University of Reading

Deep-segmentation of plantar pressure images incorporating fully convolutional neural networks

Article

Accepted Version

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Wang, D., Li, Z., Dey, N., Ashour, A. S., Moraru, L., Sherratt, S. ORCID: https://orcid.org/0000-0001-7899-4445 and Shi, F. (2020) Deep-segmentation of plantar pressure images incorporating fully convolutional neural networks. Biocybernetics and Biomedical Engineering, 40 (1). pp. 546-558. ISSN 0208-5216 doi: https://doi.org/10.1016/j.bbe.2020.01.004 Available at https://centaur.reading.ac.uk/88481/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1016/j.bbe.2020.01.004

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

University of Reading

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Full text
Accepted for Biocybernetics and Biomedical Engineering
5th Jan 2020.

Deep-segmentation of plantar pressure images incorporating fully convolutional neural networks

Dan Wang[1], Zairan Li[2], Nilanjan Dey[3], Amira S. Ashour[4], Luminita Moraru[5], R. Simon Sherratt[6], and Fuqian Shi[7]

1. Tianjin Key Laboratory of Process Measurement and Control, School of Electrical Engineering and Automation, Tianjin University, 300072, PR China
2. Wenzhou Polytechnic, Wenzhou, 325035, PR China
3. Dept. of IT, Techno India College of Technology, West Bengal, 740000, India
4. Faculty of Engineering, Tanta University, 31527, Egypt
5.Faculty of Sciences and Environment, Department of Chemistry, Physics and Environment, Dunarea de Jos University of Galati, 800008, Romania
6. Department of Biomedical Engineering, the University of Reading, RG6 6AY, UK
7. First Affiliated Hospital of Wenzhou Medical University, Wenzhou, 325035, PR China

Corresponding author Zairan Li

Abstract:
Comfort shoe-last design relies on the key points of last curvature. Traditional plantar pressure image segmentation methods are limited to their local and global minimization issues. In this work, an improved fully convolutional networks (FCN) employing SegNet (SegNet+FCN 8s) is proposed. The algorithm design and operation are performed using the visual geometry group (VGG). The method has high efficiency for the segmentation in positive indices of global accuracy (0.8105), average accuracy (0.8015), and negative indices of average cross-ratio (0.6110) and boundary F1 index (0.6200). The research has potential applications in improving the comfort of shoes.

Keywords:
Plantar pressure imaging; level set; threshold -based segment; full convolutional networks; VCG; SegNet

1 Introduction

The goal of the segmentation of plantar pressure images is to improve the wearing comfort of the shoes. In order to design a comfortable footwear product, it is necessary to consider not only the human foot structure and biomechanical properties but also the personality differences; it is also necessary to consider the comfort aspect. In the shoemaking process, the design of the shoe last directly affects the wearing comfort of the adult shoe. Information of shape, force, and trajectory

of the foot of the human body when standing or walking can be obtained by analyzing and processing the pressure on the soles of the feet. In the study of plantar pressure analysis, the emergence of plantar pressure imaging method provides new technical means for the design of shoe last. By properly analyzing the plantar pressure, some “reasonable” plantar pressure distribution can be found to avoid foot discomfort. Further, the design of the sole component is improved by analyzing the pressure of the sole, so that the weight of the human body is distributed reasonably and evenly at various mechanical support points of the foot to enhance the stability of the walking process of the human body. On the other hand, the elasticity of the arch, the overall shape, structure and shape of the sole and the heel are also closely related to the comfort of the shoes. Among them, the image of the plantar pressure is divided until the data is reduced, hereafter, the region of interest (RoIs) of images are extracted.

Due to the actual need of image annotation and foot-segment semantic aggregation, deep learning method is used to accurately segment the plantar pressure image. Traditionally, the main purpose of using the transition region-based extraction method for image segmentation, and active contour segmentation is to reduce the complexity of image operations [1]. Among them, Nasreddine Kouadria et al. used the region-based segmentation method to reduce the image compression efficiency by 50% using the Discrete Tchebichef Transform (DTT) image compression technology based on RoI, and the transition region-based approach is a hybrid segmentation technique [2]. Nilamani Bhoi et al. initially decomposed grayscale images in the wavelet-domain and applied two existed transition region methods to the approximation. The coefficients are used to extract a transition region-based feature matrix. An inverse wavelet transform is applied to the modified coefficients to obtain an edge image, which has more than one-pixel width [3]. Traditionally, the Otsu threshold is used to obtain the transition region; the wavelet domain method is used to extract a stable transition region, resulting in better segmentation. In other hand, region growing method also is a simple region-based image segmentation method. Specifically, a seed pixel is found as a growth starting point for each region that needs to be segmented. Then, the seed pixel and the pixels in the surrounding neighborhood having the similar characteristics, and the seed pixels are merged into the new seed regions; wherein the similarity can be calculated by some pre-determined growing or similarity criterion. Finally, these new pixels are regarded as new seeds to continue the process until the pixels that do not meet the conditions can be included, so that a seeded based region is grown.

Digital image segmentation algorithms are generally based on two basic characteristics of gray values, namely discontinuity and similarity. The application of this property of discontinuity is to segment the target image based on discontinuous changes in image grayscale. The main application of this property of similarity is to segment the target image into similar regions according to the implementation of the specified criteria. The region growing algorithm is based on image similarity, which refers to the similarity of image gray values. Feng et al. proposed a hybrid thermal imaging reconstruction (TIR) and automatic seeded region growing (ASRG) to treat thermal imaging of carbon fiber reinforced composites, which can significantly reduce uneven illumination and increase detection rate. Additionally, it can automatically segment defects. It also overcomes the key issue of seeded region growing (SRG), which automatically selects images, seed points and threshold growth. Probability of detection (PoD) has been derived to measure the test results, and verified the efficacy of the proposed method [4]. In image analysis field, regional growth algorithms and their improved methods have many applications, such as the treatment of breast

cancer images [5] and the automatic segmentation of cochlear nerve regions [6].

The basic idea of level set method (LSM) is to regard the target as a zero level of a function in high dimensional space, and the evolution of the target is also extended to high dimensional space. Furthermore, the evolution is carried out to obtain the boundary of the image, to achieve the purpose of image segmentation [7]. Inthiyaz et al. extracted the Gabor color texture (GCT) and color-level covariance matrix (CLCM) texture features, constructed the color discrimination texture as a knowledge-base by using convex energy function, and extracted the edgeless image activity [8]. Yang et al. combined the level set evolution (LSE) and split Bregman methods to segment the magnetic resonance (MR) images and confirmed the advantages of the proposed method for the initial contour insensitivity and robustness to noise [9]. LSM-based segmentation methods can usually be grouped into region/edge-based methods. When the region descriptor is insufficient, the former often has problems in processing an image whose object has a color intensity like the background. When the edges of the image are weak, the latter often encounters boundary leakage problems. Therefore, Zhi et al. use the significance map and color intensity as the external energy of the region to stimulate the initial level set function (LSF), and further smooth the internal energy to identify more accurate boundary location; the experimental results illustrated that the newly saliency map improves the effect of extracting objects from complex backgrounds; and the asynchronous evolution of a single LSF produces better segmentation effects. This method has advantages of flexible initialization, robust and fast convergence [10].

Basically, threshold-based segmentation, region growing, and level set segmentation cannot accurately and strictly extract the image region according to semantics; semantic segmentation is from coarse to fine image classification-primary image understanding; furthermore, more elaborate inference and understanding of the image is needed, which provides an overall understanding of the entire picture. The important content, target location and detection are not only providing the categories within the image, but also creating the location information relative to the object category. Semantic segmentation is a dense prediction of each image pixel to obtain pixel class information [11]. Image semantic segmentation is increasingly researched, of them, loss functions and error definition in semantic segmentation are the main factors of image segmentation efficiency [12]. Ates et al. proposed a novel Markov random field (MRF) framework to improve the accuracy of image parsing [13]. Image analysis under the semantic level leads to automatic extraction of image descriptions based on human perception, finally bridges the semantic differences between low level image features and high-level concepts that capture conveyance [14].

Among the deep learning methods of semantic segmentation based on image domain, RefineNet, PSPNet and ResNet are relatively classic deep learning platforms [15-17]; in this case, the visual geometry group (VGG) platform is used for algorithm design and operation. VGG-Net explored the relations between the depth of the convolutional neural network (CNN) and its performance. By repeated stacking small convolution kernels with 3×3 and maximum pooling layers of 2×2, VGG-Net successfully constructed 16-19 layers deep CNN. VGG-Net has a significantly lower error rate than the previous state-of-the-art network architecture [18-19].

2 Methods

2.1 Preliminary

2.1.1 Threshold-based segmentation

Threshold-based segmentation (TBS) is a method of segmenting images by using grayscale threshold transform. The basic idea of threshold value segmentation is to determine a threshold value, and then compare the gray value of each pixel point with the threshold value, and divide the pixel into foreground or background according to the result of the comparison. The threshold segmentation can be divided into a certain threshold and a threshold pixel comparison. And pixel categorization; where the threshold is the most important; the choice of threshold will directly affect the accuracy of image segmentation, and the resulting image description and image analysis correctness. At present, the commonly used methods for threshold-based segmentation generally have experimental methods, that is, by the observation of the human, different thresholds are tested on images with known features, and then it is observed whether the known features have been met. The shortcoming of this method is that the scope of application is narrow, and some features of the image, such as average gray scale, must be fully understood before use; and the quality of the segmented image is greatly affected by subjective limitations. Secondly, the threshold is determined according to the histogram valley. If the gray histogram of the image has obvious double peaks, the valley between the two peaks can be selected as the threshold. Let pixel (x, y) 's value be $f(x, y)$, T is a given threshold, then we have that,

$$g(x) = \begin{cases} 255 & f(x, y) \geq T \\ 0 & f(x, y) < T \end{cases} \tag{1}$$

where, $g(x)$ is a binary image after threshold. This is a single threshold-based segmentation method, which should be more extensive; but the selection of T is difficult and the deviation is large.

The threshold selection iterative method is to select a certain threshold as the initial estimate and then continuously update this estimate according to a certain rule to achieve a given condition. The key to this process is to set the iterative rules to ensure that the system converges quickly, and the results of the iteration process are better than the last iteration. A typical threshold selection iteration method is described as follows,

Step 1: set an initial threshold T;

Step 2: segment image into two region R_A and R_B using T;

Step 3: calculate the average gray value for all pixels in regions R_A and R_B, as μ_A and μ_B;

Step 4: update the thresholds by $T^k = (\mu_A + \mu_B)/2$;

Step 5: go to **Step 2** until the T^k by successive iterations is smaller than the given parameter T.

The minimum mean square error (MSE) method uses the gray level in the image as the segmentation feature; it is assumed that each gray level is an independently distributed random variable, and it is assumed that the region to be segmented in the image obeys a normal distribution and one image contains only two main gray levels: foreground and background. Let z be the gray

value and $P(z)$ be the estimated value of the gray value probability density function. If the probability density function (PDF) relatives to the grays of one background and the other corresponds to the foreground in the image, i.e. the gray value of the object, the mixed density function describing the overall gray level transformation in the image is calculated by,

$$P(z) = P_1(z) + P_2(z) \tag{2}$$

where, P_1 and P_2 are the probability of the appearance of pixels with values z in the foreground and background respectively, and satisfies $P_1 + P_2 = 1$, that is, the image pixels either belong to the foreground or belong to the background. For a given threshold - T, the probability of pre-occurrence and background classification errors is:

$$E_1(T) = \int_{-\infty}^{r} (1 - P_1(z))dz \tag{3}$$

and, we have that,

$$E_2(T) = \int_{-\infty}^{r} (1 - P_2(z))dz \tag{4}$$

and

$$E(T) = P_1 E_1(T) + P_2 E_2(T) \tag{5}$$

If theoretically satisfies a normal distribution, i.e. a Gaussian distribution, then we have that,

$$P_i(z) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(z-\mu_i)}{2\sigma_i}} \ (i = 1, 2) \tag{6}$$

if $\sigma^2 = \sigma_i^2$, then we have that,

$$T = \frac{\mu_1 + \mu_2}{2} + \frac{\sigma^2}{\mu_1 - \mu_2} \ln\left(\frac{P_2}{P_1}\right) \tag{7}$$

if $P_2 = P_1$, then we simplify (7) as,

$$T = \frac{\mu_1 + \mu_2}{2} \tag{8}$$

When the image is under thresholding, the selected segmentation threshold should maximize the difference between the average gray level of the foreground and the background region. The average gray level of the entire image. This difference can be expressed by the variance of the region, which is derived based on the principle of decision analysis least squares.

Let the pixel with gray scale i in the image be n_i, gray be $[0, L-1]$, then $N = \sum_{i=0}^{L-1} n_i$, probability for each gray $P_i = n_i / N$, and subject to $\sum_{i=1}^{L-1} P_i = 1$. By using T, separates the image gray into $[0, T-1]$ and $[T-1, L-1]$, and denoted as C_0 and C_1 respectively, then calculate the probabilities $P(C_0) = \sum_{i=0}^{T-1} P_i$ and $P(C_1) = \sum_{i=T-1}^{L-1} P_i$, mean gray is $\mu_0 = \mu(T) / P_0$ and $\mu_1 = \mu(T) / P_1$, so the maximum interclass variance of the two regions can be calculated as,

$$\sigma^2 = P_0 P_1 (\mu_0 - \mu_1)^2 \tag{9}$$

As mentioned above, let T be in each of $[0, L-1]$ to make σ^2 of maximal T, the segment is the

best threshold. This method can be used not only for the selection of a single threshold containing two regions, but also for the multi-threshold selection calculation problem.

2.1.2 Seeded region growing

Some image segmentation methods are mainly realized by finding the boundary between regions based on the discontinuity of grayscale or color properties. The region growing segmentation algorithm is a technique for directly determining regions. For example, if the criterion is a pixel intensity threshold, then knowledge of the image histogram can be used as it can be used to determine an appropriate threshold for the regional membership criteria.

Using connected neighbors to grow from seed points, the standard is set to the same pixel value. That is, continue to check the neighboring pixels of the seed point, if they have the same intensity value as the seed point, classify them as seed points. This is an iterative process until there are no changes in two successive iterations. Surely, other criteria can be developed, but the main goal is to classify the similarity of the image as a region. Its value can indicate whether the region is inside the object or closer to the edge of the object. For example, the value of each pixel in the image is replaced by its relative position value. The greater the brightness, the closer the point is to the edge. Then, several important questions about region growing can be drawn: firstly, the seed points need to be chosen appropriately, and the choice of seed points depends on the user; secondly, more image information is needed; obviously, connectivity or pixel neighbor information helps determine thresholds and seed points; thirdly, it is the determination of the minimum region threshold, that is, in the segmented image, the region in the result of the region growing method is not less than the threshold.

The splitting and merging of regions are an attempt to divide the image into uniform regions. The basic representation structure is pyramidal. From a programming point of view, any two regions that are adjacent if they are adjacent can be merged, and if the larger region meets the homogeneity criteria. However, in the image space, adjacent regions may have different parent regions or be at different levels (different sizes in the pyramid structure). When there is no further merging, the process is automatically terminated; then it will find whether the adjacent regions have similar features, and if so, merge the similar regions again, and finally achieve the purpose of image segmentation. Although usually started with a single-region hypothesis, it can also start at an intermediate level, such as 16 regions or other numbers. In the latter case, four regions can be merged into a parent region. For the sake of simplicity, it starts a single region and then the entire image. The splitting process is to process a list of the current regions, that is, regions that are defined as uneven. A region can be removed from the processing list to the region list if it is homogeneous. The method of combining region growing and region splitting saves the process of splitting, and the method of region splitting and merging can be similarly merged based on a larger similar region; however, region growing can only be grown from a single pixel (merging). Algorithms that split and aggregate repeatedly to meet the constraints.

Step 1: for any given region C_i, if $P(C_i) = 0$, then split the region into 4 connected quadrant regions;

Step 2: aggregation any region m of $P(C_i \bigcup C_m) = 1$;

Step 3: stop if aggregation or cannot split;

Region growing is shown in **Algorithm** 1

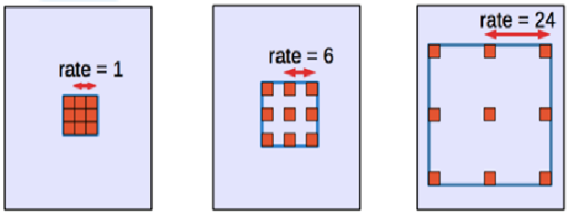
Algorithm 1 Region growing

```
INPUT: Image- I, Seed (x,y), maximal strength distance t =0.2
OUTPUT: region M
pixdist = 0; // The percentage of the distance from the latest pixel to the regional
average of the region
nb = [-1 0; 1 0; 0 -1;0 1]; //Neighbor location
// Start repositioning until the distance between the area and the positioning new
pixel becomes a certain threshold above a fixed point
WHILE (pix_dist<regmax_dist ||&&regsize<numel(I))
   FOR j=1:4,
      xn = x +nb(j, 1); yn = y +nb(j, 2); // Check neighbors inside or outside the
image
      ins=(xn>=1) &&(yn>=1) &&(xn<=I_sizes(1)) &&(yn<=I_sizes(2));
            // Add a neighbor (if in the staging area and not yet part)
      IF (ins&&(J(xn,yn)==0))
            negpos = negpos+1;
            neglist (negpos, :) = [xn yn I(xn,yn)]; J(xn,yn)=1;
      END IF
   END FOR
       IF (negpos+10>negfree)
          negfree=negfree+10000;
           neglist((negpos+1): negfree, :)=0;
 END
      //Add pixels with the intensity closest to the regional average to the region
dist = abs (neglist(1: negpos,3)-regmean);
   [pixdist, index] = min(dist);
   M (x, y) =2; regsize=regsize+1;
      // Calculate the new average of the area
   regmean= (regmean*regsize + neglist(index,3))/(regsize+1);
      // for the neighbor addition process
   x = neglist(index,1); y = neg_list(index,2);
      // Remove pixels from the neighbor (check) list
   neglist(index,:)=neglist(negpos,:); negpos=negpos-1;
END WHILE
// returns the partition region with logic matrix
M=M>1;
```

The region growing method can correctly separate regions with the same properties defined; it also can provide original images with sharp edges and good segmentation effects. The concept is simple, requiring only a small number of seeds to represent the desired attributes, then expanding the area and determining the seeds under the criteria assigned initially. Multiple criteria can be selected at the same time. However, there are also high computational costs, non-global view, and sensitivity to noise. Unless the image has a threshold and function applied, there may be a continuous path connecting colors that connect any two points in the image. Therefore, it is necessary to improve it and add a deep learning method to improve the efficiency of segmentation.

2.2 Improved full convolutional network employing SegNet

Semantic segmentation essentially performs a classification operation on each pixel of a picture. The current popular public data-set for semantic segmentation is VOC2012 and MSCOCO. Suitable traditional machine learning methods include pixel-level decision tree classification, Texton forests, and classification-based random forests. On the other hand, the current hot research is a combination of deep learning methods. A semantic learning method based on deep learning is initially popularized as a patch classification (PaC), which classifies central pixels by extracting neighborhood pixels pixel by pixel. Since the convolutional network ends at the time use full connected layers (FCL), this pixel-by-pixel segmentation method can only be used. Fully convolutional networks (FCN) removed the fully connected layer at the end. Subsequent semantic segmentation models basically adopt this FCN structure. In addition to the fully connected layer, another important issue of semantic segmentation is the pooling layer (PL). The pooling layer can further extract the abstract feature to increase the sensitive field (SF), but discard the position information of the pixel. Semantic segmentation requires the alignment of the category label and the original image. Therefore, the position information of the pixel needs to be reintroduced. There are two different architectures to solve the positioning problem of these pixels. The first one is the code-encode architecture. The encoding process gradually reduces the position information and extracts the abstract features through the pooling layer; the decoding process gradually restores the position information. There is a direct connection between the general decoding and the encoding. U-Net is the most popular in this type of architecture. The second architecture is dilated convolutions (DC), which discards the pooling layer, where the convolution kernel is 3×3. The rate is for the layer to guarantee the original receptive field, so the decision of the rate is determined by the stride value of the previous layer. Fig. 1 also shows the situation for different rates. There are many structural platforms for deep semantic segmentation, such as VGG-16 (FCN), GoogLeNet, FCN-8s, ENet bottlenek, Fast R-CNN+DeepMask, Own 3DCNN, C3D, etc.



(a) rate=1 (b) rate=6 (c) rate=24

Figure 1 Convolution operation feature map mapping process

The basic model uses the VGG-16 model, removing the FC layer, making the encoder network smaller and easier to train, while greatly reducing the size; the decoder network maps the lower-level pixels in the encoder network to the entire image; and the decoder network and the encoder network basically can be docked completely. Finally, a soft-max based multi-classification task can be performed for each pixel. The up-sampling of the decoder network uses the pool indices process, as shown in Fig. 2.

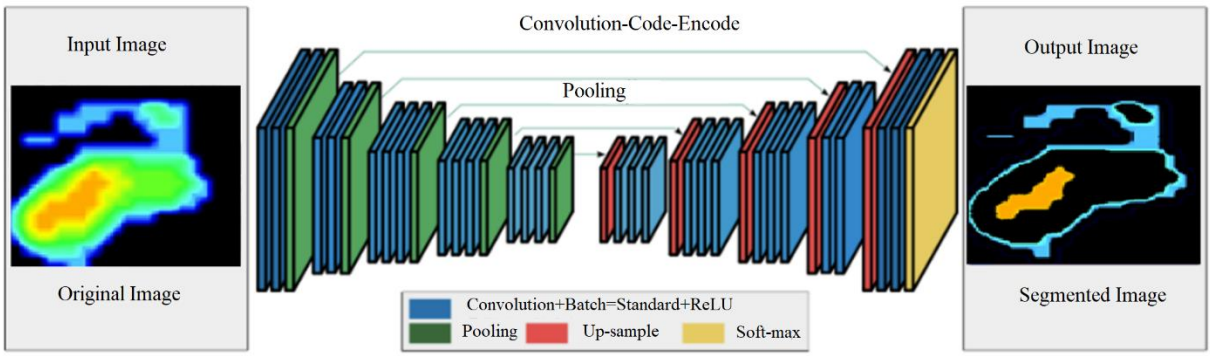


Figure 2 convolution-pooling-sampling code-encode

The model does not add parameters and does not need to participate in training and reducing memory. The decoder also does not need to store the output of the encoder, which can improve the expression of the boundary. The network structure can be extended to any encoder-decoder network structure. Fig. 3 shows the SegNet full convolutional layer segmentation method.

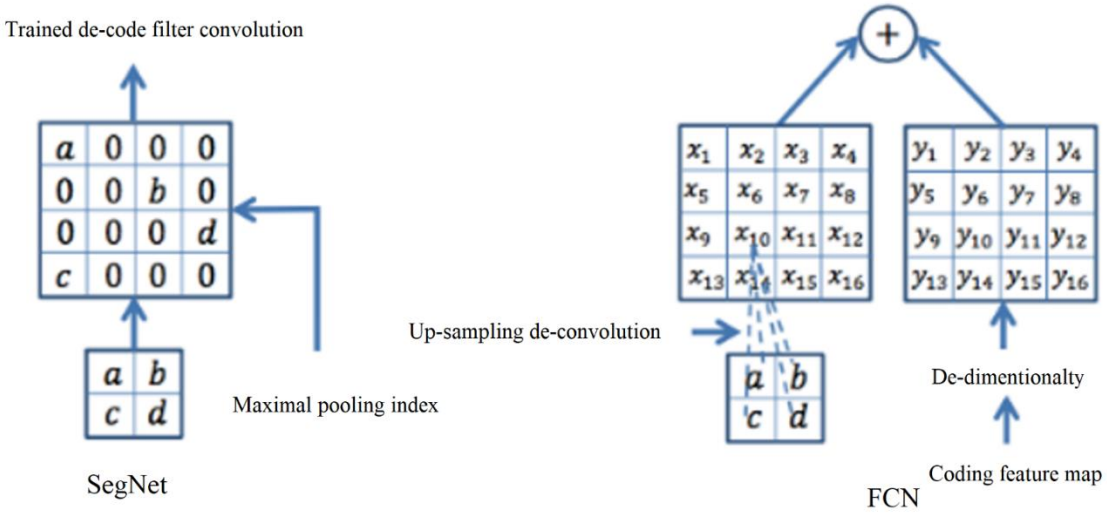


Figure 3 SegNet and FCN

Deep seeded region growing (DSRG), which is a network based on VGG-16 or Resnet-101, is also used to test the effectiveness of the method. Among them, VGG-16 changes the fully connected layer to the convolution layer, and classifies each pixel to realize classification prediction. The input to the convolutional network is an image, and the output is the segmentation result of the image. The values above the feature maps of the N H×Ws predict the probability that each pixel belongs to a certain category. After obtaining the seed region, it is now considered how to use the seed cues to train the image semantic segmentation network. Considering the uneven distribution of foreground and background seed regions, unlike the seed loss proposed in the seeding expansion and constrain loss (SECL) method, the balanced sowing loss proposed by this method (balanced sowing loss), two normalization coefficients with foreground and background respectively encourage the prediction of the segmentation network to match only the seeds [20]. The overall model is shown in Fig. 4.

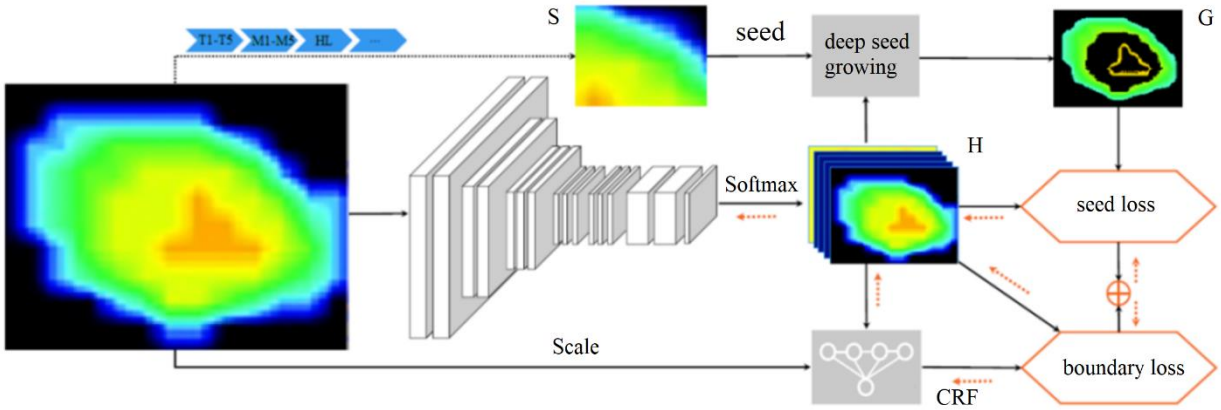


Figure 4 Application of depth seed region growth in the fundus pressure imaging set

Suppose C is the set of categories that exist in the image (except for the background), and c is the background. Suppose that S_c is a set of seed regions divided into c classes. Then, the definition of balanced seed loss can be calculated as follows,

$$\ell_{seed} = -\frac{1}{\sum_{c\in C}|S_c|}\sum_{c\in C}\sum_{u\in S_c}\log H_{uc} - \frac{1}{\sum_{c\in C}|S_c|}\sum_{c\in \bar{C}}\sum_{u\in S_c}\log H_{uc} \tag{21}$$

where, H_{uc} indicates the probability that the pixel at the position u in the segmentation graph H is predicted to be c class. In addition, the boundary loss of the Conditional Random Field (CRF) proposed in the SEC method is used to encourage the matching of the segmentation map with the object boundary ($\ell_{boundary}$). Finally, optimize the segmentation network by minimizing the loss function.

$$\ell = \ell_{seed} + \ell_{boundary} \tag{22}$$

By observing the obtained seed region, the seed region is positioned accurately, but it is very sparse. In practice, approximately 40% of pixels have labels. In order to improve the results of static supervision, iterative training can be used. In the training process, the seed region can be extended by the seed region growth method; each iteration uses the seed region extended after the last training as the initial seed region, and then training, and then expanding again based on the seed region until two The results of the iterative process are no longer updated. The theoretical basis for seed region growth is based on the presence of small, uniform regions in the image where the pixels should have the same label, i.e. pixels that are adjacent in position typically have the same category label. The method of expanding the seed firstly defines the similarity criterion P according to the relationship between the predicted probability value of the pixel in the segmentation map H generated by the segmentation and the set threshold value, and the specific formula is calculated as follows,

$$P(H_{uc},\theta_c) = \begin{cases} 1 & H_{uc} \geq 0 \; and \; c = \arg\max H_{uc'} \\ 0 & other \end{cases} \tag{23}$$

where, θ is a given probability threshold.

In experimental design, let threshold of foreground class be θ_b, and background class be θ_c firstly;

then perform segmentation H and seed region S as inputs to perform DSRG region growth. DSRG is an iterative process for each class, and the iterative growth process of class c is denoted as V_c, $C \in [0,|C|]$, where $c=0$ represents the background class. In each iteration of V_c, all locations in S_c will be accessed, and when pixel Q is accessed, the untagged pixel set in the eight-connected neighborhood of Q is represented as R. For $R_u \in R$, the probability of class c is expressed as $P(H_{uc})$ as described above. Then classify R_u according to P. Finally, after accessing all locations, all newly marked pixel heaps are added to S_c according to the rules to generate a new seed region. After changing S_c, the updated S_c will be accessed again. Otherwise, V_c stops. As the ability to segment the network increases, the number of unmarked pixels decreases and the object range is covered by the correct label. Finally, the update stops when S_c is no longer updated. The DSRG approach is basically superior to all previous techniques for weak supervision using image-level tags, and has achieved the best results on the current PASCAL VOC2012 dataset. Applying it to the semantic segmentation of the plantar pressure image to achieve a certain segmentation effect and comparison in SegNet+FCN 8s.

3 Result and Discuss

3.1 Data acquisition

Plantar pressure test and gait analysis is a biomechanical-based measurement technique that assesses and predicts possible future disease in the foot by detecting the structure of the lower extremity of the human body, providing internationally advanced techniques for scientific rehabilitation. The Foot-Scan device and gait analysis system developed by RS-Scan (RSscan International NV, Belgium) is a dynamic and static detection system for subjects. It can automatically detect the center of gravity, time measured and balance of the subject, a variety of pressure impulses and other data, and can be automated analysis system. Through the analysis of Foot-Scan software, abnormalities such as cerebral palsy and diabetic foot in gait of hip, knee and ankle can be found, which provides a scientific basis for clinical early diagnosis and corrective treatment [21-22]. The parameters for sensor and all testing device are listed in Table 1.

Table 1 RS-scan foot pressure data acquisition function parameters

Items	Description & Accuracy
size	1068 mm×418mm×12mm
number of sensors	8192（array with 128×64）
activity region of sensors	975mm×325mm
pressure	1 – 127 N/cm²+ /-0.05
temp	+15 °C - +30 °C
humidity	20% - 80%
length of Interface box cable	6000mm +/- 50mm
weight	7.7kg
size of sensors	7.62 mm×5.08 mm
type of sensors	resistive
frequency	<=500 Hz
scan accuracy	<=0.5mm
repeatable accuracy	<=0.02mm

A total of 60 volunteers (30 males, 30 females) older than 18 years old with normal non-related neurological diseases; no walking instability; abnormal gait; intermittent claudication and blurred vision; no joint serious illness; normal muscle strength; normal sputum reflex; no serious foot pain; no foot ulcers. Medical history, record age, gender, height, weight, blood pressure and other indicators are collected. Participants were asked to take off their socks, the habit of wearing shoes, fill out the relevant forms, and walk at normal speed for 10 repeated tests. The Foot-Scan system manually divides the plantar into 10 anatomical-based partitions and gives measurement data (discrete values), including the medial and lateral sides of the heel, the middle of the foot, the five metatarsals, the thumb, and the remaining four toes illustrated by Fig. 5.

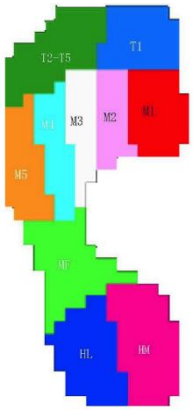


Figure 5 functional region segmentation in foot (HL-Heel Lateral, HM-Heel Medial, MF-Mid Foot, M1-5->Metatarsal 1- 5, T1-Hallux, T2-T5-Toe 2 to Toe 5)

The subject's walking process can be divided into 4 stages and 5 intervals.

(1) In the selection of the upper board method, “take the middle step” is adopted.

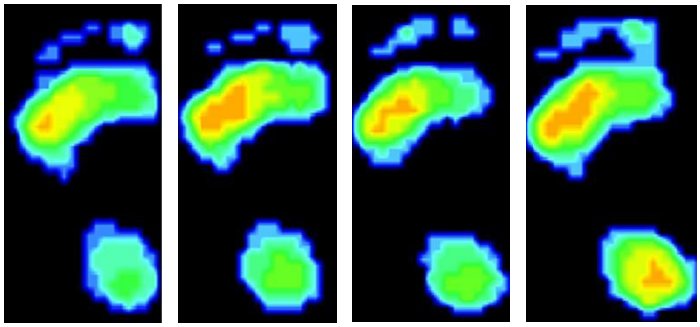
(2) In the selection of the number of tests, “take a step in the middle” to collect data 4-6 times. To obtain reliable “peak pressure” and “pressure-time integration”, “one step on board” needs to collect data at least 8 times, and “two steps on board” needs to collect data at least 5 times. The integration value needs to collect data at least 6 times. “Three steps on board” need to collect data at least 6 times.

(3) During the static balance test, pay attention to the control of the test time in four states (double open, double closed, single open, single closed). Standing on two feet is generally 20 seconds, and standing on one foot is generally 10 seconds.

(4) the image acquisition in each test is from 4 stages of initial foot contact (IFC), fore foot contact phase (FFCP), foot flat phase (FFP) and fore foot push off phase (FFPOP)

3.2 Process

To illustrate the segmentation effect, four foot-pressure imaging images were selected from a number of data-sets for segmentation (Fig. 6). To simplify the calculation, partial plantar pressure imaging data was selected. The image is first processed using the SegNet platform [23-24].



(a) (b) (c) (d)

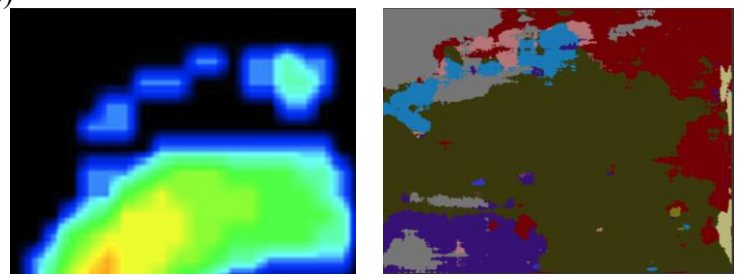
Figure 6 original image set ((a) left foot intermediate state 1 (b) left foot intermediate state 2 (c) left foot intermediate state 3 (d) left foot intermediate state 4)

The original image set is a left foot experiment and is shaped into a pulsed image in the format jpg. The imported split label is shown in Fig. 5. The operating status is shown in Table 3.

Table 3 SegNet platform training status parameters (average)

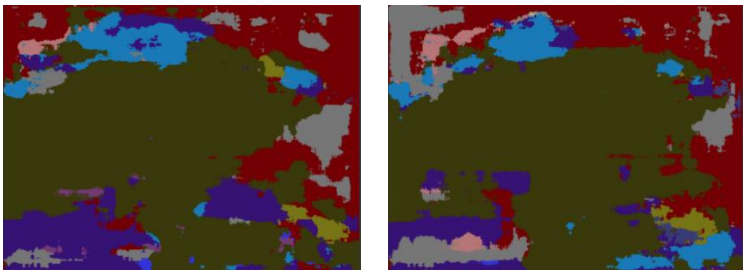
Iterations	**Time(s)**	**Mini batch accuracy**	**Mini batch loss**	**Basic learn rate**
1	00：00：07	41.22%	0.7870	0.0010
10	00：00：49	48.27%	0.8102	0.0010
15	00：01：12	58.25%	0.8185	0.0010
20	00：01：41	69.21%	0.8201	0.0010

The network structure layer [31×1], the connection number dimension is [34×2]; further, when the layer number dimension is [59×1], the connection number dimension is [66×2], the image size is $480\times640\times3$, the number of categories is 10; the coded depth is 4. After processing by the segNet platform, the phenomenon of over-segmentation is apparent, as shown in Figure 7. The forefoot is segmented (Fig. 7(a)), and the result image of Fig. 7(b) is obtained, and the over-segmentation result is generated. Further experiments are performed on other data sets, and over-segmentation is generated (Fig. 8.)

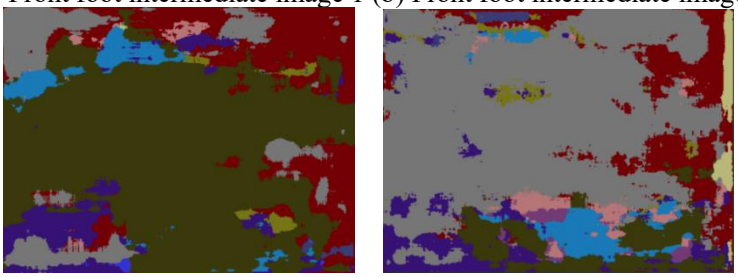


(a) Local original image of plantar pressure imaging (b) Segmentation result map

Figure 7 Partial foot pressure imaging segmentation processed by SegNet



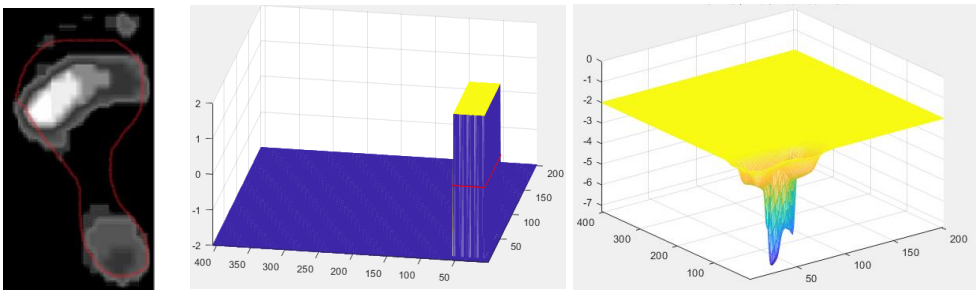
(a) Front foot intermediate image 1 (b) Front foot intermediate image 2



(c) Front foot intermediate image 3 (d) Rear foot intermediate image

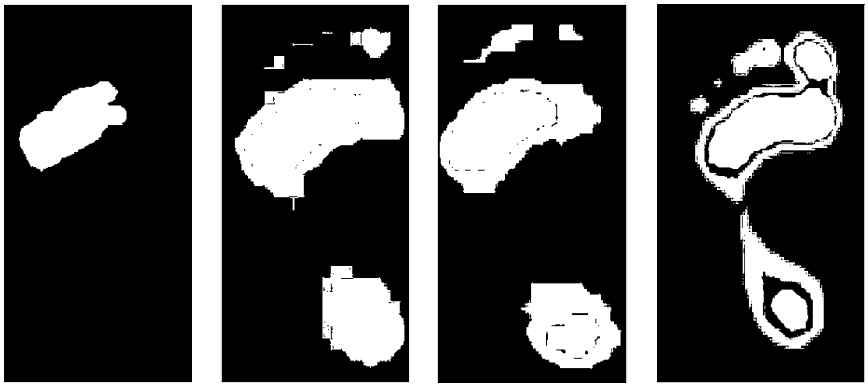
Figure 8 Over-segmentation results of the forefoot (forefoot, 1-3 subgraph) and the hindfoot (Rearfoot, 4th subgraph)

Among them, the boundary F1 index refers to the boundary F1 (BF) contour matching score indicating the degree of alignment of the prediction boundary of each class with the real boundary. More relevant to a person's qualitative assessment than IoU, BF score can be used [25]. To show the efficiency of the algorithm, the segmentation experiment was performed using the level set and the seed region growth method as shown in Fig. 9 and Fig. 10.



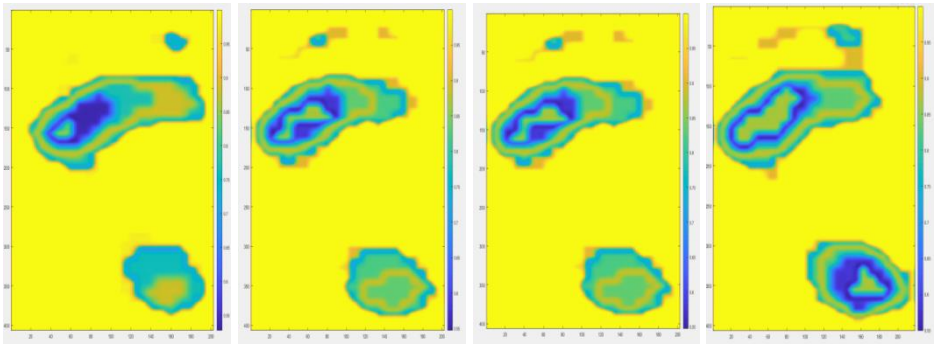
(a) Level set segmentation result graph (b) Initial level set function (c) 210 iteration level set functions

Figure 9 level set segmentation and effect



(a) initial level set (b) 50 iterations (c) 100 iterations (d) 210 iterations

Figure 10 Step-optimized single seed region growth segmentation method



(a) initial segmentation results (b) training intermediate state 1 (c) training intermediate state 2 (d) optimization segmentation results

Figure 11 SegNet+FCN 8s semantic segmentation based on VGG-16 platform

In order to illustrate the specific training process used, the segNet+FCN8s semantic segmentation of the plantar pressure image set was performed. First, this example creates a SegNet network that is initialized from the VGG-16 network. To get VGG-16, install the Deep Learning ToolboxTM model for the VGG-16 network. A CUDA-capable NVIDIATM GPU with strong computing power or a high version are suggested in the case. The parallel computing toolbox also is required, and the experimental segmentation results are shown in Fig. 11. The above methods are all domain-based segmentation methods. Not only to evaluate the segmentation effect of a single algorithm, but also to compare the segmentation effects of other kinds of deep learning, here are two aspects: one is the selection of contrast indicators, and the other is the calculation method of various indicators, including the setting of parameters. Table 4 shows the training state parameters of the SegNet+FCN 8s platform.

Table 4 SegNet+FCN 8s platform training status parameters (average)

Iterations	Time(s)	Mini Batch accuracy	Mini Batch Loss	Basic Learn Rate
1	00：00：07	47.05%	0.6852	0.0010
10	00：00：48	50.12%	0.6954	0.0010
15	00：01：11	59.51%	0.7002	0.0010

20	00：01：31	72.12%	0.7254	0.0010

4 Discussion & Statistical analysis

Convolutional neural networks can use a small amount of image preprocessing to learn abstract, essential, and deep-level image features from the original data. It has a better application of self-deep learning technology. The intelligent segmentation of images is becoming more and more sophisticated. When the number of hidden layers on the network increases to multiple layers, the powerful GPU computing capabilities have made major breakthroughs in many areas. Although the layer-by-layer initialization technology of deep neural networks needs to solve the problem of difficult training of deep networks, it can be better overcome through better preprocessing and area calculation. The convolutional neural network is a classic model generated by the combination of deep learning and image processing technology. The reason why the neural network can be combined with the image field and presents great development prospects.

The SegNet+FCN 8s parameter is set to be up-sampled by bilinear interpolation, and the parameters are fixed, that is, they do not participate in learning; the max-pooling indices are used for up-sampling, that is, they do not participate in learning; the bilinear interpolation is used for up-sampling, that is, parameter participation. Learning; use bilinear interpolation for initialization. SegNet-Basic uses an FNC-like decoder method; four encoders, four decoders, and up-samples use down-samples (indics) to connect a BN operation after each conv on the encoder/decoder. For the decoder network, the ReLU non-linear activation function and bias are not used in the conv. With a 7×7 convolution kernel, the receptive field of VGG layer 4 is 106×106, and the number of filters convolved by decoder is the same as the number of filters convolved for one encoder. SegNet-single-pass first decoding; decoder convolution kernel number is 1; the final dimension corresponds to the corresponding encoder; the conclusion is that the FN-Basic-NoDimReduction method works best.

Combined with SegNet's FCN with 8s, the semantic segmentation of level set, domain value segmentation and deep learning is evaluated, as shown in Table 5. By comparison, the proposed SegNet+FCN 8s segmentation method has global accuracy (0.8105), average accuracy (0.8015), average cross-ratio (negative index) (0.6110), and boundary F1 index (negative index) (0.6200). Both are superior to the other three methods: distance regularization level set, single seed region growth segmentation, and single segNet segmentation. Only in the weight-to-weight ratio (0.7855) is slightly inferior to single seed region growth segmentation and single SegNet segmentation; it is proved that SegNet and FCN have higher efficiency for segmentation of plantar pressure imaging.

Table 5 Evaluation index and effect of segmentation of foot pressure imaging by various network platforms

Seg. Methods	Global accuracy [26]	Mean accuracy [26]	mIoU	WIoU	F1 score
DRLSE	0.7012	0.7214	0.6251	0.7952	0.4751
Single seed region growth	0.7121	0.7254	0.7089	0.7245	0.5822

Single SegNet	0.7825	0.7721	0.6852	0.7851	0.5247
SegNet+FCN 8s	0.8105	0.8015	0.6110	0.7855	0.6200

(mIoU: Mean Intersection over Union; WIoU: Weighted Intersection over Union; F1 score: Sørensen–Dice coefficient)

5 Conclusions

From the deep learning technology, the emergence of deep neural networks makes the intelligent segmentation of images more and more fine. When the number of hidden layers of the network increases to multiple layers, the powerful GPU computing power has made major breakthroughs in many fields. Although the layer-by-layer initialization technology of deep neural network needs to solve the problem of deep network training difficulty, it can be better overcome by better preprocessing and regional calculation. The convolutional neural network is a classic model produced by the combination of deep learning and image processing technology. The reason why neural network can be combined with the image field has a huge development prospect.

This paper theoretically studies the model of the plantar pressure imaging data-set segmentation, including traditional domain value segmentation techniques, level set theory and its variants, such as distance regularization level set evolution technology, and regional growth method. Through experiments and evaluation of the segmentation effect, mainly through global accuracy, average accuracy, average cross-over ratio and so on. The proposed deep learning network-SegNet network combined with the fully connected layer network has higher segmentation efficiency for the segmentation of the foot-pressure imaging data set.

Research based on the plantar pressure data set will be further used for rehabilitation-related basic research using foot pressure for dynamic image analysis can be further developed in clinical applications. Medical research has shown that diseases such as diabetes, obesity, and Parkinson's disease can cause changes in plantar pressure. This cross-study on the customization of functional shoes and precision technology will bring shoe customization and comfort shoe design. Segmentation models based on deep neural networks are not only used in the research of plantar pressure imaging datasets. It plays a significant role, and for some small-scale, but high-quality data sets such as medical image data, including CT, magnetic resonance, etc.

In future works, the various segmentation algorithms involved in the research still need to be further improved. The training set size and training parameter settings of the full convolutional network need to be further optimized. In addition, it is also very important to use these key point sets obtained from plantar pressure imaging data to generate the required comfortable plantar surface.

References

[1] Parida P. Fuzzy clustering-based transition region extraction for image segmentation. Future Computing and Informatics Journal 2018; 3(2): 321-333.

[2] Kouadria N, Mechouek K, Harize S, et al. Region-of-interest based image compression using the discrete Tchebichef transform in wireless visual sensor networks. Computers & Electrical Engineering 2019; 73: 194-208.

[3] Parida P, Bhoi N. Wavelet based transition region extraction for image segmentation. Future Computing and

Informatics Journal 2017; 2(2): 65-78.
[4] Feng Q, Gao B, Lu P, et al. Automatic seeded region growing for thermography debonding detection of CFRP. NDT & E International 2018; 99: 36-49.
[5] Punitha S, Amuthan A, Joseph K S. Benign and malignant breast cancer segmentation using optimized region growing technique. Future Computing and Informatics Journal 2018; 3(2): 348-358.
[6] Jeevakala S, Brintha Therese A, Rangasami R. A novel segmentation of cochlear nerve using region growing algorithm. Biomedical Signal Processing and Control 2018; 39: 117-129.
[7] Liu Y, He C, Gao P, et al. A binary level set variational model with L1 data term for image segmentation. Signal Processing 2019; 155: 193-201.
[8] Inthiyaz S, Madhav B, Kishore P. Flower image segmentation with PCA fused colored covariance and Gabor texture features based level sets. Ain Shams Engineering Journal 2018; 9(4): 3277-3291.
[9] Yang Y, Tian D, Jia W, et al. Split Bregman method-based level set formulations for segmentation and correction with application to MR images and color images. Magnetic Resonance Imaging 2019; 57: 50-67.
[10] Zhi X, Shen H. Saliency driven region-edge-based top down level set evolution reveals the asynchronous focus in image segmentation. Pattern Recognition 2018; 80: 241-255.
[11] Fahad Lateef, Yassine Ruichek, Survey on semantic segmentation using deep learning techniques, Neurocomputing 2019; 338:321-348.
[12] Garcia-Garcia A, Orts-Escolano S, Oprea S, et al. A survey on deep learning techniques for image and video semantic segmentation. Applied Soft Computing 2018; 70: 41-65.
[13] Ates H F, Sunetci S. Multi-hypothesis contextual modeling for semantic segmentation. Pattern Recognition Letters 2019; 117: 104-110.
[14] khodaskar A, Ladhake S. Semantic Image Analysis for Intelligent Image Retrieval. Procedia Computer Science 2015; 48: 192-197.
[15] Kemker R, Salvaggio C, Kanan C. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. ISPRS Journal of Photogrammetry and Remote Sensing 2018; 145, Part A: 60-77.
[16] Hofbauer H, Jalilian E, Uhl A. Exploiting superior CNN-based iris segmentation for better recognition accuracy. Pattern Recognition Letters 2019; 120: 17-23.
[17] Zhang P, Liu W, Wang H, et al. Deep gated attention networks for large-scale street-level scene segmentation. Pattern Recognition 2019; 88: 702-714.
[18] Dhomne A, Kumar R, Bhan V. Gender Recognition Through Face Using Deep Learning. Procedia Computer Science 2018; 132: 2-10.
[19] Chavan T R, Nandedkar A V. AgroAVNET for crops and weeds classification: A step forward in automatic farming. Computers and Electronics in Agriculture 2018; 154: 361-372.
[20] Fang H, Lu G, Fang X, et al. Weakly and Semi Supervised Human Body Part Parsing via Pose-Guided Knowledge Transfer. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018: pp. 70-78.
[21] Dan Wang, Zairan Li, Nilanjan Dey, Amira S. Ashour, Luminita Moraru, Anjan Biswas, and Fuqian Shi Optical pressure sensors based plantar image segmenting using an improved fully convolutional network, Optik - International Journal for Light and Electron Optics2019; 179: 99-114
[22] Zairan Li, Dan Wang, Nilanjan Dey, Amira S. Ashour, R. Simon Sheratt, and Fuqian Shi, Plantar pressure image fusion for comfort fusion in diabetes mellitus using an improved fuzzy hidden Markov model, Biocybernetics and Biomedical Engineering 2019; 39: 742-752,
[23] Lye K, Chua K, Ko C. Performance of SegNet - a simulation study. Computer Communications 1987; 10(6): 297-303.
[24] Al-masni M A, Al-antari M A, Choi M, et al. Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. Computer Methods and Programs in Biomedicine 2018; 162: 221-231.
[25] Yating Chen, Gaoxiang Chen, Yu Wang, Nilanjan Dey, R. Simon Sherratt, and Fuqian Shi, A Distance Regularized Level-Set Evolution Model Based MRI Dataset Segmentation of Brain's Caudate Nucleus, IEEE Access 2019; 7: 124128-124140
[26] Rodrigues É, Conci A, Liatsis P. Morphological classifiers. Pattern Recognition, 2018, 84: 82-96.