

基于词袋模型的迁移学习算法

吴丽娜 黄雅平 郑翔

(北京交通大学计算机与信息技术学院 北京 100044)

摘要 在分类新类别图像时,词袋模型总需要重新学习视觉词典及分类器,而不能充分利用已经学习好的视觉词典。运用迁移学习的思想,提出一种视觉短语的迁移学习算法。这种视觉短语不仅包含图像的局部不变特征,而且包含特征间的空间结构信息,能更有效地描述不同类别图像之间的共同特征。在分类新类别图像时,算法通过迁移视觉短语而不是重新学习视觉词典,来完成图像分类任务。实验结果证明这种迁移算法能有效地利用已有知识,在分类新类别图像时取得很好的效果,而且还能适用于仅有少量训练样本的图像分类任务。

关键词 图像分类,词袋模型,迁移学习

中图分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2014.12.056

Novel Transfer Learning Algorithm Based on Bag-of-visual Words Model

WU Li-na HUANG Ya-ping ZHENG Xiang

(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

Abstract The bag-of-visual words model needs to learn visual vocabulary and classifier from the beginning when it learns a novel image category, and it cannot make use of learned visual vocabulary. This paper proposed a transfer learning algorithm based on visual phrases. The visual phrases contain not only local invariable features, but also local spatial information, which can describe the common characteristics among different image categories. Our algorithm can make the bag-of-visual words model obtain good performance in the novel image category by transferring visual phrases from source visual vocabulary. The experimental results validate that the algorithm effectively utilizes learned knowledge and gains better performance in the novel image category even when there are a few training images.

Keywords Image categorization, Bag-of-visual words model, Transfer learning

1 引言

近年来,词袋模型在图像分类任务中取得了良好的效果,因此得到了广泛研究和应用。但是词袋模型是一种有监督机器学习算法,其可靠性依赖于一个严格的假设:用于建立模型的训练图像样本与新的图像样本满足独立同分布的条件。因此这种学习方式存在着两个问题:

1)学好的视觉词典只能用于同分布的同类图像的分类中。在学习一个新类别时,词袋模型需要从大量的标记好的图像样本中重新学习视觉词典及分类器,而不能充分利用已经学习好的视觉词典,这限制了词袋模型的泛化能力。

2)当新类别的图像只有少量已标记的图像样本时,词袋模型所获得的视觉词典难以刻画新类别的图像的总体分布特征,分类效果不理想。

为充分利用已学好的知识,提高词袋模型在新类别图像中的分类效果,许多研究者提出基于样本迁移的算法,例如 Yao^[1]提出的 Tradaboost,该方法从多个源任务的数据集中进行采样,使得采样后的数据集合与目标任务的数据的分布相似。Lin^[2]提出一种基于 Tradaboost 的样本迁移算法,

该方法采用一种双向选择机制减小源数据中不相关或有害数据的影响。由于视觉单词是词袋模型的核心和关键,因此基于特征的迁移学习方法的研究更为广泛,例如 Rajat Raina^[3]从各种未标记图像中提取局部特征,并学习其稀疏编码表示,将得到的基应用于新类别中。Long^[4]也将从中提取的局部区域在上述得到的基上的响应作为其特征表示来进行分类,在稀疏编码的基础上实现迁移学习。Florent Perronnin^[5]把通用词典作为先验知识,然后每一个类别在此基础上再学习与类别相关的词典。谢尧芳^[6]把从未标记样本中基于稀疏编码学习到的特征表示迁移到行人分类中。这些研究充分证明,利用已有的知识有助于提高新类别的学习能力。

基于迁移学习的思想,本文提出一种视觉短语的迁移学习算法。视觉短语是包含一定特征间的空间结构信息的视觉单词组合,这种单词组合不仅包含图像的局部不变特征,而且具有一定的高级语义层次,能更有效地描述不同类别图像之间的共同特征。在学习新类别图像时,算法不重新学习新图像类的单词,而是从已有的视觉短语词典中迁移视觉短语,并通过循环迭代的方式调整所迁移的视觉短语,保留对新图像分类有益的视觉短语,构建新类别图像的单词和短语词典。算

到稿日期:2013-06-26 返修日期:2013-08-16 本文受国家自然科学基金(61273364,61272354,61105119),北京市自然科学基金(4112047),中央高校基本科研业务费专项资金(2011JBZ005)资助。

吴丽娜(1982-),女,博士生,主要研究领域为人工智能、计算机视觉、模式识别, E-mail: wulina010@aliyun.com; 黄雅平(1974-),女,博士,教授,主要研究领域为人工智能、计算机视觉、模式识别; 郑翔(1971-),女,博士,副教授,主要研究领域为法学、计算机应用。

法利用这些词典训练分类器,使分类器能够在新图像类中取得良好的分类效果。与重新学习视觉词典的分类算法相比,这种迁移算法有效地利用了已有知识,在新类别图像的训练样本较少的情况下,也能获得较好的分类效果。

2 迁移算法的基本思想

传统的机器学习假设训练数据与测试数据服从相同的数据分布,因此,当训练集和目标集的分布不同时,词袋模型算法的分类效果欠佳。而迁移学习允许训练集和目标集的样本之间存在差异,它可以从已有的数据中迁移知识,来帮助新类别的学习。这种学习方法非常符合人类的学习方式。

词袋模型用视觉单词来表示图像中存在的相似特征,因此最直观的迁移方式就是从已知类别的视觉词典中选择一些单词迁移到新类别的视觉词典中,用于新类别图像的分类任务。但是这种直接迁移单词的方式分类效果很差。因为不同的图像类别中虽然会含有相似的视觉单词,但这样的单词的语义概念很可能相距甚远,甚至可能由于单词所代表的局部特征过小而不足以表示一个语义概念,这样的单词在图像的分类过程中很可能会具有歧义性。由于“分类”是对整个图像的概括的语义理解,因此采用较低层次的局部特征难以完成图像的高级语义理解。为构建具有一定高级语义的单词,很多研究者提出了一些加入空间信息的算法,如 Lazebnik^[7]提出的空间金字塔模型;Yang^[8]改进了空间金字塔模型,采用了稀疏编码的方法;Russakovsky^[9]提出了以目标为中心的图像子区域划分,分别对各个子区域中的视觉单词进行统计;Yao^[10]在算法中考虑了子区域之间相互的空间关系。这些研究表明,在词袋模型中加入空间关系能够有效地提高词袋模型的性能。

本文提出用视觉短语来描述不同类别图像之间的共同特征,用视觉短语代替视觉单词作为迁移的对象,完成新类别图像的分类任务。视觉短语由具有空间邻近性的单词组合而成,包含了图像的局部不变特征和特征间的空间结构信息。与视觉单词相比,它能表达更丰富的图像信息,并有助于消除其中任一单词独立使用时具有的语义歧义性。

3 视觉短语词典的创建

为获取知识,本文先用词袋模型算法通过学习大量已知类别的图像样本,获得与类别相关的视觉短语词典。为避免视觉短语的规模过大,本文算法首先用 WmRMR 准则创建一个紧凑的视觉单词集合,然后加入空间信息关系,将符合规则的视觉单词组合成视觉短语。

3.1 WmRMR 准则

本文从信息论的角度出发,提出加权的最大相关最小冗余 WmRMR 准则,并以此作为视觉词典子集的评价函数,从初始视觉词典中选择对算法的分类性能贡献程度大的视觉词典子集。

WmRMR 准则不仅考虑到单词与类别之间的相关性,还考虑了视觉词典内部单词之间的相关性,通过去除对分类性能影响较小的冗余单词,保留了对图像分类有益的富有判别力的视觉词典子集。

1. 单词与类别之间的相关性

算法定义初始视觉词典 $V = [v_1, \dots, v_K]$, V 是 $K \times D$ 的

矩阵, K 是视觉词典中单词的数量, D 是图像局部特征描述子的维数。单词 v_i 和类别 $c (c = 1 \dots C)$ 的相关性定义为

$$Rlv(v_i, c) = I(v_i, c) \quad (1)$$

视觉词典 V 和类别 c 的相关性定义为 V 中每个单词和相对应的类别的互信息的平均值,即

$$Rlv(V, c) = \frac{1}{|V|} \sum_{v_i \in V} I(v_i, c) \quad (2)$$

其中,单词和类别之间的互信息 $I(v_i, c)$ 可由如下公式计算得到:

$$I(v_i, c) = \sum_{c \in \{0,1\}} p(v_i, c) \log \frac{p(v_i, c)}{p(v_i)p(c)} + \sum_{c \in \{0,1\}} p(\bar{v}_i, c) \log \frac{p(\bar{v}_i, c)}{p(\bar{v}_i)p(c)} \quad (3)$$

2. 单词之间的相关性

单词 v_i 和视觉词典内其他单词的冗余性计算如下:

$$Rdd(v_i) = \frac{1}{|V| - 1} \sum_{v_j \in V, v_i \neq v_j} I(v_i, v_j) \quad (4)$$

视觉词典中视觉单词之间的互信息的平均值作为度量视觉词典中单词冗余性的函数。可由如下公式计算:

$$Rdd(V) = \frac{1}{|V|^2} \sum_{v_i, v_j \in V} I(v_i, v_j) \quad (5)$$

其中, $I(v_i, v_j)$ 是单词 v_i 和 v_j 的互信息。

3. WmRMR 准则

为使视觉词典子集能够满足单词与类别相关性尽量大,单词之间的冗余性尽量小,我们定义了加权的最大相关最小冗余性原则(WmRMR),即最大化评价函数 $\Phi(Rlv, Rdd)$ 。计算一个单词的 WmRMR 值如下:

$$\Phi(Rlv(v_i, c), Rdd(v_i)) = \lambda Rlv(v_i, c) - (1 - \lambda) Rdd(v_i) \quad (6)$$

视觉词典子集的 WmRMR 值的计算方法如下:

$$\Phi(Rlv(V, c), Rdd(V)) = \lambda Rlv(V, c) - (1 - \lambda) Rdd(V) \quad (7)$$

WmRMR 准则选择与图像类有最大相关性且单词之间的冗余性小的视觉词典。该算法一方面尽量保留与类别相关性强的单词,另一方面通过衡量单词之间的相关性,选择出具有代表性的单词,去除冗余的单词。单词集中所有单词经过 WmRMR 准则进行相关性分析后,保留的视觉单词组合成最终的视觉词典,满足以上对视觉词典的要求。

3.2 视觉短语的创建

本文将图像的局部空间信息融入到视觉单词中,创建视觉短语。从图像中可以发现,距离较近的几个单词比远距离的几个单词有较大的概率构成一个有意义的短语。

如图 1 所示,以图中“□”代表的局部特征为例,“○”标记的是图中与“□”在空间上相近的局部特征,而“△”标记的是与“□”距离较远的局部特征。从图中可以看出,特征“□”更可能与“○”所代表的邻近的局部特征组成有意义的视觉短语,这是因为邻近的特征更可能同时位于目标或同时处于背景中。

因此,本文定义从一个单词的 R 个近邻的单词范围内寻找到与该单词能构成稳定关系的一个或多个近邻单词来组成二元或多元视觉短语。为简单起见,本文算法采用二元视觉短语。对图像中提取到的每一个局部特征 p_m ,算法同时记录了其在图像中的空间位置坐标向量 l_m 。根据局部特征在图

像中的空间位置,可以找到 R 个与之近邻的特征 $\{nb_m^{(1)}, \dots, nb_m^{(R)}\}$, 组成 R 个特征二元对 $\{(p_m, nb_m^{(1)}), \dots, (p_m, nb_m^{(R)})\}$ 来表示特征 p_m 的局部空间信息。 $nb_m^{(r)}$ ($r=1, \dots, R$) 的计算如式(8)和式(9)所示。

$$nb_m^{(1)} = \{p_j | \max_{m \neq j} dist(l_m, l_j)\} \quad (8)$$

$$nb_m^{(r)} = \{p_j | \max_{m \neq j} dist(l_m, l_j)\} \quad (9)$$

where $p_j \neq nb_m^{(1)}, \dots, nb_m^{(r-1)}, 2 \leq r \leq R$

其中, $dist(l_m, l_j)$ 为距离函数, 计算两个局部特征的空间位置的欧氏距离。



(a)



(b)

图1 图像中局部特征的相对空间关系

通过上述定义,算法会得到一个候选的视觉短语的集合,但是其中包含了很多无意义的视觉短语。为获得具有稳定的相对空间的视觉短语,算法采用动态创建视觉短语的方法训练图像分类器以去除视觉短语中 $(v(p_m), v(nb_m^{(r)}))$ 低于预先设定的阈值的二元视觉短语,保留对分类有用的视觉短语,这些视觉短语词典为迁移学习奠定了基础。

4 视觉短语的迁移学习算法

记已学好类别的视觉词典为源视觉词典,它由源视觉单词集 $V_S = [v_1^S, \dots, v_n^S]$ 和视觉短语集 $V_{SS} = [(v_1^S, v_1^S), \dots, (v_n^S, v_n^S)]$ 组成。算法为每一个图像类别分别创建视觉词典,再利用 WmRMR 准则和空间信息将视觉单词聚合成一个大的含类别信息的视觉短语词典。目标视觉词典由目标视觉单词集 $V_T = [v_1^T, \dots, v_n^T]$ 和目标视觉短语集 $V_{TT} = [(v_1^T, v_1^T), \dots, (v_n^T, v_n^T)]$ 组成。

新类别的视觉词典由两部分组成:一部分是根据新类别当前的图像样本学习得到的自有视觉词典;另一部分是从源视觉词典中迁移过来的视觉短语词典。

算法计算源视觉词典中的单词和新类别中自有单词的相

似性。自有视觉词典中视觉单词 v_i^S 和源视觉词典中的单词 v_j^T 的相似性 $\text{sim}(v_i^S, v_j^T)$ 采用欧氏距离的倒数表示,见式(10),其值越大,表明两个单词越相似。算法保留源视觉词典与自有视觉词典中的相似性小的单词和其构成的视觉短语作为迁移学习所需的候选集。

$$\text{sim}(v_i^S, v_j^T) = \frac{1}{\sqrt{\sum_{k=1}^D (v_i^S[k] - v_j^T[k])^2}} \quad (10)$$

其次,算法期望所迁移的视觉短语能有助于提高图像分类效果。为此,本文为每一个视觉短语计算其在新类别中对分类的贡献程度,为算法找到适合新类别的视觉短语提供依据。本文算法使用贝叶斯分类器对新类别图像进行分类,即

$$c_j^* = \arg \max_{j=1, \dots, C} p(c_j | I) \quad (11)$$

根据贝叶斯公式,算法计算视觉词典中每个单词和视觉短语在新类别中出现的比重。单词 v_k 在新类别 c_j 中的比重定义为:

$$p(v_k | c_j) = \frac{1 + \sum_{i=1}^n N(v_k, I_i)}{|V_T| + \sum_{k=1}^n \sum_{i=1}^n N(v_k, I_i)} \quad (12)$$

其中, $|V_T|$ 是视觉词典中单词的总数,新类别中样本总数为 n , $N(v_k, I_i)$ 为单词 v_k 在图像 I_i 中出现的次数, $\sum_{k=1}^n \sum_{i=1}^n N(v_k, I_i)$ 为所有单词出现的次数之和。

视觉短语 (v_k, v_k') 在新类别 c_j 中的比重定义为:

$$p((v_k, v_k') | c_j) = \frac{1 + \sum_{i=1}^n N((v_k, v_k'), I_i)}{|V_{TT}| + \sum_{k=1}^n \sum_{i=1}^n N((v_k, v_k'), I_i)} \quad (13)$$

同样, $|V_{TT}|$ 是视觉词典中视觉短语的总数,新类别中样本总数为 n , $N((v_k, v_k'), I_i)$ 为视觉短语 (v_k, v_k') 在图像 I_i 中出现的次数, $\sum_{k=1}^n \sum_{i=1}^n N((v_k, v_k'), I_i)$ 为所有视觉短语出现的次数之和。

算法选择具有较大 $p((v_k, v_k') | c_j)$ 值的视觉短语迁移到新类别中,加入到目标视觉词典,帮助新类别的学习。

算法的主要步骤如下:

视觉短语的迁移学习算法:

- 1) 对新类别中的图像样本集,依据 Lowe 的 SIFT 方法提取新类别图像样本中的感兴趣点集合;
- 2) 从源视觉词典中随机挑选 L 个二元视觉短语到目标视觉词典中;
- 3) 根据目标视觉词典对步骤 1) 新类别中获得的感兴趣点进行向量量化;
- 4) 对新类别的每个图像根据视觉单词和二元视觉短语计算其直方图,根据直方图训练分类器,对图像样本进行分类,并计算分类误差;
- 5) 根据式(13)计算每个视觉短语对分类的贡献程度,将评价值小于等于 δ 的视觉短语删除;
- 6) 评价值都大于 δ 或迭代次数达到 30 时算法停止,否则从源数据集中继续随机挑选视觉短语,使得目标视觉词典中有 L 个二元视觉短语,并重复步骤 2)-4)。

由上述步骤得到的迁移过来的视觉词典与新类别的自有视觉词典共同构成新类别的视觉词典,在此基础上计算图像的直方图表示并进行分类;并且算法将新类别的自有视觉单词加入到源视觉词典中,来帮助其他新类别的迁移学习。以此类推,随着新类别的增加,源视觉词典不断更新,使之能为更多的新类别提供迁移学习的候选集,帮助其完成分类。

5 实验结果

为验证基于 WmRMR 准则的词袋模型算法的有效性,本文采用图像分类和识别任务中最经典的数据库 Caltech 101 以及 Caltech256 来进行实验验证。这两个数据库都具有图像数据庞大、图像种类多、对象类内变化多样等特点,具有一定的代表性。

为验证本文算法,需要创建用于迁移的源视觉词典。算法首先从上述图像库中选取 16 类图像 (Face, Airplanes, Leopards, Schooner, Stop_sign, barrel, buddha, ceiling_fan, chair, cougar_face, cougar_body, garfield, panda, snoopy, soccer_ball, Background)。对于每一类图像都分别学习,根据视觉单词构建视觉短语集,将这些视觉词典整个共同构成源视觉词典。

5.1 本文算法与词袋模型比较

为验证本文算法,将其与标准的词袋模型算法做比较,以较有代表性的人脸类和飞机类为例来说明算法的特点。实验中首先从图像集中随机选取一部分作为训练样本,这里设定训练样本占图像集的比率 μ 值取 0.5,即训练图像集和测试图像集各占整个数据集的一半,并且两种算法均采用相同的图像样本集。词袋模型算法中获得的视觉词典是根据该类别的训练样本集重新学习得到的;迁移学习算法中获得的视觉词典由根据该类别中的图像样本独立学习到的视觉单词和从源视觉词典迁移过来的视觉短语构成,由于源视觉词典的规模很大,这里设定前者占 20%,后者占 80%。图 2 所示为两种算法分别得到的描述分类性能的 ROC 曲线,图 2(a)为人脸类的结果,图 2(b)为飞机类的结果。

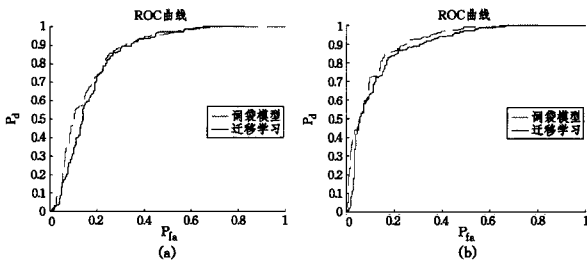


图 2 迁移学习与词袋模型分类性能对比

从图中可以看出,两种算法的分类性能相差不多,这说明,从源视觉词典中迁移到人脸类的视觉短语能够有效地帮助该类别的学习和分类。

5.2 样本数对算法的影响

当训练集中只有少量样本时,本文算法也能获得较好的分类效果。训练样本的数量按照其占该类别中总的样本数量的比例,取值范围为 10%到 50%;训练样本从该类数据集中随机选取,剩余的样本作为测试集合。两种算法在“人脸”、“飞机”和“猎豹”类的分类正确率随样本数变化的曲线图如图 3 所示,实验中其他类别也得到了类似的结果。当样本数较小时,随机选取的样本构成不同可能会带来较大的分类性能差异,为避免随机取样带来的结果偏差,实验中词袋模型算法和基于视觉短语的迁移学习算法都取 10 次实验结果的均值,且每次实验中,两种算法的训练样本数量和组成都完全一样。

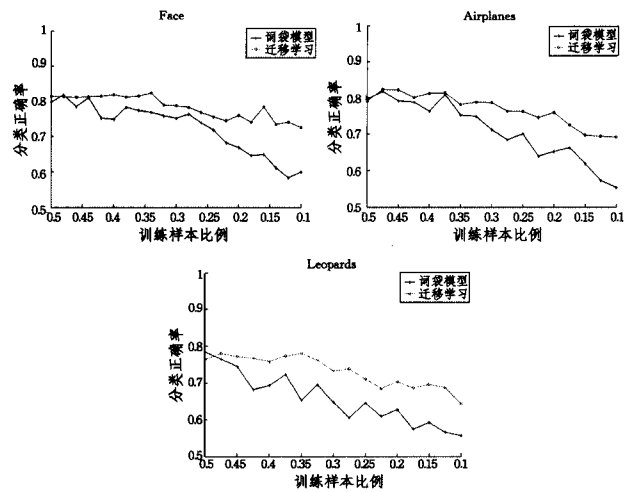


图 3 算法随训练样本数减小的分类正确率

图中横坐标表示训练样本数占该类别整个数据集的比例,纵坐标表示算法的分类正确率。从图中可以看出,当训练样本数逐渐减少时,标准的词袋模型算法的分类正确率迅速下降,而本文基于视觉短语的迁移学习算法基本保持稳定。

结束语 基于视觉短语的迁移学习算法能够利用已学习好的类别的知识帮助新类别分类学习,尤其在新类别只有少量样本的情况下,迁移学习仍然能够从已有的源视觉词典中找到对新类别分类有益的视觉短语,弥补新类别中知识的不足,帮助新类别取得较好的分类性能。而且这种迁移算法能从新图像类和已知图像类别中发现一些共同特征,这些关联信息还有助于判断类别之间的相互关系,为更高层次的语义理解提供有效信息。

本文算法中只采用了视觉单词之间的二元关系。将图像中的视觉单词与视觉短语按照一定的规则组合起来,建立多元关系的视觉单词组合,采用文本分类中的一系列有序的视觉短语组成的“句子”,来建立一个层次化的词袋模型分类算法,将是本文下一步的研究方向。

参考文献

- [1] Yao Y, Doretto G. Boosting for transfer learning with multiple sources[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2010:1855-1862
- [2] Lin D, Xing A, Zhang Z J. Double-bootstrapping source data selection for instance-based transfer learning[J]. Pattern Recognition Letters, 2013, 34(11): 1279-1285
- [3] Raina R, Battle A, Lee H, et al. Self-taught Learning: Transfer Learning from Unlabeled Data[C]//Proceedings of the 24th international conference on Machine learning, 2007:759-766
- [4] Long M, Ding G, Wang J. Transfer Sparse Coding for Robust Image Representation[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2013
- [5] Perronnin F. Universal and Adapted Vocabularies for Generic Visual Categorization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(7): 1243-1256
- [6] 谢尧芳, 苏松志, 李绍滋. 基于稀疏编码的迁移学习及其在行人检测中的应用[J]. 厦门大学学报: 自然科学版, 2010, 49(2): 186-192

(下转第 274 页)

板,能够较好地实现不同成像模式下视网膜图像中微小动脉瘤的动态匹配。

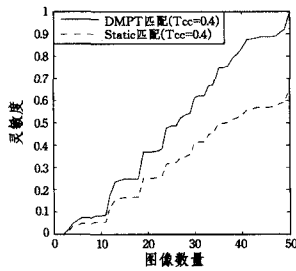


图 14 DMPT 模板和静态模板匹配灵敏度的比较($T_{cc}=0.4$)

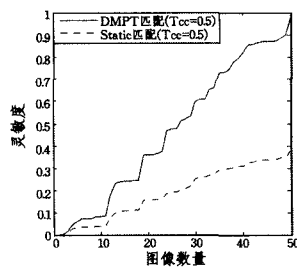


图 15 DMPT 模板和静态模板匹配灵敏度的比较($T_{cc}=0.5$)

此外,本文还通过仿真实验与文献[9]中提出的视网膜微小动脉瘤的检测算法进行了对比,对比结果如图 16 所示。图中虚线为作者实现的文献[9]中算法的 FROC 曲线,实线为本文算法的 FROC 曲线。表 3 和表 4 分别是灵敏度比较和平均假阳性比较数据,通过比较可以看出,上述算法在微小动脉瘤的检测灵敏度上平均提高了约 12.2%,并且在相同检测灵敏度时平均假阳性率降低了约 62.5%,因此可以证明本文中的算法无论是在检测灵敏度指标上还是平均假阳性指标上都优于文献[9]中的检测算法。所以本文中的算法对视网膜图像中微小动脉瘤的检测是有效的且更具优势。

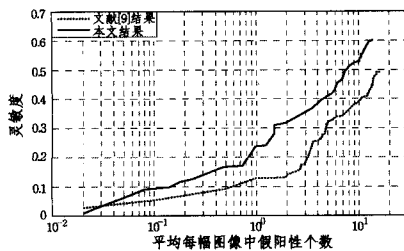


图 16 本文算法与文献[9]的比较

表 3 灵敏度比较

	文献[9]	本文	提升
灵敏度	0.481	0.603	12.2%

表 4 平均假阳性比较

	文献[9]	本文	降低
灵敏度=0.1	0.55	0.14	74.5%
灵敏度=0.2	3.25	0.84	74.2%
灵敏度=0.3	4.72	1.48	68.6%
灵敏度=0.4	10.38	4.62	55.5%
平均值	4.725	1.77	62.5%

结束语 本文主要针对微小动脉瘤的模板匹配算法和特

征提取算法做了相应的研究,由于时间有限,本文还存在一些问题需要继续深入研究和改进,例如,为了降低算法的计算复杂度,文中 DMPT 模板的陡峭度系数选用了折中取值的方法,该方法虽然能最大程度地使折中函数满足评判条件的要求,但仍与变化不定的微小动脉瘤形态产生矛盾。如何选用更为合适的陡峭度系数,有待进一步研究。

参考文献

- [1] Fong D S, Aiello L, Gardner T W, et al. Diabetic retinopathy [J]. *Diabetes Care*, 2003, 26: 226-229
- [2] Baudoin C E, Lay B J, Klein J C. Automatic detection of microaneurysms in diabetic fluorescein angiographies [J]. *Revue D'épidémiologie et de Sante Publique*, 1984, 32: 254-261
- [3] Gardner G G, Keating D, Williamson T H, et al. Automatic detection of diabetic retinopathy using an artificial neural network; a screening tool [J]. *Br. J. Ophthalmol*, 1996, 80: 940-944
- [4] Niemeijer M, Ginneke B, Mizutani V A, et al. Retinopathy online challenge; automatic detection of microaneurysms in digital color fundus photographs [J]. *IEEE transactions on medical imaging*, 2010, 29: 185-195
- [5] Fleming A D, Philip S, Goatman K A, et al. Automated microaneurysms detection using local contrast normalization and local vessel detection [J]. *IEEE transactions on medical imaging*, 2006, 25: 1223-1232
- [6] Walter T, Massin P, Erginay A, et al. Automatic detection of microaneurysms in color fundus images [J]. *Medical image analysis*, 2007, 11(6): 555-566
- [7] Niemeijer M, van Ginneken B, Staal J, et al. Automatic detection of red lesions in digital color fundus photographs [J]. *IEEE transactions on medical imaging*, 2005, 24(5): 584-592
- [8] Sinthanayothin C, Boyce J F, Williamson T H, et al. Automated detection of diabetic retinopathy on digital fundus images [J]. *Diabetic medicine*, 2002, 19: 105-112
- [9] Bob Z, Qian W, Jane X Y, et al. Detection of microaneurysms using multi-scale correlation coefficients [J]. *Pattern Recognition*, 2010, 43: 2237-2248
- [10] Bob Z, Lin Z, Lei Z, et al. Retinal vessel extraction by matched filter with first-order derivative of Gaussian [J]. *Computers in biology and medicine*, 2010, 40: 438-445
- [11] Gonzalez R C, Woods R E. *Digital Image Processing (第二版)* [M]. 阮秋琦, 阮宇智, 等译. 北京: 电子工业出版社, 2007: 496-498
- [12] Yuen S Y, Chow C K. A Genetic Algorithm That Adaptively Mutates and Never Revisits [J]. *IEEE transactions on evolutionary computation*, 2009, 13: 454-472

(上接第 263 页)

- [7] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories[C]// *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2006: 2169-2178
- [8] Yang J, Yu K, Gong Y, et al. Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. 2009;

1794-1801

- [9] Russakovsky O, Lin Y, Yu K, et al. Object-centric spatial pooling for image classification[C]// *European Conference on Computer Vision*. 2012
- [10] Yao Bang-peng, Li Fei-fei. Recognizing Human-Object Interactions in Still Images by Modeling the Mutual Context of Objects and Human Poses[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(9): 1691-1703