Sensors and Materials, Vol. 32, No. 1 (2020) 159–170
MYU Tokyo

S & M 2093

Using Fully Convolutional Networks for Floor Area Detection

Cheng-Jian Lin,[1*] Yu-Chi Li,[1] and Chin-Ling Lee[2]

[1]Department of Computer Science and Information Engineering, National Chin-Yi University of Technology, Taichung City 41107, Taiwan, ROC
[2]Department of International Business, National Taichung University of Science and Technology, Taichung City 40401, Taiwan, ROC

(Received July 29, 2019; accepted October 8, 2019)

Keywords: image sensor, fully convolutional networks, floor area detection, fuzzy integral, image segmentation

Most mobile robots use visual images to obtain information about the surrounding environment and the nonlinear diffusion method to detect candidate areas of the floor, but they could not be applied to more complicated environments. In this study, a hybrid of fully convolutional networks (FCNs) and fuzzy integral is proposed for detecting the position of the floor and nonfloor from visual images. FCN is an end-to-end, pixels-to-pixels network for semantic segmentation. Semantic segmentation aims to perform dense segmentation tasks on images and segments each pixel to a specified category. To overcome the majority decision drawback in the traditional voting method and increase the accuracy, the fuzzy integral is used for the fusion of multiple FCNs with various optimal methods. The overall accuracy, mean accuracy, and mean intersection over union (MIoU) of the proposed method are 0.9824, 0.9816, and 0.9577, respectively. The experimental results show that the proposed hybrid method has better accuracy than other methods in identifying the location of the floor area.

1. Introduction

In recent years, the applications of robot technology have become more extensive, and most of them use visual images to obtain information about the surrounding environment. To make a robot avoid obstacles and move autonomously, the robot should first recognize the surrounding environment through images. Image segmentation is a key part of the machine that separates the object area automatically from the image and identifies the object as well. Many advances in technology, such as artificial intelligence, robots, deep learning, networking, and sensors, have quickly changed human life. Image sensors have been widely used in various robots to sense environmental features in recent years, such as unmanned aerial vehicles (UAVs), unmanned ground vehicles (UGVs) including sweeping robots, Mars rover vehicles, storage robots, and robotic dogs. Therefore, in this paper, we use the vision camera in UGV robots to detect the floor area.

*Corresponding author: e-mail: cjlin@ncut.edu.tw
https://doi.org/10.18494/SAM.2020.2577

ISSN 0914-4935

There are many methods traditionally used for image segmentation. For example, Chun *et al*.[1] proposed a novel method of calculating depth information from a single indoor image through nonlinear diffusion and image segmentation, and using the nonlinear diffusion method to detect floor candidate regions. Image segmentation technology can be used to detect the floor area and calculate depth information, but it cannot be applied to more complex environments. Kumar *et al*.[2] used superpixels to segment some small obstacles. If the obstacles are not accurately distinguished from the appearance of the floor area, the line segment detection method is added to solve this problem. Finally, further processing using the Markov random field can detect obstacles on the floor but cannot accurately segment the floor area. Aggarwal *et al*.[3] proposed a first-person camera image to detect the floor area, using surface clue classification, floor position density cues, and geometric cues to synthesize a common floor area mask, using the GrabCut algorithm.[4] The universal floor area mask is subjected to multiple iterations to obtain a complete floor area. This method can detect the floor area in a variety of indoor environments but requires a high computational complexity. In 2016, DeepLabv2 was proposed by Google,[5] which used atrous convolution to increase the perception domain without increasing the number of parameters, and then adopted the fully connected condition random field to optimize the prediction results. However, the computational cost of using atrous convolution is high and it takes up a lot of memory. In 2017, Lin *et al*.[6] proposed RefineNet. The deeper layers that capture high-level semantic features can be directly refined using fine-grained features from earlier convolutions.

Although the above-mentioned traditional image segmentation methods can detect the floor area, there is a major disadvantage. That is, the floor area cannot be divided accurately. Recently, many researchers have applied deep learning networks to image segmentation to obtain good image segmentation recognition capabilities. Convolutional neural networks (CNNs) can effectively implement image classification. Since the network eventually has a fully connected layer, the original two-dimensional matrix becomes one-dimensional and loses spatial information. Finally, the output is the classification label vector. Therefore, when detecting the floor area, some special cases and misjudgments are generated. To make the mobile robot identify the walkable area accurately, in this study, we adopt a fully convolutional network (FCN) to implement the floor area detection.

For a deep learning network, there are some optimal methods for improving the network's performance. The most common method is stochastic gradient descent (SGD),[7] which randomly selects a certain number of training samples in each time for training. This method can usually learn training samples effectively, but it depends on the learning rate setting and the training time is long. Therefore, some optimization methods[8] have been proposed for solving this problem. For example, AdaGrad[9] adaptively adjusts the learning rate during the training process and uses a larger learning rate to update the training parameters. In contrast, a smaller learning rate is used to update the larger learning rate. In the middle and late stages of training, the relationship between the multiplication rate and the learning rate is approximated by the gradient of the denominator. When the gradient approaches zero, it will cause the training to end early. Adam[10] adopts an adaptive learning rate for each parameter. After the deviation correction is performed, each parameter is calculated for each iteration. The parameters have a clear range of smoothness, which makes the parameters more stable.

In this study, a hybrid of FCNs and fuzzy integral is used to detect the position of the floor and nonfloor from visual images. FCN is used as the trained end-to-end, pixels-to-pixels network for semantic segmentation. Semantic segmentation aims to perform dense segmentation tasks on images and segments each pixel to a specified category. Some optimal methods are used for improving the performance of FCN. Therefore, to overcome the majority decision drawback in the traditional voting method, the fuzzy integral is used for the fusion of multi FCNs with various optimization methods.

This paper is organized as follows. In Sect. 2, we introduce the FCN network. In Sect. 3, we present a hybrid of FCN and fuzzy integral. Experiments on floor detection are carried out to verify the validity of the proposed method, as discussed in Sect. 4. Section 5 is the discussion and Sect. 6 is the conclusion and outlines future works.

2. FCNs

In general, CNN is usually used for object detection and recognition. Its accuracy is low in the classification of image pixels. Our research requires the classification of pixels in images, which is semantic segmentation. Therefore, the traditional CNN is not suitable in this study. FCN is used to classify the pixels of images and achieves semantic segmentation.

FCN was proposed by Long *et al.*[11] in 2017 to classify images at the pixel level and end-to-end for image segmentation, pixels-to-pixels, and training using supervised learning. FCN is different from CNN and is built on a "fully convolution" network, which can input images of any size and use the deconvolution layer to map the features of the last convolutional layer. It is restored to the same size as the input image. Then, a prediction is generated for each pixel. Therefore, the final pixel classification is performed using the upsampled feature map, as shown in Fig. 1.

The network layer of FCN is mainly composed of a convolution layer, a pooling layer, an upsampling layer, an activation function layer, and a skip layer. The operation process is described as follows:

- Convolution layer

The convolution layer consists of several convolution kernels are used to extract the features of the input image. More convolutional layers can extract more complex features. Each

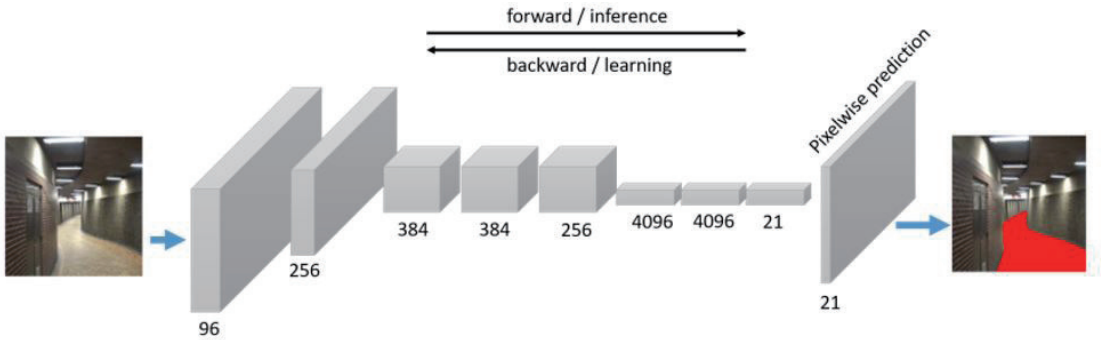


Fig. 1. (Color online) Structure of a FCN.

convolution layer is covered by a convolution kernel, and each convolution kernel has a different weight combination. A convolution operation (inner product) is performed in a sliding window manner to generate a feature map, as shown in Fig. 2.

- Pooling layer

The goal of the pooling layer is to reduce the dimensions and not lose important feature information, and to reduce the amount of subsequent large-parameter operations. The pooling process also performs a mask operation on the input matrix in a sliding window manner. The mask moving process adopts a nonoverlapping manner, indicating that each element of the input matrix will only undergo one pooling operation. The common pooling functions are "max pooling" and "average pooling", as shown in Fig. 3.

- Upsampling layer

This layer is the fully connected layer of the alternative convolutional network. The purpose of this layer is to restore the pooled output to a segmentation map of the input image. This is called deconvolution or transposed convolution. Deconvolution is similar to convolution, and involves multiplication and addition operations. A schematic diagram of upsampling is shown in Fig. 4.

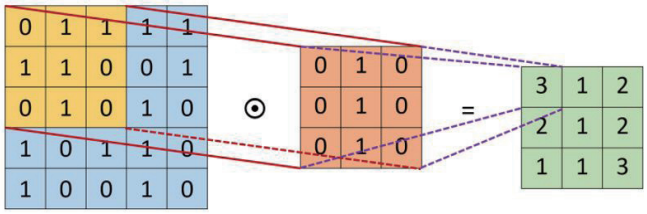


Fig. 2. (Color online) Convolution operation.

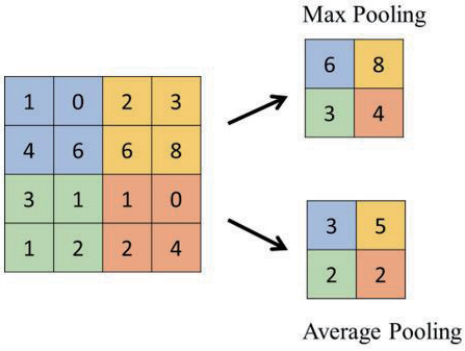


Fig. 3. (Color online) Pooling operation.

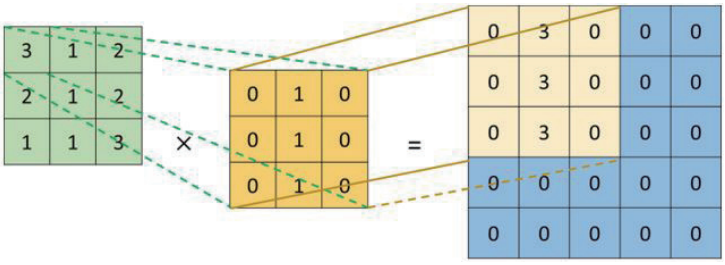


Fig. 4. (Color online) Schematic diagram of upsampling.

- Activation function layer

The goal of this layer is to provide networks with only linear combinations with nonlinear expression capabilities for solving more complex nonlinear problems. The functions employed in this layer are called nonlinear transfer functions. Common activation functions are sigmoid, tanh, and the rectified linear unit (ReLU). In a traditional neural network, the sigmoid function is usually used as the activation function but it encounters the problem of the gradient disappearing in the process of transit transfer. Therefore, ReLU is widely used as the activation function to solve this problem and can reduce the degree of overfitting, as shown in Fig. 5.

- Skip layer

When the upsampling process is completed, the result of dense prediction can be obtained and the segmentation result is rough. Then the skip structure is added, whose main function is to optimize the result. The results of different pooling layers are upsampled to optimize the output, that is, the sum of the output of the last few layers and the final output. The structure of the skip layer is shown in Fig. 6.

3. Proposed Hybrid of FCN and Fuzzy Integral Method

3.1 Three optimal methods for FCN

To improve the performance of FCN, two methods are generally adopted, one for the network structure change and the other for various optimal methods. The network structure change usually requires the addition of the hardware devices and increases the cost. Therefore, most researchers focus on the optimal method of FCN. In the field of deep learning, the choice of optimal algorithms is also a top priority of model training. Even with the same dataset and model architecture, different optimal algorithms are used to generate different training results. Gradient descent is one of the most widely used optimal algorithms in neural networks. To overcome the various drawbacks of gradient descent, several researchers have developed some optimal algorithms. In this study, we will introduce three optimal methods.

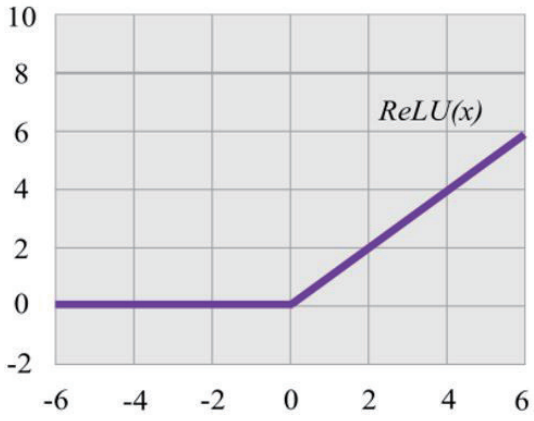


Fig. 5. (Color online) ReLU activation function.

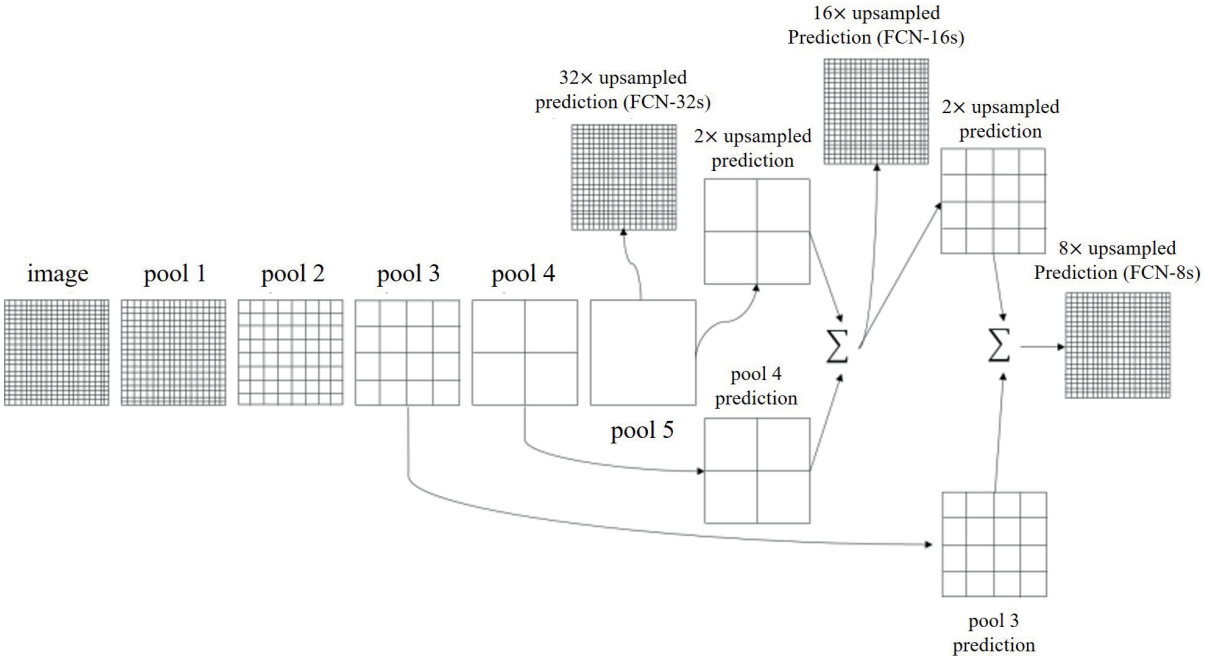


Fig. 6. Structure of skip layer.

3.1.1 SGD

The most common optimization method in the training process is SGD, which randomly selects a set of training samples each time for training. This method can usually learn training samples effectively, but is too dependent on the learning rate setting and has a longer training time.

3.1.2 AdaGrad

In the training of the neural network, the learning rate η is very important. A small learning rate will require a very long training time, and a large learning rate will oscillate during training, resulting in failure to learn correctly. The learning rate of the past optimizer is fixed. AdaGrad adjusts the learning rate η according to the gradient.

3.1.3 Adam

Adam is an optimal method that combines momentum and AdaGrad. It retains the gradient velocity adjustment of the momentum and refers to the direction of the past gradient and the learning rate of the square of the past gradient. In addition, Adam performs an "offset correction" on the parameters to determine each learning rate, which will make the parameter update more stable.

3.2 Fuzzy integral fusion for multiple FCNs

The traditional fuzzy integral can combine multiple classifiers of the same task to improve the performance.[12] In the trained classifiers, the output of each classifier is used as the input of the fuzzy integral. The fuzzy integral evaluates the output of each classifier via the fuzzy measure. Then, the final output is obtained.

To accurately detect the position of the floor area from images, we fuse multiple FCNs with the fuzzy integral. That is, the fusion of multiple optimal FCNs is used by the fuzzy integral. First, we train different FCNs with the same set of training materials, where each FCN has its own performance capabilities, and then we fuse the fuzzy integral with the good representation capabilities of each FCN to obtain better accuracy. The fuzzy integral takes the output of the multiple FCNs as an input, and refers to the performance of each FCN to calculate the output to obtain the classification results. The overall flow chart of fusing multiple FCN fusions based on fuzzy integral is shown in Fig. 7.

If $\boldsymbol{X} = \{\boldsymbol{x_i}\}_{\boldsymbol{i}=1:\boldsymbol{n}}$ represents a collection of n classifiers, the fuzzy measure $\boldsymbol{g}(\boldsymbol{x_i})$ can be used as their confidence level and to evaluate a subset. The fuzzy measure is between 0 and 1. If the fuzzy measure is equal to 1, this means that the output of this set can be fully trusted. If the fuzzy measure is equal to 0, the output of this set has no reference value. The fuzzy integral is mainly based on a fuzzy measure. Multiple rules are used to determine a decision and to obtain the final output. The fuzzy measure should satisfy three conditions.

1. $\boldsymbol{g}(\boldsymbol{X}) = 1$ indicates that when all classifier outputs are consistent, the results must be trusted.
2. $\boldsymbol{g}(\varnothing) = 0$ represents that the outputs of all classifiers are not considered; the result has no reference value.

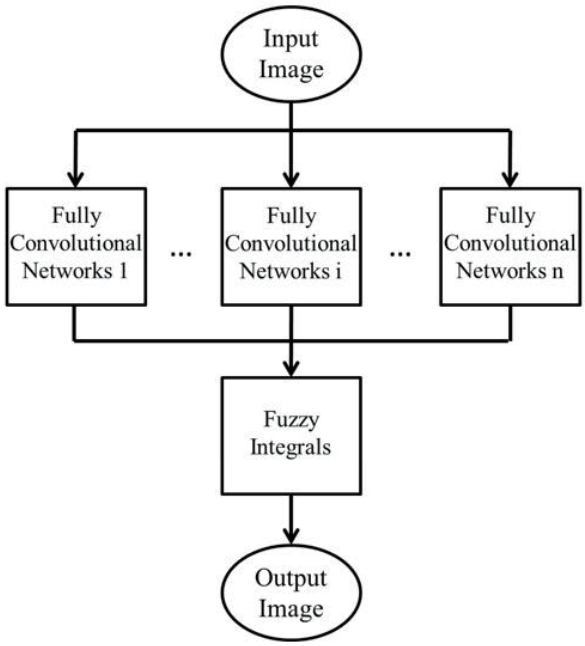


Fig. 7. Flow chart of fusing multiple FCNs based on fuzzy integral fusion.

3. The fuzzy measure must be an increasing monotonic function, namely,

$$\boldsymbol{A} \subset \boldsymbol{B} \subset \boldsymbol{X}, \text{ then } 0 \leq \boldsymbol{g}(\boldsymbol{A}) \leq \boldsymbol{g}(\boldsymbol{B}) \leq 1. \quad (1)$$

When the set in $\boldsymbol{g}()$ has only one element, this fuzzy measure is called the fuzzy density. Before calculating other fuzzy measures, the fuzzy density must be designed.

The relationship between fuzzy measure and fuzzy density is shown in Fig. 8. In the traditional fuzzy integral, the fuzzy density is designed by experts, and the intuitive cognitive fuzzy density is the confidence level of each output of the classifier. In general, the fuzzy density represents the worth value of the outputs in each network classifier. Some studies have used the accuracy of non-training data as the fuzzy density; however, the accuracy of non-training data is usually not the optimal fuzzy density.

4. Experimental Results

To verify the effectiveness of the proposed method, in this paper, we use the public MIT scene dataset and various indoor environment images as data sets to evaluate the performance of the proposed method. The experiment divides the data set into three parts: training data, verification data, and test data, to perform the floor image segmentation task, and then compares the differences of each image segmentation method. Finally, the proposed fuzzy integral method is used for image segmentation, which can improve the recognition rate.

Three metrics are used from common semantic segmentation and scene resolution evaluations, namely Pixel Accuracy (PA), Mean Pixel Accuracy (MPA), and Mean Intersection over Union (MIoU). k represents the number of pixel class and P_{ij} represents the number of pixels that belong to class i but are predicted to be in class j. P_{ii} represents the number of true positives, and P_{ij} and P_{ji} are interpreted as false positives and false negatives, respectively.

PA:

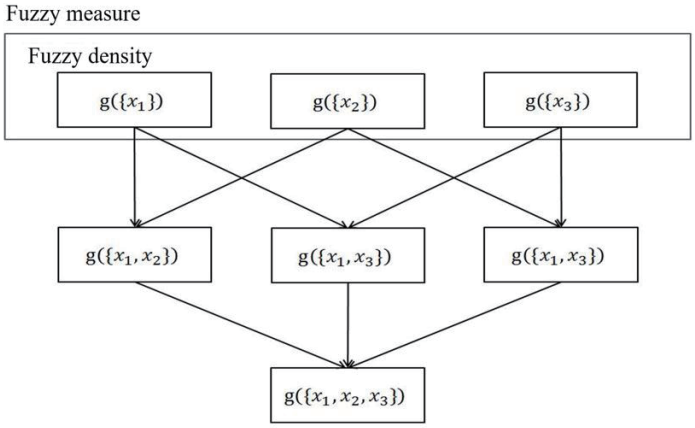


Fig. 8. Relationship between fuzzy measure and fuzzy density.

$$PA = \frac{\sum_{i=0}^{k} P_{ii}}{\sum_{i=0}^{k} \sum_{j=0}^{k} P_{ij}} \tag{2}$$

MPA:

$$MPA = \frac{1}{k+1} \sum_{i=1}^{k} \frac{P_{ii}}{\sum_{j=0}^{k} P_{ij}} \tag{3}$$

MIoU:

$$MIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{P_{ii}}{\sum_{j=0}^{k} P_{ij} + \sum_{j=0}^{k} P_{ji} - P_{ii}} \tag{4}$$

In this experiment, the different extracted information methods of FCN (i.e., FCN-8s, FCN-16s, and FCN-32s) are adopted. FCN-8s represents extracted information from pool 3, pool 4, and conv 7, and then simply adds them by upsampling. FCN-16s represents extracted information from pool 4 and conv 7, and then simply adds them by upsampling. FCN-32s represents extracted information only from conv 7 and then resizes it by upsampling. Table 1 shows the comparison results of FCN using different extracted information. Experimental results show that FCN-8s has a better performance than FCN-16s and FCN-32s in terms of overall accuracy, mean accuracy, and MIoU.

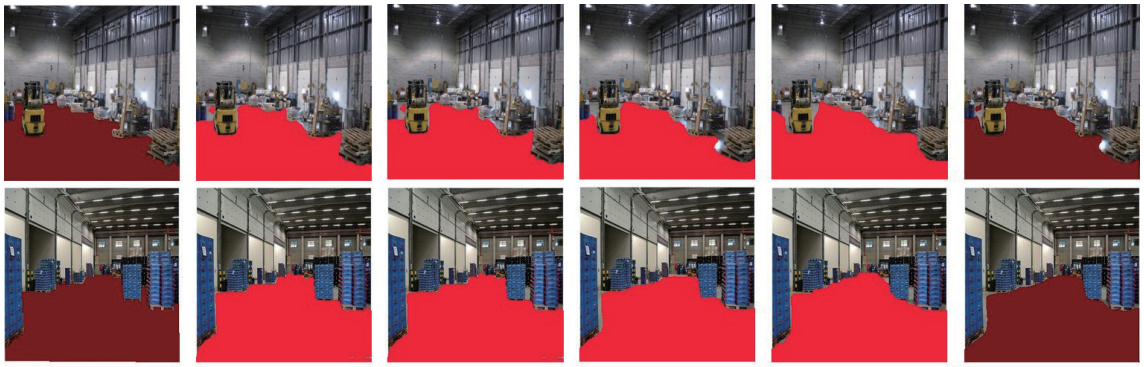
In the following experiments, we adopt FCN-8s with different optimal methods to perform the detection of floor area. Three different optimal methods are adopted, namely, Adam, AdaGrad, and SGD, and a related research method, DeepLabv2.[5] A total of 200 images are used as training data, whereas the verification data and the test data are 48 and 48 images, respectively. The number of iterations in the training process is 100000 the learning rate is 0.001, and the momentum is 0.9. Table 2 shows a comparison of the results of FCN with Adam, FCN with AdaGrad, FCN with SGD, DeepLabv2, and the proposed method. The experimental results show that the proposed method has a better overall accuracy, mean accuracy, and MIoU than FCN with Adam, FCN with AdaGrad, FCN with SGD, and DeepLabv2. Figure 9 shows a comparison of the results of ground truth, the proposed method, DeepLabv2, and the three optimal methods of FCN in the detection of floor area.

Table 1
Comparison of results of FCN using different information extraction methods.

FCN (s)	Overall accuracy	Mean accuracy	MIoU
8	0.9769	0.9661	0.9389
16	0.9752	0.9634	0.9343
32	0.9721	0.9594	0.9267

Table 2
Comparison of results of FCN with Adam, FCN with AdaGrad, FCN with SGD, DeepLabv2, and the proposed method.

Methods	Accuracy		
	Overall accuracy	Mean accuracy	MIoU
FCN with Adam[10]	0.9775	0.9660	0.9403
FCN with AdaGrad[9]	0.9768	0.9654	0.9386
FCN with SGD[7]	0.9782	0.9685	0.9423
DeepLabv2[5]	0.9713	0.9663	0.9389
Proposed method	0.9824	0.9816	0.9577



(a) (b) (c) (d) (e) (f)

Fig. 9 (Color online) Floor area detection results: (a) ground truth, (b) FCN with Adam, (c) FCN with AdaGrad, (d) FCN with SGD, (e) DeepLabv2, and (f) proposed method.

5. Discussion

A comparison of the results of different methods are shown in Fig. 9. We can observe that FCN with three different optimal methods and DeepLabv2[5] can identify the location of the floor area. However, DeepLabv2 cannot identify the details of the floor area successfully. For example, DeepLabv2 may misjudge other smoother items as the floor and detect the edge of the floor area as not being flat as shown in Fig. 9(e). Therefore, the advantage of the fuzzy integration method is added to obtain a more accurate position of the floor area, and the result of the fuzzy integral is shown in Fig. 9(f). The proposed method can accurately detect the location, details, and debris of the floor area. Table 2 shows a comparison of the experimental results. The overall accuracy, mean accuracy, and MIoU of the proposed method are 0.9824, 0.9816, and 0.9577, respectively. As a result, the proposed method has a higher accuracy rate than the other methods in the floor area detection.

6. Conclusion

In this study, we propose a hybrid of FCN and the fuzzy integral to detect the position of the floor and nonfloor from visual images. Some optimal methods are used for improving the performance of FCN. Therefore, to overcome the majority decision drawback in the traditional

voting method, the fuzzy integral is used for the fusion of multiple FCNs with various optimal methods. Experimental results show that FCN-8s has a better performance than FCN-16s and FCN-32s. Moreover, the experimental results also show that the proposed method has a higher overall accuracy, mean accuracy, and MIoU than FCN with Adam, AdaGrad, SGD, and DeepLabv2. The MIoU of the proposed method and FCN with Adam, AdaGrad, and SGD, and DeepLabv2 are 0.9577, 0.9403, 0.9386, 0.9423, and 0.9389, respectively.

In the fuzzy integral, the fuzzy density is designed by experts, and the intuitive cognitive fuzzy density is the confidence level of each output of the classifier. In future research, evolutionary methods will be used to determine the fuzzy density. In addition, we will also apply the proposed hybrid detection method to real mobile robots for identifying the floor area, and in the future develop an efficient algorithm for avoiding obstacles while the mobile robot is moving.

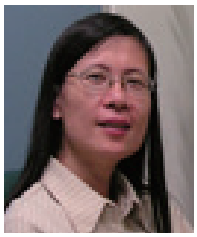
References

1 C. Chun, D. Park, W. Kim, and C. Kim: 2013 IEEE Int. Conf. Image Processing (IEEE, 2013) 3358. https://doi.org/10.1109/ICIP.2013.6738692
2 S. Kumar, M. S. Karthik, and K. M. Krishna: 2014 IEEE Int. Conf. Robotics and Automation (IEEE, 2014) 494. https://doi.org/10.1109/ICRA.2014.6906901
3 S. Aggarwal, A. M. Namboodiri, and C. V. Jawahar: 2014 22nd Int. Conf. Pattern Recognition (IEEE, 2014) 4275. https://doi.org/10.1109/ICPR.2014.733
4 Y. Boykov and V. Kolmogorov: IEEE Trans. Pattern Anal. Mach. Intell. **26** (2004) 9. https://doi.org/10.1109/TPAMI.2004.60
5 L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille: IEEE Trans. Pattern Anal. Mach. Intell. **40** (2018) 4. https://doi.org/10.1109/TPAMI.2017.2699184
6 G. Lin, A. Milan, C. Shen, and I. Reid: 2017 IEEE Conf. Computer Vision and Pattern Recognition (IEEE, 2017) 5168. https://doi.org/10.1109/CVPR.2017.549
7 Y. Liu, W. Huangfu, H. Zhang, and K. Long: IEEE Trans. Wireless Commun. **18** (2019) 7. https://doi.org/10.1109/TWC.2019.2914040
8 C. S. Asness, T. J. Moskowitz, and L. H. Pedersen: J. Finance **68** (2013) 3. https://doi.org/10.1111/jofi.12021
9 J. Duchi, E. Hazan, and Y. Singer: J. Mach. Learn. Res. **12** (2011) 2121. http://dl.acm.org/citation.cfm?id=1953048.2021068
10 Z. Chang, Y. Zhang and W. Chen: 2018 IEEE Int. Conf. Software Engineering and Service Science (IEEE, 2018) 245. https://doi.org/10.1109/ICSESS.2018.8663710
11 J. Long, E. Shelhamer, and T. Darrell: IEEE Trans. Pattern Anal. Mach. Intell. **39** (2017) 4. https://doi.org/10.1109/TPAMI.2016.2572683
12 N. Bouadjenek, H. Nemmour, and Y. Chibani: IET Biom. **6** (2017) 6. https://doi.org/10.1049/iet-bmt.2016.0140
13 B. Tong and Y. Liu: IEEE Int. Conf. Robotics and Biomimetics (IEEE, 2016) 1337. mhttps://doi.org/10.1109/ROBIO.2016.7866512
14 C. J. Lin, S. S. Kuo, and C. C. Peng: Int. J. Fuzzy Syst. **14** (2012) 3.

About the Authors

Cheng-Jian Lin received his Ph.D. degree in electrical and control engineering from National Chiao-Tung University, Taiwan, R.O.C., in 1996. Currently, he is a Lifetime Distinguished Professor of the Computer Science and Information Engineering Department, National Chin-Yi University of Technology, Taichung City, Taiwan, R.O.C. His current research interests are machine learning, pattern recognition, intelligent control, image processing, and evolutionary robots. (cjlin@ncut.edu.tw)

Yu-Chi Li received her B.S. degree from the Department of Computer Science and Information Engineering, Providence University, Taichung City, Taiwan, R.O.C., in 2018. Currently, she is a graduate student of the Computer Science and Information Engineering Department, National Chin-Yi University of Technology, Taichung City, Taiwan, R.O.C. Her current research interests are machine learning, pattern recognition, and image processing.



Chin-Ling Lee received her B.S. degree in English literature from Tamkang University, Taiwan, R.O.C., in 1986, M.S. degree in English literature from Central Missouri State University, U.S.A. in 1990, and Ph.D. degree in industrial education from National Taiwan Normal University, Taiwan, R.O.C., in 2005. From August 2003 to July 2005, she was an assistant professor in the Department of Applied Foreign Languages, Nan-Kai Institute of Technology, Nantou, Taiwan, R.O.C. Currently, she is an associate professor in the International Business Department, National Taichung University of Science and Technology. Her current research interests are English teaching, time series prediction, machine learning, and intelligent systems.