IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 30 June 2023, accepted 15 July 2023, date of publication 20 July 2023, date of current version 7 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3297141

RESEARCH ARTICLE

An Efficient Neural Network for Pig Counting and Localization by Density Map Estimation

WEI FENG[1,2], KAINING WANG[2], AND SHANGBO ZHOU[1], (Member, IEEE)

[1]College of Computer Science, Chongqing University, Chongqing 400044, China
[2]SimpleCredit Micro-Lending Company Ltd., Chongqing 401147, China

Corresponding author: Shangbo Zhou (shbzhou@cqu.edu.cn)

This work was supported by NSFC under Grant 62272070.

ABSTRACT Automatic pig counting and locating from camera images is one of the most important tasks in modern pig farming industry, which helps farmers to improve the efficiency of the livestock management in pig feeding, welfare estimation, unexpected events monitoring and etc. Due to the complex and diverse pigpen environment, complicated distributions of pig population and various motions of live pigs, traditional image processing techniques are not effective in counting and locating pigs in crowds. Thus, this task relies on manual labor heavily which is time-consuming and error-prone. In this paper, we propose an efficient and accurate pig counting method for top-view surveillance images in large-scale, crowded feeding scenes. The proposed method is composed of a novel density map generator and a density map estimation network architecture. The pigs in images are expressed as ellipses and their group density is generated with elliptical 2D Gaussian distribution. The proposed network is designed with efficient hybrid blocks including selective kernel convolution and vision transformer for feature extraction and density map regression. The total number of pigs in one image can be calculated by summing entire values in the density map. We also apply a modified K-means clustering algorithm on the density map to locate pig targets. To verify the effectiveness and precision of the proposed method, we evaluate our proposed method on our testing dataset. The Mean Absolute Error of counting numbers on testing images is 0.726. Due to lightweight design by using depth-wise separable convolutions and hybrid-vit blocks, our proposed method has very fast inference speed and will reduce dependency on computing resources substantially when deployed in pig farms. Based on the accurate estimated density maps of the testing images, pig locating can also achieve pretty good results. The modified K-means clustering algorithm proposed in this paper obtain target locating with 88.22% precision and 86.02% recall respectively. These results indicate our proposed method can accurately count pigs in piggery by density map estimation and locate pig targets even in crowded situations.

INDEX TERMS Pig counting, pig locating, selective kernel convolution, vision transformer, density map estimation.

I. INTRODUCTION

Counting and locating pigs with automatic methods in livestock environment is very critical for large-scale pig farming industry. Frequently obtaining pig numbers in each pen is essential since pigs are often moved or divided into different barns due to growth, size, behavior and etc. It is important for farmers to prepare appropriate amount of fodder and adjust pig numbers according to the scale of pigpens to ensure the welfare of animals. On the other hand, it is possible to detect unexpected things such as missing pigs, dying or disease pigs which stay still for long times. However, counting and inspecting pigs in pigpen by human is huge amount of labor in large-scale pig farming industry. Frequent contact between human and pigs will also increase the risk of cross-species transmission. Thus, it is promising to adopt new techniques for the automatic, non-contacting management of pig farming which would greatly lower the management cost of manpower.

The rapid development of image and video surveillance technologies as well as computer vision techniques enables researchers to count, locate and track pigs in pigpens by

The associate editor coordinating the review of this manuscript and approving it for publication was Yiming Tang.

bypassing high-cost and inefficient manual labors. The related researches using traditional image processing techniques usually segment the images into blobs with pig targets, and then transform them into ellipses using the blob pixels or border pixels. An ellipse can be simply represented using five parameters but it appropriates the edge of pig body sufficiently. Ellipse fitting methods for pig locating can be traced back to the research decades ago [1] and have been developed and improved in the following years. Many recent researches have applied ellipse fitting approaches in the detection and location of pigs successfully [2], [3], [4]. However, the traditional pig counting, detection and segmentation algorithms encounter great challenge because of the complicated housing environment, changeable illumination, various distribution of group density, as well as overlapping and occlusion. In recent years, deep neural networks have been proven to be powerful to extract features automatically in complicated scenes and widely used in many fields of agriculture [5], [6] including pig farming industry [7], [8]. Detection-based methods can be used for object counting [9], [10], but their performance will decline as the objects become more and more crowded. In contrast, density-based methods [11], [12] are very suitable for counting in crowds. The counting accuracy of density-based methods depends on the density descriptions and feature maps extracted. Most density-based methods obtain ground-truth density maps by convolving object position with Gaussian kernels and ignore the shape and size of objects. This method was adopted in the research of Tian et al. for pig counting [13]. While this representation method is not reasonable as the targets are usually very large in images which we couldn't neglect their sizes and shapes. Deep neural network architectures have been shown to be effective for feature extraction from images and be very suitable for the following tasks such as classification, regression. One of the most representative algorithms for crowd counting is Multi-Column Convolutional Neural Network (MCNN) [14]. The backbones of semantic segmentation networks are very suitable for the regression tasks in density counting such as FCN [15], DeepLabV3 [16]. Except for CNN based networks, transformer-based networks have shown great potential in image processing tasks since ViT [17]. SegFormer [18] is one of the efficient transformer-based networks for semantic segmentation and its backbone is very suitable for density regression tasks. Inspired by the aforementioned methods, we propose a novel density map estimation method for pig counting and locating in crowds.

The rest of the paper is organized as follows: Section II describes our data collection and data augmentation of our dataset, the method of pig labeling and ground-truth density map generation, the structure of our proposed model and the postprocess algorithm of density maps for pig counting and locating in details. In section III, we introduce the experiments implemented and the results obtained, we compare the density estimation and counting performance, inference speed of our proposed method with several competing methods. The robustness of the proposed network is also verified in this section. Discussion and conclusion are summarized in Section IV and V.

This paper makes several contributions to the literature. First, a more reasonable density map generator considering shapes and sizes of pigs is proposed to generate density maps instead of convolving pig position with Gaussian kernels. Second, we design a high-efficient deep neural network to obtain density maps of the input images. Postprocess is applied on the output density maps to obtain pig numbers and locations. Experiments are conducted and the results indicate our method is accurate, efficient and robust. Finally, our approach is distinct from similar studies by counting and locating pigs in crowds using density maps which takes positions and shapes of pigs into consideration by ellipse-fitting.

II. MATERIALS AND METHODS

A. DATASET AND DATA AUGMENTATION

The dataset for this work has been acquired from 4 pigpens in a very large-scale pig farm. The pigs in those pigpens are not usually the same via time due to growth or piggery adjustment. The images used in our dataset are captured by 4 top-view fisheye cameras installed on the roof of 4 pigpens. Two of the pigpens are relatively small and lack of sunlight, less pigs are feeding in these 2 pigpens. The other two are large and have sufficient natural illumination, the density of pig groups in these 2 pigpens are relatively higher. We collect these pictures in natural feeding conditions and no special enhancements, such as constant illumination, marks to distinguish pigs from background, etc., are prepared. We take pictures of these pigpens at random times each day, and totally capture 12 images of pigs one day. This procedure of data collection has lasted for 183 days. The original resolution of the images from this camera is 1080 × 720 pixels. To enhance the variety of background, we crop each image with a random size range from 0.4 to 0.8 of the image size to obtain sample images for our dataset. Figure 1 shows some cropped images used in this paper. Images in the first column are captured from the smaller two pigpens at 9:40am, 17:50pm, respectively. The other images are captured from the larger two pigpens at 11:20am, 19:00pm, 14:30pm and 10:00am, from left to right and top to bottom, respectively.

Totally we have obtained 2196 images for our dataset and the number of pigs in each image is in a range from 5 to 46, averagely 22. To avoid overfitting of the network, data augmentation is implemented while the training batches are loaded. The data augmentation is a combination of vertical flipping with 0.5 probability, horizontal flipping with 0.5 probability, randomly changing the brightness, contrast and saturation of the image. Due to the different sizes of images in dataset, we pad the images by filling zero values to generate batches with the same height and width. The height and width of images in batches are all divisible by 16.

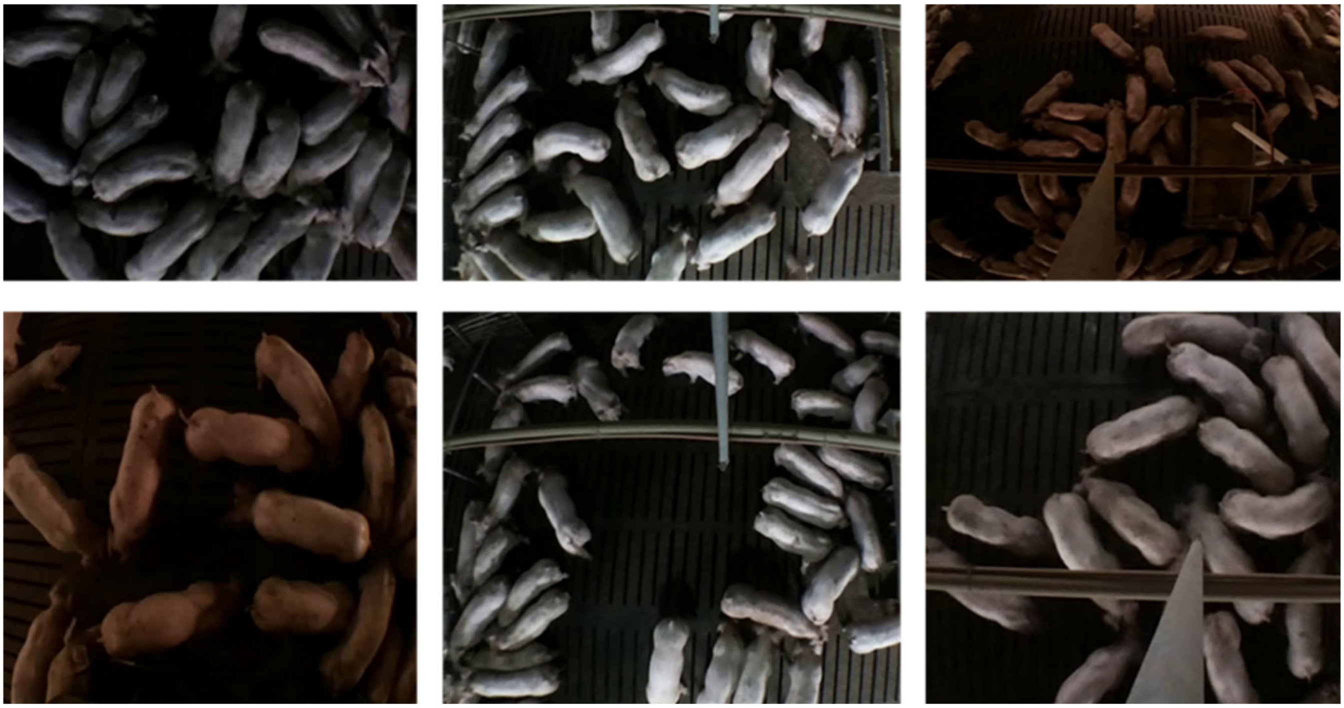

FIGURE 1. Examples of images in our dataset.

B. DATA LABELING AND GROUND-TRUTH DENSITY MAP GENERATION

In this paper, we mark more than 5 points at the boundary of a pig target and then apply a direct least square ellipse fitting method to obtain an ellipse for this pig. We record the five parameters of the fitted ellipse including centroid (x_c, y_c), semi-axis lengths (R_a, R_b), counterclockwise angle θ from the horizontal axis to the major axis. The position of a pig is represented by the ellipse center and its shape and size are expressed as the area of the ellipse. This manual work will continue until all pigs in one image are labelled by ellipses and their parameters are recorded.

In our proposed method, the density maps are designed at a size with 1/8 of the original image sizes. So, the recorded parameters of ellipses should be resized to the same scale. For one pig in the image, its density representation is described as in (1) and (2).

$$G(x;\mu,\sum)=\frac{|\mathrm{Z}|}{2\pi}e^{-\frac{(x-\mu)^T\sum^{-1}(x-\mu)}{2}} \tag{1}$$

$$\mu=(x'_c,y'_c)=(x_c/8,y_c/8) \tag{2}$$

$|Z|$ is the normalization factor to ensure the sum density values of one pig is 1.0. The isocontours of the G function are all elliptically shaped in a ratio of R_a/R_b. Based on the 3σ rules in Gaussian distribution, the standard deviation on the orientation of major/minor axis would be described as in (3) and (4).

$$\sigma_1=R'_a/3=R_a/24 \tag{3}$$

$$\sigma_2=R'_b/3=R_b/24 \tag{4}$$

And the covariance matrix Σ for the Gaussian distribution is represented by the parameters of ellipse as in (5).

$$\begin{aligned}\sum&=\begin{bmatrix}\sigma_{11}^2 & \sigma_{12}\\ \sigma_{21} & \sigma_{22}^2\end{bmatrix}\\ &=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}\sigma_1^2 & 0\\ 0 & \sigma_2^2\end{bmatrix}\begin{bmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{bmatrix}\end{aligned} \tag{5}$$

The range of the Gaussian distribution is set from x_c - $3\sigma_{11}$ to x_c+ $3\sigma_{11}$ in the horizontal direction and y_c - $3\sigma_{22}$ to y_c+ $3\sigma_{22}$ in the vertical direction respectively. Repeat the above steps until all pigs in one image are represented by elliptical 2D Gaussian distributions, the ground-truth density heat map is then generated using (6).

$$D(x)=\sum_{p_i\in P}G(x;p_i) \tag{6}$$

P represents the collection of all fitting ellipses for pigs in one image. The total number of pigs in the image can be calculated by summing up all values in the density map as in (7).

$$N=\sum_{x\in I}D(x) \tag{7}$$

Fig.2 (a), (b) shows a brief example of data labelling and ground-truth density map generation of our proposed method for one image in dataset, Fig.2 (c), (d) shows the density map generated by convolving with constant Gaussian kernels and adaptive Gaussian kernels [19] in contrast. It is obvious that the density map generated by our proposed method had huge

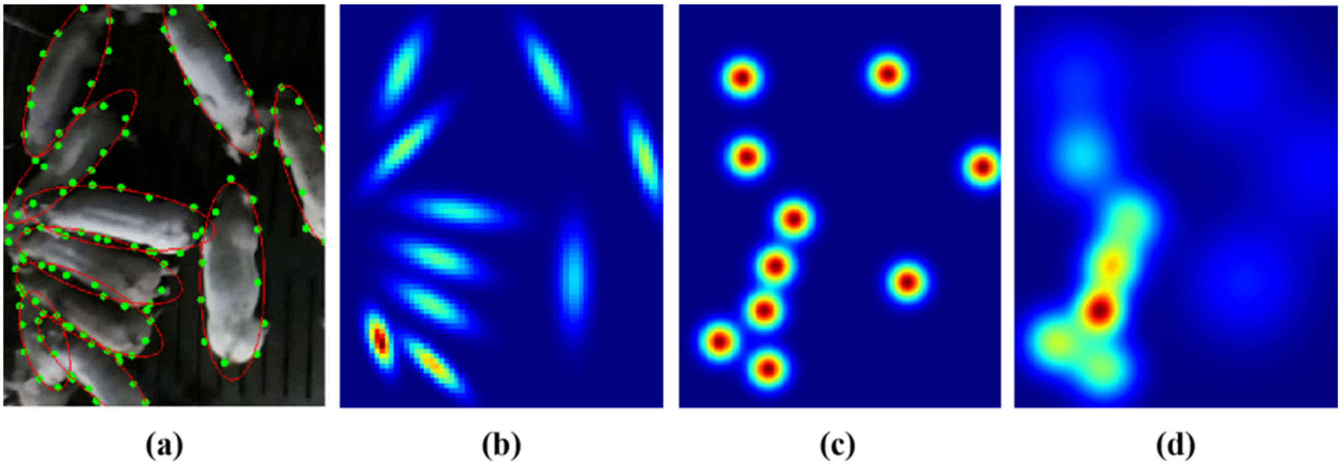

FIGURE 2. (a) pig labeling and ellipse fitting, (b) ground-truth density map generated by our proposed method, (c) ground-truth density map generated by convolving with constant Gaussian kernels, (d) ground-truth density map generated by convolving with adaptive Gaussian kernels.

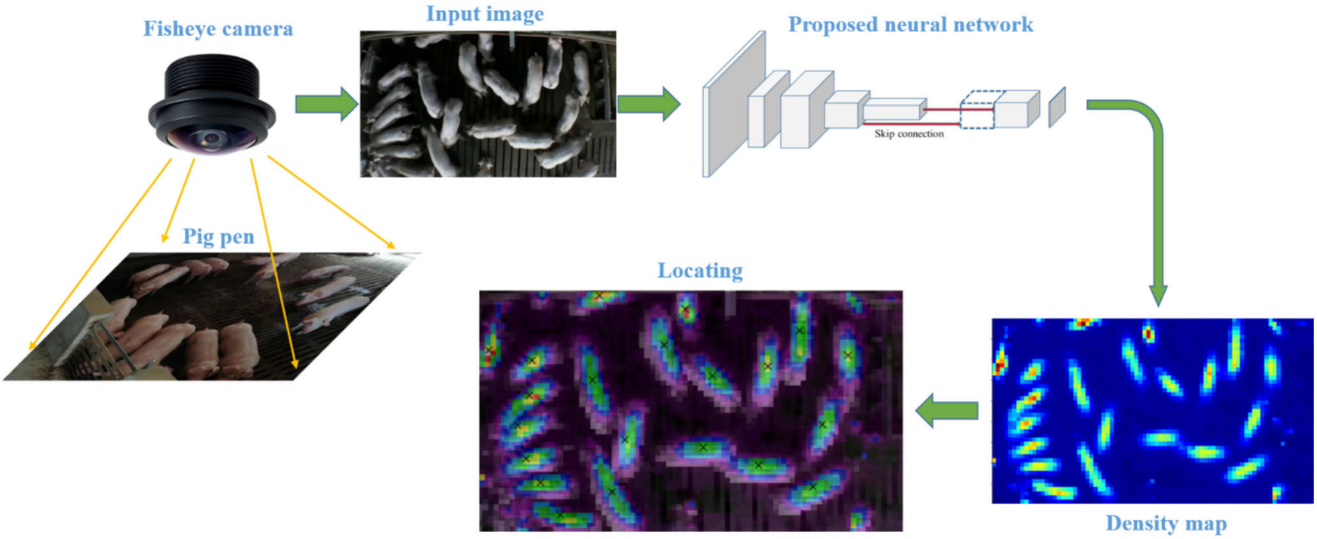

FIGURE 3. The sketch of the proposed method.

advantage over the traditional method on the representation of locations and shapes of pigs.

C. THE PROPOSED METHOD

In the proposed method, a more reasonable density map generator is designed for pig images to generate density maps instead of convolving pig position with Gaussian kernels. We take the shapes and sizes of pigs into account by ellipse fitting and ground-truth density maps generating using 2D Gaussian distributions. Then we design a high-efficient deep neural network to obtain accurate density maps of the input images. Since we can obtain accurate density distribution of pigs with centers and shapes, a modified K-means clustering method was applied on density maps to locate pigs. The sketch of our proposed method is shown as in Figure 3.

For our proposed neural network, it is composed of a feature extractor and a density map regression head. The feature extractor draws out image features by CNN modules in shallow layers and efficient vision transformer modules in deeper layer. In the first two stages, a modified depth-wise selective kernel convolution (DWSKConv) block has been stacked to extract features with larger sizes. The DWSKConv blocks use selective kernel convolution [20] which can aggregate features extracted by kernels with different sizes and adaptively adjust the receptive field size of neurons. The depth-wise and point-wise convolution and the reversed bottleneck structure in the DWSKConv block can reduce computational cost significantly without performance degradation. The sketch of DWSKConv is shown in Figure 4 (b). In the last stage, a hybrid vision transformer (hybrid-vit) block is stacked to dig features with global receptive field. The hybrid-vit

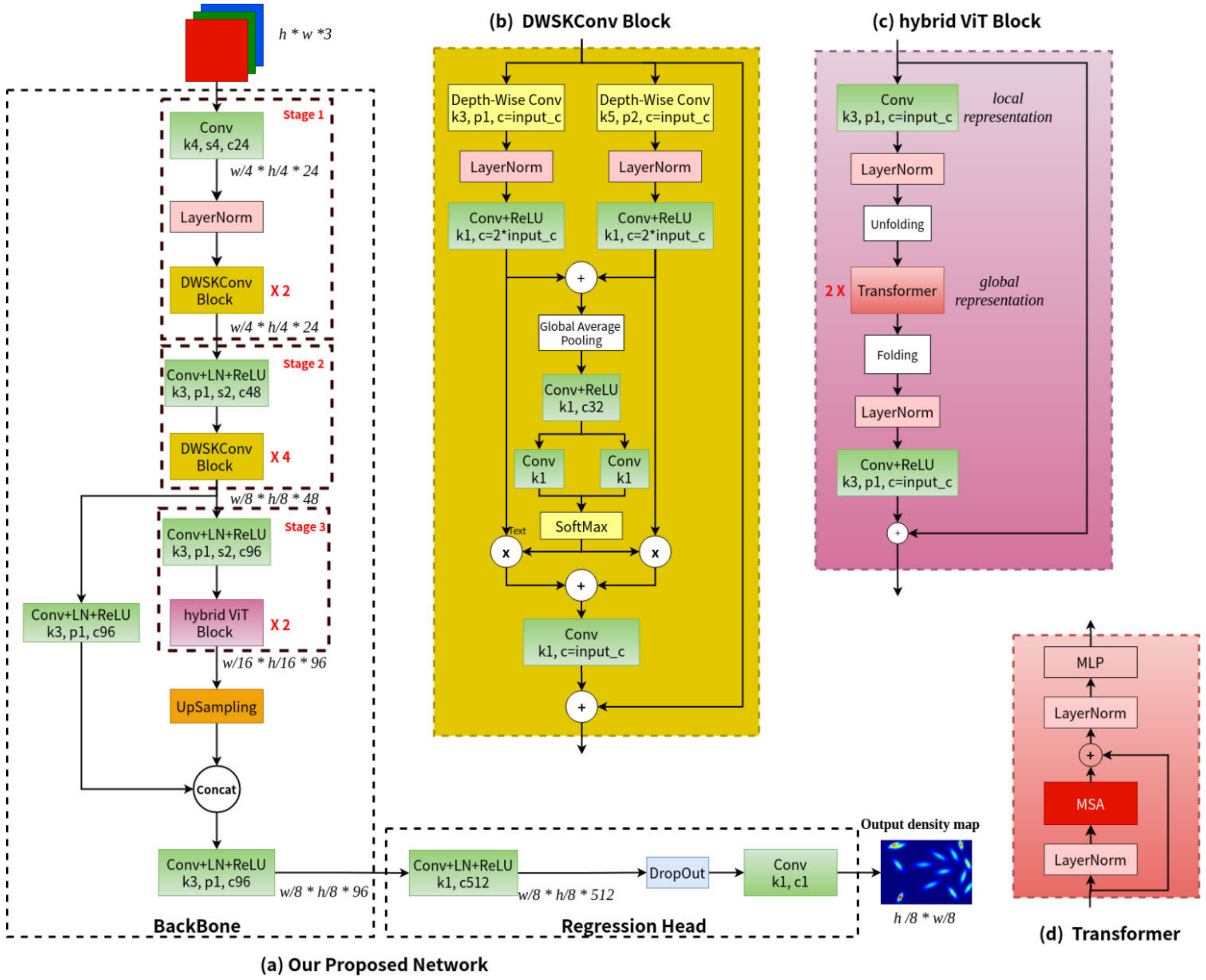

FIGURE 4. (a) The sketch of our proposed neural network architecture. (b) The DWSKConv Block. (c) The hybrid-vit Block. (d) Transformer architecture.

block [21] uses a convolution with filters of 3 × 3 to obtain the local representation of the feature maps at the start. Then the output feature maps are partitioned by windows with a size of 3 × 3. Since the tokens in each partition windows have a local connection by CNN already, the following efficient vision transformer blocks calculate multi self-attention of tokens only on the same position of each partition windows. This operation is implemented by folding, vision transformer and unfolding. Figure 5 gives a brief example of folding and unfolding. This operation reduces the computation to 1/9 compared with multi-head self-attention calculation of all tokens and can still obtain a perfect global representation of features. After the global representation of features, convolution with filters of 3 × 3 is implemented. The sketch of hybrid-vit is shown in Figure 4 (c) (d). The size of the output feature maps by the last stage is 1/16 of original image size, upsampling of high-level features by bilinear interpolation with a factor of 2. Feature fusion from deep and shallow levels by concatenation and convolution is implemented to utilize features from different levels. The output feature maps are then fed into a simple regression head built by CNNs to get the predicted density maps of the input images. The architecture of the proposed network and the output sizes of key layers are illustrated as in Figure 4 (a).

The proposed network combines CNNs and vision transformer blocks to extract features in different receptive fields from local to global. Considering the cost of deploying the model in pig farms, our multi-receptive-model is designed by using efficient convolution blocks and lightweight vision transformer blocks as basic building blocks. Due to the depth-wise separable selective kernel convolution blocks in the first two stages and the efficient hybrid-vit blocks in the last stage, the computational cost of the proposed model decreases significantly.

In order to obtain the precise density map and counting number of pigs, pixel-wise loss and counting loss are

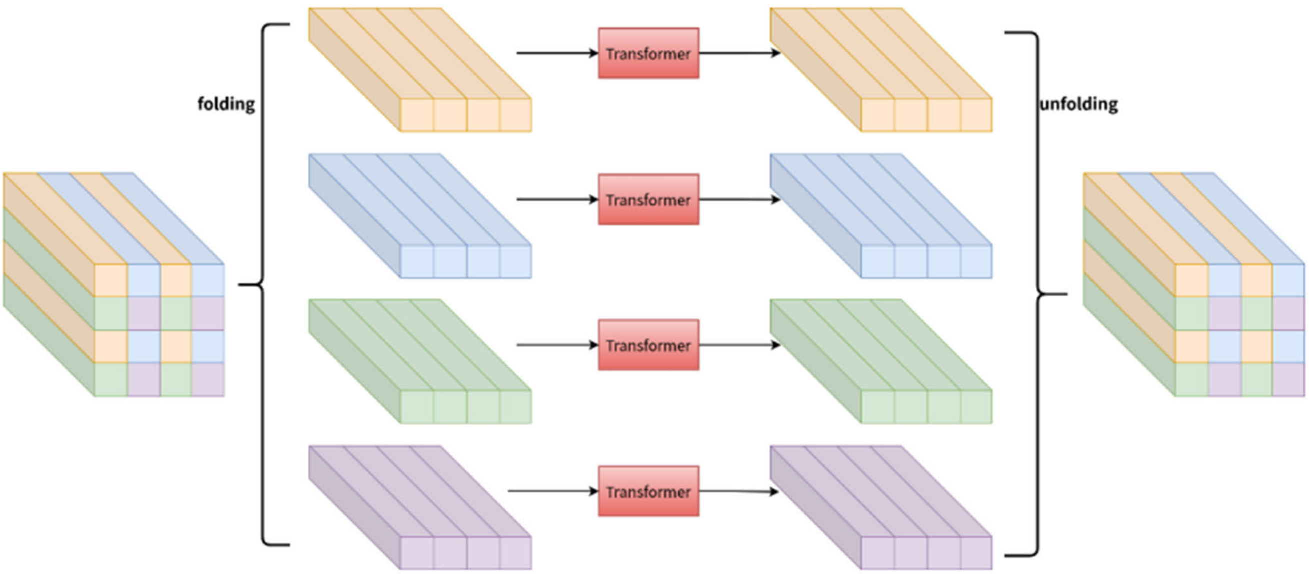

FIGURE 5. A simple example of folding and unfolding operations with window size of 2 × 2 in the hybrid-vit block.

combined into a comprehensive weighted loss function shown as in (8), (9) and (10).

$$l_{pixel} = \frac{1}{N_{image}} \sum_{(x,y)} \left[D(x,y) - \hat{D}(x,y) \right]^2 \quad (8)$$

$$l_{count} = \frac{1}{N_{image}} \left[\sum_{(x,y)} D(x,y) - \sum_{(x,y)} \hat{D}(x,y) \right]^2 \quad (9)$$

$$loss = l_{pixel} + \lambda l_{count} \quad (10)$$

where $D(x, y)$ represents the ground-truth density maps and $\hat{D}(x, y)$ represents the output density maps of the proposed neural network. The pixel-wise loss ensures the output density map is a pixel level reproduction of the ground-truth map and the counting loss tunes the network for the accuracy of counting. Since the counting loss is much larger than pixel-wise loss and highly positive correlated to pixel-wise loss, the hyper parameter λ is used to balance the impact of pixel-wise loss and counting loss in training process. The impact of λ value on the performance of the proposed network would be discussed in the following experimental section.

At test stage, the given test images are firstly padded with zero values, the heights and widths of test images are all divisible by 16 and then fed into our network. The density map of the test image is obtained by cropping the output into 1/8 of the original image size.

Detecting and locating objects directly from density map was first proposed for partially-occluded small instances [22] with an integer programming algorithm. For large objects such as pigs in crowed scenes, the integer programming algorithm is not accurate enough. Since the output density maps of our proposed network have well-defined information of target centers and boundaries similar to the ground-truth density maps, target locating directly from the density maps using clustering can also gain perfect performance. In this paper, we propose a modified K-means clustering where pixels are weighted based on their density values. The center of each cluster can be calculated by all pixels in each cluster as in (11).

$$\left(C_x, C_y\right) = \left(\frac{\sum_{(x,y)\in C} x\hat{D}(x,y)}{\sum_{(x,y)\in C} \hat{D}(x,y)}, \frac{\sum_{(x,y)\in C} y\hat{D}(x,y)}{\sum_{(x,y)\in C} \hat{D}(x,y)} \right) \quad (11)$$

This procedure would lead clustering centers shifting towards higher value region and getting more accurate target positions. To accelerate the iterative process and reduce computation, we segment the density map to several connected components and calculate local peak pixels as initial clustering centers. The sketch of our locating algorithm is described as follows:

1) Binarize the density map with a threshold and find all connected regions R;
2) For each region R_i in the collection R, find all local peaks P and determine clustering number by summing up all density values in R_i, obtain the initial N clustering centers from P if the number of points in P is equal or larger than N, otherwise randomly choose the remaining centers in R_i;
3) Calculate the distances between each pixel to clustering centers and divide pixels into clusters corresponding to the nearest center;
4) Update centers by weighted average of the pixels in clusters according to (11);

TABLE 1. Configurations of the experiments.

CPU+RAM	Intel i7-11700F (16 cores) +32GB
GPU	4 * NVIDIA GeForce GTX 1080 Ti (11GB)
Machine learning framework	Pytorch 1.9.0
Operating System	Ubuntu 18.04
Size of training dataset	1600
Training dataset augmentation	Yes
Training batch size per GPU	4
Size of testing dataset	596
Testing dataset augmentation	NO

TABLE 2. Performance of the proposed network on varying values of balancing parameter lambda of loss function.

λ	MAE	MSE
0	0.982	0.298
0.001	0.908	0.309
0.002	0.812	0.310
0.005	0.726	0.317
0.01	0.796	0.319
0.02	0.862	0.334
0.05	0.977	0.345
0.1	1.055	0.351

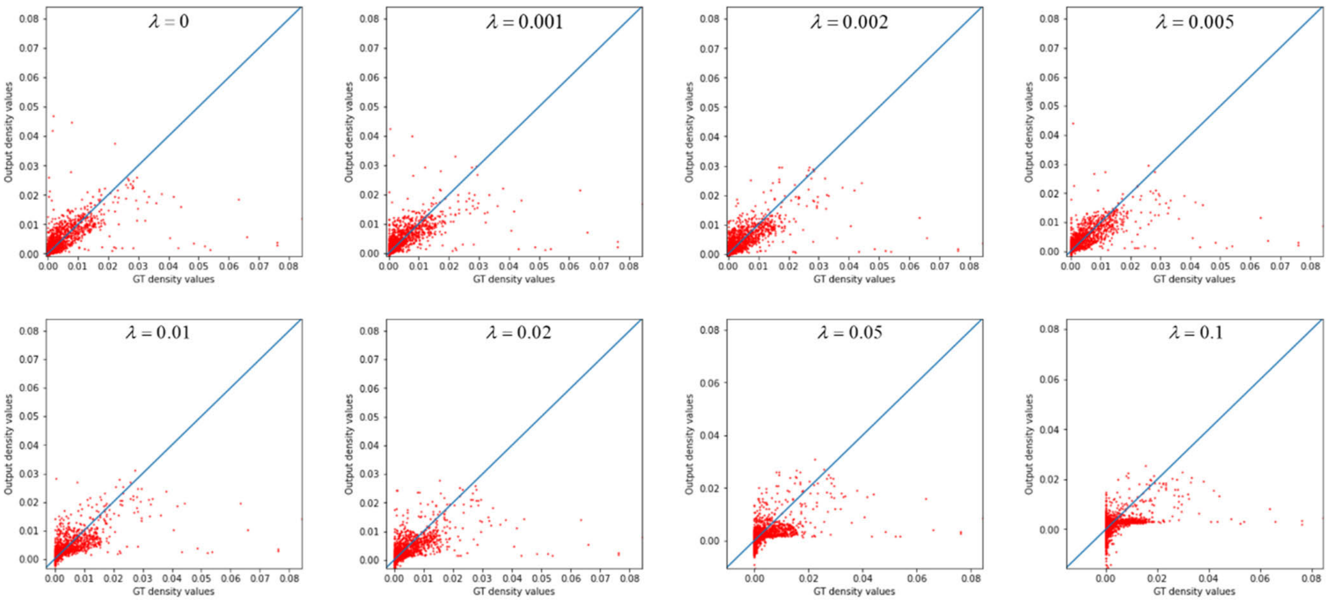

FIGURE 6. Scatter plot of pixel-wise values between the ground-truth density map and the predicted density map of our proposed network with various loss functions.

5) Repeat step 3 and step 4 until the centers of clusters stop moving, and output the centers of clusters.

III. EXPERIMENTS AND RESULTS

A. EXPLANATION OF THE EXPERIMENTS

To verify the performance of our proposed neural network on the collected dataset, we apply mean absolute error (MAE) to evaluate the performance of pig counting and mean square error (MSE) to evaluate the performance of pig density map estimation. They can be described as in (12) and (13).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\sum D_i(x,y) - \sum \hat{D}_i(x,y)\right| \tag{12}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\sum (D_i(x,y) - \hat{D}_i(x,y))^2 \tag{13}$$

The training and testing of the neural network are carried out on a desktop computer. The related configurations are summarized in TABLE 1.

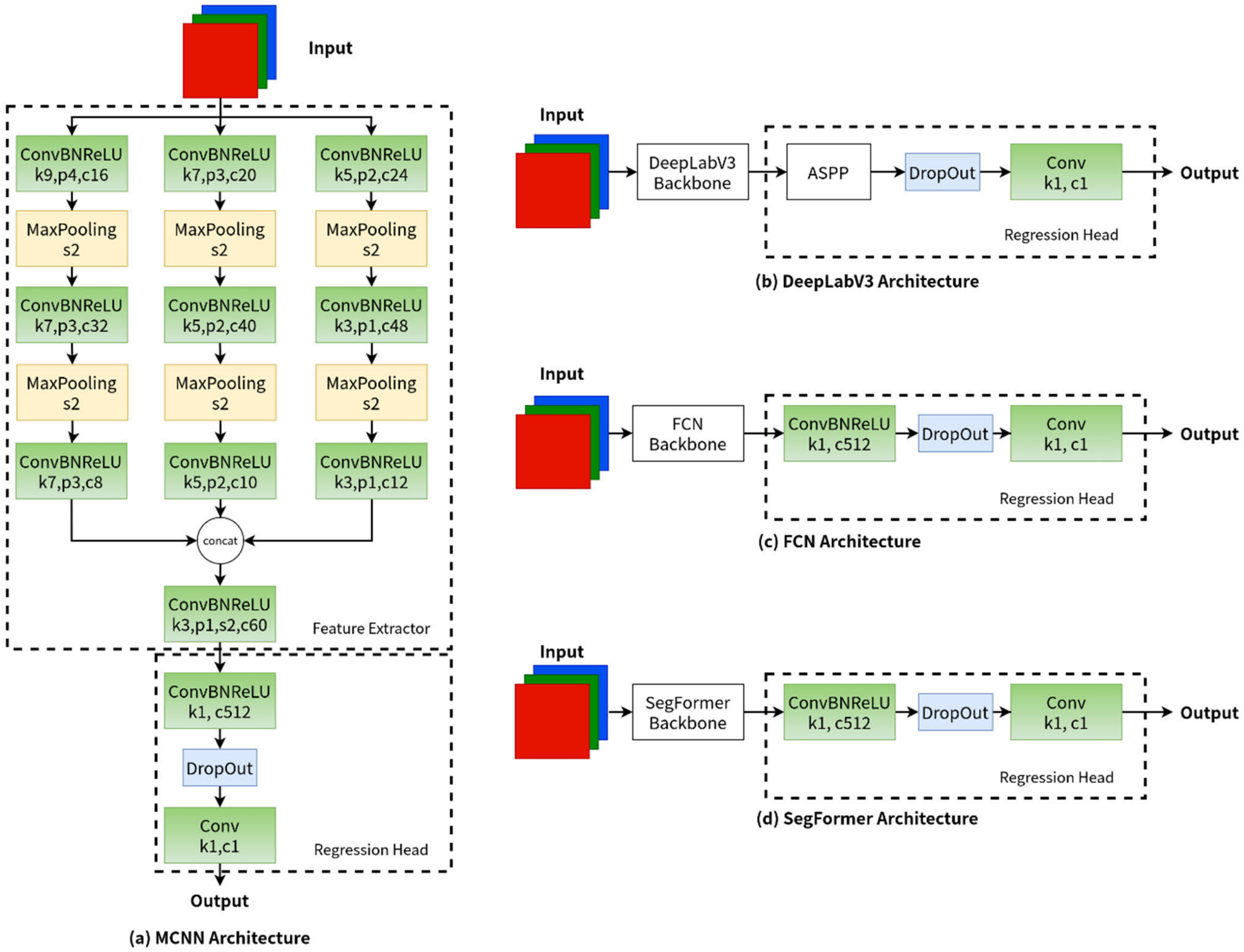

FIGURE 7. (a) MCNN architecture, (b) DeepLabV3 architecture, (c) FCN architecture, (d) SegFormer architecture.

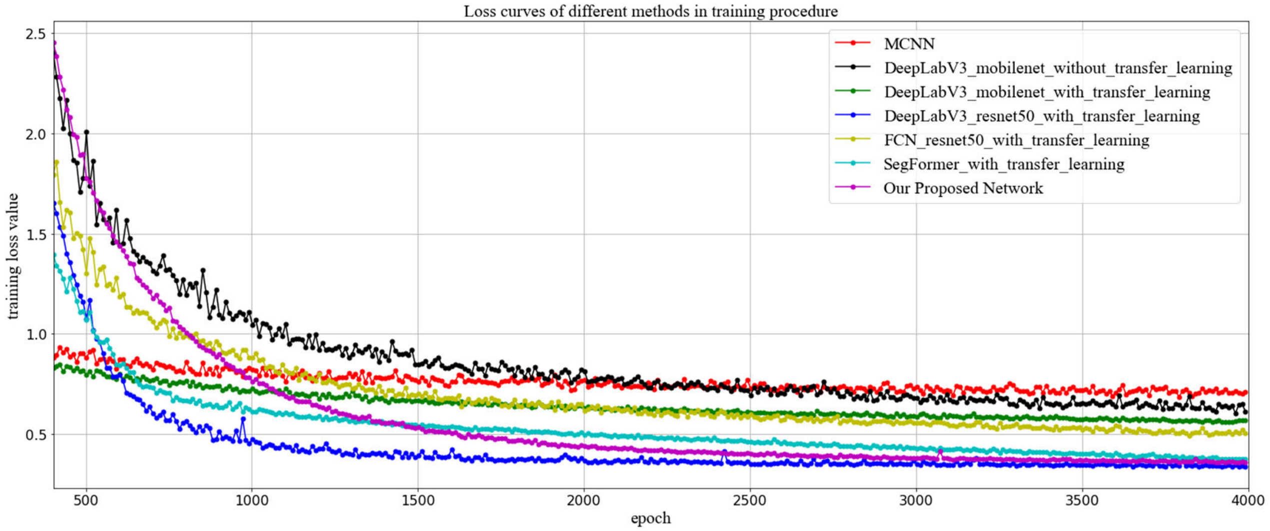

FIGURE 8. Loss on training dataset of several methods.

We adopt full images instead image patches to train and test the proposed neural network. In order to avoid overfitting, two-dimensional dropout is used while training the network. The dropout probability is set as 0.5. The parameters of the

TABLE 3. Comparison of computational complexity and counting performance of different models.

Method	Transfer Learning	Parameters (M)	Inference Time (s)	MAE	MSE
MCNN	No	0.14	0.0124	1.561	0.441
DeepLabV3 (mobilenetv3_large)	No	1.10	0.0133	1.394	0.403
DeepLabV3 (mobilenetv3_large)	Yes	1.10	0.0133	1.099	0.363
FCN (resnet50)	Yes	3.29	0.0734	0.854	0.337
DeepLabV3 (resnet50)	Yes	3.96	0.1032	0.730	**0.314**
SegFormer (mit_b3)	Yes	4.85	0.0709	0.735	0.322
Our proposed network	No	1.42	**0.0077**	**0.726**	0.317

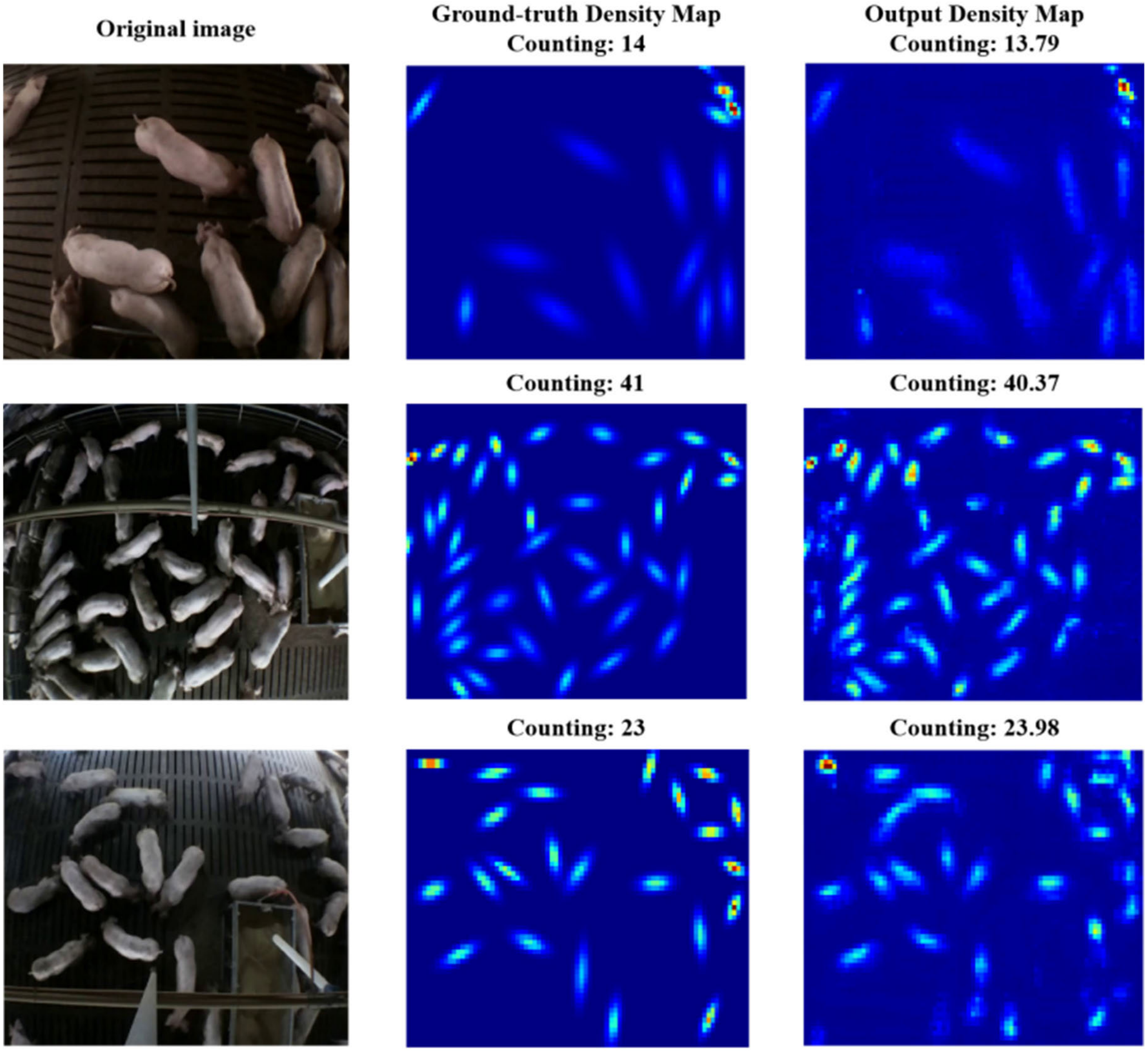

FIGURE 9. Visualization of the performance of our proposed method on pig density estimation and counting.

network are initialized by He initialization [23] and optimized using Adam with a learning rate of 8×10^{-5}. The first and second order moment calculation of the optimizer are set as 0.9 and 0.99 respectively. In the training process, we evaluate the performance of the network on testing dataset every 5 training iterations and select the model with best performance. To find the most reasonable loss function for back propagation in training process, we apply a series of λ values in the training process of the proposed neural network. The performances on testing dataset of the proposed network with different loss functions are listed in TABLE 2.

To better display the pixel-wise performance of the network with various λ values, a scatter plot between the

TABLE 4. Performance of our proposed method on pig number validation.

Validation case	Training images	Testing images	MAE	MSE
Training on data with more pigs, testing on data with less pigs.	1257	939	0.819	0.328
Training on data with less pigs, testing on data with more pigs.	939	1257	0.826	0.331

TABLE 5. Performance of our proposed method on pig density validation.

Validation case	Training images	Testing images	MAE	MSE
Training on data with more crowded pigs, testing on data with less crowded pigs.	824	1372	0.893	0.340
Training on data with less crowded pigs, testing on data with more crowded pigs.	1372	824	0.998	0.359

TABLE 6. Statistics of the performance of pig locating using our proposed clustering method on testing dataset.

Ground-truth	Predict results	
	Positive	Negative
Positive	11586	1883
Negative	1547	/

ground-truth density values and predicted density values are shown in Figure 6.

From Table 2 and Fig.6, it is obvious that with the increase of λ value, the counting performance of the proposed network raises first and then falls down, while the pixel-wise performance decreases. Comprehensively considering all results of the experiments, the weighted parameter λ for loss function is reasonable to set in a range between 0.002 to 0.01. In this paper, we adopt $\lambda = 0.005$ in all the training procedures.

B. PERFORMANCE COMPARISON WITH OTHER DENSITY REGRESSION NETWORK ARCHITECTURES

To demonstrate the superior performance of our proposed neural network, we compare the performance of our network with several outstanding density regression networks. MCNN is one of the well-known density estimation networks which is used for various applications. Since it is lightweight and efficient, the performance of MCNN on our dataset is applied as the baseline. As introduced in section II, many semantic segmentation algorithms are very suitable for regression tasks. In this work, we implement several segmentation neural networks to evaluate their performance on our dataset for pig density estimation and pig counting, including FCN with backbone of resnet50, DeepLabV3 with backbone of mobilenetv3_large and resnet50, SegFormer with backbone of mit_b3. The schematic diagram of these algorithms is shown as Figure 7.

Since some of these semantic segmentation networks are heavyweight and difficult to train properly on small dataset. We use transfer learning on these models to obtain their best performance on our dataset very fast. The training steps of transfer learning are shown as follows:

1) Load pre-train weight data;
2) Freeze all parameters in backbone and only update parameters in regression head within several training epochs;
3) Release frozen parameters and update all parameters in the network with a very small learning rate.

The losses in the training process are shown in Figure 8 and the performance of all algorithms on density map estimation and pig counting are shown in TABLE 3. All results are achieved by averaging data with several training and testing processes. The inference times of all methods are obtained by feeding input images with resolution of 640 × 480 pixels on a NVIDIA GeForce GTX 1080Ti.

From Fig.8 and Table 3, we could clearly see that all loss curves converge very well, training with pre-trained weights could accelerate convergence speed and enhance the performance of DeepLabV3 (mobilenetv3_large). Our proposed method has a slight advantage on counting performance and very similar density regression performance compared with DeepLabV3 (resnet50) and SegFormer (mit_b3), better than the other models. On the other hand, the model complexity and computational cost of our proposed neural network are much less than the other methods, the inference speed of our proposed method is nearly ten times faster than DeepLabV3 (resnet50) and SegFormer (mit_b3). These experimental results show great superiority of our proposed method on pig density map estimation and counting with our collected dataset.

To directly display the performance of our proposed method on pig density estimation and counting, we randomly

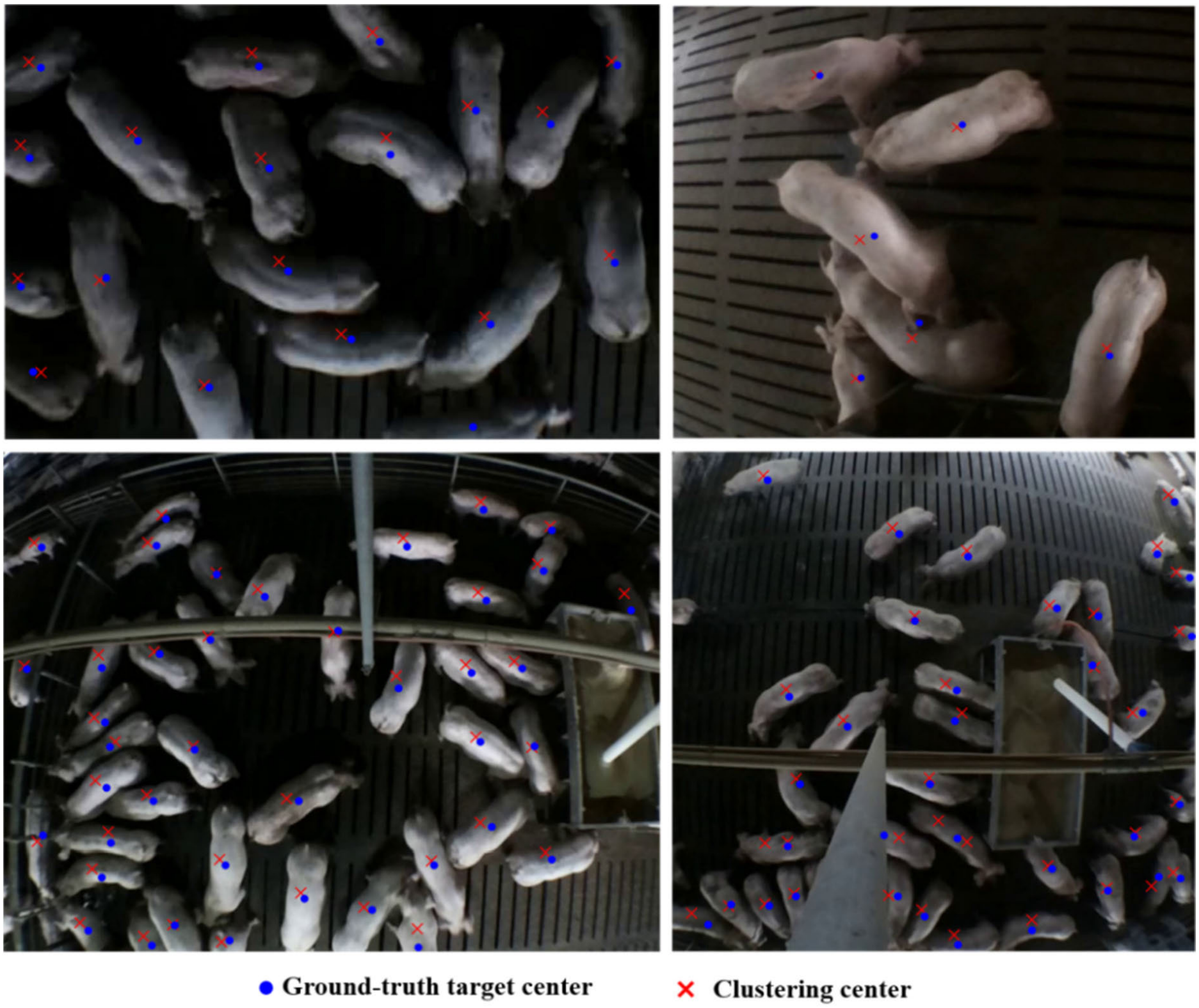

FIGURE 10. Examples of the locating performance of our proposed method.

select several images in test dataset and feed them into the proposed method. Figure 9 displays the results of the input images obtained by our proposed method.

C. CROSS-VALIDATION OF OUR PROPOSED METHOD FOR PIG DENSITY PREDICTION AND COUNTING

To verify the robustness and reliability of our proposed method under various cases, we apply two cross-validations for our proposed network. One is pig number validation and another is pig density validation. In the pig number validation procedure, we divide the dataset into two parts: images with more than 20 pigs and otherwise. We train our network using one part and test on the other one. The result of pig number validation of our proposed method is shown as in TABLE 4.

In the pig density validation procedure, we still divide our dataset into two parts: images with more crowded pigs and otherwise. We train our network using one part and test on the other one. The result of pig density validation of our proposed method is shown in TABLE 5.

The performance of our proposed method also achieves very good performance on pig density estimation and counting in both two cross-validation procedures. Even though the performance is slightly worse due to the different data distribution of the training and testing dataset. And pig density distribution seems to have larger influence on the performance of our proposed method. The MAEs in both cross-validations are all smaller than 1.0 which indicates the robustness and reliability of our proposed method under various cases.

D. PERFORMANCE ON PIG LOCATING USING THE OUTPUT DENSITY MAP

After we obtain the accurate density map of the testing images by our proposed neural network, modified K-means clustering is applied on these density maps to get the locations of pigs in images. Precision and recall of targets locating are computed to evaluate the performance of the clustering algorithm. A target center (x_i, y_i) located by the proposed clustering method is judged as a true positive locating for one pig target (cx_j, cy_j, a_j, b_j, θ) only if the following two conditions are satisfied:

1) The distance between (x_i, y_i) and (cx_j, cy_j) is the minimum one;
2) (x_i, y_i) is located within the core area of the ellipse which is defined as an ellipse with parameter (cx_j, cy_j, $0.5a_j$, $0.5b_j$, θ).

Otherwise, the unpaired locating points are treated as false positives, the unpaired ground-truth ellipses are treated as false negatives. Table 6 demonstrates the locating performance of the modified clustering method on the output density maps generated by our proposed neural network.

Then we could calculate the precision, recall and F1 score of the locating algorithm as in (14), (15) and (16).

$$precision = \frac{TP}{TP + FP} = 88.22\% \tag{14}$$

$$recall = \frac{TP}{TP + FN} = 86.02\% \tag{15}$$

$$F_1 = \frac{2PR}{P + R} = 0.8711 \tag{16}$$

Due to the precise density map estimation of our proposed neural network, pig locating with our modified clustering method could also achieve good performance on precision and recall. To directly display the performance of our proposed method on pig locating, we randomly select several images in test dataset and feed them into the proposed method. Figure 10 displays the results of the input images obtained by our proposed method.

IV. DISCUSSION

Counting and locating pigs with automatic methods in livestock environment is very critical for large-scale pig farming industry. In this paper, we have proposed a novel solution for pig counting and locating. And we have achieved a performance with MAE of 0.726 on pig counting, 88.26% precision and 86.02% recall on pig locating regardless of pigs in crowed scenes. For comparison, Tian et al. [13] achieved a MAE value of 1.69 on dataset collected by themselves and 2.78 on dataset from internet. In our research, we focus on counting and locating pigs in large-scale farms using surveillance videos from top-view and achieve a much lower MAE value. Jensen and Pedersen [24] achieved very similar results with ours on pig counting with their model and dataset by directly outputting pig numbers in images. In contrast, the images in our dataset contain more pigs and more crowed scenes. We aim at not only counting pigs in images, but also locating each pig. Counting and locating pigs in large-scale pig farms regardless of crowds accurately and quickly demonstrates the uniqueness of our research in pig farming industry.

V. CONCLUSION

In this paper, we have proposed a novel and efficient method for pig counting and locating from top-view images using density map estimation method. We propose a simple and accurate density map generation algorithm for pig images by approximating pigs as ellipses. Then an efficient neural network composed of depth-wise separable SK blocks and hybrid-vit blocks is designed to obtain density maps and count pig numbers from input images. The output density maps are processed by our modified clustering method for locating pigs. The results on our testing images demonstrate that our proposed solution has achieved a MAE of 0.726 on pig counting, 88.26% precision and 86.02% recall on pig locating regardless of pigs in crowed scenes. Due to the lightweight design of our proposed method by using depth-wise separable convolutions and efficient hybrid-vit blocks, the running speed of our proposed is very fast compared with several outstanding density regression networks, nearly 10 times faster than DeepLabV3 (resnet50) and SegFormer (mit_b3) according to our experiments. To verify the robustness of our proposed neural network, we conduct experiments of pig number cross-validation and pig density cross-validation. The results of both cross-validation show a good performance under various cases. Our proposed method could make a contribution to pig counting and locating in modern pig farming industry. However, this method is mainly suitable for those images captured from top-view cameras. For images from side-view cameras, pigs aren't appropriately fitted by ellipses which would lead the failure of our proposed method. Counting and detecting using images captured from various views will be considered in our future work.

ACKNOWLEDGMENT

The authors would like to thank SimpleCredit Micro-lending Company Ltd., for providing the materials and facilities of this research. They also thank MMSegmentation: OpenMMLab Semantic Segmentation Toolbox and Benchmark (http://github.com/open-mmlab/mmsegmentation) for the open-source project and algorithm code.

REFERENCES

[1] N. J. B. McFarlane and C. P. Schofield, "Segmentation and tracking of piglets in images," *Mach. Vis. Appl.*, vol. 8, no. 3, pp. 187–193, May 1995.

[2] L. Xuan and L. Shiyao, "Development and experiment of closed performance measuring station for breeding pig integrating body size information," *Trans. Chin. Soc. Agricult. Machinery*, vol. 53, no. 7, pp. 267–274, 2022.

[3] J. Brünger, I. Traulsen, and R. Koch, "Model-based detection of pigs in images under sub-optimal conditions," *Comput. Electron. Agricult.*, vol. 152, pp. 59–63, Sep. 2018.

[4] M. Lu, T. Norton, A. Youssef, N. Radojkovic, and D. Berckmans, "Extracting body surface dimensions from top-view images of pigs," *Int. J. Agricult. Biol. Eng.*, vol. 11, no. 5, pp. 182–191, 2018.

[5] L. Liu, R. Wang, C. Xie, R. Li, F. Wang, and L. Qi, "A global activated feature pyramid network for tiny pest detection in the wild," *Mach. Vis. Appl.*, vol. 33, no. 5, Sep. 2022.

[6] J. Sun, X. He, M. Wu, X. Wu, J. Shen, and B. Lu, "Detection of tomato organs based on convolutional neural network under the overlap and occlusion backgrounds," *Mach. Vis. Appl.*, vol. 31, no. 5, Jul. 2020.

[7] S. Wang, G. Sun, B. Zheng, and Y. Du, "A crop image segmentation and extraction algorithm based on mask RCNN," *Entropy*, vol. 23, no. 9, p. 1160, Sep. 2021.

[8] A. Feng, H. Li, Z. Liu, Y. Luo, H. Pu, B. Lin, and T. Liu, "Research on a Rice counting algorithm based on an improved MCNN and a density map," *Entropy*, vol. 23, no. 6, p. 721, Jun. 2021.

[9] D. Xiao, A. Feng, and J. Liu, "Detection and tracking of pigs in natural environments based on video analysis," *Int. J. Agricult. Biol. Eng.*, vol. 12, no. 4, pp. 116–126, 2019.

[10] C. Liu, J. Su, L. Wang, S. Lu, and L. Li, "LA-DeepLab V3+: A novel counting network for pigs," *Agriculture*, vol. 12, no. 2, p. 284, Feb. 2022.

[11] S. Kumagai, K. Hotta, and T. Kurita, "Mixture of counting CNNs," *Mach. Vis. Appl.*, vol. 29, no. 7, pp. 1119–1126, Oct. 2018.

[12] M. D. S. de Arruda, L. P. Osco, P. R. Acosta, D. N. Gonçalves, J. Marcato Junior, A. P. M. Ramos, E. T. Matsubara, Z. Luo, J. Li, J. D. A. Silva, and W. N. Gonçalves, "Counting and locating high-density objects using convolutional neural network," *Exp. Syst. Appl.*, vol. 195, Jun. 2022, Art. no. 116555.

[13] M. Tian, H. Guo, H. Chen, Q. Wang, C. Long, and Y. Ma, "Automated pig counting using deep learning," *Comput. Electron. Agricult.*, vol. 163, Aug. 2019, Art. no. 104840.

[14] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 589–597.

[15] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[16] S. Kim and M. Kim, "Learning of counting crowded birds of various scales via novel density activation maps," *IEEE Access*, vol. 8, pp. 155296–155305, 2020.

[17] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang, and D. Tao, "A survey on vision transformer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 87–110, Jan. 2023.

[18] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Álvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," in *Proc. NeurIPS*, 2021, pp. 1–21.

[19] V. Lempitsky and A. Zisserman, "Learning to count objects in images," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 1–9.

[20] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 510–519.

[21] J. Yi, Z. Shen, F. Chen, Y. Zhao, S. Xiao, and W. Zhou, "A lightweight multiscale feature fusion network for remote sensing object counting," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5902113.

[22] Z. Ma, L. Yu, and A. B. Chan, "Small instance detection by integer programming on object density maps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3689–3697.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[24] D. B. Jensen and L. J. Pedersen, "Automatic counting and positioning of slaughter pigs within the pen using a convolutional neural network and video images," *Comput. Electron. Agricult.*, vol. 188, Sep. 2021, Art. no. 106296.

WEI FENG received the B.S. and Ph.D. degrees in instrument science and technology from the University of Science and Technology of China, Hefei, China, in 2018. He is currently a Postdoctoral Researcher with the College of Computer Science, Chongqing University, Chongqing, China. His current research interests include image processing, deep neural networks, and agricultural application.

KAINING WANG received the B.S. degree in mathematics from the University of Science and Technology of China, in 1982, and the Ph.D. degree in mathematics from the University of Notre Dame, in 1992. He is currently an independent chief consultant in the area of artificial intelligence theory and applications. He has published more than 40 technical articles and coauthored a book in the area of artificial intelligence and control theory.

SHANGBO ZHOU (Member, IEEE) received the B.S. degree from the Department of Mathematics, Guangxi University for Nationalities, in 1985, the M.S. degree from the Department of Mathematics, Sichuan University, in 1991, and the Ph.D. degree from the School of Electronic and Information Engineering, University of Electronic Science and Technology of China, in 2003.

In 1991, he was with the Chongqing Aerospace Mechanical and Electrical Design Institute, where he was the Deputy Manager, the Manager of the Computer Engineering Department, and the Deputy Director of the Military Products Institute. In 2003, he was with the School of Computer Science, Chongqing University. He has presided one National Key Research and Development Project, one National Natural Science Foundation Project, one Chongqing Basic Science and Frontier Technology Research Project, and participated in several national key research and development projects related to computer applications. His current research interests include video and image signal processing, artificial neural network systems, physical engineering calculation, and computer simulation technology.