Proceedings of the 3rd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2025.CH23274

Improving U-net Model for Pulmonary Nodule Segmentation Through Attention Mechanism and Post-processing Module

Yuhan Cao

Yunnan University, Kunming, China
caoyuhan@stu.ynu.edu.cn

Abstract: Pulmonary nodule segmentation plays a crucial role in the early detection and diagnosis of lung cancer, significantly impacting patient outcomes. The U-net model has emerged as a useful architecture in medical image processing like pulmonary nodule segmentation, gaining widespread popularity. However, U-net suffers from poor fine segmentation capabilities and the presence of noise in the segmentation results. In this study, we propose an Enhanced U-net model to improve segmentation results. The Enhanced U-net uses ECA and CRF modules. ECA module can make the model focus more on important features and improve the fine segmentation ability of the model. Meanwhile, CRF module allows the model to further refine the results, reduce the noise and boundary discontinuities. Utilizing the LIDC dataset, we evaluate the model's performance through indicators such as recall, IoU, Dice score. Our findings show that Enhanced U-net can achieve the best performance among all U-net based models. And Enhanced U-net can significantly improve segmentation outcomes for small or blurred nodules.

Keywords: Pulmonary nodule, Semantic segmentation, U-net, Attention mechanism

1. Introduction

Pulmonary nodules are small abnormal growths in lung tissue that can indicate the early stages of lung cancer [1]. Early and precise diagnosis of these nodules is needed for effective treatment. However, the manual segmentation of pulmonary nodules in medical imaging is a labor-intensive and time-consuming process typically performed by radiologists. This task not only requires significant expertise but is also subject to variability among practitioners, leading to inconsistent results that can adversely affect diagnostic accuracy.

Computed Tomography (CT), a medical imaging technique, can provide cross-sectional images of the specific parts of a patient. CT scans are typically presented in DICOM format, which includes pixel data as well as critical metadata such as patient information and imaging protocols. The high resolution and detailed structure of CT images make them invaluable for diagnosing pulmonary nodules. However, the complexity and volume of data can overwhelm manual analysis, necessitating the development of automated segmentation methods.

Despite advancements in computational techniques, manual segmentation remains fraught with challenges. One significant issue is the difficulty in achieving fine edge segmentation because of the difference in nodule density, size, shape, and the presence of surrounding tissues that can obscure clear delineation. Furthermore, pulmonary nodules may appear across multiple slices in a three-dimensional space, complicating the learning process for neural networks that rely on consistent features across different layers. This spatial correlation is vital for capturing the complete morphology of nodules, yet many traditional segmentation methods struggle to integrate information effectively across CT layers.

In addition to these challenges, automatic segmentation methods are often limited by data sets. The reliance on limited training data can lead to overfitting, particularly when using complex models, thereby reducing the model's generalization capability. This paper aims to evaluate and compare the effectiveness of different advanced neural network models and finally propose an Enhanced U-net used in image segmentation. The models used for comparison include 2D U-net, 3D U-net, Efficient Channel Attention (ECA) networks, Squeeze-and-Excitation (SE) module, and Conditional Random Field (CRF) layers. By leveraging the Lung Image Database Consortium (LIDC-IDRI) dataset [2], we seek to enhance the accuracy and reliability of pulmonary nodule segmentation. And ultimately make contributions to improved clinical decision-making and patient outcomes.

2. Related works

The field of pulmonary nodule segmentation has evolved significantly over the past few decades, addressing various challenges through the innovative application of imaging technologies and computational methods. Kostis et al. [3] use an adaptive threshold method to initially separate pulmonary nodules, and then use morphological algorithms to improve the segmentation results, which can initially realize the segmentation task of lung nodules. Dehmeshki et al. [4] propose an adaptive segmentation method based on the contrast between pulmonary nodules and background. However, the segmentation results of the above methods are limited by the setting of initial conditions, and the segmentation effect in the edge region is not good.

The introduction of convolutional neural networks (CNNs) revolutionizes the field of medical image analysis, providing a more robust framework for classification. Kumar et al. [5] implement CNNs to classify benign and malignant pulmonary nodules. Long et al. [6] propose a fully convolutional neural network (FCN) to perform image segmentation. Instead of convolutional layers, they use fully connected layers to enhance the performance of the network. Ronneberger et al. [7] propose the U-net architecture based on FCN. U-net's unique encoder-decoder structure is crucial for extracting detailed features while preserving spatial information, which is vital for accurately delineating nodule boundaries in medical images. The strengths of the U-net model are its effectiveness in localizing features and its capacity to work well with limited training data. However, its reliance on 2D representations may limit its performance in capturing volumetric structures inherent in CT scans.

To address this limitation, Ahmed Abdulkadir et al. [8] introduce the 3D U-net architecture. This extension significantly enhances the model's capabilities by allowing it to process volumetric data directly, capturing richer contextual information and consequently improving segmentation accuracy in complex medical imaging scenarios. While the 3D U-net excels at recognizing spatial hierarchies, it comes with increased computational demands and memory usage, which can pose challenges in resource-constrained environments.

In a different approach, Jie Hu et al. [9] propose the Squeeze-and-Excitation (SE) module, which optimizes feature representation by modeling interdependencies among convolutional features. This innovation leads to more insightful learning processes, enhancing the model's sensitivity to salient features. However, the integration of the SE module can introduce additional computational overhead, potentially requiring more extensive resources.

To counterbalance this computation cost, Qilong Wang et al. [10] introduce the Efficient Channel Attention (ECA) module. ECA maintains the ability to capture cross-channel interactions while circumventing the dimensionality reduction challenges found in traditional attention mechanisms.

This results in a more computationally efficient model, but its efficacy may vary across different tasks compared to more complex attention mechanisms.

3. Methodology

To propose a robust neural network architecture, we test several variants of U-net and analyze the improvement strategy according to the segmentation results. To solve the problem of insufficient boundary feature extraction and the issue of noise, we add attention mechanism and post-processing module. The Enhanced U-net shows better performance than original models.

3.1. Data pre-processing

The CT images in the LIDC dataset are typically stored in DICOM format. DICOM files contain several information, such as patient information, and image orientation, making them vital for clinical applications.

For data preprocessing, we extract the CT images corresponding to each pulmonary nodule and resize them into either 2D segments of 32*56 pixels or 3D volumes of 24*32*56 pixels based on the nodule's center position. The processed images are divided into three subsets with the following distribution: 70% of the data is allocated for training, 10% for testing, and 20% for validation. To optimize runtime performance during model training, the images are saved in the more efficient .npy format, allowing us to bypass the DICOM reading process later.

For the training dataset, we horizontally flipped about 50% of the images and generated a random number in the range [0,90) for each image as the angle by which the image should be rotated.

3.2. Framework of Enhanced U-net

The proposed Enhanced U-net framework integrates the original 2D U-Net architecture with the ECA and CRF module. This combination aims to significantly improve the segmentation performance of pulmonary nodules by addressing insufficient boundary feature extraction and the issue of noise.

The base architecture of the Enhanced U-net retains the characteristic encoder-decoder structure of the original U-Net, which effectively captures multi-scale features through downsampling and upsampling pathways. The encoder progressively reduces the spatial dimensions while increasing the feature depth, allowing for the extraction of high-level contextual information. In contrast, the decoder reconstructs the segmentation map by merging features from the encoder, facilitating precise localization.

To enhance the extraction of boundary features, we introduce the ECA module to the encoder. The ECA module dynamically adjusts the channel-wise attention weights based on the learned importance of features, enabling the model to focus on critical areas. This increased focus on relevant features helps to overcome the challenge of insufficient boundary delineation.

Following the segmentation by the U-Net, the CRF module is employed as a post-processing step to refine the output. The CRF model leverages contextual information from neighboring pixels, allowing for the correction of misclassified areas and the smoothing of segmentation boundaries. This refinement step is particularly crucial in reducing noise and enhancing the spatial coherence of the segmented regions, leading to more accurate and clinically relevant results.

The framework of Enhanced U-net is shown in Figure 1.

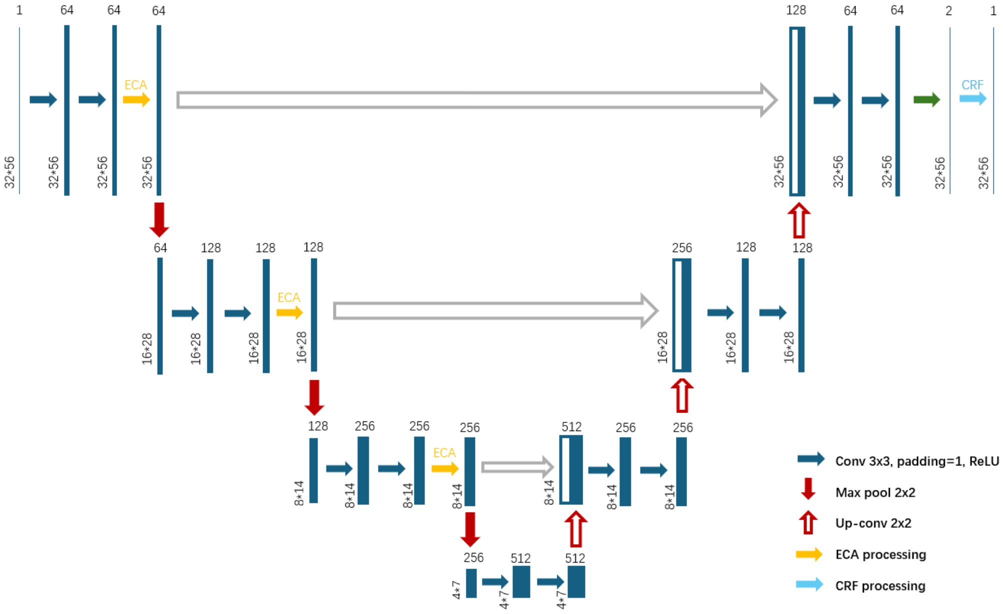

Figure 1: Architecture of Enhanced U-net

3.3. Improving fine segmentation with ECA module

The original U-net exhibits limitations in its ability to achieve fine segmentation at the boundary of pulmonary nodules. To address this concern, we introduce the ECA module, which allows the model to focus more on important features of pulmonary nodules, thereby improving the extraction of detailed boundary information.

ECA module begins with the global average pooling operation to summarize the feature responses of each channel. Then ECA module uses a one-dimensional convolutional layer with a 1D kernels of size 3 to capture channel dependencies and create attention weights of each channel. By applying these weights to the original feature maps, more important features can be paid more attention.

By using 1D convolution to model channel dependencies ECA module can avoid feature dimensionality reduction, thus maintaining more information from the feature maps. This ability to capture local inter-channel relationships leads to improved model performance without the overhead of fully connected layers.

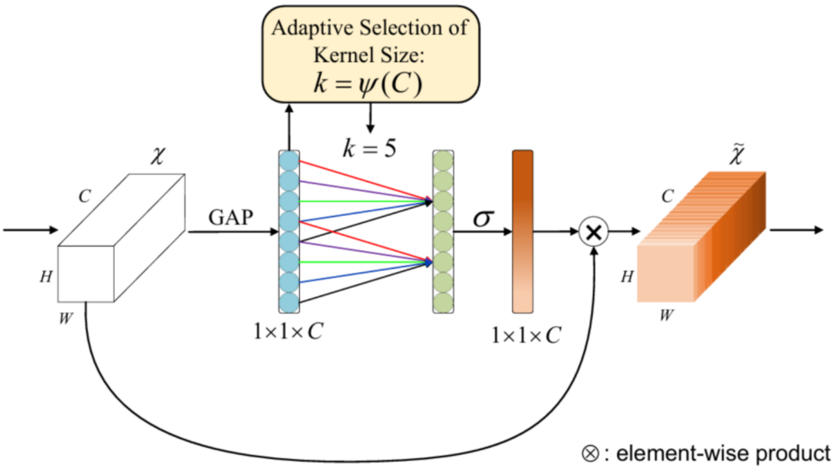

Figure 2: Effective Channel Attention module [9]

ECA module is incorporated into the U-Net architecture by positioning it before the maximum pooling operations in the encoder section. After each convolution operation in the encoder, the ECA module computes the importance of each channel based on the global contextual information, which can emphasize the most informative channels while suppressing less relevant ones. So the ECA module can enhance the feature representation, thereby facilitating improved capture of fine details and boundaries.

3.4. Post-processing with CRF module

The segmentation results of U-net with ECA may contain a small amount of noise. So, we use CRF module to further optimize the output. As a probabilistic graphical model, CRF module is used to improve segmentation outcomes in image segmentation.[11] CRFs can leverage the spatial coherence between neighboring pixels and enforce consistency within the segmented regions. By modeling the relationships and contextual information of pixels, CRFs can address the issues of noise and inaccuracies in preliminary segmentation results.

A CRF operates on the predictions of a deep learning model. First, U-net performs a preliminary segmentation and outputs the probability that each pixel belongs to a different class. Then the CRFs captures the interactions between pairs of pixels, encouraging neighboring pixels to have similar labels based on spatial proximity. CRFs need to minimize the energy function of the image, where the energy function is given by the formula (1).

$$E(y|x) = \sum_i U(y_i, x_i) + \sum_{(i,j)\in \mathrm{N}} V(y_j, x_j) \tag{1}$$

Where y denotes the label configuration, x represents the observed input features, i indexes individual pixels, and N denotes pairs of neighboring pixels.

In the Enhanced U-net, CRF module is positioned after the segmentation predictions are generated. By leveraging local contextual relationships, CRF module can enhance segmentation results and yield more coherent and precise segmentations.

4. Experimental results

4.1. Dataset and implementation details

We use Python as our programming language. The experimental environment is built based on PyTorch. The computer configuration is as follows: The operating system is Windows11; NVIDIA GeForce RTX 4060 LapTop GPU with 16GB RAM. The system memory is 16GB; AMD Ryzen 7 7735H with Radeon Graphics.

In this paper, all CT images are screened from the LIDC-IDRI [2] database and cut into 32*56 pixels or 24*32*56 pixels according to the central position of nodules, which were used as datasets for 2D and 3D models respectively. Then, according to the doctor's annotation of pulmonary nodule outline in XML, a pulmonary nodule outline label is generated, and the result is shown in Figure 3.

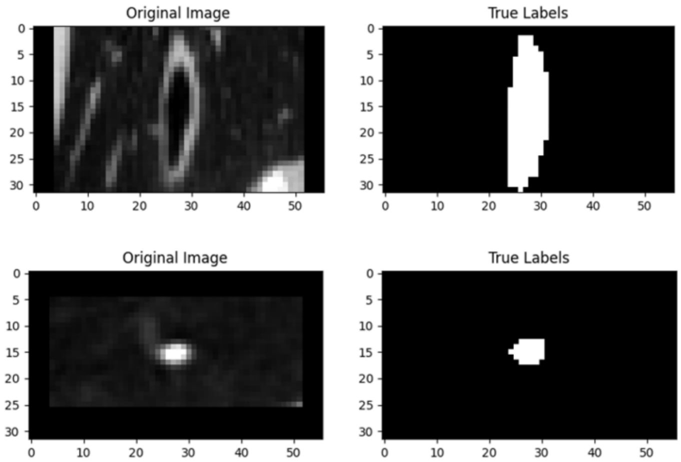

Figure 3: Experimental dataset

When choosing the loss function, we evaluate several loss functions, including Cross-Entropy Loss, Dice Loss, and Tversky Loss. Ultimately, we find that Weighted Cross-Entropy Loss yields the best performance. To address the issue of imbalanced pixel distribution between the background and the nodule, we test different weightings within the Cross-Entropy Loss function. And the best weight is [0.2, 1] for [background, nodule], respectively. Weighted Cross-Entropy Loss is given by the formula (2).

$$L = -\frac{1}{\mathrm{N}} \sum_{i=0}^{N} w_{yi} \cdot (y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \tag{2}$$

In the training process of the neural network, the optimizer is Adam optimizer. The batch size is 2. The momentum parameter is 0.95. The learning rate is 10^{-4}. Training iterations are selected as 50 rounds.

4.2. Evaluation index

We use the following evaluation indexes: recall, intersection over union(IoU), and dice score. The indexes above are calculated using the following formula.

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{3}$$

$$\mathrm{IoU} = \frac{1}{2} \cdot \left(\frac{\mathrm{TP}}{\mathrm{TP+FP+FN}} + \frac{\mathrm{TN}}{\mathrm{TN+FP+FN}} \right) \tag{4}$$

$$\mathrm{DiceScore} = \frac{\mathrm{TP}}{\mathrm{TP}+\frac{1}{2}\cdot\mathrm{FP}+\frac{1}{2}\cdot\mathrm{FN}} \tag{5}$$

Recall refers to the ration of correctly predicted nodule area and the summary nodule area. IoU calculates the ratio between the area of intersection and union between the predicted result and the ground truth. Dice Score assesses how closely the predicted result aligns with the ground truth, which is particularly valuable in scenarios with imbalanced classes. All the indicators are between 0 and 1. Higher values of the above indicators means a better model.

4.3. Experimental results and analysis

In order to accurately describe the performance of Enhanced U-net, we use the following models for comparison: 2D U-net, 3D U-net, 2D U-net with SE module, 2D U-net with ECA module, 2D U-net with CRF module and Enhanced U-net.

Table 1 shows the segmentation results of different networks on the test set. From the result, 3D U-net results are worse than 2D U-net results. That is because the number of background pixels in the data used by 3D U-net is much larger than the number of nodule pixels, so the current pixel is more likely to be predicted as the background when forecasting. Besides, after the addition of SE module in 2D U-net, the segmentation effect is almost the same as that of the original network or even slightly decreased. That is because SE module makes the model more complex, but the size of the data set is small, which may cause overfitting. After the CRF module is added to the original 2D U-net, the performance is significantly improved. The reason is that CRF considers global image information to enhance segmentation, while U-net can only capture local features and cannot use global context information.

The Enhanced U-net proposed in this paper has the best result of IoU and Dice Score among all models. That is because ECA module empowers the network by refining its feature representation capabilities, allowing it to mitigate the influence of less informative background areas. Concurrently, CRF module enhances the spatial coherence of the segmentation results by incorporating global contextual information, enabling the model to reduce boundary discontinuity. Consequently, the Enhanced U-net demonstrates the best performance among the evaluated networks.

Table 1: Results of different segmentation networks

Network	Recall	IoU	Dice Score
U-net	0.9050	0.7757	0.7638
3D U-net	0.9064	0.7496	0.7651
U-net+SE	0.9115	0.7681	0.7661
U-net+ECA	0.9246	0.7881	0.7703
U-net+CRF	0.8711	0.7835	0.7709
Enhanced U-net	0.9086	0.7908	0.7801

The results of different segmentation networks can be seen in Figure 4. In the figure, the pulmonary nodule image is in column 1, the gold standarder outline marked by doctors is in column 2, and the following columns are results using different models.

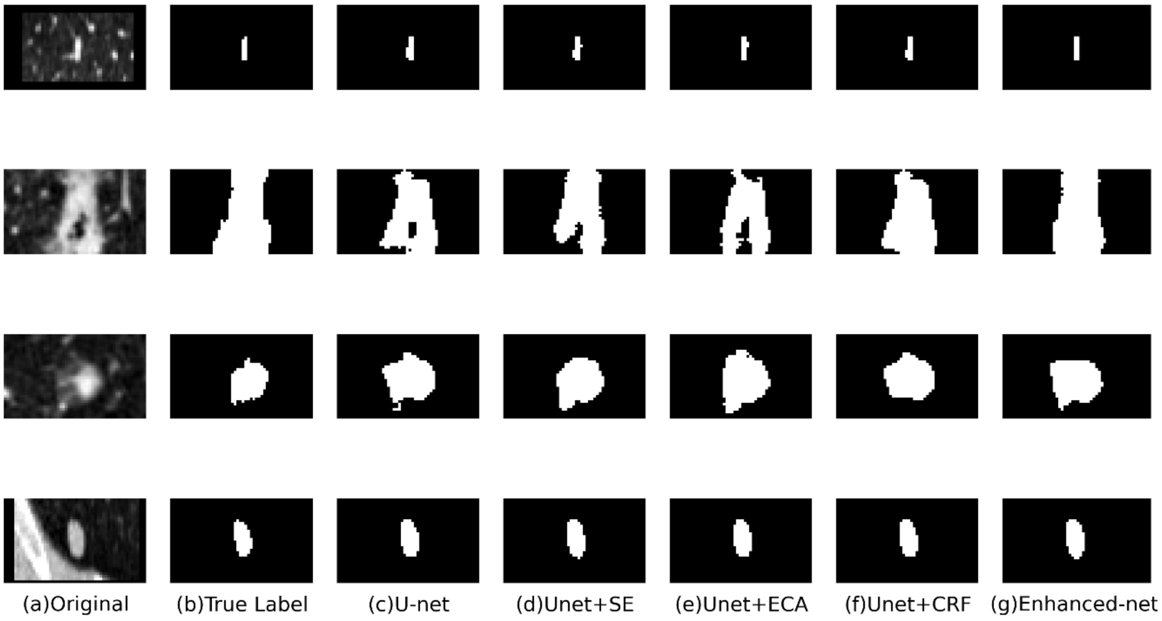

Figure 4: Comparison of different networks' results

As illustrated in Figure 4, the segmentation results from the original U-Net exhibit noticeable noise, particularly in the boundary regions. After integrating the ECA module, the accuracy of edge segmentation is improved, although some noise still persists. Further improvement is achieved with the introduction of the CRF module, which significantly reduces noise and substantially enhances the overall segmentation results. By combining the strengths of both the ECA and CRF modules, the Enhanced U-Net demonstrates the best segmentation performance, effectively minimizing noise while achieving better boundary accuracy.

5. Conclusion

In this paper, we propose an Enhanced U-net used for pulmonary nodule segmentation. The Enhanced U-net uses ECA and CRF modules. The ECA module enables the model to concentrate on the most significant features, enhancing its fine segmentation capability. At the same time, the CRF module facilitates the refinement of results, minimizing noise and discontinuities at the boundaries. The combination of ECA and CRF module can significantly optimize U-net segmentation results, which are suitable for pulmonary nodal segmentation tasks. However, the model in this paper still has some limitations. When constructing 2D data set, only one image was selected for each nodule, while other slices were ignored. As a result, the data set was small and the performance on complex networks was deficient. In the future, we plan to explore the use of transformers to enhance the segmentation performance, because transformers has achieved excellent segmentation performance in the current image segmentation task. Different data augmentation techniques are employed to increase the size of the training dataset so that it can support larger model architecture.

References

[1] *Chon A, Balachandar N, Lu P. Deep convolutional neural networks for lung cancer detection[J]. Standford University, 2017: 1-9.*

[2] *Armato III, Samuel G. et al., "Data From LIDC-IDRI." The Cancer Imaging Archive, 2015, doi: 10.7937/K9/TCIA.2015.LO9QL9SX.*

[3] *Kostis W J, Reeves A P, Yankelevitz D F, et al. Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images[J]. IEEE transactions on medical imaging, 2003, 22(10): 1259-1274.*

[4] *Dehmeshki J, Amin H, Valdivieso M, et al. Segmentation of pulmonary nodules in thoracic CT scans: a region growing approach[J]. IEEE transactions on medical imaging, 2008, 27(4): 467-480.*

[5] *Kumar D, Wong A, Clausi D A. Lung nodule classification using deep features in CT images[C]//2015 12th conference on computer and robot vision. IEEE, 2015: 133-138.*

[6] *Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.*

[7] *Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer International Publishing, 2015: 234-241.*

[8] *Çiçek Ö, Abdulkadir A, Lienkamp S S, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]//Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19. Springer International Publishing, 2016: 424-432.*

[9] *Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.*

[10] *Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11534-115 42.*

[11] *Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging[J]. arXiv preprint arXiv:1508.01991, 2015.*