



基于Capsule网络的甲骨文构件识别方法

鲁绪正, 蔡恒进, 林莉

引用本文:

鲁绪正, 蔡恒进, 林莉. 基于Capsule网络的甲骨文构件识别方法[J]. 智能系统学报, 2020, 15(2): 243–254.

LU Xuzheng, CAI Hengjin, LIN Li. Recognition of Oracle Radical based on the Capsule network[J]. *CAAI Transactions on Intelligent Systems*, 2020, 15(2): 243–254.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.201904069>

您可能感兴趣的其他文章

生成对抗网络辅助学习的舰船目标精细识别

Fine-grained inshore ship recognition assisted by deep-learning generative adversarial networks
智能系统学报. 2020, 15(2): 296–301 <https://dx.doi.org/10.11992/tis.201901004>

反馈式K近邻语义迁移学习的领域命名实体识别

Domain-named entity recognition based on feedback K-nearest semantic transfer learning
智能系统学报. 2019, 14(4): 820–830 <https://dx.doi.org/10.11992/tis.201804013>

基于改进卷积神经网络的多标记分类算法

A multi-label classification algorithm based on an improved convolutional neural network
智能系统学报. 2019, 14(3): 566–574 <https://dx.doi.org/10.11992/tis.201804056>

卷积神经网络的贴片电阻识别应用

Chip resistance recognition based on convolution neural network
智能系统学报. 2019, 14(2): 263–272 <https://dx.doi.org/10.11992/tis.201710005>

基于卷积神经网络的盲文音乐识别研究

Research on braille music recognition based on convolutional neural networks
智能系统学报. 2019, 14(1): 186–193 <https://dx.doi.org/10.11992/tis.201805002>

基于卷积特征和贝叶斯分类器的人脸识别

Face recognition based on convolution feature and Bayes classifier
智能系统学报. 2018, 13(5): 769–775 <https://dx.doi.org/10.11992/tis.201706052>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.201904069

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20190828.1430.004.html>

基于 Capsule 网络的甲骨文构件识别方法

鲁绪正, 蔡恒进, 林莉

(武汉大学 计算机学院, 湖北 武汉 430072)

摘要: 甲骨文作为中国最早的成形文字系统, 具有重要的文化和学术价值。研究甲骨文构件和其构形系统是破译未识别的甲骨文的重要方向, 但是甲骨文构件的标记工作只能由资深专家来完成, 并且需要耗费大量时间和精力。针对这些问题, 提出了一种基于 Capsule 网络和迁移学习的模型 OracleNet, 可以自动识别并标记甲骨文字形中包含的构件; 同时, 构建了包含标记的甲骨文字形和构件数据集, 用于模型的训练和评估。实验结果显示, OracleNet 模型对甲骨文构件的预测精确度达到了 60% 以上, 其中 Top5 精确度达到了 71.56%, 验证了模型的有效性。

关键词: 甲骨文; 甲骨文构件识别; 卷积神经网络; Capsule 网络; 动态路由算法; 迁移学习; 多目标识别; 图像识别
中图分类号: TP319.4 **文献标志码:** A **文章编号:** 1673-4785(2020)02-0243-12

中文引用格式: 鲁绪正, 蔡恒进, 林莉. 基于 Capsule 网络的甲骨文构件识别方法 [J]. 智能系统学报, 2020, 15(2): 243-254.

英文引用格式: LU Xuzheng, CAI Hengjin, LIN Li. Recognition of Oracle Radical based on the Capsule network[J]. CAAI transactions on intelligent systems, 2020, 15(2): 243-254.

Recognition of Oracle Radical based on the Capsule network

LU Xuzheng, CAI Hengjin, LIN Li

(School of Computer Science, Wuhan University, Wuhan 430072, China)

Abstract: As the earliest shaped character system in China, the inscriptions on bones or tortoise shells of the Shang Dynasty (c. 16th-11th century BC) have important cultural and academic values. The research on the constructional element, i.e., Oracle Radical, and configuration system of the inscriptions on bones or tortoise shells of the Shang Dynasty is a vital direction to identify the unrecognized Oracle Graphics. However, the marking of Radicals can only be done by experienced experts; moreover, it will take a considerable amount of time and effort. To solve these problems, we proposed a model, i.e., OracleNet, based on the Capsule network and transfer learning, which can automatically identify the Oracle Radical contained in a graphic. At the same time, we built a labeled Oracle Graphics dataset and a labeled Radicals dataset, which were used for training and evaluating the model. The experiment showed that the OracleNet had more than 60% precision for recognizing Radicals in a graphic and the Top 5 precision reached 71.56%, which verified validity of the model.

Keywords: inscriptions on bones or tortoise shells of the Shang Dynasty; Oracle Radical recognition; convolutional neural networks; Capsule network; dynamic routing algorithm; transfer learning; multi-target recognition; image recognition;

甲骨文作为汉字鼻祖, 记录了商朝的经济和政治情况。其不仅是一种文化的符号、文明的标志, 还复原了殷商历史的框架, 将我国信史时代向前推进了近五个世纪^[1]。自 1899 年甲骨文出土后, 甲骨文已逐渐渗透到历史学、艺术史、科技史等多个相关学科领域, 其传承具有重大文化及学

术意义。目前, 已识别的甲骨文字形有 2 400 余个, 未识别的仍有 2 500 余个, 因此甲骨文的相关研究任重道远。

研究表明, 甲骨文已经具有了一套相对完整的构形体系, 这套构形体系将甲骨文分为了字形和构件。组成甲骨文字形的构件通过一些固定的构造规律相互影响、关联, 在此基础上进行相互区分, 并且最终组合起来构建出了一个有序的构

收稿日期: 2019-05-05. 网络出版日期: 2019-08-28.

通信作者: 蔡恒进. E-mail: hjcai@whu.edu.cn.

形系统,而甲骨文构件精确地反映出了该系统的性质和特点^[2]。所以,对于甲骨文识别相关工作来说,甲骨文构件的研究是其基础,具有重要意义。

许多学者从甲骨文构件的角度出发,以加快未知字形的解读工作,如周新伦等^[3]基于图论对甲骨文的笔画特点进行分析,从而识别甲骨文字形,实现了早期的尝试;而李锋等^[4]则是对甲骨文的图特征进行了提取,以提升甲骨文字形识别的准确度;李东琦等^[5]实现了一个基于甲骨文构件的编码器,将甲骨文进行了数字化;高峰等^[6]通过建立甲骨文语义构件向量,结合 Hopfield 网络的识别结果的方法来匹配甲骨拓片或者照片中的模糊字;吴琴霞等^[7]通过仿射变换复用构件生成甲骨文字形,为以后的语义构件统计打下了基础;而顾绍通利用分型集合的原理将甲骨文字形进行拆分、描述,然后再对甲骨文字形进行识别,达到了一定的效果^[8]。

然而,通过构件研究甲骨文,首先需要专家进行构件标注,不仅需要相关专业知识,还要耗费大量的时间、人力和精力。针对上述问题,本文提出了一种新的通过机器识别甲骨文构件的方法,建立了一个基于 Capsule 网络和迁移学习的模型 OracleNet,通过甲骨文构件数据集对其进行训练,可以同时识别甲骨文字形中多个构件同时进行识别。

1 甲骨文数据集

本文尝试通过深度学习 (deep learning) 的方法来识别甲骨文字形中包含的构件,而深度学习模型对于数据样本的准备有以下要求:

- 1) 数据样本图片需要清晰可见且无重叠、模糊,图片大小、颜色、格式需要统一;
- 2) 字形识别任务为有监督学习的分类任务,所以全部训练数据都需要被正确标记;
- 3) 本文需要提取甲骨文构形规律和构件特征,所以需要被正确标记的甲骨文字形数据集;
- 4) 数据样本需要满足一定数量,否则深度学习模型难以收敛。

鉴于目前并没有公开的较为完整的数字化甲骨文字形集,为了便于后续研究,本文通过对经典书籍《甲骨文字编》^[9]中扫描的甲骨文字形进行预处理、标记,建立了两个类似于 MINSIT 数据集的已标记甲骨文数据集。

这两个数据集不仅限于用于图像分类识别任务,还可以应用于其他甲骨文领域的研究中,如甲骨文字形破译等。本文希望通过这个数据集简化研究人员对于甲骨文数据收集、处理、筛选的时间,更多专注于研究本身。

两个数据集分别为: Oracle-250 数据集和 Rad-

ical-148 数据集。










1.1 Oracle-250 数据集

目前已发现的殷商甲骨文字形共有近 5 000 个,其中包括已识别甲骨文字形 2 304 个,未识别甲骨文字形 2 523 个。由于甲骨文字形中大部分字形出现率极低,样本数很少,本文只选择甲骨文字形中频率最高的 250 个字形来构成 Oracle-250 数据集。

此外,由于各个类别的分布不均匀,本文通过网络获取和数据增强算法,对已识别的原始数据集进行了扩充,同时对 250 个类别的样本数量进行平衡。最终得到的 Oracle-250 数据集中的甲骨文字形样本总数为 92 160 个。

在本数据集中,并没有对甲骨文图片的大小设置标准,可以在后续神经网络中进行调整。在已识别的甲骨文字形数据中,标记了每个字形所包含的构件,Oracle-250 中的部分样本如表 1 所示。

表 1 Oracle-250 部分样本及其所含构件
Table 1 Samples of Oracle-250 and containing radicals

甲骨文字形	索引	构件	索引	构件	索引
	37		65		0
	187		60		8
	76		73		9

1.2 Radical-148 数据集

殷商甲骨文是具有系统的成熟文字,每个字都是由构件组成或其本身就是构件。通过专家学者多年的全面考察,甲骨文的基础构件共有 412 个,这些基础构件一般来说可以独立使用于记录语言之中,也可以和其他的基础构件一起组成新的甲骨文字形。

若按照是否构成完整字形来分类,甲骨文构件中 278 个为成字构件,而其他的 134 个为非字构件。而如果按照甲骨文构件行使的功能来分类,甲骨文构件中有 356 个用于表形功能,有 125 个用于表义,113 个用于示音,18 个用于标示功能。其中,有一部分成字构件独立成字时自身就具有了表形、表义或示音功能,而当这些构件用于组成新的甲骨文字形的时候,可能会行使其他的功能,而和自身原本的功能不同。

《甲骨文字编》中选取了主要的成字构件,并将这些构件分为了 148 类。Radical-148 数据集也参考了这种分类方式,是已标记的构件图片的集合,包含 148 个构件类别。

为了进一步扩充 Radical-148 数据集,本文从字源网站获取了更多的构件字形图片对数据集进

行填充,但这些字形出现的频率不同,导致了数据集数量不平衡。

为解决这一问题,本文邀请了 700 名志愿者分别对部分甲骨文构件进行手写仿写。志愿者进行仿写的内容是经过计算和设计的,由随机从原始数据库中选出的甲骨文构件字符组成,而每个类型的构件的仿写总量是根据原始类别而计算得出的。因此,每位志愿者的仿写内容是不一样的,针对同一构件,不同仿写者所参照的图片也是不一样的,这样可以保证收集到的数据的多样性最大,同时又解决了原始数据不均衡的问题。

如图 1 所示各为一张仿写问卷征集表,图 2 所示为一份仿写结果。



图 1 Radical-148 数据集问卷征集表
Fig. 1 Questionnaire of Radical-148

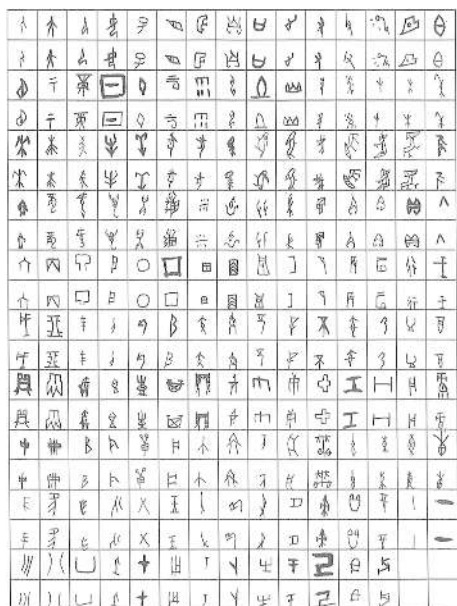


图 2 已填写的 Radical-148 数据集征集表
Fig. 2 Filled questionnaire of Radical-148

最终,本文得到 148 个类的甲骨文构件图片共计 108 989 张。扩充后的 Radical-148 数据集如表 2 所示。

表 2 Oracle-250 部分样本及其所含构件
Table 2 Samples of Oracle-250 and their containing radicals

汉字	Unicode	图片数量	原始图片样本	手写扩充样本
人	4EBA	956		
大	5927	947		
止	6B62	692		
女	5973	611		

2 Capsule 网络

Capsule 网络是 Geoffrey Hinton 在 2017 年提出的一种新型的神经网络^[10],与 LeNet^[11]、Alex-Net^[12] 等传统卷积神经网络 (convolutional neural networks, CNN) 不同, Capsule 神经网络设计了一种新型的神经元——Capsule(胶囊),来表示一些特定的实体类型 (Entity)。

Capsule 网络的设计受到了人类视觉生物学原理的启发。在传统的 CNN 架构中,图像整体全部传入模型进行训练,通过卷积运算将图像中一个位置获得的优秀权重值知识转化到其他位置上,并通过池化 (Pooling) 运算逐渐缩小特征图 (Feature Map),从而为模型带来不变性。

但是,这样的计算过程和人类视觉生物学原理有所不同,人类视觉往往会通过注视一些较小的区域 (注视点) 序列来降低需要进行高分辨处理的画面面积,而不会去同时关注图像的全部内容。Hinton 假设人类的多层视觉系统可以在每个注视点上创建一种类似于解析树 (Parse Tree) 的机制,对于每个注视点,其解析树由指定多层神经网络进行构建,每一层会被分割为许多神经元小组,它们被称为 Capsule。

本文使用 Capsule 网络来进行甲骨文构件识别的原因是:

1) Capsule 网络可以方便地通过单目标 (Single-Object) 数据集进行训练,而在测试时可以进行高精度的多目标 (Multi-Object) 预测^[13],这种方式和本文的甲骨文识别任务相契合:通过单目标的甲骨文构件数据集训练模型,然后在甲骨文字形数据集上进行多目标预测,判断每个甲骨文字形中包含的构件;

2) 相对于其他的多目标预测模型,如 RCNN^[14]、Fast-RCNN^[15] 等, Capsule 网络无需准备多标签数

据集,同时也不需要训练数据中目标的位置信息进行标记,从而可以极大程度上降低数据准备的工作量。

2.1 Capsule

Capsule 是一组新型的神经元,该神经元使用激活向量 (Activity Vector) 代替传统 CNN 中的标量特征值,用“向量输入/输出”取代了以往的“标量输入/输出”。

每个 Capsule 单元表示了特定实体类型的实例化参数,如色彩、位置、纹理、大小、方向等,通过激活向量的长度来表示某个特定实体出现的概率,通过激活向量的方向来表示该实体所对应的更高层级属性。

Capsule 网络提供了一种基于聚类思想来代替池化完成特征整合的全新方案, Capsule 本身也可以通过动态路由算法带来类似于池化的不变性。其优势在于,相对于传统的池化算法, Capsule 保留了全部的图像特征,其表达能力也更强。

2.2 squashing 激活函数

Capsule 网络将激活向量视作逻辑单元,其长度表示特定实体出现的概率,以判断图像中实例化实体的存在与否。为此, Capsule 网络需要使用一种新型的非线性激活函数: squashing 函数,其表示为

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

式中: \mathbf{v}_j 表示第 j 个 Capsule 的输出向量; \mathbf{s}_j 表示输入向量的总和。

和传统神经网络中常使用的 sigmoid 函数类似, squashing 函数使激活向量的方向不变,将较短和较长的激活向量的长度分别缩放至 0 和 1 附近,并将向量长度控制在 0~1。通过 squashing 激活函数, Capsule 实现了用激活向量的模长来表示实体出现的概率。

2.3 动态路由算法

Capsule 网络通过动态路由算法 (dynamic routing algorithm) 来实现对更高层级的 Capsule 实体 (或属性) 的聚类,并将子 Capsule 输出向量中包含的特征信息传送到合适的父类 Capsule 中。

在全连接形式的 Capsule 网络中,对于除了第一层 Capsule 之外的所有层级,每层 Capsule 的输出在传入下一层 Capsule 之前,都需要乘以一个权值矩阵 \mathbf{W}_{ij} 进行变换:

$$\hat{\mathbf{u}}_{ji} = \mathbf{W}_{ij} \mathbf{u}_i \quad (1)$$

式中: \mathbf{W}_{ij} 为待训练的权值矩阵,通过反向传播梯度更新进行训练; \mathbf{u}_i 为输入向量。

总输入 \mathbf{s}_j 为下层 Capsule 中所有输出向量 $\hat{\mathbf{u}}_{ji}$ 的加权和:

$$\mathbf{s}_j = \sum_i c_{ij} \hat{\mathbf{u}}_{ji}$$

式中: c_{ij} 为耦合系数 (coupling coefficients),由迭代的动态路由选择过程确定。

Capsule_{*i*} 和其上层所有的 Capsule 之间的耦合系数的和为 1,并且由“softmax”路由算法决定:

$$\sum_j c_{ij} = 1$$

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})}$$

式中: b_{ij} 是 Capsule_{*i*} 和上层 Capsule_{*j*} 相耦合的对数先验概率,初始化为 0。

对数先验概率 b_{ij} 和权值矩阵 \mathbf{W}_{ij} 分别通过动态路由算法和梯度回传有区别地进行训练,它们之和两个 Capsule 的位置与类型相关,而和输入图像的内容无关。

耦合系数 c_{ij} 通过测量上层的每一个 Capsule_{*j*} 的输入 \mathbf{v}_j 和下层的每一个 Capsule_{*i*} 的预测一致性进行迭代更新,并被加入 b_{ij} 中。本文中该预测一致性被设置为标量乘积 $\hat{\mathbf{u}}_{ji} \cdot \mathbf{v}_j$,为对数似然度。

Capsule 网络动态路由算法的流程如图 3 所示。

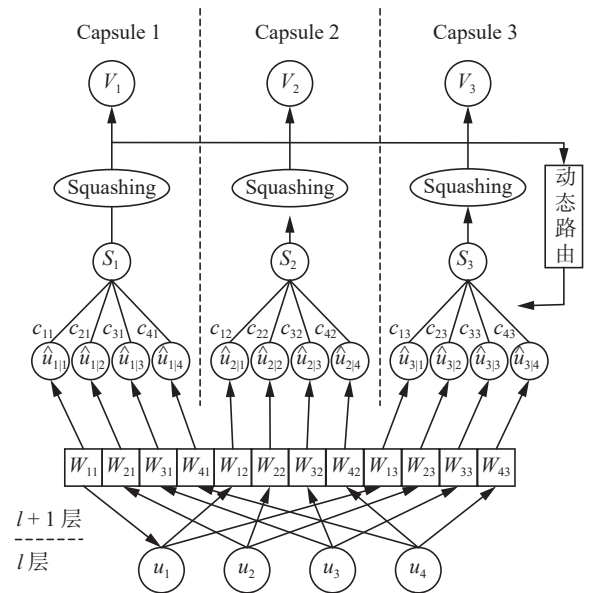


图 3 Capsule 网络的动态路由算法

Fig. 3 Dynamic routing algorithm of Capsule network

在卷积形式的 Capsule 层中,每一个 Capsule 单元都是一个卷积单元而不是全连接单元,因此每一个 Capsule 将输出一个向量组而不是单个输出向量。

动态路由算法的步骤如算法 1 所示。

算法 1 动态路由算法

输入 第 l 层 Capsule 的输出向量组 u_i

输出 第 $l+1$ 层 Capsule 的输出向量组 v_j

1) 对于第 l 层中的所有 Capsule i 以及第 $l+1$ 层中的所有 Capsule j : 初始化 $b_{ij} = 0$

2) 执行循环, 迭代 r 次:

3) 对于第 l 层中的所有 Capsule i : $c_i \leftarrow \text{softmax}(b_i)$

4) 对于第 $l+1$ 层中的所有 Capsule j : $s_j \leftarrow \sum_i c_{ij} \hat{u}_{ji}$

5) 对于第 $l+1$ 层中的所有 Capsule j : $v_j \leftarrow \text{squash}(s_j)$

6) 对于第 l 层中的所有 Capsule i 以及第 $l+1$ 层中的所有 Capsule j : $b_{ij} \leftarrow b_{ij} + \hat{u}_{ji} \cdot v_j$

7) 返回 v_j

2.4 损失函数

在进行分类时, Capsule 网络通过计算不同类别所对应的激活向量的模长来表示某个类型实体存在的概率。当网络检测到了某种实体时, 概率趋近于 1, 而当网络认为某种实体不存在时, 概率趋近于 0。同时, 为了支持多目标识别, Capsule 网络的损失函数采用了支持向量机 (support vector machine) 中经常使用的 Margin loss, 其表达式为

$$L_c = T_c \max(0, m^+ - \|v_c\|)^2 + \lambda (1 - T_c) \max(0, \|v_c\| - m^-)^2$$

式中: c 是分类的类别; T_c 为分类的指示函数, c 类存在时为 1, 否则为 0; m^+ 与 m^- 分别代表了上

下边界值; λ 通过对不存在的类别进行加权, 从而阻止网络在刚开始学习的过程中缩小所有 Capsule 的激活向量的长度。

本文中, m^+ 设置为 0.9, m^- 设置为 0.1, λ 设置为 0.5。

2.5 重构

传统的卷积神经网络通常使用 Dropout^[16] 作为正则化方法以降低过拟合风险, 而 Capsule 网络使用一种重构 (Reconstruction) 结构进行正则化。

重构结构的原理类似于自动编码器 (Autoencoder) 中的解码器 (Decoder) 部分, 在训练过程中, 忽略其他所有 Capsule, 只使用正确类别所对应的 Capsule 的激活向量重新构建出原始图像。通过重构图像与原始输入图像计算得到重构损失 (Reconstruction Loss), 并且将该损失计入模型总损失中进行梯度回传更新。重构结构鼓励 Capsule 的激活向量对图像进行更为宏观的表征, 使激活向量包含更多的有用信息, 从而达到正则化的目的。

3 OracleNet

3.1 OracleNet 架构

针对甲骨文构件识别任务, 本文建立了一个基于卷积神经网络和 Capsule 网络的端对端的模型, 命名为 OracleNet, 其架构如图 4 所示。

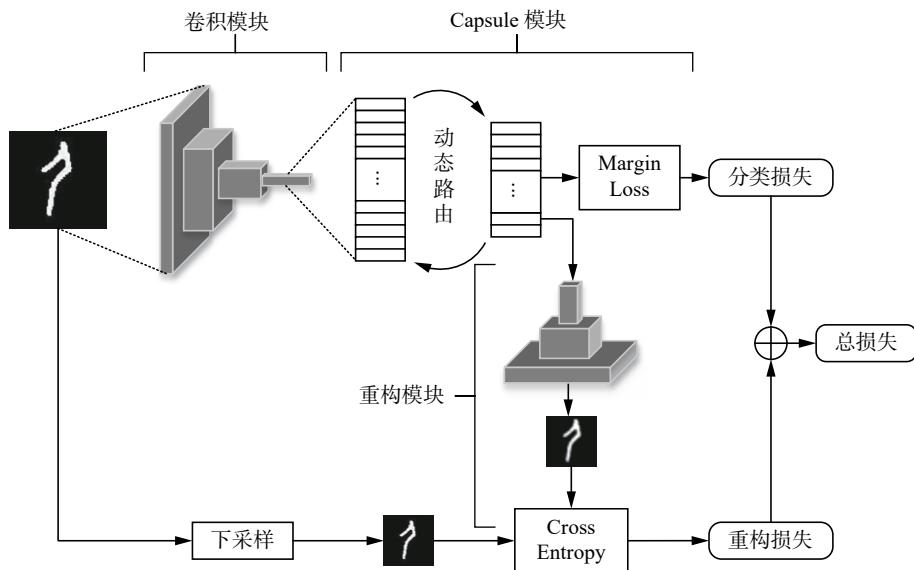


图 4 OracleNet 架构

Fig. 4 Architecture of OracleNet

3.1.1 卷积模块

OracleNet 首先通过多层卷积层来提取图片

中的空间特征信息, 每层卷积层都采用 ReLU^[17] 激活函, 并且使用批标准化 (Batch Normaliza-

tion)^[18]进行正则化。

3.1.2 Capsule 模块

在卷积层后分别加入两层 Capsule, 使用 squashing 函数作为激活函数。输出层 Capsule 数量为 148 个, 每个 Capsule 对应一个甲骨文构件。两层 Capsule 之间的参数在整个模型前向传播时通过动态路由算法进行更新, 路迭代次数为 3 次。

在本文实验中发现, 除了输出层之外的 Capsule 采用修改后的 squashing 激活函数会使结果有所提升, 如式 (2) 所示:

$$v_j = \frac{\|s_j\|^2}{0.5 + \|s_j\|^2} \frac{s_j}{\|s_j\|} \quad (2)$$

这个函数的特点是在向量长度很接近于 0 时起到放大作用, 而不像原来的函数压缩全局向量。

此外, 为节省计算资源, 本文使用权值共享代替全连接形式的 Capsule, 如图 5 所示。

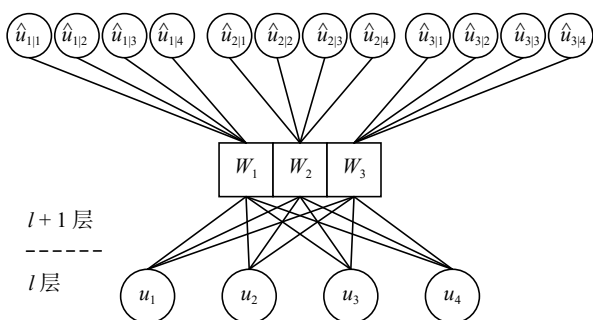


图 5 权值共享形式的 Capsule
Fig. 5 Weights sharing Capsule

所有输入向量共享一个权值矩阵可以很大程度上减少参数量级, 从而降低模型过拟合的风险, 并且压缩计算所需的内存和时间。在权值共享形式下, 式 (1) 修改为式 (3) 形式:

$$\hat{u}_{ji} = W_j u_i \quad (3)$$

3.1.3 重构模块

在输出 Capsule 之后建立重构模块。本文尝试了多种重构方式, 包括全连接、卷积+上采样、反卷积等, 最终选择使用多层反卷积的结构, 对于甲骨文构件数据来说效果最优。每层反卷积层均采用 ReLU 激活函数, 并且在每个反卷积层后设置 Batch Normalization 批标准化层。

3.1.4 损失函数与优化器

OracleNet 的分类损失使用 Margin Loss, 而重构损失使用交叉熵 (Cross Entropy), 为了让两者保持同一量级, 将重构损失按照一定比例进行缩小, 缩小比例取决于重构的图片尺寸。最后将分类损失和重构损失相加构成总损失, 并通过 Adam

优化器^[19]进行梯度回传更新。

3.2 基于迁移学习的 OracleNet

鉴于 Capsule 网络的计算量相比传统 CNN 结构提升了数倍, 为了节省计算资源, 本文采用了迁移学习的方式来训练 OracleNet 中的卷积模块。本文分别使用已在 ImageNET 数据集上预训练过的 InceptionV3、ResNet50 和 Xception 架构进行迁移学习训练。

Inception 架构由 Szegedy 等^[20]于 2014 年提出, 该架构使用一种特殊的 Inception 模块充当“多级特征提取器”, 分别使用不同大小的卷积核进行卷积, 并把卷积输出串联起来作为下一层的输入, 进一步提高了在 ImageNet 数据集上的分类效果。

ResNet 架构由 He 等^[21]于 2015 年提出, 与传统的顺序 CNN 网络架构不同, 其加入了恒等映射层, 从而让网络在深度增加情况下而避免“退化”现象, 同时压缩了参数量级。

Xception 架构由 Chollet 等^[22]于 2016 年提出, 它是 Inception 架构的扩展, 用深度可分离的卷积代替了标准的 Inception 模块, 更大程度地减少了模型参数, 并且在 ImageNet 上获得了更高的准确率。

迁移学习训练分为两步, 如图 6 所示。

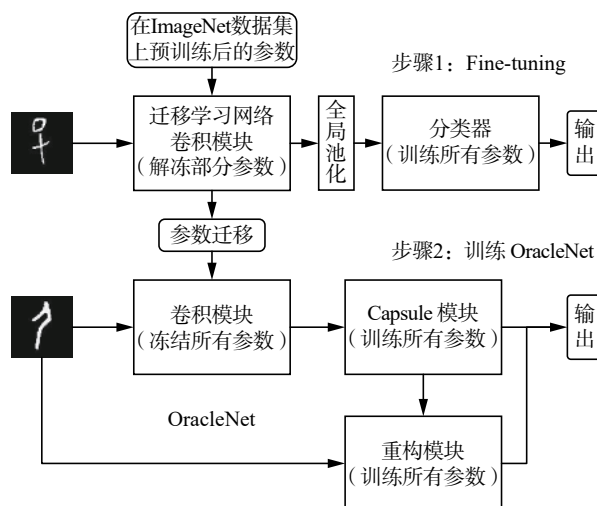


图 6 基于迁移学习的 OracleNet
Fig. 6 OracleNet based on transfer learning

1) 使用已经在 ImageNet 数据集上预训练完成的迁移学习网络 (InceptionV3、ResNet50 和 Xception 架构), 解冻其最后 10 层卷积层的参数, 在全局池化层后新添加由多层全连接层构成的分类器, 并使用甲骨文部数据集对其进行微调 (Fine-tuning) 训练。

2) 在训练 OracleNet 时, 将 1) 中的预训练网

络卷积部分的参数迁移至 OracleNet 的卷积模块,并且冻结其所有参数。使用甲骨文部数据集对 Capsule 模块和重构模块进行训练。

3.3 使用 OracleNet 预测甲骨文字形中的构件

使用训练完成的 OracleNet 预测甲骨文字形中包含的构件。

将甲骨文字形图片输入已训练的模型,得到 Capsule 输出层的预测向量组。计算每个类别对应的激活向量的模长,得到每个类别的预测概率,然后判断甲骨文字形中包含的构件,其流程如图 7 所示。

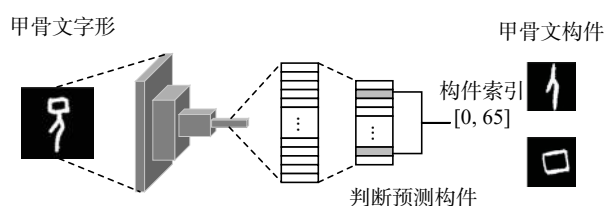


图 7 甲骨文构件识别流程

Fig. 7 Process of oracle radical recognition

本文通过两种方式来对甲骨文字形中包含的构件进行判断,生成最终的预测构件列表:

1) 不考虑置信度: 选取预测概率中最高的 n 个类别生成预测构件列表;

2) 考虑置信度: 设置置信度阈值 a , $0\% < a < 100\%$ 。选取所有预测概率大于 a 的类别生成预测构件列表。

4 实验

4.1 数据预处理

鉴于甲骨文构件原始数据集大小不足,且各类样本数不均衡,本文通过数据增强 (data augmentation) 方法^[23],对原始图片进行旋转、平移、反转、放缩等随机干扰,生成新的样本,从而扩充原始数据集。

将扩充后的数据集样本进行归一化处理,放缩至 $(0, 1)$ 区间,并按照 4:1 的比例进行训练集-测试集拆分,同时对样本标签进行独热编码 (One-Hot Encoding)。

甲骨文构件的原始数据集包含 148 类构件,共 108 989 张样本图片。在此基础上使用不同的图片采样尺寸进行数据增强,最终建立了 3 个新的甲骨文构件数据集,分别为 Radical-2000-28、Radical-10000-56 和 Radical-10000-224,如表 3 所示。

表 3 经过预处理的甲骨文构件数据集
Table 3 Preprocessed oracle radical dataset

	Radical-2000-28	Radical-10000-56	Radical-10000-224	Oracle-224
训练集样本数	236 800	1 184 000	1 184 000	—
验证集样本数	59 200	296 000	296 000	—
测试集样本数	—	—	—	92 160
图片尺寸	28×28	56×56	224×224	224×224

其中 Oracle-224 为已识别并被标记的甲骨文字形,作为甲骨文构件识别的最终测试集,也经过了上述步骤中同样的归一化和独热编码处理。

4.2 训练 OracleNet

4.2.1 基于迁移学习的 OracleNet 训练

OracleNet 使用甲骨文构件数据集进行训练。本文基于 TensorFlow 平台进行 OracleNet 的代码实现,实验设备为双路 NVIDIA GeForce GTX 1080Ti。

本文采用迁移学习的方式来训练 OracleNet 中的卷积模块,具体训练步骤如下。

1) 使用 Radical-10000-224 数据集对已在 ImageNET 数据集上预训练过的 InceptionV3、Res-

Net50 和 Xception 模型进行 Fine-tuning 训练。

2) 使用甲骨文构件训练集 Radical-2000-28、Radical-10000-56、Radical-10000-224 等数据集,对 OracleNet 的 Capsule 模块和重构模块进行训练。训练周期为 20 个 epoch(训练周期),学习率设置为 0.001,每 2 000 个批次后衰减为之前的 96%。

此外,作为对照,本文额外实现并训练了一个基线模型 (baseline model),其采用 Hinton 提出的 CapsNet 架构,包含 2 层卷积层和 2 层 Capsule 层,并使用 3 层全连接层作为重构模块。

OracleNet 在甲骨文构件训练集上的训练结果如表 4 所示。

表4 OracleNet 训练的的验证集准确率
Table 4 Validation accuracy of OracleNet training

训练集	模型	Top1准确率/%	Top5准确率/%
2000-28	基线模型	65.59	71.32
10000-56	InceptionV3迁移	76.86	83.53
10000-56	ResNet50迁移	77.42	84.61
10000-56	Xception迁移	79.67	87.26

从训练的验证结果来说:

1) 迁移学习带来的提升是明显的, 相对于基线模型提升了10%以上的准确率, 因为其将在ImageNET数据及上已经学得的知识迁移到了甲骨文任务中识别中, 同时极大程度上节省了计算资源和时间;

2) Xception 架构优于 InceptionV3 和 ResNet50 架构, 这是因为它对传统的 Inception 模块进行了优化, 并且加入了类似于 ResNet 中的恒等

映射层, 具有更佳的特征提取能力。

4.2.2 模型融合

本文使用了模型融合(model ensemble)的方法提升模型鲁棒性, 主要使用以下两种方法进行融合。

1) 在训练过程中, 将不同超参数下各个模型的预测结果(概率)进行加权平均融合;

2) 在进行预测时, 从原始图片中随机裁剪出若干不同大小、不同位置的图片, 并将这些图片的预测结果进行融合。

此外, 如果在预测时考虑置信度, 若直接采用加权平均的方法进行模型融合会减小方差, 使预测概率大幅降低至置信度以下, 反而使精确度下降。所以, 针对考虑置信度的模型融合方式为: 每个类别从所有模型的预测结果中选取概率最高的值, 作为该类别的预测概率, 并最终生成融合后的结果。

使用模型后的结果如表5所示。

表5 模型融合后的 OracleNet 验证集准确率
Table 5 Validation accuracy of OracleNet with ensemble

训练集	模型	Top1准确率/%	Top5准确率/%
2000-28	基线模型	65.59	71.32
10000-56	Xception迁移	79.67	87.26
10000-56	Xception迁移+5模型融合	80.24	87.72
10000-56	Xception迁移+10模型融合	80.90	88.88
10000-224	Xception迁移+10模型融合	82.71	90.32

可以清楚地看到, 模型融合带来了1%以上的小幅提升, 这是因为它通过降低预测结果的方差, 使模型系统的鲁棒性得到了提升。不同超参数下的模型预测的结果中, 对于发生分歧的类别, 模型融合会降低对应的概率; 而对于大部分模型都一致的类比, 模型融合会保留其原有的判别(如激活或不激活)。通过模型融合, 降低了模型预测结果中的部分噪声, 提高了模型预测的稳定性。

此外, 使用 Radical-10000-224 数据集可以获得更高的准确率, 因为该数据集中每张图片所含像素是 Radical-10000-56 的16倍, 包含了更多的图像信息, 也为重构模块带来了更好的正则化效果。

4.2.3 训练时间对比

不同模型的训练时间对比如表6所示。

表6 模型融合后的 OracleNet 验证集准确率
Table 6 Training time of OracleNet

训练集	模型	训练时间/h
2000-28	基线模型	0.5
10000-56	InceptionV3迁移	5
10000-56	ResNet50迁移	5
10000-56	Xception迁移	5
10000-56	Xception迁移+5模型融合	25
10000-56	Xception迁移+10模型融合	50
10000-224	Xception迁移+10模型融合	140

可以看到, 使用更大尺寸的图像可以带来更好的训练效果, 这是因为更高的分辨率包含了更多的特征信息, 也为重构模块带来了更好的正则化效果。采用模型融合的方式也提升了模型的鲁棒性并带来了准确度的提升。

但是, 大尺寸和模型融合的代价是计算时间

的成倍增长,使用最大尺寸训练集下、Xception 迁移和 10 模型融合的累计时间达到了 140 h。

4.3 甲骨文构件识别

通过已识别并标记的甲骨文字形数据集 Oracle-224 对 OracleNet 进行测试评估。

甲骨文构件识别是典型的多分类任务,通过精确度 (Precision) 对预测结果进行评估,其计算方式为

$$\text{Precision} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}} \quad (4)$$











表 7 OracleNet 在 Oracle-224 字形测试集上的精确度

Table 7 Precision on Oracle-224 oracle graphic test set

训练集	模型	考虑置信度50%的 精确度/%	不考虑置信度的Top2 精确度/%	不考虑置信度的Top5 精确度/%
Radical-2000-28	基线模型(baseline)	44.76	51.26	56.02
Radical-10000-56	Xception迁移	53.23	63.06	68.14
Radical-10000-56	Xception迁移+5模型融合	58.92	63.77	69.61
Radical-10000-56	Xception迁移+10模型融合	61.51	63.35	69.94
Radical-10000-224	Xception迁移+10模型融合	62.39	65.82	71.56

表 8 考虑置信度 50% 的预测结果示例

Table 8 Prediction results considering a confidence level of 50%

字形索引	字形	真实构件标签	预测构件标签	预测构件示例图片
109		[1, 8]	[0, 1, 8, 65]	
158		[1, 60]	[1, 59, 60, 133]	
542		[2, 9]	[9, 78]	
793		[25]	[25, 117, 80]	
1 041		[3, 42]	[37]	

考虑置信度带来的好处是,可以展现出 Capsule 网络“真正”认为输入字形中存在的构件类别,更符合人类的认知方式。可以看到,当字形结构关系较为简单时,有着优秀的预测性能;但是,当字形较为复杂时,所有类别的概率都会降低,经常只输出 1 个预测类别甚至没有预测结

式中: TruePositive 为真正例; FalsePositive 为假正例。统计过程中,对于一个真实甲骨文字形样本,若其真实构件标签之一出现在预测构件列表中,即认为该标签被成功预测,判为真正例,否则将该标签判为假正例。

各模型在 Oracle-224 测试集上的预测结果如表 7 所示。








随机选取部分考虑置信度 50% 的测试结果如表 8 所示。

果。此外,考虑置信度时没有优秀的模型融合方案,如果直接进行平均融合会使全部概率大幅降低至置信度以下,即使按照本文设计的融合方法,也会部分丧失模型融合带来的鲁棒性。

随机选取部分不考虑置信度的 Top2 的测试结果如表 9 所示。

表9 不考虑置信度的 Top2 预测结果示例







Table 9 Top2 prediction results without considering a confidence level

字形索引	字形	真实构件标签	预测构件标签	预测构件图片
75		[0, 1]	[0, 1]	
167		[2, 8]	[8, 19]	
337		[0, 5]	[5, 64]	
900		[8, 28, 60]	[60, 111]	
1 041		[0, 37]	[0, 37]	

可以看到, 不考虑置信度时, 会强制模型输出概率最高的若干个类别, 即使在所有概率都很低的情况下也可以进行预测。这种方式不论是对简单还是复杂的字形结构都有着优秀的识别能力, 鲁棒性比考虑置信度时更高。但是, 其缺点是会输出冗余的类别, 干扰人们进一步的工作。如 Top5 预测虽然准确度最高, 但是其预测的类别数远多于真实标签中的类别数, 相关工作还需要从预测的 5 个类别中再次进行人工筛选。

此外, 部分甲骨构件本身包含了其他构件的图形, 如表 10 所示。

表 10 部分存在包含关系的构件
Table 10 Radicals with inclusive relations

索引	构件图片	包含的构件索引	构件图片
4		[65, 133, 134]	
60		[59, 133]	
143		[98, 133, 134, 144]	

这种包含关系会给模型带来困扰, 因为 Capsule 网络在进行预测时由对应各个构件的激活向量的模长进行判断, 并输出其检测到的所有构件。如果一个构件存在包含关系, Capsule 网络可能会错误地检测到所有包含的图形, 并将其所包含的其他构件一起输出。

存在包含关系的构件的概率大小不一, 所以很难通过对预测概率排序的方式来进行选择, 而如果将全部构件都输出, 则会造成预测结果的冗余, 不利于进一步工作。如何在存在此类包含关系的众多构件中判断正确的类别是一个难点, 也是本文未来进一步研究的方向。

5 结束语

本文提出了一种基于深度学习的机器识别甲骨文字形中包含构件的方法。首先, 通过扫描和手工仿写的方式, 分别建立了已标记的甲骨文字形数据集和构件数据集。接着, 本文分析了 Capsule 网络的原理, 以及其对于甲骨文构件识别任务的优势所在, 并且构建了一个基于 Capsule 网络的甲骨文构件识别模型 OracleNet。最后, 本文使用之前构建的数据集对 OracleNet 进行了验证, 通过迁移学习的方式使用甲骨文构件对其进行训练, 训练过程中在验证集上的 Top5 准确率达到 90% 以上; 在甲骨文字形数据集上进行了最终测试和评估, 得到了高于 70% 的精确度。实验表明, 基于 Capsule 网络的 OracleNet 可以高效地对甲骨文字形中所含构件进行识别, 并给出候选构件列表。

本文的贡献在于, 通过将甲骨文构件标记的工作自动化, 可以给甲骨文相关工作者高质量的指导意见, 并且在很大程度上降低他们的工作量, 从而将研究精力转向更深的领域。同时, 对甲骨文构件的识别也提升了破译未识别甲骨文字形的可能性。本文的另一个贡献是建立了完善且公开的甲骨文字形和构件数据集。这两个数据集不仅可以对甲骨文相关工作者提供帮助, 简化他们的数据处理相关工作, 也可以用于各类图像识别模型的验证, 从而帮助研究者优化算法。

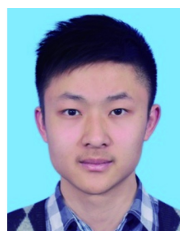
本文验证了 Capsule 网络 and 传统 CNN 相比的优势: 其实体 (或属性) 的概念引入到了网络中, 从而使不同 Capsule (或神经元) 之间具有更近似于人类知识体系的联系。本文中实验也表明了 Capsule 网络有效地学习到了甲骨文构件的特征信息, 以及构件和字形之间的构形关系。在今后的工作中, 可以继续研究如何利用这些知识去破译剩余的 2 500 多个未被识别的甲骨文字形。

参考文献:

- [1] 朱彦民. 从甲骨文说到中国文化自信 [J]. 殷都学刊, 2018, 39(3): 23-34.
ZHU Yanmin. From the oracle to Chinese culture confidence[J]. Yindu journal, 2018, 39(3): 23-34.
- [2] 竺海燕. 甲骨构件与甲骨文构形系统研究 [D]. 上海: 华东师范大学, 2005.
ZHU Haiyan. Research of the structural element and structural system of the oracle-bone inscriptions[D]. Shanghai: East China Normal University, 2005.
- [3] 周新伦, 李锋, 华星城, 等. 甲骨文计算机识别方法研

- 究[J]. 复旦学报(自然科学版), 1996, 35(5): 481–486.
- ZHOU Xinlun, LI Feng, HUA Xingcheng, et al. A method of Jia Gu Wen recognition based on a two-level classification[J]. Journal of Fudan University (Natural Science), 1996, 35(5): 481–486.
- [4] 李锋, 周新伦. 甲骨文自动识别的图论方法[J]. 电子科学学报, 1996, 18(S1): 41–47.
- LI Feng, ZHOU Xinlun. Recognition of Jia Gu Wen based on graph theory[J]. Journal of electronics, 1996, 18(S1): 41–47.
- [5] 李东琦, 刘永革. 基于构件的甲骨文字编码器设计与实现[J]. 科技创新导报, 2010(15): 18.
- LI Dongqi, LIU Yongge. Design and implementation of radical-based oracle encoder[J]. Science and technology consulting herald, 2010(15): 18.
- [6] 高峰, 吴琴霞, 刘永革, 等. 基于语义构件的甲骨文模糊字形的识别方法[J]. 科学技术与工程, 2014, 14(30): 67–70, 86.
- GAO Feng, WU Qinxia, LIU Yongge, et al. Recognition of fuzzy inscription character based on component for bones or tortoise shells[J]. Science technology and engineering, 2014, 14(30): 67–70, 86.
- [7] 吴琴霞, 栗青生, 高峰. 基于语义构件的甲骨文字库自动生成技术研究[J]. 北京大学学报(自然科学版), 2014, 50(1): 161–166.
- WU Qinxia, LI Qingsheng, GAO Feng. Study on the technique of automatic generation of oracle characters based on semantic component[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2014, 50(1): 161–166.
- [8] 顾绍通. 基于分形几何的甲骨文字形识别方法[J]. 中文信息学报, 2018, 32(10): 138–142.
- GU Shaotong. Identification of oracle-bone script fonts based on fractal geometry[J]. Journal of Chinese information processing, 2018, 32(10): 138–142.
- [9] 李宗焜. 甲骨文字编[M]. 北京: 中华书局, 2012.
- [10] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules[C]//Proceedings of the 31st Conference on Neural Information Processing Systems. Long Beach, USA, 2017: 3856–3866.
- [11] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278–2324.
- [12] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, USA, 2012: 1097–1105.
- [13] HINTON G, SABOUR S, FROSST N. Matrix capsules with EM routing[C]//Proceedings of the 6th International Conference on Learning Representations, ICLR. 2018.
- [14] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 580–587.
- [15] GIRSHICK R. Fast R-CNN[C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, USA, 2015: 1440–1448.
- [16] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The journal of machine learning research, 2014, 15(1): 1929–1958.
- [17] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks[C]//Proceedings of the 14th International Conference on Artificial Intelligence and Statistics. Fort Lauderdale, USA, 2011: 315–323.
- [18] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[EB/OL]. [2018-10-15] <https://arxiv.org/abs/1502.03167v1>.
- [19] KINGMA D P, BA J. Adam: a method for stochastic optimization[EB/OL]. [2018-10-05] <https://arxiv.org/abs/1412.6980>.
- [20] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 2818–2826.
- [21] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 770–778.
- [22] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 1251–1258.
- [23] PEREZ L, WANG J. The effectiveness of data augmentation in image classification using deep learning[EB/OL]. [2018-09-05] <https://arxiv.org/abs/1712.04621>.

作者简介:



鲁绪正, 硕士研究生, 主要研究方向为机器学习、深度学习、人工智能。



蔡恒进, 教授, 博士生导师, 主要研究方向为区块链技术、人工智能、软件工程、金融信息工程。卓尔智联研究院执行院长, 中国科学院深圳先进技术研究院多媒体集成技术研究中心客座研究员, 中国人工智能和大数据百人会专家委员, 中国通信工业协会区块链专业委员会副主任委员。著作《机器崛起前传——自我意识与人类智慧的开端》获得 2017 年吴文俊人

工智能科学技术奖。发表学术论文 100 余篇, 授权国家专利 4 项。



林莉, 硕士研究生, 主要研究方向为人工智能。

第 16 届中国智能系统会议 (CISC2020) The 16th Chinese Intelligent System Conference

中国智能系统会议是由中国人工智能学会智能空天系统专业委员会发起的系列学术会议, 其宗旨是为本领域的专家学者、研究生以及工程技术人员提供一个学术交流的平台, 以推动我国智能系统相关理论、技术与应用的发展。第 16 届中国智能系统会议 (CISC2020) 将于 2020 年 10 月 24-25 日在广东省深圳市召开。本次会议由中国人工智能学会主办, 中国人工智能学会智能空天系统专业委员会协办, 哈尔滨工业大学(深圳)与北京航空航天大学联合承办。会议论文集将由 Springer 出版社在 Lecture Notes in Electrical Engineering 系列正式出版, EI 收录。

投稿要求:

1. 论文未曾在国内外杂志或会议上发表。
2. 稿件写作必须使用英文, 并严格按照 LaTeX 模板要求进行排版。
3. 论文采用网上投稿, 投稿系统网址为

<https://easychair.org/conferences/?conf=cisc20200>

论文投稿截止日期: 2020 年 5 月 31 日

论文录用通知日期: 2020 年 6 月 20 日

会议注册/终稿提交截止日期: 2020 年 7 月 10 日