基于核化双线性卷积网络的细粒度图像分类

葛疏雨,高子淋,张冰冰,李培华
(大连理工大学信息与通信工程学院,辽宁大连 116024)

摘 要: 双线性卷积网络(Bilinear CNN,B-CNN)在计算机视觉任务中有着广泛的应用. B-CNN 通过对卷积层输出的特征进行外积操作,能够建模不同通道之间的线性相关,从而增强了卷积网络的表达能力. 由于没有考虑特征图中通道之间的非线性关系,该方法无法充分利用通道之间所蕴含的更丰富信息. 为了解决这一不足,本文提出了一种核化的双线性卷积网络,通过使用核函数的方式有效地建模特征图中通道之间的非线性关系,进一步增强卷积网络的表达能力. 本文在三个常用的细粒度数据库 CUB-200-2011、FGVC-Aircraft 以及 Cars 上对本文方法进行了验证,实验表明本文方法在三个数据库上均优于同类方法.

关键词: 核化双线性聚合; 双线性卷积网络; 端到端学习; 细粒度图像分类
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2019)10-2134-08
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2019.10.015

Kernelized Bilinear CNN Models for Fine-Grained Visual Recognition

GE Shu-yu, GAO Zi-lin, ZHANG Bing-bing, LI Pei-hua
(*School of Information and Communication Engineering, Dalian University of Technology, Dalian, Liaoning* 116024, *China*)

Abstract: The bilinear convolutional neural network (B-CNN) has been widely used in computer vision. B-CNN can capture the linear correlation between different channels by performing the outer product operation on the features of the convolutional layer output, thus enhancing the representative ability of the convolutional network. Since the non-linear relationship between the channels in the feature map is not taken account of, this method cannot make full use of the richer information contained between the channels. In order to solve this problem, this paper proposes a kernelized bilinear convolutional neural network employing the kernel function to effectively capture the non-linear relationship between the channels in the feature map, and further enhancing the representative ability of the convolutional network. In this paper, the method is evaluated on three common fine-grained benchmarks CUB-200-2011, FGVC-Aircraft and Cars. Experiments show that our method is superior to its counterparts on all three benchmarks.

Key words: kernelized bilinear pooling; bilinear convolution neural network; end to end learning; fine-grained visual recognition

1 引言

2012 年,Krizhevsky 等人[1]提出 AlexNet 深度卷积网络模型,在 ImageNet[2]大规模图像识别任务上成功应用. 伴随着深度学习、卷积网络技术的不断发展,深度卷积网络在计算机视觉领域得到广泛的应用,如图像检索[3]、场景解析[4]、目标跟踪[5]等. 在细粒度图像识别领域,深度卷积网络也得到广泛地研究与应用. 细粒度图像识别区别于通用图像识别,是对粗粒度大类别的目标物体进行精细的子类别识别,由于在细粒度图像识别中,类间差异更小且更容易受到姿势、视角与位置等因素的影响,因此细粒度图像识别任务更具有挑战性. 邹承明等人[6]将卷积网络提取的图像表达与传统 SIFT 特征经编码之后得到的图像表达进行融合,用于细粒度图像识别. 2015 年,Lin 等人[7]提出了双线性卷积网络(B-CNN),通过对卷积层输出的特征图进行外积操作建模了特征图中通道之间的线性相关,并进行端到端的联合优化学习,在细粒度图像识别任务上取

收稿日期:2018-09-03;修回日期:2019-01-14;责任编辑:孙瑶
基金项目:国家自然科学基金(No. 61471082)

得了优异的性能. 由于无法捕捉特征图中通道之间的非线性关系,该方法没有充分地挖掘卷积网络的表达能力.

针对上述问题,本文提出了一种核化双线性卷积网络模型,通过使用核函数的方式简洁有效地建模特征图中通道之间的非线性关系,同时实现端到端的联合优化学习,进一步增强卷积网络的表达能力. 本文提出的核化双线性卷积网络架构如图 1 所示. 该网络由卷积层、核化双线性聚合模块以及 softmax 分类器三个部分组成,卷积层用于对一幅图像提取局部特征得到特征图. 核化双线性聚合模块共包含三个结构层,分别为通道二范数归一化层、核化双线性聚合层以及矩阵幂正规化层[8,9]. 通道二范数归一化层即对特征图中每个通道 $\boldsymbol{x}_i$ 进行二范数的归一化:$\boldsymbol{x}_i \leftarrow \boldsymbol{x}_i / \| \boldsymbol{x}_i \|$. 核化双线性聚合层通过对输入的特征图进行核化双线性聚合以建模特征图中通道间的非线性关系,得到该幅图像的表达 $\boldsymbol{P}$. 矩阵幂正规化层是对核化双线性聚合输出的矩阵 $\boldsymbol{P}$ 进行指数幂的操作:$\boldsymbol{P} \leftarrow \boldsymbol{P}^{\alpha}$. 由于矩阵幂正规化层输出的矩阵为对称矩阵,本文将向量化的矩阵上三角部分作为图像的最终表达. 网络架构的第三部分由全连接层与 softmax 层构成,对得到的图像表达进行分类. 为加速网络收敛,本文在全连接层之前加入批正则化(Batch Normalization,BN)[10]层.

本文内容分为以下四个部分:第一部分主要介绍在细粒度图像识别领域中的相关工作;第二部分介绍本文所提出的核化双线性卷积网络;第三部分为实验,实验首先在 FGVC-Aircraft[11]细粒度图像数据库上对本文方法进行参数评估,其次在三个细粒度数据库 CUB-200-2011[12]、FGVC-Aircraft 以及 Cars[13]上将本文方法与当前主流方法进行比较,验证本文方法的有效性;最后一部分为结论,总结本文工作并明确下一步研究工作.

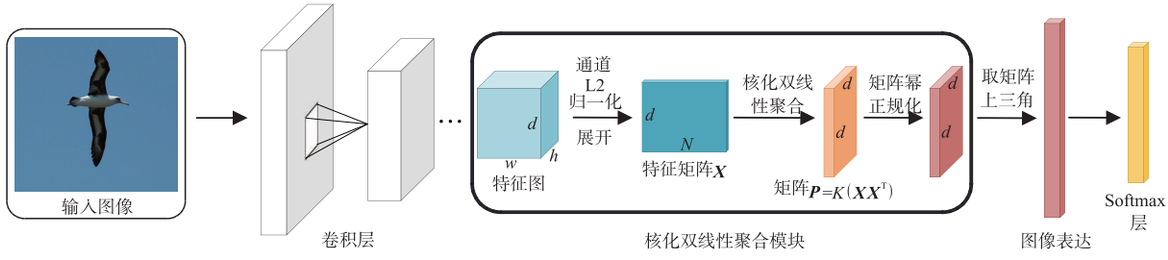


图1　核化双线性卷积网络架构图

2　相关工作

双线性卷积网络由 Lin 等人[7]于 2015 年提出,该方法通过对卷积网络中最后一个卷积层输出的特征图进行外积聚合,建模了特征分布的二阶统计信息,并作为图像的最终表达. 外积聚合如式(1)所示.

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} \tag{1}$$

其中 $\boldsymbol{X} \in \mathbb{R}^{d \times N}$,$\boldsymbol{Y} \in \mathbb{R}^{d \times d}$. 2017 年,Lin 等人[9]在双线性卷积网络的基础上,提出了改进的双线性卷积网络(Improved B-CNN),通过对外积聚合矩阵 $\boldsymbol{Y}$ 引入矩阵平方根正规化:$\boldsymbol{Y} \leftarrow \boldsymbol{Y}^{1/2}$,使得对于特征分布二阶统计信息的建模过程更为鲁棒,进一步提升网络性能. Li 等人[8]首次将矩阵幂正规化应用于大规模图像分类任务,并将矩阵幂正规化与矩阵对数正规化进行详细地比较与分析,提出了 MPN-COV 卷积网络模型,在 ImageNet[2]大规模数据集上取得了优异的性能.

围绕双线性卷积网络的压缩,2016 年,Gao 等人[14]使用张量速写(Tensor Sketch)与随机麦克劳林(Random Maclaurin)两种算法分别对外积聚合的矩阵 $\boldsymbol{Y}$ 进行低维近似,提出了压缩的双线性卷积网络(Compact Bilinear Pooling,CBP),在性能几乎不损失的情况下,较好地减少了双线性卷积网络的参数,同时降低图像表达的维度. Li 等人[15]通过对双线性卷积网络中全连接层的卷积核进行分解,提出了一种基于矩阵低秩分解的因子分解双线性网络(Factorized Bilinear Network,FBN),减少了双线性卷积网络的参数量与计算量.

针对卷积网络中特征分布建模的问题,2017 年,Cui 等人[16]提出了核聚合卷积网络(Kernel Pooling,KP),该方法并没有直接使用核运算,而是从核函数的特征映射函数角度,通过使用特征的 1 至 p 阶张量乘法近似高斯核函数的特征映射函数,并对特征进行建模作为图像的最终表达,取得了较好的性能. Wang 等人[17]提出全局高斯分布嵌入神经网络(Global Gaussian Distribution Embedding Network,G^2DeNet),网络核心是利用提出的可训练的高斯分布嵌入层对卷积层输出的特征分布进行高斯建模得到一幅图像的表达. 该方法同时考虑了高斯分布的黎曼流形结构,在图像区域识别以及细粒度图像识别任务中均取得了优异的性能.

本文从建模特征图中通道之间相互关系的角度,提出一种核化双线性卷积网络模型,通过使用核函数的方式简洁有效地建模通道之间的非线性关系并实现端到端的联合优化学习,取得了优异的性能.

3 核化双线性卷积网络

本节将重点介绍提出的核化双线性卷积网络. 首先,3.1 节介绍作为基线方法的双线性卷积网络,并指出该方法存在的不足. 针对上述不足,3.2 节将重点介绍本文提出的三种核化双线性聚合方式,并分别给出前向与反向传播的公式推导.

3.1 双线性卷积网络

令 $\boldsymbol{X}$ 表示一幅图像的特征图经展开后的特征矩阵,$\boldsymbol{X}\in\mathbb{R}^{d\times N}$,其中 N 表示特征图中包含的特征数目,d 表示特征图的通道数目即特征的维度,令 $\boldsymbol{f}_j$ 表示特征矩阵中第 j 个特征,$\boldsymbol{f}_j\in\mathbb{R}^{d\times 1}$. 双线性卷积网络通过对特征矩阵中的特征 $\boldsymbol{f}_j$ 进行外积聚合得到图像的表达,如式(2)所示.

$$\boldsymbol{Y}=\sum_{j=1}^{N}\boldsymbol{f}_j\boldsymbol{f}_j^{\mathrm{T}}=\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}=\begin{bmatrix}\langle\boldsymbol{x}_1,\boldsymbol{x}_1\rangle & \langle\boldsymbol{x}_1,\boldsymbol{x}_2\rangle & \cdots & \langle\boldsymbol{x}_1,\boldsymbol{x}_d\rangle\\ \langle\boldsymbol{x}_2,\boldsymbol{x}_1\rangle & \langle\boldsymbol{x}_2,\boldsymbol{x}_2\rangle & \cdots & \langle\boldsymbol{x}_2,\boldsymbol{x}_d\rangle\\ \vdots & \vdots & \ddots & \vdots\\ \langle\boldsymbol{x}_d,\boldsymbol{x}_1\rangle & \langle\boldsymbol{x}_d,\boldsymbol{x}_2\rangle & \cdots & \langle\boldsymbol{x}_d,\boldsymbol{x}_d\rangle\end{bmatrix}\tag{2}$$

式中,外积聚合矩阵 $\boldsymbol{Y}\in\mathbb{R}^{d\times d}$,$\boldsymbol{x}_i$ 表示 $\boldsymbol{X}$ 中的第 i 行即特征图中第 i 个通道,$\boldsymbol{x}_i\in\mathbb{R}^{1\times N}$. 从式中可以看出,双线性卷积网络通过外积聚合得到的矩阵 $\boldsymbol{Y}$ 中每个元素均为通道之间的内积,从而可以捕捉特征图中通道之间的线性关系. 在双线性卷积网络架构中,该方法对矩阵 $\boldsymbol{Y}$ 经向量化后得到的向量 $\boldsymbol{z}$ 依次进行带符号的元素级平方根正则化 $\boldsymbol{z}\leftarrow\mathrm{sign}(\boldsymbol{z})\sqrt{|\boldsymbol{z}|}$ 以及二范数归一化 $\boldsymbol{z}\leftarrow\boldsymbol{z}/\|\boldsymbol{z}\|$,并作为图像的最终表达送至 softmax 层进行端到端的联合优化学习.

Lin 等人[9]在双线性卷积网络架构中引入矩阵平方根正规化,提出了改进的双线性卷积网络,进一步提升性能. 矩阵平方根正规化[8,9]的操作如下:由于经过外积聚合得到的矩阵 $\boldsymbol{Y}$ 为对称(半)正定矩阵,记矩阵 $\boldsymbol{Y}$ 的奇异值分解为 $\boldsymbol{Y}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{T}}$,其中 $\boldsymbol{\Lambda}=\mathrm{diag}(\lambda_1,\cdots,\lambda_d)$ 即对角元为 λ_i 的对角矩阵. 则对于矩阵 $\boldsymbol{Y}$ 的幂正规化可以表示为 $\boldsymbol{Y}^{\alpha}=\boldsymbol{U}\boldsymbol{\Lambda}^{\alpha}\boldsymbol{U}^{\mathrm{T}}$,其中 $\boldsymbol{\Lambda}^{\alpha}=\mathrm{diag}(\lambda_1^{\alpha},\cdots,\lambda_i^{\alpha})$,当式中 α 为 0.5 时,即对矩阵 $\boldsymbol{Y}$ 的进行平方根正规化.

3.2 核化双线性聚合

针对双线性卷积网络中仅能建模通道间线性关系的不足,本文提出了一种核化双线性卷积网络,该网络的关键部分在于使用本文提出的核化双线性聚合替代双线性卷积网络中的外积聚合,通过使用核函数的方式直接有效地建模特征图中通道间的非线性关系,更为充分地利用通道间蕴含的丰富信息. 核化双线性聚合与外积聚合的比较如图 2 所示.

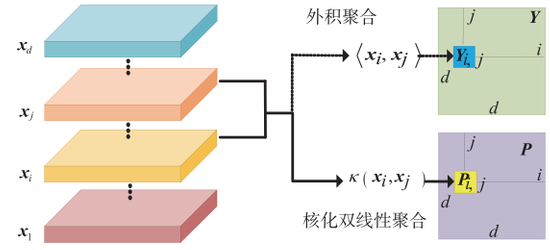


图2 核化双线性聚合与外积聚合的比较

令 $\boldsymbol{X}$ 表示卷积层输出的特征图经展开后的特征矩阵,$\boldsymbol{X}\in\mathbb{R}^{d\times N}$,$\boldsymbol{P}$ 为经过核化双线性聚合得到的矩阵,$\boldsymbol{P}\in\mathbb{R}^{d\times d}$. 核化双线性聚合可由式(3)表示.

$$\boldsymbol{Y}=K(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})=\begin{bmatrix}\kappa(\boldsymbol{x}_1,\boldsymbol{x}_1) & \kappa(\boldsymbol{x}_1,\boldsymbol{x}_2) & \cdots & \kappa(\boldsymbol{x}_1,\boldsymbol{x}_d)\\ \kappa(\boldsymbol{x}_2,\boldsymbol{x}_1) & \kappa(\boldsymbol{x}_2,\boldsymbol{x}_2) & \cdots & \kappa(\boldsymbol{x}_2,\boldsymbol{x}_d)\\ \vdots & \vdots & \ddots & \vdots\\ \kappa(\boldsymbol{x}_d,\boldsymbol{x}_1) & \kappa(\boldsymbol{x}_d,\boldsymbol{x}_2) & \cdots & \kappa(\boldsymbol{x}_d,\boldsymbol{x}_d)\end{bmatrix}\tag{3}$$

其中 $\boldsymbol{x}_i$ 为特征图中的第 i 个通道,$\boldsymbol{x}_i\in\mathbb{R}^{1\times N}$,$\kappa$ 为本文所使用的核函数. 本文分别提出指数核化双线性聚合、多项式核化双线性聚合以及 sigmoid 核化双线性聚合,用于通道之间非线性关系的建模并实现端到端的联合优化学习. 以下将分别介绍上述三种核化双线性聚合方式及其正向与反向传播的公式推导.

本文使用 Ionescu 等人[18]提出的矩阵反向传播的梯度求导方法,对核化双线性聚合层进行反向传播的梯度求导. 记矩阵 $\boldsymbol{A}=\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$,则核化双线性聚合的公式可以表示为 $\boldsymbol{P}=K(\boldsymbol{A})$. 令 $\partial l/\partial\boldsymbol{A}$,$\partial l/\partial\boldsymbol{P}$ 分别表示损失函数 l 对于矩阵 $\boldsymbol{A}$ 与 $\boldsymbol{P}$ 的梯度,则由 $\partial l/\partial\boldsymbol{A}$ 可得损失函数 l 对于特征矩阵 $\boldsymbol{X}$ 的梯度如式(4)所示.

$$\frac{\partial l}{\partial\boldsymbol{X}}=\left(\frac{\partial l}{\partial\boldsymbol{A}}+\left(\frac{\partial l}{\partial\boldsymbol{A}}\right)^{\mathrm{T}}\right)\boldsymbol{X}\tag{4}$$

根据链式法则,在对三种核化双线性聚合方式的反向梯度推导过程中,只需分别求出损失函数 l 对矩阵 $\boldsymbol{A}$ 的梯度 $\partial l/\partial\boldsymbol{A}$.

指数核化双线性聚合 指数核函数如式(5)所示.

$$\kappa(\boldsymbol{x}_i,\boldsymbol{x}_j)=\exp(\beta\cdot\langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle)\tag{5}$$

其中 $\beta>0$. 由式(5)可以得出指数核化双线性聚合的公式如下:

$$\boldsymbol{P}=\exp(\beta\cdot\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})=\exp(\beta\cdot\boldsymbol{A})=[\exp(\beta\cdot\langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle)]_{d\times d}\tag{6}$$

β 为可调的标量参数. 反向传播过程中,损失函数 l 对于矩阵 $\boldsymbol{A}$ 的梯度求导公式如下:

$$\frac{\partial l}{\partial\boldsymbol{A}}=\beta\cdot\exp(\beta\cdot\boldsymbol{A})\circ\frac{\partial l}{\partial\boldsymbol{P}}=\beta\cdot\boldsymbol{P}\circ\frac{\partial l}{\partial\boldsymbol{P}}\tag{7}$$

其中“∘”表示两个矩阵对应元素的乘积,即哈达马积.

多项式核化双线性聚合　为避免与前文中矩阵幂正规化:$\boldsymbol{Y}^{\alpha}=\boldsymbol{U}\boldsymbol{\Lambda}^{\alpha}\boldsymbol{U}^{\mathrm{T}}$ 相混淆,定义 $\boldsymbol{Y}^{(\alpha)}$ 表示对于任意矩阵 $\boldsymbol{Y}$ 进行元素级的 α 幂操作. 多项式核函数如式(8)所示.

$$\kappa(\boldsymbol{x}_i,\boldsymbol{x}_j)=(\langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle+1)^c \tag{8}$$

其中 c 是多项式的次数,为可调的标量参数. 则多项式核化双线性聚合如式(9)所示.

$$\begin{aligned}\boldsymbol{P}&=(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}+\boldsymbol{1}_{d\times d})^{(c)}=(\boldsymbol{A}+\boldsymbol{1}_{d\times d})^{(c)}\\&=[(\langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle+1)^c]_{d\times d}\end{aligned} \tag{9}$$

其中 $\boldsymbol{1}_{d\times d}$ 表示元素均为 1 的 d 维方阵. 其梯度反向传播公式如下:

$$\begin{aligned}\frac{\partial l}{\partial\boldsymbol{A}}&=c\cdot(\boldsymbol{A}+\boldsymbol{1}_{d\times d})^{(c-1)}\circ\frac{\partial l}{\partial\boldsymbol{P}}\\&=c\cdot\boldsymbol{P}^{\left(\frac{c-1}{c}\right)}\circ\frac{\partial l}{\partial\boldsymbol{P}}\end{aligned} \tag{10}$$

sigmoid 核化双线性聚合　sigmoid 核函数如式(11)所示.

$$\kappa(\boldsymbol{x}_i,\boldsymbol{x}_j)=\tanh(\theta\cdot\langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle+\gamma) \tag{11}$$

其中 θ 为幅度调节参数,$\theta>0$,γ 为位移参数,$\gamma<0$[19]. 则 sigmoid 核化双线性聚合公式如下:

$$\begin{aligned}\boldsymbol{P}&=\tanh(\theta\cdot\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}+\gamma\cdot\boldsymbol{1}_{d\times d})\\&=\tanh(\theta\cdot\boldsymbol{A}+\gamma\cdot\boldsymbol{1}_{d\times d})\\&=[\tanh(\theta\cdot\langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle+\gamma)]_{d\times d}\end{aligned} \tag{12}$$

其梯度反向传播公式如式(13)所示.

$$\begin{aligned}\frac{\partial l}{\partial\boldsymbol{A}}&=\theta\cdot(1-(\tanh(\theta\cdot\boldsymbol{A}+\gamma\cdot\boldsymbol{1}_{d\times d}))^{(2)})\circ\frac{\partial l}{\partial\boldsymbol{P}}\\&=\theta\cdot(1-\boldsymbol{P}^{(2)})\circ\frac{\partial l}{\partial\boldsymbol{P}}\end{aligned} \tag{13}$$

综上所述,本文提出的三种核化双线性聚合方法如表 1 所示. 在核化双线性卷积网络架构中,由核化双线性聚合得到的矩阵 $\boldsymbol{P}$,经过矩阵的幂正规化[8,9]:$\boldsymbol{P}\leftarrow\boldsymbol{P}^{\alpha}$,并取输出矩阵的上三角部分,经向量化后得到向量 $\boldsymbol{z}$,作为图像的最终表达送至 softmax 分类器实现端到端的联合优化学习.

表 1　本文提出的三种核化双线性聚合

核化双线性聚合	前向传播	反向传播
指数核化双线性聚合	$\boldsymbol{P}=\exp(\beta\cdot\boldsymbol{A})$	$\frac{\partial l}{\partial\boldsymbol{A}}=\beta\cdot\boldsymbol{P}\circ\frac{\partial l}{\partial\boldsymbol{P}}$
多项式核化双线性聚合	$\boldsymbol{P}=(\boldsymbol{A}+\boldsymbol{1}_{d\times d})^{(c)}$	$\frac{\partial l}{\partial\boldsymbol{A}}=c\cdot\boldsymbol{P}^{\left(\frac{c-1}{c}\right)}\circ\frac{\partial l}{\partial\boldsymbol{P}}$
sigmoid 核化双线性聚合	$\boldsymbol{P}=\tanh(\theta\cdot\boldsymbol{A}+\gamma\cdot\boldsymbol{1}_{d\times d})$	$\frac{\partial l}{\partial\boldsymbol{A}}=\theta\cdot(1-\boldsymbol{P}^{(2)})\circ\frac{\partial l}{\partial\boldsymbol{P}}$

4　实验

数据库　实验部分在三个细粒度图像识别数据库 CUB-200-2011[12]、FGVC-Aircraft[11] 以及 Cars[13] 上对本文方法进行评估. CUB-200-2011 数据库共包含来自 200 个鸟类物种的 11788 张图像,其中 5994 张训练与验证图像,5794 张测试图像. FGVC-Aircraft 数据库包括 1 万张图像共 100 种飞机型号,其中训练与验证图像共 6667 张,测试图像为 3333 张. Cars 数据库具有 196 个汽车类别共 16185 张图像,其中训练与验证图像共 8144 张,测试图像为 8041 张. 数据库实例如图 3 所示. 本文实验中均未采用边界框(Bounding Box)等额外标注信息. 实验使用 MatConvNet[20] 工具包完成.

CUB-200-2011　　FGVC-Aircraft　　Cars

图3　数据库示例图像

实验参数设置　核化双线性卷积网络的架构如图 1 所示. 本文选用在 ImageNet[2] 数据库上预训练的 VGG-M[21] 与 VGG-VD16[22] 网络模型作为基础骨干网络用于提取图像的特征图,并嵌入本文提出的核化双线性聚合模块在目标数据集上进行端到端的微调. 实验分别使用网络模型中 Conv5(VGG-M)与 Conv5_3(VGG-VD16)卷积层(不包含 ReLU[23] 非线性层)的输出作为一幅图像的特征图,其中特征图的通道数目均为 512. 本文采用与改进的双线性卷积网络[9] 相同的图像预处理方式:其中 CUB-200-2011 以及 Cars 数据库采用同样的图像预处理方式,即保持图像的长宽比,并将图像的短边缩放至 448 后中心裁剪出 448×448 的区域作为输

入图像;对于 FGVC-Aircraft 数据库,将图像缩放至 512×512 后中心裁剪出 448×448 的区域作为输入图像.输入图像经过卷积层后输出的特征图大小分别为 27×27×512(VGG-M)与 28×28×512(VGG-VD16),并通过核化双线性聚合模块得到图像的最终表达,其中矩阵幂正规化层[8,9]中参数 α 设为 0.5. 本文使用带动量的批处理随机梯度下降算法,动量设置为 0.9,权重衰减为 5×10^{-4},对于 VGG-M、VGG-VD16 网络训练时的批大小分别设置为 40 与 20,在三个细粒度数据库上以 8×10^{-4} 的学习率训练 30~40 个 epoch,数据增广方式为图像的随机水平翻转. 测试阶段,本文同样采用与改进的双线性卷积网络[9]相同的处理方式,即使用线性支持向量机(Support Vector Machine,SVM)替代核化双线性卷积网络中的 softmax 分类器对测试图像进行分类,并取该测试图像及其水平翻转图像的预测分数的均值作为最终分类结果.

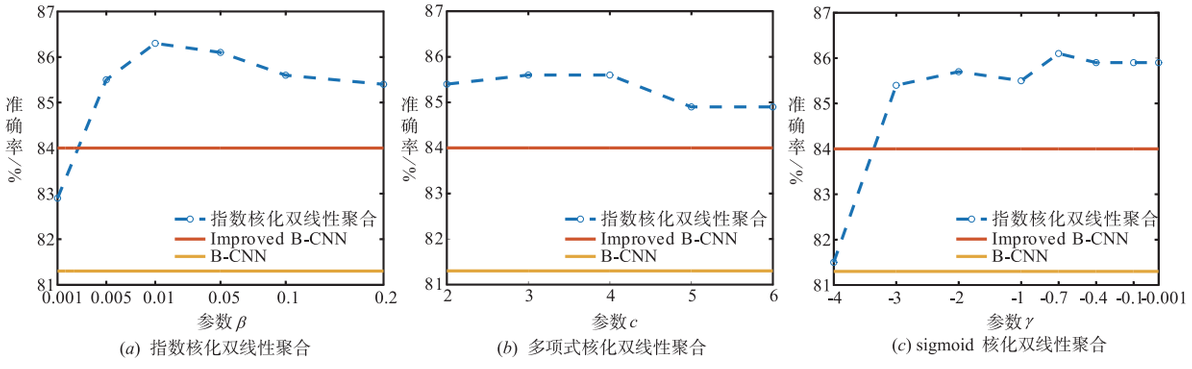
4.1 三种核化双线性聚合的评估

实验部分首先针对本文提出的三种核化双线性聚合方法进行参数评估,实验在 FGVC-Aircraft 数据库上进行,选用 VGG-M 网络模型作为基础骨干网络. 实验中,评估结果同时与本文相关的两个基线方法双线性卷积网络(B-CNN)[7]以及改进的双线性卷积网络(Improved B-CNN)[9]进行性能比较,基线方法同样采用 VGG-M 网络模型作为基础骨干网络. 考虑到文献[7]在 FGVC-Aircraft 上图像预处理方式与本文及文献[9]不一致,为公平比较,本文中 B-CNN 的实验结果均引自文献[9].

指数核化双线性聚合的参数评估 指数核化双线性聚合:$\boldsymbol{P}=[\exp(\beta\cdot\langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle)]_{d\times d}$,实验针对参数 β 进行评估. 实验结果如图 4(a)所示,其中参数 β 的取值依次为[0.001,0.005,0.01,0.05,0.1,0.2]. 当 β 取 0.001 至 0.01 时,网络的性能逐步提升,并当 β 等于 0.01 时,性能达到最优为 86.3%;当 β 为 0.2 时,性能下降为 85.4%. 图 4(a)中作为基线方法的双线性卷积网络与改进的双线性卷积网络的性能分别为 81.3% 与 84.0%. 当参数 β 为 0.01 时,网络的性能分别领先基线方法 5% 与 2.3%. 下述实验中,指数核化双线性卷积网络中的参数 β 固定为 0.01.

多项式核化双线性聚合的参数评估 多项式核化双线性聚合:$\boldsymbol{P}=[(\langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle+1)^c]_{d\times d}$,实验针对参数 c 即多项式的次数进行评估. 实验结果如图 4(b)所示,其中多项式次数 c 的取值分别为[2,3,4,5,6]. 当 c 为 2 时,网络的性能为 85.4%. c 取 3 时,网络性能提升至 85.6%. 进一步增加 c 至 4 时,性能仍为 85.6%. 而当 c 取 5 时,性能下降为 84.9%. 由上述实验结果可以得出,对于多项式核化双线性聚合,当 c 取 3 时,可以较好的建模特征图中通道间的高阶非线性关系. 然而高阶信息通常较难估计从而可靠性较差,因此当进一步增加 c 至 5 时,引入了更高阶的非线性关系,并同时可能带来信息的冗余,进而造成网络性能的下降. 下述实验中,多项式核化双线性聚合中参数 c 固定为 3.

sigmoid 核化双线性聚合的参数评估 sigmoid 核化双线性聚合:$\boldsymbol{P}=[\tanh(\theta\cdot\langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle+\gamma)]_{d\times d}$,实验主要针对核函数中位移参数 γ 进行评估,幅度调节参数 θ 固定为 1. 实验结果如图 4(c)所示,其中位移参数 γ 的取值依次为[−4,−3,−2,−1,−0.7,−0.4,−0.1,−0.001]. 当 γ 取 −4 至 −0.7 时,网络的性能逐渐提升,并当 γ 取 −0.7 时性能最优为 86.1%,此时网络的性能分别领先双线性卷积网络与改进的双线性卷积网络 4.8% 与 2.1%. 当 γ 大于 −0.7 时,网络性能下降,γ 为 −0.4 时,网络性能为 85.9%. 下述实验中,sigmoid 核化双线性卷积网络中位移参数 γ 固定为 −0.7,幅度调节参数 θ 固定为 1.



(a) 指数核化双线性聚合 (b) 多项式核化双线性聚合 (c) sigmoid 核化双线性聚合

图 4 三种核化双线性聚合的评估

三种核化双线性聚合的性能比较 三种核化双线性聚合方式在 FGVC-Aircraft 数据库上的性能比较如表 2 所示. 本文所提出的三种核化双线性聚合方法均领先基线方法并取得优异的性能,其中指数核化双线性聚

合方法性能为 86.3%,分别领先双线性卷积网络与改进的双线性卷积网络 5%与 2.3%.

表 2　三种核化双线性聚合在 FGVC-Aircraft 上性能比较

	骨干网络模型	方法	准确率/%
本文方法	VGG-M	指数核化双线性聚合	**86.3**
		多项式核化双线性聚合	85.6
		sigmoid 核化双线性聚合	86.1
基线方法	VGG-M	B-CNN	81.3
		Improved B-CNN	84.0

同时,本文对于三种核化双线性聚合相较于 Improved B-CNN 的性能提升进行显著性检验. 在显著性水平 $\alpha=0.05$ 的情况下,指数核化双线性聚合、多项式核化双线性聚合以及 sigmoid 核化双线性聚合在每类上的识别率相对 Improved B-CNN 方法分别提升 2%、1.1%、1.9%.

4.2　与当前方法的比较

本节实验进一步在三个细粒度图像数据库上对本文方法进行全面的评估,并与当前主流方法进行比较. 实验结果如表 3 所示,其中本文方法的设置如下:使用 VGG-VD16[22] 网络模型作为基础骨干网络,指数核化双线性聚合中参数 β 取 0.01,多项式核化双线性聚合中多项式次数 c 固定为 3,sigmoid 核化双线性聚合中幅度参数 θ 为 1,位移参数 γ 取 -0.7.

表 3　与当前方法的比较

准确率/%	骨干网络模型	CUB	Aircraft	Cars
指数核化双线性聚合	VGG-VD16	85.9	**91.3**	92.6
多项式核化双线性聚合	VGG-VD16	85.5	89.8	91.8
sigmoid 核化双线性聚合	VGG-VD16	**86.1**	90.8	**92.8**
B-CNN[7]	VGG-VD16	84.0	86.9	90.6
Improved B-CNN[9]	VGG-VD16	85.8	88.5	92.0
G^2DeNet[17]	VGG-VD16	**87.1**	89.0	92.5
KP[16]	VGG-VD16	86.2	86.9	92.4
MoNet[24]	VGG-VD16	86.4	89.3	91.8
CBP[14]	VGG-VD16	84.0	–	–
FBN[15]	VGG-VD16	82.9	–	–
LRBP[25]	VGG-VD16	84.2	87.3	90.9
RA-CNN[26]	VGG-VD19	85.3	–	92.5
BoostCNN[27]	B-CNN	86.2	88.5	92.1
ST-CNN[28]	BN-Inception	84.1	–	–

由表 3 实验结果所示,本文方法在三个细粒度图像识别数据库上均优于基线方法(B-CNN,Improved B-CNN),并在 FGVC-Aircraft 以及 Cars 数据库上取得了领先的性能. 指数核化双线性卷积网络在 FGVC-Aircraft 上取得 91.3%的准确率,分别领先 B-CNN 与 Improved B-CNN 方法 4.4%和 2.8%. Sigmoid 核化双线性卷积网络,在 Cars 上取得 92.8%的准确率,相较于 B-CNN 与 Improved B-CNN 方法分别提升 2.2%和 0.8%;并在 CUB-200-2011 上达到 86.1%的准确率,优于 B-CNN 方法 84.0%的准确率以及 Improved B-CNN 方法 85.8%的准确率. 表中 G^2DeNet、KP 以及 MoNet[24] 与 B-CNN 为同类方法,其中 G^2DeNet 在 CUB-200-2011 上取得最优性能为 87.1%. MoNet 通过引入子矩阵平方根结构层解决了 G^2DeNet 方法中无法使用低维近似的问题,在三个细粒度图像数据库上的准确率分别为:86.4%、89.3%与 91.8%. CBP、FBN 以及 LRBP[25] 均为双线性卷积网络的压缩算法. 由表 3 可知,LRBP 为三种方法中性能最优,该方法通过对协方差进行低秩近似进而减小了计算的复杂度,在三个细粒度图像数据库上分别取得了 84.2%、87.3%与 90.9%的性能,均落后于本文方法. RA-CNN[26] 使用 VGG-VD19 网络模型,并在卷积网络中引入注意力机制,通过递归的方式学习出更具判别力的感兴区域以及更强的区域特征表达,在 CUB-200-2011 与 Cars 上取得 85.3%与 92.5%的准确率. BoostCNN[27] 方法通过增强多个在不同尺度下训练的 B-CNN 网络模型,在 CUB-200-2011 上取得了 86.2%的准确率,然而该方法复杂度较高同时速度较慢. ST-CNN[28] 方法通过在卷积网络架构中引入空间变换模块从而实现数据的空间变换与对齐,在 CUB-200-2011 上取得了 84.1%的准确率.

5　结论

本文提出了一种核化双线性卷积网络模型,通过使用核函数的方式直接有效地建模特征图中通道间的非线性关系,并实现端到端的联合优化学习,得到更具判别力的图像表达. 实验表明本文所提出的核化双线性卷积网络在三个细粒度图像数据库上均取得了优异的性能. 在未来的工作中,可以将本文方法应用至更多计算机视觉任务中,同时进一步考虑将本文提出的核化双线性聚合应用于其他卷积网络架构中,如 ResNet[29]、Inception[30] 等.

参考文献

[1] KRIZHEVSKY A,SUTSKEVER I,HINTON G E. Imagenet classification with deep convolutional neural networks [A]. Advances in Neural Information Processing Systems

[C]. Lake Tahoe:NIPS Foundation,2012. 1097 – 1105.

[2] DENG J,DONG W,SOCHER R,et al. Imagenet:A large-scale hierarchical image database[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C]. Florida:IEEE Press,2009. 248 – 255.

[3] 柯圣财,赵永威,李弼程,等. 基于卷积神经网络和监督核哈希的图像检索方法[J]. 电子学报,2017,45(1):157 – 163.
KE Sheng-cai,ZHAO Yong-wei,LI Bi-cheng,et al. Image retrieval based on convolutional neural network and kernel-based supervised hashing [J]. Acta Electronica Sinica, 2017,45(1):157 – 163. (in Chinese)

[4] 王泽宇,吴艳霞,张国印,等. 基于空间结构化推理深度融合网络的 RGB-D 场景解析[J]. 电子学报,2018,46(5):1253 – 1258.
WANG Ze-yu,WU Yan-xia,ZHANG Guo-yin,et al. RGB-D scene parsing based on spatial structured inference deep fusion networks[J]. Acta Electronica Sinica,2018,46(5):1253 – 1258. (in Chinese)

[5] 李康,李亚敏,胡学敏,等. 基于卷积神经网络的鲁棒高精度目标跟踪算法[J]. 电子学报,2018,46(9):2087 – 2093.
LI Kang,LI Ya-min,HU Xue-min,et al. Robust and accurate object tracking algorithm based on convolutional neural network[J]. Acta Electronica Sinica,2018,46(9):2087 – 2093. (in Chinese)

[6] 邹承明,罗莹,徐晓龙. 基于多特征组合的细粒度图像分类方法[J]. 计算机应用,2018,38(7):1853 – 1856,1861.
ZOU Cheng-ming,LUO Ying,XU Xiao-long. Fine-grained image classification method based on multi-feature combination [J]. Journal of Computer Applications, 2018, 38(7):1853 – 1856,1861. (in Chinese)

[7] LIN T Y,ROYCHOWDHURY A,MAJI S. Bilinear CNN models for fine-grained visual recognition[A]. Proceedings of IEEE International Conference on Computer Vision [C]. Santiago:IEEE Press,2015. 1449 – 1457.

[8] LI P,XIE J,WANG Q,et al. Is second-order information helpful for large-scale visual recognition[A]. Proceedings of IEEE International Conference on Computer Vision [C]. Venice:IEEE Press,2017. 2070 – 2078.

[9] LIN T Y,MAJI S. Improved bilinear pooling with CNNs [A]. British Machine Vision Conference [C]. London: British Machine Vision Association,2017. 1 – 12.

[10] IOFFE S,SZEGEDY C. Batch normalization:Accelerating deep network training by reducing internal covariate shift [A]. International Conference on Machine Learning[C]. Lille:ACM,2015. 448 – 456.

[11] MAJI S,RAHTU E,KANNALA J,et al. Fine-Grained Visual Classification of Aircraft [OL]. https://arxiv.org/abs/1306.5151,2013.

[12] WAH C,BRANSON S,WELINDER P,et al. The Caltech-Ucsd Birds-200-2011 Dataset[R]. Technical report,Caltech. 2011.

[13] KRAUSE J,STARK M,DENG J,et al. 3D object representations for fine-grained categorization[A]. Proceedings of IEEE International Conference on Computer Vision Workshops[C]. Portland:IEEE Press,2013. 554-561.

[14] GAO Y,BEIJBOM O,ZHANG N,et al. Compact bilinear pooling[A]. Proceedings of IEEE Conference on Computer Vision and attern Recognition[C]. Las Vegas:IEEE Press,2016. 317 – 326

[15] LI Y,WANG N,LIU J,et al. Factorized bilinear models for image recognition[A]. Proceedings of IEEE International Conference on Computer Vision[C]. Venice:IEEE Press,2017. 2098 – 2106.

[16] CUI Y,ZHOU F,WANG J,et al. Kernel pooling for convolutional neural networks [A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C]. Honolulu:IEEE Press,2017. 3049 – 3058.

[17] WANG Q,LI P,ZHANG L. G2DeNet:Global Gaussian distribution embedding network and its application to visual recognition[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition[C]. Honolulu: IEEE 2017. 2730 – 2739.

[18] IONESCU C,VANTZOS O,SMINCHISESCU C. Training deep networks with structured layers by matrix backpropagation[OL]. https://arxiv.org/abs/1509.07838,2015.

[19] LIN H T,LIN C J. A Study on Sigmoid Kernels for SVM and the Training of Non-PSD Kernels by SMO-Type Methods[R]. Technical Report,Nat'l Taiwan Univ,2003.

[20] VEDALDI A,LENC K. Matconvnet:convolutional neural networks for matlab[A]. ACM International Conference on Multimedia[C]. Brisbane:ACM,2015. 689 – 692.

[21] CHATFIELD K,SIMONYAN K,VEDALDI A,et al. Return of the devil in the details:Delving deep into convolutional nets[A]. British Machine Vision Conference[C]. Nottingham:British Machine Vision Association,2014. 1 – 12.

[22] SIMONYAN K,ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[A]. International Conference on Learning Representations[C]. San Diego,2015. 1 – 14.

[23] NAIR V,HINTON G E. Rectified linear units improve restricted boltzmann machines[A]. International Conference on Machine Learning[C]. Haifa:ACM,2010. 807 – 814.

[24] GOU M,XIONG F,CAMPS O,et al. MoNet:Moments embedding network[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C]. Salt

Lake City:IEEE Press,2018. 3175 – 3183.
[25] KONG S, FOWLKES C. Low-rank bilinear pooling for fine-grained classification[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C]. Honolulu:IEEE Press,2017. 7025 – 7034.
[26] FU J,ZHENG H,MEI T. Look closer to see better:Recurrent attention convolutional neural network for fine-grained image recognition[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition[C]. Honolulu:IEEE Press,2017. 4476 – 4484.
[27] MOGHIMI M, BELONGIE S J, SABERIAN M J, et al. Boosted convolutional neural networks[A]. British Machine Vision Conference[C]. York:British Machine Vision Association,2016. 1 – 13.
[28] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial transformer networks[A]. Advances in neural information processing systems[C]. Montreal:MIT Press, 2015. 2017 – 2025
[29] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition[C]. Las Vegas:IEEE Press,2016. 770 – 778.
[30] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition[C]. Boston: IEEE Press,2015. 1 – 9.

作者简介

葛疏雨　男,1994 年 2 月出生于安徽宿州. 大连理工大学信息与通信工程学院硕士研究生. 主要研究方向为计算机视觉、深度学习.
E-mail:gsy@ mail. dlut. edu. cn

高子淋　女,1995 年 6 月出生于黑龙江哈尔滨. 大连理工大学信息与通信工程学院硕士研究生. 主要研究方向为深度学习、计算机视觉.
E-mail:gzl@ mail. dlut. edu. cn

张冰冰　女,1990 年 5 月生于辽宁沈阳. 大连理工大学信息与通信工程学院博士研究生. 主要研究方向为深度学习、视频行为识别.
E-mail:icyzhang@ mail. dlut. edu. cn

李培华(通信作者)　男,1971 年 7 月出生于黑龙江安达. 2003 年获得哈尔滨工业大学的计算机应用技术博士学位. 现为大连理工大学信息与通信工程学院教授、博士生导师,主要研究方向为统计学习、计算机视觉与模式识别.
E-mail:peihuali@ dlut. edu. cn