



南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

本科生毕业设计(论文)

题 目: 基于邻近坐标注意力神经网络的甲骨文识别算法研究

姓 名: 孙清扬

学 号: 11912236

系 别: 计算机科学与工程系

专 业: 计算机科学与技术

指导教师: 刘 江

年 月 日

诚信承诺书

1. 本人郑重承诺所呈交的毕业设计(论文)，是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料均真实可靠。

2. 除文中已经注明引用的内容外，本论文不包含任何其他人或集体已经发表或撰写过的作品或成果。对本论文的研究作出重要贡献的个人和集体，均已在文中以明确的方式标明。

3. 本人承诺在毕业论文(设计)选题和研究内容过程中没有抄袭他人研究成果和伪造相关数据等行为。

4. 在毕业论文(设计)中对侵犯任何方面知识产权的行为，由本人承担相应的法律责任。

作者签名：

_____年____月____日

基于邻近坐标注意力神经网络的甲骨文识别算法研究

孙清扬

(计算机科学与技术系 指导教师：刘江)

[摘要]：甲骨文(Oracle Bone Inscriptions, OBIs)是中国古代象形文字的一种，也是许多亚洲文字的前身。甲骨文的研究对于了解商周历史，以及汉字起源有重要意义。甲骨文的识别是甲骨文研究的关键研究内容。目前，该工作主要依靠甲骨学专家的人工完成，该方式效率较低且消耗人力资源。因此，研究甲骨文的自动识别技术很有价值。近年来，深度学习方法已逐渐被应用于甲骨文识别，但这些方法往往不能充分利用甲骨文字符的字形特征信息。

针对上述问题，本文对甲骨文自动识别开展了研究，研究内容概括如下：

(1)本文提出了一种新的邻近坐标注意力(Local Cross-position Interaction, LCI)模块，该模块通过利用甲骨文的字形结构特征，动态调整卷积神经网络(Convolutional Neural Network, CNN)中特征图的相对重要性。LCI 模块通过方向池化提取字形的上下文信息，然后通过局部空间邻近交互构建字形上下文特征之间的局部依赖关系。

(2)本文将邻近坐标注意力模块与残差模块结合，形成残差-邻近坐标注意力模块，并通过叠加多个残差-邻近坐标注意力构建邻近坐标注意力网络(Local Cross-position Interaction Network, LCINet)。

(3)本文构建了甲骨文图像数据集 OBI316 以验证 LCINet 的有效性。该数据集类别数达 316，样例数达 17886。在 OBI316 数据集和 CIFAR

10/100 数据集上的综合实验表明，本文提出的 LCINet 优于基准方法和最先进的基于注意力的网络。

[关键词]： 甲骨文；字符识别；邻近坐标；卷积神经网络；注意力机制

[ABSTRACT]: Oracle Bone Inscriptions (OBIs) are a type of ancient Chinese hieroglyphs, which are precursors of many Asian characters. The study of oracle bone inscriptions is of great importance for understanding the history of the Shang and Zhou dynasties, as well as the origins of Chinese characters. The recognition of OBIs is a key part of the study of oracle bone inscriptions. At present, this work mainly has to be executed manually by oracle experts, which is inefficient and labour-intensive. Therefore, it is of great value to research automatic oracle bone recognition techniques. Recently, deep learning methods have been gradually applied to OBIs recognition, but these methods often fail to leverage the glyphological feature information of oracle characters.

In view of the above-mentioned problems, this thesis conducts research on automatic OBIs recognition. The main work of this thesis is summarized as follows:

(1) This thesis proposes a novel Local Cross-position Interaction (LCI) module, which dynamically adjusts the relative importance of feature maps in convolutional neural networks (CNNs) by exploiting the potential of the glyphological information. LCI extracts the glyphological context information by orientation pooling, then constructs the local dependencies between glyphological context features via a local cross-position interaction.

(2) This thesis combines the LCI module with the residual module to form the Residual-LCI module and build a Local Cross-position Interaction Network (LCINet) by stacking multiple Residual-LCI modules.

(3) This thesis constructs an OBIs dataset named OBI316 with 316 classes and 17886 instances to verify the effectiveness of LCINet. The comprehensive experiments on an OBI316 dataset and CIFAR10/100 datasets demon-

strate that the proposed LCINet outperforms baselines and state-of-the-art attention-based networks.

[Keywords]: Oracle Bone Inscription; Character Recognition; Local Cross-position Interaction; Convolutional Neural Network; Attention Mechanism

目录

1. 绪论.....	9
1.1 甲骨文的起源与甲骨学的发展.....	9
1.2 甲骨文字形结构特征.....	10
1.3 相关工作.....	11
1.3.1 甲骨文识别.....	11
1.3.2 注意力机制.....	12
1.4 论文主要研究内容及章节安排.....	12
1.4.1 研究内容.....	12
1.4.2 章节安排.....	13
2. 甲骨文数据集的构建.....	14
2.1 数据集图像来源与预处理.....	14
2.2 构建实验数据集.....	14
3. 邻近坐标注意力网络.....	15
3.1 挤压激励注意力.....	15
3.2 邻近坐标注意力模块.....	16
3.2.1 坐标方向池化.....	17
3.2.2 局部空间邻近交互.....	17
3.3 网络结构.....	18
4. 实验结果及分析.....	19
4.1 实验设置.....	19

4.2 评价标准.....	20
4.3 实验结果.....	20
4.3.1 OBI316 数据集的性能验证.....	20
4.3.2 CIFAR10/100 数据集的性能验证.....	21
4.3.3 k_1 与 k_2 对网络性能的影响.....	22
5. 总结与展望.....	23
5.1 总结.....	23
5.2 展望.....	23
参考文献.....	25

1. 绪论

1.1 甲骨文的起源与甲骨学的发展

汉字是汉语的记录符号，历经了千年演变，蕴含着中华民族的历史血脉与文化底蕴。甲骨文是迄今为止我国发现的最早的成熟文字系统，出现于商周时期，因镌刻于龟甲、兽骨之上而得名^[1]，如图 1 所示。甲骨文常用以商周王室贵族的占卜与记事^[2]，记录了距今约 3600 多年的殷商时期的历史，为历史学家提供广泛的研究史料^[3,4]。甲骨文的研究不仅对了解中国古代历史文化十分重要，而且对汉字乃至全世界文字的溯源与传承有非比寻常的意义。2017 年 12 月 26 日，甲骨文成功入选“世界记忆名录”，甲骨文的重要性已经得到中国乃至全世界的公认^[5]。

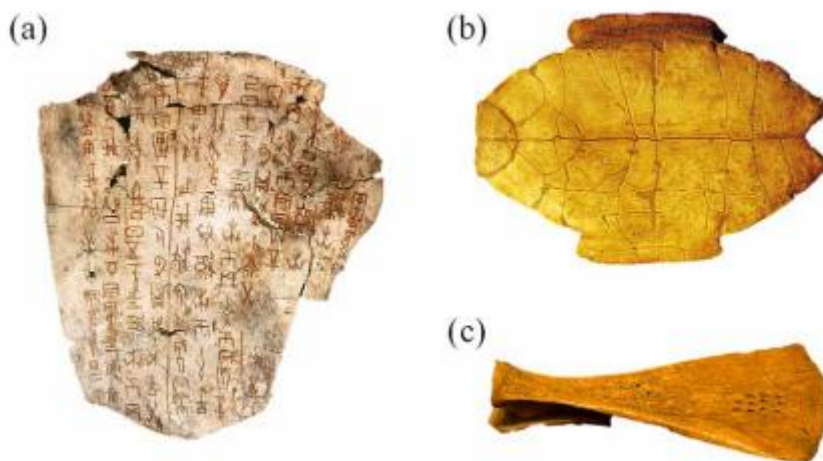


图 1' 甲骨文文字载体 (a) 肩胛骨 (b) 龟甲 (c) 牛骨

由于甲骨文字体繁琐、书写不便，随着历史的演进，甲骨文曾一度失传，在多个朝代不为人所知，直到清朝末期甲骨文才首次被发现。1899 年，清末官员、金石学家王懿荣在中药材“龙骨”中首次发现甲骨文，这一重大发现随即引起中外学术界轰动。百年间，曾有数十次以甲骨文相关历史材料为目标的考古发掘活动。迄今为止，作为甲骨文主要发掘地的河南省安阳市殷墟已经出土了超过 15 万片的字甲骨^[6]。

自甲骨文首次出土之后，历经古今中外的甲骨学者对甲骨文的不断研究与探索，甲骨学不断发展^[7]。其中，甲骨文释读是甲骨学重要分支之一，这指的是确定一个甲骨文字的音、形、义。甲骨文的释读对破解商周史料、探索文字演化都

有十分重要的意义。然而作为拥有超过 3000 年历史的文字体系，对甲骨字符的释读与破译存在一定难度。在目前已挖掘的 15 万余片甲骨中，约有 4500 个不同的字符，然而其中仅有约 2000 个字符被成功释读^[8]。

甲骨文识别指确定甲骨文字符类别，甲骨文正确的分类是释读的重要先决条件。目前甲骨文识别主要依赖于学科专家的识别工作，往往需要专家有丰富的文字学、历史学有关知识与经验。这种专家识别方式的成本较高且效率较低。

如果要提高甲骨文研究的工作效率，实现甲骨文的自动识别技术是十分必要的。近年来，随着人工智能技术的兴起，特别是深度学习已经成功应用于很多领域，如图像分类与识别。将人工智能技术应用于甲骨文识别，可以很好地利用计算机的高算力，节省人力资源与时间资源，从而极大地提高甲骨文识别效率，助力甲骨学的研究工作。然而，现有的基于深度学习的甲骨文识别方法主要关注构建卷积神经网络(Convolutional Neural Network, CNN)的结构上，鲜有研究工作关注甲骨文的字形结构特征。然而这些字形特征是历史上众多学者反复研究、探讨所形成的甲骨文释读方法体系中的重要组成部分，且如今也为甲骨文研究学者所常用。因此，将字形结构信息引入甲骨文识别框架中，有助于甲骨文识别效果的进一步提升。

1.2 甲骨文字形结构特征

甲骨文的考释是复杂的过程，有时一个甲骨文字的释读过程需要数十年的时间^[9]。文字学考释领域已形成多套成体系方法与多部代表性著作，如唐兰《古文字学导论》^[10]中介绍了三种甲骨文考释方法：因袭对比法、推勘法与偏旁分析法。其中，偏旁分析法^[11]指将汉字整体分解成偏旁。由于偏旁是构成整字的低一层的单元，且具有一定的文字学意义，因此一个未识别的字符可以当做偏旁的组合。如图 2 所示，将图 2(b)视为未识别的甲骨文字符，可将其分为上、下两部分。由于这两部分所代表的字符已被释读，因此可以推断出未识别字符的类别。

此外，汉字具有间架结构信息。间架结构指汉字的笔画组合成字的形式与规律，可大致分为 4 类 13 种，如上下结构、左右结构、半包围结构等，如图 3 所示，图 3(a)为上下结构汉字，图 3(b)为左右结构，图三(c)为半包围结构汉字。

偏旁分析法与间架结构信息分别从局部与整体的角度识别甲骨文字，这可以

为甲骨文字自动识别网络的设计带来启发。

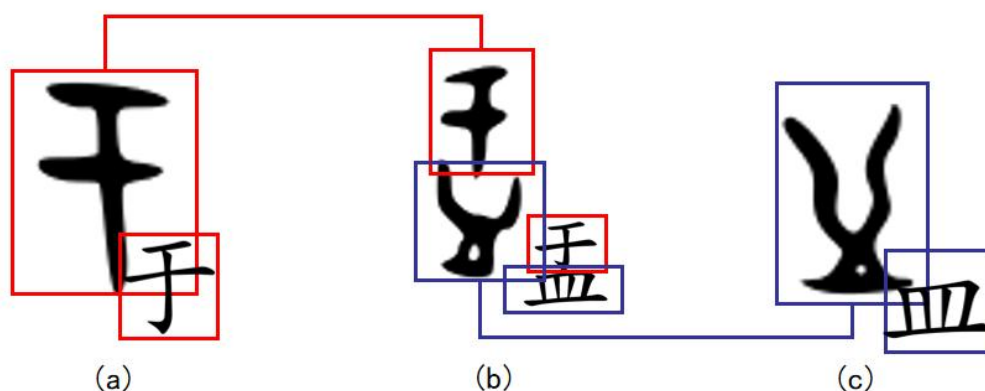


图 2 根据偏旁分析法将未识别的甲骨文字分解

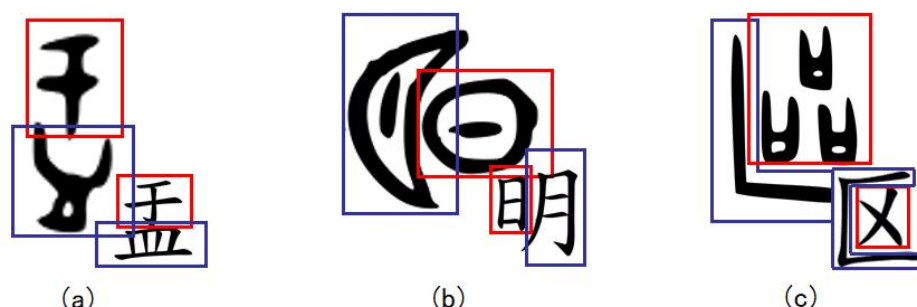


图 3 汉字的间架结构

1.3 相关工作

1.3.1 甲骨文识别

多年来，研究人员已通过多种方法对甲骨文识别进行了尝试。Li 等人提出了一种甲骨文编码技术，将字符视为无向图，并提取其拓扑特征^[12]。栗青生等提出了一种将甲骨文转化为有标签无向图的方法，并基于甲骨文对应的有标签无向图的拟邻接矩阵编码，进而基于图同构算法识别字符^[13]。Meng 等人基于模板匹配研究了甲骨文识别技术，并在提取文字特征时使用四向扫描标记法(four-directional scan labeling)降低噪声^[14]。刘永革等使用支持向量机(Support Vector Machine, SVM)技术进行甲骨文识别^[15]。

近来，随着深度学习的快速发展，一些研究人员尝试基于深度神经网络来识别甲骨文。Meng 等人通过扫描甲骨文书籍，并使用算法进行数据增强，从而构建了实验的甲骨文数据集，之后使用预训练的 AlexNet 进行甲骨文识别^[16]。Liu 等人基于深度卷积神经网络解决了一个不完整的甲骨文图像识别问题^[17]。Fujikawa

等采用 YOLO(You Only Look Once)检测龟甲拓片上的甲骨文,之后通过 Mobile-Net 识别检测的甲骨文^[18]。Yue 等人提出了一个改良的 Wasserstein GAN(WGAN)来生成甲骨文图像进行训练,然后用一个基于 CNN 的神经网络进行甲骨文识别^[19]。Gao 等人基于 VGG16 完成甲骨文识别任务,且该模型可以较好地识别甲骨文变体^[20]。然而,现有的框架未能利用甲骨文的字形特征,如汉字间架结构特征与偏旁部首信息。因此,甲骨文识别工作的效率和准确性还可以进一步提高。这些汉字结构特征可以通过注意力机制来被网络学习。

1.3.2 注意力机制

众所周知,人脑视觉系统具有复杂的机制^[21,22],该机制可以将人脑注意力转移到视觉输入中更重要的区域。基于这一观察,研究者发明了模仿人类视觉系统这种特性的算法,注意力机制被引入计算机视觉。这种将注意力转移到图像最重要区域而忽略无关部分的方法称为注意力机制。它已被广泛应用于许多视觉任务中,例如:图像分类、目标检测、语义分割、医疗图像处理等。

对于图像分类任务,已有一些被广泛使用的注意力模块。挤压激励注意力(Squeeze-and-Excitation, SE)^[23]将全局信息压缩,之后构建通道间的依赖关系并校准通道的权重。Wang 等人提出了高效通道注意力(Efficient Channel Attention, ECA)^[24],该方法提取通道邻近信息,避免引入冗余的通道全局依赖信息,从而改进了 SE。Gao 等人引入了全局二阶池化卷积网络(Global Second-Order Pooling Convolutional Networks, GSop-Net)通过协方差矩阵的高阶统计量有效地增强了全局信息的表示^[25]。基于风格校准模块(Style-based Recalibration Module, SRM)在提取全局信息时,引入风格特征,从而增强了神经网络的特征表达能力^[26]。不同于上述方法,本文提出的方法充分利用了甲骨文的字形特征,从而提高甲骨文识别效果。

1.4 论文主要研究内容及章节安排

1.4.1 研究内容

本文的总体目标是针对当前大规模识别甲骨文工作的问题需求,结合卷积神经网络与注意力机制的相关技术,对甲骨文识别进行深入研究。根据该目标,本

文的主要研究内容如下：

1.本文提出了一种新的注意力模块：邻近坐标注意力模块，该模块可以将甲骨文的字形特征信息引入特征图中，进而提高卷积神经网络的甲骨文识别性能。

2.本文通过将邻近坐标注意力模块与残差块(Residual Block)结合，并叠加多个残差邻近坐标注意力模块构建了一个邻近坐标注意力网络来完成甲骨文识别任务。邻近坐标注意力网络可以有效地学习甲骨文字形特征信息，从而提高识别效果。

3.本文构建了一个甲骨文数据集，OBI316，来验证邻近坐标注意力网络的有效性。通过 OBI316 和 CIFAR10/100 数据集上的实验表明，本文提出的邻近坐标注意力网络优于其他先进的注意力网络。

1.4.2 章节安排

第一章首先介绍了甲骨文的起源与甲骨学的发展；然后介绍了传统的甲骨文释读方法中利用的甲骨文字形结构特征；之后阐述了甲骨文自动识别与注意力机制的相关工作；最后概述了本文的研究内容与章节安排。

第二章介绍了甲骨文数据集 OBI316 的构建。首先介绍了该数据集中图像的来源与预处理方式；其次介绍了为进行实验而对数据集进行的分割。

第三章介绍了邻近坐标注意力。首先回顾了经典注意力模块：挤压激活注意力的实现方式；之后介绍了邻近坐标注意力模块的设计；最后描述了邻近坐标注意力网络的结构。

第四章介绍了实验结果。该章节首先介绍了实验设置，然后介绍了实验结果的评价标准，最后分别介绍了邻近坐标注意力网络在 OBI316、CIFAR10/100 数据集上与先进注意力网络的性能比较，之后对邻近坐标注意力网络中的超参数进行了消融实验。

第五章总结了本文的主要研究内容，并对未来的研究方向进行展望。

2. 甲骨文数据集的构建

2.1 数据集图像来源与预处理

本论文数据集的图像来源为一个公开的甲骨文图像网站，字源网站¹。该网站由汉字爱好者汉字叔叔创建，收录了部分汉字的甲骨文、金文与小篆图像。该数据集噪声较少、文字清晰。本论文编写了爬虫代码从该网站获取了共 558 个字符的 18568 张图像。

然而该数据集也存在一些不足。部分样本存在随机墨点噪音，为提高后续模型训练的效果，本研究使用 `opencv` 库中的中位数去噪算法进行预处理，处理前后效果如图 5。

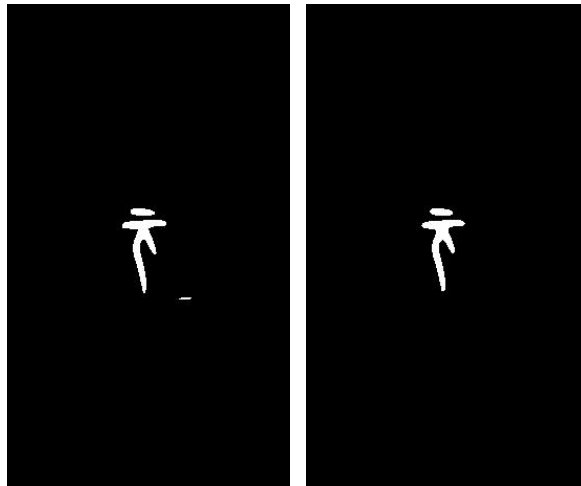


图 4 数据预处理前(左)与数据预处理后(右)的对比

2.2 构建实验数据集

通过编写脚本统计该原始数据集中各字符包含的图片数量，并按图片数量分类对字符分类，统计结果如图 5。其中，仅包含小于等于 2 张图片的字符占 20%，共 113 类字符；包含 3 或 4 张图片的占 12%，共 70 类字符；包含 5~9 张的占 11%，共 60 类字符；而包含大于等于 10 张图片的字符占 57%，共 316 类字符。可见，该原始数据集的数据分布不均匀，有些字符只包含少量图像。因此，本论文舍弃了图像少于 10 张的甲骨文字符类，构建了实验用的甲骨文图像数据集 OBI316。该数据集包含 316 个甲骨文字符的 17886 张图像。在实验中，本论文将 OBI316

¹ <https://hanziyuan.net>

分成三个不相关的数据集：训练集、验证集与测试集。其中，训练集有 12522 张图片，验证集有 1788 张图片，测试集有 3576 张图片。

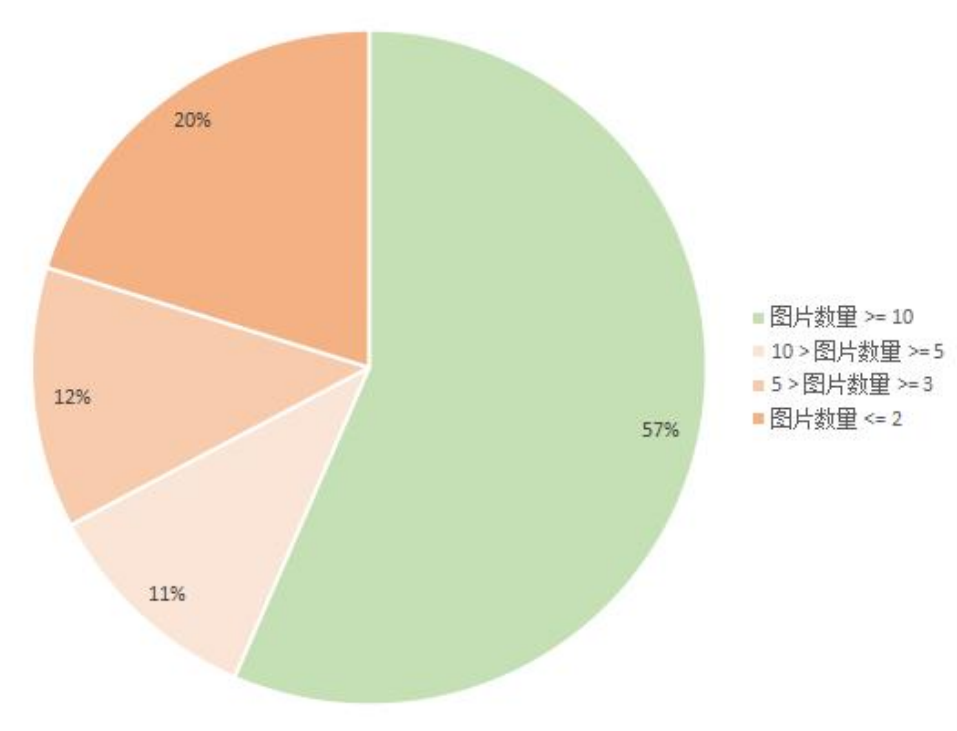


图 5 原始数据集的数据分布

3. 邻近坐标注意力网络

3.1 挤压激励注意力

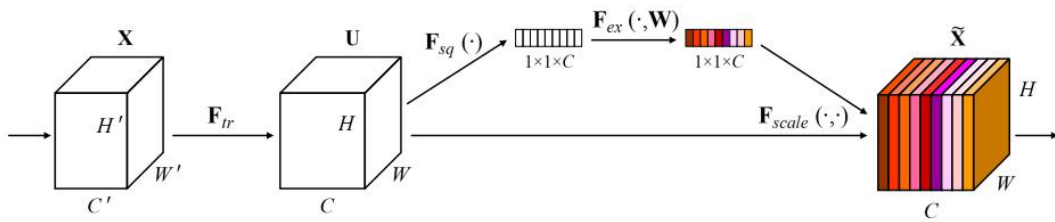


图 6 挤压激励注意力的结构图

挤压激励注意力是一种常见的注意力模块，其可以构建通道之间的相关性，并重新校准各通道权重。该注意力包含两个运算符：挤压运算符与激励运算符。假设输入特征图 $X \in \mathbb{R}^{C \times H \times W}$ (其中 C, H, W) 与分别表示通道数，特征图的高度和宽度)。挤压运算符通过全局平均池化(Global Average Pooling, GAP)压缩并提取各通道特征图的全局信息，该过程可以表述为：

$$z_c = \frac{1}{H \times W} \sum_{j=1}^H \sum_{i=1}^W X_c(i, j) \quad (1)$$

其中 z_c 是第 c 个通道的全局特征。之后，激励运算符对各通道间的全局依赖关系进行建模，并生成重校准的特征图，该过程可表示为：

$$\hat{X} = \sigma(W_2(\text{ReLU}(W_1(z)))) \cdot X \quad (2)$$

其中 \cdot 是逐通道乘积， σ 表示 sigmoid 激活函数， W_1 和 W_2 是两个全连接(Fully-Connected, FC)层的权重。挤压激励注意力有效地提取并构建了通道间的相关关系，然而却压缩了全局上下文特征，忽略位置信息，而特征图的位置信息对甲骨文的识别具有重要意义。接下来，本文将提出一种新的注意力模块，该模块可以充分地利用甲骨文图像的位置信息。

3.2 邻近坐标注意力模块

如图 7(a)所示，本文提出邻近坐标注意力模块包括两个运算符：坐标方向池化与局部空间邻近交互。邻近坐标注意力模块将 $X \in \mathbb{R}^{C \times H \times W}$ 作为输入，并输出一个具有位置特征增强表示的特征图 $Y \in \mathbb{R}^{C \times H \times W}$ ，其尺寸与 X 相同。

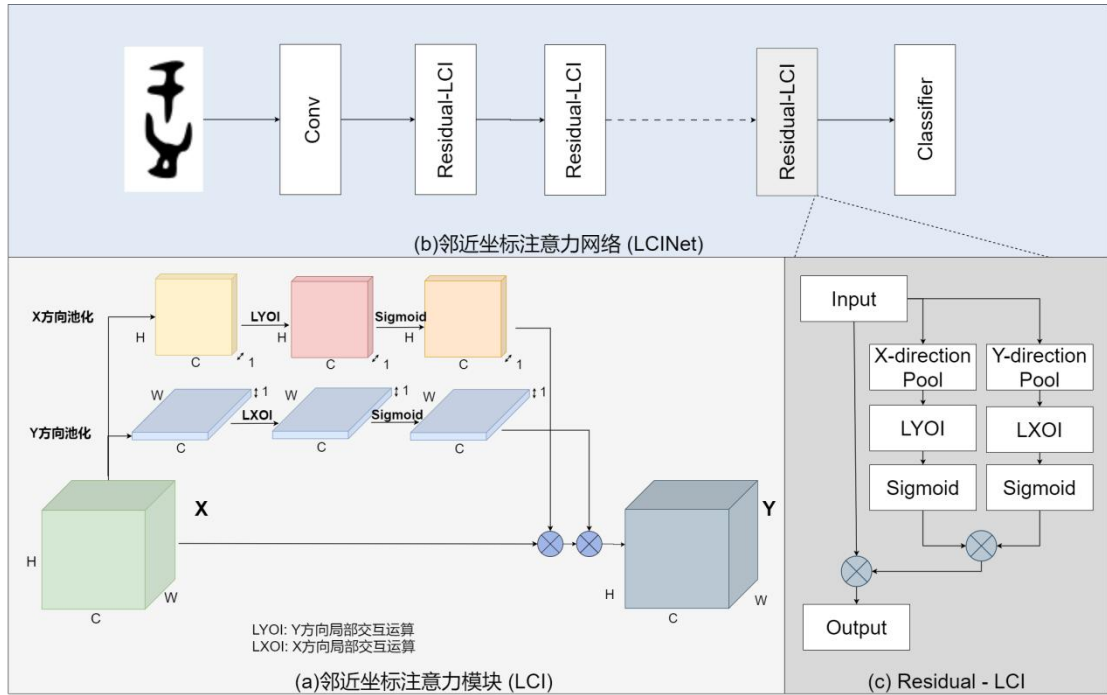


图 7 面向甲骨文识别的邻近坐标注意力网络

3.2.1 坐标方向池化

全局平均池化被广泛应用于全局信息的提取。然而这种信息提取方式对于甲骨文图像而言,会导致坐标轴方向信息的损失,导致网络学习字形特征效果较差。本文提出了一种坐标方向池化的方法来提取垂直 X 方向和水平 Y 方向上的坐标方向上下文信息。对于输入 X , 坐标方向池化包含两种方向池化操作: X 方向池化和 Y 方向池化, 分别提取垂直方向和水平方向的上下文特征, 可以表述为:

$$z_c^w = \frac{1}{H} \sum_{j=1}^H X_c(w, j) \quad (3)$$

$$z_c^h = \frac{1}{W} \sum_{i=1}^W X_c(i, h) \quad (4)$$

其中 $z_c^w \in \mathbb{R}^{C \times 1 \times W}$ 表示沿 X 方向的通道描述符, 该描述符中包含坐标方向信息。 $z_c^h \in \mathbb{R}^{C \times H \times 1}$ 为沿 Y 方向的通道描述符。

与挤压激活注意力所使用的全局平均池化方式不同, 本文提出的坐标方向池化运算符分别沿着 X、Y 两个方向提取空间信息。为了增强这两个包含坐标方向信息的通道描述符中的甲骨文字形特征信息, 抑制冗余信息, 本文提出了一种局部空间邻近交互运算符。SE 中的 FC 层构建了全局的通道间的依赖关系, 而本文的局部空间邻近交互运算符仅在其周围 k 个相邻位置中建立局部的空间依赖关系。

3.2.2 局部空间邻近交互

综上所述, 公式(3)与公式(4)通过坐标方向池化编码了具有坐标方向信息的空间信息。根据甲骨文的字形特征, 本文认为该空间信息中的每个特征均与其周围 k 个相邻位置存在空间依赖关系, 而编码这种空间依赖关系有助于提升甲骨文识别网络的性能。因此, 本文设计了一种局部空间邻近交互运算符, 该运算符包含 X 方向局部交互运算(Local X-orientation Interaction, LXOI)与 Y 方向局部交互运算(Local Y-orientation Interaction, LYOI), 如图 7(a)所示。

给定水平坐标方向上下文特征 Z^H , X 方向局部交互运算可以构建该坐标方向每个特征与其周围 k 个邻居之间的关系。本文通过在不同阶段采用自适应核大

小 k_1 与 k_2 的快速一维深度卷积(Depthwise Convolution, DW-Conv)^[27]来实现这一点。以卷积核大小 k_1 为例, LXOI 可以表示为:

$$g_c^h = \sigma(DC1D_{k_1}(z_c^h)) \quad (5)$$

其中 g_c^h 是注意力权重, σ 是 sigmoid 函数, $DC1D_{k_1}$ 是卷积核为 k_1 的一维深度卷积。

同样, Y 方向局部交互运算的实现方式如下:

$$g_c^w = \sigma(DC1D_{k_1}(z_c^w)) \quad (6)$$

此处 g_c^w 表示注意力权重。本文将在在第四章的实验中讨论自适应的不同尺寸的深度卷积核大小 k_1 和 k_2 对 LXOI 和 LYOI 的影响。

最后, 含有位置信息的特征图 $Y \in \mathbb{R}^{C \times H \times W}$ 中可由如下得:

$$y_c = x_c \cdot g_c^w \cdot g_c^h \quad (7)$$

其中 y_c 是第 c 个通道的输出。

3.3 网络结构

本文的研究内容是构建一个可以充分利用字形信息的甲骨文识别网络, 以提高甲骨文识别性能。为了验证本文提出的邻近坐标注意力模块的有效性, 本文以 ResNet18 和 ResNet50^[28]作为网络骨架以进行实验。目前最先进的注意力方法通常是建立在 ResNet 上的, 因此本文也选择 ResNet 作为网络的骨架。

本文将邻近坐标注意模块与残差块结合, 构造了残差-邻近坐标注意力模块, 如图 7(b)所示。本文提出的邻近坐标注意力网络(Local Cross-position Interaction Network, LCINet)由多个残差-邻近坐标注意力模块堆叠而成, 如图 7(c)所示。此外, 该网络采用 Softmax 函数作为分类器, 采用交叉熵损失函数为损失函数。表 1 列出了 LCINet18 与 LCINet50 的结构。LCINet18 包含四个阶段, 分别由 {2,2,2,2} 个对应的残差-邻近坐标注意力模块构成; LCINet50 有四个阶段, 分别由 {3,4,6,3} 个对应的残差-邻近坐标注意力模块构成。

表 1 LCINet18 与 LCINet50 网络结构表

阶段	输出尺寸	LCINet18	LCINet50
开始	112×112	7×7 , 64, 步长=2	
	56×56	3×3 , 最大池化, 步长=2	
1	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \\ LCI, k_1 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \\ LCI, k_1 \end{bmatrix} \times 3$
2	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \\ LCI, k_1 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \\ LCI, k_1 \end{bmatrix} \times 4$
3	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \\ LCI, k_2 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \\ LCI, k_2 \end{bmatrix} \times 6$
4	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \\ LCI, k_2 \end{bmatrix} \times 2$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \\ LCI, k_2 \end{bmatrix} \times 3$
结束	1×1	GAP, FC, softmax	

4. 实验结果及分析

4.1 实验设置

本文使用 PyTorch 实现所有的方法。在训练中, 本文采用 SGD 优化器作为默认设置的优化器(动量为 0.9, 权值衰减为 $1e^{-4}$), 初始学习率 0.015, 训练周期 150。当训练周期超过 10 时, 将学习率设置为 0.0015。批大小设置为 32。图像大小设置为 224×224 像素。采用随机水平翻转、随机垂直翻转和随机旋转方法对训练图像进行增强。

本文使用 SE^[23], SRM^[26], 坐标注意力(Coordinate Attention, CA)^[29], 卷积块注意力组件(Convolutional Block Attention Module, CBAM)^[30], ECA^[24], 临床意识注意力(Clinical-awareness attention, CCA)^[31]来验证本文提出的方法的有效性。

4.2 评价标准

本文采用四个常用的指标来评估方法的性能:准确率(Accuracy, ACC)、精确率(Precision, PR)、F1、kappa 系数值(kappa)。这些评价指标可根据以下公式来计算:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$PR = \frac{TP}{TP + FP} \quad (9)$$

$$Se = \frac{TP}{TP + FN} \quad (10)$$

$$F_1 = \frac{2 \times PR \times Se}{PR + Se} \quad (11)$$

其中 TP 为真阳性(True Positive), TN 为真阴性(True Negative), FP 为假阳性(False Positive), FN 为真阴性(True Negative)。此外,准确率表示图像被正确识别的百分比;精确率表示一个算法预测为正值的样本占实际为正值的比例;F1 可以看作是敏感度和精确率的调和平均;kappa 可以用来评估分类任务模型,特别是对于分布不均匀的数据集。

4.3 实验结果

4.3.1 OBI316 数据集的性能验证

本文首先将邻近坐标注意力网络与其他先进的注意力在 OBI316 数据集上进行比较,实验以 ResNet18 和 ResNet50 为网络骨架。

实验结果如表 2 所示。在与复杂度相似的注意力网络的比较中,邻近坐标注意力网络稳定地高了甲骨文识别性能。LCINet50 比 CA-ResNet50 节省了 15.54% 的参数量,却在四个评估指标(ACC, PR, Se, Kappa)上较 CA 提高超过 1.1%。LCINet50 与 SE-ResNet50 复杂度相当,且使用参数量略少,却在四个评估指标(ACC, PR, Se, Kappa)上较 SE 提高了 1.2%。实验结果进一步表明,将甲骨文的字形结构特征通过坐标方向池化与局部空间邻近交互的方式引入卷积神经网络的特征表示中,可以较好地提高甲骨文识别网络的性能。

表 2 不同注意力机制网络在 OBI316 数据集上分类结果对比

方法	ACC	PR	F1	Kappa	Params	FLOPs
ResNet18	89.36	90.01	89.76	89.94	11.24M	1.82G
SE-ResNet18	89.67	89.77	89.63	89.84	11.43M	1.82G
SRM-ResNet18	89.56	90.12	89.45	89.78	11.34M	1.82G
CA-ResNet18	89.78	89.99	90.34	89.67	11.47M	1.83G
CBAM-ResNet18	89.94	89.65	90.32	90.24	11.43M	1.83G
ECA-ResNet18	90.06	89.97	90.32	90.16	11.34M	1.82G
CCA-ResNet18	90.03	89.76	89.54	90.10	13.43M	1.83G
LCINet18	90.64	90.54	90.77	90.46	11.34M	1.83G
ResNet50	89.99	89.34	89.76	90.01	24.15M	4.13G
SE-ResNet50	90.27	90.54	90.61	89.89	26.68M	4.14G
SRM-ResNet50	90.46	90.68	91.01	90.97	24.15M	4.13G
CA-ResNet50	90.54	90.68	90.64	89.98	27.96M	4.20G
CBAM-ResNet50	90.97	90.31	91.15	91.01	26.69M	4.14G
ECA-ResNet50	90.76	90.10	90.54	90.94	24.16M	4.14G
CCA-ResNet50	90.85	91.04	91.71	91.16	24.53M	4.20G
LCINet50	91.67	92.01	91.81	91.89	24.20M	4.15G

4.3.2 CIFAR10/100 数据集的性能验证

为了进一步验证本文提出的邻近坐标注意力网络的性能，本文在 CIFAR-10/100 数据集^[32]上进行了实验。这两个数据集由 5 万张训练图像和 1 万张测试图像组成，每张图像的尺寸为 32×32 像素。CIFAR-10 有 10 个类，CIFAR-100 有更细粒度的 100 个类。CIFAR-10/100 常用于验证分类网络性能。

在训练中，本文使用 SGD 优化器训练。训练周期为 200，批量大小设置为 128，该设置与^[30]对应。初始学习率设置为 0.1，每 40 个周期降低 10 倍。本文将每个图像填充 4 个值为零的像素，然后随机将填充后的图像裁剪回原始图像大小作为数据增强方式。如表所示，与其他先进的注意力网络相比，本文提出的邻近坐标注意力网络在效率与有效性之间保持了较好的平衡。

表 3 不同注意力机制网络在 CIFAR10/100 数据集上分类结果对比

方法	CIFAR-10			CIFAR-100		
	ACC	Params(M)	GFLOPs	ACC	Params(M)	GFLOPs
ResNet18	93.02	11.17	0.557	74.56	11.22	0.557
SE-ResNet18	94.84	11.26	0.557	75.19	11.31	0.557
SRM-ResNet18	93.32	11.18	0.557	73.89	11.22	0.557
CA-ResNet18	95.21	11.33	0.558	77.73	11.36	0.558
CBAM-ResNet18	95.19	11.26	0.557	77.82	11.31	0.557
ECA-ResNet18	93.12	11.18	0.557	74.43	11.23	0.557
LCINet18	95.44	11.18	0.558	78.91	11.23	0.558
ResNet50	93.62	23.52	1.311	78.51	23.71	1.312
SE-ResNet50	95.35	26.06	1.316	79.28	26.24	1.316
SRM-ResNet50	94.05	23.53	1.312	76.93	23.71	1.312
CA-ResNet50	95.40	27.34	1.346	79.87	27.52	1.347
CBAM-ResNet50	95.70	26.05	1.319	80.13	26.24	1.319
ECA-ResNet50	94.00	23.53	1.313	78.07	23.71	1.314
LCINet50	95.47	23.58	1.318	80.29	23.76	1.318

4.3.3 k_1 与 k_2 对网络性能的影响

如表所示，本文提出的邻近坐标注意力网络包含超参数 k_1 和 k_2 。为探究上述参数对邻近坐标注意力网络性能的影响，本实验将 LCINet18 和 LCINet50 的不同阶段设置中的 k_1 和 k_2 设置为 3~7 的值。从表 4 的性能结果来看，设置 $k_1=5$ 、 $k_2=3$ 时，LCINet18 和 LCINet50 取得最好的性能；其他设置会导致性能略微降低。实验结果表明，自适应地设置 k_1 和 k_2 对于建模空间邻近关系是十分重要的。

表 4 LCINets 的不同 k 设置的分类结果对比

方法	设置	ACC	PR	F1	Kappa
LCINet18	$k_1 = 3, k_2 = 3$	90.31	89.98	90.27	90.12
	$k_1 = 5, k_2 = 5$	90.42	90.37	90.82	90.78
	$k_1 = 7, k_2 = 7$	89.98	90.21	89.76	89.93
	$k_1 = 5, k_2 = 3$	90.64	90.54	90.77	90.46
	$k_1 = 7, k_2 = 3$	90.14	90.00	89.54	89.96
	$k_1 = 7, k_2 = 5$	89.37	89.64	89.20	89.13
LCINet50	$k_1 = 3, k_2 = 3$	91.04	90.78	90.67	90.94
	$k_1 = 5, k_2 = 5$	91.36	91.18	91.74	91.41
	$k_1 = 7, k_2 = 7$	90.78	90.42	91.01	90.78
	$k_1 = 5, k_2 = 3$	91.67	92.01	91.81	91.89
	$k_1 = 7, k_2 = 3$	91.14	90.75	91.01	90.96
	$k_1 = 7, k_2 = 5$	90.49	90.71	91.34	91.01

5. 总结与展望

5.1 总结

本文提出了一种邻近坐标注意力网络(LCINet)来实现甲骨文的自动识别。在 LCINet 中, 通过引入甲骨文的文字特征信息构建了一个邻近坐标注意力 LCI 模块, 引导网络关注甲骨文图像中重要的特征表示, 忽略冗余的特征表示。此外, 本文还建立了一个甲骨文数据集 OBI316 以验证 LCINet 的有效性。OBI316 和 CIFAR10/100 数据集上的实验结果表明, 本文提出的方法优于最先进的注意力网络。本文的工作证明了, 在甲骨文识别任务中, 进一步挖掘字符的字形特征信息是极具启发意义的。

5.2 展望

本文对甲骨文识别进行了初步的研究, 且证明了邻近坐标注意力网络对于甲骨文识别的有效性, 然而仍然存在一定不足, 在未来的工作中可以被进一步优化

和完善。

首先，甲骨文的数据集仍需扩大。可以进一步搜集各数据源的甲骨文数据，同时可利用深度学习的数据增广方法有效扩充甲骨文数据集。

其次，本文所收集的甲骨文图片数据来源较为单一，且图像质量较好。后期需进一步采集甲骨文拓片等未经人工处理的图像，研究图像去噪算法，以及验证与拓展算法在噪声较多的甲骨文图像上的鲁棒性。

最后，甲骨文识别中语义分析是不可或缺的。考古专家释读甲骨文字时往往结合上下文，因此将甲骨文字形结构特征与甲骨文语义信息结合，可能会进一步提上甲骨文的识别效果。

参考文献

- [1] 王梦茹.基于 VGG-Siam 网络的甲骨文文字演化规律研究[D].郑州: 河南大学,2022.
- [2] 张国硕.商族的起源与商文化的形成[J].殷都学刊,1995(02):1-5.
- [3] 连劭名.卜辞所见商代的自然崇拜和巫术[J].故宫博物院院刊,2000(03):64-69.
- [4] 王佳欣.从甲骨文字形考察殷商社会的婚丧习俗[J].丽水学院学报,2014,36(01):48-51.
- [5] 林冬阳.中华瑰宝 历久弥新——写在甲骨文入选《世界记忆名录》之际[J].档案与建设,2018(02):67-68.
- [6] 胡厚宣.八十五年来甲骨文材料之再统计[J].史学月刊,1984(05):17-24.
- [7] 郭沫若.甲骨文合集[M].中华书局, 1978: 1-47
- [8] Cheung C. The Chinese History That Is Written in Bone [EB/OL]. <https://www.sapiens.org/archaeology/chinese-oracle-bones-history/>. 2018.1.23.
- [9] 裘锡圭. 殷墟甲骨文字考释(七篇)[J]. 湖北大学学报(哲学社会科学版), 1990(01): 50-57.
- [10] 唐兰. 古文字学导论[M]. 齐鲁书社, 1981.
- [11] 刘钊. 古文字构形学[M]. 古文字构形学, 2006: 23-64.
- [12] Li F, Woo P. The coding principle and method for automatic recognition of Jia Gu Wen characters[J]. International Journal of Human-Computer Studies, 2000, 53(2): 289-299.
- [13] 栗青生,杨玉星,王爱民. 甲骨文识别的图同构方法[J]. 计算机工程与应用, 2011, 47(8): 112-114.
- [14] Meng L, Fujikawa Y, Ochiai A, et al. Recognition of oracular bone inscriptions using template matching[J]. International Journal of Computer Theory and Engineering, 2016, 8(1): 53.
- [15] 刘永革, 刘国英.基于 SVM 的甲骨文字识别 [J]. 安阳师范学院学报, 2017, 2: 54 - 56.
- [16] Meng L, Kamitoku N, Yamazaki K. Recognition of oracle bone inscriptions using deep learning based on data augmentation[C]//2018 Metrology for Archaeology and Cultural Heritage (MetroArchaeo). IEEE, 2018: 33-38.
- [17] Liu M, Liu G, Liu Y, et al. Oracle bone inscriptions recognition based on deep convolutional neural network[J]. Journal of Image and Graphics, 2020, 8(4): 114-119.

- [18] Yue X, Li H, Fujikawa Y, et al. Dynamic dataset augmentation for deep learning-based oracle bone inscriptions recognition[J]. *ACM Journal on Computing and Cultural Heritage*, 2022, 15(4): 1-20.
- [19] Fujikawa Y, Li H, Yue X, et al. Recognition of oracle bone inscriptions by using two deep learning models[J]. *International Journal of Digital Humanities*, 2022: 1-15.
- [20] Gao J, Liang X. Distinguishing oracle variants based on the isomorphism and symmetry invariances of oracle-bone inscriptions[J]. *IEEE Access*, 2020, 8(99): 1.
- [21] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis [J]. *IEEE Transactions on pattern analysis and machine intelligence*, 1998, 20(11): 1254-1259.
- [22] Hayhoe M, Ballard D. Eye movements in natural behavior[J]. *Trends in cognitive sciences*, 2005, 9(4): 188-194.
- [23] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018: 7132-7141.
- [24] WANG Q, WU B, ZHU P, et al. ECA-Net: efffficient channel attention for deep convolutional neural networks, 2020 IEEE[C]//*CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. 2020.
- [25] Gao Z, Xie J, Wang Q, et al. Global second-order pooling convolutional networks[C]//*Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*. 2019: 3024-3033.
- [26] Lee H J, Kim H E, Nam H. Srm: A style-based recalibration module for convolutional neural networks[C]//*Proceedings of the IEEE/CVF International conference on computer vision*. 2019: 1854-1862.
- [27] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. *arXiv preprint arXiv:1704.04861*, 2017.
- [28] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016: 770-778.
- [29] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021: 13713-13722.
- [30] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//*Proceedings of the European conference on computer vision (ECCV)*. 2018: 3-19.

- [31] Zhang X, Xiao Z, Hu L, et al. CCA-Net: Clinical-awareness attention network for nuclear cataract classification in AS-OCT[J]. Knowledge-Based Systems, 2022, 250: 109109.
- [32] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images[J]. 2009.