

# 基于模型重建的深度卷积网络权值可视化方法

刘嘉铭\* 邢孟道 符吉祥 徐丹

(西安电子科技大学雷达信号处理国家重点实验室 西安 710071)

**摘要:** 针对深度卷积网络原理分析的问题, 该文提出一种基于模型重建的权值可视化方法。首先利用原有的神经网络对测试样本进行前向传播, 以获取重建模型所需要的先验信息; 然后对原本网络中的部分结构进行修改, 使其便于后续的参数计算; 再利用正交向量组, 逐一地计算重建模型的参数; 最后将计算所得的参数按照特定的顺序进行重排列, 实现权值的可视化。实验结果表明, 对于满足一定条件的深度卷积网络, 利用该文所提方法重建的模型在分类过程的前向传播运算中与原模型完全等效, 并且可以明显观察到重建后模型的权值所具有的特征, 从而分析神经网络实现图像分类的原理。

**关键词:** 深度卷积网络; 可视化; 模型重建

中图分类号: TN957.51

文献标识码: A

文章编号: 1009-5896(2019)09-2194-07

DOI: 10.11999/JEIT180916

## A Method to Visualize Deep Convolutional Networks Based on Model Reconstruction

LIU Jiaming XING Mengdao FU Jixiang XU Dan

(National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China)

**Abstract:** A method for visualizing the weights of a reconstructed model is proposed to analyze a deep convolutional network works. Firstly, a specific input is used in the original neural network during the forward propagation to get the prior information for model reconstruction. Then some of the structure of the original network is changed for further parameter calculation. After that, the parameters of the reconstructed model are calculated with a group of orthogonal vectors. Finally, the parameters are put into a special order to make them visualized. Experimental results show that the model reconstructed with the proposed method is totally equivalent to the original model during the forward propagation in the classification process. The feature of the weights of the reconstructed model can be observed clearly and the principle of the neural network can be analyzed with the feature.

**Key words:** Deep convolutional networks; Visualization; Model reconstruction

### 1 引言

在2012年ILSVRC(Imagenet Large Scale Visual Recognition Challenge)比赛中, 以卷积神经网络(Convolutional Neural Networks, CNN)为模型的Alexnet<sup>[1]</sup>, 对于测试样本, 分类器输出最有可能的5个分类结果, 而测试样本不是其中任何一类的概率只有15.3%, 远超过第2名的26.2%, 取得了比赛冠军。从那以后, 深度学习的研究迎来了又一个高潮。先后有VGGNet<sup>[2]</sup>, Googlenet<sup>[3]</sup>, ResNet<sup>[4]</sup>

等网络结构相继提出。

一般来说科学研究是一个理论向实践转化的过程, 可深度学习却是一个例外。深度学习的实践发展非常迅速, 恰恰相反, 理论研究却显得有些不足<sup>[5]</sup>, 神经网络的隐藏层就好比一个黑箱<sup>[6]</sup>, 虽然效果很好但原理却非常不清楚。到目前为止, CNN的理论研究主要局限于理解CNN中不同层发生的情况, 即特征可视化。具有代表性的特征可视化方法主要包括Zeiler等人<sup>[7]</sup>提出的反卷积技术以及Zhou等人<sup>[8]</sup>提出的类别激活映射方法。关于权值的可视化最早是在Alexnet中直接可视化了第1个卷积层的卷积核, 由于卷积核的尺寸是 $11 \times 11$ , 所以还能看出一定的纹理信息, 但是对于结构复杂的卷积网络中 $5 \times 5$ ,  $3 \times 3$ 甚至是 $1 \times 1$ 的卷积核<sup>[9]</sup>, 由于尺寸过小, 分析会很困难。深度卷积网络模型结构复

收稿日期: 2018-09-21; 改回日期: 2019-02-19; 网络出版: 2019-03-21

\*通信作者: 刘嘉铭 liujiaming@stu.xidian.edu.cn

基金项目: 国防科技卓越青年人才基金(2017-JCJQ-ZQ-061)

Foundation Item: The National Defense Science and Technology Excellent Youth Talent Foundation of China (2017-JCJQ-ZQ-061)

杂，如卷积结构具有“稀疏连接”、“权值共享”的特性，相比于全连接结构降低了参数规模；为了缓解过拟合的问题，引入了dropout结构等，这些网络结构的调整目的是为了降低训练难度，提高准确率，但如果分析的问题不是神经网络的训练过程，而是网络的训练结果即网络实现分类预测的原理，那么复杂的训练模型会加大分析的难度。因此本文提出一种将原本复杂的网络训练模型重建为简单的传递模型的算法，新模型在网络前向传播的过程中与原始模型完全等效，并借助重建后的模型分析神经网络实现图像分类的原理，为未来的模型调整与参数优化提供理论基础。本文中出现的正向传播都是指利用训练好的网络模型进行分类测试的正向传播过程，不考虑网络的训练过程。

## 2 卷积神经网络的传递模型

CNN是一种多隐藏层的网络结构，主要包括卷积层、池化层、全连接层和分类层，除此以外，还有激活函数以及正则化等结构。网络结构表示为

$$\mathbf{y} = f_{\mathbf{x}}(\mathbf{x}) \quad (1)$$

其中， $\mathbf{x}$ 表示卷积神经网络的输入，长度为单个测试样本的像素数； $\mathbf{y}$ 表示网络的输出，长度为网络的分类数； $f_{\mathbf{x}}(\cdot)$ 表示神经网络对测试样本的变换，由神经网络训练好的权值和测试样本共同决定。

针对式(1)，本文提出如下假设：在神经网络进行前向传播的过程中，如果网络内部任意结构(如卷积、池化、激活函数等)的每个输出可以用式(2)表示，则式(1)表示的整个网络的模型结构可以表示成式(3)的形式

$$h(\mathbf{x}_0) = \mathbf{w}_0^T \mathbf{x}_0 + b_0 \quad (2)$$

$$\mathbf{y} = f_{\mathbf{x}}(\mathbf{x}) = \mathbf{w}_{\mathbf{x}}^T \mathbf{x} + \mathbf{b}_{\mathbf{x}} \quad (3)$$

式(2)中 $h(\mathbf{x}_0)$ 是标量函数，表示神经网络中某一结构的输出； $\mathbf{x}_0$ 表示该结构的输入矢量； $\mathbf{w}_0$ 分两种情况，既可以是不变的常数权矢量，也可以是通过用测试样本作为网络输入获取前向传播的先验信息后，利用该先验信息来确定具体数值的权矢量，后面会详细讨论； $b_0$ 是标量，表示偏置。

将式(3)称为卷积神经网络的传递模型，其中 $\mathbf{w}_{\mathbf{x}}$ 是模型重建后的权矢量， $\mathbf{b}_{\mathbf{x}}$ 是偏置矢量，这2个矢量都是由原模型训练好的权值和输入的测试样本共同决定。

深度卷积神经网络内部参数较多，同时卷积层还存在稀疏连接、权值共享的情况，导致内部权值的连接十分复杂，如果直接分析网络内部权值的连接来进行模型重建势必会非常困难，因此本文提出一种利用正交向量组计算模型参数的方法。

### 2.1 正交向量组求解系统响应

对于一个线性系统，假定其响应函数如式(4)所示

$$\mathbf{y} = \mathbf{w}^T \mathbf{x} + \mathbf{b} \quad (4)$$

其中， $\mathbf{w}$ 、 $\mathbf{b}$ 均为常数矢量。式(4)的具体形式为

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nm} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \quad (5)$$

对于式(5)所示的响应函数，可以考虑利用正交向量组求解系统响应。正交向量组如图1所示。

$$\underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}}_m$$

图1 正交向量组

首先将长度为 $m$ 的全0向量作为系统输入，计算0输入响应 $\mathbf{b}$ 。之后再图1中的第1列向量作为输入 $\mathbf{x}$ 代入式(5)中，通过输出矢量 $\mathbf{y}$ 和预先计算好的 $\mathbf{b}$ ，根据 $w_{i1} = y_i - b_i (i = 1, 2, \dots, n)$ 求得 $\mathbf{w}^T$ 第1列的值；然后将图1中的第2列向量作为输入，可以求得 $\mathbf{w}^T$ 第2列的值；依此类推，通过重复该步骤 $m$ 次，可以求得 $\mathbf{w}$ 的全部值。

需要注意的是，由于神经网络并不是简单的线性系统，计算0输入响应比较困难(某些正则化结构会引入额外的0输入响应)，而且本文主要是对模型重建后的权值进行可视化分析，重点在 $\mathbf{w}$ 上，所以在后面模型重建的理论分析与实验中不考虑 $\mathbf{b}$ 的影响，所有的偏置直接置0处理。

### 2.2 网络结构调整

由于式(3)中 $\mathbf{w}_{\mathbf{x}}$ 会随着神经网络的输入矢量 $\mathbf{x}$ 而改变，所以不能直接用正交向量组进行求解。但是在神经网络的一次前向传播计算中，输入是唯一确定的， $\mathbf{w}_{\mathbf{x}}$ 也不会改变。因此可以考虑以某一固定输入作为先验条件的情况下，通过调整网络的内部结构，使得 $\mathbf{w}_{\mathbf{x}}$ 变为固定值，之后便可以使用正交向量组求解该输入条件下的 $\mathbf{w}_{\mathbf{x}}$ 。

“固定” $\mathbf{w}_{\mathbf{x}}$ 就是要确定式(2)中每个结构的 $\mathbf{w}_0$ ，实际上式(2)既是整个网络模型是否能用式(3)来表示的判定条件，也是网络内部每个结构调整后的结

果。对于具体的不同结构,式(2)也有着不同的具体形式。

### 2.2.1 卷积层、平均池化层和全连接层的调整

由于卷积、平均池化和全连接在不考虑偏置的情况下都是线性运算,对应式(2)中的 $w_0$ 仅由神经网络的训练结果决定,而与网络前向传播的测试样本无关,所以卷积、平均池化和全连接这3种结构不需要修改。实际上,不仅仅上述的3种结构,神经网络中所有线性运算结构(如dropout<sup>[10]</sup>)都可以直接保留。

### 2.2.2 最大池化层的调整

与平均池化不同,最大池化是非线性运算,所以需要调整。最大池化原理如图2所示。

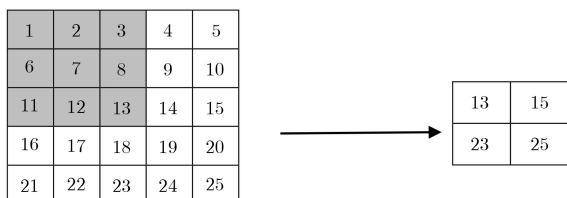


图2 最大池化

假定最大池化层的输入矢量 $x$ 的最大值仅出现1次,那么运算过程可以表示为

$$y = \begin{bmatrix} 1_{[x_1=\max(x_1,x_2,\dots,x_n)]} \\ 1_{[x_2=\max(x_1,x_2,\dots,x_n)]} \\ \vdots \\ 1_{[x_n=\max(x_1,x_2,\dots,x_n)]} \end{bmatrix}^T \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad (6)$$

其中, $1_{[\cdot]}$ 是逻辑函数,当 $[\cdot]$ 内的表达式为“真”,则函数值为1,否则为0。相应地,式(2)中的 $w_0$ 可以表示为

$$w_0 = \begin{bmatrix} 1_{[x_1=\max(x_1,x_2,\dots,x_n)]} \\ 1_{[x_2=\max(x_1,x_2,\dots,x_n)]} \\ \vdots \\ 1_{[x_n=\max(x_1,x_2,\dots,x_n)]} \end{bmatrix} \quad (7)$$

利用测试样本传递至最大池化层处的输入特征图 $x$ ,计算 $w_0$ 的值,利用该 $w_0$ 将最大池化层调整成式(2)的形式,这样在后续利用正交向量组计算模型参数的过程中,式(2)的 $w_0$ 将变为固定值,不会随该层的输入 $x_0$ 而改变。

以上只是理论分析,实际上最大池化层的模型调整还有更简便的方法。最大池化层在前向传播中通常会记录该层的输出在输入特征图中的具体位置,该位置信息会运用于后续的反向传播计算中。因此,在用正交向量组计算模型参数时,当输入向量以特征图的形式传递至最大池化层时,利用先验

位置信息,直接将输入特征图上对应位置的值输出即可。

### 2.2.3 激活函数的调整

在常用的激活函数中仅有ReLU<sup>[11]</sup>函数(包括LReLU, PReLU以及RReLU)能够写成式(2)的形式,尽管这对于本文所提模型重建算法的实际应用是一个不小的限制,但是深度网络通常都会利用ReLU作为激活函数来缓解“梯度消失<sup>[12]</sup>”的问题,所以本文所提算法在以ReLU作为激活函数的网络原理分析中仍有一定的意义。

ReLU函数的表达式为

$$y = \max(0, x) \quad (8)$$

改写成式(2)的形式

$$y = \max(0, x) = 1_{[\max(0,x)>0]}x \quad (9)$$

对应式(2)中的 $w_0 = 1_{[\max(0,x)>0]}$ (由于激活函数的输入为标量,所以此处的 $w_0$ 为标量),在用正交向量组计算模型参数的过程中,利用先验信息计算 $w_0$ 的值,将式(8)改写成式(2)的形式。

更一般地,对于ReLU函数的变形(如LReLU等),表达式为

$$y = \max(ax, x) \quad (10)$$

其中, $0 < a < 1$ ,改写成式(2)的形式

$$y = \max(ax, x) = a^{1_{[\max(ax,x)>x]}}x \quad (11)$$

对应式(2)中的 $w_0 = a^{1_{[\max(ax,x)>x]}}$ ,调整方式与ReLU相同。

在网络最终的分层中通常会采用softmax激活函数计算分类值,对于分类数为 $k$ 的网络,类别 $i$ 激活值的具体表达式为

$$y_i = \frac{\exp(x_i)}{\sum_{m=1}^k \exp(x_m)} \quad (12)$$

式(12)具有如下性质:如果类别 $i$ 的激活值 $y_i = \max(y_1, y_2, \dots, y_k)$ ,那么相对应的输入 $x_i = \max(x_1, x_2, \dots, x_k)$ 。由此可见,最终的分类结果不会被softmax激活函数改变,softmax更多是运用于训练过程中分类损失的计算,由于本文是对已经训练好的模型进行分析,所以直接将softmax分类层的输入作为原网络分类的输出。

### 2.2.4 正则化结构调整

随着神经网络研究的不断深入,除了以上的基本结构外,有些神经网络还包括一些辅助训练的结构,主要目的是加快网络训练速度、降低调参难度以及缓解“过拟合”等。这些结构被称为正则化结构,比较具有代表性的是批量归一化<sup>[13]</sup>(Batch

Normalization, BN), BN的原理本文不做详细介绍, 主要过程为

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i \quad (13)$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad (14)$$

$$E[x] = E_B[\mu_B] \quad (15)$$

$$\text{Var}[x] = \frac{m}{m-1} E_B[\sigma_B^2] \quad (16)$$

$$y = \frac{\gamma}{\sqrt{\text{Var}[x] + \varepsilon}} \cdot x + \left( \beta - \frac{\gamma E[x]}{\sqrt{\text{Var}[x] + \varepsilon}} \right) \quad (17)$$

BN结构的提出者Sergey指出BN在神经网络分

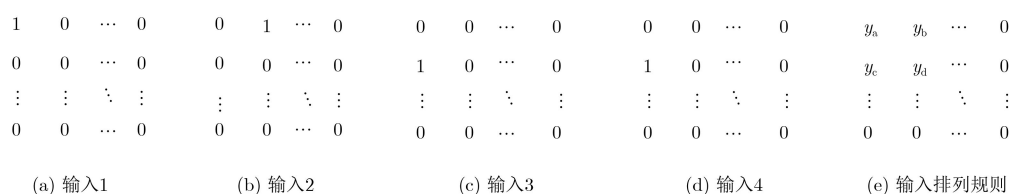


图3 权值的可视化排列

将图1中的每个列向量重新排列为类似图3(a)的数组, 分别输入到调整后的模型中计算输出, 当所有的输出y计算完毕后, 按照分类的不同分别进行排列, 可以得到n个与输入完全等维的数组, n的值取决于网络的输出类别数。最终排列好的数组本质上是原始神经网络每一个分类的输出相对于输入(用于获取先验信息的测试样本)的响应, 实现了权值的可视化。

基于模型重构的权值可视化方法步骤如下:

- (1) 判断神经网络模型的内部结构是否满足式(2)的模型重构条件;
- (2) 对于可重建模的神经网络, 利用测试样本进行前向传播, 获取先验信息;
- (3) 根据先验信息, 按照不同的结构计算w<sub>0</sub>, 将这些结构调整式为式(2)的形式;
- (4) 利用正交向量组计算模型调整后的输出y;
- (5) 输出计算完毕后按照图3的方式进行排列, 完成重建后模型的权值可视化。

### 3 实验结果与分析

#### 3.1 样本及神经网络训练模型介绍

训练以及测试样本是雅克-42和安-26飞机(图4)的雷达回波数据采用常规的距离-多普勒(R-D)成像算法生成的ISAR图像<sup>[14]</sup>以及这两类飞机的少量光

学预测的前向传播过程中是线性运算, 所以在计算重建后的模型参数的过程中, BN结构不需要修改, 不过需要注意的是式(17)中β -  $\frac{\gamma E[x]}{\sqrt{\text{Var}[x] + \varepsilon}}$ 为偏置项, 在参数计算中需要置0。

#### 2.3 权值的可视化排列

前面介绍了正交向量组计算模型参数的方法, 但实际上卷积神经网络的输入并不是1维向量, 而是2维或3维数组。将图3(a), 3(b), 3(c), 3(d)4个数组分别输入到调整后的网络模型中, 可以分别得到4个列向量, 每个列向量的长度均等于网络输出的类别数, 针对同一个分类输出(即不同列向量的相同位置的元素)的y<sub>a</sub>, y<sub>b</sub>, y<sub>c</sub>, y<sub>d</sub>4个值, 将它们按照特定的规则(如图3(e)所示, 输入中“1”的位置就是对应输出y所在的位置)排列。

学图像。雷达载频为5.5 GHz, 带宽400 MHz, 分辨率0.375 m。训练集两类飞机图片共计327张, 其中雅克-42飞机有161张, 安-26飞机166张。卷积神经网络采用去掉LRN以及dropout, 输出类别为2类的Alexnet模型, 并且直接用softmax分类器的输入作为网络的输出。样本的尺寸为227×227×1, 训练与测试的图像进行去均值处理, 训练过程每32个样本进行一次迭代, 学习率为0.01, 训练过程的学习曲线如图5所示。

#### 3.2 原始模型与传递模型分类输出比较

以10个不同的测试样本采用4000次迭代后原模型的权值分别进行重建实验, 比较原模型与传递模型分类输出结果。

表1中前5列是安-26飞机图像的输出结果, 后5列是雅克-42飞机图像的输出结果。可以看出, 在



(a) 雅克-42 (b) 安-26

图4 训练样本

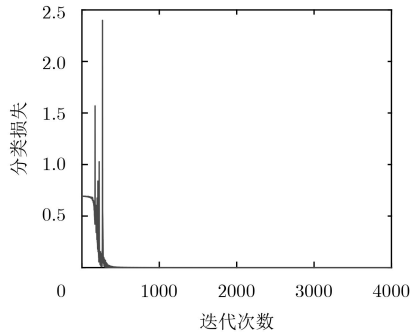


图5 学习曲线

保留2位小数的情况下，两种模型重建前后的输出完全相同。实际上，如果保留4位小数，第6列后的值为10.6519, -10.8071, 10.6517, -10.8070，重建后的传递模型与原模型输出有极小的差异。原因在于模型的训练与测试均利用了GPU加速，程序中调用了cuda架构的库函数cublasSgemm，该函数的功能是计算矩阵相乘，虽然速度快但精度不算高。实验结果表明，在不考虑计算误差的情况下，对于每一个样本，传递模型在分类预测的前向传播中与原模型是完全等效的。

### 3.3 传递模型的权值可视化

图6与图7是用4000次迭代训练的原模型权值进行重建的可视化结果。图6(a)是测试样本安-26的ISAR图像，图6(b)是输出类别为雅克-42对应的权值可视化结果，图6(c)是输出类别为安-26对应的权值可视化结果。从中可以看出两个分类输出对应于测试样本不同区域的激活与抑制的结果，其中深色表示抑制，浅色表示激活。图7是测试样本雅克-42的权值可视化结果。

图8与图9是用4000次迭代训练的原模型采用反卷积算法的可视化结果。图8(a), 图8(b), 图8(c), 图8(d), 图8(e)依次为Alexnet模型从输入层开始，以安-26为测试样本的5个卷积层特征图的反卷积结果。图9是雅克-42反卷积的结果。

以往采用反卷积的权值可视化方法只能分析神经网络提取了输入图像的特征，但是对分类结果并不敏感，本文提出的可视化方法可以区分不同分类的输出对输入图像响应，并且该响应既包括对输入图像的激活，也包括对输入图像的抑制。更为重要的是，以往的采用特征重构<sup>[15]</sup>的可视化方法最终

表1 两种模型分类输出比较

输出结果	安-26飞机图像					雅克-42飞机图像				
	原模型	-13.45	-9.22	-8.14	-6.75	-5.63	10.65	11.01	5.93	9.12
重建后模型	13.42	9.30	8.23	6.83	5.69	-10.81	-11.15	-6.01	-9.22	-7.38

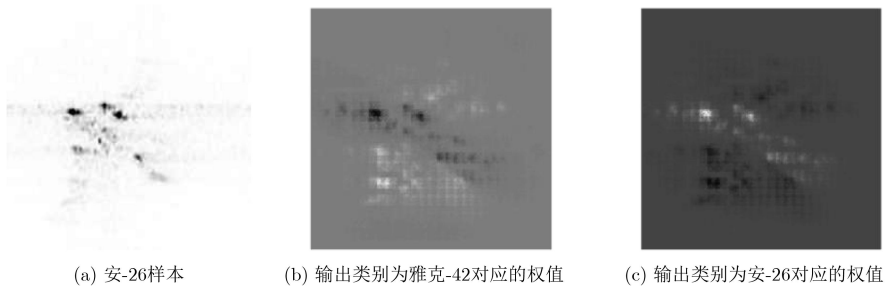


图6 安-26作为样本输入的权值可视化

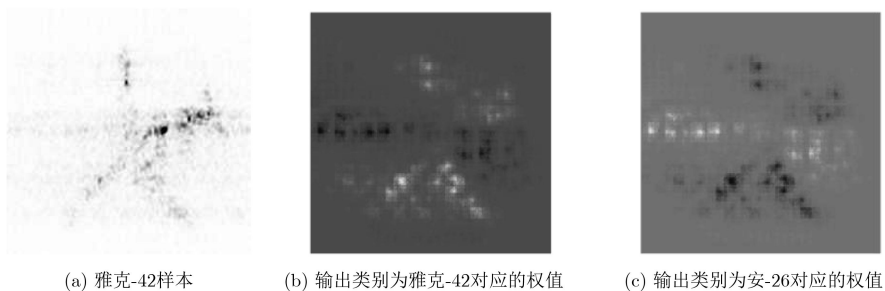


图7 雅克-42作为样本输入的权值可视化

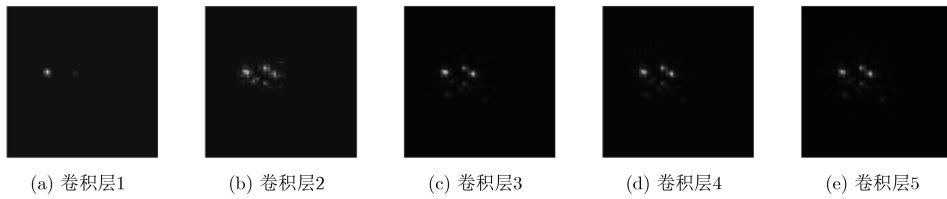


图 8 安-26为样本的反卷积结果

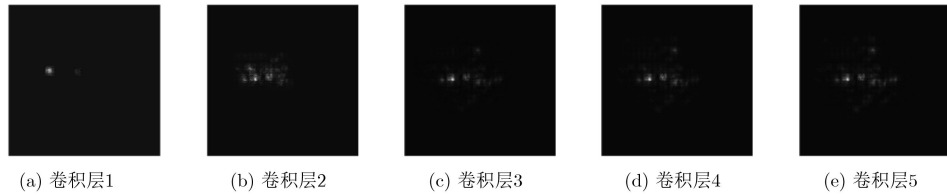


图 9 雅克-42为样本的反卷积结果

只能定性地分析原图中哪些部分被激活，而本文模型重建后的权值是具体的数值，可以定量地计算原图激活与抑制的程度，这是以往的算法所无法实现的。

此外，本文所提模型重建算法不仅可以分析网络输出对原始输入图像的响应，也可以分析输出相对于网络内部隐藏层特征图的响应。具体方法为将原网络待分析的隐藏层之前的结构全部移除，将该隐藏层的特征图作为网络的测试样本输入，以获取剩余结构重建所需的先验信息，然后采用前面提到的正交向量组计算模型参数，实现对任意隐藏层特

征图的响应进行分析。由此可见，本文所提算法具有较高的灵活性。

图10(a)，图10(b)，图10(c)，图10(d)，图10(e)分别是原训练模型迭代100, 200, 300, 400, 500次的模型重建结果。从图10(a)中可以看出在训练刚开始时重建的传递模型权值没有明显的特征，随着训练的迭代次数不断增多，图10(b)的权值已经有了明显的变化。从图5的学习曲线中可以看出，当训练达到300次左右时，分类损失已经是一个较低值，图10(c)往后传递模型的权值变化较小。图11是训练模型相邻100次迭代的进行模型重建后权值

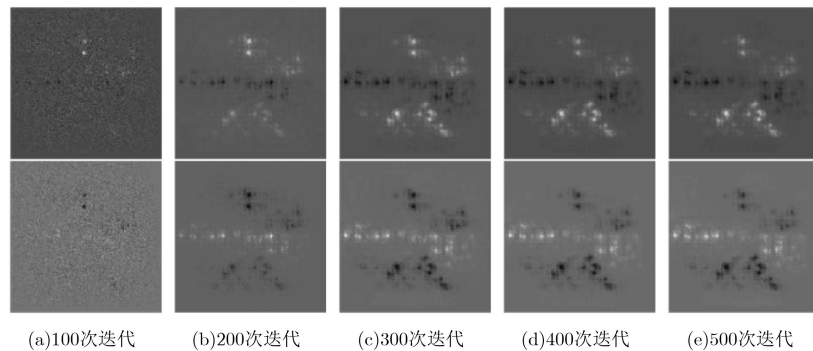


图 10 间隔100次迭代的重建模型的权值可视化

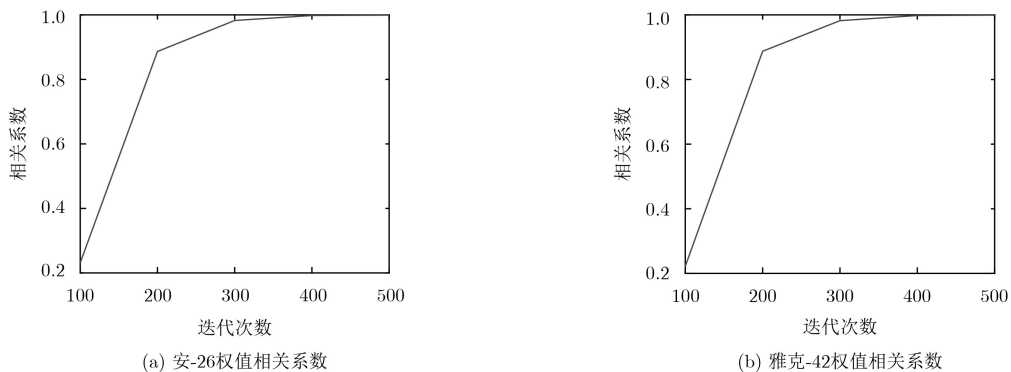


图 11 相邻100次迭代后重建模型权值的相关系数

的相关系数, 300次迭代后, 权值的相关系数趋近于1, 模型的权值变化平稳。

#### 4 结束语

与特征重构的思路不同, 本文从权值重构的角度出发, 提出了一种基于模型重建的权值可视化方法。只要满足一定的条件, 就可以对任意结构的网络进行模型重建, 不受网络深度等其他因素的影响。通过对重建后模型的权值进行可视化, 得到网络输出相对于输入的响应, 该响应表示图像中不同区域的激活与抑制, 神经网络由此实现了对图像进行分类的功能。并且本文所提算法可以定量地计算测试样本激活与抑制的程度, 在这一点上要优于传统的特征重构方法, 为以后的模型调整和参数优化提供理论基础。

#### 参考文献

- [1] FARAHBAKHS E, KOZEGAR E, and SORYANI M. Improving Persian digit recognition by combining data augmentation and AlexNet[C]. Iranian Conference on Machine Vision and Image Processing, Isfahan, Iran, 2017: 265–270.
  - [2] HOU Saihui, LIU Xu, and WANG Zilei. DualNet: Learn complementary features for image recognition[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 502–510.
  - [3] SZEGEDY C, LIU Wei, JIA Yangqing, *et al.* Going deeper with convolutions[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 1–9.
  - [4] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778.
  - [5] 王俊, 郑彤, 雷鹏, 等. 深度学习在雷达中的研究综述[J]. 雷达学报, 2018, 7(4): 395–411. doi: [10.12000/JR18040](https://doi.org/10.12000/JR18040).  
WANG Jun, ZHENG Tong, LEI Peng, *et al.* Study on deep learning in radar[J]. *Journal of Radars*, 2018, 7(4): 395–411. doi: [10.12000/JR18040](https://doi.org/10.12000/JR18040).
  - [6] PUNJABI A and KATSAGGELOS A K. Visualization of feature evolution during convolutional neural network training[C]. The 25th European Signal Processing Conference, Kos, Greece, 2017: 311–315.
  - [7] ZEILER M D and FERGUS R. Visualizing and understanding convolutional networks[C]. The 13th European Conference on Computer Vision, Zurich, Switzerland, 2014: 818–833.
  - [8] ZHOU Bolei, KHOSLA A, LAPEDRIZA A, *et al.* Learning deep features for discriminative localization[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 2921–2929.
  - [9] SUZUKI S and SHOUNO H. A study on visual interpretation of network in network[C]. 2017 International Joint Conference on Neural Networks, Anchorage, USA, 2017: 903–910.
  - [10] GAL Y and GHAMRANI Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning[C]. The 33rd International Conference on Machine Learning, New York, USA, 2016: 1050–1059.
  - [11] NAIR V and HINTON G E. Rectified linear units improve restricted Boltzmann machines[C]. The 27th International Conference on International Conference on Machine Learning, Haifa, Israel, 2010: 807–814.
  - [12] PEHLEVAN C and CHKLOVSKII D B. A normative theory of adaptive dimensionality reduction in neural networks[C]. The 28th International Conference on Neural Information Processing Systems, Montreal, Canada, 2015: 2269–2277.
  - [13] IOFFE S and SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]. The 32nd International Conference on International Conference on Machine Learning, Lille, France, 2015: 448–456.
  - [14] 王思雨, 高鑫, 孙皓, 等. 基于卷积神经网络的高分辨率SAR图像飞机目标检测方法[J]. 雷达学报, 2017, 6(2): 195–203. doi: [10.12000/JR17009](https://doi.org/10.12000/JR17009).  
WANG Siyu, GAO Xin, SUN Hao, *et al.* An aircraft detection method based on convolutional neural networks in high-resolution SAR images[J]. *Journal of Radars*, 2017, 6(2): 195–203. doi: [10.12000/JR17009](https://doi.org/10.12000/JR17009).
  - [15] NOH H, HONG S, and HAN B. Learning deconvolution network for semantic segmentation[C]. 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1520–1528.
- 刘嘉铭: 男, 1994年生, 博士, 研究方向为目标识别。  
邢孟道: 男, 1975年生, 教授, 研究方向为SAR/ISAR成像、动目标检测等。  
符吉祥: 男, 1992年生, 博士, 研究方向为ISAR成像。  
徐丹: 女, 1992年生, 博士, 研究方向为电磁特征提取。