COMPUTER ENGINEERING

UDC 004.032.26(045)
DOI:10.18372/1990-5548.85.20429

D. V. Prochukhan

A CLASSIFICATION METHOD FOR OPTICAL COHERENCE TOMOGRAPHY IMAGES BASED ON A STRUCTURE-ORIENTED ADAPTIVE NEURAL NETWORK ARCHITECTURE

[1]Kharkiv National University of Radio Electronics, Kharkiv, Ukraine
E-mail: viprochuhan@gmail.com

***Abstract*—*The method of optical coherence tomography image classification for automated diagnosis of diabetic retinopathy and diabetic macular edema is proposed in the article. An innovative adaptive multi-task deep neural network is created. It simultaneously solves the problems of pathology classification and structural feature reconstruction. The neural network uses the pre-trained EfficientNetB7 model as an encoder for efficient extraction of high-level features. The structural feature learning branch (decoder) is responsible for restoring spatial information. It increases the resolution of feature maps to the original size of 224x224 pixels with a gradual decrease in the number of filters and the use of Batch Normalization to stabilize learning. The classification branch combines semantic and structural features. It uses the channel attention mechanism for dynamic weighting of informative channels. Dropout and Batch Normalization layers are used to prevent overtraining in the classification branch. The model is optimized using a multi-task loss function. It consists of a modified loss function for classification (with class weights to balance data imbalance) and a root-mean-square error for structural loss. Training is performed using the Adam optimizer and the EarlyStopping, ModelCheckpoint, and ReduceLROnPlateau callbacks. The experiment was conducted on the OCT Image Classification dataset. Data augmentation (horizontal reflections) was performed to increase the number of images. High accuracy rates and cost functions were obtained as a result of training. The multi-task method enables the encoder to learn details and boundaries of the retina through Canny edge reconstruction. It contributes to improved classification and provides a powerful internal regularization mechanism, increasing the generalization ability of the model.*

Keywords—Artificial intelligence; image classification; deep learning; tomography; algorithm.

I. INTRODUCTION

Diagnosis of eye diseases is a critical task in modern medicine. Preserving vision and patients' quality of life depends on the timely detection of pathologies. Traditional diagnostic methods, such as visual examination and clinical tests, rely heavily on the physician's expertise and are labor-intensive. In recent years, the volume of medical data has been increasing significantly, creating a demand for new tools capable of efficiently processing this information. The integration of artificial intelligence technologies, along with machine learning and deep learning methods, provides new opportunities. These approaches automate medical workflows, enhance diagnostic accuracy and improve processing speed.

II. PROBLEM STATEMENT

The growing number of patients and the increasing volume of medical data place a significant burden on healthcare systems. The process of analyzing and interpreting diagnostic images is complex. This complexity in establishing a diagnosis often leads to delays in initiating treatment for eye diseases. Existing methods are not always capable of handling the information flow, and the risk of physician error remains high. Therefore, developing efficient, automated systems is a pressing task. Such systems would enable rapid and accurate screening. The creation and optimization of an innovative convolutional neural network architecture is a priority research goal, serving as the foundation for implementing an automated eye disease diagnostic system. Automated diagnosis of retinal diseases using optical coherence tomography is an important area of scientific research. Convolutional neural networks have demonstrated significant capabilities, surpassing traditional methods in both accuracy and efficiency. Several studies have focused on the use of architectures such as VGG, ResNet, and Inception, which effectively process large volumes of medical data. Transfer learning enables models to quickly adapt to the specific characteristics of medical images. ResNet models have been successfully applied for detecting various eye diseases in optical coherence tomography images [1], [2]. The EfficientNet model provides high performance [3]. Multitask learning allows a model to simultaneously learn to perform multiple related tasks, which improves performance and generalization ability [4]. Segmentation and classification of affected lung regions were successfully conducted in a study [5], and multitask learning has also been applied for tumor detection [6]. Edge detection is a fundamental task in computer vision, playing a key role in identifying objects and their boundaries. The Canny operator is widely used for detecting both sharp and noisy edges [7]. Neural networks have also been successfully applied to this task. Encoder-decoder architectures have proven effective in detecting blurred edges [8], [9]. Attention mechanisms have become an integral part of modern deep learning architectures, allowing models to focus on the most relevant features.

Channel attention mechanisms compute weights for each channel, enhancing important channels while suppressing less relevant ones [10], [11].

Hybrid neural networks have been successfully used for classifying images with signs of eye diseases [12] – [14].

The purpose of this work is to develop an innovative method for classifying optical coherence tomography images by creating an adaptive, structurally-oriented neural network.

III. Problem Solution

The Canny algorithm is widely used in the processing of optical coherence tomography images. The algorithm reduces noise without losing important information. It computes the gradient magnitude and direction to detect regions with significant changes in intensity. Such regions are potential edge indicators. The Canny algorithm uses two threshold values (high and low) to classify strong and weak pixels. It is also used to determine the boundaries between different retinal layers, which is important for detecting anomalies. An example of the Canny algorithm's application is shown in Fig. 1.

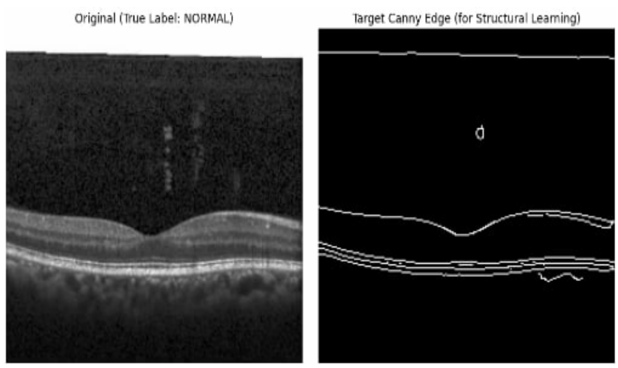

Fig. 1. Application of the Canny algorithm for image processing

Accurate boundary detection is crucial for diagnosing diabetic retinopathy and diabetic macular edema. The Canny algorithm effectively detects features of diabetic retinopathy by outlining the borders of hemorrhages, microaneurysms, and exudates, as well as masked and blurred pathological features. It also identifies signs of diabetic macular edema by detecting the inner and outer boundaries of thickened retina, which helps assess the extent of the swelling. Furthermore, it detects dark circular and oval regions with bright borders, indicating retinal detachment.

It is advisable to create a hybrid neural network integrating the Canny algorithm to solve the optical coherence tomography image classification problem. The main goal in developing the network is to combine precise edge detection with high-level feature extraction. There are methods to integrate the Canny algorithm into a convolutional neural network. The first method is use Canny as an additional input channel. The second method is use Canny in the loss function. The Canny output can regularize the loss function, encouraging the network to detect edges. The third method is use Canny as a preprocessing filter for the initial convolutional neural network layers to initialize the weights of some early convolutional layers, potentially helping the network converge faster. Fourth method is use Canny for label generation. The network is trained using labels created from Canny outputs. Fifth method is integrate Canny into hybrid architecture blocks within the convolutional neural network, where part of the processing is done with Canny and merged into the convolutional stream. Sixth method is use Canny as an attention mechanism by generating attention masks to guide the network toward regions with important edges. Seventh method is train convolutional filters to mimic Canny-like edge detection so the initial layers inherently detect similar edge features.

The second method is preferable. It can be implemented through an effective auxiliary task where the network is trained simultaneously on two related objectives: classification and reconstruction of edge maps. The model receives an additional signal emphasizing the importance of local texture information without increasing its size or adding extra branches or channels at the input. This improves sensitivity to subtle textural differences that are critical for detecting diabetic retinopathy and diabetic macular edema. The method provides a strong mechanism for internal regularization, fine-tunes the contribution of structural components, and improves the model's generalization ability in classification tasks. Three OCT image classes from the OCT Image Classification dataset were used: 3000 healthy retina images, 3000 images with signs of diabetic retinopathy, 3000 images with signs of diabetic macular edema.

The images were decoded and resized to 224×224 pixels using bilinear interpolation. Pixel values were normalized to the range [0,1]. Random horizontal flipping was applied to improve generalization and reduce overfitting, as it is a safe and effective augmentation. Vertical flipping was not used since it creates unrealistic anatomical configurations. A Canny edge map was generated for each image through the following steps: convert image to grayscale, apply noise reduction, detect edges using Canny with low and high thresholds, normalize the resulting edge map to [0,1]. The EfficientNetB7 model was used as a powerful feature extractor, applied without its top layers. Output features were passed into a structural feature learning module (decoder), which reconstructs spatial information and Canny edge maps obtained from the encoder. The decoder takes encoder features and progressively upsamples them to the original 224×224 resolution. The next part of the neural network consists of five consecutive blocks that increase the resolution. Each block doubles the spatial dimensions and contains the following layers: UpSampling2D(2, 2) for interpolation and increasing the spatial dimensions, Conv2D with a ReLU activation function and 3×3 kernels for processing newly created details and integrating information from neighboring pixels, BatchNormalization for stabilizing training by normalizing the outputs of the convolutional layer before activation.

This sequence of blocks allows the decoder to efficiently transform the compact feature representation from the encoder into a high-resolution spatial map. This map is then used to predict an edge map, similar to what is obtained using the Canny algorithm.

The final convolutional layer generates a single-channel output map. The classification branch is responsible for the main task: pathology classification. It combines information from two sources. Global Feature Averaging is applied to the output features to obtain a high-level feature vector. Global Feature Averaging is also applied to the output of the structural branch. These two feature vectors are concatenated. A channel attention mechanism is then applied to five sequential upsampling blocks, each doubling the spatial resolution, containing: UpSampling2D(2, 2), Conv2D (3×3 kernel, ReLU activation), BatchNormalization.

This sequence transforms the encoder's compact feature representation into a high-resolution spatial map for edge prediction. The final convolutional layer produces a single-channel output map. The classification branch combines two sources of information: global average pooling of high-level features, global average pooling of structural features from the decoder branch. These feature vectors are concatenated, followed by a channel attention mechanism to recalibrate channel weights, giving higher importance to the most informative channels. The recalibrated features are processed by a multilayer perceptron (two Dense layers) and passed through a sigmoid activation to produce channel weight coefficients. These coefficients are multiplied with the input features to enhance or suppress channels.

The resulting features are processed through additional Dense layers with L2 regularization, ReLU activation, Dropout and BatchNormalization to prevent overfitting. The final Dense layer outputs the probabilities for the three image classes.

The model is optimized using a composite loss function with two main components, each addressing a separate task: image classification and structural information learning.

A modified loss function is used for the classification task with class weights dynamically calculated based on the frequency of each class in the training set. This prevents larger classes from dominating and improves accuracy for smaller classes. This is the formula for the weighted cross-entropy loss function

$$L_{cls} = -\frac{1}{N}\sum_{i=1}^{N} w_{y_i} \sum_{j=1}^{C} y_{i,j} \log(\hat{y}_{i,j}), \quad (1)$$

where N is the batch size; $c = 3$ is the number of classes; $y_{i,j}$ is the true binary label for the ith sample and the jth class; $\hat{y}_{i,j}$ is the predicted probability for the ith sample to belong to the jth class; w_{y_i} is the weight assigned to the true class.

The MSE function is used for the auxiliary task of edge map reconstruction. This function measures the mean square difference between the model-predicted edge map and the target edge map. The structural loss formula is

$$L_{str} = \frac{1}{M}\sum_{k=1}^{M} (T_k - P_k)^2, \quad (2)$$

where M the total number of pixels in the edge map; T_k the pixel value at the kth position of the target

Canny map; P_k the value of the pixel at the kth position of the predicted structure map.

The formula for the general loss function is:

$$L_{total} = W_{cls} \times L_{cls} + W_{str} \times L_{str}. \quad (3)$$

In the created model the loss weights are as follows: for the classification loss (the main task) and for the structural loss.

This allows us to maintain the priority of classification, while encouraging the model to learn useful structural representations.

The model is compiled using the Adam optimizer. EarlyStopping, ModelCheckpoint, and ReduceLROnPlateau callbacks are used to improve the stability of training and prevent overtraining. High accuracy rates and loss functions are obtained.

The accuracy graph during training is presented in Fig. 2.

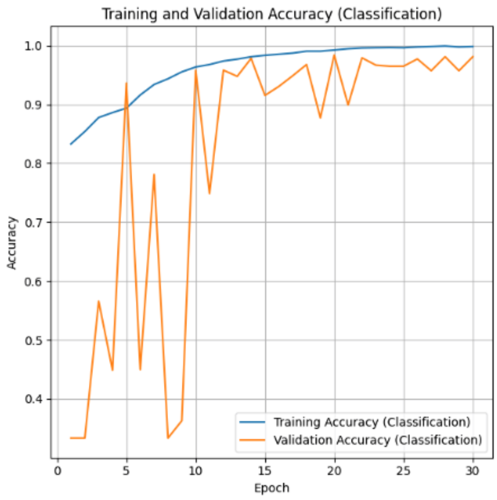

Fig. 2. Accuracy graph

Graphs of the structural loss function is presented in Fig. 3.

Fig. 3. Graphs of the structural loss function

Graphs of the classification and general loss function are presented in Fig. 4.

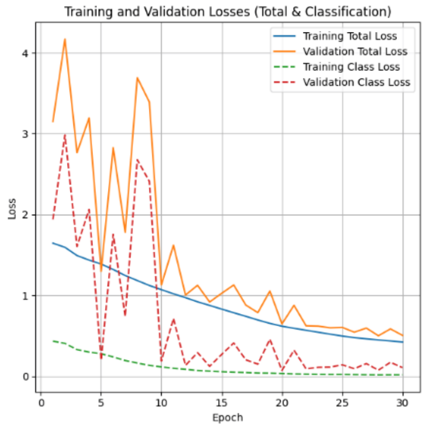

Fig. 4. Graphs of the classification and general loss function

The confusion matrix is presented in Fig. 5.

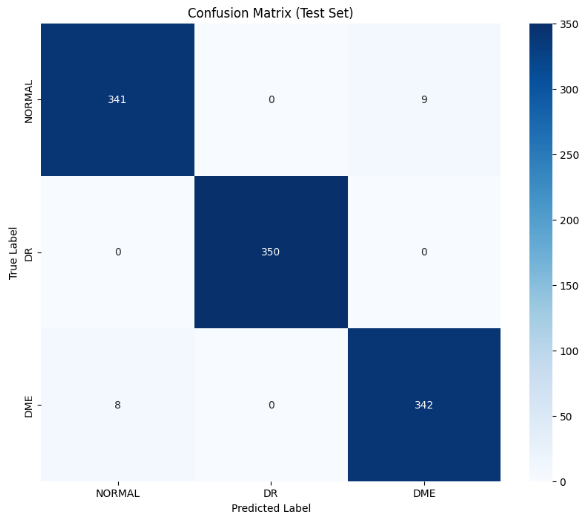

Fig. 5. Confusion matrix

The classification report is presented in Fig. 6.

```
Classification Report
              precision    recall  f1-score   support

      NORMAL       0.98      0.97      0.98       350
          DR       1.00      1.00      1.00       350
         DME       0.97      0.98      0.98       350

    accuracy                           0.98      1050
   macro avg       0.98      0.98      0.98      1050
weighted avg       0.98      0.98      0.98      1050

Weighted F1-score: 0.9838
Weighted ROC AUC: 0.9992
```

Fig. 6. Classification report

IV. CONCLUSIONS

The multi-task approach to classifying retinal optical coherence tomography images with comprehensive feature extraction has several advantages. These advantages make it more efficient compared to traditional models. The neural network simultaneously learns to perform two complementary tasks – classification and structural feature recovery. The auxiliary task of Canny-edge reconstruction prompts the encoder to learn details and boundaries in the retina. The method creates an informative representation of the image to improve classification. Adaptive use of channel attention allows the model to dynamically weight the importance of each channel. Adaptive feature integration allows the system to more effectively combine information from different sources, whic6h makes the final classification solution accurate and reliable. The use of a pre-trained EfficientNetB7 as the base encoder provided high efficiency in feature extraction. The model adapted to the unique characteristics of optical coherence tomography images by thawing and fine-tuning the layers of the base model. Intermediate integration at the level of aggregated features with an attention mechanism has been successfully applied. Edge reconstruction helps the encoder learn features for classification, and classification improves the quality of reconstruction. The use of the proposed hybrid architecture is a prospect for further research to solve other medical image processing problems.

REFERENCES

[1] H. Y. Li, D .X. Wang, L. Dong, and W.B. Wei, “Deep learning algorithms for detection of diabetic macular edema in OCT images. A systematic review and meta-analysis,” *Eur J Ophthalmol*, 2023, no. 33(1), pp. 278–290. https://doi.org/10.1177/11206721221094786

[2] S. Manikandan, R. Raman, R. Rajalakshmi, S. Tamilselvi, and R. J. Surya, “Deep learning-based detection of diabetic macular edema using optical coherence tomography and fundus images: A meta-analysis,” *Indian J Ophthalmol,* no. 71(5), pp. 1783–1796, 2023. https://doi.org/10.4103/IJO.IJO_2614_22

[3] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” *In International Conference on Machine Learning*, 2019, pp. 6105–6114.

[4] R. Caruana, “Multitask Learning // Machine Learning,” vol. 28, no. 1, pp. 41–75, 1997. https://doi.org/10.1023/A:1007379606734

[5] S. Li, “Thoracic Disease Classification and Localization Using a Multi-task Deep Learning Framework,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1657–1666. 2019.

[6] C. Chen, “Deep Learning for Lung Nodule Classification and Segmentation: A Survey,”, *Journal of Medical Systems*, vol. 43, no. 1, p. 25, 2019.

[7] J. Canny, “A Computational Approach to Edge Detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, 1996, pp. 679–698. https://doi.org/10.1109/TPAMI.1986.4767851

[8] S. Xie and Z. Tu, “Holistically-Nested Edge Detection,” *In Proceedings of the IEEE International Conference on Computer Vision,* 2015, pp. 1391–1399. https://doi.org/10.1109/ICCV.2015.164

[9] X. Xavier. “DexiNed: Dense Extreme Inception Network for Edge Detection,” *IEEE Access,* 2019, vol. 7.

[10] J. Hu, L. Shen, and G. Sun, “Squeeze-and-Excitation Networks,” *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 2018, pp. 7132–7141. https://doi.org/10.1109/CVPR.2018.00745

[11] S. Woo, J. Park, J. Lee, and I. S Kweon, “CBAM: Convolutional Block Attention Module,” *in Proceedings of the European Conference on Computer Vision (ECCV),* 2018, pp. 3–19. https://doi.org/10.1007/978-3-030-01234-2_1

[12] D. V. Prochukhan, “Osoblyvosti konkatenatsii zghortkovykh neironnykh merezh dlia skryninhu diabetychnoi retynopatii,” *Systemy obrobky informatsii,* no. 1(176), pp. 89–94. 2024. [in Ukrainian] https://doi.org/10.30748/soi.2024.176.11

[13] D. V. Prochukhan, “Syntez zghortkovykh neironnykh merezh ta dovhoi korotkochasnoi pamiati dlia detektuvannia proliferatyvnoi retynopatii,” *Visnyk Natsionalnoho tekhnichnoho universytetu “KhPI”. Seriia: Informatyka ta modeliuvannia*, vol. 1, no. 1–2 (11–12), pp. 76–86, 2024. [in Ukrainian]. https://doi.org/10.20998/2411-0558.2024.01.06

[14] V. Sheketa, V. Pikh, M. Slabinoha, and Y. Striletskyi, “Metodolohiia optymizatsii neironnoi merezhi z intehratsiieiu pertseptronnykh komponentiv realizovanykh na plis,” *Herald of Khmelnytskyi National University. Technical Sciences,* no. 351, pp. 415–427, 2025. [in Ukrainian].

Received June 11, 2025

Prochukhan Dmytro. ORCID 0000-0002-4622-1015. Postgraduate student.
Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.
Education: National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine, (2019).
Research interests: deep learning, artificial intelligence.
Publications: 55.
E-mail: viprochuhan@gmail.com

Д. В. Прочухан. Метод класифікації зображень оптичної когерентної томографії на основі структурно-орієнтованої адаптивної нейронної мережі

В статті запропоновано метод класифікації зображень оптичної когерентної томографії для автоматизованої діагностики діабетичної ретинопатії та діабетичного макулярного набряку Створено інноваційну адаптивну багатозадачну глибоку нейронну мережу. Вона одночасно вирішує задачі класифікації патологій та реконструкцію структурних ознак. Нейронна мережа використовує попередньо навчену модель EfficientNetB7 як енкодер для ефективного екстракції високорівневих ознак. Гілка вивчення структурних ознак (декодер) відповідає за відновлення просторової інформації. Вона збільшує роздільну здатність карт ознак до вихідного розміру 224x224 пікселів з поступовим зменшенням кількості фільтрів та використанням Batch Normalization для стабілізації навчання. Гілка класифікації об'єднує семантичні та структурні ознаки. Вона застосовує механізм канальної уваги для динамічного зважування інформативних каналів. Шари Dropout та Batch Normalization використані для запобігання перенавчанню використані в класифікаційній гілці. Модель оптимізується за допомогою багатозадачної функції втрат. Вона складається з модифікованої функції втрат для класифікації (з ваговими коефіцієнтами класів для балансування дисбалансу даних) та середньоквадратичної помилки для структурної втрати. Навчання відбувається з використанням оптимізатора Adam та функцій зворотного EarlyStopping, ModelCheckpoint і ReduceLROnPlateau. Експеримент проведено на наборі даних OCT Image Classification. Аугментація даних (горизонтальні віддзеркалення) проведена для збільшення кількості зображень. Високі показники точності та функції витрат отримано в результаті навчання. Багатозадачний метод надає можливості енкодеру вивчати деталі та границі сітківки через реконструкцію Canny-країв. Він сприяє покращенню класифікації та забезпечує потужний механізм внутрішньої регуляризації, підвищуючи узагальнюючу здатність моделі.

Ключові слова: штучний інтелект; класифікація зображень; глибоке навчання; томографія; алгоритм.

Прочухан Дмитро Володимирович. ORCID 0000-0002-4622-1015.
Харківський національний університет радіоелектроніки, Харків, Україна Аспірант.
Освіта: Національний технічний університет «Харківський політехнічний інститут», (2019)
Напрям наукової діяльності: глибоке навчання, штучний інтелект
Кількість публікацій: 55.
E-mail: viprochyhan@gmail.com