Webology (ISSN: 1735-188X)
Volume 18, Number 6, 2021

Attention Mechanism In Segmentation And Classification Of Optical Coherence Tomography Images

M.Nagoor Meeral[1] , Dr. S.Shajun Nisha[2] , Dr. M.Mohamed Sathik[3]

[1]Ph.D Research Scholar, PG & Research Department of Computer Science, Sadakathullah Appa College, Rahmath Nagar, Tirunelveli, India. Affiliation of Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli 627012. Reg No.19211192282029

[2]Research Supervisor, Assistant Professor & Head, PG & Research Department of Computer Science, Sadakathullah Appa College, Rahmath Nagar, Tirunelveli, India.

[3]Principal, Sadakathullah Appa College, Rahmath Nagar, Tirunelveli, India.

Abstract

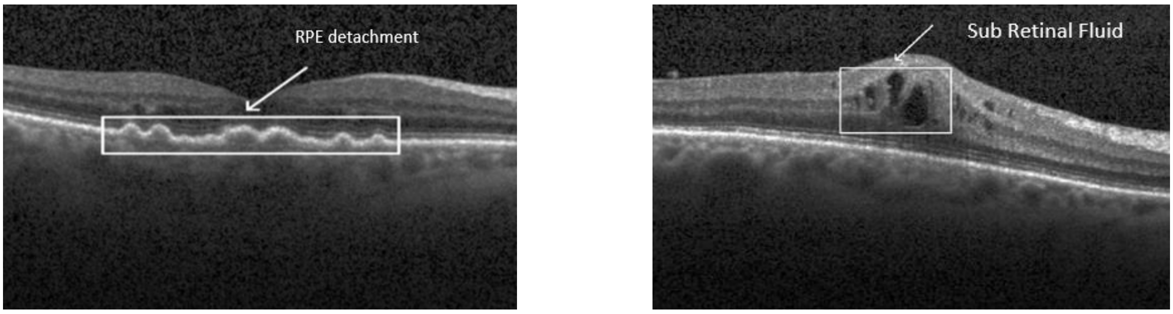
Humans have an inherent ability to locate important areas in complicated situations. Attention mechanism is a simple method ,included in the field of computer vision based on the human visual system. An attention mechanism is a dynamic weight adjustment procedure based on the input image's attributes. Many visual tasks, including as image classification, object recognition, semantic segmentation, video interpretation, 3D vision, multimodal tasks, and self-supervised learning, have seen considerable success using attention mechanisms. Recent research work focus on using attention mechanism in medical image analysis. Many macular diseases are classified using deep learning algorithms results in misinterpretation of class labels. The attention mechanism is leveraged in this field to diagnose retinal disorders with improved accuracy. We present a detailed evaluation of numerous attention processes in computer vision, the application of attention module in OCT image interpretation. Moreover, challenges and future research directions are also discussed in this paper.

Keywords: Attention, OCT, Computer vision, Retinal diseases.

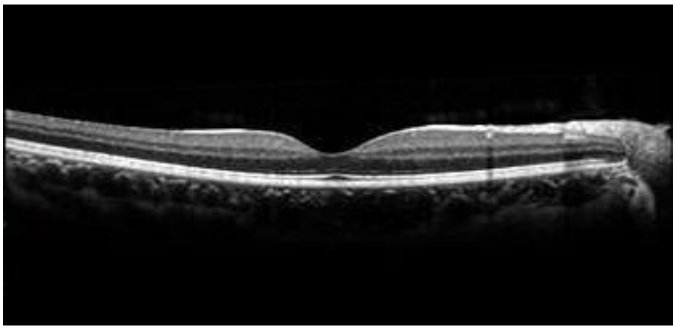
I. Introduction:

The most important retinal region responsible for central vision is Macula. The number of diseases affecting the macular region are Age-related Macular Degeneration (AMD),Diabetic Macular Edema (DME),Macular Hole(MH), Macular Pucker (MP) etc. These diseases causes severe structural changes in the retinal layers. The cross section of these retinal regions can be captured by standard imaging modality like Spectral Domain Optical Coherence Tomography(SD-OCT).SD-OCT is a non-contact interferometric modality which is capable of acquiring micrometric resolution of retinal structures. It can clearly manifest the biomarkers like hard exudates, sub retinal fluids, haemorrhages, wool spots etc[1][2].

In the recent decades, many computer aided techniques are available in the literature for OCT image classification. However, these methods are based on handcrafted feature extraction process. In [3],OCT volumes are classified using histogram of gradient informations, but the method requires retinal flattening and ROI extraction. In [4],linear binary patterns are used for classification. These conventional methods are database specific and semi-automatic in nature. It is avoided using deep convolutional neural network(CNN) which uses sequences of processing layers and yields better accuracy for the application of medical image classification. In [5] ,A CNN model is implemented to classify OCT volumes in surrogated images. In[6], a combination of CNN models are utilized for classification problems. In[7] Iterative Fusion CNN is implemented for OCT image classification of macular pathologies. These techniques focuses more on unwanted regions ,provides inaccurate results for multi class classification problems.



(a)



(b)

Fig 1 (a) (b)OCT images with pathological changes (b) Normal OCT image

To reduce these shortcomings, introduction of attention modules into CNN are employed in many OCT classification problems. For biomedical imaging, attention has been used for report generation [12], disease classification [8], organ segmentation [9] [10]and localization [11]. The main contributions of this paper are :

1. To present a systematic review of attention mechanism in computer vision problems and its brief description
2. Applications of attention mechanism in Macular disease Diagnosis

3. Future research in visual attention.

II Organization of the paper:

The rest of the paper is organized as follows. Section II comprises attention mechanism and its categories, Section III explains applications of attention mechanism in OCT images, Section IV illustrates challenges and Future Work , Section V includes Conclusion

III Attention Mechanism:

In the context of deep learning, attention is a technique that has the ability to choose and concentrate on relevant informations. In other words, attention is a method that tries to enhance the important features of an image while fading out the non-relevant information. The attention module is incorporated in pre trained networks to improve its efficiency. The general form of attention mechanism is

Attention=f(g(x),x)

Where g(x) refers the generating attention, f(g(x),x) represents the processing of input x on the basis of g(x) .The basic structure of attention module is shown in Fig 2.
The attention module consists of a simple 2D-convolutional layer, MLP(in the case of channel attention), and sigmoid function at the end to generate a mask of the input feature map.

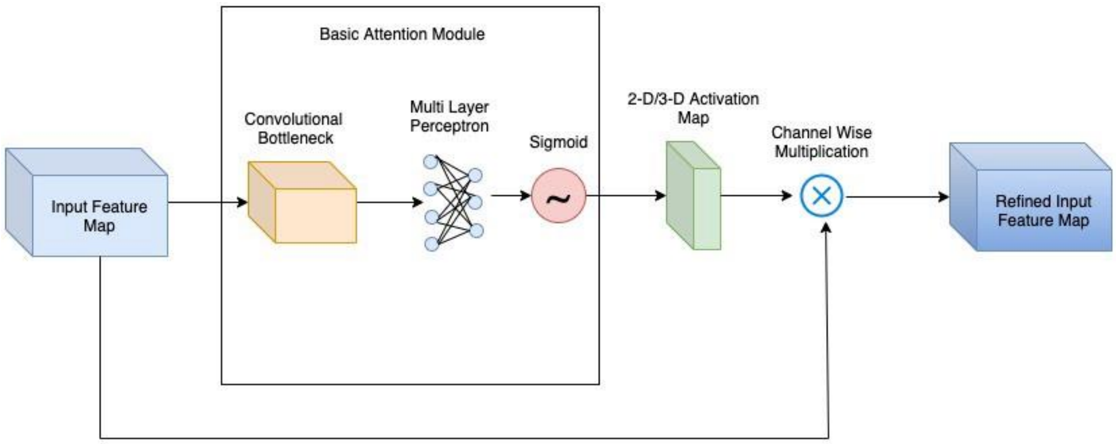


Fig 2. Basic Structure of Attention Module

The attention mechanism is divided into the step-by-step computations of the alignment scores, the weights and the context vector: There are three steps in computation of attention mechanism. They are context vector, alignment score and the soft max function. The context vector takes input from the previous layers and computes the probability distribution. It helps to capture global informations of the images. The context vectors are the sum of feature vectors of the input image, weighted by the attention coefficients, given by:

$$c_v = \sum_{i=1}^{n} af$$

Then the alignment score is generated for the encoded context vectors using the formula

$$a = \mathrm{align}(y, x)$$

Where y and x are the input feature maps to be aligned. There are different alignment score functions which is illustrated in table 1

Table 1. Different attention score functions.

Name	Alignment score function
Additive [13]	$\mathrm{score}(s_t,h_i)=v_a^T \tanh(W_a[s_t:h_i])$
Dot-product [14]	$\mathrm{Score}(s_t,h_i)= s_t^T.h_i$
Scaled Dot Product [15]	$\mathrm{Score}(s_t,h_i)= s_t^T.h_i/\sqrt{n}$
General [14]	$\mathrm{Score}(s_t,h_i)= s_t^T W_a h_i$

The soft max operation for the alignment score is calculated using the [13]

$$\mathrm{Softmax} = \frac{\exp(x)}{\sum \exp(x)}$$

3.1 Types of attention mechanism:

3.1.1 Self Attention:

When the receptive field is small in the convolution and pooling operation, the global information loss will occur. In order to resolve this issue,self attention mechanism is introduced. It helps to employ large receptive fields with no additional computational cost. Self attention focuses on long range dependencies between image regions. It is the mechanism of executing training with similar pixels and rejecting varying pixels to enhance the performance. Self attention mechanism computes the weighted average of all pixel to predict the corresponding value of pixel x in the ith position[16].The self attention mechanism is depicted in Fig 3.

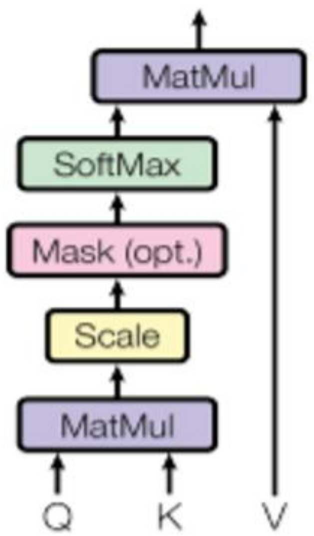


Fig 3. Self Attention

In an input feature map, the corresponding value of pixel X_i is assumed to be obtained by the weighted average of all X_j features.

$$y_i = \frac{1}{C(x)} \sum\nolimits_{\forall j} f(x_i, x_j) g(x_j)$$

The above formula shows the calculation method of the i^{th} output position. f is a function to calculate the relationship between i and all j, g is a transformation function to calculate the output characteristic representation of j position, and C (x) is the normalization factor [15]

3.1.2 Soft and Hard Attention:

- Soft Attention: It learns the alignment weights and placed "softly" over all patches in the source image. The model is smooth and differentiable. It is expensive when the source input is high.
- Hard Attention: It only selects one patch of the image to attend to at a time. It requires lesser calculations. But the model is non-differentiable and requires complicated network to train[17].

3.1.3 Global/Local Attention:

Global attention attends all the input feature maps and calculates the global align weights. Local Attention is different from Global Attention Model in a way that in Local attention model only a few patches from source is used to calculate the align weights. Local attention is a combination of hard and soft attentions. Like hard attention, it focuses on a subset. Like soft attention, it's differentiable and hence easier to implement and train. It's computationally simpler than global or soft attentions [14].The illustration of Global and Local attention is shown in Fig 4.

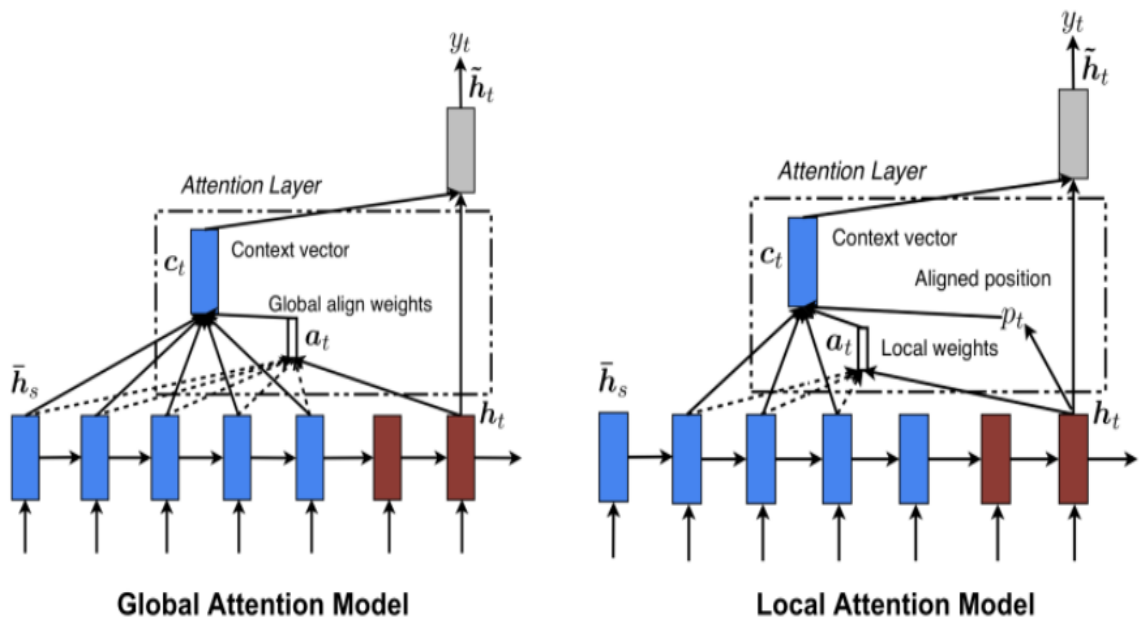


Fig 4 Global(Left) and Local attention(Right)

3.1.4 Convolutional Block Attention Module (CBAM):

The Convolutional Block Attention module is proposed in [18].It mainly focuses on extracting deep representations along spatial and channel axes. It consists of two sub modules arranged sequentially like Spatial Attention Module (SAM) and the Channel Attention Module (CAM)[18].

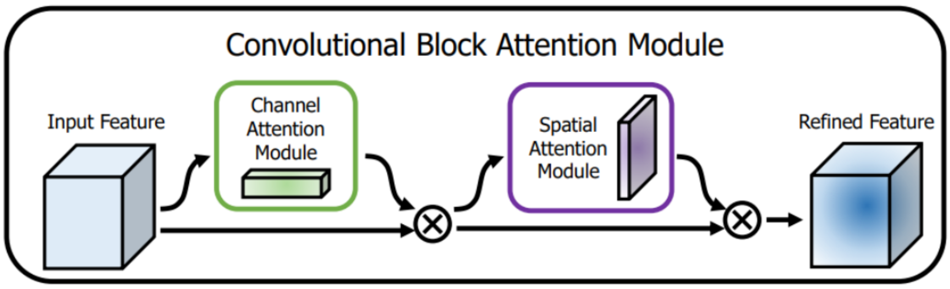


Fig 5. Convolutional Block Attention Module(CBAM)

Spatial attention Module

Spatial is the domain space associated with each feature maps. Spatial attention mechanism generates the mask of the feature maps that improves the representation of the whole image. This enhanced features will be fed as input to the subsequent convolution layers. Spatial attention module is introduced in Convolutional Block Attention Module (CBAM).This module uses the relationship of features between the spatial informations to produce a spatial attention map. It concentrates on the region which contains significant features and contribute to the specific applied task. The spatial attention module consists of average pooling and max pooling operations on channels and aggregate them to obtain an efficient feature representations. Finally, convolution layer is applied to the output spatial attention map[18].The operation is expressed mathematically as

$$M_s(F) = \sigma(f^{mxm}([Avgpool(F); Maxpool(F)])$$

Where F is the feature maps and f refers convolution operation. The illustration of patial attention is shown in Fig 6.

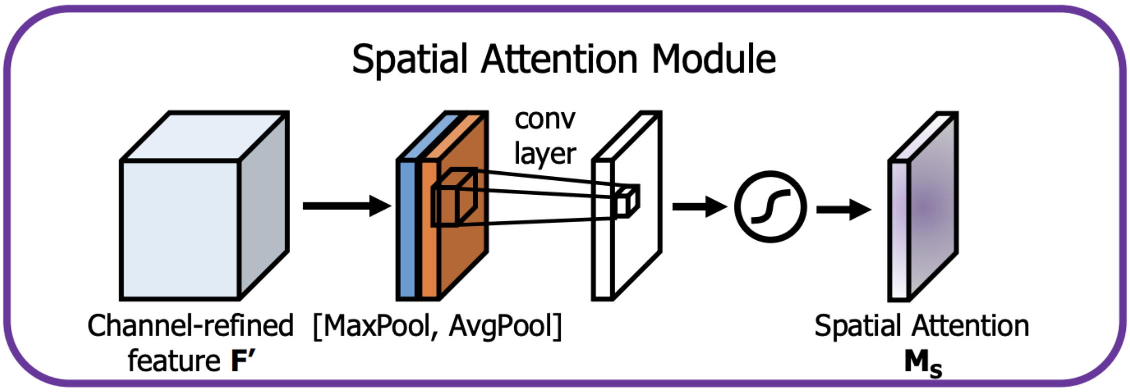


Fig 6. Spatial Attention Module

Channel Attention Module:

Channel attention focuses on significant channels which gives more contribution for better output results. After convolution operation, weights are assigned to the different channels. The channel with higher weight value will be the most important channel. The channel attention module aims to utilizing relationship between the channel features to generate attention map. It aggregates the spatial informations of feature map by implementing average pooling and max pooling operations. The context descriptor produced will be fed into shared Multi Layer Perceptron with a hidden layer. The output is then applied for element wise summation[18].The operation of channel attention is shown in Fig 7.The operation of channel attention module is expressed as

$$M_c(F) = \sigma(MLP([Avgpool(F) + Maxpool(F)])$$

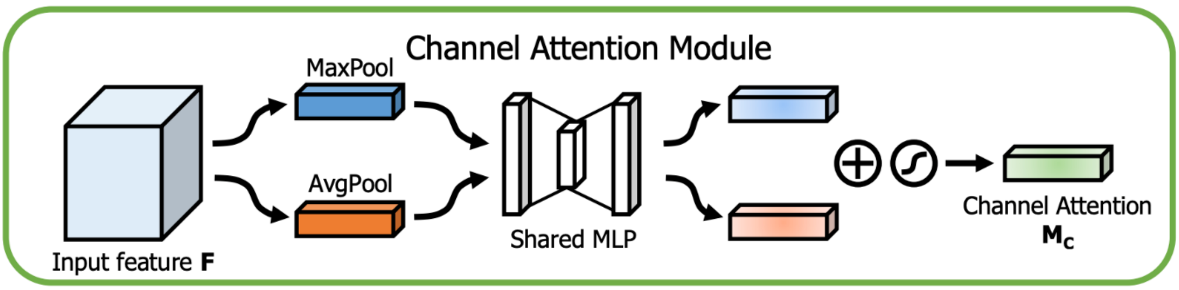


Fig 7. Channel Attention Module

3.1.5 Multi head Attention:

This type of attention accepts multiple heads and executed parallely. The individual outputs will be concatenated and linear transformation is applied to produce the output in required size[15].The operation of Multihead attention can be expressed as

Multihead $(Q,K,V)=[head_1,head_2,\ldots.head_h]W_0$

where W is the learnable weights. The operation of Multi head attention is shown in Fig 8

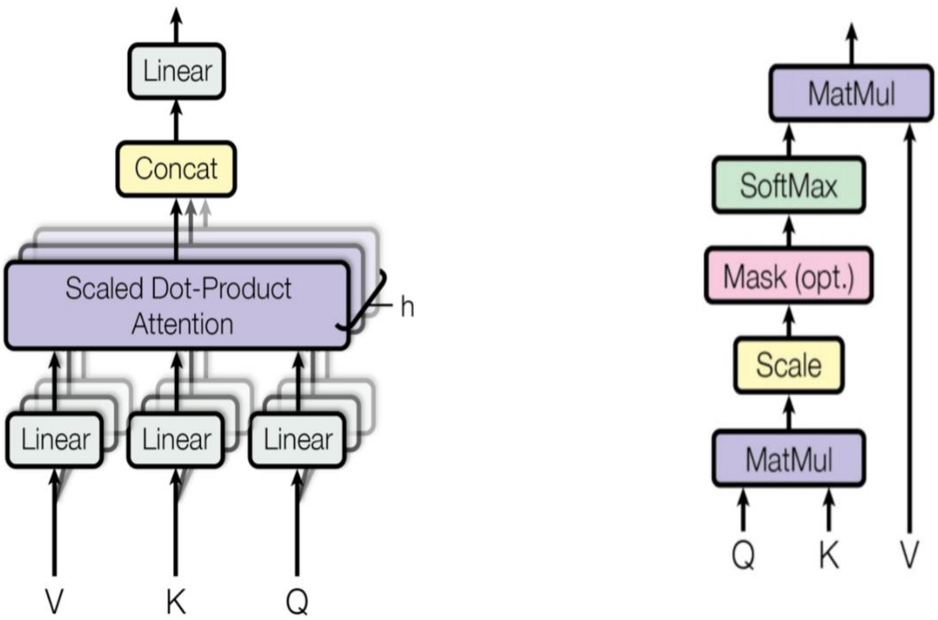


Fig 8 Multi Head Attention(Left) ,Scalar Dot Product (Right)

3.2 Application of Attention Mechanism in OCT image Interpretation:

In the recent years, The attention mechanism is incorporated in OCT image interpretation to focus more on the pathological changes in the retina. It helps to better diagnosis of diseases.

Ref.	Attention block	Purpose	Dataset and results	Description
[19]	Layer Integrated attention block	Classification of CNV, DRUSEN, DME, Normal	Accuracy: UCSD:88.4% HUCM:89.9%	The model incorporates layer Integrated block to process extracted ILM-RPE layer and focusing on meaningful lesion related information

[20]	Multi level dual attention	Classification of AMD, DME, Normal	Accuracy: NEH dataset:95.57%	Two types of attention multi level and self attention is included in the architecture to focus higher entropy regions at different level of feature subspaces.
[21]	Hybrid attention	Classification of AMD, DME, DRUSEN, NORMAL	Accuracy: NEH dataset: 99.76% Kermany dataset: 96.51%	Parallel spatial and channel attention is utilized in this network to extract key features of retinal informations and alleviate the background informations.
[22]	Residual Attention (RA) module	Classification of CNV, DME, DRUSEN, NORMAL	Accuracy: Kermany dataset: 99.5%	A residual attention module is incorporated with Inception network to distinguish OCT images. It need no additional labelling informations and efficiently learns the sensitive OCT informations.
[23]	Joint attention network	Classification of AMD, DME, Normal	Duke: Accuracy:100%, 99.36%, 99.68% OCT2017:92.40%95.60%,77.40%	Both supervised and unsupervised learning informations are aggregated using joint attention network. The architecture overcomes the spatial information loss.
[24]	Self attention mechanism	Classification of AMD,DME and Normal	Accuracy: Duke:97% NEH:95%	Spatial informations are extracted using CNN and the output is aggregated with attention module to generate distinguishable higher level representations
[25]	Informative Attention Convolutional Neural Network (IA-net)	Choroid Neo vascularization Segmentation	Dice coefficient: 0.8862	The model pays high attention to CNV feature maps .It increases the discriminative ability by including attention module

				for improving segmentation performance.
[26]	Skip Connection Attention	Choroid layer Segmentation	Dice coefficient: 0.951	A novel Skip Connection Attention (SCA) module which is integrated into the U-Shape architecture is proposed. The main function of the SCA module is to capture the global context in the highest level to provide the decoder with stage-by-stage guidance, to extract more context information and generate more consistent predictions for the same class targets.
[27]	Context Attention fusion	Segmentation of Intraretinal fluid (IRF), subretinal fluid (SRF) and pigment epithelial detachment (PED).	Dice coefficient: 74.64%	AF-Net proposes the context shrinkage encode (CSE) module and context pyramid guide (CPG) module to extract and fuse global context information. The CSE module embedded in the encoder path can ignore redundant information and focus on useful information by a shrinkage function. Besides, the CPG module is inserted between the encoder and decoder, which can dynamically fuse multi-scale information in high-level features.
[28]	Group-wise attention fusion network	Choroid segmentation	Dice coefficient: 95.21	GAF-Net proposes a group-wise channel module (GCM) and a group-wise spatial module (GSM) to fuse group-wise information. The GCM uses channel information to

				guide the fusion of group-wise context information, while the GSM uses spatial information to guide the fusion of group-wise context information.
[29]	Channel attention module	Segmentation of hyper-reflective foci (HF)	Dice coefficient: 73.26	There are two contribution in proposed architecture (1) multi-scale convolution based on dilated convolution is utilized to achieve adaptive receptive fields of the images. (2) In order to ignore irrelevant information and focus on key information in the channels, the channel attention module is embedded in the model.
[30]	Task specific attention	Retinal layers Segmentation	Dice coefficient: 0.912	The Task-specific attention module uses the advanced global responses to guide the model to learn the task tailored part from the shared features.
[31]	Attention Fully connected odule	Segmentation of Geographic atropy	Dice coefficient: 84.70	To extract available multi-scale features, a Scaling and Up Sampling (SUS) module is designed to balance the information content between features of different scales. To capture more discriminative features, an Attentional Fully Connected (AFC) module by incorporating attention mechanism is introduced.It enhances the significant informative features and suppress less useful ones. Based on the location cues, the final GA

				region prediction is obtained by the projection segmentation of MS-CAM

IV Challenges and future trends:

Extensive research works are required for spatial attention. Channel attention is well suited for classification purposes. but spatial attention lacks efficiency for semantic segmentation, object detection etc. Spatial attention evaluates where to pay attention whereas channel attention considers on what regions the attention has to pay. There is no generic model that utilizes all these characteristics of attention mechanism. Attention-based models are usually shown as attention maps. However, rather than providing a clear insight, this can merely provide an intuitive sense of what is going on. A better understanding of how procedures function, including failure mechanisms, is required. Attention models that are more characterisable and interpretable might be more frequently used. Attention models are well suited with large scale apre trained models for different inputs. Thus these pre trained model should be investigated further. A complex varied attention modules are difficult to apply. So. It is essential to identify simple and efficient model to deploy [32].

V Conclusion:

In the age of deep learning, attention mechanisms have become an essential method in the computer vision applications. This paper summarizes the popular attention mechanisms and its operation. The paper also explains applications of attention mechanism in OCT image interpretation for segmentation and classification of retinal diseases. Finally, the paper discusses challenges intention mechanism and future research directions. This research paper will give an overview to the researchers to get insights about attention mechanism and its importance in disease diagnosis.

References:

1. M. Schmitt, "Optical coherence tomography (OCT): a review," in IEEE Journal of Selected Topics in Quantum Electronics, vol. 5, no. 4, pp. 1205-1215, July-Aug. 1999, doi: 10.1109/2944.796348.
2. Bhende M, Shetty S, Parthasarathy MK, Ramya S. Optical coherence tomography: A guide to interpretation of common macular diseases [published correction appears in Indian J Ophthalmol. 2018 Mar;66(3):485]. Indian J Ophthalmol. 2018;66(1):20-35. doi:10.4103/ijo.IJO_902_17.
3. Srinivasan PP, Kim LA, Mettu PS, Cousins SW, Comer GM, Izatt,JA, Farsiu S. Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. Biomed Opt Exp. 2014;5(10):3568–77.
4. Wang Y, Zhang Y, Yao Z, Zhao R, Zhou F. Machine learning based detection of age-related macular degeneration (amd) and diabetic macular edema (dme) from optical coherence tomography (OCT) images. Biomed Opt Exp. 2016;7(12):4928–40.

5. Rong Y, Xiang D, Zhu W, Yu K, Shi F, Fan Z, Chen X. Surrogate assisted retinal OCT image classification based on convolutional neural networks. IEEE J Biomed Health Inf. 2018;23(1):253–63.
6. Li F, Chen H, Liu Z, dian Zhang X, shan Jiang M, zheng Wu Z,qian Zhou K. Deep learning-based automated detection of retinal diseases using optical coherence tomography images. Biomed Opt Express 2019; 10(12): 6204–6226. https:// doi. org/ 10. 1364/ BOE.10. 006204. http:// www. osapu blish ing. org/ boe/ abstr act. cfm? URI= boe- 10- 12- 6204
7. Fang L, Jin Y, Huang L, Guo S, Zhao G, Chen X. Iterative fusion convolutional neural networks for classification of optical coherence tomography images. J Vis Commun Image Rep. 2019;59:327–33.
8. Guan, Q., Huang, Y., Zhong, Z., Zheng, Z., Zheng, L., & Yang, Y. (2020). Thorax disease classification with attention guided convolutional neural network. Pattern Recognition Letters, 131, 38-45..
9. Wang, Y., Zhou, Y., Shen, W., Park, S., Fishman, E. K., & Yuille, A. L. (2019). Abdominal multi-organ segmentation with organ-attention networks and statistical fusion. Medical image analysis, 55, 88-102.
10. Lei, Y., Wang, T., Tian, S., Fu, Y., Patel, P., Jani, A. B., ... & Yang, X. (2021, February). Multi-organ segmentation of male pelvic CT using dual attention networks. In Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling (Vol. 11598, p. 115981L). International Society for Optics and Photonics.
11. Cai, J., Lu, L., Harrison, A. P., Shi, X., Chen, P., & Yang, L. (2018, September). Iterative attention mining for weakly supervised thoracic disease pattern localization in chest x-rays. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 589-598). Springer, Cham..
12. Xue, Y., Xu, T., Rodney Long, L., Xue, Z., Antani, S., Thoma, G. R., & Huang, X. (2018, September). Multimodal recurrent model with attention for automated radiology report generation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 457-466). Springer, Cham.
13. Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. Ar Xiv preprint arXiv:1409.0473.
14. Luong, M. T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. Ar Xiv preprint arXiv:1508.04025.
15. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.
16. Cheng, J., Dong, L., & Lapata, M. (2016). Long short-term memory-networks for machine reading. arXiv preprint arXiv:1601.06733.
17. Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., ... & Bengio, Y. (2015, June). Show, attend and tell: Neural image caption generation with visual attention. In International conference on machine learning (pp. 2048-2057). PMLR.

18. Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV) (pp. 3-19).
19. Huang, L., He, X., Fang, L., Rabbani, H., & Chen, X. (2019). Automatic classification of retinal optical coherence tomography images with layer guided convolutional neural network. IEEE Signal Processing Letters, 26(7), 1026-1030.
20. Mishra, S. S., Mandal, B., & Puhan, N. B. (2019). Multi-level dual-attention based CNN for macular optical coherence tomography classification. IEEE Signal Processing Letters, 26(12), 1793-1797.
21. Xu, L., Wang, L., Cheng, S., & Li, Y. (2021). MHA Net: A hybrid attention mechanism for retinal diseases classification. Plos one, 16(12), e0261285.
22. Shang, Z., Fu, Z., Liu, C., Xie, H., & Zhang, Y. (2018, December). Potential of attention mechanism for classification of optical coherence tomography images. In 2018 IEEE Visual Communications and Image Processing (VCIP) (pp. 1-4). IEEE.
23. Kamran, S. A., Tavakkoli, A., & Zuckerbrod, S. L. (2020, October). Improving robustness using joint attention network for detecting retinal degeneration from optical coherence tomography images. In 2020 IEEE International Conference On Image Processing (ICIP) (pp. 2476-2480). IEEE.
24. Das, V., Prabhakararao, E., Dandapat, S., & Bora, P. K. (2020). B-Scan attentive CNN for the classification of retinal optical coherence tomography volumes. IEEE Signal Processing Letters, 27, 1025-1029.
25. Xi, X., Meng, X., Qin, Z., Nie, X., Yin, Y., & Chen, X. (2020). IA-net: informative attention convolutional neural network for choroidal neovascularization segmentation in OCT images. Biomedical Optics Express, 11(11), 6122-6136.
26. Mao, X., Zhao, Y., Chen, B., Ma, Y., Gu, Z., Gu, S., ... & Liu, J. (2020, July). Deep Learning with Skip Connection Attention for Choroid Layer Segmentation in OCT Images. In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 1641-1645). IEEE.
27. Ye, Y., Chen, X., Shi, F., Xiang, D., Pan, L., & Zhu, W. (2021, February). Context attention-and-fusion network for multiclass retinal fluid segmentation in OCT images. In Medical Imaging 2021: Image Processing (Vol. 11596, p. 1159622). International Society for Optics and Photonics.
28. Cheng, X., Chen, X., Feng, S., Zhu, W., Xiang, D., Chen, Q., ... & Shi, F. (2020, March). Group-wise attention fusion network for choroid segmentation in OCT images. In Medical Imaging 2020: Image Processing (Vol. 11313, p. 1131332). International Society for Optics and Photonics.
29. Zhu, L., Zhu, W., Feng, S., & Chen, X. (2020, March). Fully automated segmentation of hyper-reflective foci in OCT images using a U-shape network. In Medical Imaging 2020: Image Processing (Vol. 11313, p. 1131308). International Society for Optics and Photonics.
30. Cao, J., Liu, X., Zhang, Y., & Wang, M. (2020, October). A Multi-task Framework for Topology-guaranteed Retinal Layer Segmentation in OCT Images. In 2020 IEEE

International Conference on Systems, Man, and Cybernetics (SMC) (pp. 3142-3147). IEEE.

31. Ma, X., Ji, Z., Niu, S., Leng, T., Rubin, D. L., & Chen, Q. (2020). MS-CAM: Multi-scale class activation maps for weakly-supervised segmentation of geographic atrophy lesions in SD-OCT images. IEEE Journal of Biomedical and Health Informatics, 24(12), 3443-3455.
32. Hafiz, A. M., Parah, S. A., & Bhat, R. U. A. (2021). Attention mechanisms and deep learning for machine vision: A survey of the state of the art. Ar Xiv preprint arXiv:2106.07550.