

深度卷积神经网络鉴别正交特征生成及其应用

杨 勃, 邵泉铭, 李文彬, 郭观七, 方 欣

(湖南理工学院信息与通信工程学院, 湖南岳阳 414000)

摘 要: 针对现有深度卷积神经网络在小样本学习时的泛化性问题, 本文提出一种鉴别正交特征生成方法. 该方法通过正则化技术对网络非负中间层特征输出的异类正交度和同类相关度进行优化, 生成具有稀疏特性的网络中间层鉴别正交特征. 为有效调节稀疏度以控制网络容量, 采用正则化系数自适应调节方式逼近预设特征稀疏度目标. 为提高特征生成计算效率, 进一步设计了随机 2 类别鉴别正交特征生成反向传播规则. 随后在数据集 MNIST 上进行了小样本手写体数字识别对比实验, 验证了本文方法的稀疏度调节特性和网络表达容量控制能力. 通过反卷积可视化, 进一步发现本文方法还具有衍生出的局部鉴别区域聚焦特性. 最后, 将鉴别正交特征生成卷积网络应用到老年痴呆症 3D 磁共振影像分析上. 实验结果表明, 本文方法用于老年痴呆症诊断, 不仅诊断效果更好, 而且利用其良好的局部聚焦性, 还成功定位了老年痴呆症与健康对照组典型差异脑区.

关键词: 深度卷积网络; 鉴别正交特征生成; 脑影像分析; 核磁共振;

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2018)10-2376-08

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2018.10.010

Deep Convolutional Neural Networks Controlled by Discriminatively Orthogonal Feature Generation and Its Application

YANG Bo, SHAO Quan-ming, LI Wen-bin, GUO Guan-qi, FANG Xin

(School of Information & Communication Engineering, Hunan Institute of Science and Technology, Yueyang, Hunan 414000, China)

Abstract: To improve the generalization of deep convolutional neural networks (CNN), we proposed a discriminatively orthogonal feature generation method. By regularizing nonnegative outputs, orthogonal degree and correlation degree were optimized simultaneously, which helps to generate discriminatively orthogonal and sparse features. To adjust sparse degree for controlling network capacity, the technique of auto-adjusting regularization coefficient was proposed. To improve computational efficiency, a stochastically 2-class discriminatively orthogonal feature generation rule was further designed. Subsequently, a comparative experiment was conducted on handwritten digit set MNIST. In this experiment, the sparsity adjustment property of our method was verified. By means of deconvolution technique for visualization, it was further found that our method has a good property of focusing on local discriminant areas. Finally, our method was applied to Alzheimer's Disease MRI image analysis. The experimental results showed that our method outperforms some other representative methods and locates the importantly discriminant brain regions successfully.

Key words: deep convolutional neural networks; discriminatively orthogonal feature generation; brain image analysis; nuclear magnetic resonance;

1 引言

近年来,随着计算能力不断提高,大样本数据集出现,以深度神经网络为代表的深度学习方法在图像、语音、自然语言处理等诸多方面取得了极大进展^[1,2].其中,深度卷积网络以其权重共享,图像局部分析等优良

特性,在诸多图像大数据集上进行的多种视觉任务^[3-6]上均取得了明显优于传统浅层机器学习方法的效果.

作为一类具有庞大参数数量的复杂机器学习方法,深度学习效果与数据集样本量密切相关^[7,8].当面向样本量较少的学习任务时,过于复杂的深度学习方法通常极易出现过拟合现象,表现甚至不如传统浅层

机器学习方法.

小样本学习是统计机器学习的一个基本研究内容,也是实际应用场景不能回避的一个重要问题.以往研究成果表明^[9,10],采用诸如数据降维、容量控制、迁移学习等策略可改善小样本学习效果.为提高泛化性,这些策略也被应用于深度学习.降维方面,采用如深度孪生卷积网络,匹配网络等深度网络模型进行度量学习^[11,12],提取低维流形特征,可改善识别泛化性能.容量控制方面,Hinton 提出 Dropout 方法^[13],通过随机置 0 扰动,防止过拟合;此外,对网络权重参数引入 L2 范数正则化项^[14],也是神经网络经典的容控手段.迁移学习方面,Yosinski^[15]、Azizpour^[16]等通过在 ImageNet, ILS-VRC 2013 等大数据集上迁移实验,验证了迁移学习对小样本视觉任务的有效性.

本文以磁共振影像分析为背景,尝试从容量控制角度,继续对小样本任务下的深度卷积神经网络进行研究.考虑到小样本下的过拟合是由网络本身表达容量和现有样本量不匹配所致,本文提出一种鉴别正交特征生成方法,强制网络指定层生成的特征满足同类强相关和异类正交的表达容量控制要求,通过控制特征稀疏度改善泛化性能.为验证鉴别正交特征生成有效性,我们在 Caffe 平台^[17]上对其进行了实现,采用 LeNet 经典卷积神经网络架构首先在 MNIST 数据集上进行了小样本实验.对比结果表明,小样本情形下鉴别

正交特征生成方法能够显著提高分类效果,具有较强的网络容量控制能力.进一步通过反卷积可视化发现,鉴别正交特征生成方法可控制生成特征的稀疏性,并具有很强的局部鉴别区域定位能力.最后,我们将鉴别正交特征生成方法应用到基于 3D 脑磁共振影像分析的老年痴呆症自动诊断中,一方面进一步提高了诊断正确率,另一方面基于反卷积重构技术实现了老年痴呆病人与健康对照组典型差异脑区的定位.

2 深度卷积神经网络

第一个提出并实际应用的卷积神经网络是用于手写数字识别的多层卷积网络 LeNet-5^[18].该网络由 2 个卷积层,2 个池化层,2 个全互联层以及 1 个分类器层组成,如图 1 所示.



图1 LeNet-5架构示意图

LeNet-5 给出了一个应用于自然图像的标准多层多尺度分析架构.受此启发,随后出现了多种卷积神经网络架构.

第一个真正取得突破的深度卷积神经网络则是 AlexNet^[3],其架构示意图如图 2 所示.



图2 AlexNet架构示意图

相比于 LeNet-5 架构,除整体架构更深之外,AlexNet 还引入了 ReLU 非线性映射层,LRN 归一化层,并使用了 Dropout 和 Data augmentation 策略.其中,ReLU 函数的引入不仅克服了经典 sigmoid 函数的饱和问题,而且因其非负特性,使特征表达在一定程度上具有稀疏性,为加速收敛和改善分类效果提供了支持.

近年来,深度卷积网络向极深层发展,出现了诸如 GoogleNet^[19],ResNet^[20]等极深层卷积神经网络架构,并在大规模数据集上取得了优异表现.

3 卷积神经网络鉴别正交特征生成

卷积网络中间层特征通常通过分类损失函数优化过程中的反向传播算法逐渐生成.目前常用的分类器和损失函数是 softmax 分类器和互熵损失函数.

对一个 D 类别分类任务,训练集为 $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{1 \leq i \leq N}$,其中 $\mathbf{x}_i, \mathbf{y}_i$ 分别为第 i 个训练样本和对应的 D 维类别标签向量,设 softmax 分类器输出为 $\{\mathbf{o}_i\}_{1 \leq i \leq N}$,此时互熵损失函数 $L1$ 定义为

$$\min L1 = - \sum_{i=1}^N \log(\mathbf{o}_{i, D(i)}) \quad (1)$$

其中, $D(i)$ 是第 i 个训练样本类别序号, $\mathbf{o}_{i, D(i)}$ 是该样本在其类别序号对应的维度上的 softmax 分类器输出.

ImageNet 大数据集下卷积网络中间层特征可视化结果表明^[21],从底层到顶层,各层特征呈现逐步抽象和稳定的现象,有效改善分类效果.该实验直观验证了卷积网络特征生成对视觉任务的重要性.

对小样本任务,卷积网络面临的一个主要困境是无法充分进行特征学习,从而导致过拟合.有文献表明^[21-23],具有表达稀疏特性的深度网络往往具有很好的容量控制能力,能改善网络的过拟合现象.为进一步约束卷积网络中间层特征生成,我们利用非负向量正交稀疏性,在互熵损失基础上,采用正则化策略设计了一个基于相关性度量的鉴别正交特征生成损失项.

3.1 非负向量的正交稀疏性

对 l 维向量 $\boldsymbol{\alpha} = (\alpha_1 \cdots \alpha_l)^T$,我们将该向量稀疏度 $sp(\boldsymbol{\alpha})$ 定义为

$$sp(\boldsymbol{\alpha}) = \frac{card(\{\alpha_i | \alpha_i = 0\}_{1 \leq i \leq l})}{l} \quad (2)$$

其中, $card(\{\alpha_i | \alpha_i = 0\}_{1 \leq i \leq l})$ 表示向量 $\boldsymbol{\alpha}$ 中 0 元素的数目.

对 l 维向量 $\boldsymbol{\alpha}, \boldsymbol{\beta}$, 其相关系数可定义为

$$\rho(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\boldsymbol{\alpha}^T \boldsymbol{\beta}}{\sqrt{\|\boldsymbol{\alpha}\|^2 \|\boldsymbol{\beta}\|^2}} \quad (3)$$

由相关系数定义(3)出发, 我们可得到如下描述 $\boldsymbol{\alpha}, \boldsymbol{\beta}$ 之间稀疏度关系的简单性质.

性质 1 若 l 维非负向量 $\boldsymbol{\alpha}, \boldsymbol{\beta}$ 相互正交, 即 $\rho(\boldsymbol{\alpha}, \boldsymbol{\beta}) = 0$, 此时若稀疏度 $sp(\boldsymbol{\alpha}) = e$, 则稀疏度 $sp(\boldsymbol{\beta}) \geq 1 - e$.

证明 由 $\boldsymbol{\alpha}, \boldsymbol{\beta}$ 非负性, 若 $\boldsymbol{\alpha}, \boldsymbol{\beta}$ 相互正交, 则对任意同位置向量元素 α_i, β_i , 一定满足 $\alpha_i \times \beta_i = 0$. 这意味着 α_i 不为 0 时, β_i 一定为 0, 因此 $sp(\boldsymbol{\alpha}) = e$ 意味着 $sp(\boldsymbol{\beta}) \geq 1 - e$ 一定成立.

性质 1 表明, 在非负定义域内, 向量 $\boldsymbol{\beta}$ 可以采用与向量 $\boldsymbol{\alpha}$ 的正交性来控制其稀疏度.

3.2 卷积网络鉴别正交特征生成损失

设 D 类别训练样本通过卷积网络生成的第 s 层特征为 $\{\mathbf{h}_i^{D(i)}(s)\}_{1 \leq i \leq N}$. 其中, $\mathbf{h}_i^{D(i)}(s)$ 为第 i 个训练样本的第 s 层特征, 该样本属于第 $D(i)$ 个类.

假定所有第 s 层特征满足非负特性, 即 $\mathbf{h}_i^{D(i)}(s) \geq 0$, 此时属于第 j 个类和不属于第 j 个类的训练样本的第 s 层特征均值向量 $\mathbf{m}_j(s), \tilde{\mathbf{m}}_j(s)$ 可分别定义为

$$\begin{cases} \mathbf{m}_j(s) = (1/N_j) \sum_{D(i)=j} \mathbf{h}_i^{D(i)}(s) \\ \tilde{\mathbf{m}}_j(s) = (1/(N - N_j)) \sum_{D(i) \neq j} \mathbf{h}_i^{D(i)}(s) \end{cases} \quad (4)$$

其中 N_j 为第 j 个类训练样本数.

对于第 j 个类的任意一个训练样本的第 s 层特征 $\mathbf{h}^j(s)$, 我们将其与 j 类和非 j 类的相关性度量分别定义为该特征与 $\mathbf{m}_j(s)$ 和 $\tilde{\mathbf{m}}_j(s)$ 的相关系数

$$\begin{cases} \rho(\mathbf{h}^j(s), \mathbf{m}_j(s)) = \frac{\mathbf{h}^j(s)^T \mathbf{m}_j(s)}{\sqrt{\|\mathbf{h}^j(s)\|^2 \|\mathbf{m}_j(s)\|^2}} \\ \rho(\mathbf{h}^j(s), \tilde{\mathbf{m}}_j(s)) = \frac{\mathbf{h}^j(s)^T \tilde{\mathbf{m}}_j(s)}{\sqrt{\|\mathbf{h}^j(s)\|^2 \|\tilde{\mathbf{m}}_j(s)\|^2}} \end{cases} \quad (5)$$

基于上述相关性度量, 我们设计如下添加 s 层鉴别正交特征生成损失正则化项的损失函数

$$\min L = L1 + c(s) * (\rho_B(s) - \rho_W(s)) \quad (6)$$

其中, $\rho_B(s) = (1/N) \sum_{i=1}^N \rho(\mathbf{h}_i^{D(i)}(s), \tilde{\mathbf{m}}_{D(i)}(s))$, $\rho_W(s) = (1/N) \sum_{i=1}^N \rho(\mathbf{h}_i^{D(i)}(s), \mathbf{m}_{D(i)}(s))$ 分别为平均类间相关度和平均类内相关度, $c(s)$ 为第 s 层正则化系数.

式(6)中引入基于相关度的鉴别特征生成损失, 要

求中间层特征平均类内相关度 $\rho_W(s)$ 大且平均类间相关度 $\rho_B(s)$ 小. $\mathbf{h}^j(s), \mathbf{m}_j(s), \tilde{\mathbf{m}}_j(s)$ 非负, 因此 $1 \geq \rho(\mathbf{h}^j(s), \tilde{\mathbf{m}}_j(s)) \geq 0, 1 \geq \rho(\mathbf{h}^j(s), \mathbf{m}_j(s)) \geq 0$. 此时 $\rho_B(s)$ 尽可能小则意味着类间特征之间尽可能正交, 从而生成第 s 层的鉴别正交特征. 进一步, 假定 $\tilde{\mathbf{m}}_j(s)$ 稀疏度小, 由性质(1)可知, 通过正交化约束可使 $\mathbf{h}^j(s)$ 稀疏度变大.

3.3 鉴别正交特征生成损失反向传播计算

令 $L2 = c(s) * (\rho_B(s) - \rho_W(s))$, 式(6)对第 s 层特征 $\mathbf{h}_i^{D(i)}(s)$ 求导, 可得如下梯度计算公式

$$\frac{dL}{d\mathbf{h}_i^{D(i)}(s)} = \frac{dL1}{d\mathbf{h}_i^{D(i)}(s)} + \frac{dL2}{d\mathbf{h}_i^{D(i)}(s)} \quad (7)$$

其中, $\frac{dL1}{d\mathbf{h}_i^{D(i)}(s)}$ 是互熵损失函数对 $\mathbf{h}_i^{D(i)}(s)$ 的梯度.

$$\begin{aligned} \frac{dL2}{d\mathbf{h}_i^{D(i)}(s)} &= \frac{\frac{c(s)}{N} \tilde{\mathbf{m}}_{D(i)}(s)}{\sqrt{\|\tilde{\mathbf{m}}_{D(i)}(s)\|^2 \|\mathbf{h}_i^{D(i)}(s)\|^2}} \\ &\quad - \frac{\frac{c(s)}{N} \mathbf{m}_{D(i)}(s)}{\sqrt{\|\mathbf{m}_{D(i)}(s)\|^2 \|\mathbf{h}_i^{D(i)}(s)\|^2}} \\ &\quad - (\rho(\mathbf{h}_i^{D(i)}(s), \tilde{\mathbf{m}}_{D(i)}(s)) \\ &\quad - \rho(\mathbf{h}_i^{D(i)}(s), \mathbf{m}_{D(i)}(s))) \frac{c(s) \mathbf{h}_i^{D(i)}(s)}{\|\mathbf{h}_i^{D(i)}(s)\|^2} \\ &\quad + \frac{c(s)}{N - N_{D(i)}} \sum_{D(j) \neq D(i)} \rho(\mathbf{h}_j^{D(j)}(s), \tilde{\mathbf{m}}_{D(j)}(s)) \\ &\quad \cdot \left(\frac{\mathbf{h}_j^{D(j)}(s)}{\tilde{\mathbf{m}}_{D(j)}(s)^T \mathbf{h}_j^{D(j)}(s)} - \frac{\tilde{\mathbf{m}}_{D(j)}(s)}{\|\tilde{\mathbf{m}}_{D(j)}(s)\|^2} \right) \\ &\quad - \frac{c(s)}{N_{D(i)}} \sum_{i=1}^{N_{D(i)}} \rho(\mathbf{h}_i^{D(i)}(s), \mathbf{m}_{D(i)}(s)) \\ &\quad \cdot \left(\frac{\mathbf{h}_i^{D(i)}(s)}{\mathbf{m}_{D(i)}(s)^T \mathbf{h}_i^{D(i)}(s)} - \frac{\mathbf{m}_{D(i)}(s)}{\|\mathbf{m}_{D(i)}(s)\|^2} \right) \end{aligned}$$

其中, $N_{D(i)}$ 为第 $D(i)$ 类的样本数.

引入 s 层鉴别正交特征生成损失正则化项后, 卷积网络反向传播过程中, 对第 s 层特征的反向传播计算采用式(7)实现.

3.4 鉴别正交特征损失正则化系数自动调节

引入鉴别正交损失后, 生成特征呈稀疏特性. 此时, 特征稀疏程度与正则化系数 $c(s)$ 值大小密切相关. 为精准调节对非负输出层进行鉴别正交操作得到的稀疏特征的稀疏度, 以达到控制网络容量的目的. 我们设计如下正则化系数 $c(s)$ 更新公式

$$c(s) = c(s) \times \frac{1 - sp(\mathbf{h}(s))}{1 - sp} \quad (8)$$

其中, $sp, sp(\mathbf{h}(s))$ 分别为预设的稀疏度阈值和当前第 s 层输出的平均稀疏度.

观察式(8)可知,若当前特征平均稀疏度大于预设阈值,则 $c(s)$ 减小,意味着降低鉴别正交特征生成比重;反之则 $c(s)$ 增大,意味着进一步强化鉴别正交特征生成。

进一步,我们提出如下通过自动调节正则化系数 $c(s)$ 控制第 s 层鉴别正交特征稀疏度的方法:

第 1 步 网络优化初始,预设第 s 层特征的稀疏度阈值 sp 和正则化系数 $c(s)$ 的初始值, $i = 1$;

第 2 步 第 i 个 flop 训练结束,统计当前第 s 层特征的平均稀疏度 $sp(\mathbf{h}(s))$;

第 3 步 依据式(8)更新 $c(s)$;

第 4 步 $i = i + 1$,若满足训练终止条件,则终止,否则转第 2 步。

3.5 随机 2 类别鉴别正交反向传播

由式(7)计算可知,计算均值 $\tilde{\mathbf{m}}_{D(i)}(s)$ 需要进行 $N - N_{D(i)}$ 次求和操作。特别是对多类任务(类别数大于 2),在整个迭代反向传播过程中该求和操作计算成本较大。此外,直接采用式(5)优化类间相关度 $\rho_B(s)$,在每次迭代中均强制第 $D(i)$ 类样本和其他所有非 $D(i)$ 类的样本均值 $\tilde{\mathbf{m}}_{D(i)}(s)$ 尽量正交,特征生成限定条件过强,实践中发现极易出现特征过度稀疏情况。鉴于此,对 D 类别任务($D > 2$),我们设计了如下随机 2 类别鉴别正交反向传播。

每一次反向传播,随机生成 2 个类别号 $i, j (i, j \leq D, i \neq j)$ 。此时,非 i 类样本和非 j 类样本均值向量分别为 $\tilde{\mathbf{m}}_i(s) = \mathbf{m}_j(s), \tilde{\mathbf{m}}_j(s) = \mathbf{m}_i(s)$ 。

进一步我们设计如下随机 2 类鉴别正交反向传播规则:

对 $\mathbf{h}^i(s), \mathbf{h}^j(s)$,使用式(7)反向传播;

对 $\mathbf{h}^k(s) (k \neq i, j)$,仅使用 $\frac{dL1}{d\mathbf{h}^k(s)}$ 反向传播。

4 实验结果与分析

4.1 实验描述

实验分别在 MNIST 手写体数字数据集和 Alzheimer's Disease Neuroimaging Initiative (ADNI) 脑影像公共数据集上进行。其中, MNIST 数据集包含 10 类总计 6 万个训练样本和 1 万个测试样本。ADNI 数据集我们仅使用其中的 3D sMRI 结构磁共振影像。影像数据通过去除头动等异常数据等标准预处理后包含老年痴呆症患者样本 122 个,健康对照组样本 120 个。

对 MNIST 数据集,我们使用引入非线性映射层的 LeNet 网络作为基本架构进行对比实验。该架构在原始 LeNet5 网络(图 1 所示)的 pool1, pool2 层和 ip1 层后分别添加 ReLU 非线性层 ReLU1, ReLU2 和 ReLU3,使上述三层特征输出非负。对比实验分别在基本 LeNet 网

络,基于 LeNet 网络架构的孪生网络,引入 Dropout 的 LeNet 网络和鉴别正交特征生成 LeNet 网络上进行。其中,微调阶段孪生网络卷积层学习率统一降低为 0.001。对引入 Dropout 的 LeNet 网络,其 Dropout 层紧邻放置在 ReLU 层后,Dropout 层随机置 0 率预设为 0.5。网络采用随机梯度下降算法进行反向传播,其 batchsize = 100(每类 10 个样本)。

对于 ADNI 数据集,我们采用以上应用于 MNIST 上同样的 LeNet 网络架构和参数,唯一不同的是该网络改为适用于 3D 影像数据的 3D 卷积网络。第 1 卷积层 3D 卷积核尺寸为 $5 * 5 * 5$,输入通道为 1,输出通道为 20;第 2 卷积层 3D 卷积核尺寸为 $5 * 5 * 5$,输入通道为 20,输出通道为 50,并设定 batchsize = 20(每类 10 个样本)。

所有网络均采用 Caffe^[24] 实现,使用 GTX 980Ti 6GB 显卡运行。针对本文提出的鉴别正交特征生成,我们在 Caffe 下设计和实现了一个 Dis 层,该层正向传播时为一个恒等层,输出等于输入,其反向传播执行本文提出的随机 2 类别鉴别正交反向传播。实验中,鉴别正交特征生成卷积网络的 2 个 Dis 层 Dis1 和 Dis2 分别紧邻放置在 ReLU1 层和 ReLU2 层之后,正则化系数 $c(s)$ 统一预设为 $c(s) = 0.01$ 。稀疏度目标阈值统一在 $\{0.5, 0.6, 0.7, 0.8, 0.9\}$ 中选取。

我们在实验中还对网络中间层特征进行了重构输入可视化研究。重构采用文献[21]提出的反卷积方法(<https://github.com/piergiaj/caffe-deconvnet>)。对于 3D 卷积网络,其反卷积重构可视化程序,我们在原始代码上改写实现。

4.2 MNIST 数据集实验

为测试鉴别正交特征生成方法在小样本下的有效性,我们从 6 万个训练样本中随机抽取 100/200/300/1000/2000/3000 个样本作训练样本(每类 10/20/30/100/200/300 个),最后在 1 万个测试样本上进行分类测试。实验重复 20 次,结果如表 1 所示。

表 1 不同训练样本数目下的平均正确率表

训练样本数	测试正确率 / % (平均值 ± 标准差)			
	基本 LeNet 网络	LeNet 孪生网络	Dropout LeNet 网络	鉴别正交特征生成 LeNet 网络
100	80.3 ± 1.1	81.3 ± 1.2	83.8 ± 1.0	87.5 ± 0.7
200	86.8 ± 1.4	87.0 ± 1.0	90.1 ± 1.2	91.5 ± 0.9
300	89.5 ± 0.8	89.6 ± 0.7	92.2 ± 0.4	92.8 ± 0.5
1000	93.9 ± 0.3	93.9 ± 0.9	96.0 ± 0.4	96.0 ± 0.6
2000	95.7 ± 0.2	95.9 ± 0.5	97.2 ± 0.3	96.7 ± 0.3
3000	96.6 ± 0.2	96.7 ± 0.3	97.9 ± 0.3	97.3 ± 0.3

由表 1 可看出,相比基本网络,在小样本任务上,孪生网络,Dropout 网络和我们的鉴别正交特征生成卷积网络均有不同程度的改善.特别当在训练样本很少,数目为 100、200 时,在平均正确率上鉴别正交特征生成卷积网络相对于其他网络优势明显.但随着样本数逐渐增加,鉴别正交特征生成网络优势逐渐缩小,在样本数为 2000/3000 时,Dropout 网络表现要略微优于鉴别正交特征生成卷积网络.

为进一步说明鉴别正交特征生成的优势和机理,我们对本实验中的两种极端样本情况(样本数分别为

100/3000)下基本网络和我们的鉴别正交特征生成卷积网络进行了更深入的比较分析.首先我们对比并绘制了训练测试损失曲线,如图 3 所示.

由图 3 可看出,100 个训练样本时,两种网络模型训练损失收敛曲线无明显差异,但测试损失变化情况明显不同.基本网络在训练早期其测试损失急速下降后出现一定程度反向增长然后维持平稳;鉴别正交特征生成网络测试损失下降速度不如基本网络,但整个训练过程中测试损失保持下降趋势,且最终测试损失明显低于基本网络,泛化性更好.

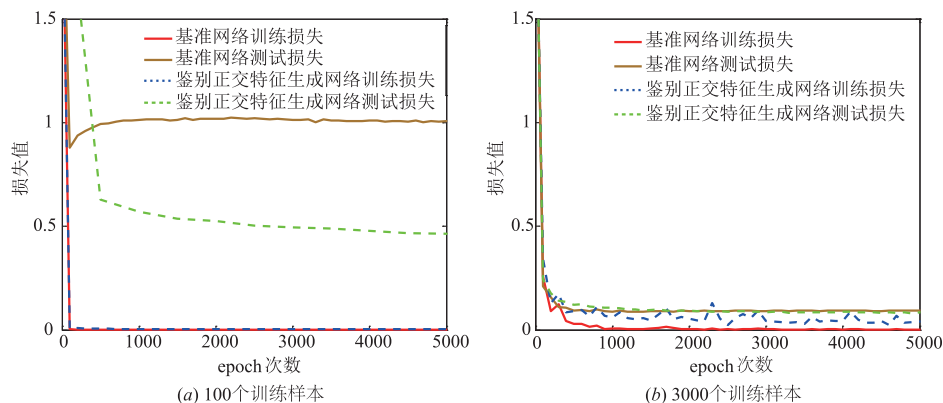


图3 训练、测试损失收敛曲线图

3000 个训练样本时,相比于基本网络,鉴别正交特征生成网络具有更高的训练损失和相对更低的测试损失,表明鉴别正交特征生成网络具有更强的网络容量控制能力.

以上实验结果均是对稀疏度目标阈值进行网格搜索后的最优结果.网格搜索在随机抽取的包含 10000 个样本(每类 1000 个样本)的验证集上进行,结果如图 4

所示.

图 4 表明,100 个训练样本下,稀疏度目标阈值对网络性能影响较大.随着训练样本数持续增加,稀疏度目标阈值对网络性能影响逐渐减小,同时最优稀疏度目标阈值也在逐渐减小.由图 4 可知,100/3000 个训练样本下的稀疏度目标阈值最终分别确定为 0.8/0.7.

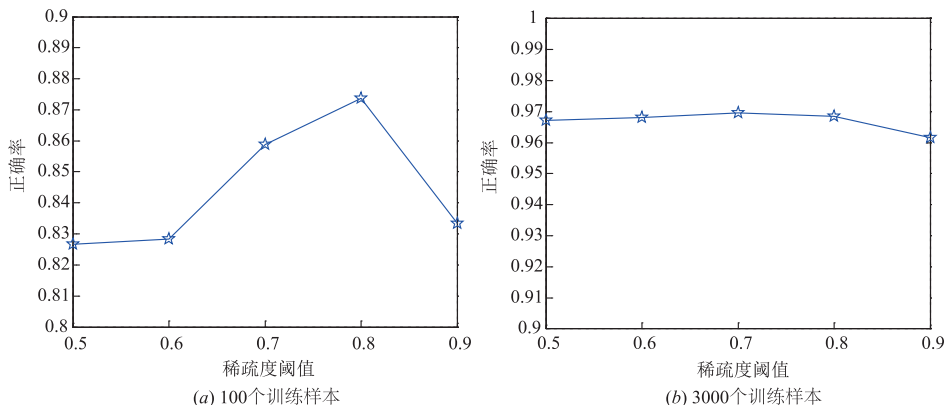


图4 稀疏度目标阈值网格搜索图

进一步,我们对网络中 ReLU1/ReLU2 非线性层特征输出真实稀疏度情况进行了统计,如图 5 所示.

图 5 表明,鉴别正交特征生成网络真实稀疏度基本达到了预设稀疏度目标阈值.与基本网络相比,鉴别

正交特征生成网络具有更大的稀疏度.

最后,我们对特征进行可视化分析.从测试样本中随机挑选 0~9 数字各 1 个,通过可视化技术对 Pool2 层输出到全互联层的特征进行反向重构,如图 6 所示.

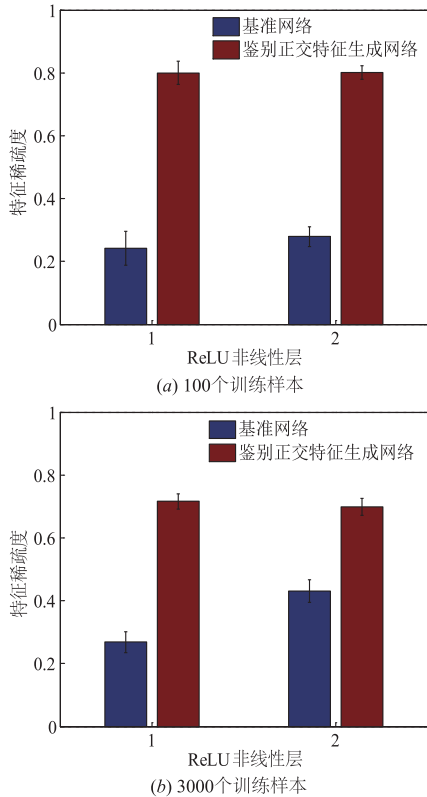


图5 ReLU层特征稀疏度统计图

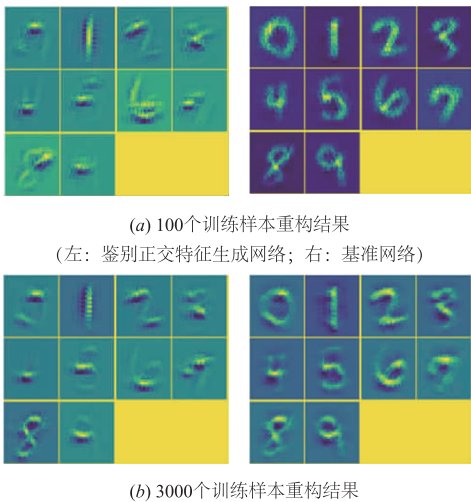


图6 MNIST数据Pool2层输出反向重构图

观察图 6, 相比于基本网络, 更具稀疏度的鉴别正交特征生成网络更关注局部鉴别区域, 对其他区域有明显抑制作用, 具有很强的聚焦性. 此外, 我们发现 100/3000 个训练样本下, 鉴别正交特征重构的激活区域比较稳定, 而原始网络特征重构的高亮局部区域差异更明显, 且 3000 个样本下原始网络特征重构的高亮区域与鉴别正交特征重构的激活区域趋于一致.

4.3 老年痴呆症磁共振影像分析实验

我们将鉴别正交特征生成网络应用到基于 3D sMRI 脑灰质影像分析的老年痴呆症自动诊断任务上. 实验在 ADNI 数据集上进行.

原始 3D sMRI 脑影像的 T1 像使用 SPM8 和 VBM8 工具包在 MATLAB 2013b 平台下进行预处理. 首先使用 VBM8 将全脑影像标准化到 the standard Montreal Neurological Institute (MNI) 空间, 然后将其分割成灰质、白质和脑脊液三个部分. 实验中我们仅使用灰质影像进行诊断和分析老年痴呆症病人与正常人在灰质皮层上的显著差异脑区. 因此, 进一步使用 SPM8 单独对灰质影像进行标准化, 并进行尺寸为 $8\text{mm} \times 8\text{mm} \times 8\text{mm}$ 的窗口平滑操作. 原始影像经一系列预处理操作, 最终得到脑灰质影像, 如图 7 所示.

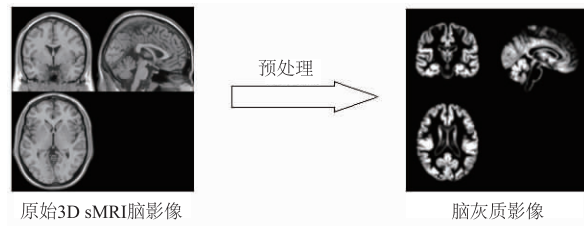


图7 脑影像预处理结果示意图

预处理后脑灰质影像尺寸为 $121 \times 145 \times 121$. 为降低计算成本, 进一步降采样至 $32 \times 32 \times 32$ 维, 作为 3D 卷积网络的最终输入.

首先, 我们进行老年痴呆症病人-正常人分类实验. 实验采用 5 重交叉验证方式, 每重训练使用 3 份样本, 验证使用 1 份, 测试使用最后 1 份. 5 重交叉验证的平均分类正确率结果如表 2 所示.

表 2 ADNI 分类平均分类正确率表

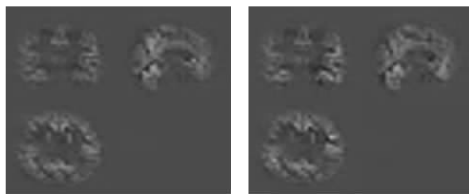
平均测试正确率 / % (平均值 \pm 标准差)				
文献[25]	文献[26]	文献[27]	基本 LeNet 网络	鉴别正交特征生成 LeNet 网络
87.0 \pm 4.6	90.2 \pm 5.3	92.4 \pm 5.3	90.4 \pm 5.7	94.8 \pm 5.0

表 2 中不仅给出了基本 LeNet 网络和鉴别正交特征生成 LeNet 网络的分类结果, 还列出了来自文献[25~27]的 ADNI 数据集上的分类结果. 其中文献[25,26]使用的是浅层方法, 文献[27]使用的是自行设计的卷积网络. 此外, 表中文献[27]结果是 MRI 和 PET 多模态融合后的最好结果, 优于单独 MRI 影像结果.

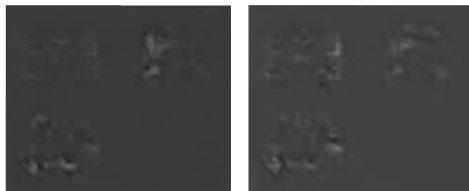
表 2 表明, 相比与基本网络, 鉴别正交特征生成网络取得了 4.4% 的平均正确率提升. 此外, 与其他重要文献报道的分类实验结果相比, 鉴别正交特征生成网络也取得了更好的效果.

进一步, 我们采用反卷积技术由 Pool2 层输出反向

重构脑灰质 3D 影像,可视化结果如图 8 所示.



(a) 基本网络重构结果
(左: 老年痴呆症; 右: 健康对照组)



(b) 鉴别正交特征生成网络重构结果
(左: 老年痴呆症; 右: 健康对照组)

图8 ADNI数据Pool2层输出反向重构图

由可视化对比结果可以看出,采用鉴别正交特征生成网络相比与基本网络具有更好的聚焦性,因此能够更好的定位具有典型差异的重要脑区.

最后,我们使用 xjview 工具包对鉴别正交特征生成网络发现的重要脑区进行自动标记和分析,结果如图 9 所示.

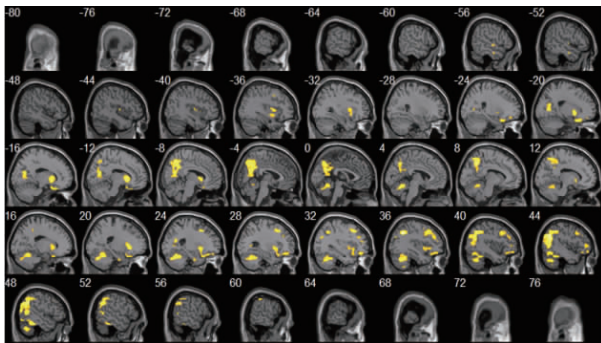


图9 鉴别脑区定位图

通过 xjview 聚类分析功能,最终找到的老年痴呆症病人与正常人的典型差异脑区主要有:额内侧回,杏仁体,内嗅皮层,海马结构.该分析结果与文献[24,28,29]报道结果相吻合.

5 结束语

本文提出一种鉴别正交特征生成方法,以改善深度卷积网络在小样本任务上的学习效果.该方法通过引入关于中间层特征的同类/异类相关度正则化项,生成鉴别正交特征;采用正则化系数自适应调整技术,达到调节特征稀疏度以控制网络表达容量的目的.为提高计算效率,本文还设计了一个随机 2 类别鉴别正交反向传播规则.为验证本文方法的有效性,在 MNIST 上

进行了小样本学习对比实验.实验结果表明本文方法具有较高的稀疏度和良好的聚焦性,能够更好地定位典型差异区域,改善分类性能.最后,将本文方法应用于老年痴呆症脑磁共振影像分析上,获得了更好的分类效果,并成功定位了老年痴呆症病人与对照组的典型差异脑区.

参考文献

- [1] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [2] Schmidhuber J. Deep learning in neural networks: An overview[J]. Neural Networks, 2015, 61(1): 85-117.
- [3] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[A]. International Conference on Neural Information Processing Systems[C]. Doha, Qatar: Curran Associates Inc, 2012. 1097-1105.
- [4] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[A]. Computer Vision and Pattern Recognition[C]. USA, Boston: IEEE, 2015. 3431-3440.
- [5] Dai J, Li Y, He K, et al. R-FCN: Object Detection via Region-based Fully Convolutional Networks[Z]. Preprint arXiv:1605.06409, 2016.
- [6] Lin K, Yang H F, Hsiao J H, et al. Deep learning of binary hash codes for fast image retrieval[A]. Computer Vision and Pattern Recognition Workshops[C]. USA, Boston: IEEE, 2015. 27-35.
- [7] Halevy A, Norvig P, Pereira F. The unreasonable effectiveness of data[J]. IEEE Intelligent Systems, 2009, 24(2): 8-12.
- [8] Sun C, Shrivastava A, Singh S, et al. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era[Z]. Preprint arXiv:1707.02968, 2017.
- [9] Vapnik V N. Statistical Learning Theory[M]. [S. l.]: Publishing House of Electronics Industry, 2015.
- [10] 庄福振, 罗平, 何清, 等. 迁移学习研究进展[J]. 软件学报, 2015, 26(1): 26-39.
Zhuang Fu-Zhen, Luo Ping, He Qing, et al. Survey on transfer learning research[J]. Journal of Software, 2015, 26(1): 26-39. (in Chinese)
- [11] Koch G. Siamese Neural Networks for One-Shot Image Recognition[D]. University of Toronto, 2015.
- [12] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning[A]. Proceedings of Advances in Neural Information Processing Systems[C]. Spain, Barcelona, 2016. 3630-3638.
- [13] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A

- simple way to prevent neural networks from overfitting [J]. *Journal of Machine Learning Research*, 2014, 15 (1):1928 – 1958.
- [14] Hinton G E. Connectionist learning procedures [J]. *Artificial Intelligence*, 1989, 40(1–3):185 – 234.
- [15] Yosinski J, Clune J, Bengio Y, et al. How transferable are features in deep neural networks? [A]. *International Conference on Neural Information Processing Systems* [C]. Canada, Montreal, 2014. 3320 – 3328.
- [16] Azizpour H, Razavian A S, Sullivan J, et al. From generic to specific deep representations for visual recognition [A]. *Computer Vision and Pattern Recognition Workshops* [C]. USA, Boston: IEEE, 2015. 36 – 45.
- [17] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional Architecture for Fast Feature Embedding [Z]. Preprint arXiv:1408.5093, 2014.
- [18] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. *Proceedings of the IEEE*, 1998, 86(11):2278 – 2324.
- [19] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [A]. *Computer Vision and Pattern Recognition Workshops* [C]. USA, Boston: IEEE, 2015. 1 – 9.
- [20] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [A]. *Computer Vision and Pattern Recognition Workshops* [C]. USA, Las Vegas: IEEE, 2016. 770 – 778.
- [21] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks [A]. *European Conference on Computer Vision* [C]. Swiss, Zurich: 2014, 8689:818 – 833.
- [22] 孙晓, 潘汀. 基于兴趣区域深度神经网络的静态面部表情识别 [J]. *电子学报*, 2017, 45(5):1189 – 1197. SUN Xiao, PAN Ting. Static facial expression recognition system using ROI deep neural networks [J]. *Acta Electronica Sinica*, 2017, 45(5):1189 – 1197. (in Chinese)
- [23] 胡正平, 陈俊岭. 多层融合深度局部 PCA 子空间稀疏优化特征提取模型 [J]. *电子学报*, 2017, 45(10):2383 – 2389. HU Zheng-ping, CHEN Jun-ling. Feature extraction model based on multi-layered deep local subspace sparse optimization [J]. *Acta Electronica Sinica*, 2017, 45(10):2383 – 2389. (in Chinese)
- [24] Ewers M, Walsh C, Trojanowski J Q, et al. Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance [J]. *Neurobiology of Aging*, 2012, 33(7):1203 – 1214.
- [25] Westman E, Muehlboeck J S, Simmons A. Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion [J]. *Neuroimage*, 2012, 62(1):229 – 238.
- [26] Liu M, Zhang D, Shen D. Ensemble sparse classification of Alzheimer's disease [J]. *Neuroimage*, 2012, 60(2):1106 – 1116.
- [27] Li R, Zhang W, Suk H I, et al. Deep learning based imaging data completion for improved brain disease diagnosis [A]. *International Conference on Medical Image Computing and Computer-Assisted Intervention* [C]. USA, Cambridge: Springer, 2014. 305 – 312.
- [28] Mueller S G, Weiner M W, Thal L J, et al. The Alzheimer's disease neuroimaging initiative [J]. *Neuroimaging Clinics of North America*, 2005, 15(4):869 – 869.
- [29] Johnson N A, Jahng G H, Weiner M W, et al. Pattern of cerebral hypoperfusion in Alzheimer disease and mild cognitive impairment measured with arterial spin-labeling MR imaging: initial experience [J]. *Radiology*, 2005, 234(3):851 – 859.

作者简介



杨 勃 男, 1974 年出生于湖南岳阳. 现为湖南理工学院副教授、硕士生导师. 主要研究方向为统计机器学习、深度学习和脑影像分析.
E-mail: ybmengshen@163.com



邵泉铭 男, 1991 年出生于四川巴中. 现为湖南理工学院硕士研究生. 主要研究方向为深度学习.
E-mail: 783295011@qq.com



方 欣 (通信作者) 男, 1970 年出生于湖南岳阳. 现为湖南理工学院副教授. 主要研究方向为统计机器学习.
E-mail: fashion90@163.com