Growth prediction of tomato seedlings based on causal LSTM and GAN

Hongduo Zhang, Yutaka Kaizu*, Kenichi Furuhashi, Heming Hu, Kenji Imou

(*Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1, Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan*)

Abstract: The stable production of seedlings is very important for seedling growers. Predicting the growth of seedlings helps growers promptly adjust management strategies and production expectations. Traditional methods rely on historical growth data or assess current plant physiological parameters to estimate future growth. This study aims to predict future images directly from historical growth images of tomato seedlings. Specifically, a dataset of 10-d image sequences of tomato seedlings was collected. Then, an algorithm based on several neural networks was applied to predict the images of the next 5 d based on the images of the first 5 d. The algorithm was composed of a causal long short-term memory (LSTM) unit, a gradient highway unit (GHU), and a pix2pix unit. The experimental results showed that the introduction of a Generative Adversarial Network (GAN) further enhanced the clarity and realism of the predicted images, ensuring higher quality and more accurate visual results. From the perspective of image similarity, the average mean squared error (MSE) reached 394.97 and the average structural similarity (SSIM) reached 0.90 over 5 d. From the perspective of biological information, the average prediction errors of the plant area were 1.7, 1.4, 1.5, 0.9, and 3.2 cm^2 over the 5 d, and the average prediction errors of plant height were 1.7, 1.9, 4.6, 6.9, and 4.5 mm, respectively. The extracted biological information such as plant area and height showed good following performance compared with the real growth information. The research results show that predicting future plant images from historical images has the potential to become a useful tool for nursery growers to adjust management strategies and production expectations.

Keywords: growth prediction, tomato seedlings, LSTM, GAN, deep learning

DOI: 10.25165/j.ijabe.20251803.8757

Citation: Zhang H D, Kaizu Y, Furuhashi K, Hu H M, Imou K. Growth prediction of tomato seedlings based on causal LSTM and GAN. Int J Agric & Biol Eng, 2025; 18(3): 51–57.

1 Introduction

Tomatoes are one of the most important varieties of vegetable in Japan[1], and many tomato seedlings are produced by seedling growers and supplied to farmers for cultivation. Tomato seedlings play a critical role in ensuring successful cultivation, as they are produced by specialized seedling growers and then supplied to farmers. However, the growth and quality of these seedlings are highly influenced by various environmental factors, such as light, temperature, humidity, and fertilizer concentration. Ensuring production stability for seedling growers remains a key challenge, requiring effective solutions to optimize management strategies[2]. For instance, adjusting nutrient solution formulas based on growth stages has been shown to promote yield and quality in greenhouse strawberry production, a practice that could similarly benefit tomato seedling cultivation[3]. To ensure the stability of production, researchers have proposed various methods for seedling growers to adjust management strategies. There are two main types of methods: one is to use nonimage data to predict the final yield through artificial neural networks, and the other is to use nondestructive measurement methods, mainly image data, to obtain real-time plant growth and disease statuses.

Methods that use nonimage data to predict seedling growth mainly make use of an artificial neural network. Using the SVM method, the apparent photosynthetic rate of the tomato canopy can be predicted[4]. Moriyuki and Fukuda[5] used leaf size and circadian rhythms to predict early seedling fresh weight gain. Temperature, solar irradiance, and vapor pressure parameters can be used to predict the leaf area index and daily dry weight through a Bayesian network[6]. The field canopy coverage index can be predicted by combining image monitoring data and climate environment through neural networks[7]. The real-time acquisition of biological information and health status by nondestructive measurement is realized mainly through image processing technology. For example, the image processing method was used to detect the nodes of tomato seedlings and estimate the lengths between nodes. This method provided a basis for the evaluation of the growth and vitality of tomato seedlings[8]. Relying on long-term collected public datasets, researchers have proposed deep learning-based methods to classify tomato leaves with different diseases[9-12]. With the introduction of a real-time detection framework, it is possible to locate pests and diseased parts of tomato seedlings growing in a greenhouse[13,14]. Based on the greenhouse monitoring platform[15] combined with the convolutional neural network[16], the detection accuracy of tomato leaf diseases can be improved. However, methods that rely on static data do not take into account real-time environmental changes, which makes it difficult to apply to complex and dynamic agricultural environments. Methods that rely on images are usually used to assess the current status and have no predictive ability. Addressing these limitations requires combining the advantages of both methods to achieve dynamic and predictive modeling of seedling growth.

One of the achievements of machine vision tasks in recent

Received date: 2023-12-26 **Accepted date:** 2024-12-02

Biographies: Hongduo Zhang, PhD candidate, research interest: machine vision and controlled environment agriculture, Email: zhang-hongduo@g.ecc.u-tokyo.ac.jp; **Kenichi Furuhashi**, Asst. Professor, research interest: biofuel production and agricultural machinery, Email: kfuruhashi@g.ecc.u-tokyo.ac.jp; **Heming Hu**, PhD, research interest: precision spraying and agricultural machinery, Email: heming950211@g.ecc.u-tokyo.ac.jp; **Kenji Imou**, Professor Emeritus, research interest: agricultural machinery and biomass energy, Email: k-imou@g.ecc.u-tokyo.ac.jp.

***Corresponding author: Yutaka Kaizu**, Assoc. Professor, research interest: agricultural robotics and renewable energy. 1-1-1 Yayoi, Bunkyo, Tokyo, 113-8657, Japan, Tel: +81-3-5841－5358, Email: kaizu@g.ecc.u-tokyo.ac.jp.

years is to predict future image frames from past image frames in video footage based on long short-term memory (LSTM) technology. It has been successfully applied to automatic navigation[17], human behavior prediction[18], and weather forecasting[19]. Likewise, if it is possible to predict future images from past images of tomato seedlings, LSTM will help seedling growers adjust management strategies and production expectations to maintain seedling production stability. Based on public datasets such as Komanatsuna[20] and Arabidopsis[21], researchers have proposed methods to predict seedling growth images at an early stage from the top view of plants[22-24]. However, due to the short sampling time interval of these datasets, these studies achieved image prediction after a few hours and were only evaluated using image similarity metrics. However, image growth prediction over a longer time needs to be implemented, and the accuracy of the method needs to be measured with biological information.

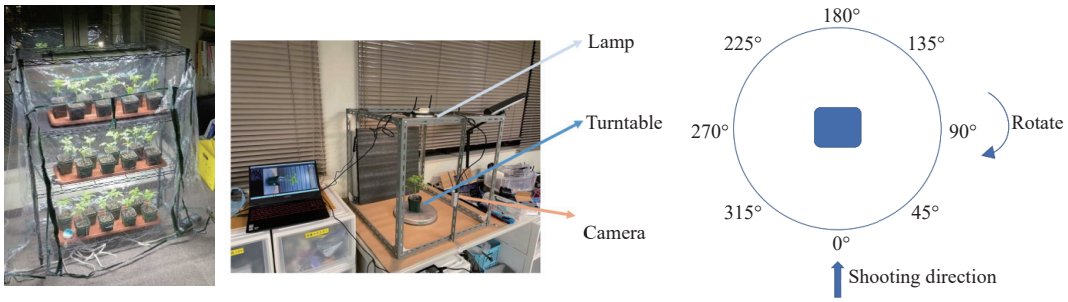
In this research, this study aims to predict future 5-day images of tomato seedlings using images from the previous 5 d. This can allow seedling growers to perform forecasting and adjust management strategies and production expectations. The objectives of this study are to be presented as follows: (1) To build an efficient neural network for image prediction of tomato seedling growth from 5 d into the future; and (2) to judge whether the prediction is effective from two aspects: image similarity and biological information.

2 Materials and methods

2.1 Materials

2.1.1 Data collection

Materials used were provided by a seedling grower (Evergreen Fujimi K. K., Gunma, Japan) and included a total of 24 fresh tomato grafting seedlings (grafting is performed in the early stages of tomato seedlings, when they have 2.5-3 leaves) with the root stock of Cady 1 and the scion of Momotaro 8. The front of each pot was labeled from 1 to 24. To obtain image sequences of seedling growth, a simple greenhouse and photographic platform were built. As shown in Figure 1a, the simple greenhouse had three levels, and each level could hold six tomato seedlings. There were LED lights at the top of each level to maintain 12 h of light per day. Each tomato seedling was supplied with 100 mL of water per day. The temperature was maintained at 25°C during the day and 15°C at night by air conditioning. As shown in Figure 1b, the photo platform consists of a turntable, a camera (RealSense D435, Intel, Santa Clara, CA), and a circular LED lamp. As shown in Figure 1c, the turntable angle was labeled from 0° to 315°. When collecting data, this study placed the seedling pot at the center of the turntable to align the front label of the pot with the 0° label. Then, the turntable was rotated to the next labels in turn, and images were taken from each angle. Figure 2 shows examples of images taken from 0° and 45°.



a. Simple greenhouse b. Photo platform c. Photo method

Figure 1 Process of collecting data

a. Taken from 0° b. Taken from 45°

Figure 2 Photo examples

For 10 consecutive days, the seedlings were photographed sequentially starting at 10:00 am and 10:00 pm each day, and the photographing of the 24 plants, in all eight directions, was completed at approximately 12:00 am and 12:00 pm each day. A total of 3840 images were collected.

2.1.2 Preprocessing

To reduce the amount of computation and the interference of noise, the obtained images were preprocessed as shown in Figure 3. First, each image was cropped from 720×1080 pixels to 720×720 pixels. The selection of the region of interest (ROI) was based on the principle of ensuring that the growth of the seedling was always within the image. The cropped image was then processed with the excess green (ExG) index, which removed the background and retained only the seedlings. This index is generally used for background segmentation when the target is green plants. Then, the images were resized to a size of 256×256. After preprocessing, 1 pixel corresponded to 0.94 mm, and the entire image area corresponded to 576 cm^2 in reality.

Because of the nyctinastic movement of plants[25], leaves of tomato seedlings closed at night and opened in the morning, and the dataset was divided into Night and Day groups. Taking every 10 images as a sequence, there were 24 seedlings, each seedling contained eight angles, and a total of 192 image sequences could be obtained for both Night and Day. Three seedlings were randomly selected from 24 seedlings as the test set, and the remaining 21 seedlings were used as the training set. Thus, in each group, a total of 168 sequences were used for training, and 24 sequences were used for testing. Figures 4a and 4b show examples of growth sequences of the same plant during the day and at night. Specifically, this study aims to predict the images of the next 5 d (t = 6\sim 10) based on the images of the first 5 d (t = 1\sim 5).

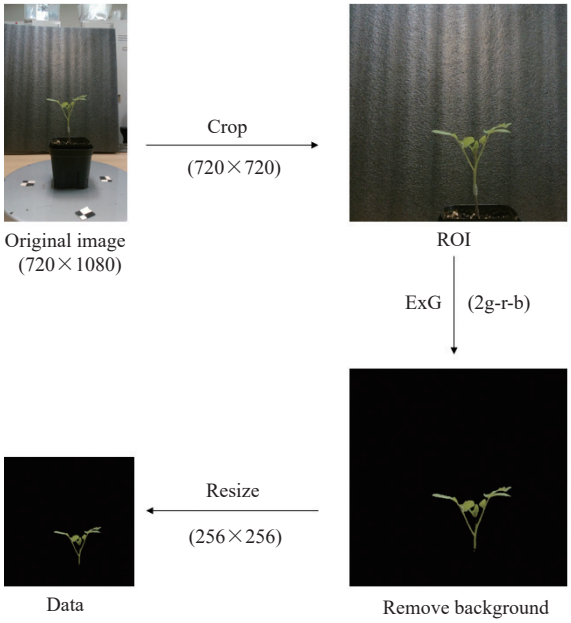
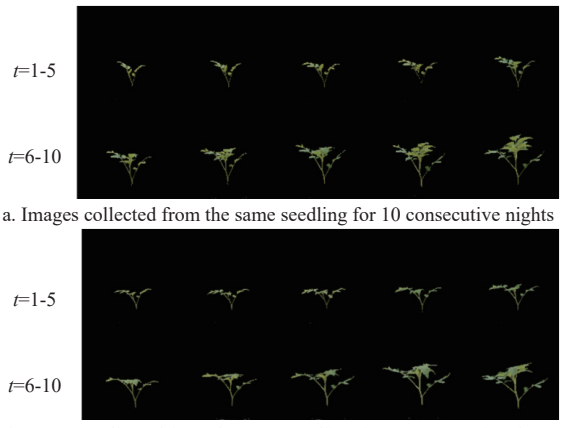


Figure 3 Schematic diagram of image preprocessing flow



a. Images collected from the same seedling for 10 consecutive nights

b. Images collected from the same seedling for 10 consecutive days

Figure 4 Sequence examples

2.2 Method of prediction

2.2.1 Causal LSTM

LSTM is a recurrent neural network (RNN) architecture that is usually composed of forget gates, memory gates, and output gates and can be used to process time-series data[26]. It has been applied to radar echo prediction[27], traffic flow prediction[28], and video prediction[29]. The causal LSTM unit[30] shown in Figure 5 is connected in a cascaded way through gate operations and contains the spatial memory M_t^k and the temporal memory C_t^k. This design incorporates a deeper network structure compared to a standard LSTM network, thereby enhancing its capacity to extract spatiotemporal features more effectively. The formulas can be written as follows:

$$\begin{pmatrix} g_t \\ i_t \\ f_t \end{pmatrix} = \begin{pmatrix} tanh \\ \sigma \\ \sigma \end{pmatrix} W_1 * \left[X_t, H_{t-1}^k, C_{t-1}^k\right] \tag{1}$$

$$C_t^k = f_t \odot C_{t-1}^k + i_t \odot g_t \tag{2}$$

$$\begin{pmatrix} g_t' \\ i_t' \\ f_t' \end{pmatrix} = \begin{pmatrix} tanh \\ \sigma \\ \sigma \end{pmatrix} W_2 * \left[X_t, C_t^k, M_t^{k-1}\right] \tag{3}$$

$$M_t^k = f_t' \odot tanh\left(W_3 * M_t^{k-1}\right) + i_t' \odot g_t' \tag{4}$$

$$H_t^k = tanh\left(W_4 * \left[X_t, C_t^k, M_t^k\right]\right) \odot tanh\left(W_5 * \left[C_t^k, M_t^k\right]\right) \tag{5}$$

where, t represents the timestamp and k represents the k^{th} causal LSTM module in the t timestamp. New temporal memory and spatial memory are generated by an input gate i_t, forget gate f_t, and input modulation gate g_t. The symbol $*$ represents convolution, $\odot$ represents elementwise multiplication, and $W_{1\sim5}$ represents the convolutional filter. The output of this timestamp H_t^k is determined jointly by M_t^k and C_t^k.

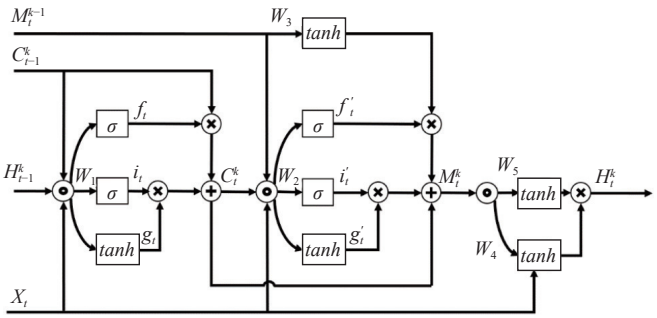


Figure 5 Architecture of causal LSTM

2.2.2 Gradient highway unit

In the traditional RNN structure, when the network is deep, the gradient is difficult to transmit, so the features of the raw input are easily forgotten. The gradient highway unit (GHU), as shown in Figure 6, is designed to enhance the raw input features at each timestamp[31]. It is connected between the first and second causal LSTM modules and traverses the entire network, which allows it to effectively transfer long-term skip-frame relationships.

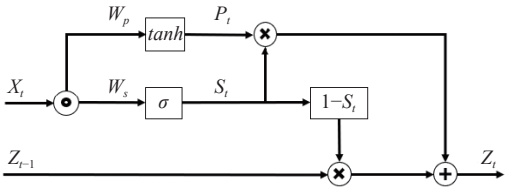


Figure 6 Architecture of gradient highway unit

The formulas of GHU can be written as follows:

$$Z_t = S_t \odot P_t + (1 - S_t) \odot Z_{t-1} \tag{6}$$

$$P_t = tanh\left(W_{px} * X_t + W_{pz} * Z_{t-1}\right) \tag{7}$$

$$S_t = \sigma\left(W_{sx} * X_t + W_{sz} * Z_{t-1}\right) \tag{8}$$

where, S_t, called the switch gate, can adaptively learn the transformed input P_t and the hidden states Z_t.

2.2.3 Pix2pix GAN

The pix2pix module as shown in Figure 7 is a kind of generative adversarial network (GAN) that can transfer one kind of image style to another[32]. It is made up of a generator and discriminator. The generator is used to generate a new image, while the discriminator is used to determine the gap between the generated image and the target image. The generator is trained by the discriminator to improve its performance. The original intention of using the pix2pix module was as follows: When the input image is processed multiple times by the causal LSTM module, the final output result might be blurred, which is difficult for a human to recognize as a seedling. The pix2pix module can convert blurred

images into more realistic and detailed images. Therefore, it can improve the image quality of the final prediction.

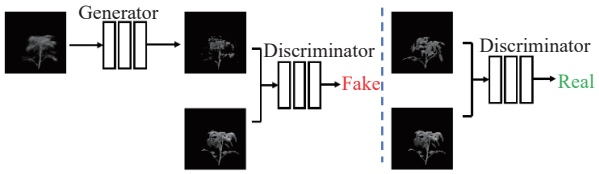


Figure 7 Schematic diagram of pix2pix GAN

2.2.4 Structure of the whole network

Figure 8 shows the structure of the entire network, which is mainly composed of three modules: causal LSTM, gradient highway unit (GHU), and pix2pix. *d* at the bottom represents the timestamp; in chronological order, images are input from the position of each timestamp. For example, the image of the first day is input from the position of t=1. Each image is processed by four superimposed causal LSTM modules to update spatial memory M_t^k and temporal memory C_t^k. The k^{th} causal LSTM modules of adjacent timestamps are connected, and the output at the previous timestamp is connected to the first causal LSTM module at the next timestamp. The GHU module traverses the entire network and accepts the output of the first causal LSTM at each timestamp. The pix2pix module is set in the output stage, accepts the output of the causal LSTM, and outputs the final predicted image. The output of each timestamp can be regarded as a prediction for the next day. For example, the output at t=7 is used as the predicted image on the eighth day. After the training is completed, in actual use, the pictures of the seedlings in the first 5 d are input, and through this network, the images of the next 5 d can be obtained.

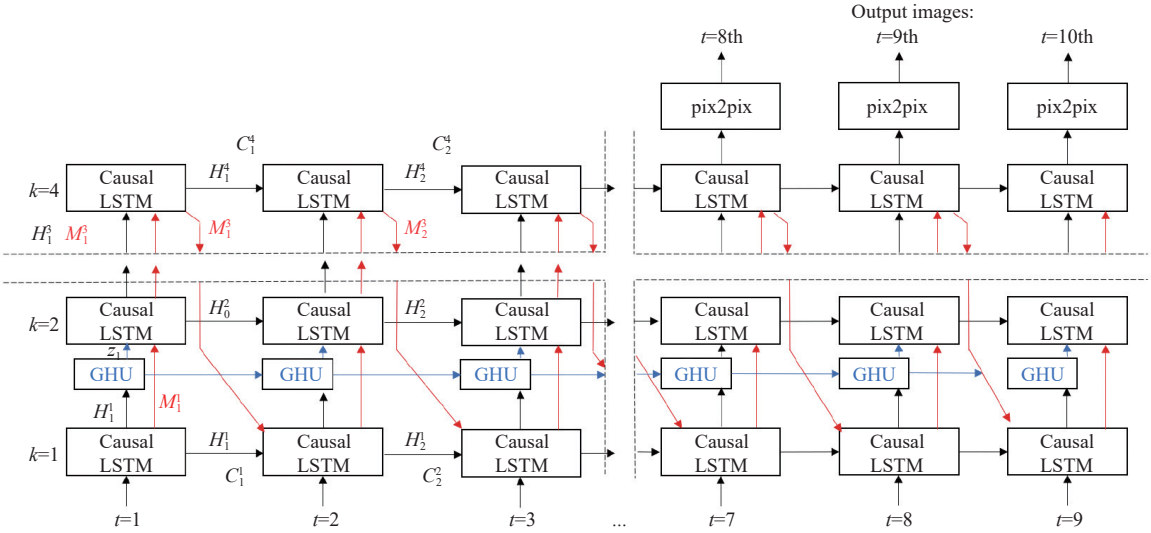
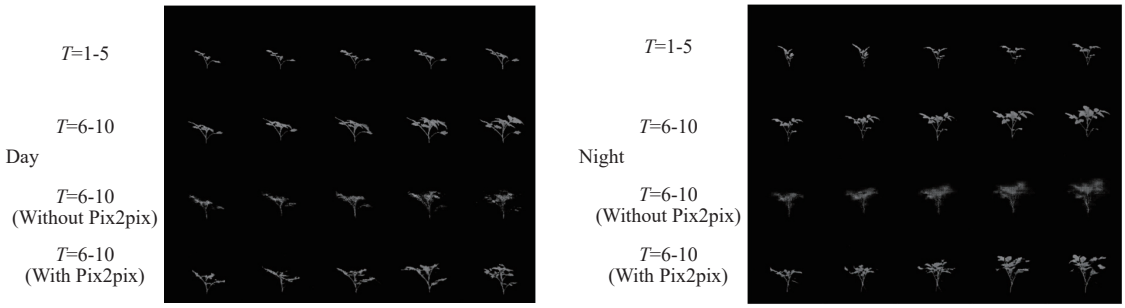


Figure 8 Network structure diagram of the proposed method

3 Results and discussion

3.1 Image similarity

This study's process is to input the image sequence of the first 5 d of the test dataset, and output the image sequence of the last 5 d through the proposed method. Figure 9 shows the predicted images obtained when inputting daytime images and nighttime images. Obvious nyctinastic movement can be seen from the comparison between the two sets of images. The first row contains the first 5-d images of this tomato seedling, which were used as input. The second row is the real growth image of this tomato seedling. The third row is the prediction result without using the pix2pix unit, and the fourth row is the final images predicted by the method proposed in this paper. Although the image without the pix2pix unit resulted in the gradual growth of the seedling, the generated images were relatively blurry, and biological information could not be clearly obtained from them. However, the method presented in this paper could significantly improve the quality of the generated images, and each image could be recognized as a seedling by a human.



The 1st row is the input, which is the image of the previous 5 d. The 2nd row is the real growth image. The 3rd row is the prediction without pix2pix. The 4th row is the prediction result combined with pix2pix.

Figure 9 Growth prediction example

To measure the similarity of the predicted image and ground truth, the mean square error (MSE) and structural similarity (SSIM) were used. MSE aims to evaluate the squared difference between pixels[33] and is an objectively quantified error. The formula for MSE

is as follows:

$$\mathrm{MSE}=\frac{1}{M\cdot N}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}[I(i,j)-K(i,j)]^2 \tag{9}$$

where, I represents the ground truth image; K represents the predicted image; M is the pixel size of the image height; N is the pixel size of the image width; i and j represent the position of the pixel along the x- and y-axes in the image coordinate system, respectively; and $I(i,j)$ and $K(i,j)$ represent the binary number of pixels at (i,j) in the ground truth image and predicted image, respectively. If the image size is 256, the MSE ranges from 0 to 65536. When the MSE is equal to 0, the two images can be considered the same.

Unlike MSE, SSIM considers the brightness, contrast, and overall structure of images to evaluate the similarity between images[34]. The formula for SSIM is as follows:

$$\mathrm{SSIM}=\frac{(2\mu_I\mu_K+C_1)\cdot(2\sigma_{IK}+C_2)}{(\mu_I^2+\mu_K^2+C_1)\cdot(\sigma_I^2+\sigma_K^2+C_2)} \tag{10}$$

where, I represents the ground truth image; K represents the predicted image; and μ represents the luminance information calculated by averaging all pixel values; σ_I and σ_K represent the standard deviations of the luminance in I and K, respectively, quantifying the contrast of an image; σ_{IK} represents the covariance of pixel values between I and K, reflecting the structural similarity and correlation; C is a constant used to prevent division by zero, $C_1=(K_1L)^2$, $C_2=(K_2L)^2$; K_1 and K_2 are traditionally set to 0.01 and 0.03, respectively; and L represents the dynamic range of the pixel values (for grayscale images, $L=255$). SSIM ranges from 0 to 1, while when SSIM is equal to 1, two images can be considered the same.

Figure 10 shows the MSE and SSIM values at each time stamp. As with other forecasting tasks, errors accumulate over time. Within the predicted 5 d, the MSE increases from 148.27 to 396.69 without GAN and from 252.11 to 582.53 with GAN. The SSIM decreases from 0.93 to 0.87 without using GAN, while after using GAN, the value decreases from 0.93 to 0.85. When GAN was used, the generated image was closer to the real image. To display the similarity more vividly, two sets of image examples with different similarity values were selected in Figure 11, where the left side of each group is a real image and the right side is a predicted image. The left picture is a group with higher similarity, an MSE of 173.79 and an SSIM of 0.94, and the right picture is a group with lower similarity, an MSE of 532.72 and an SSIM of 0.88.

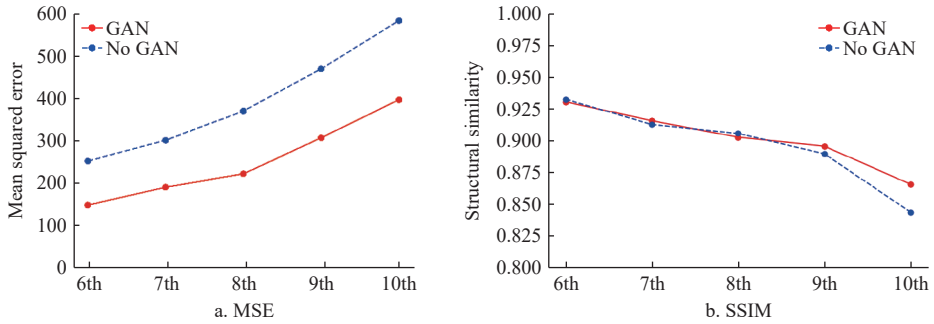


Figure 10 Similarity index between predicted images and real images in the next 5 d

MSE=173.79, SSIM=0.94
a. High similarity, better matching of branch direction and leaf area

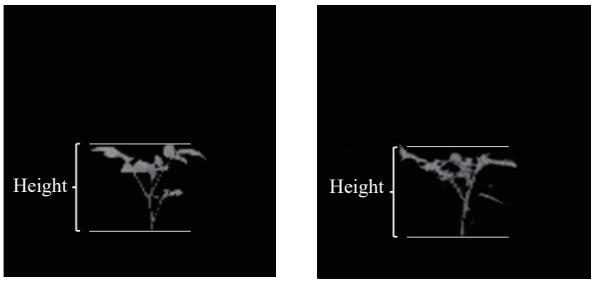
MSE=532.72, SSIM=0.88
b. Low similarity, branch direction and leafarea are difficult to match with those of the real image

Figure 11 Examples of images with different similarity values

For the growth of seedlings, from the perspective of image sequences, the morphological changes of images not only depend on the growth time of seedlings but also depend on the mutual constraints between various parts. In reality, the variety of seedlings and environmental factors also affects the morphological changes of seedlings. The method proposed in this paper essentially realizes the prediction of the overall morphological changes of seedlings, as well as the prediction of height and plant area, but the generated images still have some errors. In terms of image quality, if predictions are made from only one angle, as the seedling grows, some parts of the leaves will be revealed, and some parts will be occluded. In addition, different parts of the seedling and different leaf angles respond to different trends. These factors can cause random and unexpected effects on forecast accuracy. Therefore, to further improve the accuracy, the idea of image segmentation can be combined to distinguish different parts and calculate the changes in each part. Furthermore, the time interval of the image sequence in the dataset can be appropriately shortened.

3.2 Biological information

In addition to calculating the image similarity, it is equally important to extract biological information from the predicted images. The plant area and the height were selected as parameters and compared the error between these parameters of the predicted image and the real image to measure the effectiveness of the method proposed in this paper. As shown in Figure 12, ImageJ software was used to measure the pixel value of the canopy from the bottom as the height, and counted the white pixel value of the image as the plant area, and converted it into the actual height and area according to the 1 pixel value = 0.94 mm calibrated in the preprocessing process.



a. True image b. Predicted image

Figure 12 Schematic diagram of biological information

Figure 13 shows the daily changes in the area and height of the seedlings, in which the red curves are the true value and contain the data of 10 d. The predicted values start from the 6th day and are represented by the blue curve. Table 1 shows the error between the true value and the value obtained from the predicted images for each day. The relative errors of the plant area were 1.7, 1.4, 1.5, 0.9, and 3.2 cm^2, and the relative prediction errors of the height were 1.7, 1.9, 4.6, 6.9, and 4.5 mm over the 5 d.

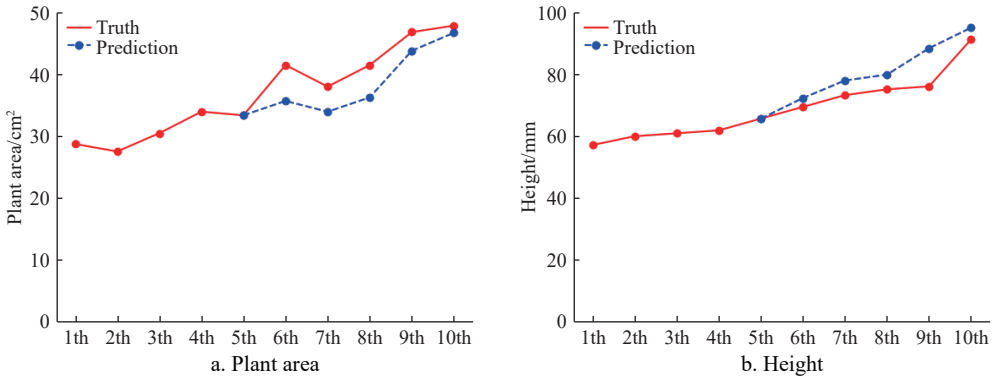


Figure 13 Comparison of biological information extracted from predicted and real images

Table 1 Relative error of biological information extracted from predicted and real images

	6th	7th	8th	9th	10th
Plant area/cm^2	1.7	1.4	1.5	0.9	3.2
Height/mm	1.7	1.9	4.6	6.9	4.5

From Figure 13 and Table 1, it can be found that the predicted values of plant area and height can follow the changes of the real values well, but at the same time, it can be also observed that the predicted plant area value is always smaller than the actual value, while the predicted plant height value is larger than the actual measurement value. These trends may be due to the limitations of the model and data representation. For underestimation of plant area, the main reason may be that the model mainly considers the changes in image edge features and structural characteristics rather than accurate area calculations, and as the plant grows, the leaves will curl and extend, which is difficult to learn from images of a single angle. Regarding the overestimation of plant height, the model may be biased towards vertical structures, such as stems, because they have strong and easily recognizable features in the image. This may cause the model to overemphasize these structures when predicting height. This leads to over-prediction.

To solve these problems, two aspects can be considered. 1) From the perspective of output values, after verification with larger-scale data, introducing adjustment factors based on average underestimation or overestimation can reduce errors. 2) From the perspective of improving generative neural networks, combining, for example, 3D structural data or multi-angle images to learn more comprehensive plant growth characteristics can reduce errors.

4 Conclusions

In this study, a 10-d growth image sequence dataset of 24 tomato seedlings at eight angles was obtained, and a deep learning approach based on causal LSTM, GHU, and pix2pix for predicting future 5-day images from the previous 5 d was proposed. This method of predicting future growth images based on historical seedling image sequences has the potential to become a basis for plant factories to adjust production expectations and management strategies. From the evaluation of imaging quality, image similarity, and biomass tracking performance, the following conclusions can be drawn:

(1) Compared with the traditional method of using the LSTM module to form the direct prediction of the recurrent neural network, in the output stage, the use of the GAN module can effectively improve the image quality.

(2) The image similarity gradually accumulates errors with the increasing number of prediction days. The MSE changed from 148.27 on day 6 to 346.69 on day 10, while the SSIM went from 0.92 on day 6 to 0.857 on day 10.

(3) The plant area and plant height extracted from the predicted image showed the same trend as the real values; the relative prediction errors of the plant area were 1.7, 1.4, 1.5, 0.9, and 3.2 cm^2; and the relative prediction errors of the height were 1.7, 1.9, 4.6, 6.9, and 4.5 mm over the 5 d.

Acknowledgement

Thanks to Kobuna Yukihiro for providing tomato seedlings and planting methods for this study.

[References]

[1] Higashide T, Yasuba K I, Suzuki K, Nakano A, Ohmori H. Yield of Japanese tomato cultivars has been hampered by a breeding focus on flavor. HortScience Horts, 2012; 47: 1408–1411.

[2] Luna-Maldonado A I, Vidales-Contreras J A, Rodríguez-Fuentes H. Editorial: Advances and trends in development of plant factories. Front. Plant Sci., 2016; 7: 1848.

[3] Yu W Z, Zheng J F, Wang Y L, Ji F, Zhu B Y. Adjusting the nutrient solution formula based on growth stages to promote the yield and quality of strawberry in greenhouse. Int J Agric & Biol Eng, 2023; 16(2): 57–64.

[4] Yin J, Liu X Y, Miao Y L, Gao Y, Qiu R C, Zhang M, et al. Measurement and prediction of tomato canopy apparent photosynthetic rate. Int J Agric & Biol Eng, 2019; 12(5): 156–161.

[5] Moriyuki S, Fukuda H. High-throughput growth prediction for *Lactuca sativa* L. seedlings using chlorophyll fluorescence in a plant factory with artificial lighting . Frontiers in plant science, 2016; 7: 394.

[6] Kocian A, Massa D, Cannazzaro S, Incrocci L, Di Lonardo S, Milazzo P, Chessa S. Dynamic Bayesian network for crop growth prediction in greenhouses. Comput. Electron. Agric., 2020; 169: 105167.

[7] Wang X, Yang Y, Zhao X, Huang M, Zhu Q B. Integrating field images and microclimate data to realize multi-day ahead forecasting of maize crop coverage using CNN-LSTM. Int J Agric & Biol Eng, 2023; 16(2): 199–206.

[8] Yamamoto K, Guo W, Ninomiya S. Node detection and internode length estimation of tomato seedlings based on image analysis and machine learning. Sensors, 2016; 16: 1044.

[9] Mohanty S P, Hughes D P, Salathé M. Using deep learning for image-based plant disease detection. Front. Plant Sci., 2016; 7: 1419.

[10] Agarwal M, Gupta S K, Biswas K K. Development of efficient CNN model

for tomato crop disease identification. Sustain. Comput. Inform. Syst., 2020; 28: 100407.

[11] Yuan Y, Chen L, Ren Y C, Wang S M, Li Y. Impact of dataset on the study of crop disease image recognition. Int J Agric & Biol Eng, 2022; 15(5): 181–186.

[12] Baheti H, Thakare A, Bhople Y, Darekar S, Dodmani O. Machine learning algorithm for detection and classification of tomato plant leaf disease. In: 2022 IEEE 7th International Conference for Convergence in Technology (I2CT). IEEE, Mumbai, India, 2022; pp.1–7.

[13] Liu J, Wang X. Tomato diseases and pests detection based on improved yolo V3 convolutional neural network. Front. Plant Sci., 2020; 11: 898.

[14] Rustia D J A, Chao J J, Chiu L Y, Wu Y F, Chung J Y, Hsu J C, et al. Automatic greenhouse insect pest detection and recognition based on a cascaded deep learning classification method. J. Appl. Entomol., 2021; 145: 206–222.

[15] Jin X, Zhu X W, Ji J T, Li M Y, Xie X L, Zhao B. Online diagnosis platform for tomato seedling diseases in greenhouse production. Int J Agric & Biol Eng, 2024; 17(1): 80–89.

[16] Li P, Zhong N, Dong W, Zhang M, Yang D T. Identification of tomato leaf diseases using convolutional neural network with multi-scale and feature reuse. Int J Agric & Biol Eng, 2023; 16(6): 226–235.

[17] Jeong Y, Kim S, Yi K. Surround vehicle motion prediction using LSTM-RNN for motion planning of autonomous vehicles at multi-lane turn intersections. IEEE Open J. Intell. Transp. Syst., 2020; 1: 2–14. DOI: 10.1109/OJITS. 2020.2965969.

[18] Tang Y, Ma L, Liu W, Zheng W S. Long-term human motion prediction by modeling motion context and enhancing motion dynamic. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence. AAAI Press, Stockholm, Sweden, 2018; pp.935–941. DOI: 10.48550/arXiv. 1805.02513.

[19] Karevan, Z, Suykens J A K. Transductive LSTM for time-series prediction: An application to weather forecasting. Neural Networks, 2020; 125: 1–9.

[20] Uchiyama H, Sakurai S, Mishima M, Arita D, Okayasu T, Shimada A, Taniguchi R I. An easy-to-setup 3D phenotyping platform for Komatsuna dataset. In: Proceedings of the IEEE international conference on computer vision workshops, 2017; pp.2038–2045.

[21] Scharr H, Pridmore T, Tsaftaris S A. Computer vision problems in plant phenotyping, CVPPP 2017: Introduction to the CVPPP 2017 workshop papers. In: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW). IEEE, Venice, Italy, 2017; pp.2020–2021. DOI: 10.1109/ICCVW.2017.236

[22] Sakurai S, Uchiyama H, Shimada A, Taniguchi R I. Plant growth prediction using convolutional LSTM. In: Kerren, A., Hurter, C., Braz, J. (Eds.), VISIGRAPP 2019 - Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. SciTePress, Portugal, 2019; pp.105–113. DOI: 10.5220/0007404901050113.

[23] Yasrab R, Zhang J, Smyth P, Pound M P. Predicting plant growth from time-series data using deep learning. Remote Sensing, 2021; 13: 331.

[24] Kim T, Lee S H, Kim J O. A novel shape based plant growth prediction algorithm using deep learning and spatial transformation. IEEE Access, 2022; 10: 37731–37742.

[25] Ueda M, Asano M, Sawai Y, Yamamura S. Leaf-movement factors of nyctinastic plant, *Phyllanthus urinaria* L., the universal mechanism for the regulation of nyctinastic leaf-movement. Tetrahedron, 1999; 55: 5781–5792.

[26] Li Y, Zhu Z, Kong D, Han H, Zhao Y. EA-LSTM: Evolutionary attention-based LSTM for time series prediction. Knowl-Based Syst, 2019; 181: 104785.

[27] Liu J, Xu L, Chen N. A spatiotemporal deep learning model ST-LSTM-SA for hourly rainfall forecasting using radar echo images. J. Hydrol., 2022; 609: 127748.

[28] Fu R, Zhang Z, Li L. Using LSTM and GRU neural network methods for traffic flow prediction. In: 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), IEEE, Wuhan, China, 2016; pp.324-328. DOI: 10.1109/YAC. 2016.7804912.

[29] Oprea S, Martinez-Gonzalez P, Garcia-Garcia A, Castro-Vargas J A, Orts-Escolano S, Garcia-Rodriguez J, Argyros A. A review on deep learning techniques for video prediction. IEEE Trans. Pattern Anal. Mach. Intell., 2022; 44: 2806–2826.

[30] Wang Y, Gao Z, Long M, Wang J, Yu P S. PredRNN++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. In: Dy J, Krause A. (Eds.) Proceedings of the 35th International Conference on Machine Learning. PMLR, Proceedings of Machine Learning Research, 2018; pp.5123–5132.

[31] Srivastava R K, Greff K, Schmidhuber J. Training very deep networks. In: Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2. MIT Press, Montreal, Canada, 2015; pp.2377–2385. DOI: 10.48550/ arXiv.1507.06228.

[32] Isola P, Zhu J Y, Zhou T, Efros A A. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017; pp.1125–1134. DOI: 10.1109/CVPR.2017.632

[33] Palubinskas G. Image similarity/distance measures: What is really behind MSE and SSIM? Int. J. Image Data Fusion, 2017; 8: 32–53.

[34] Bakurov I, Buzzelli M, Schettini R, Castelli M, Vanneschi L. Structural similarity index (SSIM) revisited: A data-driven approach. Expert Syst. Appl., 2022; 189: 116087.