

# 基于深度特征融合的图像分类算法的研究

蔡志锋<sup>1</sup> 袁宝华<sup>2\*</sup> 刘广海<sup>3</sup>

<sup>1</sup>( 三江学院计算机科学与工程学院 江苏 南京 210000)

<sup>2</sup>( 南京理工大学泰州科技学院计算机系 江苏 泰州 225300)

<sup>3</sup>( 广西师范大学计算机科学与信息工程学院 广西 桂林 541004)

**摘 要** 提出一种基于深度特征融合的图像分类方法。通过不同的深度学习预训练网络来获取图像的高层语义特征;采用 Weighted Discriminant Correlation Analysis( WDCA) 方法提取其转换矩阵及其融合矩阵;通过支持向量机分类器进行分类识别。在 Caltech 256 标准数据库上的实验表明,该方法不但能够有效地优化整合不同的深度特征,而且能够有效地降低特征的冗余信息,从而使融合后的特征具有很强的鉴别能力和低维特点。

**关键词** 深度特征 图像分类 特征融合

中图分类号 TP391.41

文献标志码 A

DOI: 10.3969/j.issn.1000-386x.2020.10.028

## IMAGE CLASSIFICATION METHOD BASED ON DEEP FEATURE FUSION

Cai Zhifeng<sup>1</sup> Yuan Baohua<sup>2\*</sup> Liu Guanghai<sup>3</sup>

<sup>1</sup>( College of Computer Science and Engineering , Sanjiang University , Nanjing 210000 , Jiangsu , China)

<sup>2</sup>( Department of Computer Science and Technology , Taizhou Institute of Sci. & Tech. , Nanjing University of Science and Technology , Taizhou 225300 , Jiangsu , China)

<sup>3</sup>( College of Computer Science and Information Engineering , Guangxi Normal University , Guilin 541004 , Guangxi , China)

**Abstract** This paper presents an image classification method based on deep feature fusion. The high-level semantic features of images were obtained through different deep learning pre-training networks; the weighted discriminant correlation analysis( WDCA) method was used to extract the transformation matrix and its fusion matrix; the support vector machine classifier was used for classification and recognition. Experiments on the Caltech 256 standard database show that our method not only can effectively optimize the integration of different deep features , but also effectively reduce the redundant information of features , so that the fused features have strong discriminating ability and low dimensional characteristics.

**Keywords** Deep feature Image classification Feature fusion

## 0 引 言

基于内容的图像分类主要通过图像的视觉特征信息对图像进行分类,是近年来计算机视觉、人工智能等领域中一个非常热门的研究课题。传统的图像分类方式通常使用统计学习算法进行图像分类,例如:支持向量机<sup>[1]</sup>( SVM)、K 均值聚类<sup>[2]</sup>和最近邻距离<sup>[3]</sup>等。但

上述算法未曾考虑图像的高层语义信息,因此经常产生错分、漏分等现象,分类精度较低。

深度学习网络是近十年来兴起的一种方法,已经应用到图像的各个领域并取得了令人惊喜的进展,如:目标识别<sup>[4]</sup>、图像检索<sup>[5]</sup>和图像分类<sup>[6-7]</sup>等。深度学习通过多个隐含层的学习,从大规模数据中逐层进行学习和特征提取,寻找数据最优的抽象表达方法,从而提高分类或识别的准确率。但深度网络模型需要大

收稿日期:2019-07-05。国家自然科学基金项目(61866005)。蔡志锋,高级实验师,主研领域:人工智能。袁宝华,副教授。刘广海,教授。

量的数据进行训练,耗时较长,对存储要求也相对较高。对于小数据量的数据集,直接采用 CNN 进行训练,往往容易造成过拟合现象,因此通常采用预训练的 CNN 网络。输入图像经过深度学习预训练网络,将卷积层或者连接层的响应值作为该图像的深度特征,然后通过训练 SVM 等常规分类器对场景图像进行分类。

数据模态的多样性必然导致特征抽取方式的多样性,每种特征抽取方式都有其关注的侧重点,比如:纹理特征描述符只关注纹理而不关心颜色,因此单个特征通常只表征某一方面的信息。特征融合优势是明显的,因为同一模式所提取的不同特征向量反映了模式的不同特性,对它们进行优化组合,既保留了参与融合的多组特征的有效鉴别信息,又消除了特征向量之间的冗余信息。深度学习方法已经深入到在计算机的各个领域中,研究适用于深度学习的信息融合技术很有必要。融合方法通常分为像素级、特征层以及决策层的融合<sup>[8]</sup>。特征层的融合<sup>[9-11]</sup>相比其他两种融合更有效,因为其融合后的特征通常包含更多丰富的信息,从而提高识别率。Miao 等<sup>[9]</sup>通过 ResC3D 深度学习网络提取深度特征,然后采用典型相关分析方法进行特征融合,用在手势识别上取得了不错的效果。Haghighat 等<sup>[10]</sup>考虑分类中的类别信息,提出鉴别相关分析的方法用于多模态的特征融合,该方法能够有效地消除类间的相关性,限制类内的相关性。Chaib 等<sup>[11]</sup>采用 VGG 预训练网络不同全连接层的特征,通过判别相关分析(Discriminant Correlation Analysis, DCA)进行特征融合,最后通过 SVM 进行分类。借鉴文献[9-11]的思路,本文采用不同的预训练网络来提取图像的高层语义特征,然后进行特征融合,最后采用 SVM 分类器分类。

本文比较了两种不同的深度特征获取策略:(1)采用同一预训练深度学习网络的不同层特征的特征融合;(2)采用多个深度学习网络的全连接层特征的特征融合。同时,通过增加权重,重新定义 DCA 方法中的类间散度矩阵,使其能够对那些类别差异较小的类别也有较好的区分度。该方法可以充分利用深层网络结构的优势,获取图像的高层语义特征,提高分类结果的准确性,同时考虑了深度特征融合,能更有效地挖掘高层语义特征,大幅降低训练时间并提高分类精度。构建多个预训练 CNN 的深度特征的特征融合,将其应用到图像分类中,实验结果表明该特征融合方法获得的深度特征相比于单个深度学习特征,具备强大的特征表征能力和低维特性,从而提高图像分类性能。

## 1 相关知识

### 1.1 特征融合

典型相关分析(Canonical Correlation Analysis, CCA)通常是用来解决两组随机向量之间相互关系的统计方法,其目的是寻找两组投影方向,使两个随机向量投影后的相关性达到最大<sup>[12]</sup>。然而,CCA 的主要缺点是其忽略了数据集中的类别信息。近年来,DCA 的出现克服了 CCA 中忽略类别信息的缺点,能够将同类样本特征相关性最大化,同时最小化不同类样本特征之间的相关性,有利于提高分类性能。

假设  $D$  为数据集, $C$  为类别数, $X$  为特征集,则  $x_{ij} \in X$  表示第  $j$  类中的第  $i$  个图像。假设  $\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^n x_{ij}$  表示第  $i$  类的平均值, $\bar{x}$  表示所有类的平均值:

$$\bar{x} = \frac{1}{n_i} \sum_{i=1}^c n_i \bar{x}_i \quad (1)$$

类间散度矩阵定义为:

$$S_{\text{bx}(p \times q)} = \Phi_{\text{bx}} \Phi_{\text{bx}}^T \quad (2)$$

式中: $\Phi_{\text{bx}} = [\sqrt{n_1}(\bar{x}_1 - \bar{x}), \sqrt{n_2}(\bar{x}_2 - \bar{x}), \dots, \sqrt{n_c}(\bar{x}_c - \bar{x})]$ 。

$\Phi_{\text{bx}} \Phi_{\text{bx}}^T$  的特征向量能够通过  $\Phi_{\text{bx}}^T \Phi_{\text{bx}}$  的特征向量映射得到。为了类间可分离,将  $\Phi_{\text{bx}}^T \Phi_{\text{bx}}$  对角化:

$$P^T \Phi_{\text{bx}}^T \Phi_{\text{bx}} P = \Lambda \quad (3)$$

式中: $P$  是正交向量矩阵; $\Lambda$  是由非负特征值组成的对角矩阵。式(3)可表示为:

$$\varphi^T \Phi_{\text{bx}}^T \Phi_{\text{bx}} \varphi = \Lambda_{(r \times r)} \quad (4)$$

式中: $\varphi_{(c \times r)}$  是最大非零的特征向量矩阵; $r$  表示矩阵的秩。

$S_{\text{bx}}$  最大  $r$  个特征向量能够通过映射  $\varphi \rightarrow \Phi_{\text{bx}} \varphi$  得到:

$$(\Phi_{\text{bx}} \varphi)^T S_{\text{bx}} \Phi_{\text{bx}} \varphi = \Lambda_{(r \times r)} \quad (5)$$

设  $W_{\text{bx}} = \Phi_{\text{bx}} \varphi \Lambda^{-\frac{1}{2}}$ , 则:

$$W_{\text{bx}}^T S_{\text{bx}} W_{\text{bx}} = I \quad (6)$$

$$X'_{(r \times n)} = W_{\text{bx}}^T_{(r \times p)} \cdot X_{(p \times n)} \quad (7)$$

第二特征集  $Y$  也采用上述方法得到:

$$W_{\text{by}}^T S_{\text{by}} W_{\text{by}} = I \quad (8)$$

$$Y'_{(r \times n)} = W_{\text{by}}^T_{(r \times p)} \cdot Y_{(p \times n)} \quad (9)$$

式中: $r$  为转换后特征的秩。

$$r \leq \min(c-1, \text{rank}(X), \text{rank}(Y)) \quad (10)$$

由于将  $X$  和  $Y$  转换到了  $X'$  和  $Y'$ ,使得类间具有最小相关性。接下来  $X'$  和  $Y'$  两个特征集具有最大相关性。将  $X'$  和  $Y'$  构成的协方差矩阵( $S'_{xy} = X'Y'^T$ ) 对角化,采用奇异值分解方法对角化  $S'_{xy}$ 。

$$S'_{xy(r \times r)} = U \Sigma V^T \Rightarrow U^T S'_{xy} V = \Sigma \Rightarrow$$

$$(U \Sigma^{-\frac{1}{2}})^T S'_{xy} (V \Sigma^{-\frac{1}{2}}) = I \quad (11)$$

令  $W_{cx} = U \Sigma^{-\frac{1}{2}}$  和  $W_{cy} = V \Sigma^{-\frac{1}{2}}$  得到:

$$DX = W_{cx}^T X' = W_{cx}^T W_{bx}^T X = W_x^T X \quad (12)$$

$$DY = W_{cy}^T Y' = W_{cy}^T W_{by}^T Y = W_y^T Y \quad (13)$$

式中:  $W_x = W_{bx} W_{cx}$  和  $W_y = W_{by} W_{cy}$  分别表示  $X$  和  $Y$  的最终转换矩阵。

## 1.2 深度特征网络

在过去的十年中,已经开发出几种用于大规模图像分类和物体检测的 CNN 模型,例如: Alexnet<sup>[13]</sup>, VGG<sup>[14]</sup>, ResNet<sup>[7]</sup>, Inception-v3<sup>[16]</sup>, GoogLeNet<sup>[17]</sup> 和 Inception-ResNet-v2<sup>[18]</sup> 等。不同的预训练网络具有不同的特性,它们都具备了提取强大且丰富特征的能力,从而使得其能迁移学习到其它领域中。不同的深度学习网络具有不同的网络结构,下面介绍几种典型的深度网络。

(1) AlexNet: 由 Alex Krizhevsky 设计的卷积神经网络,共八层,前五层是卷积层,后三层是全连接层。在网络中使用非饱和 ReLU 激活函数,其能够提供比 tanh 和 sigmoid 等激活函数更好的训练性能。实验中,网络的输入图像大小通常为  $227 \times 227$ ,提取第一个全连接层特征作为图像特征,其维数为 4 096 维。

(2) ResNet: 2015 年提出的深度卷积网络,当年在 ImageNet 竞赛中获得图像视觉挑战中三项任务的冠军。通过增加网络的深度来提高识别率,从而解决网络深度增加带来的退化问题,使网络更容易优化。实验中,提取残差网络的全连接层作为图像特征,其维数为 1 000 维。

(3) VGG: 在 AlexNet 网络的基础上开发的,其具有良好的泛化能力。VGG 网络由 conv、pool、fc 和 softmax 层组成。它的主要贡献是使用一个非常小的  $3 \times 3$  卷积内核进行网络设计,并将网络深度增加到 16 或 19 层。在实验中,采用第一个全连接层的特征来作为图像特征,其维数为 4 096 维。

(4) GoogLeNet: 由 Szegedy 等构造的深度学习网络,其网络结构稀疏且具备高计算性能。该模型通过构造 Inception 模块和均值池化来代替全连接层来减少模型参数规模。在网络设计之初,研究人员就考虑了计算效率和实用性,从而让 GoogLeNet 能够在不同设备上运行。实验中,获取最后一个池化层特征来作为图像特征向量,其维数是 1 024 维。

几种典型的深度学习网络参数如表 1 所示。

表 1 典型的深度学习网络参数

深度网络	层数	输入大小	全连接层维数
AlexNet	8	227 227	4 096
VGG	16 或 19	224 224	4 096
GoogLeNet	22	224 224	1 024
Inception-v3	48	299 299	1 000
ResNet101	101	224 224	1 000
Inception-ResNet-v2	164	299 299	1 000

与大多数基于 SIFT、SURF 和 HOG 等低层特征的场景分类方法相比,本文提出的框架是基于不同预训练 CNN 模型的深度特征的融合。将不同预训练 CNN 模型的全连接层作为输入的特征向量,能够描述图像场景的重要特征。

## 2 基于深度特征融合的图片分类算法

### 2.1 Weighted DCA(WDCA)

DCA 算法考虑了类别信息,能够使同类特征相关最大化,并且不同类的特征相关最小化,但是其不能很好区分类间距离较小甚至重叠的类别。DCA 的类间散度矩阵如式(2)所示,对于那些类间距离越大的类别,其散度矩阵中对应的值越大,反之,则越小。这样会导致过分强调那些离散度大的类的作用而忽略了离散度较小的类。

为此,通过增加权重来约束,降低那些离散度大的类的影响,提高离散度较小类的作用。重新定义 DCA 算法中的类间散度矩阵为:

$$S_{bx(p \times p)} = \sum_{i=1}^c n_i W(\bar{x}_i, \bar{x}) (\bar{x}_i - \bar{x}) (\bar{x}_i - \bar{x})^T =$$

$$W(\bar{x}_i, \bar{x}) \Phi_{bx} \Phi_{bx}^T \quad (14)$$

式中:  $\Phi_{bx} = [\sqrt{n_1}(\bar{x}_1 - \bar{x})], [\sqrt{n_2}(\bar{x}_2 - \bar{x})], \dots, [\sqrt{n_c}(\bar{x}_c - \bar{x})]$ ;  $W(\bar{x}_i, \bar{x}) = \frac{\sum_{j=1}^p e^{-|(\bar{x}_i)_j - (\bar{x})_j|}}{p}$ ;  $W$  为权重函数,

取值范围是(0,1),表示样本  $x_i$  的中心  $\bar{x}_i$  与总体样本的中心  $\bar{x}$  的接近程度,越接近,则权重越大,反之,则越小。与中心接近的类获得较大的权重,更远离中心  $\bar{x}$ ,而那些远离中心的类权重更小,远离类的影响被削弱了。

### 2.2 基于 WDCA 的图片分类算法

目前,常用的特征融合方法通常有:简单叠加和串行连接。简单叠加方法难以反映多特征之间的差异性,而串行连接方法通常会产生高维的融合特征,其包

含了过多的冗余信息,导致计算效率降低。因此本文将 WDCA 的方法应用于深度网络的高层语义特征的融合,其不但能够有效地表征不同的语义特征之间差异,而且能够有效地降低特征融合后的维度,从而大大地节约计算成本。本文采用两种策略来进行深度网络的高层语义特征获取,并在实验中进行比较分析。

第一种策略:首先输入图像通过同一类型的卷积神经网络,然后分别提取不同层的深度特征作为图像特征,对其采用 WDCA 的方法进行融合。

第二种策略:首先输入图像通过两个不同类型的卷积神经网络提取到深度特征,然后对深度特征采用 WDCA 的方法进行融合。

两种策略仅在于深度学习特征的获取方式不同,后续特征融合的步骤相同。采用上述策略,特征融合后的特征维数为  $100 \times 2$ ,维数大大降低,使训练时间大大缩短,最后通过 SVM 分类器进行识别,如图 1 所示。

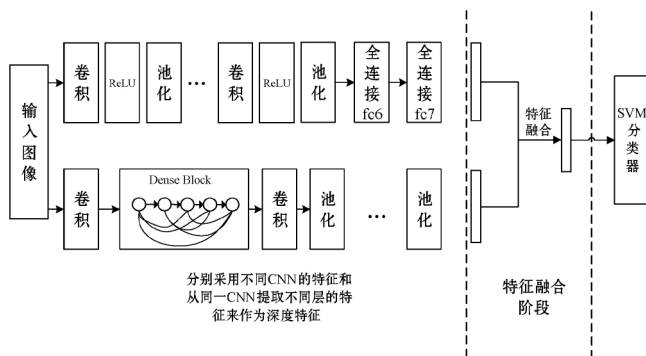


图1 基于深度特征融合的图片分类框架

选择不同深度网络的全连接层或者同一深度网络的不同层作为图像的深度特征,然后通过 WDCA 方法对两个不同的深度特征进行融合。由于  $X$  和  $Y$  的维数通常远大于类别数,如式(10)所示,则融合后的特征维数最大为  $C-1$ ,  $C$  为类别数。采用两种形式来表示单个图像的特征:

$$Z_1 = \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} W_x^T X \\ W_y^T Y \end{bmatrix} = \begin{bmatrix} W_x & 0 \\ 0 & W_y \end{bmatrix}^T \begin{bmatrix} X \\ Y \end{bmatrix} \quad (15)$$

此为串行形式,融合后的特征维数  $2 \times (C-1)$ 。

$$Z_2 = u + v = W_x^T X + W_y^T Y = \begin{bmatrix} W_x \\ W_y \end{bmatrix}^T \begin{bmatrix} X \\ Y \end{bmatrix} \quad (16)$$

此为求和形式,融合后的特征维数为  $C-1$ 。

基于深度特征融合的图片分类算法步骤如下:

(1) 利用不同的预训练网络,分别提取训练集图像的全接连层的深度特征,构成图像深度特征矩阵  $X_{m \times p} = [\varphi_1 \ \varphi_2 \ \cdots \ \varphi_m]$  和  $Y_{m \times q} = [\psi_1 \ \psi_2 \ \cdots \ \psi_m]$ 。

(2) 根据 WDCA 特征融合算法,利用式(12)和式(13)将矩阵  $X$  和  $Y$  融合后分解为  $DX$ 、 $DY$  以及投影矩

阵  $W_x$ 、 $W_y$ ;根据式(15)或式(16)组合成训练图像最终的特征融合向量  $Z_1$  或  $Z_2$ 。

(3) 利用不同的预训练网络,提取测试图像的深度特征矩阵  $TX_{n \times p} = [\varphi_1 \ \varphi_2 \ \cdots \ \varphi_n]$  和  $TY_{n \times q} = [\psi_1 \ \psi_2 \ \cdots \ \psi_n]$ 。分别将其投影到对应的特征空间  $W_x$ 、 $W_y$ ,得到测试图像的深度融合特征  $A$  和  $B$  特征向量:

$$A = W_x TX \quad (17)$$

$$B = W_y TY \quad (18)$$

(4) 根据式(15)或式(16)将测试图像的特征  $A$  和  $B$  组合成测试图像最终的特征融合向量  $TZ_1$  或  $TZ_2$ 。

(5) 采用 SVM 分类器进行图像分类。

## 3 实验结果

### 3.1 实验平台

为了验证特征融合方法的有效性,本文在公开的 Caltech 256 数据集上,对算法进行评估。在分类任务中,采用 SVM 分类器,使用 LIBSVM 库<sup>[1]</sup>,并通过五个交叉验证选择正则参数。实验中计算机配置如下: Intel Core I7-4710Mq CPU @ 2.5 GHz  $\times$  8,内存 12 GB,无 GPU。实验中软件环境为 MATLAB 2018b,使用的深度学习网络架构均来自 MATLAB 2018b 自带的深度学习包。

对于第一种策略,采用 VGG 预训练网络,提取 fc6、fc7 层特征进行特征融合;对于第二种策略,分别采用 VGG(简称 V)、resnet101(简称 R)、GoogLeNet(简称 G)和 inceptionresnetv2(简称 I)四种不同结构的预训练网络进行特征融合。

Caltech 256 数据库是 Li 等<sup>[18]</sup>在 Caltech 101 数据集的基础上进行的扩展,分为 256 个不同的对象类别,共有 30 607 幅图像。Caltech 256 数据库选自 Google Image 数据集,总共分为 256 个类别,每个类别包含的图像数量为 80 到 827 幅不等,每幅图像的尺寸大小不等,图 2 为 Caltech 256 部分示例图。

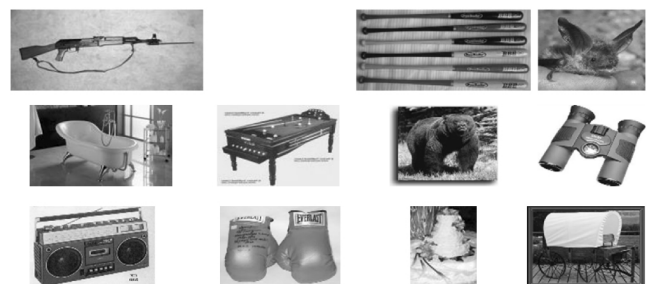


图2 Caltech 256 部分图像

### 3.2 实验分析

在使用不同的卷积神经网络处理之前,需要根据

不同的深度 CNN 网络的要求对输入图像大小进行调整。为了验证 WDCA 方法对于深度特征融合的有效性,分别采用单独的深度学习网络提取特征,然后采用不同深度网络的深度特征融合,最后采用 SVM 分类器进行分类识别,分类结果如表 2 所示。

表 2 不同的深度特征的图像分类结果 %

模型	无融合	DCA	WDCA
VGG_fc6	80.02	-	-
VGG_fc7	77.36	-	-
ResNet101	81.23	-	-
inceptionresnetv2	82.46	-	-
GoogLeNet	75.90		
G&R	-	83.52	84.61
I & R	-	85.69	87.23
V_fc6 & fc7		81.19	82.51

可以看出,与采用单个深度学习网络结果相比,本文提出的两种不同策略的深度特征融合方法的识别率有较好的提升,说明融合后的深度特征更能够描述场景,有利于提高场景的识别率。同时,WDCA 的特征融合效果要优于 DCA,这主要因为其平衡了类别差异的影响,类别差异小的类别能够增加其类间距离,而那些类别差异较大的类别,降低其类间距离,也不影响其区分度。

此外,从不同的深度学习框架提取的特征融合策略要优于从同一深度学习框架提取的特征融合,这主要因为不同的深度学习框架其特征的互补性要优于同一个深度学习网络的不同层。

不同的深度学习框架提取的特征,识别率也不相同,这说明不同的深度网络提取的特征各有差异。从融合后的结果来看,其不同的深度网络特征尽管不同,但是具有一定的互补性,因此融合后的特征表达能力要强于单个特征。GoogLeNet 预训练网络的特征识别率最低,但是融合后的特征识别率提升幅度最大。这也说明融合前的各自特征表达能力强,不代表融合后的特征表达能力一定强。融合后的特征表达能力强弱主要取决于融合前特征的互补性。

为了验证特征融合方法的执行效率,比较了单个深度网络的图像分类方法和基于 WDCA 深度特征融合的图像分类方法的训练时间和测试时间,结果如表 3 所示。可以看出,经本文方法融合后的特征维数仅为 200,远小于全连接层的 4 096 维特征,因此其训练时间和测试时间也大幅降低。特征融合方法是对深度特征进行优化组合,既保留了参与融合的深度特征的

有效鉴别信息,又消除了特征向量之间的冗余信息。

表 3 不同方法的训练时间和测试时间的比较 s

特征向量维数	训练时间	测试时间
4 096(深度特征)	18.263	0.103
200( WDCA)	1.441	0.007

4 结 语

针对复杂场景下的图像分类问题,本文提出基于深度学习预训练网络对场景进行特征学习,然后基于权重的 DCA 方法进行特征融合,最后通过 SVM 分类器进行图像场景的分类识别。传统的特征融合方法不仅容易导致维数增高,而且冗余信息较多。深度特征虽然表达特征能力较强,但是其维数较高且包含冗余信息,不同的深度网络结构提取的特征的表达能力也各不相同。本文提出的深度特征融合方法不仅能够有效地优化不同的深度 CNN 网络组合,而且能够有效地消除冗余信息,其特征维数也大大降低,在提高图像分类识别率的同时又减少了样本训练和测试时间。实验结果验证了本文方法的有效性和正确性。

参 考 文 献

[1] Chang C C, Lin C J. LIBSVM: a library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology 2011( 5) : 27.

[2] Jain A K. Data clustering: 50 years beyond K-means[J]. Pattern Recognition Letters 2010 31( 8) : 651 - 666.

[3] Zhang M L, Zhou Z H. ML-KNN: A lazy learning approach to multi-label learning[J]. Pattern Recognition 2007 40( 7) : 2038 - 2048.

[4] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision 2015 115( 3) : 211 - 252.

[5] 包青平,孙志锋. 基于度量学习的服装图像分类和检索[J]. 计算机应用与软件 2017 34( 4) : 255 - 259.

[6] 周忠义,吴谨,朱磊. 基于多路特征融合和深度学习的雾霾图像分类[J]. 计算机应用与软件 2018 35( 10) : 205 - 210 231.

[7] 赵浩如,张永,刘国柱. 基于 RPN 与 B-CNN 的细粒度图像分类算法研究[J]. 计算机应用与软件 2019 36( 3) : 210 - 213 264.

[8] Sun Q S, Zeng S G, Liu Y, et al. A new method of feature fusion and its application in image recognition[J]. Pattern Recognition 2005 38( 12) : 2437 - 2448.

( 下转第 243 页)

匹配。另外,设计了结构更加精简的风格转换网络,从而可以加快图像风格化速度的同时有效提升图像风格化的效果。从图像风格化效果的主观评价和图像风格化时间的客观评价两方面进行分析,结果表明 FSTSC 方法优于现有的一些方法。

下一步的研究工作是解决在图像风格化过程中当内容图像与风格图像的语义内容差别过大时,由于很难保证内容与风格的完美匹配所导致的在一定程度上影响生成图像视觉观感的问题。

## 参 考 文 献

- [1] Gatys L A, Ecker A S, Bethge M. A neural algorithm of artistic style[EB]. arXiv: 1508.06576 2015.
- [2] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks [C]//Computer Vision and Pattern Recognition ( CVPR ). IEEE 2016.
- [3] Li C, Wand M. Combining markov random fields and convolutional neural networks for image synthesis [C]//Computer Vision and Pattern Recognition. IEEE 2016.
- [4] Johnson J, Alahi A, Li F F. Perceptual losses for real-time style transfer and super-resolution [C]//European Conference on Computer Vision 2016.
- [5] Ulyanov D, Lebedev V, Vedaldi A, et al. Texture networks: feed-forward synthesis of textures and stylized images [C]//International Conference on Machine Learning ( ICML ) 2016.
- [6] Gatys L A, Ecker A S, Bethge M, et al. Controlling perceptual factors in neural style transfer [C]//Computer Vision and Pattern Recognition. IEEE 2017.
- [7] 吴联坤. 基于 TensorFlow 分布式与前景背景分离的实时图像风格化算法[D]. 杭州: 浙江大学 2017.
- [8] Briechele K, Hanebeck U D. Template matching using fast normalized cross correlation[J]. Proceeding of SPIE on Optical Pattern Recognition XII 2001 4387: 95 – 102.
- [9] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence 2018 40 ( 4 ): 834 – 848.
- [10] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//Computer Vision and Pattern Recognition. IEEE 2016.
- [11] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision [C]//Computer Vision and Pattern Recognition. IEEE 2016.
- [12] Radford A, Metz L, Chintala S, et al. Unsupervised representation learning with deep convolutional generative adversarial networks [C]//International Conference on Learning Representations 2016.
- [13] 操江峰. 基于深度学习的图像与视频风格化研究与实现 [D]. 北京: 中国科学院工程管理与信息技术学院 2017.
- [14] Chikontwe P, Lee H J. Towards robust face sketch synthesis with style transfer algorithms [M]//IT Convergence and Security 2017 2018: 172 – 179.
- [15] 赵梦. 基于主成分分析法的图像质量评价方法研究[D]. 杭州: 杭州电子科技大学 2013.

( 上接第 179 页 )

- [9] Miao Q, Li Y, Ouyang W, et al. Multimodal gesture recognition based on the ResC3D network [C]//2017 IEEE International Conference on Computer Vision Workshops ( IC-CVW ). IEEE, 2018: 3047 – 3055.
- [10] Haghighat M, Abdel M M, Alhalabi W. Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition [J]. IEEE Transactions on Information Forensics and Security 2016 11 ( 9 ): 1984 – 1996.
- [11] Chaib S, Liu H, Gu Y, et al. Deep feature fusion for VHR remote sensing scene classification [J]. IEEE Transactions on Geoscience and Remote Sensing 2017 55 ( 8 ): 4775 – 4784.
- [12] Hotelling H. Relations between two sets of variates [J]. Biometrika 1936 28 ( 3/4 ): 321 – 377.
- [13] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. Advances in neural information processing systems 2012 25 ( 2 ): 1097 – 1105.
- [14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB]. [2019 – 07 – 05]. arXiv: 1409.1556 2014.
- [15] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision [C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition 2016: 2818 – 2826.
- [16] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions [C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition 2015: 1 – 9.
- [17] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning [C]//The Thirty-First AAAI Conference on Artificial Intelligence 2017: 4278 – 4284.
- [18] Li F F, Fergus R, Perona P. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories [J]. Computer vision and Image Understanding 2007 106 ( 1 ): 59 – 70.