Visual Communications and Image Processing (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/vcip.2023.020101
ISSN 2522-6770 Vol. 2 Num. 1

Zero-RADCE: Zero-Reference Residual Attention Deep Curve Estimation for Low-Light Historical Tibetan Document Image Enhancement

Qinghua Zhao, Weilan Wang*

Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou, China
**Corresponding author*

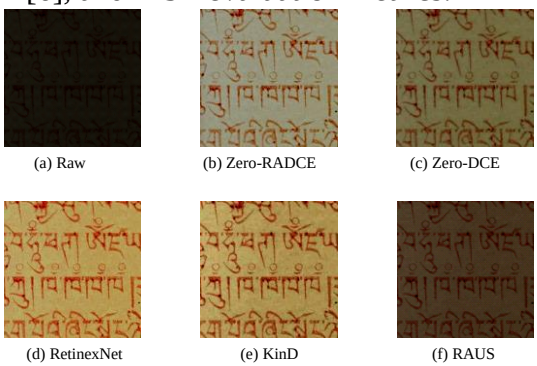
Keywords: Low-light enhancement, historical tibetan document image, zero-DCE

Abstract: To improve the reading experience and the performance of subsequent document analysis and recognition algorithms, the background of low-light document image is expected to be smoothed and its foreground text information needs to be highlighted. The enhanced document images can improve the reading experience and the subsequent document analysis and recognition algorithms. In this paper, first construct a low-light historical Tibetan document image dataset. then improve the deep curve estimation network of Zero-DCE by using encoder-decoder architecture, residual network, and spatial attention mechanism; finally extract the text information features in low-light historical Tibetan document images by specially designed Gaussian and Laplace filters for improving the spatial consistency loss, which not only achieves low-light image enhancement but also improves the spatial consistency loss. Experiments show that the proposed method achieves better results in both quality and quantitative evaluations for low-light historical Tibetan document image enhancement. Meanwhile, the training parameters are reduced by 39.52% and the Flops are reduced by 70.63% compared to the original network.

1. Introduction

The number of historical Tibetan documents is huge and covers many fields such as history, politics, and economy [1]. In the context of current information technology, digital archiving not only enables readers to access efficiently, but also helps researchers better organize and mine information. Since most of the historical Tibetan document images are large in size, the document images are mainly acquired by capturing photographs, and when the historical Tibetan document images are photographed in a low-light environment or under technical limitations, the captured images are darker and the textual details are hidden in the dark. Low-light historical Tibetan document image enhancement is very necessary, which not only can improve readers' reading experience, but also lay a good foundation for subsequent research work such as layout analysis and recognition. Figure 1 shows the enhancement effects of a set of different networks under very low-light conditions and blurred images of historical Tibetan documents. Compared with the state-of-

the-art networks Zero-DCE [2], RetinexNet [3], KinD [4], and RAUS [5], our method Zero-RADCE improves the brightness of images with smoother backgrounds, and its results achieve the highest scores in PSNR, SSIM [6], and MSE evaluation metrics.



(a) Raw (b) Zero-RADCE (c) Zero-DCE

(d) RetinexNet (e) KinD (f) RAUS

Figure 1: Image enhancement effect of different networks on low-light historical Tibetan document images.

In this study present a Zero-Reference Residual Attention Deep Curve Estimation (Zero-RADCE) for low-light historical Tibetan document image enhancement. Zero-DCE is used for low-light enhancement and formulates the enhancement task as an image-specific curve estimation problem. The network model takes a shimmering image as input, generates a high-order curve as output, and uses curve to adjust the dynamic range of the input to obtain an enhanced image. At the same time, Zero-DCE designs a set of microscopic non-reference loss functions spatial consistency loss, exposure control loss, color constancy loss, and light smoothness loss to achieve unsupervised training. In order to enhance historical Tibetan document images with Low-light and highlight the textual effects at meantime, the main contributions of Zero-RADCE proposed in this paper are as follows.

(1) The use of encoder-decoder architecture and residual network in the curve estima-tion network. This operation brings adaptive adjustment capability and maintains the consistency of global light to the network, and reduces the amount of training parameters by 39.52%, which greatly improves the training speed and also enhances the enhancement effect.

(2) Adding spatial attention mechanism. The spatial attention mechanism can add a certain weight to the text information, which can enhance the text information in the historical Tibetan document images while weakening the irrelevant back-ground information.

(3) Improving spatial consistency loss. After the input image is passed through Gaussian operator, Laplace operator, and max pooling operations, the background noise of the input image is eliminated and the text information is highlighted, and then the processed image is subjected to spatial consistency loss operation with the enhanced image.

The improved Zero-RADCE achieves advanced results in both qualitative and quantitative metrics for low-light historical Tibetan document image enhancement, and more importantly, good results can be obtained in subsequent binarization work using only simple traditional algorithms. In addition, RADCE is a lightweight network with few computational resources, fast speed and high performance.

2. Related Work

Low-light image enhancement algorithm can be divided into traditional method and deep learning method. Traditional low-light image enhancement methods usually adopt histogram equalization [7], Gamma correction [8], Retinex [9], and other algorithms, but images enhanced by these methods may have a lot of noise. In recent years, there have been a lot of achievements in

low-light image enhancement based on neural network algorithms. Most algorithms based on CNN require training of paired data sets such as RetinexNet and KinD. In addition, there are some networks that do not rely on training pairs of data sets, such as Zero-DCE and RUAS. However, these methods are mainly oriented to scene images, and it is difficult to achieve satisfactory visual effects due to problems such as noise, blurred text, and uneven background in document images.

Most CNN-based solutions rely on supervised training of paired data. Typically, paired data is collected through automatic light attenuation, changes to camera settings during data capture, or synthetic data through image modification. Wei et al. used deep neural network simulation of Retinex's model (RetinexNet) to make low-light images and normal images share network parameters and decompose to get the same reflection image to ensure the difference between the two light images. Then, based on the decomposition results, the network is trained to enhance the illuminated image, and the brightness of the enhanced results is significantly improved. Yonghua Zhang and their team created a network called Kindling the Darkness (KinD), which is a simple and effective method inspired by the Retinex theory. KinD decomposes the image into two parts to enhance its quality. The illumination component is responsible for conditioning the light, and the other reflectivity component is responsible for removing degradation. However, when the above-mentioned methods deal with real low-light historical Tibetan document images, especially when the image contains a lot of noise, the enhancement result will have serious color distortion, and the noise will be obviously amplified.

Unsupervised network model also plays a significant role in the application of image enhancement. Chunle Guo and their team introduced the Zero-Reference Deep Curve Estimation method (Zero-DCE) for enhancing low-light images. Zero-DCE uses deep networks to estimate image-specific curves for light enhancement. In contrast, the DCE-net adjusts an image to estimate the dynamic range of pixels and higher-order curves.Curve estimation is specifically designed to take into account the range of pixel values while maintaining monotonicity and differentiability. Zero-DCE is a technique for image enhancement that doesn't need paired or unpaired data during training. The method uses non-reference loss functions to measure enhancement quality and drive learning. It also uses simple nonlinear curve mapping for intuitive image enhancement. This was developed by Risheng Liu and their team.RUAS builds upon the Retinex rule, which is a principle for color image enhancement that seeks to separate the illumination and reflectance components of an image. RUAS establishes models to characterize the intrinsic underexposed structure of low-light images and then unrolls their optimization processes to construct a holistic propagation structure. This allows for a more efficient and effective enhancement of low-light images in real-world scenarios.

3. Methodolgy

Figure 2 shows the framework of Zero-RADCE. By incorporating encoder-decoder [10], residual network [11], and spatial attention mechanism in Cbam [12], deep curve estimation network is used to estimate a set of optimal fitting light enhancement curve (LE curve) of an input image. LE curve is shown in Equation (1):

$$LE_n(\mathrm{X}) = LE_{n-1}(\mathrm{X}) + \mathcal{A}(X)LE_{n-1}(\mathrm{X})\big(1 - LE_{n-1}(\mathrm{X})\big) \tag{1}$$

The $\mathcal{A}$ parameter map in the Zero-DCE method is pixel-wise and has the same size as the input image. Each pixel in the input image corresponds to a curve in the A map with the best-fitting alpha to adjust its dynamic range. These curves are then used for pixel-wise adjustment of the input's dynamic range to obtain an enhanced image.

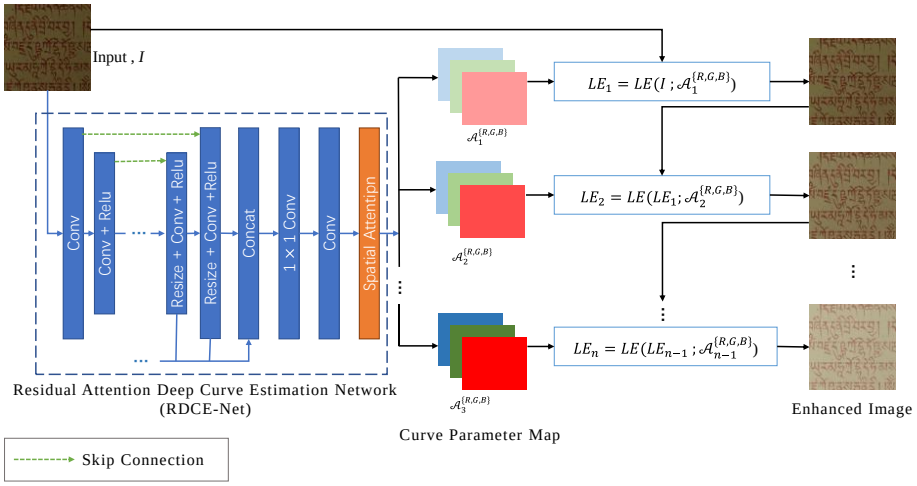


Figure 1: The framework of Zero-RADCE.

To enable zero-reference learning, Zero-DCE uses a set of differentiable non-reference loss functions: The spatial consistency loss in image enhancement aims to maintain spatial coherence in the enhanced image. It achieves this by preserving the difference of neighboring regions between the input image and its enhanced version; The Exposure control loss is used to control exposure level; The loss of color constancy is used to correct the possible color deviation in the enhanced image. The illumination smoothing loss is used to preserve the monotonic relationship between neighboring pixels.

3.1. RADCE-Net

The EnhanceNet-Like structure in RetinexNet and spatial attention mechanism are adopted in RADCE-Net to obtain the mapping between the training input image and its best-fitting curve parameters. The input is a low-light image and the output is a set of pixel level curve parameter mappings for the corresponding higher order curve. EnhanceNet-Like structure includes encoder-decoder and residual network. Encoder-decoder can obtain context information in a large area, which brings adaptive adjustment ability to the network. Meanwhile, it can further reduce the number of parameters and achieve faster operation efficiency. Residual network can reduce information loss in neural network training while maintain the consistency of global illumination. The spatial attention mechanism can pay attention to the Tibetan information in the image, add higher weight to the Tibetan part, and better highlight the text information in the enhanced image.

After the low-light historical Tibetan document images enters Zero-RADCE, it first goes through a convolution kernel of 3×3, padding 1 and stride 1. Then, after three layers of size 3×3, padding 1, and stride 2 followed by ReLU activation function, it performs jump connection and element-by-element summing for three layers of up-sampling block, down-sampling block and up-sampling block of corresponding size, make network learning residual. Then, multi-scale splicing is introduced to adjust the up-sampling blocks of different sizes to the same size through nearest neighbor interpolation, and the feature graphs are spliced to simplify the cascading features into 32 channels through a 1×1 convolution layer. Finally, the 24 pixel-level parameter maps are obtained through 3×3 convolution layer.

The feature of spatial attention to promote the key expression, is essentially the spatial information of original image through space conversion module, transform to another space and keep the key information, for each location weighting mask and weighted output, thereby enhancing the text information in the historical Tibetan document images document image and weaken the

background region is not relevant information. Figure 3 shows the spatial attention mechanism model. The 24 pixel-level curve parameter maps are generated into 2 channel feature maps through max-pooling and average pooling between channels. Then, the feature maps generated by sigmoid activation function are multiplied element by element by convolution kernel with a size of 7×7, padding 3, and stride 1. Finally, 24 weighted pixel-level curve parameter maps are obtained and used for 8 iterations to generate enhanced images.

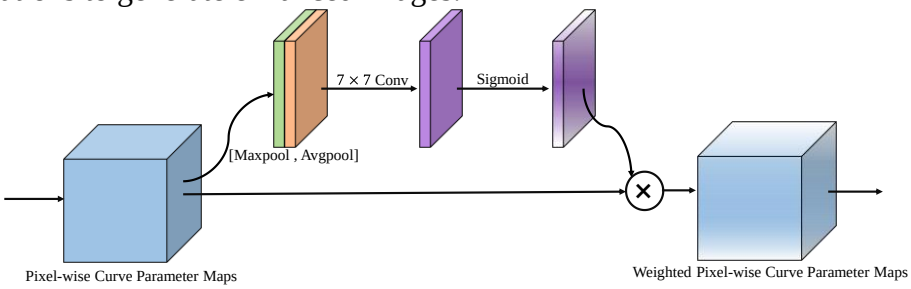


Figure 3: Spatial attention.

3.2. Loss of Spatial Consistency

L_{spa} promotes the spatial consistency of the enhanced image by calculating the difference between the neighboring regions of the input image and its enhanced image:

$$L_{spa} = \frac{1}{K}\sum_{j\in\Omega(i)}^{K}\left(\left|\left(Y_i - Y_j\right)\right| - \left|\left(I_i - I_j\right)\right|\right)^2 \tag{2}$$

$$L'_{spa} = \frac{1}{K}\sum_{j\in\Omega(i)}^{K}\left(\left|\left(Y_i - Y_j\right)\right| - \left|\left(I'_i - I'_j\right)\right|\right)^2 \tag{3}$$

Equation (2) and Equation (3) are the original and improved spatial consistency loss, respectively. Where K is the number of local regions, the size of local regions is set to 4×4, and Ω (i) is the four neighboring regions (up, down, left and right) centered at the region i. Y, I and I' are denoted as the local average intensity values of the enhanced image, the input image and the input image after processing (Gaussian operator, Laplacian operator, max-pooling, and up-sampling).

The Laplacian operator is an image enhancement algorithm based on second-order differential, which emphasizes the sudden change of gray level in the image, and does not emphasize the area where the gray level changes slowly. Therefore, it can effectively enhance Tibetan text edge and give Tibetan text information higher weight.

Sharpening often enhances the noise while enhancing the edge. In order to achieve a better Tibetan extraction effect and minimize the noise, this paper first smoothed the input image and then sharped to enhance the Tibetan edges and details. Gaussian smoothing can reduce and suppress image noise. The weight of the center point of the Gaussian template is the largest. With the increase of the distance from the center point, the weight decreases rapidly, so as to ensure that the center point looks close to the point closer to it and achieve a more natural smoothing effect. Laplace's sharpening effect will also cause double edges and hollow characters in Tibetan, so 2×2 max-pooling is used to solve the problem of hollow characters. Equation (4) and Equation (5) are a group of Gauss and Laplacian operators. After many experiments, they are suitable for smoothing and sharpening historical Tibetan documents images in low-light conditions.

$$W_1 = \frac{1}{273}\begin{bmatrix} 1 & 4 & 7 & 4 & 1 \\ 4 & 16 & 26 & 16 & 4 \\ 7 & 26 & 41 & 26 & 7 \\ 4 & 16 & 26 & 16 & 4 \\ 1 & 4 & 7 & 4 & 1 \end{bmatrix} \tag{4}$$

$$W_2 = \begin{bmatrix} 4 & 16 & 4 \\ 16 & -80 & 16 \\ 4 & 16 & 4 \end{bmatrix} \tag{5}$$

Figure 4 is a diagram of L_{spa} and $L_{spa}^{'}$. Due to the low-light input image, there are problems such as low contrast between text and background and background noise, which are not conducive to text highlighting and background noise reduction in enhanced image. Therefore, Equation (2) can better promote the spatial consistency of enhanced image and provide motivation for spatial attention mechanism.

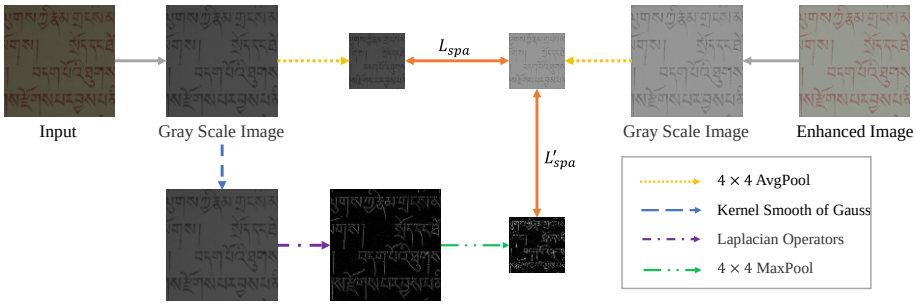


Figure 4: L_{spa} and L'_{spa} operation diagram.

4. Experiments

Zero-RADCE is an unsupervised network trained with multiple exposure images to maximize the wide dynamic range adjustment capability. To this end, an image of historical Tibetan document, usually 5500×1500 in size, is cut into multiple 512×512 subgraphs in this paper, and low-light and overexposed subgraphs are included in the data set. The dataset consisted of 3000 images with different exposures, which were randomly divided into 2320 training sets and 680 validation sets. Zero-RADCE was trained using an NVIDIA GeForce RTX 2080 Ti GPU, using Python 3.7, CUDA 10.2, Pytorch 1.6.0, and other related function libraries.In the training process of the model, the filter weights of each layer were initialized with the standard zero mean and 0.02 standard deviation Gaussian function. A batch size of 8 was used, and the entire training process took approximately 1 hour, with a total of 200 rounds of training.

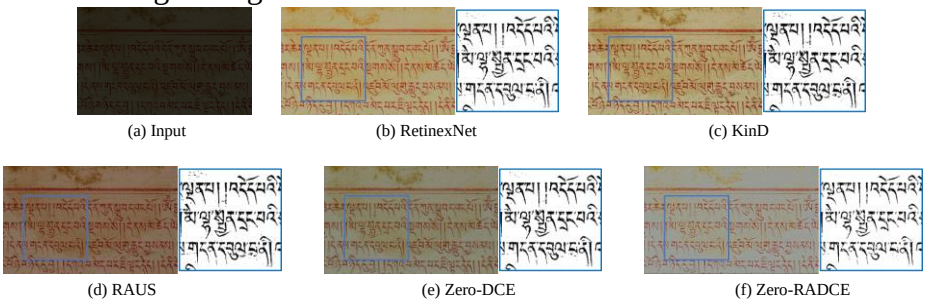
4.1. Benchmark Evaluations

This paper compare Zero-RADCE with several state-of-the-art methods: two CNN-based supervised approaches RetinexNet and KinD, and two CNN-based unsupervised approaches RAUS and Zero-DCE, both subjective and objective experimental studies using publicly available source codes and standard image datasets.

4.1.1. Subjective Evaluation

Figure 5 shows the enhanced images of low-light historical Tibetan document images by different methods. The blue box area is the binarization effect of Otsu [13]. Zero-RADCE minimizes the number of smudges in the document image background while RetinexNet, KinD,

RAUS, and Zero-DCE do not attenuate the smudges. This problem not only affects the visual effects but can be seen in binary images. Zero-RADCE enhanced image binarization greatly reduces the noise around the text compared with other methods, which can provide a solid foundation for subsequent document image recognition.



(a) Input (b) RetinexNet (c) KinD

(d) RAUS (e) Zero-DCE (f) Zero-RADCE

Figure 5: Ablation study of low-light historical Tibetan document images. The blue boxes indicate that there are significant differences in the binarization of the corresponding parts.

At the same time, subjective evaluation is also carried out to quantify the subjective visual quality of each method. This paper use different methods to enhance low-light historical Tibetan documents, and display the enhanced results on the screen and provide input images for reference. Fifteen subjects from different academic backgrounds rated the visual quality of the enhanced images on a scale of 10. Subjects used the following indicators as evaluation criteria: 1) Whether there was a significant contrast between the text and the background in the results; 2) Whether there is uneven texture and noise in the background; 3) whether the results contain over-/under-enhanced regions. The average subjective scores are reported in Table 1, Zero-RADCE had the highest average score in image enhancement of historical Tibetan document iamges.

Table 1: Subjective evaluation score.

Method	Retinex-Net	KinD	RAUS	Zero-DCE	Zero-RADCE
Average score	7.9	7.3	7.1	7.8	8.2

4.1.2. Objective Evaluate.

To quantitatively compare the enhancement performance of different methods on low-light historical Tibetan document images, Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM), and Mean Square Error (MSE) measures are used for full-reference image quality assessment. The following data are the average values obtained by testing the image results of 22 low-light historical Tibetan document images. In Table 2, the proposed Zero-RADCE achieves the best values under all cases.

Table 2: Objective evaluations in terms of full-reference image quality assessment metrics.

Method	PSNR ↑	SSIM ↑	MSE ↓
RetinexNet	24.83	0.93	93.24
KinD	20.27	0.90	100.20
RAUS	14.70	0.86	114.83
Zero-DCE	20.39	0.94	103.21
Zero-RADCE	24.96	0.96	82.97

It's worth not that when the input image size is 256×256 ×3, the number of model parameters in Zero-RADCE is reduced from 79416 to 48026 compared with Zero-DCE, and the NUMBER of Flops is reduced from 5.21G to 1.53G, realizing a more lightweight network.

5. Conclusion

This paper proposed Zero-RADCE for low-light historical Tibetan document image enhancement. In the network model, the encoder-decoder structure, residual network, and spatial attention mechanism are adopted, and the loss of spatial consistency is improved. It can not only enhance low-light document images, but also highlight text information and reduce the existing stain. The experiment proves the superiority in the improved Zero-RADCE method of image processing of historical Tibetan document images with low-light. In future work, will further try to improve the fad problem of the original text.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 61772430 and Grant 62166036, in part by the Program for Leading Talent of State Ethnic Affairs Commission, in part by the Program for Innovative Research Team of State Ethnic Affairs Commission (SEAC) ([2018]98).

References

[1] De, S., Geng, G. Y. X.: Research on the sharing model of Tibetan literature resources under the network environment. Tibetan Studies in China 02, 202-206(2013).
[2] Guo, C., Li, C., J, Guo, et al.: Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1780-1789. (2020).
[3] Wei, Chen, et al.: Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560. (2018).
[4] Zhang, Y., Zhang, J., Guo, X: Kindling the darkness: A practical low-light image enhancer. In: Proceedings of the 27th ACM international conference on multimedia, pp. 1632-1640. (2019).
[5] Liu, R., Ma, L., Zhang, J., Fan, X., Luo, Z.: Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10561-10570. (2021).
[6] Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing 13(4), 600-612. (2004).
[7] Pizer, S. M., Amburn, E. P., Austin, J. D., et al.: Adaptive histogram equalization and its variations. Computer vision, graphics, and image processing 39(3), 355-368. (1987).
[8] Farid, H.: Blind inverse gamma correction. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society 10(10), 1428. (2001).
[9] Land, E. H.: The Retinex Theory of Color Vision. Scientific american 237(6), 108-129. (1977).
[10] Ronneberger, O., Fischer, P., Brox, T: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Springer, Cham (2015).
[11] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 770-778 (2020).
[12] Woo, S., Park, J., Lee, J. Y., Kweon, I. S.: Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision, pp. 3-19. (2018).
[13] Akagic, A., Buza, E., Omanovic, S., Karabegovic, A.: Pavement crack detection using Otsu thresholding for image segmentation. In: 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics, pp. 1092-1097. (2018).