



(12) 发明专利

(10) 授权公告号 CN 110059741 B

(45) 授权公告日 2022.12.02

(21) 申请号 201910298278.4

G06V 10/80 (2022.01)

(22) 申请日 2019.04.15

G06V 10/774 (2022.01)

(65) 同一申请的已公布的文献号

G06V 10/764 (2022.01)

申请公布号 CN 110059741 A

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

(43) 申请公布日 2019.07.26

(56) 对比文件

(73) 专利权人 西安电子科技大学

CN 108985316 A, 2018.12.11

地址 710071 陕西省西安市太白南路2号

CN 108985377 A, 2018.12.11

(72) 发明人 高大化 马欣睿 刘丹华 张中强

CN 108830243 A, 2018.11.16

石光明 冯依好 牛毅 林杰

US 2019034800 A1, 2019.01.31

王晓甜

审查员 牛力

(74) 专利代理机构 陕西电子工业专利中心

61205

专利代理师 程晓霞 王品华

(51) Int. Cl.

G06V 10/82 (2022.01)

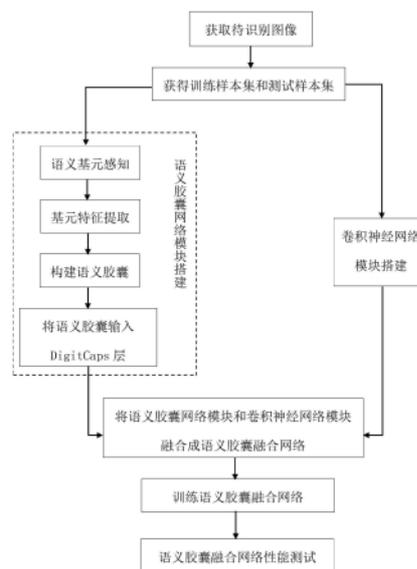
权利要求书2页 说明书10页 附图3页

(54) 发明名称

基于语义胶囊融合网络的图像识别方法

(57) 摘要

本发明公开了一种基于语义胶囊融合网络的图像识别方法,通过将神经网络与图像语义信息进行融合,解决了现有技术中神经网络依赖大量样本训练、网络结构复杂、可解释性差、空占度高、能耗高的问题。实现的步骤是:获取待识别图像;获得训练样本集和测试样本集;语义胶囊网络模块搭建;卷积神经网络模块搭建;将语义胶囊网络模块和卷积神经网络模块融合成语义胶囊融合网络;训练语义胶囊融合网络;语义胶囊融合网络性能测试。本发明提出了语义胶囊的概念,设计了语义胶囊融合网络。本发明具有在少量训练样本上识别准确率更高、网络参数量更少、网络可解释性强、能耗更低的优点。可用于大量简单图像的识别处理。



1. 一种基于语义胶囊融合网络的图像识别方法,其特征在於,包括有如下步骤:

(1) 获取待识别图像:获取待识别图像并确认其类别标签;

(2) 获得训练样本集和测试样本集:在待识别图像中随机选取80%的图像以及对应的类别标签,组成训练样本集,将剩余20%的图像和对应的类别标签,组成测试样本集;

(3) 语义胶囊网络模块搭建:输入待识别图像,对图像进行语义基元感知,将感知获得的语义基元进行特征提取,将获取的特征值整合成向量的形式即语义胶囊,然后将语义胶囊输入胶囊网络CapsNet中的最高层DigitCaps层,通过动态路由机制进行决策;语义胶囊网络模块的输出为 $N \times 1$ 的向量,向量值范围为 $(0, 1)$,对应 N 个分类类别;

(4) 卷积神经网络模块搭建:搭建一个六层的卷积神经网络,其结构依次为:第一卷积层,第一最大池化层,第二卷积层,第二最大池化层,第一全连接层,第二全连接层;将待识别图像输入搭建好的卷积神经网络模块,得到的卷积神经网络模块的输出为 $N \times 1$ 的向量,向量值范围为 $(0, 1)$,对应 N 个分类类别;

(5) 将语义胶囊网络模块和卷积神经网络模块融合成语义胶囊融合网络:设计一个线性函数,将语义胶囊网络的输出向量与卷积神经网络的输出向量同时输入线性函数进行计算,计算结果即为语义胶囊融合网络的输出;

(6) 训练语义胶囊融合网络:将训练样本集中的每张图像输入搭建好的语义胶囊融合网络,利用损失函数公式计算语义胶囊融合网络的输出值与训练样本实际类别标签之间的损失值,按照极小化误差的方法反向传播调整语义胶囊融合网络的参数,直到训练样本集所有图像的损失值都对语义胶囊融合网络的参数进行了调整,重复本步骤1万次,得到训练好的语义胶囊融合网络;

(7) 语义胶囊融合网络性能测试:将测试样本集输入到训练好的语义胶囊融合网络中,得到最终测试样本集图像的识别结果。

2. 根据权利要求1所述的基于语义胶囊融合网络的图像识别方法,其特征在於,步骤3中所述的语义胶囊网络模块搭建,具体包括有如下步骤

(3a) 语义基元感知:对输入的待识别图像感知语义基元;

(3b) 基元特征提取:每个语义基元都通过一个8维的向量表达其特征,这些语义基元的特征向量则为语义胶囊,对于封闭轮廓基元,选择中心点坐标,长短轴长度以及相对于水平轴的旋转角度构建特征向量,对于直线段基元,选择起点,中点和终点横纵坐标,长度以及相对于水平轴的旋转角度构建特征向量,对于弧线段基元,选择起点,中点和终点坐标构建特征向量;

(3c) 构建语义胶囊:对于每张图像,感知4个直线段基元,2个封闭轮廓基元和2个弧线段基元,则这8个基元的8维特征向量即为8个语义胶囊,每张图像一共有8个语义胶囊,语义胶囊的每一维都具有特定的含义即语义;

(3d) 将语义胶囊输入digitCaps层:将获得的语义胶囊输入CapsNet中的digitCaps层,通过 8×16 维权重矩阵将8维输入空间映射到16维输出空间,由此输出 N 个16维向量,使用非线性函数squashing

$$\mathbf{v}_j = \frac{\|\mathbf{s}_j\|^2 \mathbf{s}_j}{1 + \|\mathbf{s}_j\|^2 \|\mathbf{s}_j\|}$$

确保短向量的长度能够缩短到几乎等于零,长向量的长度压缩到接近但不超过1;其中 \mathbf{v}_j 为第j个胶囊的输出向量, \mathbf{s}_j 为上一层所有胶囊输出到当前层胶囊j的向量加权和;语义胶囊网络输出为 $N \times 1$ 的向量,对应N个分类类别。

3. 根据权利要求1所述的基于语义胶囊融合网络的图像识别方法,其特征在于,步骤(3a)中所述语义基元感知的步骤如下:

3a1) 图像细化:通过图像细化步骤获取图像骨架,便于更高效的完成基元感知;

3a2) 感知封闭轮廓基元:在细化图像中获取封闭轮廓基元,封闭轮廓包括圆形和椭圆形;若图像中有封闭轮廓,则将感知到的封闭轮廓从对应的细化图像中移除,若没有,则保持细化图像不变;

3a3) 模板匹配:感知直线段和弧线段基元;通过与角度,大小,长度不等的各种形态的标准直线段和弧线段图像进行模板匹配,选取匹配值最高的标准图像的图像类型即为感知到的基元类型;对于每张图像,感知4个直线段基元,2个封闭轮廓基元和2个弧线段基元。

4. 根据权利要求1所述的基于语义胶囊融合网络的图像识别方法,其特征在于,步骤5中所述的将语义胶囊网络模块和卷积神经网络模块融合成语义胶囊融合中的线性函数如下:

用 p_i 表示语义胶囊网络第i个类别的输出值,用 q_i 表示语义胶囊网络第i个类别的输出值,设计一个线性函数 $f(\cdot)$ 来计算两个网络的融合结果 o_i :

$$o_i = f(p_i, q_i) = p_i \cdot \beta_i + q_i \cdot \gamma_i$$

再对 o_i 进行softmax即得到网络对各类别的预测概率,其中, β_i, γ_i 代表两个模块的预测概率结果融合参数,通过网络训练学习得到, $i=0, 1, \dots, N, N$ 表示分类的总类别数。

5. 根据权利要求1所述的基于语义胶囊融合网络的图像识别方法,其特征在于,步骤6中所述的训练语义胶囊融合网络中利用的损失函数如下:

网络的损失函数为边缘损失函数:

$$L_i^{scf} = H_i \max(0, 0.9 - o_i)^2 + 0.5 \times (1 - H_i) \max(0, o_i - 0.1)^2$$

$$L^{scf} = \sum_{i=0}^N L_i^{scf}$$

其中, o_i 是语义胶囊融合网络的输出向量, H_i 是样本标签经过one-hot后得到的向量,向量各维度取值范围为(0, 1), $i=0, 1, \dots, N, N$ 表示分类的总类别数。

基于语义胶囊融合网络的图像识别方法

技术领域

[0001] 本发明属于计算机视觉技术领域,更进一步涉及图像识别,具体是一种基于语义胶囊融合网络的图像识别方法。主要用于针对大量的简单图像识别。

背景技术

[0002] UCLA的朱松纯教授等人在其发表的论文“Learning AND-OR Templates for Object Recognition and Detection”中利用图像模板结合与或结点构建模型实现目标识别与探测。它是一个层次化的可重构图像模板,通过概率的方法学习目标的结构语义组成。

[0003] 2017年Hinton和他的团队发表的文章“Dynamic Routing Between Capsules”里介绍了胶囊网络,胶囊层之间反应的是部分与整体之间的关系,而其中的激活神经元由标量变为可解释的向量,向量每一维反映图像中特定实体的各种属性,如姿态(位置,大小,方向)纹理,色相等。处理了简单对象和复杂对象之间的空间层次表示问题。这使得CapsNet在MNIST和识别高度重叠的数字上获得了优异的性能。

[0004] 以上几种研究都在模仿人类的决策机制和视觉识别过程,但是,由于缺失类似人类的利用先验知识的高效机制,相应的神经网络往往需要大量的训练样本数据和相当长的训练过程,由此变得十分复杂,并且效率低,能耗高。

发明内容

[0005] 本发明的目的在于克服上述现有技术的不足,提出一种更少样本,更快速度,解释性更强的基于语义胶囊融合网络的图像识别方法。

[0006] 本发明是一种基于语义胶囊融合网络的图像识别方法,其特征在於,包括有如下步骤:

[0007] (1) 获取待识别图像:获取待识别图像并确认其类别标签;

[0008] (2) 获得训练样本集和测试样本集:在待识别图像中随机选取80%的图像以及对应的类别标签,组成训练样本集,将剩余20%的图像和对应的类别标签,组成测试样本集;

[0009] (3) 语义胶囊网络模块搭建:输入待识别图像,对图像进行语义基元感知,将感知获得的语义基元进行特征提取,将获取的特征值整合成向量的形式即语义胶囊,然后将语义胶囊输入胶囊网络CapsNet中的最高层DigitCaps层,通过动态路由机制进行决策;语义胶囊网络模块的输出为 $N \times 1$ 的向量,向量值范围为(0,1),对应N个分类类别;

[0010] (4) 卷积神经网络模块搭建:搭建一个六层的卷积神经网络,其结构依次为:第一卷积层,第一最大池化层,第二卷积层,第二最大池化层,第一全连接层,第二全连接层;将待识别图像输入搭建好的卷积神经网络模块,得到的卷积神经网络模块的输出为 $N \times 1$ 的向量,向量值范围为(0,1),对应N个分类类别;

[0011] (5) 将语义胶囊网络模块和卷积神经网络模块融合成语义胶囊融合网络:设计一个线性函数,将语义胶囊网络的输出向量与卷积神经网络的输出向量同时输入线性函数进行计算,计算结果即为语义胶囊融合网络的输出;

[0012] (6) 训练语义胶囊融合网络:将训练样本集中的每张图像输入搭建好的语义胶囊融合网络,利用损失函数公式计算语义胶囊融合网络的输出值与训练样本实际类别标签之间的损失值,按照极小化误差的方法反向传播调整语义胶囊融合网络的参数,直到训练样本集所有图像的损失值都对语义胶囊融合网络的参数进行了调整,重复本步骤1万次,得到训练好的语义胶囊融合网络;

[0013] (7) 语义胶囊融合网络性能测试:将测试样本集输入到训练好的语义胶囊融合网络中,得到最终测试样本集图像的识别结果。

[0014] 本发明能够将人类知识和经验与传统神经网络进行结合,以此达到更快的训练速度,更少的训练样本以及更强的可解释性。

[0015] 本发明与现有方法相比具有如下优点:

[0016] 第一,由于本发明提出了语义基元的概念,通过感知图像语义基元提取图像语义特征信息,这一方法相比胶囊网络中使用卷积层提取特征的传统方法,本发明方法提取到的图像语义特征具有特定的语义含义,即为可解释的特征,克服了传统神经网络提取出的特征无含义,无法解释从而导致网络识别过程无可解释性的问题,使得图像识别过程具有可解释性。

[0017] 第二,由于本发明将人类先验知识与神经网络进行融合,提前融入语义信息,将图像的大量特征信息提前赋予网络,大大减少了网络通过学习的方法所需提取的特征量,克服了传统神经网络需要通过大量训练样本训练大量参数来学习特征的问题,精简了网络结构,减少了网络所需的参数量以及训练网络所需的样本量,大大提升了网络的训练速度和识别效率,大程度得节省了空间占用度,降低了能耗。

[0018] 第三,由于本发明学习人类决策和视觉识别机制的特点,使得网络的学习和识别更加贴合人脑学习的过程,在一定程度上克服了深度学习黑盒子模型的弱点,网络的可操控性得以大大提高。

附图说明

[0019] 图1是本发明的结构流程图;

[0020] 图2是本发明的图像识别方法模型结构示意图;

[0021] 图3是本发明的相同参数量不同训练样本量下本方法与原胶囊网络识别准确率对比图;

[0022] 图4是本发明的相同训练样本量不同网络参数量下本发明与原胶囊网络识别准确率对比图。

具体实施方式

[0023] 下面结合附图,对本发明的步骤作详细描述。

[0024] 实施例1

[0025] 目前,有很多深度学习算法,例如CNN, GAN, CapsNet等都在计算机视觉领域做出了显著贡献。然而,这类神经网络的优良性能均依托于大量训练样本数据,这种大数据驱动的方法有很多弊端,如:需求大量存储空间,能耗高,获取数据具有难度以及需要大量的人工数据标签注释工作。此外,鉴于神经网络的黑盒模型特点,很难对网络学习到的特征进行解

释,也难以对网络结构和参数的设定进行解释,因此无法预测网络性能。基于以上两点,需要对当前模型进行改进。客观上急需寻找一种方法,可以将人类知识和经验与传统神经网络进行结合,以此达到更快的训练速度,更少的训练样本以及更强的可解释性。

[0026] 本发明正是在此领域进行尝试和创新,提出一种基于语义胶囊融合网络的图像识别方法,参见图1,包括有如下步骤:

[0027] (1) 获取待识别图像:获取待识别图像并确认其类别标签。

[0028] (2) 获得训练样本集和测试样本集:在待识别图像中随机选取80%的图像以及对应的类别标签,组成训练样本集,将剩余20%的图像和对应的类别标签,组成测试样本集。

[0029] (3) 语义胶囊网络模块搭建:输入待识别图像,对图像进行语义基元感知,对感知获得的语义基元进行特征提取,将获取的特征值整合成向量的形式即语义胶囊,然后将语义胶囊输入胶囊网络CapsNet中的最高层DigitCaps层,通过动态路由机制进行决策,完成语义胶囊网络模块的搭建。语义胶囊网络模块的输出为 $N \times 1$ 的向量,向量值范围为(0, 1),对应 N 个分类类别。

[0030] (4) 卷积神经网络模块搭建:搭建一个六层的卷积神经网络,其结构依次为:第一卷积层,第一最大池化层,第二卷积层,第二最大池化层,第一全连接层,第二全连接层,完成卷积神经网络模块的搭建。将待识别图像输入搭建好的卷积神经网络模块,得到的卷积神经网络模块的输出为 $N \times 1$ 的向量,向量值范围为(0, 1),对应 N 个分类类别。

[0031] (5) 将语义胶囊网络模块和卷积神经网络模块融合成语义胶囊融合网络:为了更大程度发挥语义胶囊网络模块和卷积神经网络模块各自在图像识别目的中的作用,设计一个线性函数,将语义胶囊网络的输出向量与卷积神经网络的输出向量同时输入线性函数进行计算,计算结果即为语义胶囊融合网络的输出,通过训练线性函数中的参数,使得语义胶囊网络模块和卷积神经网络模块两模块的融合更加灵活机动,融合效果更好。

[0032] (6) 训练语义胶囊融合网络:将训练样本集中的每张图像输入搭建好的语义胶囊融合网络,利用损失函数公式计算语义胶囊融合网络的输出值与训练样本实际类别标签之间的损失值,按照极小化误差的方法反向传播调整语义胶囊融合网络的参数,直到训练样本集所有图像的损失值都对语义胶囊融合网络的参数进行了调整,完成一次参数学习,重复本步骤1万次,得到训练好的语义胶囊融合网络。

[0033] (7) 语义胶囊融合网络性能测试:将测试样本集输入到训练好的语义胶囊融合网络中,得到最终测试样本集图像的识别结果。

[0034] 实现本发明的技术思路是,通过图像基元感知和特征提取构建基元胶囊,结合CapsNet中的DigitCaps层构成语义胶囊网络模块,再辅助以卷积神经网络模块进行图像识别。整体的网络结构称为语义胶囊融合网络,在这个网络中,卷积神经网络模块通过学习表达语义基元难以描述的特征辅助语义胶囊网络模块从而获得更高的识别性能。

[0035] 本发明通过搭建语义胶囊网络模块,搭建卷积神经网络模块,再将语义胶囊网络模块和卷积神经网络模块进行融合构建一个语义胶囊融合网络,通过训练语义胶囊融合网络确定网络参数并进行性能测试,给出了一个基于语义胶囊融合网络的图像识别的新的技术方案,能够在更少的训练样本,更简单的网络结构下获得更优的性能,具有更好的可解释性,更快的速度,更小的空间占有度和更低的能耗。

[0036] 实施例2

[0037] 基于语义胶囊融合网络的图像识别方法同实施例1,步骤3中所述的语义胶囊网络模块搭建,具体包括有如下步骤

[0038] (3a) 语义基元感知:对输入的待识别图像感知语义基元,用语义基元的形式描述图像,获取图像的语义信息。

[0039] (3b) 基元特征提取:每个语义基元都通过一个8维的向量表达其特征。对于封闭轮廓基元,选择中心点坐标,长短轴长度以及相对于水平轴的旋转角度构建特征向量。对于直线段基元,选择起点,中点和终点横纵坐标,长度以及相对于水平轴的旋转角度构建特征向量。对于弧线段基元,选择起点,中点和终点坐标构建特征向量。向量不足8维时,剩余维度处补零。通过本步骤,语义基元特征的每一维都将具有特定的语义信息。

[0040] (3c) 构建语义胶囊:受到胶囊网络CapsNet提出胶囊神经元的启发,本发明提出语义胶囊的概念,对特征向量的每一维赋予特定的语义特征信息。对于每张图像,感知4个直线段基元,2个封闭轮廓基元和2个弧线段基元。则这8个基元的8维特征向量即为8个语义胶囊,每张图像一共获得8个语义胶囊,语义胶囊的每一维都具有特定的含义即语义,由于通过8个语义胶囊描述一幅图像的方法,每张图像的大量语义特征信息都得到了表达,因此仅通过这些语义信息,大部分图像就可得到正确的识别结果。

[0041] (3d) 将语义胶囊输入DigitCaps层:将获得的8维语义胶囊输入CapsNet中的DigitCaps层,通过 8×16 维权重矩阵将8维输入空间映射到16维输出空间,由此输出N个16维向量。通过动态路由机制并使用非线性函数squashing

$$[0042] \quad \mathbf{v}_j = \frac{\|\mathbf{s}_j\|^2 \mathbf{s}_j}{1 + \|\mathbf{s}_j\|^2 \|\mathbf{s}_j\|}$$

[0043] 确保短向量的长度能够缩短到几乎等于零,长向量的长度压缩到接近但不超过1。其中 \mathbf{v}_j 为第j个胶囊的输出向量,其每一维数值的取值范围为(0,1),表示对各个类别判断概率,其长度取值范围为(0,1), \mathbf{s}_j 为上一层所有胶囊输出到当前层胶囊j的向量加权和。语义胶囊网络输出为 $N \times 1$ 的向量,对应N个分类类别。

[0044] 本发明中的语义胶囊网络通过感知各类语义基元并获取特征的方式成功替换掉胶囊网络中通过卷积层提取特征的步骤,仅仅需要进行DigitCaps层的搭建和参数训练即可简单完成识别过程,相比胶囊网络,大大减少了参数训练量,并且语义胶囊的引入使得胶囊的可解释性增强。语义胶囊网络将图像识别过程从数据驱动转为知识驱动,不同于传统的深度学习黑盒模型,使得识别过程可以为人类所理解。另外,语义胶囊的引入,使得识别网络对训练样本量的依赖降低,在更少训练样本的情况下会获得更高的识别准确率。

[0045] 实施例3

[0046] 基于语义胶囊融合网络的图像识别方法同实施例1-2,步骤(3a)中所述语义基元感知的步骤如下:

[0047] 3a1) 图像细化:通过图像细化步骤获取图像骨架,使得图像结构更加清晰,便于更高效的完成基元感知。

[0048] 3a2) 感知封闭轮廓基元:在细化图像中获取封闭轮廓基元。封闭轮廓包括圆形轮廓和椭圆形轮廓;若图像中有封闭轮廓,则将感知到的封闭轮廓从对应的细化图像中移除,

若没有,则保持细化图像不变。

[0049] 3a3) 模板匹配:感知直线段和弧线段基元;通过与角度,大小,长度不等的各种形态的标准直线段和弧线段图像进行模板匹配,选取匹配值最高的标准图像的图像类型作为感知到的基元类型。本例中,对于每张图像,感知4个直线段基元,2个封闭轮廓基元和2个弧线段基元。在实际操作中,根据所识别图像的结构特征,可以对此参数进行调整。

[0050] 本发明感知图像中的语义基元,能够清晰地利用感知到的语义基元准确的重构出原图像,损失的图像信息较少,对于准确识别图像类别起到了必不可少的作用。另外,利用本发明,可以准确并全面的获取图像的语义信息,成功的将图像在计算机中单一利用RGB方式存储的信息转化为语义信息,本发明这种新型编码方式更加符合人类大脑在认知事物中的编码过程。

[0051] 实施例4

[0052] 基于语义胶囊融合网络的图像识别方法同实施例1-3,步骤5中所述的将语义胶囊网络模块和卷积神经网络模块融合成语义胶囊融合中的线性函数如下:

[0053] 用 p_i 表示语义胶囊网络第 i 个类别的输出值,用 q_i 表示语义胶囊网络第 i 个类别的输出值,设计一个线性函数 $f(\cdot)$ 来计算语义胶囊网络模块和卷积神经网络模块的融合结果 o_i :

$$[0054] \quad o_i = f(p_i, q_i) = p_i \cdot \beta_i + q_i \cdot \gamma_i$$

[0055] 再对两个网络的融合结果 o_i 进行softmax操作即得到语义胶囊融合网络对待识别图像各类别的预测概率。其中, β_i, γ_i 代表两个模块的预测概率结果融合参数,通过网络参数训练学习得到,其中,参数 β_i 是语义胶囊网络模块的融合参数,参数 γ_i 是卷积神经网络模块的融合参数, $i=0, 1, \dots, N, N$ 表示分类的总类别数。

[0056] 本发明该线性函数的目的在于发扬语义胶囊网络模块和卷积神经网络模块在图像识别中所起的积极作用,同时抑制两模块在图像识别中的消极作用,最大程度的提升语义胶囊融合网络的整体网络性能,相比于提前人工固定参数的融合方式,本发明通过学习确定参数能够更加准确,灵活性更高,由此,本发明方法可以获得更强的网络识别性能,使得识别性能达到最大化。

[0057] 实施例5

[0058] 基于语义胶囊融合网络的图像识别方法同实施例1-4,步骤6中所述的训练语义胶囊融合网络中利用的损失函数如下:

[0059] 本发明方法将损失函数设定为边缘损失函数:

$$[0060] \quad L_i^{scf} = H_i \max(0, 0.9 - o_i)^2 + 0.5 \times (1 - H_i) \max(0, o_i - 0.1)^2$$

$$[0061] \quad L^{scf} = \sum_{i=0}^N L_i^{scf}$$

[0062] 此函数的输出值表示网络判定结果与图像正确标签之间的损失值。其中, o_i 是语义胶囊融合网络的 N 维输出向量,向量每一维数值表示对于该维度对应类别的预测概率。 H_i 是样本标签经过one-hot后得到的向量,向量维度为 N ,向量各维度取值范围为 $(0, 1), i=0, 1, \dots, N, N$ 表示分类的总类别数。本发明将胶囊网络的边缘损失函数用于语义胶囊融合网络的损失函数的表达中,解决了语义胶囊融合网络损失函数的设定,更加匹配网络结构。

[0063] 本发明是基于语义胶囊融合网络的图像识别方法。本发明可将人类知识与神经网络进行结合,达到在更少的训练样本,更高的训练速度和更强的网络可解释性下达到更好的图像识别效果。

[0064] 下面给出一个更加详尽的例子,对本发明进一步说明。

[0065] 实施例6

[0066] 基于语义胶囊融合网络的图像识别方法同实施例1-5,参照附图1,本发明实现的步骤如下。

[0067] 步骤1,获取待识别图像

[0068] 获取待识别图像并确认其类别标签。

[0069] 步骤2,获得训练样本集和测试样本集

[0070] 在待分类图像中随机选取80%的图像以及对应的类别标签,组成训练样本集,将剩余20%的图像和对应的类别标签,组成测试样本集。

[0071] 步骤3,语义胶囊网络搭建

[0072] (3a) 语义基元感知:对输入的待识别图像感知语义基元,用语义基元的形式描述图像,获取图像的语义信息。

[0073] 3a1) 图像细化:通过图像细化步骤获取图像骨架,使得图像结构更加清晰,便于更高效的完成基元感知。

[0074] 3a2) 感知封闭轮廓基元:在细化图像中获取封闭轮廓基元。封闭轮廓包括圆形轮廓和椭圆形轮廓;若图像中有封闭轮廓,则将感知到的封闭轮廓从对应的细化图像中移除,若没有,则保持细化图像不变。

[0075] 3a3) 模板匹配:感知直线段和弧线段基元;通过与角度,大小,长度不等的各种形态的标准直线段和弧线段图像进行模板匹配,选取匹配值最高的标准图像的图像类型做为感知到的基元类型;对于每张图像,感知4个直线段基元,2个封闭轮廓基元和2个弧线段基元。

[0076] 本步骤方法感知图像中的语义基元,能够清晰地利用感知到的语义基元准确的重构出原图像,损失的图像信息较少,对于准确识别图像类别起到了必不可少的作用。另外,利用本步骤的方法,可以准确并全面的获取图像的语义信息,成功的将图像在计算机中单一利用RGB方式存储的信息转化为语义信息,这种新型编码方式更加符合人类大脑在认知事物中的编码过程。

[0077] (3b) 基元特征提取:每个语义基元都通过一个8维的向量表达其特征。这些语义基元的特征向量则为语义胶囊。对于封闭轮廓基元,选择中心点坐标,长短轴长度以及相对于水平轴的旋转角度构建特征向量。对于直线段基元,选择起点,中点和终点横纵坐标,长度以及相对于水平轴的旋转角度构建特征向量。对于弧线段基元,选择起点,中点和终点坐标构建特征向量,不足8位处补零。通过本步骤,语义基元特征的每一维都将具有特定的语义信息。

[0078] (3c) 构建语义胶囊:受到胶囊网络CapsNet提出胶囊神经元的启发,本发明提出语义胶囊的概念,对特征向量的每一维赋予特定的语义特征信息。对于每张图像,感知4个直线段基元,2个封闭轮廓基元和2个弧线段基元。则这8个基元的8维特征向量即为8个语义胶囊,每张图像一共获得8个语义胶囊,语义胶囊的每一维都具有特定的含义即语义,由于通

过8个语义胶囊描述一幅图像的方法,每张图像的大量语义特征信息都得到了表达,因此仅通过这些语义信息,大部分图像就可得到正确的识别结果。

[0079] (3d) 将语义胶囊输入DigitCaps层:将获得的8维语义胶囊输入CapsNet中的DigitCaps层,通过 8×16 维权重矩阵将8维输入空间映射到16维输出空间,由此输出N个16维向量。通过动态路由机制并使用非线性函数squashing

$$[0080] \quad \mathbf{v}_j = \frac{\|\mathbf{s}_j\|^2 \mathbf{s}_j}{1 + \|\mathbf{s}_j\|^2 \|\mathbf{s}_j\|}$$

[0081] 确保短向量的长度能够缩短到几乎等于零,长向量的长度压缩到接近但不超过1。其中 \mathbf{v}_j 为第j个胶囊的输出向量,其每一维数值的取值范围为(0,1),表示对各个类别判断概率,其长度取值范围为(0,1), \mathbf{s}_j 为上一层所有胶囊输出到当前层胶囊j的向量加权和。语义胶囊网络输出为 $N \times 1$ 的向量,对应N个分类类别。

[0082] 该语义胶囊网络通过感知各类语义基元并获取特征的方式成功替换掉胶囊网络中通过卷积层提取特征的步骤,仅仅需要进行DigitCaps层的搭建和参数训练即可简单完成识别过程,相比胶囊网络,大大减少了参数训练量,并且语义胶囊的引入使得胶囊的可解释性增强。语义胶囊网络将图像识别过程从数据驱动转为知识驱动,不像传统的深度学习黑盒模型,这一方法使得识别过程可以为人类所理解。

[0083] 步骤4,卷积神经网络搭建

[0084] 搭建一个六层的卷积神经网络,其结构依次为:第一卷积层,第一最大池化层,第二卷积层,第二最大池化层,第一全连接层,第二全连接层。将图像输入搭建好的卷积神经网络。设置胶囊网络中各层的参数:将第一卷积层的特征映射图设置为64个,卷积核大小为 5×5 像素,步长为1个像素;将第二卷积层的特征映射图设置为128个,卷积核大小为 5×5 像素,步长为1个像素;将第一全连接层神经元个数设置1024个,第二全连接层神经元个数设置为N。本配置方式是网络的初始配置结构。

[0085] 得到的卷积神经网络的输出为 $N \times 1$ 的向量,向量值范围为(0,1),对应N个分类类别。

[0086] 鉴于卷积神经网络模块在语义胶囊融合网络中的辅助识别作用,本步骤仅仅搭建了结构较为简单,参数量较少的卷积神经网络,用更少的网络参数量达到更高的识别准确率,从而减少不必要的时间空间损耗。

[0087] 步骤5,将语义胶囊网络和卷积神经网络融合成语义胶囊融合网络

[0088] 用 p_i 表示语义胶囊网络第i个类别的输出值,用 q_i 表示语义胶囊网络第i个类别的输出值,设计一个线性函数 $f(\cdot)$ 来计算两个网络的融合结果 o_i :

$$[0089] \quad o_i = f(p_i, q_i) = p_i \cdot \beta_i + q_i \cdot \gamma_i$$

[0090] 再对 o_i 进行softmax操作即得到网络对各类别的预测概率。其中, β_i, γ_i 代表两个模块的预测概率结果融合参数,通过网络参数训练学习得到, $i=0,1,\dots,N$,N表示分类的总类别数。

[0091] 该线性函数的目的在于发扬语义胶囊网络模块和卷积神经网络模块在图像识别中所起的积极作用,同时抑制两模块在图像识别中的消极作用,最大程度的提升语义胶囊

融合网络的整体网络性能,相比于提前人工固定参数的融合方式,通过学习确定参数能够更加准确,灵活性更高,由此,本发明方法可以获得更强的网络识别性能。

[0092] 步骤6,语义胶囊融合网络参数训练

[0093] 受到胶囊网络的启发并且更好的匹配网络结构,本发明方法将损失函数设定为边缘损失函数:

$$[0094] \quad L_i^{scf} = H_i \max(0, 0.9 - o_i)^2 + 0.5 \times (1 - H_i) \max(0, o_i - 0.1)^2$$

$$[0095] \quad L^{scf} = \sum_{i=0}^N L_i^{scf}$$

[0096] 此函数的输出值表示网络判定结果与图像正确标签之间的损失值。其中, o_i 是语义胶囊融合网络的N维输出向量,向量每一维数值表示对于该维度对应类别的预测概率。 H_i 是样本标签经过one-hot后得到的向量,向量维度为N,向量各维度取值范围为(0,1), $i=0, 1, \dots, N$, N表示分类的总类别数。

[0097] 步骤7,语义胶囊融合网络性能测试

[0098] 将测试样本集输入到训练好的语义胶囊融合网络中,得到最终识别结果。

[0099] 参照本发明的图像识别方法模型结构示意图的附图2,对本发明图像识别方法进行进一步的说明。

[0100] 本发明的图像识别结果主要融合两个识别预测值得到,其中一个识别预测值是使用语义胶囊网络模块获取预测值,另一个识别预测值是由普通卷积层、最大池化层,全连接层组成的简单卷积神经网络模块输出的预测值。对两个预测值使用步骤5中的融合公式进行计算得到语义胶囊融合网络的图像识别预测结果。

[0101] 综上所述:本发明采用基于语义胶囊融合网络的图像识别方法,克服现有神经网络需要依赖大量样本、网络可解释性差、网络结构复杂,能耗高的问题,有效将人类先验知识融入传统神经网络中,提高数据的利用率,获得更强的网络可解释性,减少网络结构的复杂度和参数量,提高了网络训练的速度,降低了能耗,加快了网络的识别效率。

[0102] 下面结合仿真实验对本发明的技术效果做进一步说明。

[0103] 实施例7

[0104] 基于语义胶囊融合网络的图像识别方法同实施例1-6,

[0105] 仿真实验条件:本发明的仿真实验的硬件平台为:Inter Xeon E5-2678M CPU@2.50GHz、128GB RAM。本发明的仿真实验的软件使用tensorflow。

[0106] 仿真内容:

[0107] 本发明的仿真实验是采用本发明和现有技术胶囊网络分别对美国手写数字图像集MNIST进行识别。其中,训练集由50000张图像和对应的类别标签组成,测试集由10000张图像和对应的类别标签组成。共计10类手写数字图像。

[0108] 本发明用到的现有技术对比识别方法为Sara Sabour等人在“Dynamic Routing Between Capsules, Neural Information Processing Systems 2017”中提出的图像识别方法,简称原胶囊网络识别方法。

[0109] 仿真结果分析

[0110] 为了验证本发明的高效性和良好的分类性能,采用以下三个现有指标来评价本发

明仿真实验的结果：第一个评价指标是本发明中的网络 and 原胶囊网络使用的训练样本量，训练样本量越少，说明方法越高效。第二个评价指标是识别准确率，表示正确识别的样本占所有样本的比例，该值越大，说明识别效果越好，对每张图像来说，正确识别是指模型预测类别与图像标签相同。第三个评价指标是本发明的网络 and 原胶囊网络所需训练的网络参数量，该值越小，说明方法计算量小、效率越高。

[0111] 参见图3，图3为相同参数量不同训练样本量下本方法与原胶囊网络识别准确率对比图。其横坐标表示训练样本量，纵坐标表示识别准确率。其中，虚线表示原胶囊网络在不同训练样本量下识别准确率的变化情况，实线表示语义胶囊融合网络在不同训练样本量下识别准确率的变化情况。从图3可以看到原胶囊网络在减少训练样本量的情况下，识别准确率下降明显，而本发明方法在减少训练样本量的情况下，识别准确率可以基本保持。当训练样本量降至2000张时，本发明方法相比原胶囊网络的识别准确率高出将近1.2%。

[0112] 表1. 相同参数量不同训练样本量下本发明与原胶囊网络识别准确率对比表

训练样本数据总量	本发明方法识别准确率	原胶囊网络识别准确率
10000张	99.1313%	98.8682%
5000张	99.0202%	98.5377%
4000张	98.7778%	98.2772%
3000张	98.6566%	97.8766%
2000张	98.4848%	97.3357%

[0114] 表1为本发明方法与原胶囊网络在网络参数量保持一致时，分别在不同大小的训练样本量上训练后，测试集识别准确率的比较情况，此时本发明方法和原胶囊网络的参数量为初始参数量，即网络初始配置时的参数量大小。表1给出了与图3相一致的精确数值表达，由表1也可以看出，本发明能够在更少的训练样本量的情况下，获得更高的识别准确率。

[0115] 通过引入语义基元和语义胶囊的概念，仅需要少量的图像就可以获取MNIST数据集图像的主要结构特征，利用这些特征，本发明方法可以对大部分MNIST图像做出正确的判断。然而，原胶囊网络需要大量的训练样本才能达到相同的识别准确率。因此，在少量训练样本量的情况下，本发明方法可以达到更高的识别准确率。

[0116] 实施例8

[0117] 基于语义胶囊融合网络的图像识别方法同实施例1-6，仿真条件和内容同实施例7

[0118] 参见图4，图4为相同训练样本量不同网络参数量下本发明与原胶囊网络识别准确率对比图。其横坐标表示网络参数量，纵坐标表示识别准确率。从图4可以看到原胶囊网络在减少网络参数量的情况下，识别准确率下降明显，而本发明方法的识别准确率可以基本保持。当网络参数量降至初始配置网络参数量的1/75时，本发明方法相比原胶囊网络的识别准确率高出0.72%。

[0119] 表2为本发明方法与原胶囊网络在训练样本量保持一致时，分别在不同参数量的网络配置下训练后，测试集识别准确率的比较情况，此时训练样本量为10000张图片。表2给出了与图4相一致的精确数值表达，由表2也可以看出，本发明能够在更少的训练样本量情况下，获得更高的识别准确率。

[0120] 表2. 相同训练样本量不同网络参数量下本发明与原胶囊网络识别准确率对比表

网络参数量	本发明方法识别准确率	原胶囊网络识别准确率
-------	------------	------------

初始参数量	99.1313%	98.8682%
初始参数量1/2	99%	98.4475%
初始参数量1/21	99.0606%	98.3776%
初始参数量1/75	98.9192%	98.1971%

[0122] 通过感知语义基元,本发明方法已经提取出MNIST数据集10种数字的共同特征,所以仅仅有少量的细节特征需要通过基于学习的神经网络方法来提取。因此,本发明方法仅仅需要少量的网络参数量来达到优秀的识别性能。目前,在基于传统神经网络的方法进行图像识别过程中,为了达到理想的识别效果,往往需要设计庞大的神经网络结构,导致大量的网络参数量需要训练,本发明引入语义胶囊的概念,提前赋予网络图像的语义特征,减少了网络所需学习的特征,从而大大降低了网络所需训练的参数量,提高了训练速度,降低了网络的空间占用,减少了能源损耗,使网络的图像识别速度更快。

[0123] 简而言之,本发明公开的一种基于语义胶囊融合网络的图像识别方法,通过将神经网络与图像语义信息进行融合,克服了现有技术中神经网络依赖大量样本训练、网络结构复杂、可解释性差、空间占有度高,能耗高的问题。本发明实现的步骤是:(1)获取待识别图像;(2)获得训练样本集和测试样本集;(3)语义胶囊网络模块搭建;(4)卷积神经网络模块搭建;(5)将语义胶囊网络模块和卷积神经网络模块融合成语义胶囊融合网络;(6)训练语义胶囊融合网络;(7)语义胶囊融合网络性能测试。本发明提出了语义胶囊的概念,设计了语义胶囊融合网络。本发明具有在少量训练样本上识别准确率更高、网络参数量更少、网络可解释性强、能耗更低的优点。可用于大量简单图像的识别处理。

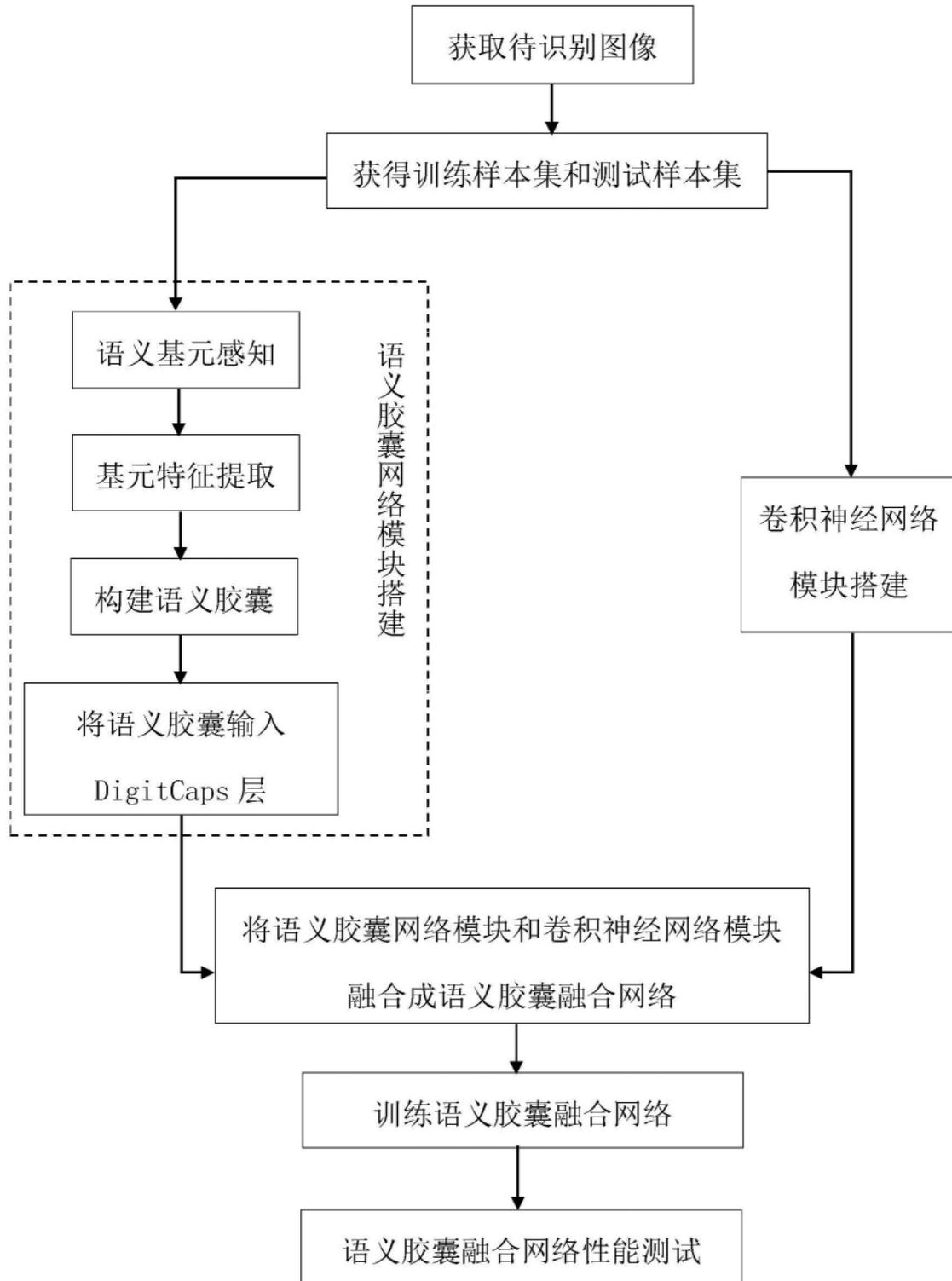


图1

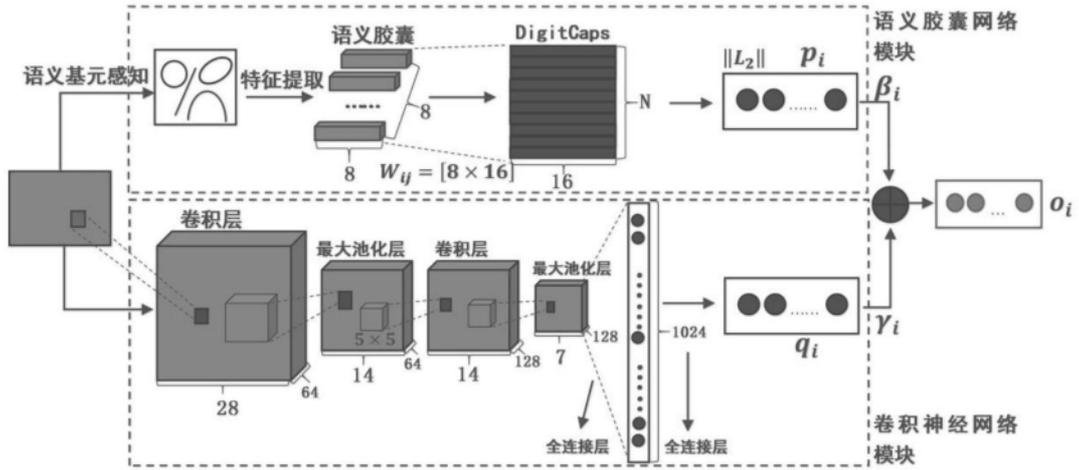


图2

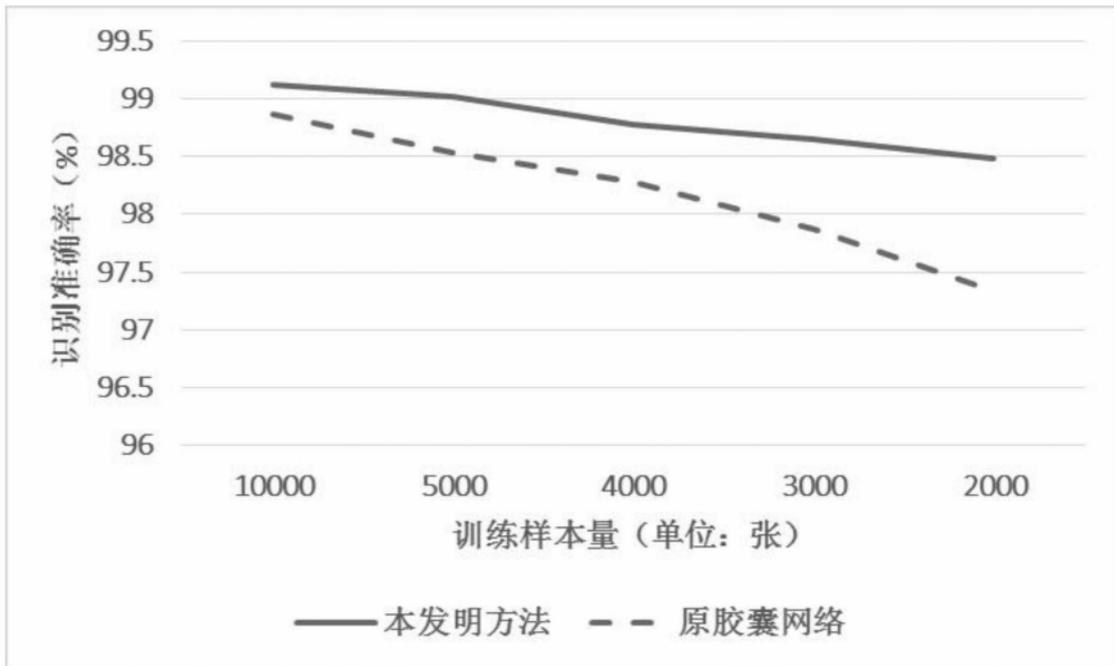


图3

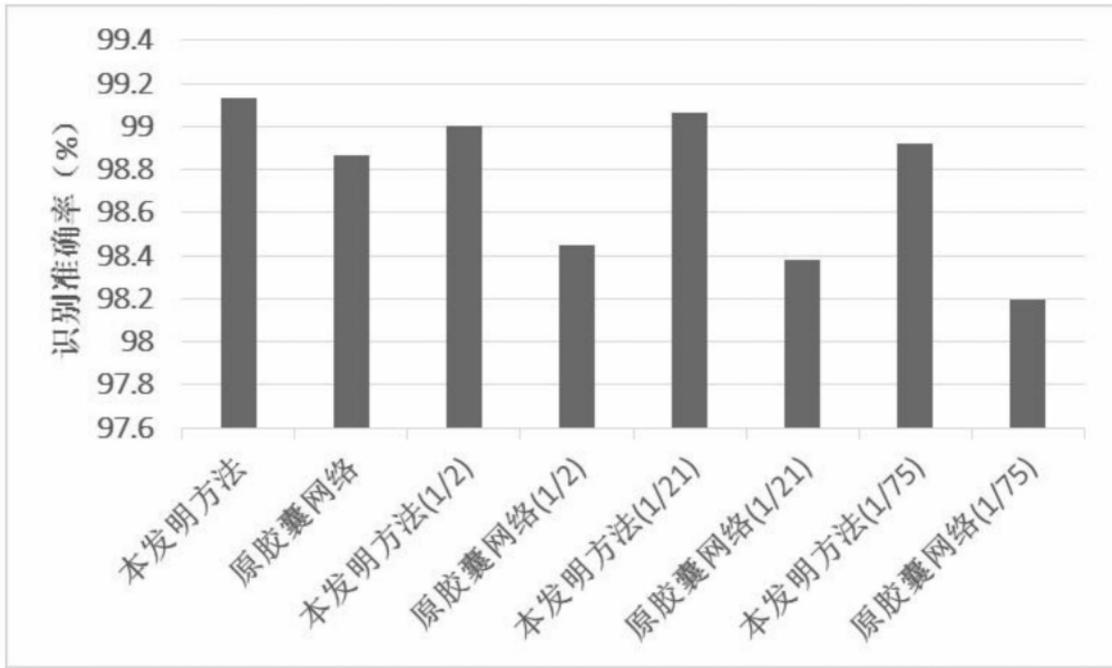


图4