is applied as a regression network, the occupancy rate of specific objects in an image can be represented by the network output without going through multiple calculation steps.

The objective of this study was to predict the occupancy rate of corn that has altered due to grazing by twenty gestating sows in a mature cornfield by capturing images with a camera-equipped UAV. A large amount of data is required to effectively train CNN-based deep learning, However, only a limited number of images were captured by the UAV so a data augmentation method that can effectively increase the data was proposed. Various CNNs were used as regression networks for comparison, and the applicability and scalability of deep learning were verified.

MATERIALS AND METHODS

Animal care

The present experiment was reviewed and approved by the Institutional Animal Care and Use Committee of North Carolina A&T University (IACUC: 12-003.0).

Study design and site

The images used for the analysis were taken at a swine unit located within the University Farm of North Carolina A&T State University (Greensboro, NC, USA; 36°4′16.63″N, 79°43′33.02″E). A 50×100 m^2 grazing area was established for twenty pregnant sows that were allowed to graze pasture two weeks prior to their expected delivery date. The grazing area was planted with corn crops. The climate in this location is classified as a humid subtropical climate (Köppen climate classification), with hot and humid summers and mild winters. The average annual precipitation is around 107 cm. The sows were given access to slightly less than standard National Research Council balanced rations (2-3 kg/day) considering the consumption of corn in the pasture, but the water was provided *ad libitum*.

Data collection

A Phantom 2 Vision model UAV manufactured by DJI® with a quad-rotor system consisting of four propellers was used in this study. Including the camera, the maximum takeoff weight is 1.3 kg, and it can fly for about 25 minutes using a 5,200 mAh lithium polymer battery. It has a remote-control range of up to 300 m and is equipped with a 1/2.3″ high-resolution 14 megapixels camera sensor with a fixed-focus wide-angle lens with a 120° FOV (Field of View) and a focal length of 28 mm. The UAV is equipped with an automatic flight control device, and a 2.4 GHz wireless remote controller was used for takeoff and landing as well as manual control of the aircraft. Please refer to the research article by Oh et al. [2] for the detailed specifications of the UAV used in this study.

Ten aerial images were taken using the UAV from a height that allowed the entire grazing area to be captured in a single image. The image data were captured on days 2, 3, 4, 5, 7, 9, 10, 11, 12, and 14 after releasing the gestating sows onto the cornfield from September 1^{st} to September 13^{th}, 2015, excluding days with rain. Also, the images were captured around 10:00 AM without the need for additional lighting, and an effort was made to minimize the effect of shadows caused by the sun. In addition, a GPS attached to the UAV was used to attempt to maintain the same altitude and position for each image.

Image Preprocessing

Convolutional Neural Network

CNN is a multi-layered artificial neural network structure that is widely used for image recognition. It consists of a sequence of convolutional, non-linear, and pooling layers, followed by a fully connected layer that produces the final output. As the input image passes through the convolutional layers, specific features of the target object are revealed, and the final output is produced by the fully connected layer. The output layer can classify objects or produce regression values. Figure 1 represents the typical basic structure of CNN. The design of the layer structure can greatly affect the accuracy of the output and computation time. In particular, LeNet [5], developed by LeCun in the late 1990s, served as the basis for modern CNNs and had a significant impact on contemporary image recognition methods. CNNs convolve the entire image and intermediate feature maps to learn the various features of the objects in an image. As a result, CNNs make it relatively easy to find object features compared to traditional methods that require direct differentiation of objects. Additionally, CNNs can even identify features that are imperceptible to the human eye, resulting in very high recognition accuracy

[14]. CNNs continue to improve their performance in the field of image recognition, and the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), which provides a common dataset for benchmarking machine learning and computer vision models, has further accelerated the development of CNN models through competition [15].

The first winner, AlexNet [16], expanded the input image size from 32×32 in LeNet to 224×224, increasing the model size, but solved the potential problem of overfitting by applying dropout layers and significantly improved its accuracy from 73.8% to 83.7% by applying the Rectified Linear Unit (ReLU) activation function to the ImageNet tests. VggNet [17] achieved remarkable results with an accuracy of 93.2% by increasing the number of convolutional filters and expanding the layer structure while unifying the convolution filter size to 3×3 to reduce computation. However, there was no significant improvement compared to having 16 layers. GoogLeNet [6] improved the inefficient structure, and could model deeper than VggNet by using an inception module that included a 1×1 convolution filter asymmetrically connected and layers that were not fully connected, resulting in a smaller model size and faster computation.

ResNet [7] recognized that designing a deeper layer structure to improve accuracy decreased performance and it achieved a higher performance by using residual learning. Residual learning can model with deeper layer structures by directly transmitting the next layer by skipping the adjacent convolutional layers without compromising the model's generalization performance. Therefore, CNNs have a significant impact on recognition accuracy and computation time depending on how the layer structure is modeled and can vary widely depending on the field of application, requiring extensive research under various conditions.

Image collection and preprocessing

Images taken by the UAV are shown on days 3 and 10 after sows were released into the field in Figure 2. The images encompassing the entire cornfield have a resolution of

4,384×3,288 pixels, but the edges of the cornfield appear distorted into a fish-eye image. Fisheye images make the subject appear more prominent and can capture a wide range of backgrounds at the same time. However, the exact size cannot be recognized in such a distorted image because the object's size is distorted. Additionally, the camera mounted on the UAV could not consistently capture the cornfield at a fixed position and height, resulting in inconsistent left and right edges in the images, and images that included other objects outside of the cornfield. Therefore, image preprocessing was required so that the images only included the cornfield to accurately compare the corn occupancy rates of the images.

An example of image preprocessing steps for an image taken on day four after sows were released into a cornfield is illustrated in Figure 3. The original image on day four distorted into a fish-eye image is in Figure 3(a). The corrected result, as shown in Figure 3(b), is obtained by applying the correction method proposed by Scaramuzza [18], which is the best known in the field of computer vision, to the distorted fish-eye image. The method is computer vision's most representative and commonly used fisheye correction algorithm. The camera extrinsic and intrinsic parameters must be obtained to connect the 3D world coordinate points to the 2D image. World coordinate points are converted to camera coordinates using external parameters. Camera coordinates are mapped to the image plane using internal parameters. Still, a bird's eye view transformation is also required because the image is not captured at the center of the field. Fig. 3(c) shows an image containing only cornfields by transforming the bird's eye view and cropping the image. After completing the image preprocessing step, an example of 10 images is shown in Figure 4. The image generation was achieved by cutting the region of interest to 3584×1792 pixels centered on the cornfield, ensuring that no other objects were in the image.

Data Augmentation

Much training data is required to train a deep learning network. Different corn object images are needed to recognize and classify common corn objects according to the size, shape, illumination, and shadow state of corn objects. The ten transformed images were too few to be used as training and testing data for the deep learning networks, and the size of the corn objects was too small to extract sufficient features. Limited training data can lead to overfitting during network training, which can have a significant impact on performance. Fortunately, the resolution of the final images was much larger than the typical input resolution required for deep learning networks. A very high input resolution in a deep learning network increases the number of internal parameters of the network, resulting in a much longer training time, and increases the network processing time, resulting in learning and results processing difficulties. In addition, the accuracy of learning and prediction is affected by the very small size of the object (corn) in the image. To resolve this issue, the object in the image can be enlarged by cropping the network input resolution to be the size used for typical network training data. Data augmentation addresses the shortcomings of small training datasets by increasing the size of a training dataset by reflecting the characteristics of the data. Basic data augmentation can be performed on an image through various image processing techniques. A commonly used technique for data augmentation is to apply transformations that alter the physical form of the image, such as flipping the image horizontally or vertically, or rotating the image.

In this study, two different data augmentation techniques were performed for network training. First, the images were segmented to crop them to the appropriate size for a deep learning input image. Then the segmented images were randomly selected as raw training data, and data augmentation was performed by flipping and rotating the images.

Image segmentation is the process of dividing an image into smaller parts that are suitable for use as deep learning input images and increasing the amount of image data for deep learning network training. Figure 5 shows the image segmentation and augmentation process. The ten transformed images were divided into eight horizontal and four vertical sections to obtain 320 segmented images, which were used as training and testing data. Among them, 48 images were selected randomly as raw training images, and 6,912 training images were generated by flipping and rotating the images.

In the agricultural field, images acquired by UAVs are advantageous because they can undergo data augmentation by rotating the images. For recognizing objects in fields other than agriculture, small-angle rotation transformations are mainly applied. Therefore, the corn images captured by the UAV can still be used for analysis even if they are rotated 180 degrees. To increase the number of training images, 48 raw training images were horizontally flipped to create 96 images, and the resulting images were augmented by rotating them in 5-degree increments to produce 6,912 images. The 6912 training images generated by data augmentation are sufficient to train a CNN to predict the occupancy of a corn field.

Convolutional Neural Network (CNN) Training

Calculation of corn occupancy rate for training image

This study aims to represent the process of cornfield degradation by gestation sows as numerical data using the degree of corn occupancy rate. The degree of corn occupancy rate needs to be known for each training image to train the deep learning network. To calculate the degree of occupancy of corn, corn objects are first labeled with three states (CI, CD, and CS) by corn field experts. CI represents the intact state where the corn has not been eaten or damaged by pigs, CD represents the state where the corn is damaged by pigs, and CS represents the state where pigs have eaten most the corn and only the cob remains. Table 1 shows an example of the boundary boxes labeled with three states for one of the 48 training images.

After labeling the corn state for any training image, the occupied area of the corn state is calculated. ACI is the area of corn labeled CI , ACD is the area of corn labeled CD, and ACS is the area of corn labeled CS. Therefore, the corn occupancy area (ACI_i, ACD_i, ACS_i) according to the corn state in the i^{th} image is calculated as in Equation (1), and the total corn occupancy rate (ACT_i) is represented as in Equation (2).

$$ACI_i = \frac{\sum_{j=1}^{n1} ACI_{ij}}{scale}, \quad ACD_i = \frac{\sum_{j=1}^{n2} ACD_{ij}}{scale}, \quad ACS_i = \frac{\sum_{j=1}^{n3} ACS_{ij}}{scale} \tag{1}$$

where, i is the number of the image, j is the number of CI, CD, and CS in the image, and n1, n2, and n3 represent the maximum number of CI, CD, and CS in the image, and ACI_{ij}, ACD_{ij}, and ACS_{ij} are the area of CI_{ij}, CD_{ij}, and CS_{ij}, respectively. And the $scale$ is set to ensure that the occupancy rate of ACT_i does not exceed 1.

$$ACT_i = w_1 \times ACI_i + w_2 \times ACD_i + w_3 \times ACS_i \tag{2}$$

where, w_1, w_2, and w_3 are weights for each corn state.

Table 2 shows an example of corn occupancy rates for the training images. The $scale$ is set to 11,000, and the weights $[w_1, w_2, w_3]$ are set to [1, 0.5, 0.2].

Convolutional Neural Network architecture and training

CNN is the most widely used multi-layer structure for image recognition, along with various models such as AlexNet, GoogLeNet, VggNet, and ResNet [6,7,15,16]. A CNN structure using four CNNs was implemented to represent the degree of the corn occupancy rate and to verify the potential of deep learning. CNNs have shown good results in image classification and recognition and also can output regression values. Figure 6 shows a rough CNN structure for outputting regression values. The input for the CNN is an image with 448×448 pixels, and the output is the degree of corn occupancy rate for the input image as ACI_i, ACD_i, ACS_i and ACT_i. The network output applied a regression layer to produce regression values and applied a ReLU layer to eliminate negative output values. The network was trained with a dataset consisting of 6,912 images created by image augmentation and calculated data on the degree of corn occupancy rate.

Proposed system

The proposed system aimed to predict the degree of corn occupancy rate for corn images on a specific date after learning seven different types of CNN trained on a learning dataset created by data augmentation. Figure 8 illustrates the proposed system flow for a corn image taken on a specific date. A distorted fisheye image captured by the UAV on a specific date was corrected to 3584×1792 pixels and divided into 32 (4×8) partitions, which were then sequentially fed into the CNNs. The CNNs learned from the sequentially inputted images, extracted the characteristics of the three corn states, and produced the degree of occupancy rate of the three corn states as outputs for each image. The occupancy rate of the entire cornfield on a specific date is ACT, which is the average value of ACT_i over 32 sequentially inputted image outputs, as shown in Equation (3). The ACI, ACD, and ACS values in the network output can be used in Equation (2) to calculate the new occupancy rate by applying different weights depending on the state of corn.

$$ACI = \sum_{i=1}^{32} ACI_i/32, \quad ACD = \sum_{i=1}^{32} ACD_i/32$$
$$ACS = \sum_{i=1}^{32} ACS_i/32, \quad ACT = \sum_{i=1}^{32} ACT_i/32 \tag{3}$$

where, ACI, ACD, and ACS represent the occupancy rate of CI, CD and CS for a corn image on a specific date, while ACT represents the overall occupancy rate of corn.

RESULTS

When there are a large number of images captured by a UAV, it is easier to apply a deep learning system, but when there is a very limited number of images, it is difficult to apply a deep learning system. In addition, it is difficult to train deep learning when the captured images are distorted and not taken from the same location. The proposed system aimed to predict the occupancy rate of corn using a small number of distorted images captured by a UAV. Therefore, bird's-eye view images were used to correct the distorted fisheye images and extract corn field images containing no other object images. In addition, due to the lack of training images, the corn field images were divided into 32 parts and image augmentation was performed by randomly selecting raw training images and rotating and flipping them. Seven types of CNNs were trained using the 6,912 augmented training images, and the occupancy rate according to the corn states and the overall occupancy rate of corn on a specific date were predicted using the trained CNNs. AlexNet, GoogLeNet, Vgg16, Vgg19, ResNet50, and ResNet101 were applied to the structure of the CNN, and the applicability and scalability of deep learning were confirmed.

Figure 7 shows the degree of corn occupancy rate for all images. The CNN used a network provided by Matlab® [19] and the output network was configured to suit the proposed purpose. The same initial learning rate was set to 0.0001 to evaluate the performance of seven types of CNN. Adam Optimizer was used for learning, and network learning was performed with a maximum epoch of 500 and a mini-batch size of 32. The hardware used for the experiment was an Intel i9-12900 CPU and an NVIDIA RTX-A6000 graphics accelerator.

Table 3 shows the results of training the seven types of CNN. All networks were trained five times, and the performance evaluation index was obtained by averaging the values. The root mean square errors (RMSE) show that ResNet50 has the smallest learning error of the networks, while GoogLeNet has the largest. The network learning time varies depending on the size of the network, with AlexNet taking the shortest time and ResNet101 taking the longest time. Thus, AlexNet and the ResNet series were found to be advantageous in learning in terms of RMSE and learning time.

Figures 9 to 12 indicate the occupancy rate according to the corn state and the total corn occupancy rate in order of date using the ten cornfield images input into the CNN. Overall, the graphs displayed similar trends. Figures 9(a)-12(a) show that the occupancy rate of undamaged corn (ACI) decreases gradually over time. Figures 9(b)-12(b) show that the occupancy rate of corn damaged by sows (ACD) sharply increases initially and then decreases rapidly from day 4. Figures 9(c)-12(c) show that the occupancy rate of corn damaged by sows gnawing at the corn, leaving only a stump (ACS) also sharply increased until day 4 and then decreased gradually. Figures 9(d)-12(d) show that the overall occupancy rate of corn by date decreases exponentially over time across all networks. AlexNet, Vgg16, and Vgg19 show particularly good prediction accuracy compared to the other networks. GoogLeNet showed that the occupancy rates of ACD on day 3 and day 5 were slightly higher than on day 4, and the rates of ACS and ACT were slightly higher on days 11 and 12 compared to day 10. ResNet50 and ResNet101 generally showed good prediction accuracy, but their predictions were slightly higher on day 14. Overall, the CNNs demonstrated excellent prediction accuracy, confirming the potential and scalability of deep learning. The proposed method effectively estimated the occupancy rate of a limited number of cornfield photos, and there is a high potential for expanding it into other areas of livestock farming in the future.

DISCUSSION

Deep learning has been validated for its performance in various fields, and it has also demonstrated high recognition accuracy and detection time in agricultural applications, such as pest and disease diagnosis and prediction, fruit detection and maturity determination, and yield prediction [4-8,16,20,21]. In one agricultural application, Priyadharshini et al. [9] classified corn leaves into four states, three based on diseases that appear on corn leaves and one normal state. They trained a modified version of LeNet [5] on the Plantvillage dataset and achieved a high accuracy of over 97%. Koirala et al. [10] suggested the Mango YOLO (You Only Look Once) network detects mangoes in real time and achieves excellent real-time performance with an F1 score of 0.97. Fu et al. [11] used the YOLOv4 network to accurately detect various sizes and shapes of bananas in harsh environments such as orchards. They demonstrated better detection speed and accuracy than tests performed on the YOLOv3 network. Kitano et al. [12] used U-Net to predict the growth of corn at an early stage from images of corn fields taken by a UAV(Unmanned Air Vehicle). Mota-Delfin et al. [13] also used the YOLO method to effectively detect corn in cornfields with large numbers of weeds in the background and predict the yield. Oh et al. [2] trained a YOLOv4 network using a small number of images of cornfields and calculated the occupancy rate of cornfields by detecting corn objects.

The rationale behind choosing these specific models (AlexNet, GoogLeNet, VggNet, ResNet) was based on their well-established performance and effectiveness in various computer vision tasks. These models have been widely used and tested in different research and industrial applications, demonstrating state-of-the-art results in image recognition and classification tasks. Among the deep learning currently studied, the structures of CNNs to which regression can be applied represent the four network types tested in this study. Therefore, this experiment included the above four types and subtypes, and it is confirmed that deep learning shows robust performance not only for object classification tasks but also for regression. AlexNet was one of the pioneering deep learning models that gained significant attention after winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012. Its success was attributed to its deep architecture and the use of ReLU activation functions. GoogLeNet, also known as Inception, introduced the concept of inception modules, which allowed the network to capture features at multiple scales. This architecture proved to be highly efficient and achieved outstanding performance on ILSVRC in 2014. VggNet, short for Visual Geometry Group Network, is known for its simple and uniform architecture with a deep stack of 3x3 convolutional layers. Despite its straightforward design, VggNet showed impressive results in ILSVRC 2014. ResNet (Residual Network) addressed the problem of vanishing gradients in very deep networks by introducing skip connections or residual blocks. This innovation enabled the successful training of extremely deep models, with ResNet becoming the winning model of ILSVRC 2015. Given their track record of success and the depth of their architectures, these models were chosen as they provided a strong foundation for comparison in this study of deep learning applicability and scalability.

In a pig grazing area, the decrease in the occupancy rate of undamaged corn over time could be attributed to several factors related to pig behavior. Pigs are known to forage and consume plants, including corn, as part of their diet. Over time, as pigs continue to graze in the area, they may consume or damage some of the undamaged corn plants, leading to a decrease in their occupancy rate. Pigs exhibit grazing behavior, preferring corn varieties, resulting in a higher rate of damage to corn plants, leading to a decline in their occupancy rate of intact corn. Pigs might cause physical damage to corn plants by trampling on them or rooting around the area. Such damage can hinder the growth and survival of corn plants, contributing to the decrease in their occupancy rate over time.

The sharp increase followed by a decrease in the occupancy rate of corn damaged by sows can be explained by several factors related to sow behavior. When sows are introduced to the grazing area, they might initially exhibit increased feeding activity and target the readily available and easily accessible corn plants. This initial feeding frenzy could lead to a sharp increase in the occupancy rate of damaged corn. Also, the grazing area was limited, the concentrated feeding activity of sows at the beginning could lead to a quick increase in the occupancy rate of damaged corn. In competition with other sows, if sows initially focus on consuming only the intact corn ears, leaving behind the corn stalks, their main interest may shift elsewhere afterwards exploring other areas of the grazing field or shifting their focus to alternative food sources such as pellet feed. The combination of sow behavior mentioned above can lead to the observed pattern of an initial increase and subsequent decrease in the occupancy rate of corn damaged by sows, and this aligns with the results we have analyzed through images in this study.

CONCLUSION

Deep learning has proven its performance in various fields and has demonstrated high recognition accuracy and detection time in agricultural applications such as pest and disease diagnosis and prediction. Most yield predictions involve designing a CNN-based object detector for an image, counting the number of detected objects, or calculating the number of pixels occupied by objects to predict yield or occupancy. These methods require several computational steps besides a detector, and their accuracy strongly depends on whether the object features of interest are visible in the image. However, in addition to object detection and classification, CNNs can be applied to function approximation and regression problems. Therefore, if CNN is used as a regression network, the occupancy of a specific object in an image can be expressed as a network output without going through several calculation steps for object classification. This study applied the four most widely known networks (AlexNet, Vgg16, Vgg19, and GoogLeNet) as regression networks to predict the market share according to corn condition and total corn share in day order.

In conclusion, this study emphasizes the importance of accurately measuring and addressing land degradation concerns associated with pasture-based pig farming. The proposed methodology offers an effective means to evaluate the extent of land degradation with a limited number of cornfield photos showing excellent prediction accuracy, and confirms the potential and scalability of deep learning. By taking proactive steps towards mitigating land degradation, the pasture-based pig farming sector can continue to thrive while preserving the environment and promoting socio-economic well-being.

CONFLICT OF INTEREST

We certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

REFERENCES

1. Szyndler-Nędza M, Nowicki J, MałOpolska M. The production system of high quality pork products – an example. Ann. Warsaw Univ. of Life Sci. – SGGW, Anim. Sci. 2019;58(2): 181–198.

2. Oh S, Park H, Park J. Estimating vegetation index for outdoor free-range pig production using YOLO. Journal of Animal Science and Technology. 2023;65(3): 638-651.

3. Oh S, Park H, Jung Y, Park J. 2023. Estimating vegetation index for outdoor free-range pig production. Korean Journal of Agricultural Science. 2023;50: 141-153.

4. Voulodimos A, Doulamis N, Bebis G, Stathaki T. Recent developments in deep learning for engineering applications. Computational intelligence and neuroscience, 2018; 8141259.

5. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Communications of the ACM. 2017;60(6): 84-90.

6. Zou Z, Chen K, Shi Z, Guo Y, Ye J. Object detection in 20 years: A survey. Proceedings of the IEEE. 2023;111(3): 257-276.

7. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations. 2015;arXiv:1409.1556.

8. Alibabaei K, Gaspar PD, Lima TM, Campos RM, Girão I, Monteiro J, Lopes CM. A Review of the Challenges of Using Deep Learning Algorithms to Support Decision-Making in Agricultural Activities. Remote Sensing. 2022;14(3): 638.

9. Priyadharshini RA, Selvaraj A, Madakannu A. Maize leaf disease classification using deep convolutional neural networks. Neural Computing and Applications, 2019;31(12): 8887-8895.

10. Koirala A, Walsh KB, Wang Z, McCarthy C. Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of 'MangoYOLO'. Precision Agriculture. 2019;20(6): 1107-1135.

11. Fu L, Duan J, Zou X, Lin J, Zhao L, Li J, Yang Z. Fast and accurate detection of banana fruits in complex background orchards. IEEE Access. 2020;8: 196835-196846.

12. Kitano BT, Mendes CCT, Geus AR, Oliveira HC, Souza JR. Corn plant counting using deep

learning and UAV images. IEEE Geoscience and Remote Sensing Letters. 2019. doi: 10.1109/LGRS.2019.2930549.

13. Mota-Delfin C, López-Canteñs GJ, López-Cruz IL, Romantchik-Kriuchkova E, Olguín-Rojas JC. Detection and Counting of Corn Plants in the Presence of Weeds with Convolutional Neural Networks. Remote Sensing. 2022;14(19): 4892.

14. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. Journal of big data. 2019;6(1): 1-48.

15. Taylor L, Nitschke G. Improving deep learning with generic data augmentation. in 2018 IEEE Symposium Series on Computational Intelligence (SSCI). 2018;arXiv:1708.06020

16. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998;86(11): 2278-2324.

17. Wang X. Improving Bag-of-Deep-Visual-Words model via combining deep features with feature difference vectors. IEEE Access. 2022;10: 35824-35834.

18. Scaramuzza D, Martinelli A, Siegwart R. A flexible technique for accurate omnidirectional camera calibration and structure from motion. Fourth IEEE International Conference on Computer Vision Systems (ICVS'06), New York, NY, USA, 2006;45.

19. The Math Works, Inc. MATLAB. R2021a, The Math Works, Inc., 2020. Computer Software. www.mathworks.com

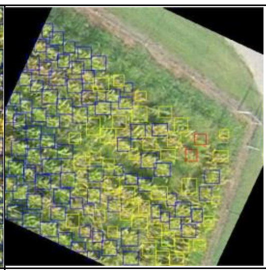
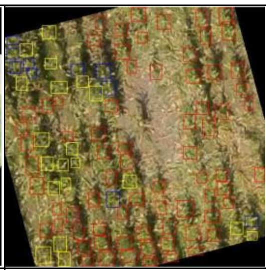
20. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015;arXiv:1409.4842.

21. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016;arXiv:1512.03385.

Table 1. Image label

Label	Color	Corn description	Sample
CI	Blue	Intact corn	
CD	Yellow	Damaged corn	
CS	Red	Corn with stubble	

Table 2. The occupancy rate of corn

Sample Images				
$\boldsymbol{ACI_i}$	0.975	0.369	0.036	0
$\boldsymbol{ACD_i}$	0.026	0.164	0.080	0
$\boldsymbol{ACS_i}$	0	0.007	0.288	0.005
$\boldsymbol{ACT_i}$	0.988	0.452	0.134	0.001

$\boldsymbol{ACI_i}$ **:** The occupancy rate of intact corn; $\boldsymbol{ACD_i}$ **:** The occupancy rate of damaged corn corn; $\boldsymbol{ACS_i}$ **:** The occupancy rate of corn with stubble; $\boldsymbol{ACT_i}$ **:** The occupancy rate of corn in all conditions

Table 3. Training results

Network(CNN)	RMSE	Training Time	Training Performance
AlexNet	0.16	74 min. 48 sec.	Very good
GoogLeNet	0.19	107 min. 16 sec.	Little good
VggNet16	0.14	355 min. 11 sec.	Good
VggNet19	0.14	414 min. 32 sec.	Good
ResNet18	0.14	88 min. 41 sec.	Very good
ResNet50	0.05	345 min. 45 sec.	Very good
ResNet101	0.07	623 min. 37 sec.	Good

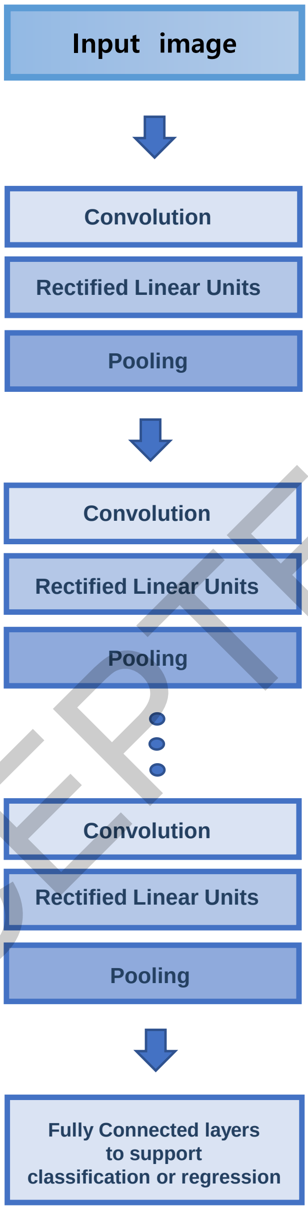
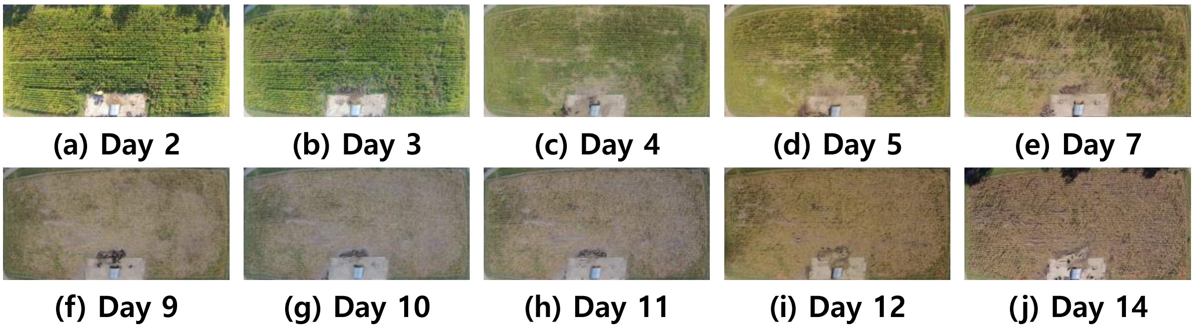


Figure 1. Structure of CNN

(a) Day 3 **(b) Day 10**

Figure 2. Examples of images taken by the UAV

Figure 3. Example of image correction.



(a) Day 2 (b) Day 3 (c) Day 4 (d) Day 5 (e) Day 7

(f) Day 9 (g) Day 10 (h) Day 11 (i) Day 12 (j) Day 14

Figure 3. Images after pre-processing.

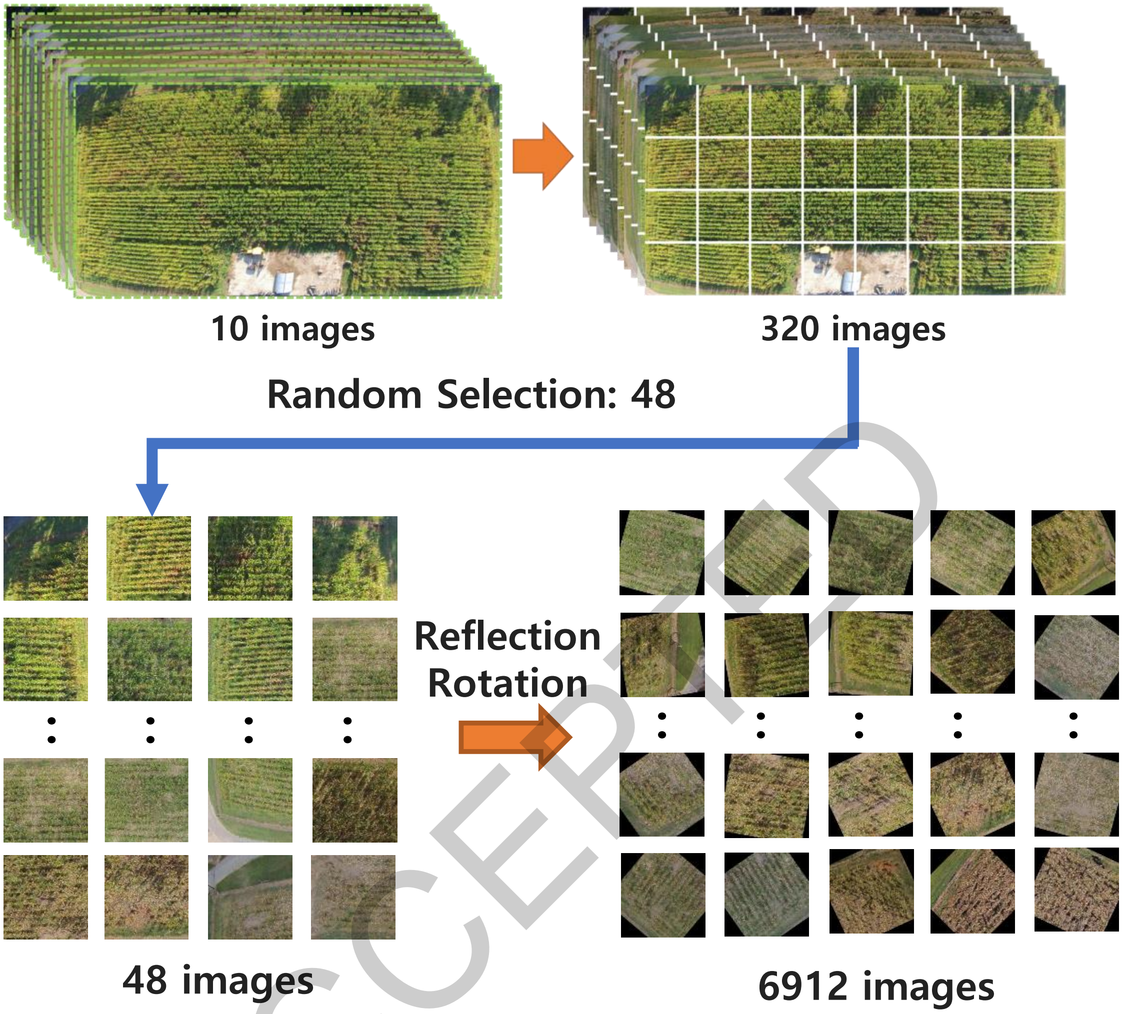


Figure 5. Data augmentation

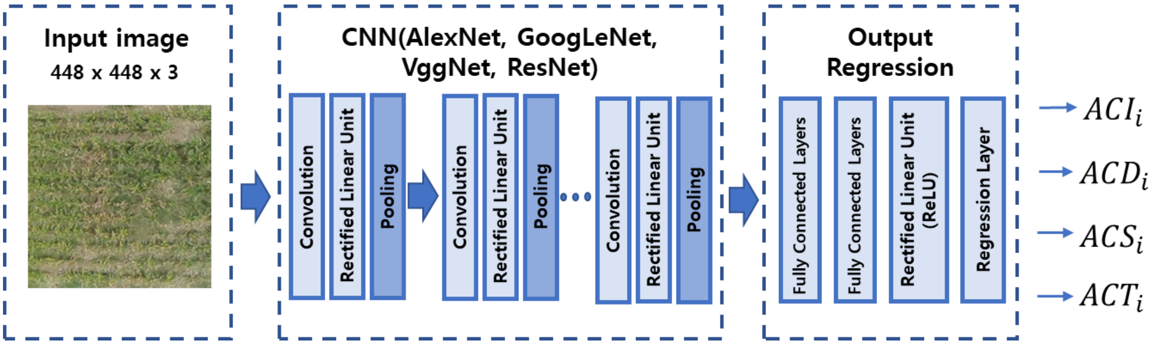
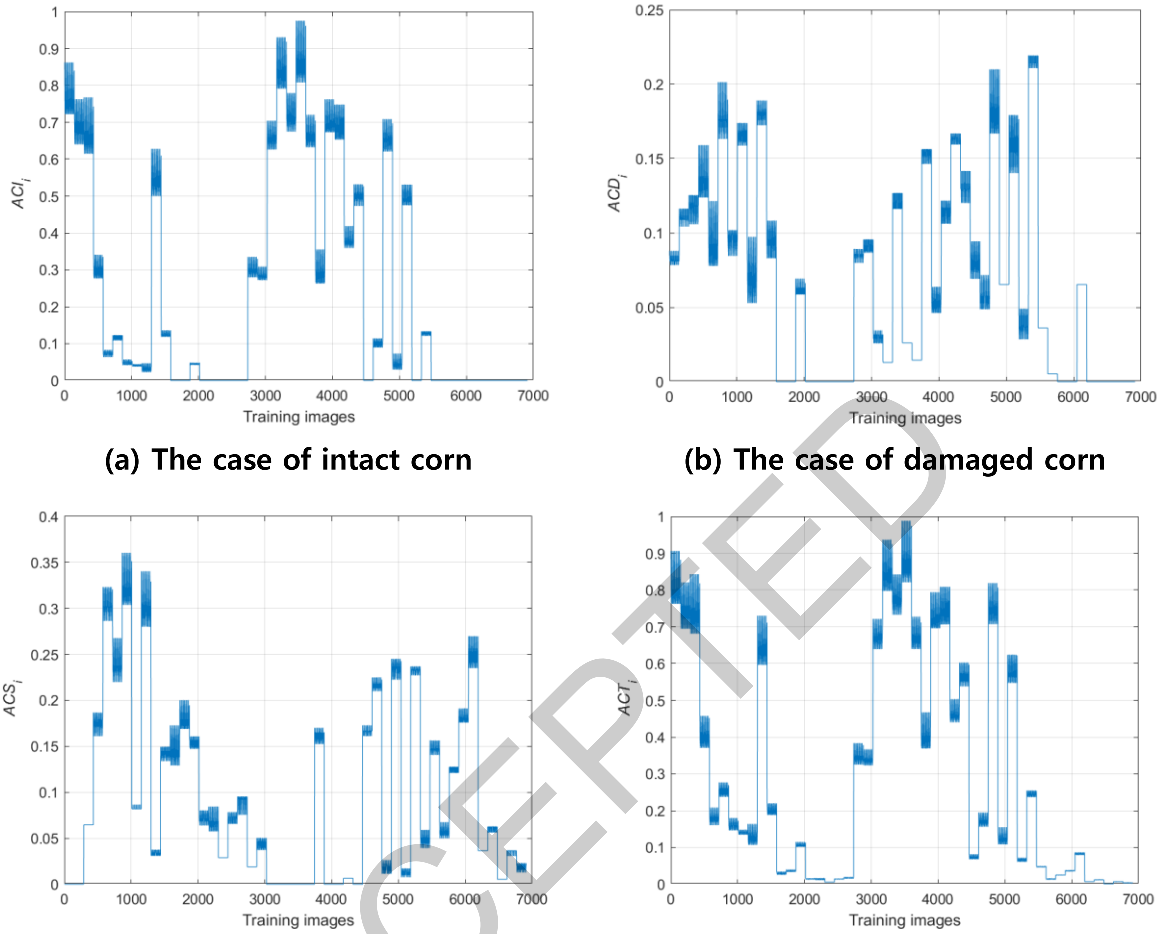


Figure 6. The CNN structure for regression



(a) The case of intact corn (b) The case of damaged corn

(c) The case of corn with stubble (d) The occupancy rate of corn

Figure 7. The degree of corn coverage in the training data

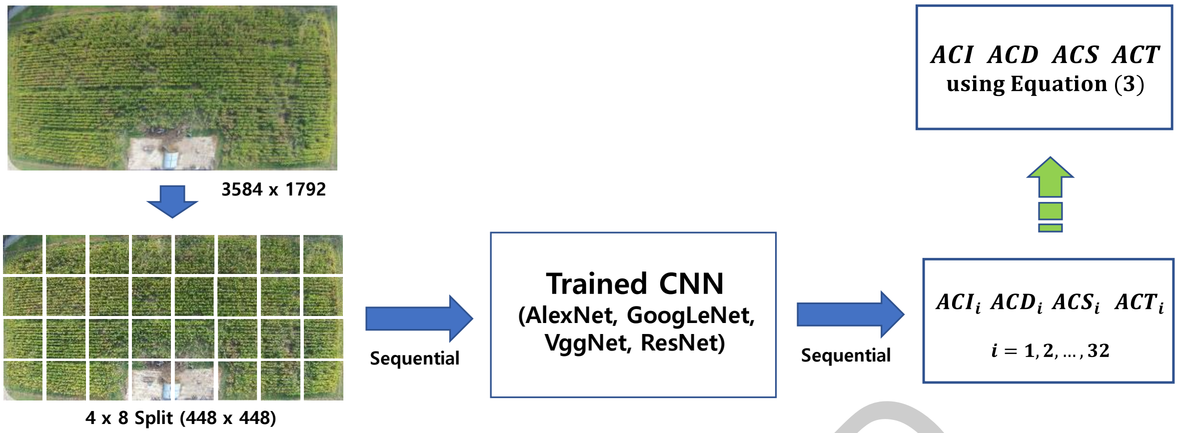
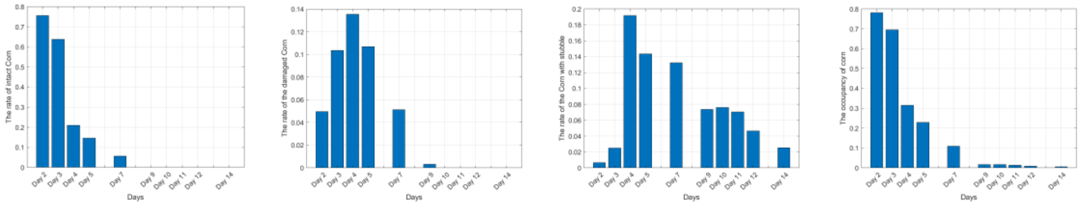
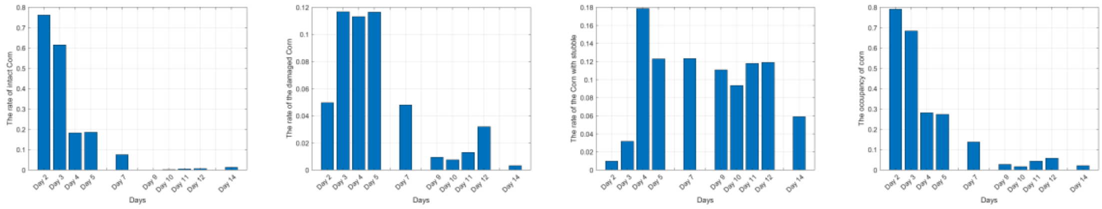


Figure 8. System configuration



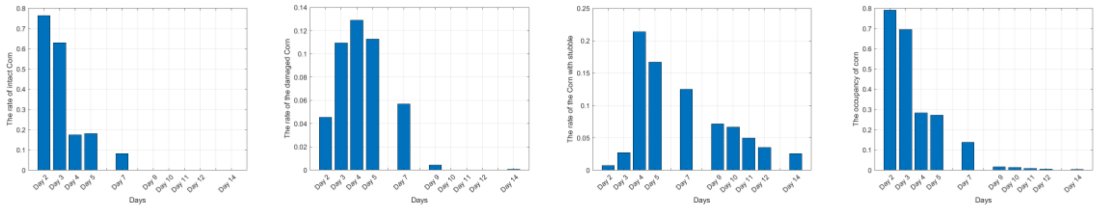
(a) ACI (b) ACD (c) ACS (d) ACT

Figure 9. The occupancy rate of corn by date(AlexNet)



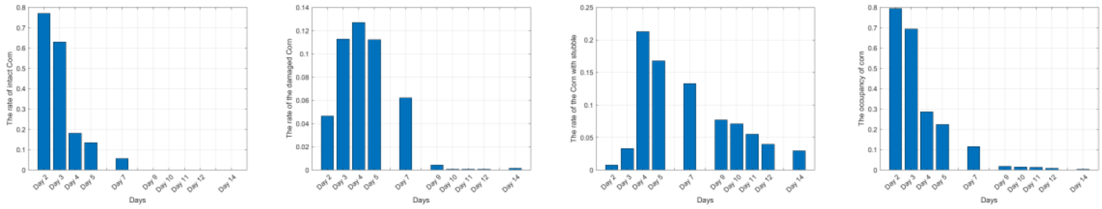
(a)ACI (b) ACD (c) ACS (d) ACT

Figure 10. The occupancy rate of corn by date (GoogLeNet)



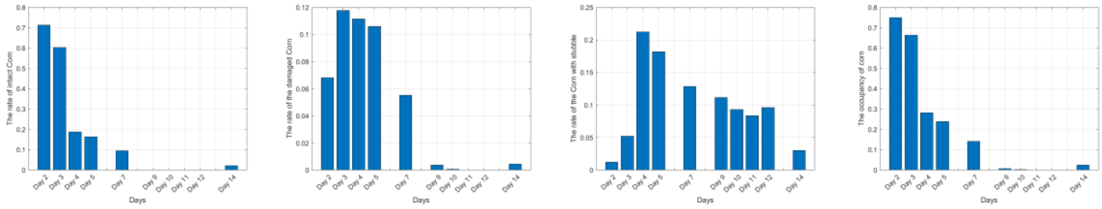
(a)ACI (b) ACD (c) ACS (d) ACT

Figure 11. The occupancy rate of corn by date (Vgg16)



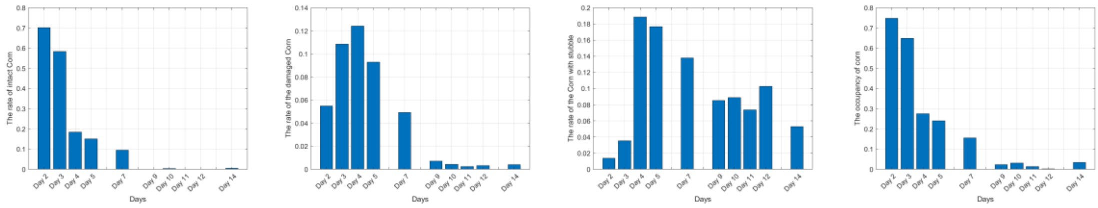
(a)ACI (b) ACD (c) ACS (d) ACT

Figure 12. The degree of occupancy of corn by date(Vgg19)



(a) ACI (b) ACD (c) ACS (d) ACT

Figure 13. The degree of occupancy of corn by date(ResNet50)



(a) ACI (b) ACD (c) ACS (d) ACT

Figure 14. The degree of occupancy of corn by date(ResNet101)