Using PSPNet and UNet to analyze the internal parameter relationship and visualization of the convolutional neural network

Wei Wang

Abstract

Convolutional neural network (CNN) has achieved great success in many fields, but due to the huge number of parameters, it is very difficult to study. Then, can we start from the parameters themselves to explore the relationship between the internal parameters of CNN? This paper proposes to use the convolution layer parameters substitution with the same convolution kernel setting to explore the relationship between the internal parameters of CNN and proposes to use the CNN visualization method to check the relationship. Using the visualization method, the forward propagation process of CNN is visualized. It is an intuitive representation of how CNN learns. According to the experiments, this paper believes that 1. Residual layer parameters of ResNet are correlated, and some layers can be substituted for each other; 2. Image segmentation is a process of first learning image texture features and then locating and segmentation.

Introduction

Convolution neural network on autopilot, image segmentation, image super-resolution directions have been a huge success, but has long been CNNs are run as a black box, people also have a series of confusion: convolution kernels as the core of CNN are there a relationship between different layers of convolution kernels, how convolution neural network works or how his work process, etc.

In order to explore these problems, people put forward various deep learning interpretation methods such as:

1. Deep learning is explained from the perspective of quantum mechanics.
2. Through the visualization of CNN feature map, the change of feature map is analyzed to explain the deep learning. However, these articles all have a defect, that is, the abstract interpretation, not intuitive, and there is no validity index of the feature maps.

The architectures of CNNs varies greatly, but there are about two basic forms in the segmentation field: one is the encoder-decoder form of UNet; the other is the method of feature extraction using the ResNet network. Two representative networks, UNet and PSPNet (Because it contains the ResNet structure) were selected for this article. They were used to perform myocardial segmentation tasks to explore the relationship between internal parameters of the convolutional neural network. How the features change in the forward propagation process of CNN were given by visualization.

Visualization analysis is an important method in CNN, but the previous article visualization is to manually select a special feature map for visualization, which is not generic. In this paper, by using the method of feature map summation, each feature map is summed on channel dimension, and then the result of summation is visualized. This paper analyzes how CNN works by visualizing features change from the bottom to the top of layers.

Contributions of this paper:

1. The correlation between the convolution kernels in the ResNet layer blocks is proposed, and the proofs are given, and the evaluation standard of the correlation is given.
2. Propose a new CNN feature map visualization method.
3. It is proposed that ResNet convolution kernel parameters in the same layer blocks are correlated.

The structure of this paper is as follows:

1. Theory: Gives the mathematical form of ResNet structure and the mathematical representation after convolution kernels replacement.
2. Data: Describes the Data set used in this article.
3. Layer block convolution kernels replacement: UNet and PSPNet network convolution kernels in the same layer blocks were replaced respectively to test the correlation of network convolution kernels.
4. ResNet34-ResNet18 results: Replace part of PSPNet-ResNet18 layers with part of PSPNet-ResNet34 layers.
5. Visualization: Visualization of the UNet and PSPNet forward propagation characteristics, respectively. The influence of convolution kernels replacement is analyzed by visualization method.
6. Discussion: Analyzed the possible questions in the article?

Theory

Mathematical representation of a ResNet residual unit:

\[ x_{l+1} = \sigma \ast (x_l + BN(\sigma(BN(x_l \ast w'_l) \ast w''_l))) \]

Replacing one of the convolution kernels in residual unit with one of the convolution kernels in the layer block can be written as:

\[ x_{l+1} = \sigma \ast (x_l + BN(\sigma(BN(x_l \ast w_n) \ast w''_l))) \]
\[ x_{l+1} = \sigma \ast (x_l + BN(\sigma(BN(x_l \ast w'_l) \ast w_n))) \]

Where, \(x_l\) represents the feature map of the input of the \(l\)th residual unit, \(x_{l+1}\) represents the input at the \((l+1)\)th layer or the feature map of the output of \(l\)th residual unit. \(w'_l\) and \(w''_l\) represent the first convolution kernel and the second convolution kernel of the \(l\)th residual unit respectively, and \(\sigma\) represents Relu activation function. \(w_n\) represents the convolution kernel to be replaced. \(\ast\) represents the convolution operation. \(x'_l\) and \(x''_l\) respectively represents the output features map after parameter replacement. BN means Batch-normalization.

If \(x_l \approx x'_l\) or \(x_l \approx x''_l\), we think \(w_n\) has the same function as \(w'_l\) or \(w''_l\), and vice versa.

**Data**

The data set used in this experiment is T1\(^{1}\) cardiac segmentation data set, which contains 210 persons and corresponding cardiac segmentation labels of each person with 55 images. The first 60 people are selected as the training set and the last 30 as the validation set. The Dice baseline of UNet and PSPNet are 0.8645 and 0.8723.

**Layer block convolution kernels replacement**

In this paper, UNet and PSPNet networks were used to segment T1 cardiac image. The structure of UNet network is as follows:

![UNet Network Structure](image)

Figure 1: UNet network structure: layer1-5 represents the layer block 1-5, feature1-10 represents the output feature maps that need to be visualized, 0,1,2 represents the 0th,1st,2nd convolution in the layer block respectively.

Replace the convolution kernel of the convolution layer in the same layer block, as follows: Layer1 [0]=Layer1[1]. The results are then tested with other parameters and validation sets unchanged. The results are as follows:

| Layer block 1 | Layer block 2 | Layer block 3 |
|---------------|---------------|---------------|
| i | j | Dice | i | j | Dice | i | j | Dice |
| 0 | 1 | 0.04 | 0 | 1 | 0.04 | 0 | 1 | 0.00 |
| 0 | 2 | 0.15 | 0 | 2 | 0.05 | 0 | 2 | 0.00 |
| 1 | 0 | 0.28 | 1 | 0 | 0.06 | 1 | 0 | 0.00 |
| 1 | 2 | 0.01 | 1 | 2 | 0.02 | 1 | 2 | 0.11 |
| 2 | 0 | 0.28 | 2 | 0 | 0.00 | 2 | 0 | 0.00 |
| 2 | 1 | 0.20 | 2 | 1 | 0.00 | 2 | 1 | 0.00 |

| Layer block 4 | Layer block 5 |
|---------------|---------------|
| i | j | Dice | i | j | Dice |
| 0 | 1 | 0.01 | 0 | 1 | 0.00 |
| 0 | 2 | 0.31 | 0 | 2 | 0.00 |
| 1 | 0 | 0.00 | 1 | 0 | 0.00 |
| 1 | 2 | 0.03 | 1 | 2 | 0.00 |
| 2 | 0 | 0.01 | 2 | 0 | 0.00 |
| 2 | 1 | 0.00 | 2 | 1 | 0.02 |

Table 1: Replacement results of UNet convolution kernel: Layer block 1-5 represents Layer1-5 of the network, i and j represent the original layer and the target layer respectively, corresponding to W of 0-2 convolutional layer in Layer 1-5. Such as Layer1 \([i] = Layer1 \([j]\).
Firstly, parameters replacement was carried out for the residual units of ResNet module in PSPNet. That is to replace the convolution layers with the same convolution kernel settings in ResNet. Under the condition that other parameters and validation set are unchanged, the result was tested directly on the validation set.

The parameters were replaced layer by layer. \(i\) represents the replaced layer and \(j\) represents the original layer, namely \(\text{Layer1}[i]=\text{Layer1}[j]\).

The experimental results are as follows:

Let’s replace the convolution kernels in each layer block with the same one. \(j\) represents the original layer. Such as \(\text{Layer1}[1,2n]=\text{Layer1}[j]\).

\[\text{RMSE:RMSE(F1,F1'),RMSE(F2,F2'),RMSE(F3,F3'),RMSE(F4,F4')}\]

\(F1,F1', F2,F2', F3,F3', F4,F4'\) need to be summed in the channel dimension, \(F1,F1', F2,F2', F3,F3', F4'\) correspond to the original feature map of Feature1-4 and the feature map after the convolution kernels replacement.

Table 3: The result of Layer1-4 parameters are all replaced by one of the layer in Layer1-4 separately.

| Layer block | \(j\) | Dice | RMSE  |
|-------------|-------|------|-------|
| Layer block 1 | 0.00 | 0.83 | 2.852 |
| Layer block 2 | 0.00 | 0.86 | 5.918 |
| Layer block 3 | 0.00 | 0.86 | 6.469 |
| Layer block 4 | 0.00 | 0.86 | 6.576 |

It can be seen from the above results that the convolution kernels in the UNet network layer blocks have no obvious correlation. The convolution kernels in the ResNet layer blocks are correlated. When only one layer of convolution kernels is replaced, RMSE and Dice change very little after the convolution kernels were replaced except Layer4. When all convolution kernels were replaced, the Layer2 result still changed little. Therefore, this paper believes that residual convolution kernels in the same ResNet layer block have the same function or different residual convolution kernels in the layer block acts on the feature map and the results are almost the same. With the deepening of the layers, the difference of the results of the residual convolution kernels acting on the feature maps increases gradually.

### ResNet34-ResNet18 results

The ResNet34 convolution kernels in PSPNet-ResNet34 was used to replace the convolution kernels of the corresponding layers in PSPNet-ResNet18, and then the results were tested directly on the validation set, as shown below:

Table 4: The results of PSPNet-ResNet 34 replacing the convolution kernels of PSPNet-ResNet18.

| \(j\) | Dice | RMSE  |
|------|------|-------|
| Layer1 | 0.00 | 0.83 |
| Layer2 | 0.00 | 0.86 |
| Layer3 | 0.00 | 0.86 |
| Layer4 | 0.00 | 0.86 |

The corresponding layer is as follows:

\('layer1.0.conv1.weight','layer1.0.conv2.weight', 'layer1.1.conv1.weight', 'layer1.1.conv2.weight'\)
Table 2: Left to right respectively represent the results after the convolution kernels of each layer inside Layer1-4 are replaced. i represents the original layer, j represents the replaced layer; RMSE from left to right: RMSE(F1,F1'), RMSE(F2,F2'), RMSE(F3,F3'), RMSE(F4,F4'). F1,F2,F3,F4,F4' need to be summed in the channel dimension, F1,F2,F3,F4,F4' correspond to the original feature map of Feature1-4 and the feature map after the convolution kernels replacement. Dice represents the result on the validation set after the convolution kernels replacement. Dice results after the convolution kernels of each layer in Layer Leyer 3:

| Layer Block 3 | RMSSE | Dice | Layer Block 4 | Leyer4.0.conv0.weight | Layer4.0.conv0.weight | Layer4.0.conv0.weight |
|---------------|-------|------|---------------|----------------------|----------------------|----------------------|
| 0             | 2     | 1.701| 1             | 0.836                | 0.836                | 0.836                |
| 1             | 2.250| 1.676| 2             | 0.845                | 0.845                | 0.845                |
| 2             | 2.801| 0.837| 3             | 0.837                | 0.837                | 0.837                |
| 3             | 3.367| 0.845| 4             | 0.845                | 0.845                | 0.845                |
| 4             | 3.923| 0.854| 5             | 0.854                | 0.854                | 0.854                |
| 5             | 4.480| 0.863| 6             | 0.863                | 0.863                | 0.863                |
| 6             | 5.046| 0.872| 7             | 0.872                | 0.872                | 0.872                |
| 7             | 5.610| 0.881| 8             | 0.881                | 0.881                | 0.881                |
| 8             | 6.177| 0.890| 9             | 0.890                | 0.890                | 0.890                |
| 9             | 6.743| 0.900| 10            | 0.900                | 0.900                | 0.900                |

The visualization of all feature map requires summing in the channel dimension firstly.

**Visualization**

The visualization of all feature map requires summing in the channel dimension firstly.

**UNet**

As can be seen from the above figure, UNet encoder is mainly responsible for learning pattern features of images, while decoder is responsible for segmentation.
PSPNet
As can be seen from the above figure, in the first three figures, we mainly learn the texture features of the original image, then locate it in the fourth figure and segment it in the fifth figure. Therefore, image segmentation is a process of first texturizing the feature, learning image texture, locating, and then realizing segmentation based on localization.

Layer4 results of feature maps visualization after ResNet34 replaced ResNet18 convolution kernel:

It can be seen from the above feature map that, except for Layer4, the residual convolution kernels in the ResNet layer block of different layers are replaceable, while the fourth layers block are replaceable in the identity mapping layer. Combined with the result of Layer block convolution kernels replacement of PSPNet, it shows that the identity mapping layers in the PSPNet network are mainly responsible for saving the complete image information, while the front part of the residual layers is responsible for learning the image features. In combination with UNet feature map changes, the more complete the image information is, the better the segmentation result will be. In PSPNet, the first three layer blocks mainly learn the image features, while the fourth layer block begins the segmentation task, so the integrity of the information should be maintained in the front. The fourth layer block begins to segment, and the segmentation is mainly aimed at the segmentation position, so the complete information of the image is no longer important, as long as the segmentation position information is sufficient. In front of PSPNet, it is mainly the learning of image information, which needs the integrity of information, so the identity mapping layers cannot be replaced, and the fourth layer block starts to learn segmentation, so the residual layer cannot be replaced.

Discussion
In section Layer block convolution kernels replacement, PSPNet Dice changes are basically consistent with RMSE changes, but some results are inconsistent with RMSE changes. Visualize the feature map for the fact that these RMSEs in the results do not fully and accurately reflect the results. The reason was found in the red box part in the figure above, because the purpose of the network was to segment the myocardium. However, although RMSE was low in some feature map, the extraction of partial feature information of myocardium was insufficient, so the Dice value decreased as a result. Although RMSE value is high in some characteristic maps, the location and information extraction of myocardium are sufficient, so the Dice value is high.

Conclusion
1. The residual convolution kernels in the network layer blocks of ResNet are correlated, that is, the convolution kernels can be replaced with each other. In this paper, there are two reasons: one is that the residual layers in the same
layer block perform the same or similar task to some ex-
tent; Second, CNN is fault-tolerant. 2. UNet, PSPNet image
segmentation networks are a process of first learning texture
features, then positioning, and finally segmentation. With-
out loss of generality, this paper believes that CNNs are a
process of first analyzing texture features and then locating
segmentation for segmentation tasks. 3. With the deepening
of layers, the convolution kernels replaceable capability in
layer blocks become worse and worse, so this paper believes
that with the deepening of layers, the task becomes more tar-
geted, which can also be seen from the visualization of the
feature map.

References
[1] Fahmy, A. S.; El-Rewaidy, H.; Nezafat, M.; Nakamori,
S.; and Nezafat, R. 2019. Automated analysis of cardio-
vascular magnetic resonance myocardial native T1 map-
ping images using fully convolutional neural networks.
Journal of Cardiovascular Magnetic Resonance 21(1).
[2] He, K.; Zhang, X.; Ren, S.; and Sun, J. 2015. Deep
Residual Learning for Image Recognition. arXiv e-
prints arXiv:1512.03385.
[3] Levine, Y.; Yakira, D.; Cohen, N.; and Shashua, A.
2017. Deep Learning and Quantum Entanglement: Funda-
mental Connections with Implications to Network
Design. arXiv e-prints arXiv:1704.01552.
[4] Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-Net:
Convolutional Networks for Biomedical Image Seg-
mentation. arXiv e-prints arXiv:1505.04597.
[5] Zeiler, M. D.; and Fergus, R. 2013. Visualizing and
Understanding Convolutional Networks. arXiv e-prints
arXiv:1311.2901.
[6] Zhao, H.; Shi, J.; Qi, X.; Wang, X.; and Jia, J.
2016. Pyramid Scene Parsing Network. arXiv e-prints
arXiv:1612.01105.