

Deep neural networks based wrist print region segmentation and classification

Hasan Erdiñç Koçer¹, Kerim Kürşat Çevik^{2,*}

¹Department of Electrical and Electronics Engineering, Faculty of Technology, Selcuk University, Konya, Turkey, ekocer@selcuk.edu.tr, ORCID: 0000-0002-0799-2140

²Department of Management Information Systems, Faculty of Applied Sciences, Akdeniz University, Antalya, Turkey, kcevik@akdeniz.edu.tr, ORCID: 0000-0002-2921-506X

ABSTRACT

In recent years, biometric recognition based systems have become widespread. One of these is wrist-based recognition systems. In this study, wrist print based recognition system was developed by using near infrared (NIR) camera. Totally 220 NIR camera images taken from 10 for each both hands of 11 people. The obtained data set is allocated 70% (154 images) for training and 30% (66 images) for testing. The wrist regions are labeled on the training set images. Data sets were created with two different labeling methods. In the first data set, only the wrist regions were labeled and it was aimed to segment the wrist region from the image. In the second data set, the wrist images were labeled according to 22 classes and these classes were tried to be predicted. The labeled data was trained with YOLOV2 architecture supported by ResNet50 one of the deep neural network models. The trained model was tested with the remaining 30% of the data set. In the test process, the wrist region was determined in the NIR images with the trained model. As a results of the study, it was seen that the wrist regions were correctly detected in all first data set test images and the mean value of obtained similarity rates was 95.26%. In the test results of the second dataset, 92.43% classification success was obtained. Therefore, it can be said that the deep learning architectures ResNet and YOLO are effective in the segmentation of the wrist region.

ARTICLE INFO

Research article

Received: 5.01.2021

Accepted: 30.04.2021

Keywords:

wrist print recognition,
deep neural networks,
near-infrared camera,
YOLO

*Corresponding author

1. Introduction

Biometric recognition based systems are known as most secure systems. What makes biometric systems important in this way is that the security parameter they use is personal and there is no risk of theft and cannot be copied. Although there are many biometric recognition methods, fingerprint recognition, iris recognition, vein recognition and face recognition are mostly used today.

Vein recognition is a high-security biometric recognition approach based on human vascular structure. In this biometric identification approach, finger veins, hand veins, palm veins and wrist veins are used as biometric parameters. The identification systems which use wrist vein structure as biometric input are called wrist print recognition. In wrist print recognition system, the human wrist image is captured by using near-infrared (NIR) camera and illumination and then the wrist region is segmented for determining the wrist vein information. Then, the wrist print is used for identification. The wrist image acquisition unit used in this study is given in Figure 1.

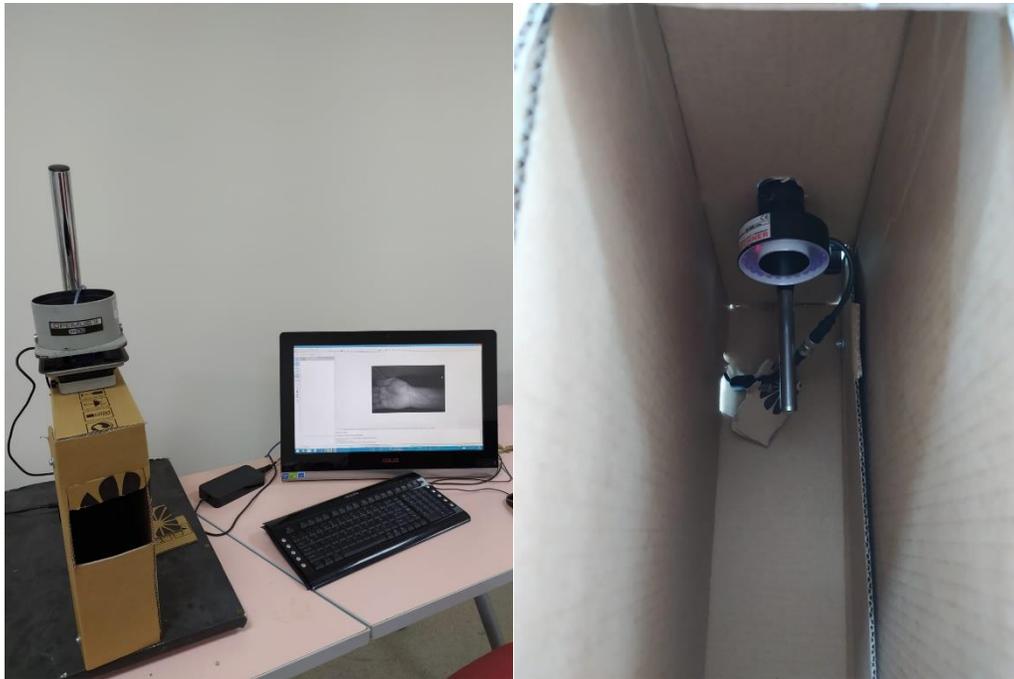


Figure 1. A wrist image acquisition unit

Although many studies have not been done on wrist vein recognition, we come across studies using different approaches in the literature. We can evaluate previous studies on the subject in 2 groups. The first one is hand, palm and finger vein recognition studies (will not be mentioned in this article), and the other is the wrist vein recognition works. Hartung et al. (2011) performed spectral minutiae extraction in the identification of wrist vessels [1]. In preprocessing stage, binarization and Skeletonization were implemented after noise reduction and edge enhancing. Convolution approach and Location-Based Spectral Minutiae Representation were used for extracting the feature. In this study, Hausdorff distance, Modified Hausdorff, Similarity-based Mix-matching, SML correlation and SML fast rotate methods were used for feature matching [2]. Das et al. (2014), performed the feature matching process with Support Vector Machines. Dense Local Binary Pattern approach was applied for extracting the feature [3]. In 2018, Nikisins et al. studied wrist vein recognition with cross-correlation-based comparison [4]. Adaptive histogram equalization [5] and Discrete Meyer Wavelet techniques [1] were used in preprocessing phase. Then Hessian matrix was implemented and the wrist vein features were extracted in this study. In 2020, Garcia-Martin et al. were developed portable contactless system for wrist vascular biometric identification. a new contactless database has been collected with the software algorithm TGS-CVBR®. In this work, Preprocessing and Identification Software for Contactless Vascular Biometric Recognition (PIS-CVBR®), is used for recognition the wrist images [6]. In this study three different wrist image datasets (UC3M-CV1, UC3M and PUT databases) were used and three different feature extraction methods (Scale-Invariant Feature Transform – SIFT, Speeded Up Robust Features – SURF and Oriented FAST and Rotated BRIEF (ORB) were applied to the wrist images. These wrist recognition researches are given in

Table 1 including authors, year, methods, database information and best error rates in percentages.

Table 1. Literature overview for wrist print recognition

Researcher(s)	Year	Methods used in classification	Dataset (NIR Imaging)		Best Error Rate(s)
			Subjects	Images	
Hartung et al. [1]	2011	SML Correlation	29	348	6.13 % 5.90 %
Das et al. [3]	2014	SML fast rotate Support Vector Machine	50	1200	0.79 %
Nikisins et al. [4]	2018	Cross Correlation with FFT	29	348	3.58 %
Garcia-Martin et al. [7]	2020	Brute Force Matcher	50	1200	8.59%
		Fast Library for Approximate Nearest Neighbors	50	1200	0.08%
			121	605	15.91%

As it is seen from this table, various methods were used in classification process. We have implemented Deep Neural Networks for recognition the wrist image, so a new approach is applied for wrist print recognition. In this study, the wrist vein region was detected and marked from the hand and wrist image taken using a near infrared camera. In this context, a software based on deep learning has been developed including ResNet50 and YOLO algorithms. The performance of the system was evaluated comparatively by testing the segmentation and classification process on 220 wrist images taken from 22 human wrists.

2. Deep neural networks (DNN)

Deep neural networks (DNN) can recognize objects without being affected by different properties such as different positions,

directions and camera angles and environmental factors such as lighting. A deep learning algorithms are trained on tagged images. An architecture of DNN is shown in Figure 2.

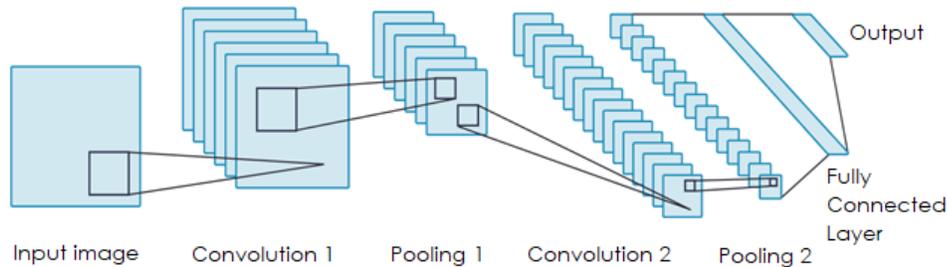


Figure 2. A Deep neural networks architecture.

The input usually consists of images or signals. In the convolution layers, the filtering process is applied to the 3-dimensional matrices in the previous layer. The number of filters used constitutes the depth (the size of the 3rd dimension) of the convolution layer. In the pooling layer, size reduction is applied. In the section called fully connected layer, classical artificial neural network operations are performed. The output can be defined as a vector with the length of the defined class. The number of layers in the architecture and the filter size and number can be changed by the user to suit the application. In addition, the performance may vary depending on the number of images and iterations to be used for training. If the number of training iterations is too high, both the training time increases and the model moves towards memorization. As the rate of misclassification in data that the model has not seen before will increase as a result of memorization, accuracy decreases. In the study, it was tried to determine the wrist region from the images taken by using the ResNet50 and YOLO architectures.

2.1. ResNet50

The Residual Network (also known as ResNet) uses redundant blocks with multiple layers to reduce training error. Typical ResNet models are implemented with double- or triple- layer skips that contain nonlinearities (ReLU) and batch normalization in between. During training, the weights adapt to mute the upstream layer, and amplify the previously-skipped layer. In the context of residual neural networks, a non-residual network may be described as a plain network [8, 9].

The standard ResNet architecture consists of 152 layers and has a higher classification success than other DNN architectures. The ResNet50 architecture used in this work, on the other hand, consists of 34 layers. Both layer blocks in the network are replaced by a three-layered residual block. The residual block characterization can be given as:[10, 11]

$$a^{[l+2]} = g(a^{[l]} + z^{[l+2]}) \quad (1)$$

Here, a is the activations (outputs) of neurons in layer l , g is the activation function for layer l . In this equation, l is used for clarity.

2.2. You Only Look Once (YOLO)

YOLO is an algorithm for object detection using convolutional neural networks (CNN). It stands for "You Only Look Once", which means "Just Look Once". The reason for choosing this name is that the algorithm is fast enough to detect objects in one go. The most important feature that distinguishes YOLO from other algorithms is its ability to detect objects in real time. YOLO uses Darknet library and has a very fast running architecture. The reason it is so fast is that it can guess objects and coordinates by passing any picture through the neural network in one go. In other words, YOLO takes an approach that solves a single regression problem while performing the estimation process [12, 13].

When we look at YOLO's algorithm, we first see that it divides the input picture into grids of certain sizes (3x3, 5x5, 17x17 etc.). It is then checked whether the searched object is within the grid. The object detected in the similarity check is scored and marked. Then non maximum suppression is applied, which evaluates the score of each marked object, and those below a certain confidence score are separated [14, 15].

3. The proposed method

There are 3 main elements in the proposed approach. These can be listed as collecting image data set, labeling the wrist area and performing deep learning based training / testing processes. While performing the deep neural network for detection and classification of wrist images, Yolo and ResNet50 architectures were implemented. The flow chart of the system designed within the scope of this study is shown in Figure 3.

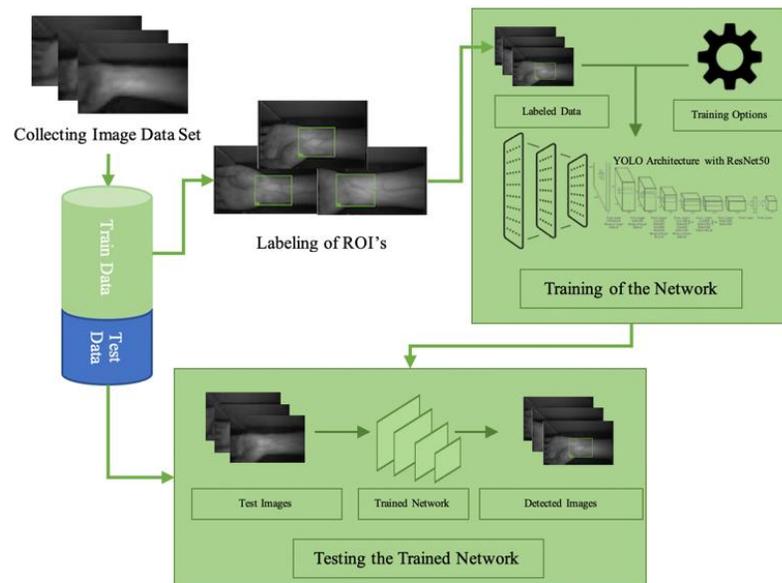


Figure 3. Flow chart of the designed system.

The study is primarily provided within the scope of the data collection. Two wrist images of 11 people were taken with a near infrared (NIR) camera. In the study, a NIR camera with a resolution of 752x480 pixels, a monochrome (gray level) and an infrared light source operating at 880nm wavelength was used to capture the images. By taking 10 images for both wrists of each person, a total of 220 images were provided in the data set. The images obtained are gray level with a resolution of 752x480 pixels in RGB format. Images consist of 8 bits as color depth. The sample images used in the designed system are shown in Figure 4. In the second stage, the data set is divided into training and test data. As used frequently in the literature, the obtained images were determined as homogeneous 70% (154 images) for training and the remaining 30% for testing.

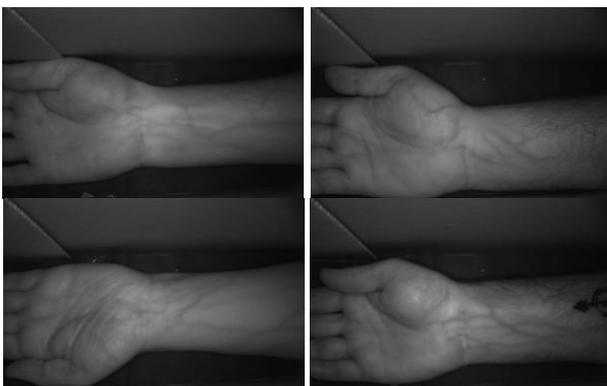


Figure 4. Image set examples.

In the next stage, Region of Interest (ROI) areas are labeled on the images (70%) allocated for training. In order to define the

wrist print region, ROI is labeled as 224x224 dimensional boxes where the veins on the wrist are clearly visible. Image Labeler plugin of Matlab software was used for this labeling process. An interface of Image Labeler is shown in Figure 5. The remaining test data were transferred to the test stage without labeling.

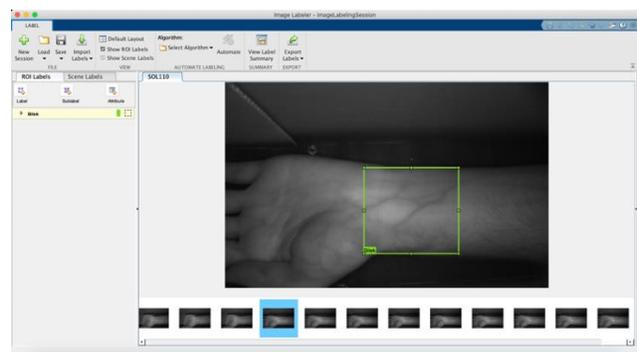


Figure 5. Image labeling process.

Data sets were created with two different labeling methods. In the first data set, only the wrist regions were labeled and it was aimed to segment the wrist region from the image (One class='Wrist'). In the second data set, the wrist images were labeled according to 22 classes and these classes were tried to be predicted (Twenty two classes='Wrist 1', 'Wrist 2'... 'Wrist 22'). In the study, both segmentation and classification performance in images were tried to be measured with these two data sets. After the data labeled and prepared, the labeled data was trained with Deep Neural Network - (DNN) with the software realized within the scope of the study. In the training phase of the network, the images labelled were given for 1 class (wrist) of

output as an introduction to the neural network built from YOLO architecture powered by ResNet50 and training has been done. Layer structure of the designed network is given in Table 2. As seen in Table 2, the input images are 224x224 in size and have 3 layers. After the convolution process, the sub-image with 112x112x64 activation was obtained. After the activation

process, the activation number of 55x55x64 was obtained by max pooling process. Convolution was made using Yolo2 in the output layer. For the other data set used in the study (Twenty two classes dataset), the 150th output layer (YOLO v2 Output) has 22 neurons.

Table 2. Layer structure of the designed network

	Name	Type	Activations	Learnables	Total Learnables
1	Input_1 224x224x3 images	Image Input	224x224x3	-	0
2	Conv1 64 7x7x3	Convolution	112x112x64	Weights 7x7x3x64 Bias 1x1x64	9472
3	bn_conv1 Batch Normalization	Batch Normalization	112x112x64	Offset 1x1x64 Scale 1x1x64	128
4	activation_1_relu ReLU	ReLU	112x112x64	-	0
5	max_pooling2d_1 3x3 max pooling	Max Pooling	55x55x64	-	0
6	res2a_branch1 256 1x1x64 conv	Convolution	55x55x256	Weights 1x1x64x256 Bias 1x1x256	16640
7	bn2a_branch1 Batch Normalization	Batch Normalization	55x55x256	Offset 1x1x256 Scale 1x1x256	512
8	res2a_branch2a 64 1x1x64 conv	Convolution	55x55x64	Weights 1x1x64x64 Bias 1x1x64	4160
9	bn2a_branch2a Batch Normalization	Batch Normalization	55x55x64	Offset 1x1x64 Scale 1x1x64	128
10	activation_2_relu ReLU	ReLU	55x55x64	-	0
			...		
141	activation_40_relu ReLU	ReLU	14x14x1024	-	0
142	yolo2Conv1 1024 3x3x1024 conv	Convolution	14x14x1024	Weights 3x3x1024x1024 Bias 1x1x1024	9438208
143	yolo2Batch1 Batch Normalization	Batch Normalization	14x14x1024	Offset 1x1x1024 Scale 1x1x1024	2048
144	yoloRelu1 ReLU	ReLU	14x14x1024	-	0
145	yolo2Conv2 1024 3x3x1024 conv	Convolution	14x14x1024	Weights 3x3x1024x1024 Bias 1x1x1024	9438208
146	yolo2Batch2 Batch Normalization	Batch Normalization	14x14x1024	Offset 1x1x1024 Scale 1x1x1024	2048
147	yoloRelu2 ReLU	ReLU	14x14x1024	-	0
148	yoloClassConv 4 1x1x1024 conv	Convolution	14x14x42	Weights 42 Bias 1x1x42	43050
149	yolo2Transform	Yolo V2 Transform Layer	14x14x42	-	0
150	yolo2OutputLayer	Yolo V2 Output Layer	-	-	0

ResNet50 was created in a 150-layered structure because of integration with YOLO. Since the input data of the ResNet network is 224x224x3, the images in the data set have been adapted to this resolution. The training parameters of the designed deep neural network were given in Table 3. The output size is 1 for segmentation of the wrist region and the output size is 22 for classification of the wrist data. The optimizer used in this study is Stochastic gradient descent with momentum. It is an iterative method for optimizing an objective function with suitable smoothness properties. The iteration number for segmentation and classification process is limited to 1300 and 2600, respectively.

Table 3. Training parameters for the proposed method

Option	Segmentation Data Set	Classification Data Set
Epoch (Iteration)	100 (1300)	200 (2600)
Batch Size	16	16
Learning Rate	0,001	0,001
Optimizer	SGDM	SGDM
Input Size	224x224x3	224x224x3
Output Size	1 (Wrist)	22 (Wrist)
Verbose	True	True
Verbose Frequency	1	1
Shuffle	Never	Never

The computer on which the designed model is trained has the following hardware configuration: Intel i7-2600 3.40 GHz processor running under 64 Bit Windows 10 operating system, 22 GB RAM, GTS450 GDDR5 1GB 128Bit Nvidia GeForce DX11 Graphics Card hardware. In addition, Matlab software was used in coding, training and testing of the model.

The change of Root Mean Square Error (RMSE the square root of the difference between the value obtained and the value obtained during the training of the network) and LOSS (expressing how different the estimate made by the model from the actual value) values are given in Figure 6. In addition, the training time of the segmentation network is 03:46:10, the resulting minimum RMSE value of the network is 0.07 and LOSS value is 5.2e-03. The training time of the classification network is 05:27:30, the resulting minimum RMSE value of the network is 0.077 and LOSS value is 0.088.

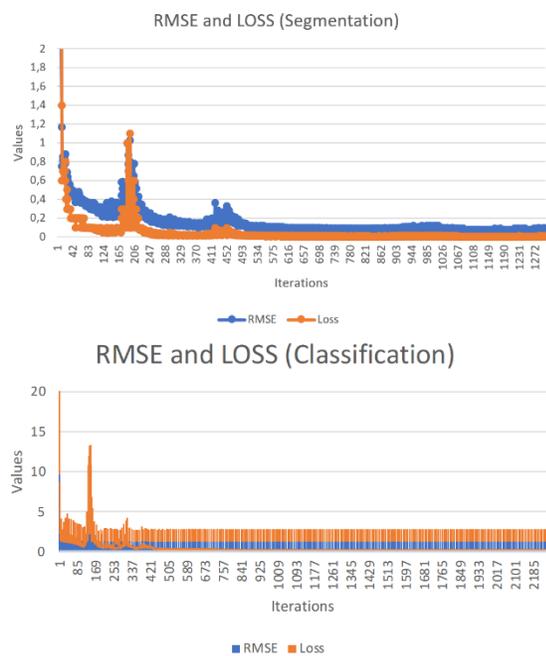


Figure 6. Change of RMSE and LOSS values during training (Segmentation and Classification).

4. Results and discussion

For the evaluation of the study, the test process was carried out by using images (30%) that were not shown to the network model whose training was completed. Test images were given as an input to the network and the wrist areas were determined. In determining the region, the minimum value of similarity ratio was taken as 0.70 and the highest similarity value was determined as the result. Sample test images including the detected and marked boxes are given in Figure 7.

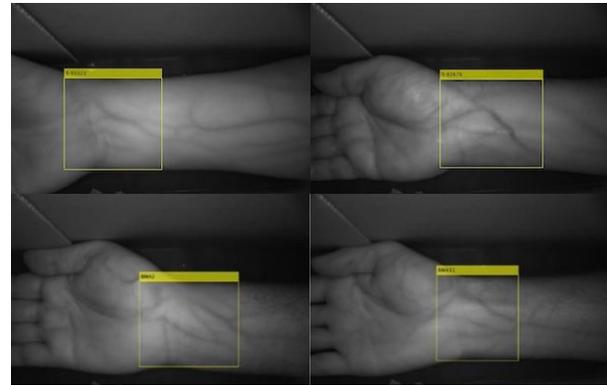


Figure 7. Detection results

The percentage values for segmentation of the wrist area, which are formed by input (total of 66 images in the test process), are shown in Figure 8(a). The process of determining the wrist area on an image takes approximately 0.42 seconds. It took approximately 30 seconds to determine the ROI areas of all test images and the ROI's were saved as a file.

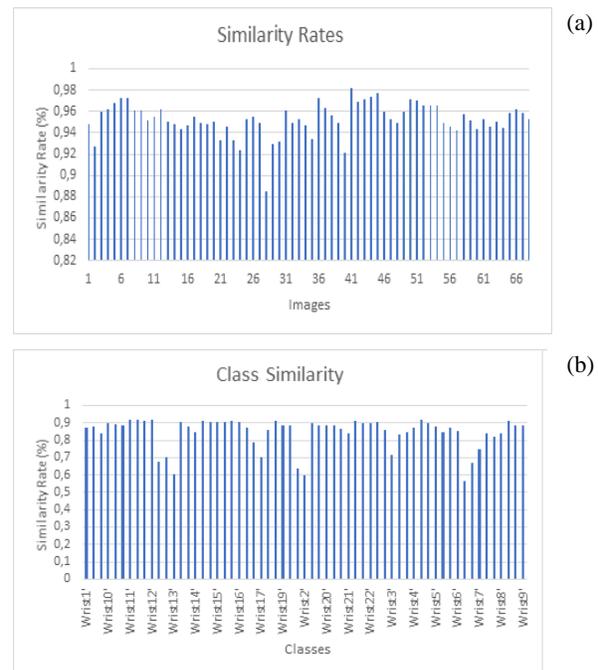


Figure 8. Similarity rates for segmentation (a) and classification (b)

The percentage values for classification, which are formed by input (total of 66 images in the test process), are shown in Figure 8(b). The process of classification on an image takes approximately 0.37 seconds. It took approximately 30 seconds to determine the ROI areas of all test images and the ROI's were saved as a file.

5. Conclusion

In this study, a deep learning based classification approach is implemented for wrist print recognition. YOLO architecture with ResNet50 is used for training process in deep learning. As a result of the study, it was seen that the wrist regions were segmented correctly in all test images and the average value of the obtained similarity rates was 95.26%. As a result of the classification process, 61 out of 66 images were classified correctly. Therefore, 92.43% classification successes were obtained. Therefore, it can be said that the deep learning architectures ResNet50 and YOLO are effective in the segmentation of the wrist region.

The obtained error rate for segmentation and classification is 7.57% in our approach. When we compare the values we obtained as a result of this study with previous studies on wrist recognition, it is seen that an average performance is achieved. The main reason for this is the small number of images (7 images per wrist) used for training in the study. It is known that the higher the number of entrance images in deep learning approaches gives the higher the classification performance. Therefore, it is thought that the success will increase by increasing the number of samples.

Acknowledgements

This study was presented as an oral presentation at the 4th International Conference on Engineering Technologies (ICENTE) held in Konya on 19-21 November 2020.

References

- [1] Hartung D., Olsen M.A., Xu H., Busch C., "Spectral minutiae for vein pattern recognition," in 2011 International Joint Conference on Biometrics (IJCB), IEEE, (2011), 1-7.
- [2] Wang L., Leedham G., Cho S.-Y., "Infrared imaging of hand vein patterns for biometric purposes," IET computer vision, 1, 3, (2007), 113-122.
- [3] Das A., Pal U., Ballester M.A.F., Blumenstein M., "A new wrist vein biometric system," in 2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM), IEEE, (2014), 68-75.
- [4] Nikisins O., Eglitis T., Anjos A., Marcel S., "Fast cross-correlation based wrist vein recognition algorithm with rotation and translation compensation," in 2018 International Workshop on Biometrics and Forensics (IWBF), IEEE, (2018), 1-7.
- [5] Uriarte-Antonio J., Hartung D., Pascual J.E.S, Sanchez-Reillo R., "Vascular biometrics based on a minutiae extraction approach," in 2011 Carnahan Conference on Security Technology, IEEE, (2011), 1-7.
- [6] Chen H., Lu G., Wang R., "A new palm vein matching method based on ICP algorithm," in Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, (2009), 1207-1211.
- [7] Garcia-Martin R., Sanchez-Reillo R., "Wrist Vascular Biometric Recognition Using a Portable Contactless System," Sensors, 20, 5, (2020), 1469.
- [8] Wikipedia. "Residual neural network." Wikipedia. https://en.wikipedia.org/wiki/Residual_neural_network (accessed 15.11.2020, 2020).
- [9] He K., Zhang X., Ren S., Sun J., "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, (2016), 770-778.
- [10] Amidi A., Amidi S., Vlachakis D., Megalooikonomou V., Paragios N., Zacharaki E.I., "EnzyNet: enzyme classification using 3D convolutional neural networks on spatial representation," PeerJ, 6, (2018), e4750.
- [11] Amidi S., Amidi A., Vlachakis D., Paragios N., Zacharaki E.I., "Automatic single-and multi-label enzymatic function prediction by machine learning," PeerJ, 5, (2017), e3095.
- [12] Kuyumcu B., Buluz B., Kömeçoğlu Y., "Author Identification in Turkish Documents with Ridge Regression Analysis," in 2019 27th Signal Processing and Communications Applications Conference (SIU), 2019: IEEE, pp. 1-4.
- [13] Redmon J., Farhadi A., "YOLO9000: better, faster, stronger," in Proceedings of the IEEE conference on computer vision and pattern recognition, (2017), 7263-7271.
- [14] Nguyen C.C.et al., "Towards real-time smile detection based on faster region convolutional neural network," in 2018 1st International Conference on Multimedia Analysis and Pattern Recognition (MAPR), IEEE, (2018), 1-6.
- [15] Redmon J., Divvala S., Girshick R., Farhadi A., "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, (2016), 779-788.