Deep Neural Network-Based System for Autonomous Navigation in Paddy Field

SHYAM P. ADHIKARI[1], **GOOKHWAN KIM**[2], **AND HYONGSUK KIM**[1,3], **(Senior Member, IEEE)**
[1]Division of Electronics and Information Engineering, Chonbuk National University, Jeonju 54896, South Korea
[2]Division of Smart Farm Development, Rural Development Administration, Jeonju 54875, South Korea
[3]Core Research Institute of Intelligent Robots, Chonbuk National University, Jeonju 54896, South Korea

Corresponding author: Hyongsuk Kim (hskim@jbnu.ac.kr)

ABSTRACT This paper presents a novel vision based approach for detecting rows of crop in paddy field. The precise detection of crop row enables a farm-tractor to autonomously navigate the field for successful inter-row weeding. While prior works on crop row detection rely primarily on various image based features, a deep neural network based approach for learning semantic graphics to directly extract the crop rows from an input image is used in this work. A deep convolutional encoder decoder network is trained to detect the crop lines using semantic graphics. The detected crop lines are then used to derive control signal for steering the tractor autonomously in the field. The results demonstrate that the proposed method is able to detect the rows of paddy accurately and enable the tractor to navigate autonomously along the crop rows even with a simple proportional only controller.

INDEX TERMS Convolutional encoder-decoder network, crop line detection, semantic graphics, vision based control.

I. INTRODUCTION

The increase in global population has led to an increase in the demand of agricultural food products to feed them. With limited availability of resources, ramping up food production to meet the ever-increasing demand is a challenging task. Researchers, engineers and farmers have come up with several ingenious solutions like better farming techniques, precision farming, farm automation etc. to overcome these challenges. Farming and most of the associated tasks are highly labor-intensive. Though human population has been increasing there has been a constant decline in the share of labor force working in agriculture [1] due to the labor-intensive and repetitive nature of the work. While much of the agricultural tasks have already been mechanized resulting in reduced human labor, researchers have been working towards reducing the reliance on human labor with automation and keeping it to minimal.

With the advancement in robotics, robots have been widely used in the farm and have been crucial in improving crop

The associate editor coordinating the review of this manuscript and approving it for publication was Mostafa Rahimi Azghadi.

productivity and reducing human labor. Farm automation with robots is a promising area that has the potential to overcome the challenges facing agriculture while keeping human involvement to minimal. Accurate machine guidance is one of the crucial factors determining the success of autonomous farm robots.

The recent advancements in deep neural networks (DNNs) have made profound impact in different areas like autonomous navigation [2], computer-assisted diagnosis [3], [4], speech recognition [5] etc. DNNs have also emerged as a promising technique with potential to take automation in agriculture to the next level. DNNs have been used extensively to automate different agricultural tasks such as plant recognition [6], crop type classification [7], plant disease classification [8], weed identification [9], [10] and land cover classification [7], [11] etc. The semi-constrained nature of agricultural farm makes it comparatively easier for the adoption of DNN, however it has its own challenges. The similar shape, texture and color of crops and weeds makes it difficult for the DNN system to discriminate them properly, resulting in reduced classification accuracy. There is severe overlap between the crops and weeds in the field which results in

occlusion. Occlusion is a challenging phenomenon for vision based system which leads to reduced performance.

In this paper we propose to use a DNN based system for detecting the rows of crop in row-transplanted paddy field using semantic graphics, and demonstrate that the detected crop rows can be used to guide a tractor to navigate autonomously in the field.

FIGURE 1. Convolutional encoder-decoder network based detection of crop lines using semantic graphics.

II. RELATED RESEARCH

Recently, autonomous agricultural robots have been widely adopted to increase crop productivity and improve labor efficiency. Navigation systems are a crucial part of such autonomous robots. Different solutions to the navigation problem are discussed in the literature, however computer vision based systems are more popular due to low cost, easy handling and wide availability of vision based sensors. Accurate crop row detection to guide the robot is one of the most important problems for computer vision based navigation of agricultural robots.

Previous works on detecting crop rows using vision based system primarily detect the position of the crops using different handcrafted features. Sogaard and Olsen [12] computed an indicator for living plant tissue by utilizing the color channels of a RGB image and estimated the center line of crops from the distribution of the living tissue indicator. Bakker et al. [13] used the living tissue indicator and Hough transform to extract straight line representing the rows of crops. Montalvo et al. [14] proposed to use vegetation index derived from the RGB channels to segment the image. Prior knowledge about the arrangement of crops like the number of rows, expected location of each crop row and approximate region of interest were utilized to extract the crop lines using linear regression. Choi et al. [15] utilized morphological characteristic of leaves converging towards the direction of central stem to estimate the central region of rice plant, and used Hough transform and regression to extract the crop line. Jiang et al. [16] used living tissue index to segment the image and extracted feature points from the binary image using a sliding window approach. Hough transform was then used to extract the candidate crop lines and vanishing point was utilized to remove the false crop lines.

Methods based on manual features work well under controlled conditions, however they can fail to work in real farm conditions. Methods based on color index can only work well in the absence of inter-row weed as the vegetation index or living tissue index of weeds is similar to that of crop. The presence of weeds and challenging natural conditions like shades or light reflection affects the extraction of binary morphological features, which ultimately leads to inaccurate crop line extraction. Guerrero et al. [17] recognized the difficulty in discrimination of crops and weeds by applying image segmentation techniques based on the RGB spectral components and utilized geometric constraints to locate crop rows with increased accuracy.
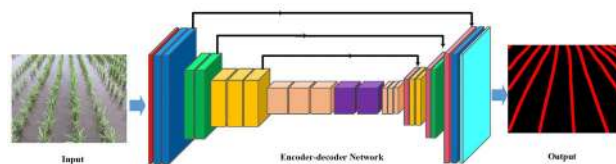
Recent advancements in neural network have shown that features learned automatically by convolutional neural networks are more robust and efficient than hand-engineered features. Methods based on CNNs have produced state-of-the-art results in different computer vision and pattern recognition problems like object detection and classification, and semantic segmentation [18]–[20]. While [21], [22] used CNN-based semantic segmentation to discriminate crops, weeds and background, the actual lines of crop are not extracted. In our previous work we presented that CNN can directly be trained to learn the concept of a crop line using "semantic graphics" [23], as shown in Fig. 1.

In this current work we extend the concept of learning semantic graphics to extract crop rows and use the rows to guide a farm tractor autonomously in a paddy field.

III. PROPOSED METHOD

In this work we take the use case of autonomous navigation of tractor in row transplanted paddy field. The successful navigation of the tractor is achieved in three steps. First a convolutional encoder-decoder network is trained using semantic graphics to detect the rows of crop. The position of the tractor wheel relative to the detected rows of paddy is then extracted using template matching. Finally, the relative positions are used to compute the steering angle and a simple proportionate control algorithm is used to drive the tractor autonomously to follow the rows without damaging the crop.

A. CROP ROW DETECTION USING SEMANTIC GRAPHIC

Semantic graphic [23] is a process of annotating an image with simple graphical sketch for easy learning by neural network. The concept of semantic graphic was introduced to simplify the annotation of images and make it less labor intensive than semantic segmentation for complex scenes. However, semantic graphics is different from semantic segmentation as it strives to annotate higher order concepts rather than semantic regions. Pixels belonging to the same semantic region can be assigned a different target category in semantic graphic. An example of annotating the rows of crops using semantic graphics is shown in Fig. 2.

Given an image of paddy field, the rows of paddy are annotated with few-pixel thick lines. The line does not necessarily cover the whole width of the row; however, it captures the human understanding about the row of crop. A convolutional encoder-decoder network is then trained to learn a mapping from the image to the crop lines, as shown in Fig. 1.
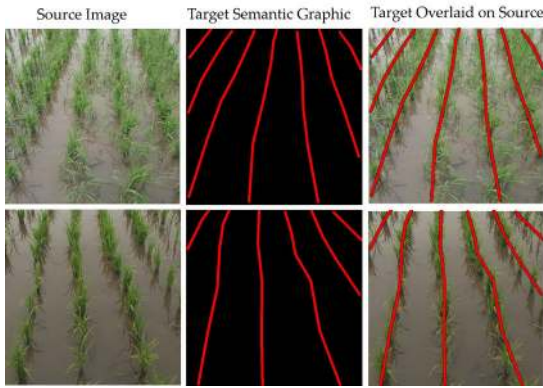
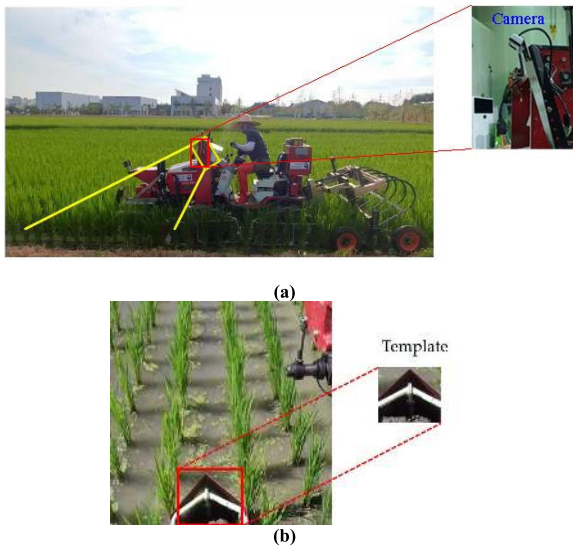**FIGURE 2.** Semantic graphic annotation of paddy lines.

FIGURE 3. (a) The experimental farm tractor with the camera mounted directly above the left wheel, (b) Field of view of the camera mounted on the tractor. A template is manually extracted from a reference image and compared against the test image by sliding it over the test image RoI to identify the matching area using normalized cross-correlation.

B. ROBOTIC PLATFORM AND CONTROL

An autonomous farm tractor, shown in Fig. 3, is used as an experimental platform for this study. The tractor consists of an onboard computer where the trained neural network model is loaded for inference and other auxiliary processing is done. A front facing camera is mounted above the left wheel of the tractor, where the field of view includes a portion of the wheel cover at the base of the image, as shown in Fig.3 (b).

1) WHEEL POSITION DETERMINATION

The placement of the camera and the presence of wheel cover with contrasting color allows us to use a simple template matching algorithm to determine the position of the wheel in the image. A portion of the wheel cover, as shown in Figure 3(b), is pre-stored in the system as a template and its location in the input image is found using normalized cross
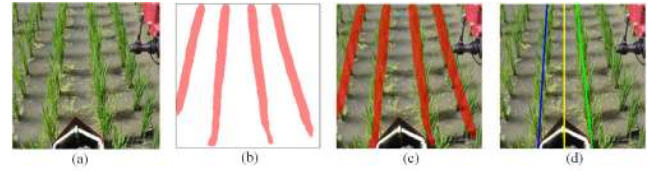

FIGURE 4. (a) Field of view of the camera mounted on the tractor, (b) the crop rows detected using the convolutional encoder-decoder network, (c) detected crop rows overlaid in the source image, (d) yellow line is the detected center line of the wheel, the blue and green lines are the extracted left and right crop rows of the host row, respectively.

TABLE 1. Sliding window algorithm for paddy line extraction.

Let,

a. (x_s, y_s) is the starting point

b. w and h are the height and width of window, respectively

c. I_b is the binarized output from neural network

1. Compute the window (W) around the starting point, the top-left and bottom right coordinates of W are $(x_s - w/2, y_s - h)$ and $(x_s + w/2, y_s)$

2. Find the index of all the paddy line pixels, *pixels_within_window (pww)*, that lie within this window. i.e. $pww = \{x, y\} \ni I_b(x, y) = 1, \; \forall (x, y) \in W$

3. Add *pww* to the global list of paddy pixels, *paddy_pixels (pp)*

4. Compute the mean of the column co-ordinates (x_{avg}) of *ppw* to obtain starting point for the subsequent window. i.e. $(x_s, y_s) = (x_{avg}, y_s - h)$

5. Repeat steps (1-4) for all valid windows

6. Fit a straight line to *paddy_pixels (pp)* to obtain the required paddy line

correlation, given as

$$R(x, y) = \frac{\sum_{x', y'} \left(T(x', y') \cdot I(x + x', y + y') \right)}{\sqrt{\sum_{x', y'} T(x', y')^2 \cdot \sum_{x', y'} I(x + x', y + y')^2}} \quad (1)$$

where, T denotes the template, and I the test image. $R(x,y)$ is computed at each pixel position and indicates the degree of similarity of the test patch with the template. The pixel position with the highest coefficient gives an estimate of the position of the wheel. As the camera is fixed to the tractor, the RoI for template matching is restricted to the lower part of the image to reduce computation time.

2) HOST ROW DETECTION

The encoder-decoder network outputs semantic lines for every visible row of paddy, as shown in Fig. 4. However, for practical purpose it is enough to detect only the host rows of paddy that lie on either side of the wheel, as shown in Fig. 4(d), to guide the tractor along the rows.

The search area is restricted to the lower half of the image to find an initial estimate of the starting points of the two host rows. A histogram of the detected paddy line is then computed, and the maxima of the histogram on the left half and the right half of the image gives the initial estimate of the starting position of the two rows. Starting from these
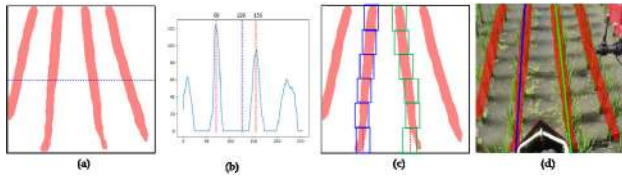
FIGURE 5. (a) The RoI is restricted to lower half of the image to compute the starting points, (b) Histogram of paddy line pixels is computed and two maxima, one on each side from the center column gives the starting point for two host rows, (c) Starting from the base of the image, sliding window algorithm is used to find pixels belonging to the corresponding paddy line, (d) Straight lines are fit to the corresponding paddy pixels to extract the two host rows.

two initial positions, a sliding window algorithm (window width: 20 pixels, height: *image_height*/6 pixels) is employed to extract the rows of paddy. The details of the sliding window algorithm are given in Table 1.

Due to the wide wheel cover of the tractor the nearby paddy plants, i.e. plants appearing near base of the image, are temporarily pushed outwards from their position. This leads to incorrect alignment of the detected lines. To avoid any errors that may be introduced by this phenomenon, the *pixels_within_window* (*pww*) of the initial window at the base of the image are not used for the fitting the final straight line. Moreover, as the long-range information of paddy row is not utilized for controlling the tractor, *pww* at the top of the image are also excluded for computing the line. The overall process of finding the two paddy rows to guide the tractor is presented in Fig. 5.

3) GUIDANCE OF TRACTOR

Once the position of the wheel and the two host rows at either side are determined, their relative positions are used to generate steering commands for the tractor to move autonomously between the rows. In this work a simple proportional control strategy is used to generate the steering commands for the tractor. As the current wheel center point p_{whl} needs to follow p_{mid}, the actual center point of the crop row, the control signal c uses a position control method that is proportional to the position difference between p_{whl} and p_{mid}. Therefore,

$$c = \alpha\,(p_{mid} - p_{whl}) \qquad (2)$$

where α is the control parameter. p_{mid} is computed from the position of the detected left (p_L) and right (p_R) rows,

$$p_{mid} = (p_R + p_L)\big/2 \qquad (3)$$

From equation (1) and (2),

$$c = \alpha\,(d_R - d_L) \qquad (4)$$

where, $d_R = p_R - p_{whl}$ and $d_L = p_{whl} - p_L$ are distances from the right and left crop row to the center of the wheel, respectively, as shown in Fig. 6. d_R and d_L are computed using the pixel distance in the image from the center of the wheel to the left and right crop rows, respectively.
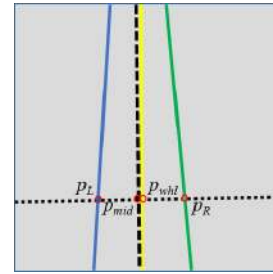
FIGURE 6. The proportionate control signal is derived from the relative position of the tractor wheel from the right and left host rows of paddy.
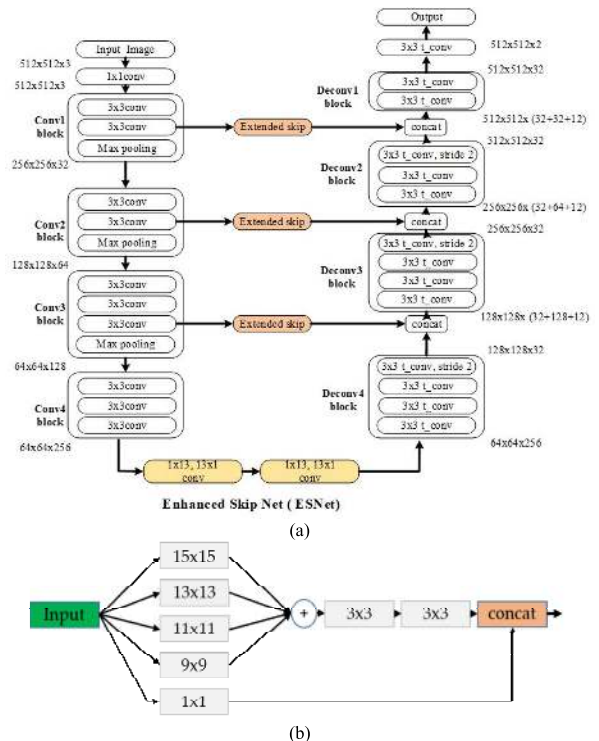

FIGURE 7. (a) ESNet: Extended Skip Net [23] for learning semantic graphics, (b) Extended Skip module consists of multi-scale convolutions to capture features at different scales.

IV. EXPERIMENTS AND RESULTS

A. DATASET AND NETWORK

The paddy line dataset [23] was used to train the neural network for detecting paddy lines. The dataset consists of 350 images of row transplanted paddy field captured while walking in between the rows of the crop. The rows of paddy were annotated with few-pixels thick "semantic lines" as shown in Figure 2. The images were down sampled to 600×600 pixels and augmented using random rotation and vertical mirroring. Finally, random crops of 512×512 were used as input to the neural network. The extended skip network was adopted in this study due to its superior performance in detecting paddy lines over other networks as demonstrated in [23]. The detailed network is reproduced in Fig. 7 for ready reference.

The network was trained from scratch on the whole paddy-line dataset using cross-entropy loss. Xavier initialization was
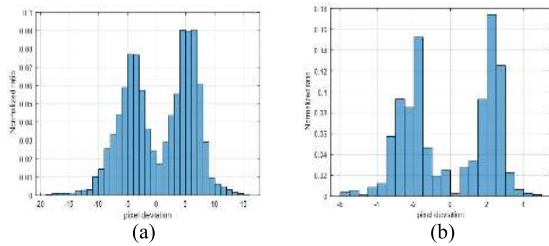
FIGURE 8. Distribution of pixel deviations (a) overall pixel deviation (*pd*), (b) *pd* for near field (lower 4 th of the image).

used for initializing and the network was trained on batch size of 5 using Adam with exponential decaying learning rate for 100 epochs. Class frequency based weighting was used to mitigate the class imbalance between the paddy-line and background classes. The network was trained with Tensorflow using Titan X GPU.

B. EVALUATION METRIC

The performance of the trained model was evaluated on the real field data by evaluating the mean pixel deviation (*mpd*) of the extracted host rows from the ground truth. If (x_p, y) is a point on the predicted line and (x_g, y) is its corresponding point on the ground truth line, the row wise pixel deviation (*pd*) and *mpd* are computed as,

$$pd = |x_p - x_g|, \qquad (5)$$

and

$$mpd = \frac{1}{N} \sum_{i=1}^{N} (pd)^i \qquad (6)$$

respectively, where N is the total number of row pixels considered in the test set.

C. RESULTS ON FIELD DATA

The weights of the trained network were stored in an industrial automation computer (Intel CoreTM i7 2.5GHz (Four cores), 32 bit, 4GB RAM) onboard the farm tractor. The RoI of the image captured by the camera mounted on the tractor was reduced, and an image of size 256×192 was used for implementing the vision based control system. The overall experiment was conducted for a single lap (45m) of the tractor in an experimental row transplanted paddy field where the inter-row distance is 30cm and separation between plants in a row is 20 cm. In this study only the near field measurement of the line was taken to compute d_R and d_L. The *mpd* for near field (i.e., the lower 1/4th of the image near the tractor) is 6.889 pixels which is equivalent to an average error 2.2 cm in real world. The *mpd* computed over the entire length of the row is 5.137 pixels. The overall and near field only distribution of pixel deviations are presented in Fig. 8(a) and Fig. 8(b), respectively. Some qualitative results of the detected paddy lines are presented in Fig. 9.

The d_R and d_L values were computed for each frame and fed as control signals to another onboard industrial automation computer. The machine control and monitoring system
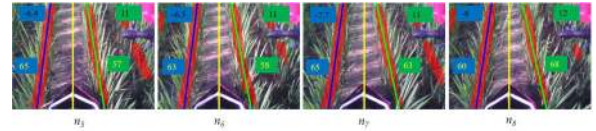
FIGURE 9. Qualitative results of paddy line detection for a section of the test field. The bold thick red lines are the binarized outputs of the neural network representing the rows of paddy, blue and green lines are the extracted left and right guidance lines, respectively, and the yellow line represents the position of the tractor wheel. The angle and pixel distance are also indicated near the respective guidance lines.
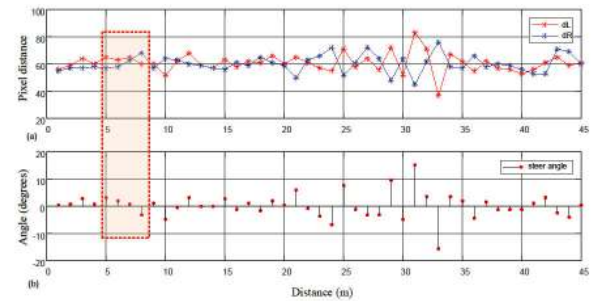
FIGURE 10. a) The computed left and right distance d_R and d_L and (b) the steering angles at different positions for a single lap of 45m in the test field.
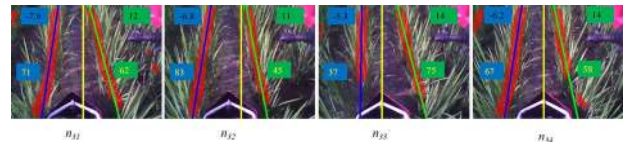
FIGURE 11. Failure of the template matching algorithm to correctly determine the position of the wheel can lead to incorrect row following. Continuous error in wheel position determination can lead to row jumping by the tractor and damage to crops.

was implemented in this computer using LabVIEW. d_R and d_L values were computed using the pixel distance measured in the row shown in Fig. 6. The d_R and d_L values computed while the tractor moved autonomously by following the rows of paddy detected using the neural network for a lap in the field are shown Fig. 10(a). The corresponding control signal applied when the tractor traversed 45m in the field is shown in Figure 10(b).

Whenever the tractor is biased towards either the left or the right side of the row, a control in the opposite direction is applied and the direction of the tractor is corrected. The images in Fig. 9 correspond a section of the field marked by the box in Fig. 10.

V. DISCUSSION

The color based simple template matching system to determine the position of wheel is prone to errors. A wrong estimation of the wheel position results in wrong control signals being sent to the tractor, which may lead to the tractor not following the host rows and possibly jump to another row. Example of a sequence where the template matching algorithm failed to determine the correct position of the wheel is presented in Fig. 11.

From Fig. 11 we see that after two successive failures in detecting the correct position of the wheel at n_{31} and n_{32},

the tractor is at the verge of jumping the correct host row at n_{33}. However, at this moment (n_{33}) the correct wheel position is determined, and a large control signal is applied in the opposite direction, shown in Figure 10(b), to restore the trajectory of the wheel, as seen in n_{34}.

Temporal tracking of the template can be employed to reduce jitters in wheel position estimation and enhance the accuracy of the system. Wheel encoders can be incorporated in the tractor to measure the actual turning angle of the wheel and make the system more robust. The ability of the tractor to navigate the field by following the rows depends mainly on the accurate detection of the crop rows. The current system in trained on a limited dataset of 350 images, which does not cover the different scenarios that can occur in a field. However, the proposed system is robust to shadows, field of view, row spacing and age of the crop as long as the rows are distinguishable. Training the network on a larger dataset is expected to significantly increase the accuracy and robustness of the system. Temporal tracking of the detected paddy lines is further expected to increase the quality of line detection.

The current system uses data from a single camera placed over the left wheel. The control of the tractor can also be improved by detecting the host rows of another wheel by using an additional camera over the right wheel. Unlike the simple proportionate only controller used in this study, a more complex PID controller or a neural network based controller can mitigate the oscillations observed in Fig. 10 (b) and smoothen the trajectory of the tractor.

The current CPU based onboard system is limited in computation. Due to the multiple convolution operations involved in the encoder-decoder network, the paddy line detection system runs at 0.5 frames per second. Due to the low frame rate, the speed for autonomous navigation of tractor is also set at a low of 0.5 m/s. The current processing time and tractor speed are compatible and practical for our experiments using a proportionate only controller in a swampy rice field. However, any additional improvement as mentioned above needs additional computation power. Considerably fast inference time can be expected if the inference is carried out in an embedded AI computing device. From our preliminary experiments, it was observed that the current system can run at 5 *fps* by replacing the onboard industrial computer implementing the line-detection sub-system with the NVIDIA Jetson TX2.

VI. CONCLUSION

A deep neural network based system for autonomous navigation of tractor in row transplanted paddy field was presented in this study. A deep convolutional neural network was used to learn the concept of rows of paddy using semantic graphic. The detected rows were then used to derive control signal for autonomous navigation of a tractor in between the crop rows. Successful autonomous navigation of the tractor for a single lap of the test field was demonstrated using a simple proportional only controller. Agricultural robots that can navigate autonomously in the field will have an enormous social and economic impact with wide ranging applications like autonomous weeding, precision spraying of nutrients, pesticides, herbicides etc. Autonomous farm systems can substantially reduce the drudgery of farmers while increasing the efficiency, productivity and quality of crops.

REFERENCES

[1] M. Roser. (2020). Employment in Agriculture. Our World in Data. Accessed: Mar. 31, 2020. [Online]. Available: https://ourworldindata.org/employment-in-agriculture

[2] B. Huval, T. Wang, S. Tandon, J. Kiske, W. Song, J. Pazhayampallil, M. Andriluka, P. Rajpurkar, T. Migimatsu, R. Cheng-Yue, F. Mujica, A. Coates, and A. Y. Ng, "An empirical evaluation of deep learning on highway driving," 2015, *arXiv:1504.01716*. [Online]. Available: https://arxiv.org/abs/1504.01716

[3] H. Yang, J.-Y. Kim, H. Kim, and S. P. Adhikari, "Guided soft attention network for classification of breast cancer histopathology images," *IEEE Trans. Med. Imag.*, early access, Oct. 17, 2019, doi: 10.1109/tmi.2019.2948026.

[4] P. Rajpurkar, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," Nov. 2017, *arXiv:1711.05225*. [Online]. Available: https://arxiv.org/abs/1711.05225

[5] D. Amodei, "Deep speech 2: End-to-end speech recognition in English and Mandarin," in *Proc. 33rd Int. Conf. Mach. Learn. (ICML)*, vol. 1, 2016, pp. 312–321.

[6] G. L. Grinblat, L. C. Uzal, M. G. Larese, and P. M. Granitto, "Deep learning for plant identification using vein morphological patterns," *Comput. Electron. Agricult.*, vol. 127, pp. 418–424, Sep. 2016, doi: 10.1016/j.compag.2016.07.003.

[7] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 778–782, May 2017, doi: 10.1109/LGRS.2017.2681128.

[8] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers Plant Sci.*, vol. 7, p. 1419, Sep. 2016, doi: 10.3389/fpls.2016.01419.

[9] M. Dyrmann, R. N. Jørgensen, and H. S. Midtiby, "RoboWeedSupport—Detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network," *Adv. Animal Biosci.*, vol. 8, no. 2, pp. 842–847, Jul. 2017, doi: 10.1017/s2040470017000206.

[10] M. Dyrmann, H. Karstoft, and H. S. Midtiby, "Plant species classification using deep convolutional neural network," *Biosyst. Eng.*, vol. 151, pp. 72–80, Nov. 2016, doi: 10.1016/j.biosystemseng.2016.08.024.

[11] D. Ienco, R. Gaetano, C. Dupaquier, and P. Maurel, "Land cover classification via multitemporal spatial data by deep recurrent neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1685–1689, Oct. 2017, doi: 10.1109/LGRS.2017.2728698.

[12] H. T. Søgaard and H. J. Olsen, "Determination of crop rows by image analysis without segmentation," *Comput. Electron. Agric.*, vol. 38, no. 2, pp. 141–158, 2003, doi: 10.1016/S0168-1699(02)00140-0.

[13] T. Bakker, "A vision based row detection system for sugar beet," *Comput. Electron. Agric.*, vol. 60, no. 1, pp. 87–95, Jan. 2008, doi: 10.1016/j.compag.2007.07.006.

[14] M. Montalvo, G. Pajares, J. M. Guerrero, J. Romeo, M. Guijarro, A. Ribeiro, J. J. Ruz, and J. M. Cruz, "Automatic detection of crop rows in maize fields with high weeds pressure," *Expert Syst. Appl.*, vol. 39, no. 15, pp. 11889–11897, Nov. 2012, doi: 10.1016/j.eswa.2012.02.117.

[15] K. H. Choi, S. K. Han, S. H. Han, K.-H. Park, K.-S. Kim, and S. Kim, "Morphology-based guidance line extraction for an autonomous weeding robot in paddy fields," *Comput. Electron. Agricult.*, vol. 113, pp. 266–274, Apr. 2015, doi: 10.1016/j.compag.2015.02.014.

[16] G. Jiang, X. Wang, Z. Wang, and H. Liu, "Wheat rows detection at the early growth stage based on Hough transform and vanishing point," *Comput. Electron. Agricult.*, vol. 123, pp. 211–223, Apr. 2016, doi: 10.1016/j.compag.2016.02.002.

[17] J. M. Guerrero, J. J. Ruz, and G. Pajares, "Crop rows and weeds detection in maize fields applying a computer vision system based on geometry," *Comput. Electron. Agricult.*, vol. 142, pp. 461–472, Nov. 2017, doi: 10.1016/j.compag.2017.09.028.

[18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[20] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988, doi: 10.1109/ICCV.2017.322.

[21] A. Milioto, P. Lottes, and C. Stachniss, "Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 2229–2235, doi: 10.1109/ICRA.2018.8460962.

[22] X. Ma, X. Deng, L. Qi, Y. Jiang, H. Li, Y. Wang, and X. Xing, "Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields," *PLoS ONE*, vol. 14, no. 4, Apr. 2019, Art. no. e0215676, doi: 10.1371/journal.pone.0215676.

[23] S. P. Adhikari, H. Yang, and H. Kim, "Learning semantic graphics using convolutional encoder-decoder network for autonomous weeding in paddy," *Frontiers Plant Sci.*, vol. 10, p. 1404, Oct. 2019, doi: 10.3389/fpls.2019.01404.

GOOKHWAN KIM received the Ph.D. degree in mechanical engineering from Kyung Hee University, South Korea, in 2010. He is currently a Researcher with the Rural Development Administration, National Academy of Agricultural Science, Jeonju, South Korea.

SHYAM P. ADHIKARI received the Ph.D. degree in electronics engineering from Chonbuk National University, Jeonju, South Korea, in 2013. He is currently a Postdoctoral Fellow with Chonbuk National University. His current research interests include memristors and deep learning for computer vision. He was a recipient of the Korea Research Fellowship, in 2015.

HYONGSUK KIM (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Missouri, Columbia, MO, USA, in 1992. From 2000 to 2002 and 2009 to 2010, he was a Visiting Scholar with the Nonlinear Electronics Laboratory, Electrical Engineering, and Computer Science Department, University of California at Berkeley, Berkeley, CA, USA. Since 1993, he has been a Professor with the Division of Electronics Engineering, Chonbuk National University, Jeonju, South Korea. His current research interests include memristors and its application to the implementation of neural networks.

• • •