On the segmentation of plantar foot thermal images with Deep Learning

Asma Bougrine, Rachid Harba, Raphael Canals, Roger Ledee, Meryem Jabloun

Univ. Orleans, PRISME, EA 4229, F45072, Orleans, France;

firstname.lastname@univ-orleans.fr

Abstract—Foot ulceration can be prevented by using thermal information of the plantar foot surface. Indeed, important indicators can be provided with a thermal infrared image. As part of a non-constraining acquisition protocol, these images are freehandedly taken with a smartphone equipped by a dedicated thermal camera. A total of 248 images have been obtained from an acquisition campaign composed of control and pathological subjects. Our aim is the segmentation of these plantar foot thermal images. To that end, we compare three different deep learning methods namely, the Fully Convolutional Networks (FCN), SegNet, U-Net, and the previously proposed prior shape active contour-based method. 80% of our database serves to train the 3 deep learning networks and 20% are used for the test. When applied to our data, results show that the SegNet method outperforms the three other methods with a Dice Similarity Coefficient (DSC) equal to 97.26%. This method also shows efficiency in segmenting both feet simultaneously with a DSC equal to 96.8% for a smartphone based plantar foot thermal analysis for diabetic patients.

Index Terms—Plantar foot thermal images, Deep Learning, prior shape active contour, image segmentation.

I. INTRODUCTION

Diabetic foot is a common disease among diabetic patients leading to foot ulceration which is the primary reason for diabetic related foot amputations. The occurrence of foot ulcer is often associated with foot hyperthermia. In [1], foot hyperthermia was defined as a temperature difference higher than 2.2°C between a foot region and the same region on the contralateral foot. According to [1], foot ulcer occurrence can be reduced by 70% if the foot hyperthermia is early detected. This information can be identified using a thermal infrared camera, as reported in several studies [2]–[5]. In fact, most of these studies require a constraining acquisition protocol which imposes on the person who participates in the acquisition to put his feet in a special device that hides all other thermal sources except those coming from the plantar foot. This process ensures good image quality and therefore a low level of segmentation complexity. Our aim is to develop new mobile and user-friendly technology to precisely analyze the plantar foot temperature in diabetic foot problem. We, therefore, free ourselves from using a complex and constraining isolation system unlike [3]–[6]. Images will be freehandly taken with a smartphone equipped by a dedicated thermal camera. The full automatic processing of the data is to be performed in the smartphone itself. It is expected that this overall protocol can be generalized in a clinical routine or even at home. The first step of this processing is a fully automatic segmentation of the plantar foot surface. Thus, the automatic segmentation of such images is difficult, since the occurrence of other thermal sources of the body. Classical segmentation methods fail to segment these specific images as demonstrated in a recent work [7]. We, therefore, presented a prior shape based active contour method that proposes to modify the snake functional of Kass *et al.* [8] by adding an extra energy term. This term guides the snake to the desired contour by minimizing a curvature difference between the snake curve and the prior shape curve. In previous work [9], we compared the proposed method to two prior shape-based active contour models. The first method is the Ahmed *et al.* method [10] which assesses the shape matching performed directly in the Fourier descriptor space. The second method is proposed by Chen *et al.* [11]; authors proposed to find the transformations (scale, rotation, and translation) such as the prior shape is closely associated with the transformed curve. When applied to our database of 50 plantar foot thermal images, results show that our proposed method outperforms the two others. The major problem with these methods is their sensitivity to the position of the initial contour. Even though the previosly proposed method is less sensitive, an automatic initialization process has to be performed to obtain a better result. Moreover, the characteristics of the images are different from person to person. Let's take the example of a healthy (or not) person with toes that are always cold. In this case, active contour-based methods may fail to find the good foot contour even with an imposed shape constraint. So that, more powerful and robust segmentation methods are required. In recent years, deep learning techniques have proven spectacular progress. Initially intended for image classification [12]–[15], they are more and more applied to a wide variety of other tasks, in particular for semantic segmentation. To the best of our knowledge, no studies have applied deep learning for diabetic plantar foot thermal images segmentation.

In the present paper, we propose to select the most suitable method for plantar foot thermal image segmentation. Thus, we compare three different deep learning networks namely, the Fully Convolutional Networks (FCN) [16], the U-Net [17] and SegNet [18]. These networks are the most successful state-of-the-art deep learning techniques for semantic segmentation. To train these networks, we launched a data acquisition campaign within the Diabetic Foot Service of the Regional Hospital of

Orleans. 198 pathological plantar foot images were acquired. A total of 248 images have been obtained thanks to our previous database of healthy persons. 80% of the total servers in the training of our three networks with a data augmentation operation. 20% is used as a test database for all compared methods. Results given by these three networks are then compared to segmentation results given by the previously proposed prior shape active contour method when applied on our test database.

The remainder of the paper is organized as follows. In Section 2, we describe our acquisition campaign and the preprocessing of our database. Section 3 details the segmentation methods, the deep learning networks, and the proposed prior shape-based active contour method. In section 4, the choice of parameters of each method is detailed. Then, qualitative and quantitative results given by the different methods are carried out to select the most suitable method for plantar foot thermal image segmentation. We test also the robustness of the best model when segmenting two feet simultaneously. Finally, conclusions and perspectives are presented in the last section.

II. MATERIALS

In this section, we first elaborate the choice of the smartphone thermal camera, describe the control group and the diabetic foot database. Then, we detail the preprocessing of the plantar foot thermal images.

A. Thermal image database

1) The chosen camera: is the FlirOne Pro thermal camera design to be plugged to a smartphone. This camera has a thermal image resolution of 160x120 pixels and a thermal sensor spectral range of 8-14 μm. FlirOne Pro can detect temperature differences of 0.1°C which is enough to detect the possible hyperthermia variations of interest.

2) Control group database: is composed of 25 healthy (non-diabetic) persons who participated in the first acquisition campaign [7]. This sample group was composed of 10 female and 15 males of staff members of Orleans University with a mean age of 34. The acquisition protocol detailed in [7] was respected.

3) Diabetic foot group database: is composed of 36 diabetic persons that participated in our acquisition campaign within the Diabetic Foot Service of the Regional Hospital of Orleans. This sample group was composed of 10 female and 26 males with a mean age of 69. This group contained diabetic foot persons who may have foot ulcers. We exclude people who have amputated foot parts. The participants have been followed over a period of 2 months. A patient may have been part of acquisition more than one time. The same acquisition protocol was respected. A total of 99 images have been acquired.

B. Database preprocessing

Images from the FlirOne Pro are in JPEG format of 640×480 pixels. Labels are displayed at the bottom and a temperature bar is displayed on the right side of the image. For all images of our database (control and diabetic foot group), the temperature bar and the labels are first removed from the image. Then, the left foot image is separated from the right foot image by splitting the image into two equal parts. The left foot is then flipped to obtain the same orientation between both foot images. The database is therefore composed of a total of 248-foot images with the same orientation. These images were manually annotated by an expert and correspond to the ground truth maps. 20% of this database corresponding to 50 images serves to validate results and to compare the methods detailed above. 80% of the database is used to train the deep learning models. This train database has occurred a data augmentation thanks to three rotations and two contrast changes to guarantee the training performance of our networks. In our context of foot segmentation, these transformations applied to the data are justified and do not change the relevant information in the image; on the contrary, they allow to make the model more robust. Finally, the overall training data is equal to 1134 images.

III. METHODS

We here review the methods to be applied to our segmentation problem. Deep learning methods are first detailed, and then we present the previously proposed prior shape-based active contour method.

A. Deep Learning methods

Deep learning algorithms have solved a variety of computer vision problems including semantic segmentation. This task consists of classifying each pixel of the image into an object instance. Deep Learning models, such as Convolutional Neural Networks (CNNs), have greatly contributed to the increase of performance on this field. CNNs are made up a series of convolution layers which extract different characteristics from the images, and fully connected layers based on the multilayer perceptron to classify an object. Semantic segmentation not only requires discrimination at pixel level but also to project the discriminative feature onto the pixel space. Thus, all semantic segmentation networks are built on two main phases. First, the encoder (downsampling) network which corresponds to a pre-trained classification network such as vgg-16 [14], ResNet [13], AlexNet [15], followed by a decoder (upnsampling) network which allows the discriminative projection of the pixel onto the pixel space. The most popular initial deep learning approaches are the Fully Convolutional Networks (FCN) and the SegNet network as they take advantage of existing CNNs as powerful models, transform them from networks which purpose is classification to make suitable for segmentation. Another example of networks, U-Net, initially proposed for biomedical image segmentation. Here we detail these three networks.

1) Fully Convolutional Networks (FCN): has been proposed by Long *et al.* [16]. Authors transformed the existing and well-known classification models into fully convolutional ones by replacing the fully connected layers with convolutional ones using 1x1 convolution. Instead of producing classification

scores as output, spatial maps are generated. In low resolution, the process produces a class presence heatmap which is up-sampled using transposed convolutions. Furthermore, the up-sampling process is, at each stage, enhanced by concatenating predicted characteristics maps from the downsampling path. In addition, to completely recover the lost spatial information in the downsampling layers, skip connection is also introduced following each convolution block.

2) SegNet: has been proposed in 2015 [18]. The encoder network of SegNet is composed of 13 convolutional layers that represent the first 13 convolutional layers of the vgg-16 network [14]. The training process is then initialized using the trained weights for classification on large data sets. The fully connected layers are eliminated and replaced by higher resolution feature maps at the deepest output of the encoder. The 13 encoder layers have their corresponding 13 decoder layers. The decoder stage of SegNet is composed of a set of upsampling and convolution layers. Each upsampling layer in the decoder stage corresponds to a max-pooling one in the encoder part. Those layers upsample feature maps using the max-pooling indices from their corresponding feature maps in the encoder phase. The upsampled maps are then convolved with a set of trainable convolutional filters to produce dense feature maps. Once the feature maps have been restored to the original resolution, they are sent to the softmax classifier to produce the final segmentation.

3) U-Net: has been proposed by O. Ronneberger *et al.* [17]. Authors have extended the FCN [16] for biological images. The encoder part of U-Net has an FCN-like architecture which extracts characteristics with 3x3 convolutions. The decoder part uses deconvolution that reduces the number of feature maps while increasing their dimensions. The feature maps cropped from the encoder part of the method are copied into the decoder to avoid losing the pattern information. In the end, the feature maps are processed with a 1x1 convolution to produce the segmentation map and finally categorize each pixel of the input image with the softmax classifier.

B. The previously proposed prior shape based method

We have proposed to modify the snake functional of Kass *et al.* [8] by adding prior shape energy. This extra energy function assesses the normalized difference between the curve curvature and the prior shape curvature during the contour evolution. It imposes the shape of the foot during the snake modifications. The total energy of the proposed model is the sum of four different energy functions including the prior shape energy E_{PS}, the internal E_{intern}, image E_{image} and external E_{con} ones. The internal energy contains two terms: length and curvature. The image energy is given by the gradient information. The external constraint is the balloon energy.

$$E_{PS} = \gamma |C_{ss}(s) - \zeta C_{ss}^*(s)|^2, \quad (1)$$

$$E_{intern} = \alpha |C_s(s)|^2 + \beta |C_{ss}(s)|^2, \quad (2)$$

$$E_{image} = -W_{edge} |\nabla I(C)|^2, \quad (3)$$

$$E_{con} = \delta |\eta(C)|^2, \quad (4)$$

where C and C^* are the snake curve and the prior shape curve, respectively. C and C^* depend both on the curvilinear abscissa s $\in[0, 1]$ and on time t. γ is the weight of the prior shape energy and $\zeta = \frac{|C_{ss}(s)|}{|C_{ss}^*(s)|}$ is the normalization factor. $C_s = \frac{\partial C}{\partial s}$, $C_{ss} = \frac{\partial^2 C}{\partial s^2}$, $\eta(C)$ is the outward unit norm to the curve C. The parameters α and β control the internal energy whereas E_{image} and E_{con} depend on parameters W_{edge} and δ respectively. As the initial contour has to grow, we choose δ to be a positive constant in order to make the balloon growing. Minimizing the total energy function of the model leads to solving the Euler Lagrange equation:

$$\alpha C_{ss} + (\beta + \gamma)C_{ssss} - \gamma\zeta C_{ssss}^* + \frac{\partial(E_{image} + E_{con})}{\partial C} = 0, \quad (5)$$

where $C_{ssss} = \frac{\partial^4 C}{\partial s^4}$ and $C_{ssss}^* = \frac{\partial^4 C^*}{\partial s^4}$. The proposed method is designed to be invariant to scale, rotation and position difference between the active contour and the prior shape one. In fact, during the snake evolution, both snake and prior shape curve are examined from the same initial point (the lowest point of the calcaneus). As a result, the method is invariant to rotation. The prior shape energy is a function of the relative positions of C and C* which avoids updating the translation parameter. Finally, we used normalized curvatures, the ζ term in equation 5, which avoids updating the scale factor.

IV. RESULTS

In this section, we first give the set of parameters needed for each method. Next is qualitative and quantitative results given by the compared methods when applied to our test database.

A. The choice of parameters

The chosen parameters are those that have given the best results after several tests. The three deep learning methods are all based on the pre-trained network model VGG-16 [14]. The encoder depth; which is the number of downsampling (or upsampling) operations, is chosen equal to 5 as the default value. We train the networks using stochastic gradient descent with momentum optimizer (SGDM). The momentum is defined as the contribution of the previous gradient step to the current iteration. We have chosen the initial value of this parameter equal to 0.9 assuming that this value works well for most problems. Effectively, this value is suitable to FCN and segNet, unlike U-Net which gave better scores with a momentum equal to 0.3. We set the initial learning rate equal to 0.01 for FCN and SegNet which is the default value for the SGDM solver. Again, a different value is used with U-Net equal to 0.0001. The stochastic gradient descent algorithm evaluates the gradient and updates the parameters using a subset of the training set (a mini-batch). The size of this mini-batch is chosen to be 4 images for the 3 networks. When the learning algorithm completely passes over the learning data set, an epoch is performed. We set the maximum number of epochs to 100. The set parameters of the prior shape snake method is α= 0.1, β= 4, W_edge= 20, δ= 0.2 and γ= 35.

B. Qualitative results

We qualitatively evaluate the performance of the 3 tested deep learning models and our prior shape active contour method. Results are presented in Fig.1 for 3 images selected from the test database. Results show that SegNet and FCN methods give both qualitatively good results compared to the prior shape-based method and the U-Net network for the 3 tested images.
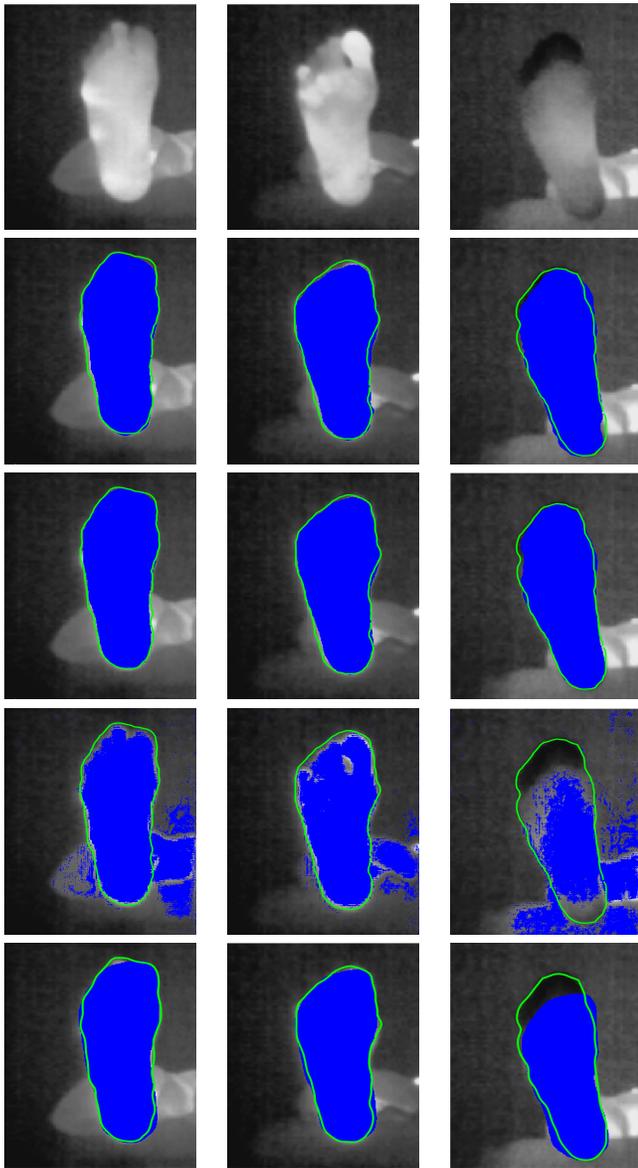

Fig. 1. Segmentation of 3 plantar foot thermal images (first row). The second row corresponds to FCN network, while the third one shows the SegNet networks results. Results given by U-Net network are in the fourth row and our prior shape active contour-based method is presented in the last row. The blue regions correspond to the segmentation results found by the methods while the green curve represents the ground truth manually annotated by the expert.

C. Quantitative results

We use the Dice Similarity Coefficient (DSC) [19] as an evaluation metric. This score assesses the similarity between the foot region given by the ground truth contour and the regions bounded by the segmentation regions given by the prior shape active contour-based method and the three deep learning networks. Table I shows the mean scores of DSC given by the methods with their respective standard deviations (STD). We clearly notice that U-Net network has not given acceptable segmentation results. We tested with several momentum and initial learning rate values. The number of epochs has also been increased to improve segmentation results, but without success. This may be due the fact that the size of the training database is not large enough for that network to learn correctly. The best segmentation result is given by the SegNet method with a DSC equal to 97.26% not too far from the result given by the FCN method.

Since our aim is to develop an automatic mobile application for the precise visualization and analysis of plantar diabetic foot temperature, the real-time processing time is important. We go further and we propose segmenting both the right and left foot simultaneously, only using the SegNet model which has learnt to segment a single foot in the image. To do this, we test SegNet model on the acquired plantar foot thermal images that contain both feet. Qualitatively SegNet succeeds in locating and segmenting the image correctly. Fig. 2 depicts segmentation results given on two different images from our test database. The mean DSC of 25 test images each containing two feet is 96.8%.
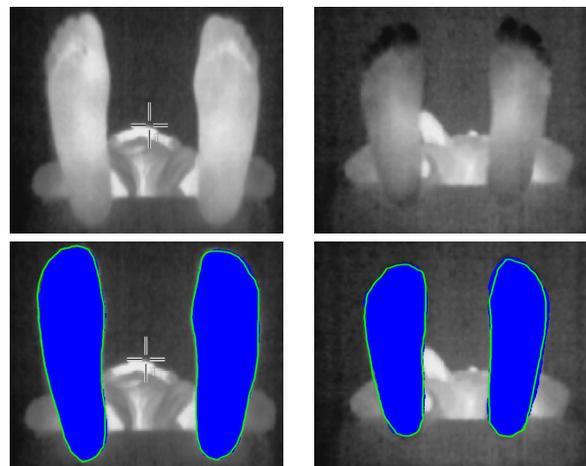

Fig. 2. Segmentation of 2 plantar foot thermal images (first row) using SegNet method. In the second row images, the blue regions correspond to the segmentation results while the green curve represents the ground truth contours.

TABLE I
DSC (±STD) OF THE SEGMENTATION METHODS.

FCN	96.16% ±0.85%
SegNet	97.26% ±0.69%
U-Net	74.35% ±9.58%
Prior shape method	94% ±2%

V. Conclusion

In the present paper, three different deep learning methods are compared to a prior shape active contour-based method for segmentation of plantar foot thermal images. We presented a new database of diabetic persons acquired within the Diabetic Foot Service of the Regional Hospital of Orleans. This campaign allows us to collect 248 plantar foot images. The comparison of these four methods carried out on our test dataset of plantar foot thermal images showed the superiority of the SegNet network with a DSC of 97.26%. This method is efficient and robust and showed its efficiency in segmenting both feet simultaneously with a DSC equal to 96.8%. This makes SegNet method suitable to a smartphone based application for plantar foot thermal analysis for diabetic patients. As a conclusion, no constraining isolation system will be needed, images will be freehandedly taken with a smartphone equipped by a dedicated thermal camera, and the fully automatic processing of the data will be performed. We intend to apply this protocol in clinical routine for the prevention of foot ulcers in diabetics based on hyperthermia identification.

Acknowledgment

References

[1] David G Armstrong, Katherine Holtz-Neiderer, Christopher Wendel, M Jane Mohler, Heather R Kimbriel, and Lawrence A Lavery. Skin temperature monitoring reduces the risk for diabetic foot ulceration in high-risk patients. *The American journal of medicine*, 120(12):1042–1046, 2007.

[2] Jaap J van Netten, Jeff G van Baal, Chanjuan Liu, Ferdi van Der Heijden, and Sicco A Bus. Infrared thermal imaging for automated detection of diabetic foot complications. 2013.

[3] Luis Vilcahuaman, Rachid Harba, Raphaël Canals, M Zequera, Carlos Wilches, MT Arista, L Torres, and H Arbañil. Detection of diabetic foot hyperthermia by infrared imaging. *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4831–4834, 2014.

[4] Yang Liu, Andrea Polo, M Zequera, Rachid Harba, Raphaël Canals, Luis Vilcahuaman, and Y Bello. Detection of diabetic foot hyperthermia by using a regionalization method, based on the plantar angiosomes, on infrared images. *2016 IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)*, pages 1389–1392, 2016.

[5] Luay Fraiwan, Mohanad AlKhodari, Jolu Ninan, Basil Mustafa, Adel Saleh, and Mohammed Ghazal. Diabetic foot ulcer mobile detection system using smart phone thermal camera: a feasibility study. *Biomedical engineering online*, 16(1):117, 2017.

[6] A Astasio-Picado, E Escamilla Martínez, A Martínez Nova, R Sánchez Rodríguez, and B Gómez-Martín. Thermal map of the diabetic foot using infrared thermography. *Infrared Physics & Technology*, 93:59–62, 2018.

[7] Asma Bougrine, Rachid Harba, Raphael Canals, Roger Ledee, and Meryem Jabloun. A joint snake and atlas-based segmentation of plantar foot thermal images. *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6, 2017.

[8] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988.

[9] Asma Bougrine, Rachid Harba, Raphael Canals, Roger Ledee, and Meryem Jabloun. A comparison of active contour prior shape segmentation methods: application to diabetic plantar foot thermal images. *8th International Conference on Digital Image Processing and Vision*, 2019.

[10] Fareed Ahmed, Huu Dien Khue Le, Julien Olivier, and Romuald Boné. An active contour model with improved shape priors using fourier descriptors. *VISAPP (1)*, pages 472–476, 2013.

[11] Yunmei Chen, Sheshadri Thiruvenkadam, Hemant D Tagare, Feng Huang, David Wilson, and Edward A Geiser. On the incorporation of shape priors into geometric active contours. *Variational and Level Set Methods in Computer Vision, 2001. Proceedings. IEEE Workshop on*, pages 145–152, 2001.

[12] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[15] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, pages 1097–1105, 2012.

[16] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. pages 3431–3440, 2015.

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241, 2015.

[18] Vijay Badrinarayanan, Ankur Handa, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. *arXiv preprint arXiv:1505.07293*, 2015.

[19] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.