

空间金字塔与局部感受野相结合的相关熵极限学习机

刘彬^{*①} 刘静^① 吴超^② 杨有恒^①

^①(燕山大学电气工程学院 秦皇岛 066004)

^②(燕山大学信息科学与工程学院 秦皇岛 066004)

摘要: 针对空间金字塔词袋模型中空间特征分布信息利用效率低, 各类特征融合不充分的问题, 该文提出空间金字塔与局部感受野相结合的相关熵极限学习机(SR-CELM)。在特征提取部分, 利用多尺度局部感受野对生成的多层级的字典特征分布图进行卷积, 并引入局部位置特征和全局轮廓特征。在特征分类部分, 提出一种新的网络以融合各部分特征。同时在传统极限学习机训练方法的基础上利用相关熵准则构建判别性约束, 推导出权重更新公式以求解网络的输出权重。为验证SR-CELM的有效性, 该文分别在数据库Caltech 101, MSRC和15 Scene上进行实验。实验表明SR-CELM能够充分利用特征中可辨识信息, 提高分类正确率。

关键词: 图像分类; 词袋模型; 局部感受野; 极限学习机; 相关熵

中图分类号: TN911.73; TP391

文献标识码: A

文章编号: 1009-5896(2021)08-2343-09

DOI: 10.11999/JEIT200562

Correntropy Extreme Learning Machine Based on Spatial Pyramid Matching and Local Receptive Field

LIU Bin^① LIU Jing^① WU Chao^② YANG Youheng^①

^①(School of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China)

^②(School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China)

Abstract: Considering the problems of inefficient use of spatial information between features and inadequate fusion of different features, a Correntropy Extreme Learning Machine based on Spatial pyramid matching and local Receptive field(SR-CELM) is proposed. In feature extraction part, multi-scale local receptive fields are used to convolve the generated multi-level dictionary feature distribution map, and local position features and global contour features are introduced. In feature classification part, a new network is proposed to fuse the features of each part. Based on the traditional extreme learning machine training method, a discriminative constraint is constructed by using the relevant entropy criterion, and the weight update formula is used to solve the output weight of the new network. In order to verify the effectiveness of the SR-CELM, experiments are performed on the databases Caltech 101, MSRC and 15 Scene. The experiments show that SR-CELM can make full use of the identifiable information in the features and improve the classification accuracy.

Key words: Image classification; Bag of words model; Local receptive field; Extreme learning machine; Correntropy

1 引言

近年来, 随着图像分类领域的飞速发展, 词袋模型^[1]作为一种简单有效的图像分类方法得到了广泛的关注和研究。该方法将图像特征量化为视觉单词, 通过统计出视觉单词的频率直方图完成图像表

示。然而, 仅统计单词特征出现次数来表示图像, 使得特征分布中的空间信息未被有效利用。因此, Lazebnik等人^[2]提出了空间金字塔词袋模型(Spatial Pyramid Matching, SPM), 将图像划分成不同尺寸的子区域分别进行直方图统计, 有效提高了词袋模型分类能力, 因此很多算法均基于SPM进行改进。

为提高分类效率, Li等人^[3]在SPM中引入极限学习机(Extreme Learning Machine, ELM)^[4], 利用其对提取出的SPM特征进行分类。由于ELM具有良好的特征, 因而广泛应用于分类、预测和回归等领域^[5-7]。然而传统ELM只能利用已提取好的特

收稿日期: 2020-06-29; 改回日期: 2020-12-05; 网络出版: 2020-12-16

*通信作者: 刘彬 liubin311@163.com

基金项目: 河北省自然科学基金(F2019203320, E2018203398)

Foundation Items: The Natural Science Foundation of Hebei Province (F2019203320, E2018203398)

征向量来训练网络,无法考虑特征的空间分布信息。因此, Huang等人^[8]提出局部感受野极限学习机(Local Receptive Fields based Extreme Learning Machine, ELM-LRF), 通过将局部感受野与ELM相结合, 使得网络可以利用空间信息以增加特征辨识度。然而对于复杂图像, 单层网络提取出的特征信息表达能力有限, 无法进一步提高网络分类能力。随着计算机性能的不不断提升, 研究学者提出了 AlexNet, GoogleNet等深度学习网络结构^[9-12], 然而其训练网络所需的计算量远超传统分类算法。而 ELM-LRF网络简单、易于训练、泛化性强, 在处理一些分类任务时同样能够获得较高精度。

由于传统ELM训练方法通常使用的均方根误差(Mean Square Error, MSE)准则对训练数据中的离群点和脉冲噪声较为敏感。因此, Xing等人^[13]引入相关熵度量作为网络损失函数, 并利用半2次优化求解满足相关熵准则的ELM, 提高了算法在噪声破坏和遮挡条件下的识别精度。吴超等人^[14]利用拉格朗日乘子法求解基于相关熵的ELM(Correntropy-based Fusion Extreme Learning Machine, CF-ELM), 并将其用于训练网络中各层权重以提高网络分类性能。然而这些基于相关熵准则的改进算法没有考虑向量之间的角度问题, 降低了网络的通用逼近能力。当网络的实际输出无限接近于期望输出时, 不仅向量模长应无限趋近, 两者所成角度也应为最小。

针对以上问题, 本文提出空间金字塔与局部感受野相结合的相关熵极限学习机(Correntropy Extreme Learning Machine based on Spatial pyramid matching and local Receptive field, SR-CELM)。

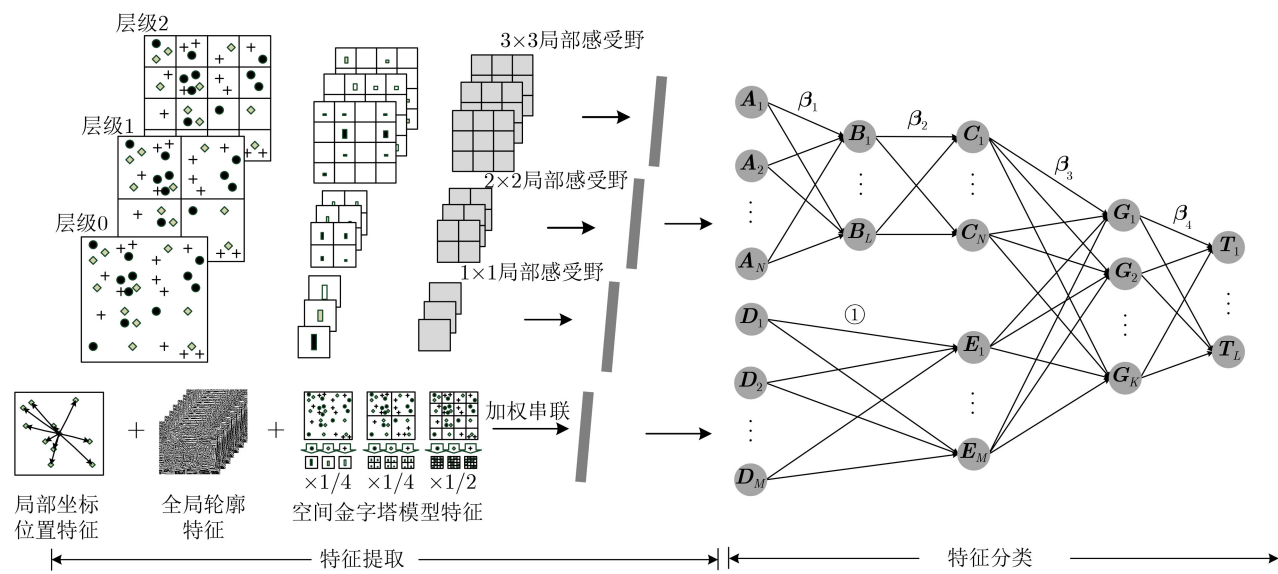
在特征提取部分, 将SPM特征与ELM-LRF充分结合, 提取出空间中的局部特征分布信息。按照字典顺序重新排列空间金字塔区域内的各个字典特征, 并统计其在每个子区域内的出现频率, 利用多尺度局部感受野提取各层级字典特征分布图的卷积特征。同时, 为提高特征多样性, 引入局部位置特征和全局轮廓特征。在特征分类部分, 提出一种新的网络对各部分特征进行编码和融合。为提高网络的鲁棒性与分类性能, 构建出基于相关熵准则的模长差值与角度约束, 推导出相关熵循环更新公式以训练网络的输出权重。最后在Caltech 101, MSRC和 15 Scene数据库上进行实验, 以验证SR-CELM的有效性。

2 空间金字塔与感受野相结合的相关熵极限学习机

为充分提取与利用图像中的可辨识特征信息, 本文提出了空间金字塔与感受野相结合的相关熵极限学习机(SR-CELM), SR-CELM主要由特征提取和特征分类两部分组成, 其整体结构如图1所示。

2.1 特征提取

为提高各子区域内特征的空间分布信息利用率, 本文将SPM特征重新排列, 如图2所示。对每个子区域内字典特征出现次数分别进行统计, 分别生成其对应的空间分布频率直方图。由于每个金字塔层级下特征分布图的尺寸不一致, 从左至右分别利用 3×3 , 2×2 和 1×1 的局部感受野对图2中3种尺寸的特征分布图进行卷积, 并将卷积后的特征矩阵串联, 得到特征矩阵 $A \in R^{N \times n}$ 。同时, 引入文献^[15]中的SPM特征、局部位置特征和全局轮廓特



① 交叉直方核映射(Histogram Intersection Kernel, HIK)

图1 SR-CELM结构图

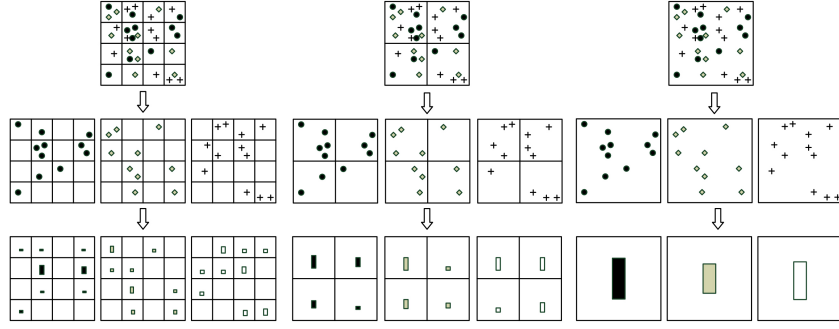


图2 不同子区域特征频率统计图

征, 并将3种特征以相应的系数串联, 得到特征组合 $D \in \mathbf{R}^{M \times n}$ 。

2.2 特征分类

为了充分融合与利用2.1节中的3种特征, 本文提出图1所示的网络结构对特征进行编码与分类。首先, 利用 β_1 对特征 A 进行编码, 令 A 作为ELM训练方法中的隐藏层输出, 按照式(1)求解权重矩阵 β_1 , 则编码后的特征为 $B = A\beta_1$, 其中 $B = [b_1, b_2, \dots, b_N] \in \mathbf{R}^{N \times l}$

$$\beta_1 = \left(A^T A + \frac{I}{K_1} \right)^{-1} A^T T \quad (1)$$

其次, 利用 β_2 对特征 B 进行加权映射。令 B 作为训练方法中的隐藏层输出, 按照式(2)求解出权重 β_2 , 编码后的特征 $C = B\beta_2$ 。为提高整体特征的表达能力, 利用交叉直方核函数对 $D \in \mathbf{R}^{N \times n}$ 进行核映射。假设输入特征为 $D = [d_1, d_2, \dots, d_N]^T \in \mathbf{R}^{N \times n}$, 对应的图像标签为 $T \in \mathbf{R}^{N \times l}$, 利用交叉直方核函数将 D 映射为 $E \in \mathbf{R}^{N \times N}$, 并将此特征与编码后的特征 C 串联以得到特征 F

$$\beta_2 = \left(B^T B + \frac{I}{K_2} \right)^{-1} B^T A \quad (2)$$

为减小串联特征维数, 对特征 F 进行压缩映射。令 F 作为隐藏层输出, 以特征 B 作为目标输出训练压缩权重 β_3 , 如式(3)所示, 用 β_3 将 F 降维成 G , 则 $G = F\beta_3$

$$\beta_3 = \left(F^T F + \frac{I}{K_3} \right)^{-1} F^T B \quad (3)$$

最后, 将编码后的特征矩阵 G 作为网络的隐藏层输出, 训练出最终分类权重矩阵 β_4 , 如式(4)所示

$$\beta_4 = \left(G^T G + \frac{I}{K_4} \right)^{-1} G^T T \quad (4)$$

由于传统ELM-LRF在训练时均使用MSE作为损失函数, 利用式(4)训练的各层权重会使网络整体易受噪声干扰。因此本文改进相关熵准则的判

别性约束, 推导出相关熵循环更新公式以求解输出权重。

2.3 相关熵准则输出权重的求解方法

首先, 利用模长差值和角度约束构建出基于相关熵的损失函数。假设输入 N 个 n 维特征 $X \in \mathbf{R}^{N \times n}$, 相对应的标签为 $T \in \mathbf{R}^{N \times l}$ 。随机设置ELM-LRF输入层与隐藏层之间的输入权重 $w \in \mathbf{R}^{n \times n_h}$, 隐藏层偏置 $b \in \mathbf{R}^{N \times n_h}$, 则隐藏层输出为 $H = g(Xw + b) \in \mathbf{R}^{N \times n_h}$, n_h 代表隐藏层特征的维数, $g(\cdot)$ 代表激活函数。本文利用相关熵准则构造出模长差值与夹角约束, 使得网络实际输出与期望输出 t_i 之间的向量模长与夹角均同时为0。为便于网络权重求解, 将求最大值改变为取其相反数的最小值, 由此可得到

$$\left. \begin{aligned} \min L &= \frac{1}{2} \|\beta\|^2 + \frac{1}{2} K \sum_{i=1}^N \xi_i^2 + \frac{1}{2} P \sum_{i=1}^N \Phi_i^2 \\ \text{s.t.} \exp \left(-\frac{(\|t_i\| - \|h_i \beta\|)^2}{2\sigma^2} \right) &= 1 - \xi_i \\ \exp \left(-\frac{1}{2\sigma^2} \left(1 - \frac{h_i \beta \cdot t_i^T}{\|h_i \beta\| \|t_i\|} \right) \right) &= 1 - \Phi_i \end{aligned} \right\} \quad (5)$$

其中, β 为网络输出权重, K, P 均为惩罚项系数, ξ_i 为相关熵误差, Φ_i 为角度误差, σ 为高斯核函数中的尺度因子。由等式约束优化(Karush Kuhn Tucker, KKT)条件可知, 式(5)的对偶优化式为

$$\begin{aligned} L &= \frac{1}{2} \|\beta\|^2 + \frac{1}{2} K \sum_{i=1}^N \xi_i^2 + \frac{1}{2} P \sum_{i=1}^N \Phi_i^2 \\ &\quad - \sum_{i=1}^N \alpha_i \left(\exp \left(-\frac{(\|t_i\| - \|h_i \beta\|)^2}{2\sigma^2} \right) - 1 + \xi_i \right) \\ &\quad - \sum_{i=1}^N \mu_i \left(\exp \left(-\frac{1}{2\sigma^2} \left(1 - \frac{h_i \beta \cdot t_i^T}{\|h_i \beta\| \|t_i\|} \right) \right) - 1 + \Phi_i \right) \end{aligned} \quad (6)$$

对式(6)中 $\beta, \xi_i, \Phi_i, \alpha_i$ 和 μ_i 变量求偏导, 使各变量偏导等于0, 并令 Q, R, Z 为对角矩阵, 化简可得

$$\beta - \mathbf{K}\mathbf{H}^T\mathbf{Q}\mathbf{H}\beta - \mathbf{P}\mathbf{H}^T\mathbf{R}\mathbf{T} + \mathbf{P}\mathbf{H}^T\mathbf{Z}\mathbf{H}\beta = 0 \quad (7)$$

当 $n \leq N$ 时

$$\beta = \left(\frac{\mathbf{I}}{P} + \mathbf{H}^T \left(\mathbf{Z} - \frac{K}{P} \mathbf{Q} \right) \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{R} \mathbf{T} \quad (8)$$

将 $\beta = \mathbf{H}^T \boldsymbol{\eta}$ 代入式(7)中, 当 $n > N$ 时可求得

$$\beta = \mathbf{H}^T \left(\frac{\mathbf{I}}{P} + \left(\mathbf{Z} - \frac{K}{P} \mathbf{Q} \right) \mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{R} \mathbf{T} \quad (9)$$

其中, 利用循环迭代的方法求解对角矩阵 \mathbf{Q} , \mathbf{R} , \mathbf{Z} 和 β 。当第1次迭代时, 令 $\mathbf{Q}_{ii} = 1$, $\mathbf{R}_{ii} = 1$, $\mathbf{Z}_{ii} = 1$, 由于当 $x \rightarrow 0$ 时, $e^x - 1 \sim x$, 可对 $\mathbf{Q}_{ii}^{\rho-1}$, $\mathbf{R}_{ii}^{\rho-1}$ 和 $\mathbf{Z}_{ii}^{\rho-1}$ 进行等价变化, 即在第 ρ 次迭代时可得到

$$\begin{aligned} \mathbf{Q}_{ii}^{\rho-1} &= \frac{1}{\sigma^2} \frac{(\|t_i\| - \|\mathbf{h}_i \boldsymbol{\beta}^{\rho-1}\|)^2}{2\sigma^2} \\ &\cdot \exp\left(-\frac{(\|t_i\| - \|\mathbf{h}_i \boldsymbol{\beta}^{\rho-1}\|)^2}{2\sigma^2}\right) \\ &\cdot \left(\frac{\|t_i\|}{\|\mathbf{h}_i \boldsymbol{\beta}^{\rho-1}\|} - 1 \right) \end{aligned} \quad (10)$$

$$\begin{aligned} \mathbf{R}_{ii}^{\rho-1} &= \frac{1}{4\sigma^4} \left(1 - \frac{\mathbf{h}_i \boldsymbol{\beta}^{\rho-1} \cdot \mathbf{t}_i^T}{\|\mathbf{h}_i \boldsymbol{\beta}^{\rho-1}\| \|\mathbf{t}_i\|} \right) \\ &\cdot \exp\left(-\frac{1}{2\sigma^2} \left(1 - \frac{\mathbf{h}_i \boldsymbol{\beta}^{\rho-1} \cdot \mathbf{t}_i^T}{\|\mathbf{h}_i \boldsymbol{\beta}^{\rho-1}\| \|\mathbf{t}_i\|} \right)\right) \\ &\cdot \frac{1}{\|\mathbf{h}_i \boldsymbol{\beta}^{\rho-1}\| \|\mathbf{t}_i\|} \end{aligned} \quad (11)$$

$$\begin{aligned} \mathbf{Z}_{ii}^{\rho-1} &= \frac{1}{4\sigma^4} \left(1 - \frac{\mathbf{h}_i \boldsymbol{\beta}^{\rho-1} \cdot \mathbf{t}_i^T}{\|\mathbf{h}_i \boldsymbol{\beta}^{\rho-1}\| \|\mathbf{t}_i\|} \right) \\ &\cdot \exp\left(-\frac{1}{2\sigma^2} \left(1 - \frac{\mathbf{h}_i \boldsymbol{\beta}^{\rho-1} \cdot \mathbf{t}_i^T}{\|\mathbf{h}_i \boldsymbol{\beta}^{\rho-1}\| \|\mathbf{t}_i\|} \right)\right) \\ &\cdot \frac{\mathbf{h}_i \boldsymbol{\beta}^{\rho-1} \cdot \mathbf{t}_i^T}{\|\mathbf{h}_i \boldsymbol{\beta}^{\rho-1}\|^3 \|\mathbf{t}_i\|} \end{aligned} \quad (12)$$

最后, $\boldsymbol{\beta}^\rho$ 可由式(13)求出

$$\boldsymbol{\beta}^\rho = \begin{cases} \left(\frac{\mathbf{I}}{P} + \mathbf{H}^T \left(\mathbf{Z}^{\rho-1} - \frac{K}{P} \mathbf{Q}^{\rho-1} \right) \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{R}^{\rho-1} \mathbf{T}, & n_h \leq N \\ \mathbf{H}^T \left(\frac{\mathbf{I}}{P} + \left(\mathbf{Z}^{\rho-1} - \frac{K}{P} \mathbf{Q}^{\rho-1} \right) \mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{R}^{\rho-1} \mathbf{T}, & n_h > N \end{cases} \quad (13)$$

由式(13)可求解出网络输出权重, 并将其更新公式应用于2.2节网络中各层权重的训练。在训练阶段, 利用式(10)–式(13)对 \mathbf{A} 编码, 将 \mathbf{A} 作为 \mathbf{H} 代入式(13)中, 得到编码权重 $\boldsymbol{\beta}_1$, 编码后的特征为 $\mathbf{B} = \mathbf{A}\boldsymbol{\beta}_1$ 。利用权重矩阵 $\boldsymbol{\beta}_2$ 将特征 \mathbf{B} 加权映射, 将式(10)–式(12)中的 \mathbf{t}_i 变为 \mathbf{x}_i , 式(13)中的 \mathbf{T} 变为 \mathbf{A} ,

可得到特征 $\mathbf{C} = \mathbf{B}\boldsymbol{\beta}_2$ 。将各类特征以特定系数串联为特征 \mathbf{D} , 将特征 \mathbf{D} 映射为特征 $\mathbf{E} = k(\mathbf{D}^{\text{train}}, \mathbf{D})$ 。将 \mathbf{C} 与 \mathbf{E} 串联生成特征矩阵 \mathbf{F} , 将 \mathbf{F} 代替式(13)中的 \mathbf{H} , \mathbf{B} 代替式(13)中的 \mathbf{T} , 求解出网络压缩权重 $\boldsymbol{\beta}_3$, 生成压缩矩阵 $\mathbf{G} = \mathbf{F}\boldsymbol{\beta}_3$ 。最后以 \mathbf{G} 替换式中的 \mathbf{H} , 利用式(13)求解网络分类权重 $\boldsymbol{\beta}_4$ 。

在测试阶段, 输入SPM特征经过卷积后生成 \mathbf{A}^{test} , 则由训练出的权重可得到 $\mathbf{B}^{\text{test}} = \mathbf{A}^{\text{test}}\boldsymbol{\beta}_1$, $\mathbf{C}^{\text{test}} = \mathbf{B}^{\text{test}}\boldsymbol{\beta}_2$, $\mathbf{E}^{\text{test}} = k(\mathbf{D}^{\text{test}}, \mathbf{D})$, $\mathbf{G}^{\text{test}} = \mathbf{F}^{\text{test}}\boldsymbol{\beta}_3$, $\mathbf{T}^{\text{test}} = \mathbf{G}^{\text{test}}\boldsymbol{\beta}_4$ 。

3 实验结果与分析

为验证所提方法的有效性, 本文从 Caltech 101, MSRC 和 15 Scene 数据库中分别提取出图像的 SPM 特征、局部位置特征和全局轮廓特征, 将特征组合后输入 SR-CELM 中。为公平比较, 所有数据库上的实验均参照文献[8]的实验步骤进行。

在特征分类部分各层权重的训练过程共需要 5 个参数, 即惩罚项系数 K 与 P , 循环次数, 局部感受野个数和相关熵参数 $d = 2\sigma^2$ 。本文在各数据集上分别进行大量实验, 确定最优参数组合。本文的实验环境为: Intel Xeon Silver 4110 @2.1 GHz 8核CPU, 64 GB RAM, MATLAB 2014a。

3.1 Caltech 101 数据库

本节使用 Caltech 101 数据库中 101 类物体图像进行实验。选取训练、测试图像方式和数量与文献[14]一致, 共使用 5948 幅图像进行实验。

针对 Caltech 101 数据库, 首先需确定最优的参数组合(惩罚项 K 与 P , 相关熵参数 d , 循环次数和感受野个数)。在实验中, K 和 P 的取值范围均为 $\{2^{-10}, 2^{-9}, 2^{-8}, \dots, 2^8, 2^9, 2^{10}\}$, d 的取值范围为 $\{0.1, 0.2, \dots, 0.9, 1\}$, 循环次数的取值范围为 $\{2, 4, 6, 8, 10, 12\}$, 其余参数设置与文献[8]一致。

首先对参数 d , K 和 P 进行优化, 当感受野个数为 1, 任意固定迭代次数为 7, 惩罚项 P 为 2^{-1} , K 和 d 变化时 SR-CELM 正确率变化曲面如图 3(a) 所示。在图 3(a) 中, K 为 2^{-10} 可获得最高正确率, 将其固定并调节惩罚项 P , P 和 d 变化时的正确率变化如图 3(b) 所示。由此可得到图 3(b) 中最优参数组合, 即 d , K 和 P 分别为 0.5, 2^{-10} 和 2^{-5} 。

在此基础上改变循环迭代次数和感受野个数所得网络分类正确率如图 4 所示。由折线图可知, 当字典维数为 400 维, 感受野个数为 21, 迭代次数为 8 时, SR-CELM 能达到最高正确率 83.72%。当字典维数为 600 维, 感受野个数为 24, 迭代次数为 9 时, 能够达到最高分类正确率 84.13%。

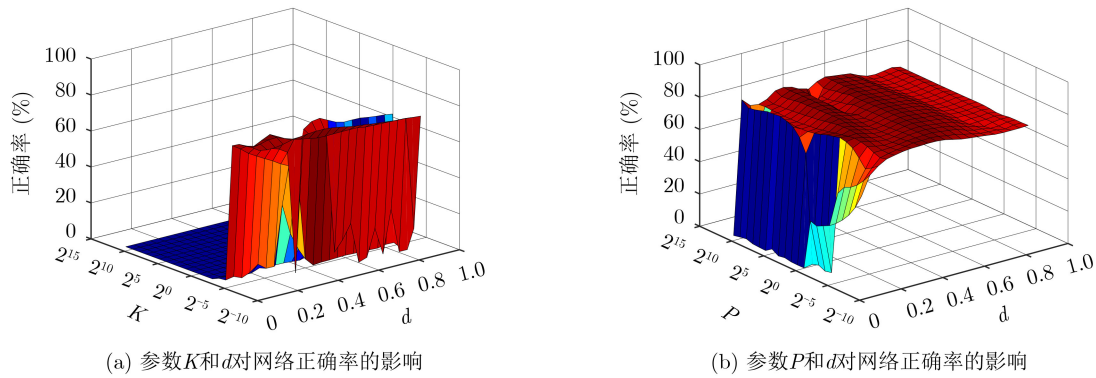


图3 不同参数变化时网络正确率曲面图

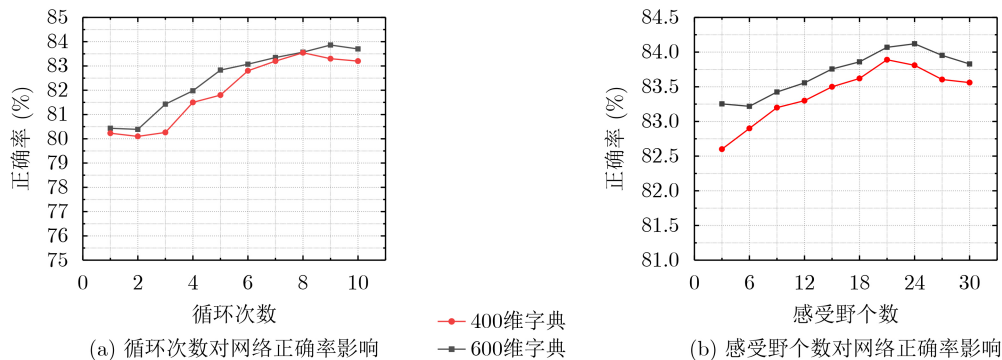


图4 不同字典维数下循环次数和感受野个数对网络正确率影响折线图

Caltech 101数据库中部分图像背景嘈杂，将图像中的嘈杂背景视为干扰噪声，选取几类背景嘈杂图像进行测试以验证网络抗干扰能力。将SR-CELM、文献[15]、CF-ELM和多尺度局部感受野极限学习机(Extreme Learning Machine with Multi-Scale Local Receptive Fields, ELM-MSLRF)^[16]的分类正确率进行比较，由图5中不同方法正确率对比可知，在嘈杂背景噪声干扰下SR-CELM具有更好的抗干扰能力。

表1列出了不同方法在Caltech 101数据库上的训练时间和正确率。由表1可知，与传统方法相比，SR-CELM分类正确率分别比SPM，支持向量机(Support Vector Machine, SVM)，核极限学习机(Kernel Extreme Learning Machine, KELM)，文献[15]和表示级特征融合的极限学习机(Fusion Extreme Learning Machine, F-ELM)高19.59%，5.79%，4.88%，5.79%和3.53%。SR-CELM因其编码特征的网络结构，正确率比文献[17]高0.23%。与深度学习网络AlexNet, GoogleNet相比，SR-

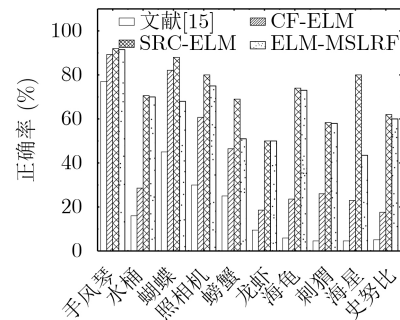


图5 4种方法对Caltech 101数据库中具有嘈杂背景图像的正确率

CELM以相对更简单、层数更少的网络获得较高分类正确率，由此可证明在物体分类的数据集上，SR-CELM能够有效利用图像中蕴含的各部分特征，提高分类正确率。

3.2 MSRC数据库

本节选择MSRC中的18个类别和MSRC-21数据库中的3个类别，共21类物体图像进行实验。选取训练、测试图像方式和数量与文献[14]一致，共使用1260幅图像进行实验。将提取出的SPM特

表1 Caltech 101正确率(%)与训练时间(s)

方法	SPM	GoogleNet ^[11]	文献[18]	文献[19]	AlexNet ^[9]	SVM ^[14]	KELM ^[14]	文献[15]	F-ELM ^[14]	CF-ELM	文献[17]	SR-CELM
字典维数	400	-	2048	2048	-	400	400	400	400	400	600	400
训练时间	13560	-	-	-	-	540.30	538.56	-	539.77	569.80	-	586.58
正确率	64.13	73.21	73.44	74.90	75.91	77.93	78.84	77.93	80.19	83.65	83.90	83.72

征、局部位置特征和全局轮廓特征按照文献[15]中 [1, 12, 0.2]的系数进行加权串联并输入网络。

当感受野个数为1, 任意固定迭代次数为6, 惩罚项 P 为 2^{-6} , K 和 d 变化时, 网络正确率变化由图6(a)所示。选取并固定此时最高正确率所对应的参数 K , 调节 P 和 d , 网络正确率变化曲面图如图6(b)所示。当参数组合 d , K 和 P 值分别为0.6, 2^{-8} 和 2^{-1} 时, 网络可达到最高正确率为95.82%。调整循环迭代次数和感受野个数, 可得到如图7所示折线图。在字典维数为400维, 迭代次数为7, 感受野个数为18时, 网络最高正确率为96.35%。当字典维数为600维, 迭代次数为8, 感受野个数为15时, 网络能

够达到最高正确率96.83%。

与3.1节相同, 选取MSRC数据库中具有嘈杂背景的5类图像进行实验。SR-CELM、文献[15]、CF-ELM和ELM-MSLRF的正确率对比如图8所示。由图8可知, SR-CELM在各图像小类的分类正确率高于其余3种方法, 证明SR-CELM在物体图像分类中能够有效减少嘈杂噪声对分类结果的影响。

表2为不同方法在MSRC数据库上的训练时间和正确率。与其余改进方法相比, 在字典维数相同时, SR-CELM分类正确率分别比文献[15], SVM和F-ELM高出2.38%, 2.22%和2.86%, 具有较高分类精度。

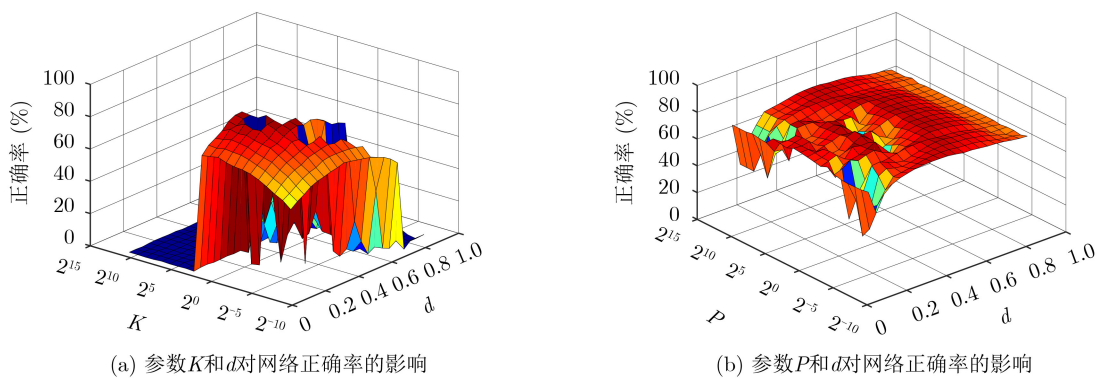


图6 不同参数变化时网络正确率曲面图

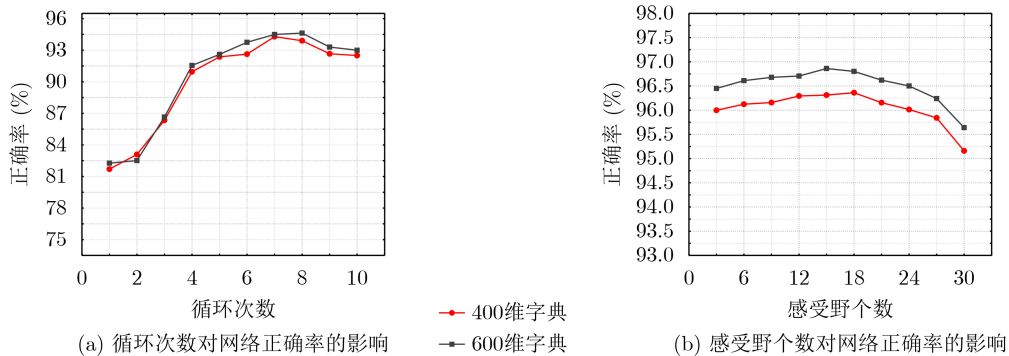


图7 不同字典维数下循环次数和感受野个数对网络正确率影响折线图

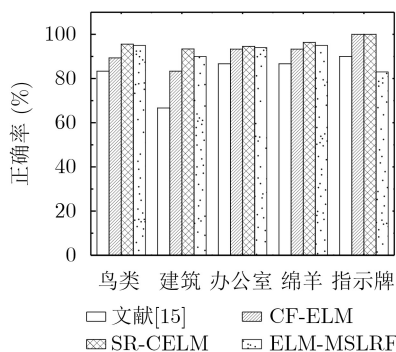


图8 4种方法对MSRC数据库中具有嘈杂背景图像的正确率

3.3 15 Scene数据库

本节利用15 Scene数据库测试SR-CELM对场景图像的分类能力, 选取训练、测试图像方式和数量与文献[14]一致, 总共使用4485幅图像进行实验。将提取出的SPM卷积特征、局部位置特征和全局轮廓特征串联, 并将这3部分特征按照[1, 0.15, 0.35]进行系数加权并输入网络。

当感受野个数为1, 迭代次数为6, 任意固定惩罚项 P 为 2^{-6} , K 和 d 变化时, SR-CELM的正确率如图9(a)所示。将 K 固定为图中最高正确率所对应的

参数值, 并调节惩罚项 P , 当 P 和 d 变化时, 网络正确率变化曲面图如图9(b)所示。当 d , K 和 P 分别为0.4, 2^{-5} 和 2^{-4} 时, 网络能达到最高正确率85.86%。

为确定循环迭代次数和感受野个数对网络正确率的影响, 分别对两个参数进行实验, 得到实验结果如图10所示。当字典维数为400维, 循环次数为6, 感受野个数为21时, 网络能够达到最高正确率88.34%; 当字典维数为600维, 循环次数为4, 感受野个数为21时, 网络能够达到最高正确率88.71%。

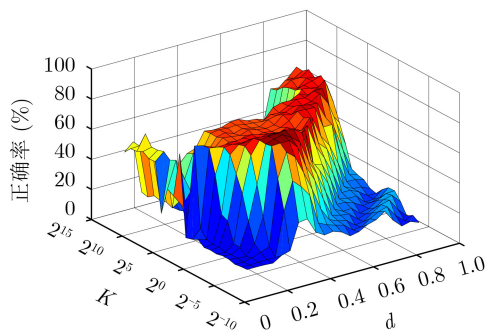
为验证SR-CELM对含嘈杂背景的场景图像的分类性能, 选取数据库中具有嘈杂背景的场景图像进行测试, SR-CELM, 文献[15], CF-ELM和

ELM-MSLRF的正确率对比如图11所示。由于不同场景类图像中可能存在相似物体, 因此在办公室这一小类上SR-CELM分类正确率略低于CF-ELM。但从整体分类正确率来说, SR-CELM仍然具有良好的抗干扰能力。

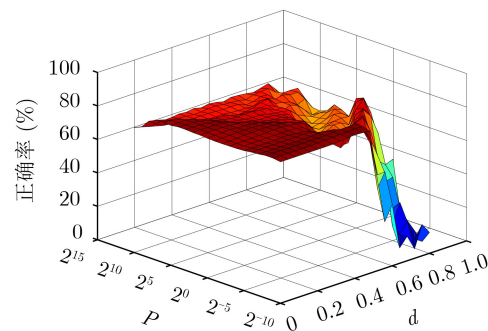
表3将本文实验结果与其他算法进行比较。在字典维数相同时, SR-CELM分类正确率分别比SVM, F-ELM高出1.88%, 4.01%, 较同系列的ELM-MSLRF具有明显优势。文献[17]比本文方法高1.39%的正确率, 然而其算法的复杂性远高于本文方法。与深度学习网络相比, SR-CELM使用更少的层数和较少的节点数, 分别比AlexNet和VGGNet高出13.8%,

表 2 MSRC正确率(%)与训练时间(s)

方法	文献[20]	文献[21]	KELM	F-ELM	文献[15]	SVM	CF-ELM	ELM-MSLRF	SR-CELM	
字典维数	265	400	400	400	400	400	400	400	600	
训练时间	-	-	17.01	17.04	-	17.11	18.20	13.24	19.59	25.46
正确率	71.00	83.50	91.74	93.49	93.97	94.13	95.76	96.01	96.35	96.83

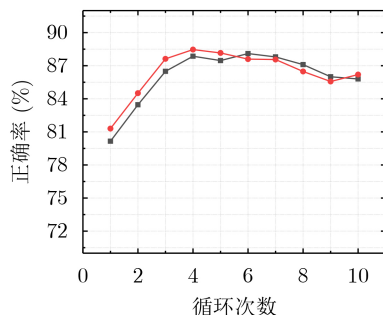


(a) 参数 K 和 d 对网络正确率的影响

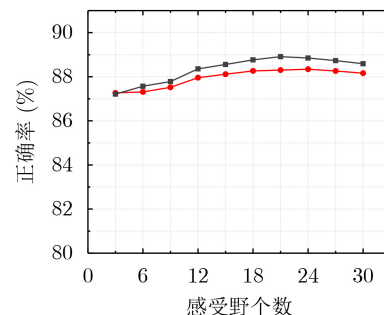


(b) 参数 P 和 d 对网络正确率的影响

图 9 不同参数变化时网络正确率曲面图



(a) 循环次数对网络正确率的影响



(b) 感受野个数对网络正确率的影响

图 10 不同字典维数下循环次数和感受野个数对网络正确率影响折线图

表 3 15 Scene正确率(%)与训练时间(s)

方法	AlexNet	文献[18]	VGGNet ^[22]	KELM	F-ELM	SVM	CF-ELM	文献[15]	GoogleNet	文献[17]	SR-CELM	
字典维数	-	1000	-	400	400	400	400	400	-	600	400	600
训练时间	-	-	-	106.11	106.41	106.25	120.10	7020	-	-	131.35	147.45
正确率	74.54	81.73	84.75	83.53	84.33	86.46	87.76	86.46	89.34	90.10	88.34	88.71

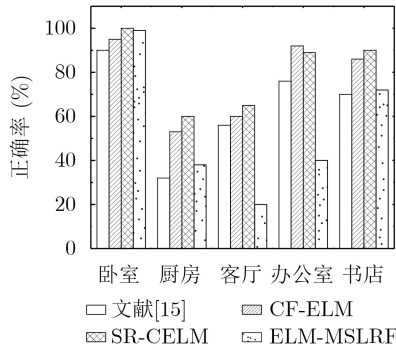


图 11 4种方法对15 Scene数据库中具有嘈杂背景图像的正确率

3.59%，略低于GoogleNet，但SR-CELM利用简单的结构模型和较快的训练速度同样能够达到较高正确率。

4 结束语

本文针对空间金字塔SPM特征空间分布信息利用不足的问题，提出了空间金字塔与感受野相结合的相关熵极限学习机，并将其用于图像分类中。通过大量实验并分析实验结果可知，在特征提取时，将空间金字塔与局部感受野相结合能进一步利用特征的空间信息，并利用所提出网络对特征进行编码，有效实现了多特征的融合。在网络权重训练时，利用相关熵准则分别构建实际输出与期望输出的角度和模长约束，提高了网络分类能力和稳定性。因此理论推导和实验结果表明，SR-CELM具有良好的鲁棒性和分类性能，但是，基于相关熵准则求解网络输出权重需要迭代调整，如何不通过迭代直接解析地求解权重是下一步研究的方向。

参考文献

- [1] CSURKA G, DANCE C R, FAN Lixin, *et al.* Visual categorization with bags of keypoints[C]. Workshop on Statistical Learning in Computer Vision, European Conference on Computer Vision, Prague, Czech Republic, 2004: 1–22.
- [2] LAZEBNIK S, SCHMID C, and PONCE J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories[C]. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, USA, 2006: 2169–2178. doi: [10.1109/CVPR.2006.68](https://doi.org/10.1109/CVPR.2006.68).
- [3] LI Qing, PENG Qiang, CHEN Junzhou, *et al.* Improving image classification accuracy with ELM and CSIFT[J]. *Computing in Science & Engineering*, 2019, 21(5): 26–34. doi: [10.1109/MCSE.2018.108164708](https://doi.org/10.1109/MCSE.2018.108164708).
- [4] HUANG Guangbin, ZHU Qinyu, and SIEW C K. Extreme learning machine: Theory and applications[J]. *Neurocomputing*, 2006, 70(1/3): 489–501. doi: [10.1016/j.neucom.2005.12.126](https://doi.org/10.1016/j.neucom.2005.12.126).
- [5] PARK J M and KIM J H. Online recurrent extreme learning machine and its application to time-series prediction[C]. 2017 International Joint Conference on Neural Networks, Anchorage, USA, 2017: 1983–1990. doi: [10.1109/ijcnn.2017.7966094](https://doi.org/10.1109/ijcnn.2017.7966094).
- [6] LIU Zheng, WEI Jin, and YING Mu. Variances-constrained weighted extreme learning machine for imbalanced classification[J]. *Neurocomputing*, 2020, 403: 45–52.
- [7] 刘彬, 杨有恒, 赵志彪, 等. 一种基于正则优化的批次继承极限学习机算法[J]. 电子与信息学报, 2020, 42(7): 1734–1742. doi: [10.11999/JEIT190502](https://doi.org/10.11999/JEIT190502).
- [8] LIU Bin, YANG Youheng, ZHAO Zhibiao, *et al.* A batch inheritance extreme learning machine algorithm based on regular optimization[J]. *Journal of Electronics & Information Technology*, 2020, 42(7): 1734–1742. doi: [10.11999/JEIT190502](https://doi.org/10.11999/JEIT190502).
- [9] HUANG Guangbin, BAI Zuo, KASUN L L C, *et al.* Local receptive fields based extreme learning machine[J]. *IEEE Computational Intelligence Magazine*, 2015, 10(2): 18–29. doi: [10.1109/mci.2015.2405316](https://doi.org/10.1109/mci.2015.2405316).
- [10] KRIZHEVSKY A, SUTSKEVER I, and HINTON G E. ImageNet classification with deep convolutional neural networks[C]. The 25th International Conference on Neural Information Processing Systems, New York, USA, 2012: 1097–1105.
- [11] SIMONYAN K and ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv: 1409.1556, 2014.
- [12] 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. doi: [10.1109/cvpr.2015.7298594](https://doi.org/10.1109/cvpr.2015.7298594).
- [13] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016: 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [14] XING Hongjie and WANG Xinmei. Training extreme learning machine via regularized correntropy criterion[J]. *Neural Computing and Applications*, 2013, 23(7/8): 1977–1986. doi: [10.1007/s00521-012-1184-y](https://doi.org/10.1007/s00521-012-1184-y).
- [15] 吴超, 李雅倩, 张亚茹, 等. 用于表示级特征融合与分类的相关熵融合极限学习机[J]. 电子与信息学报, 2020, 42(2): 386–393. doi: [10.11999/JEIT190186](https://doi.org/10.11999/JEIT190186).
- [16] WU Chao, LI Yaqian, ZHANG Yaru, *et al.* Correntropy-based fusion extreme learning machine for representation level feature fusion and classification[J]. *Journal of Electronics & Information Technology*, 2020, 42(2): 386–393. doi: [10.11999/JEIT190186](https://doi.org/10.11999/JEIT190186).

- [15] 李雅倩, 吴超, 李海滨, 等. 局部位置特征与全局轮廓特征相结合的图像分类方法[J]. 电子学报, 2018, 46(7): 1726–1731. doi: [10.3969/j.issn.0372-2112.2018.07.026](https://doi.org/10.3969/j.issn.0372-2112.2018.07.026).
LI Yaqian, WU Chao, LI Haibin, *et al.* Image classification method combining local position feature with global contour feature[J]. *Acta Electronica Sinica*, 2018, 46(7): 1726–1731. doi: [10.3969/j.issn.0372-2112.2018.07.026](https://doi.org/10.3969/j.issn.0372-2112.2018.07.026).
- [16] HUANG Jinghong, YU Zhuliang, CAI Zhaoquan, *et al.* Extreme learning machine with multi-scale local receptive fields for texture classification[J]. *Multidimensional Systems and Signal Processing*, 2017, 28(3): 995–1011. doi: [10.1007/s11045-016-0414-3](https://doi.org/10.1007/s11045-016-0414-3).
- [17] KONIUSZ P, YAN Fei, GOSSELIN P H, *et al.* Higher-order occurrence pooling for bags-of-words: Visual concept detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(2): 313–326. doi: [10.1109/TPAMI.2016.2545667](https://doi.org/10.1109/TPAMI.2016.2545667).
- [18] WANG Jinjun, YANG Jianchao, YU Kai, *et al.* Locality-constrained Linear Coding for image classification[C]. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 2010: 13–18. doi: [10.1109/CVPR.2010.5540018](https://doi.org/10.1109/CVPR.2010.5540018).
- [19] 肖文华, 包卫东, 陈立栋, 等. 一种用于图像分类的语义增强线性编码方法[J]. 电子与信息学报, 2015, 37(4): 791–797. doi: [10.11999/JEIT140743](https://doi.org/10.11999/JEIT140743).
XIAO Wenhua, BAO Weidong, CHEN Lidong, *et al.* A semantic enhanced linear coding for image classification[J]. *Journal of Electronics & Information Technology*, 2015, 37(4): 791–797. doi: [10.11999/JEIT140743](https://doi.org/10.11999/JEIT140743).
- [20] LI Weisheng, DONG Peng, XIAO Bin, *et al.* Object recognition based on the region of interest and optimal bag of words model[J]. *Neurocomputing*, 2016, 172: 271–280. doi: [10.1016/j.neucom.2015.01.083](https://doi.org/10.1016/j.neucom.2015.01.083).
- [21] KHAN R, BARAT C, MUSELET D, *et al.* Spatial histograms of soft pairwise similar patches to improve the bag-of-visual-words model[J]. *Computer Vision and Image Understanding*, 2015, 132: 102–112. doi: [10.1016/j.cviu.2014.09.005](https://doi.org/10.1016/j.cviu.2014.09.005).
- [22] 姜轩. 基于深度学习的场景图像分类算法研究[D]. [硕士学位论文], 北京邮电大学, 2019.
JIANG Xuan. Research on scene image classification algorithm based on deep learning[D]. [Master dissertation], Beijing University of Posts and Telecommunications, 2019.

刘彬: 男, 1953年生, 教授, 研究方向为计算机视觉.

刘静: 女, 1996年生, 硕士生, 研究方向为计算机视觉.

吴超: 男, 1990年生, 博士生, 研究方向为计算机视觉.

杨有恒: 男, 1996年生, 硕士生, 研究方向为机器学习.

责任编辑: 余蓉