学 校 代 码 10459
学号或申请号 201822362014377
密 级

郑州大学

专业硕士学位论文

基于深度学习的甲骨文字检测与提取技术研究

作 者 姓 名： 陈双浩
导 师 姓 名： 刘国英
专业学位名称： 工程硕士
培 养 院 系： 信息工程学院
完 成 时 间： 2021 年5月

A thesis submitted to
Zhengzhou University
for the degree of Master

Research of Oracle Bone Inscriptions Detection and Extraction Based on Deep Learning

Computer Science and Engineering
School of Computer Science and Engineering
June, 2021

学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的科研成果。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律责任由本人承担。

学位论文作者：陈22浩

日期：2021 年 6 月 1 日

学位论文使用授权声明

本人在导师指导下完成的论文及相关的职务作品，知识产权归属郑州大学。根据郑州大学有关保留、使用学位论文的规定，同意学校保留或向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅；本人授权郑州大学可以将本学位论文的全部或部分编入有关数据库进行检索，可以采用影印、缩印或者其他复制手段保存论文和汇编本学位论文。本人离校后发表、使用学位论文或与该学位论文直接相关的学术论文或成果时，第一署名单位仍然为郑州大学。保密论文在解密后应遵守此规定。

学位论文作者：陈22浩

日期：2021 年 6 月 1 日

摘要

甲骨文破译是甲骨学领域一项重要的研究内容，长期受到世界各地甲骨学研究学者的广泛关注。甲骨文字检测是字符破译的前提工作，目标是从甲骨片上定位字符的位置并确定相应类别。目前，这些工作需要甲骨学专家的参与手动完成，占用宝贵的专家资源且效率低下。另一方面，由于长期的自然腐蚀，拓片表面充斥着大量的噪声和各种各样的裂痕干扰，严重阻碍甲骨文字的可视性，不利于后续甲骨文字检测与识别工作，限制甲骨学的进一步推广。论文结合深度学习技术与机器学习理论，围绕甲骨文字检测与提取工作展开深入研究，具体内容如下：

针对甲骨文字检测，论文提出了一种简单且有效的甲骨字符检测器，该检测器采用非锚点框设计，使用自适应形状高斯核表示甲骨字符的空间区域，将甲骨字符的检测转换为对应高斯图的预测。其次，为避免部分字符之间排列紧密导致的区域重叠问题，字符区域分别以多个不同尺度的高斯核进行表示。最后，基于这些不同尺度的高斯核预测，采用一系列简单的图像后处理操作得到精确的字符边界框。

针对甲骨文字提取，论文首先构造一个像素水平的甲骨文临摹数据集，为字符提取工作提供数据基础。紧接着，构建了一个专门的甲骨字符提取模型。该模型充分利用基于分割方法的背景噪声去除能力和基于生成方法的结构信息描述能力，将甲骨字符提取任务视为图像到图像转换任务，其次，以生成对抗网络为模型的基础骨架，将分割网络嵌入编码器网络以消除拓片背景噪声的影响，以期建立更为准确的拓片图像与对应甲骨字符图像间的映射关系。最后，为获取内容完整且细节清晰的生成结果，使用全局和局部判别器对生成的甲骨字符图像进行一致性判别。

关键词： 甲骨文，深度学习，文字提取，文字检测

Abstract

Oracle deciphering is an important research work in the field of oracle bone, which has long attracted wide attention from oracle bones research scholars all over the world. Oracle Bone Inscriptions (OBIs) detection is the prerequisite for character deciphering, which the goal is to locate the position of the characters on the oracle bone rubbing and determine the corresponding category. It is also an important research content in the field of oracle bones. At present, these tasks require the participation of OBIs experts, combined with professional knowledge and experience to complete manually, occupying valuable expert resources and inefficient. On the other hand, due to long-term natural corrosion, the surface of rubbings is filled with a lot of noise and various cracks, which seriously hinder the visibility of OBIs, which is not conducive to subsequent OBIs detection and recognition, and restricts the further promotion of OBIs. This paper combines deep learning technology and machine learning theory to carry out in-depth research on the detection and character extraction of OBIs. It is hoped that it can provide convenience for OBIs research scholars and accelerate the further research and promotion of OBIs.

For the problem of OBIs detection, this paper provides a simple and effective OBIs detector, which adopts a non-anchor design and uses adaptive shape Gaussian kernel to represent the spatial region of oracle bone characters. That is to say that convert the detection of OBIs detection to the prediction of the corresponding Gaussian map. In addition, to avoid the problem of regional overlap caused by the close arrangement of some characters, the character regions are represented by multiple Gaussian kernels of different scales simultaneously. Finally, based on these Gaussian kernel predictions of different scales, a simple image post-processing operation is used to obtain accurate character bounding boxes.

For the problem of OBIs extraction, this article first builds the pixel-level OBIs dataset to provide a data basis for extraction work. Immediately afterwards, the paper constructed the first specialized network model for OBIs extraction. This model makes full use of the background noise removal ability based on the segmentation method and the structure information description ability based on the generation method, and treats

the OBIs extraction task as image-to-image conversion task, and the segmentation network is embedded in the encoder network to eliminate the influence of the rubbing background noise, in order to establish a more accurate mapping relationship between the rubbing image and the corresponding oracle bone character image. Finally, to obtain the generated results with complete content and clear details, the model combines a global discriminator and a local discriminator to determine the consistency of the generated oracle bone character images.

Key Words：Oracle Bone Inscriptions，Deep learning ,Text extraction，Text detection

目录

插图清单

表格清单

1 绪论

1.1 研究背景及意义

甲骨文是迄今为止中国已发现的最古老、体系最完整的文字之一，被认为是现代汉字的早期形式。文字作为人类文明的基石，是现代社会交流以及快速理解周围世界信息的重要工具。甲骨文记录着 3600 年前殷商时期先民的生活、思想状态、经济生产以及社会生活等方方面面，是研究中华文化乃至世界产生、发展的重要依据。因此，加强对甲骨文的研究对于了解中国以及世界的过去具有非常重要的意义。

甲骨文因刻录龟壳或兽骨上而得名，最早由河南省安阳市西北殷都区小屯村当地的村民发现。自此之后，一系列关于甲骨片的挖掘活动便逐步展开，但由于早期政府不够重视以及当时的农民文化欠缺等缘故，大量珍贵的甲骨片遭到流失和毁灭。许多的甲骨片背景被人私下挖掘，并转手贩卖给各地的古董商与收藏爱好者，使得诸多珍贵的甲骨资料遭受大量流失。据权威资料统计,超过 26700 片甲骨片流散到日本、美国、加拿大等 12 个国家。但值的庆幸的是，这些流失的甲骨资料受到了国外收藏者的重视，大部分甲骨片被完整的保存了下来，甚至一些国外学者也参与了关于甲骨文的文献篡改工作，为甲骨学的研究与发展做出了重要的贡献。截止至目前，河南殷墟发现的包含文字符号的甲骨高达 15 万余片，其中国内藏有 10 万余片，台湾藏有 3 万余片，香港藏有 100 片左右，其他国家共藏有 2700[1]片左右，这些甲骨片几乎涵盖了绝大多数的甲骨文字，成为甲骨学研究中唯一且重要的信息来源。

甲骨文发现之后，短短数年，经海内外各国学者前仆后继的研究，一系列甲骨文相关的基础性工作包括甲骨片挖掘、拼接缀合、辨伪、文字考释等发展迅速。如 1903 年刘鹗编写的《铁云藏龟》[2]一书出版，首部甲骨文著录就此诞生，是甲骨片成为了可以利用的研究资料。陈梦家在发表的《殷墟卜辞综述》[3]中探讨了甲骨时期与出土地区间的关系，并对各个时期的断代论据进一步讨论。李学勤等人在《殷墟甲骨分期研究》[4]中引入地层学信息用于辅助分期断代工作。如今，甲骨学已成为一门举世闻名的国际性学科，经过一百多年的发展，甲骨文在各个细分领域均取得一定的成果，这些成果促进了甲骨学的进一步研究，为甲骨文字

形、字义解析等工作提供巨大的帮助。

近年来，甲骨学发展迅速，在多个研究领域已取得显著的研究成果，但作为一种距今 3600 多年古文字体系，关于甲骨文的许多内容仍需进一步的探索。其中未知甲骨字考释一直是甲骨学相关学者研究的重要内容。目前为止，已有超过 15 万片的甲骨被挖掘出来，在这些甲骨片上已发现的约 5000 个甲骨文字中，仅有 2000 个左右成功破译[5]，剩下的甲骨字考释难度更大。另一方面，经过长期的自然腐蚀，甲骨片表面退化严重，上面的甲骨字符模糊不清，严重阻碍甲骨文字的可视性，不利于甲骨文字的检测、识别、以及推广等一系列工作，因此，当前对甲骨文研究仍然具有广泛的前景，有关甲骨文的潜在价值有待于进一步探究。

1.2 甲骨文研究现状

一直以来，甲骨学是一个极少数人参与的冷门学科，其自身的特殊性是导致这一问题的主要原因。甲骨学是一个交叉学科，需具备古文字学、考古学、历史学、文献学等学科知识背景，研究门槛高。其次甲骨文认读困难而没有“从书斋走向大众”，使得甲骨文仅仅停留在少数甲骨学专家的学术研究中。因此，在很长的一段时期内，甲骨学研究进展缓慢，几乎陷入停滞状态。

随着计算机技术和硬件设备的不断进步，借助计算机技术辅助甲骨文研究有望打破现有研究瓶颈，推动甲骨文研究进入全面深入发展和弘扬的新阶段。经过数十年国内外研究学者前仆后继的探索，甲骨文研究在在甲骨文的输入和可视化、检测、识别、图谱构建、语义分析等多个领域，取得显著的成果。例如在甲骨字的输入和可视化领域，刘永革构建了首个甲骨文可视化输入法，弥补了当前计算机的字符集中没有包含甲骨文字的空白[6]。顾绍通等分析了甲骨字形的拓扑结构，构建字形编码表和拼音编码表间的映射关系，并设计一种简单且有效的甲骨文输入编码方案[7]。在甲骨字识别领域，高峰等联合基于语境统计分析和 Hopfield 网络的模糊匹配识别方法，解决甲骨字样本中模糊字形不易识别的难题[8]。顾绍通等将图画性质的甲骨字转换为拓扑形式并对拓扑表达进行编码，最后，通过将甲骨字形的拓扑编码对通用甲骨文字库中字形的拓扑特征库进行配准，实现甲骨文字形的识别[9]。在甲骨文资料构建领域，毛建军等对国内外甲骨文全文数据库建设情况进行调查和分析，提出存在的问题并给出相应的建议[10]。

李志勇等人仿照知网构建体系设计，建立首个融合甲骨文、现代汉语的语义数据库，为甲骨文信息处理中未识甲骨字的语义推导和残缺甲骨拓片的文本内容整合提供解决思路[11]。在甲骨文语义分析领域，袁东等采用实例库构建、实例句相似度算法和实例检索等关键技术，提出一个基于实例的从甲骨文到对应译文的机器翻译方案[12]。熊晶等人引入计算机翻译辅助技术，将已经过甲骨文专家确认正确的现代汉语释读存储在翻译记忆库中，实现了专家知识的共享和重用[13]。

尽管当前甲骨文研究已经在多个领域取得显著的成就，但仍处于早期探索阶段，有关甲骨文相关的工作或技术探索（如甲骨文字的检测与识别）依然比较稀少。其次，大多数与甲骨文相关的工作仅仅将一些较为先进的技术方案直接运用到甲骨文研究数据上，缺乏理论上的探索，因此，有关甲骨文的研究工作有待于进一步深入探究。

1.2.1 甲骨文字检测技术

甲骨文破译是甲骨学领域一项重要的研究内容，长期受到世界各地甲骨学研究学者的广泛关注。甲骨文字检测是字符破译的前提工作，目标是从甲骨片上定位甲骨字符的位置并确定相应类别，同样也是该领域重要的研究内容。早些时期，甲骨文字检测工作需要甲骨学专家的参与，结合专业知识以及经验积累才得以完成，这个过程占用了昂贵的专家资源且效率低下。因此，传统的甲骨文检测方法已经捉襟见肘，难以取得较大的进展。

随着甲骨文研究数据的规模化和系统化，一些研究学者转变思路，借助计算机技术实现对甲骨文字型分析、语义计算的自动化，从而可能突破甲骨学研究瓶颈，进而取得新的进展。然而，不同于一般的自然场景或文本图像数据，甲骨拓片图像也面临一些特有的难题：（1）甲骨字符可分布在拓片图像任意区域，各个类别大小不一，尤其是一些被遮挡或尺度极小的目标，检测难度较大；（2）甲骨拓片图像表面退化严重，上面的文字模糊不清，并充斥大量的背景噪声；（3）由于长期掩埋以及民间私掘等缘故，出土的甲骨片存在破裂，形成裂痕，这些裂痕在纹理上与甲骨字符非常相似，难以区分；（4）甲骨字出现频率严重失衡。据统计，在 56743 个甲骨字样本中，包含 1425 个单字，其中，常用字 366 个，次常用字 500 个，罕见字 559[14]个。（5）异体甲骨字出现频繁，这些字符风格不一，差异极大；但总体上，甲骨文字可以视为一种特殊的多尺度小目标检测问题，这个问题需要克服复杂背景环境带来的文本定位困难以及字体外观多样性等挑战，

因此目标检测是与本论文研究的甲骨字符检测较为相近的视觉任务。理论上，这些领域中的方法能够很大程度上为甲骨字符的检测问题带来更多的受益和启发。近 20 年来，目标检测技术发展迅速，其方法按照时间可划分为两大类：传统的目标检测方法和基于深度学习的检测方法。

在人工智能正式介入视觉领域之前，传统的目标检测方法普遍采用滑动窗口或连通分量的检测路线，将检测过程分解为定位和分类两个阶段进行。基于滑动窗口的方法通常先使用不同比例大小的窗口在图像上密集滑动获取候选区域，然后根据人工设计的特征描述算子（如 SIFT[15]、Haar[16, 17]、HoG[18, 19]）提取区域特征并利用分类器（如 Adaboost[20]、SVM[21]）对特征进行分类，实现目标检测。而基于连通分量的检测方法则是先使用连通域提取技术（如颜色聚类、极大区域提取等）获取连通区域，然后使用人工制定的规则或使用人工设计的特征进行训练的分类器，滤除非物体目标的连通区域，完成目标检测。这些算法普遍使用手工设计的特征，针对不同类型的目标，需要设计不同的模型提取特征。2001 年，Viola 和 Jone 提出首个人脸检测器 Viola-Jone(VJ)，该算法在 Adaboot 算法的基础上，使用 Haar 特征和积分图的方法快速提取图像特征，并引入级联思想对训练出的强分类器进行关联，解决多尺度目标重复搜索问题[22]。2005 年，Girshick 等人提出了有效的行人检测方法，它们设计了一种 HOG（Histograms of Oriented Gradient），并使用局部区域的梯度方向计算结果统计作为该局部区域的特征，在传统目标检测算法种取得了优异的检测效果[23]。

然而，由于传统的目标检测算法局限于手工设计的特征，对图像语义表达能力较弱，生成候选区域需要大量的计算开销，检测效果精度较低且速度缓慢远远达不到实际应用需要，使得传统的目标检测技术研究面临着巨大的挑战。

随着信息数据的不断积累以及计算机硬件快速升级，以深度学习为基础的人工智能技术目前在学术界与工业界掀起了热潮，目标检测技术也迎来了新的突破。这类方法无需繁杂的图像预处理以及特征提取步骤，使用数据不断迭代优化，促使模型学习训练数据的内在数据分布。在目标检测领域，R-CNN[24]系列和 YoLo[25]系列是两种代表性目标检测算法。R-CNN 使用启发式算法 Selective Search[26]代替传统的滑动窗口，将目标检测分解成候选区域生成和区域检测两个子步骤，降低信息冗余的同时加快了检测速度。随后，SPP-Net[27]、Fast R-CNN[28]和 Faster R-CNN[29]等在 R-CNN 的基础上进一步改进，重点解决了候选区域输入大小固定、计算冗余和内存资源占用问题。YoLo 是另一种经典的目标检测算法，

该算法将目标检测作为一个回归问题，输入图像经过一次推理，直接得到物体的位置信息和所属类别，在维持准确率（63.4% mAP,VOC2007[30]）的前提下,达到45fps 的检测速度。YoLov2[31]借鉴了 Faster R-CNN 中的锚点机制并对 YoLo 进一步改进，在保持原有速度的同时精度得到大幅的提升。

鉴于深度学习在目标检测领域的成功应用，一些研究学者尝试将其扩展到甲骨文检测数据集上，以望能够取得更大的提升。例如李梦将 SSD300 扩展到SSD1024，构建了一个单阶段的甲骨字符检测模型[32]。王浩彬搭建了一个基于区域的全卷积网络 R-FCN 甲骨字符检测框架，并提出一个甲骨字符识别辅助检测算法，帮助检测模型减少对甲骨裂痕的误判[33]。邢济慈等人应用多个一流的目标检测算法在甲骨字符检测任务上，探索适宜甲骨字符检测的网络结构[34]。刘国英等人基于甲骨字符的数据特征，对锚点框的大小、宽高比重新设计，并提出(Spatial Pyramid Block)以稳定特征及缓解噪声干扰[35]。实验表明，相较于传统的检测方法，这些方法在甲骨文检测数据集在精度上取得显著提升。但鉴于大多数方法仅仅将一些较为经典的检测算法经过略微修改后直接应用到甲骨文研究数据上，缺乏基础理论探索，在检测精度和效率上仍然存在一定的局限性，因此甲骨文字检测算法仍然有进一步改进的空间。

1.2.2 甲骨文字提取技术

由于某些历史原因，甲骨片长久的掩埋在安阳地下的废墟中直到 120 年前才被发现，因此拓片表面不可避免的存在一定的退化，如噪声、裂痕等，这些不同程度的退化严重干扰了甲骨文字的可视性及可读性，对后续甲骨文字检测与识别等视觉任务带来极大的阻碍。甲骨文字提取任务是一种特殊的图像增强任务，通过移除复杂的甲骨背景，增强甲骨文信息的可视性，有利于甲骨文字检测与识别任务的性能提升。此外，提取的高质量甲骨文字还可作为甲骨片的临摹样本，省去了耗时耗力的人工操作，有助于甲骨学研究的开展，并对甲骨文活化与利用产生重大帮助。

截止到当前，有关甲骨文字提取相关的工作或技术探索几乎是一片空白。近几年来，随着深度学习在学术界和工业界的崛起，出现了一些在理论上能够直接或间接的用于提取拓片图像中甲骨字符的方法。这些方法大致分为两大类：基于图像生成的方法和基于图像分割的方法。

图像生成的方法（如 Pix2Pix[36]）将甲骨字符的提取视为一个图像到图像转

换任务，通过训练一个基于编码-解码的神经网络，学习拓片图像与相应字符图像间的映射，其结构如图 1.1 所示。Pix2Pix 是首个通用图像到图像转换框架，该框架巧妙的利用 GAN[37]捕获高维数据分布，实现图像着色、超分辨率等多种图像生成任务。其次，在网络结构上基于 DCGAN[38]，并借鉴 U-Net[39]采用残差连接，以捕获更多低频特征细节。稍后，Isola 等[40]对 Pix2Pix 继续进一步改进，提出判别器 PatchGAN。它避免直接对整个图像区域进行判别，从感受野角度对生成图像的局部区域进行判别，提升了判别器判别性能的同时，改善其鲁棒性。PAN[41]介绍了感知损失，代替传统图像转换模型使用 L1 和 L2 norm 进行损失计算，降低丢失高频信息造成的模糊和失真。感知损失利用隐层评价输出结果与真实标记间的差异，结合对抗损失，从各个方面进行优化，自动持续的寻找还没有被优化的输出和真实图像之间的差异。实际上，即使计算感知损失，图像的模糊问题仍未完全解决。DRPAN[42]从优化生成过程对 PAN 的模糊问题进行改善。它采用先粗糙后精炼的学习计划，先用一个网络生成含有全局结构的图片，紧接着使用另一网络找出生成图片中最不真实的区域，并使用专门的修复器（判别器）对该区域进行局部图像修复，以提升图像的局部细节。此外，针对一些特定任务，采用特殊的机制或结构也能够进一步提高图像的生成质量。Chen 等[43]提出 SketchyGAN，能够实现草图图像向真实感图像间的转换。该算法提出一种适用于生成器又适用于判别器的新型网络结构块，通过注入多尺度的输入图像来改善信息流动，以生成真实感更强的目标图像。针对人脸属性编辑问题，Liu 等[44]提出 STGAN，将差异属性向量作为输入，增强属性的灵活转换简化训练过程，设计选择性传输单元并与编码器-解码器结合，同时提高属性操作能力和图像质量。Park 等[45]提出 SPADE（Spatially Adaptive Normalization）模块，通过使用语义图来调整 Normalization 输出的结果，使其具有更好的语义信息，可应对各种使用语义图的生成任务。

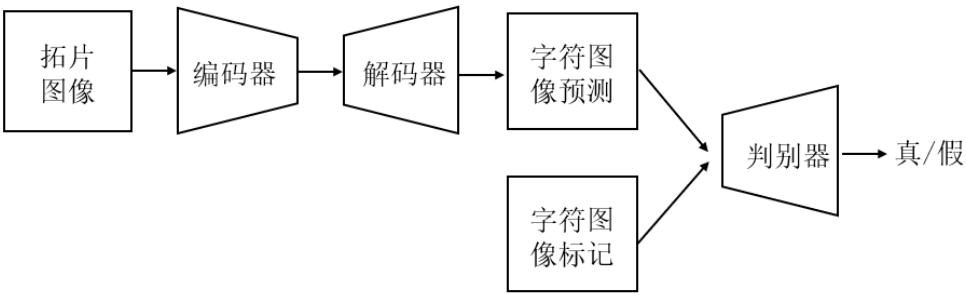


图 1.1 通用图像到图像转换模型结构

图像分割的方法（如 U-Net）将甲骨字符提取视为像素分类任务，通过对拓片图像进行逐像素分类，预测出字符在拓片图像中的所在区域，具体结构如图 1.2 所示。目前最具代表性的语义分割方法是全卷积网络（Full Convolutional Network，FCN[46]），该方法将 CNN 的全连接层改为卷积层，从而达到对图像像素进行类别预测的目的。但 FCN 通过池化操作进行下采样，导致部分空间信息丢失，并缺乏对图像空间上下文语义信息的利用。针对该问题，Chen 等人提出 DeepLab 系列。DeepLabv1[47]通过将 DCNNs 层相应和完全连接的条件随机场相结合，改善了特征图中高层特征的平移不变性，增强了对数据分层抽象的能力。DeepLabv2[48]引入了空洞空间卷积池化金子塔（Atrous Spatial Pyramid Pooling，ASPP），并行的采用多个采样率的空洞卷积进行探测，以多种比例捕捉对象及其上下文信息，最后，通过结合 DCNN 和概率图模型，继续改进分割边界结果。在 v1、v2 的基础上，DeepLabv3[49]继续对分割网络进行改进，提出的串行和并行的 ASPP 网络模块包含了不同比率的空间卷积处理与批归一化层，能够捕获更大范围的语义，整合多尺度信息。除此之外，一些语义分割方法专注设计特定的网络结构(编解码结构)以及融合经典的机制（注意力机制、生成对抗），从其它的角度提升分割的结果。PSPNet[50]、RefineNet[13]等引入空间金字塔模块和精炼模块分别融合低层和高层的上下文信息，提高网络对图像全局和局部特征的有效利用。PAN[51], DANet[52]等引入空间注意力机制，捕获重要的局部上下文依赖关系，进一步提取准确且密集的特征并进行像素分类。

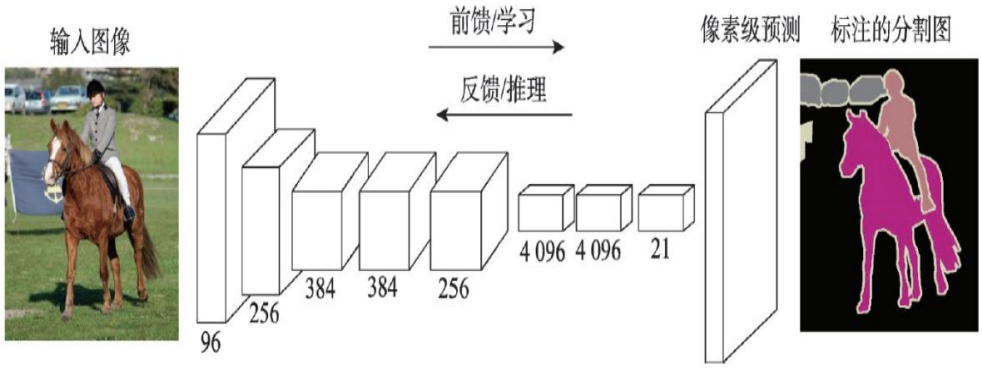
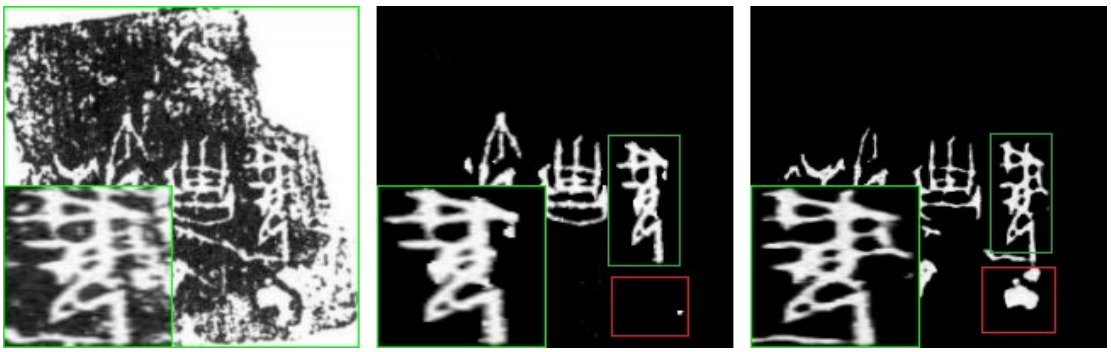


图 1.2 全卷积神经网络结构

然而，实验表明，直接将上述方法应用于甲骨文研究数据上往往存在一定的问题，具体如下：1）基于分割的方法具有较强的区分拓片图像背景和甲骨字符的能力，但得到的字符图像往往比较粗糙，存在字符笔画粘连、模糊等问题，如图 1.3（b）所示；2）基于生成的方法具有较强的结构信息描述能力，生成的甲骨字符在局部笔画细节上更为清晰，但往往会受背景噪声和裂痕的干扰，如图 1.3（c）所示。



（a）甲骨片　　（b）SegNet　　(c)Pix2Pix

图 1.3 基于分割和生成方法的甲骨字符提取结果

鉴于上述观察，本论文构建了一个全新的甲骨字符提取模型，该模型融合基于分割方法的背景噪声去除能力和基于生成方法的结构信息描述能力，其次，以生成网络为模型的基础骨架，将分割网络嵌入编码器网络以消除拓片背景噪

声的影响，以期建立更为准确的拓片图像与对应甲骨字符图像间的映射关系，最后，为获取内容完整且细节清晰的生成结果，该模型结合使用全局判别器和局部判别器对生成的甲骨字符图像进行一致性判别，能够生成更高质量的甲骨字符图像。

1.3 本文的主要工作及贡献

本文的总体目标是针对当前甲骨文检测与字符提取工作，结合深度学习技术和机器学习理论，对甲骨文检测和特征提取工作的自动化技术进行深入探索，具体的，本文的主要工作及贡献包括如下几点：

1. 本文首先构建了甲骨文字符临摹数据集，为深度学习技术应用于甲骨文字符提取工作提供数据基础。该数据集属于像素级别的甲骨文数据集，由人工使用 Photo Shop 工具逐张临摹而来，这些数据的素材主要来源于一些权威性的甲骨学著录资料，具体包括 435 对有代表性且退化严重的拓片图像记录和 3195 对使用多种扩充手段生成的与真实拓片图像记录相似的假数据记录。
2. 针对甲骨文字检测问题，论文提出一个简单且有效的甲骨字符检测器。该检测器采用非锚点框设计，使用自适应形状高斯核表示甲骨字符的空间区域，将甲骨字符的检测转换为对应高斯图的预测。其次，为避免部分字符之间排列紧密导致的区域重叠问题，字符区域分别以多个不同尺度的高斯核进行表示并采用一种渐进性尺度延伸策略沿最小尺度的高斯核预测朝向最大核进行延伸，以获取彼此相互分离的字符区域，最后，基于该相互分离字符区域，通过一系列简单的图像后处理操作获取精确的字符边界框。实验表明，在甲骨字符检测数据集上，该字符检测器实现了 83.2%的 F-Measure，大幅超越一些主流的目标检测器。
3. 作为甲骨文字检测的辅助工作，论文基于深度神经网络构建了一个端到端的甲骨字符提取模型。该模型充分利用基于分割方法的背景噪声去除能力和基于生成方法的结构信息描述能力，将甲骨字符提取任务视为图像到图像转换任务，以生成网络为模型的基础骨架，将分割网络嵌入编码器网络以消除拓片背景噪声的影响，以期建立更为准确的拓片图像与对应甲骨字符图像间的映射关系。最后，为获取内容完整 且细节清晰的生成结果，该模型结合使用全局判别器和局部判别器对生成的甲骨字符图像进行一致性判别。实验结果

表明，相比于一些经典的图像生成和分割方法，该字符提取模型能够生成更高质量的甲骨字符图像。

1.4 论文的主要内容与组织架构

第 1 部分为绪论。首先阐述了论文工作的研究背景、研究价值以及甲骨学研究核心任务—甲骨文字破译。然后，总体介绍甲骨学研究的基本状况，并重点介绍甲骨文字检测和字符提取工作，最后列出本文主要研究内容和内容结构安排。

第 2 部分为相关理论基础介绍。该章节介绍了论文工作涉及的深度学习技术基础，包括2D图像处理基础—卷积神经网络、图像生成基础—生成对抗网络。最后，目标检测作为本文工作的重点，继续对代表性的目标检测模型进行介绍。

第 3 部分是甲骨字符检测工作的详细介绍。包括当前现存的甲骨文字检测方法、论文提出的多尺度高斯核甲骨字符检测器以及对该字符检测器的实验验证。

第 4 部分主要介绍甲骨文字提取工作。该章节首先介绍了任务的应用背景，然后对字符提取工作简要分析，并引出论文构建的甲骨字符提取网络，最后通过实验结果验证该字符提取模型的有效性和优异性。

第 5 部分是论文工作的总结和未来工作的展望，包括甲骨文字检测、甲骨文字提取，指出当前这些工作的不足并讨论未来研究的努力方向。

2 相关技术基础

2.1 卷积神经网络

随着信息数据的不断累积及计算机硬件快速升级，以深度学习为基础的人工智能技术在学术界与工业界掀起了热潮，并在多个领域中都取得了巨大的成功，其中卷积神经网络（CNN）作为基础，在图像、音频、文本等诸多领域得到了广泛的应用。

卷积神经网络最早由纽约大学的 Yann LeCun 提出，并应用在手写字体识别数据集(MINST)上。其采用的局部连接和权值共享机制，一方面减少了的权值的数量使得网络易于优化，另一方面降低了过拟合的风险。但 CNN 网络主要擅长 2D 图像处理，图像可以直接作为网络的输入，避免了传统识别算法中复杂的特征提取和数据重建的过程。此外，CNN 网络还能够自行抽取图像的特征包括颜色、纹理、形状及图像的拓扑结构，在识别位移、缩放及其他形式扭曲不变性的应用上具有良好的鲁棒性和运算效率。结构上，CNN 网络是一种由多种层不断堆叠而成的网络结构，其主要的关键层有三个：卷积层、池化层，全连接层，如图 2.1 所示。然而在实际应用中，CNN 具有较强的灵活性和扩展性，针对不同的任务需求，可改变 CNN 的网络结构以实现相应的任务需求如图像生成、目标检测、图像识别。下面对这三个关键层具体介绍：

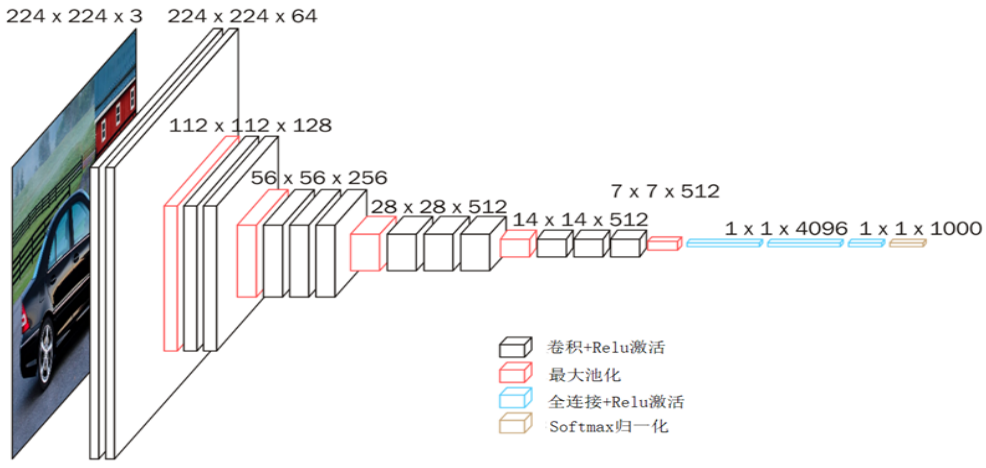


图 2.1 CNN 网络结构

（1）卷积层

卷积层是卷积神经网络的核心，大多数计算都是在卷积层中进行的，这些计算实质上是分析数学中的离散卷积。卷积运算使用卷积核矩阵遍历整个图像，对所有位置进行卷积操作，得到输出图像。假设卷积核矩阵为K是一个$s\times s$的矩阵卷积输入图像为X，在图像的(i,j)坐标出，卷积操作的输出值为：

$$\sum_{p=1}^{s}\sum_{q=1}^{s}k_{pq}\bullet x_{i+p-1,j+q-1} \tag{2.1}$$

即将卷积核矩阵和图像对应位置处的元素相乘，然后累加求和得到输出值。对整个图像从上到下、从左到右依次进行卷积，得到卷积之后的图像。

从上述介绍中可以发现，卷积层的每一个卷积核重复作用于整个感受野中，对输入图像进行卷积，卷积结果构成了输入图像的特征图，提取图像的局部特征，可以很大程度上减少模型参数。此外，卷积通过一定大小的卷积核作用在局部区域，可以很好的挖掘图像的局部特征，再加上使用不同卷积核的组合，便于模型学习更高层的语义信息，从而达到更好的效果。

(2) 池化层

池化层是一种下采样操作，可用于对输入特征向量进行降维。不同于卷积操作，池化层通常使用某一个区域用一个值代替，如最大值或平均值完成下采样操作。 除了降低图像尺寸之外，池化带来的另外一个好处是一定程度的平移、旋转不变性，因为输出值由图像的一片区域计算得到，对于小幅度的平移和旋转不敏感，但池化操作降低尺度大小的同时，也会噪声一定的细节损失，从而引发一系列的模型衰退问题。

池化主要包括均值池化和最大池化，但都是进行卷积操作之后得到的特征图进行分块。这些特征图像分别被划分成不相交的块，通过计算这些块内的最大值或平均值得到池化后的图像。均值池化和最大池化都可以完成降维，前者是线性函数，而后者是非线性函数，一般情况下最大池化有更好的效果。

(3) 全连接层

全连接层在整个卷积神经网络中起到“分类”作用，将学到的“分布式特征表示”映射到样本标记空间。一般在卷积神经网络末尾，通过全连接操作将卷积操作后的特征张量转化为若干神经元的输出，以便于损失计算，从而对权重参数

进行优化。

2.2 生成对抗网络

随着深度学习在人工智能领域不断的发展，越来越多新颖的网络模型不断的被提出。2014 年，由 Goodfellow 的生成对抗网络[37]，一致被认为 20 年来机器学习领域最让人激动的想法。生成对抗网络是一种新颖的使用机器学习的思路来解决数据生成问题的一种通用框架，其目标是生成服从某种概率分布的随机数据。凭借着强大的捕获高维数据分布的能力，迅速成为人工智能领域中的一个研究热点，目前 GAN 在多个领域得到广泛运用，如图像描述、超分辨率、文本数据生成等，随着生成对抗网络受到学术界和工业界的关注越来越多，GAN 将会在未来的研究中发挥着更加重要的作用。

生成对抗网络由一个生成模型和一个判别模型构成。生成模型用于学习真是样本数据的概率分布，并直接生成符合这种分布的数据；判别模型的任务是指导生成模型的训练，判断一个输入样本数据是真实样本还是由生成模型生成。训练过程，两个模型彼此之间相互竞争，从而分别提高它们的生成能力和判别能力。但 GAN 的原理远不止如此，下面本部分将会对 GAN 的基本原理，目标函数，以及交替优化学习进行详细的介绍。

2.2.1 基本原理

生成对抗网络可以看作是两个网络的博弈过程，生成网络通过机器不断生成新的数据，目的是愚弄判别器，希望生成的数据被认为是真实的，而判别网络通过不断学习提高判别能力，目标是尽可能的判别生成的数据是否是真实的。

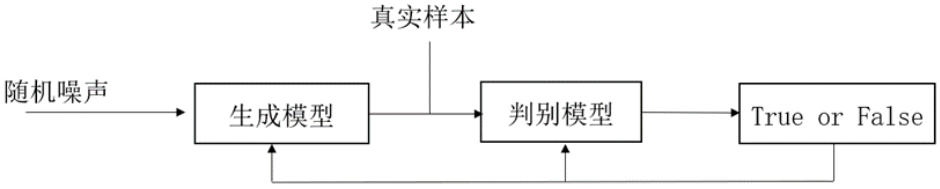


图 2.2 生成对抗网络模型框架

图 2.2 为生成网络模型基本框架，由生成模型和判别模型构成。具体的，生

成模型以随机噪声或类别之类的控制变量z作为输入，经过一个多层感知神经网络映射到一个新的数据分布，得到生成样本$G(z)$，然后这些生成样本和真实样本一起送入判别模型进行训练。判别模型是一个二分类器（一般也通过卷积神经网络实现），判定一个样本是真实的还是生成的。随着训练的进行，当生成模型产生的样本与真实样本几乎没有差别，判别模型也无法准确地判断出一个样本是真实的还是生成模型生成时系统达到平衡，训练结束（此时的分类错误率为0.5）。

2.2.2 目标函数

生成对抗网络训练的目标是让判别模型能够最大程度地正确区分真实样本和生成模型生成的样本；同时要让生成模型生成的样本尽可能地和真实样本相似。生成模型需要最小化如下目标函数：

$$\ln\left(1-D\left(G\left(\mathrm{z}\right)\right)\right) \tag{2.2}$$

这意味着如果生成模型生成的样本$G\left(\mathrm{z}\right)$和真实样本越接近，则被判别模型判断为真实样本的概率就越大，即$G\left(D(\mathrm{z})\right)$的值就越接近于 1，目标函数的值越小。对于判别模型，要让真实样本尽可能的被判定为真实的，即最大化$\ln D\left(\mathrm{x}\right)$,这意味着$D(\mathrm{x})$的值尽量接近于1；对于生成模型生成的样本尽可能的被判别为0。综合得知，生成对抗网络的总体优化目标可以定义为：

$$\min_G \max_D V\left(D,G\right)=E_{X\sim p_{data}(x)}\left[\ln D(x)\right]+E_{z\sim p_z(z)}\left[\ln(1-D(G(z)))\right] \tag{2.3}$$

2.2.3 交替优化

生成对抗网络训练时采用分阶段优化策略进行优化，生成模型和判别模型交替优化，直到达到平衡状态训练结束。完整的训练算法如下：

循环 1 $t<$ 最大迭代次数:

//第一阶段：训练判别模型

循环 2 $i<k$:

生成m个服从噪声分布$p_g(z)$的噪声数据$z_1,z_2,z_3,...,z_m$

从服从概率分布$p_{data}(x)$的样本数据中采样出m个样本 $x_1,x_2,x_3,...,x_m$

使用随机梯度上升法更新判别模型，判别模型参数的梯度计算公式为：

$$\nabla_{\theta_d}\frac{1}{m}\sum_{i=1}^{m}\left[\ln(D(x_i))+\ln\left(1-D(G(\mathrm{z}_i))\right)\right] \tag{2.4}$$

//第二阶段：训练生成模型

生成m个服从噪声分布$p_g(z)$的噪声数据$z_1,z_2,z_3,...,z_m$

使用随机梯度下降发更新生成模型，生成模型参数的梯度计算公式为

$$\nabla_{\theta_d}\frac{1}{m}\sum_{i=1}^{m}\left[\ln\left(1-D(G(\mathrm{z}_i))\right)\right] \tag{2.5}$$

其中，m是人工设定的参数，即 Mini-Batch 梯度下降法中的批量大小。外层循环里所做的工作分为两步，首先获取m个真实样本，用生成模型生成m个样本，用这 2m个样本训练判别模型。然后用生成模型生成m个样本，用这些样本训练生成模型。在第一步中，生成模型保持不变；在第二步中，判别模型保持不变。训练判别模型时采用的是梯度上升法，因为要求目标函数的极大值；训练生成模型时使用的是梯度下降法，因为要求目标函数的极小值。

2.3 常用目标检测模型

在深度学习介入计算机视觉领域之前，传统的目标检测方法主要采用滑动窗口或基于连通分量的检测路线，这类方法通常采用人工设计的特征及分类规则，鲁棒性较差，难以应对复杂的自然场景。另一方面，这类方法通常还需要大量的候选区域获取，时间复杂度高，效果差，尤其针对实时性检测问题，效果差强人意。

随着深度学习的广泛应用，目标检测迎来了迅速发展，涌出了一系列经典的检测算法，这些方法在检测性能、鲁棒性、实时性等方面均表现出优异的效果。本部分主要介绍当前经典的目标检测模型：Faster R-CNN[29]、SSD[53]、YoLo[31]等。

2.3.1 Faster R-CNN

2015 年，何凯明等人提出了更快的 Faster R-CNN，该算法在 Fast R-CNN、R-CNN 的积淀上，提出了 RPN 候选框生成算法，使得综合性能有较大提高，在检测速度方面尤为明显，甚至于在 2015 年的 ILSVRV 和 COCO 竞赛中获得多项

第一。

R-CNN 和 Fast R-CNN 的检测过程都包括候选区域生成、特征提取、候选区域分类、目标位置回归与调整 4 个步骤。Fast R-CNN 对后面的 3 步仅仅做了性能上的优化，重点解决候选区域问生成问题，这一步耗时严重且无法使用 GPU 并行优化，对检测的效率造成很大的阻碍。Faster R-CNN 将候选区域的提取用卷积网络(Region Proposal Network，RPN)实现，代替 Selective Search 算法。通过候选区域生成网络实现，候选区域生成网络与检测网络共享卷积层参数，因此，计算开销非常小。整体系统结构如图 2.3 所示。

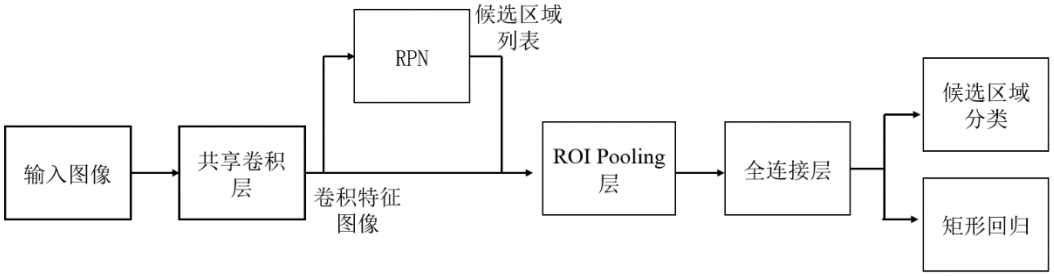


图 2.3 Faster R-CNN 模型整体系统结构

RPN 的输入为卷积特征图，输出为一系列的候选矩形框。RPN 网络对最后一个共享卷积层的特征图像进行处理，首先执行 3x3 卷积，将每个卷积核滑动位置都映射一个 256 维的特征向量，然后分别送入两个并列的 1x1 卷积层进行处理，一个分支用于目标矩形分类，确定这个区域是背景还是前景，另一个分支是对目标矩形回归，确定候选框的位置和大小。

为了检测不同大小和宽高比的目标，使用了一种成为锚点（Anchor）的机制。卷积特征图像中的每个位置对应原始图像中的多个矩形框，这些矩形以该点为中心，有不同的大小和宽高比，不同矩形大小是为了检测不同大小的目标，不同的宽高比是为了检测不同形状的目标。例如：如果每个位置的锚点有 k 种不同的大小，每种大小有 n 种不同的宽高比，因此，在每个位置产生出 nxk 个候选框。

RPN 使用反向传播算法和随机梯度下降法训练，为了实现 RPN 前面的卷积层和检测网络的卷积层参数共享，采用了交替训练的策略，具体做法如下：

（1）用 ImageNet 模型初始化，训练 RPN 的参数。

（2）用 ImageNet 模型初始化，使用上一步 RPN 产生的候选区域作为输入，训

练 Fast R-CNN 网络，此时两个网络每一层的参数不共享。

（3）使用第（2）步的 Fast R-CNN 网络参数初始化一个新的 RPN，但设置 RPN、Fast R-CNN 共享的那些卷积层的学习率为 0，仅更新 RPN 特有的那些层重新训练，此时两个网络已经共享了所有公共的卷积层。

（4）固定共享的那些网络层，把 Fast R-CNN 特有的层也加入进来，形成一个统一的网络，继续训练，微调 Fast R-CNN 特有的层。

Faster R-CNN 作为成熟的两阶段检测框架，仍存在非常大的不足，有待于进一步的改进，例如 2016 年 Dai 等[54]提出 R-FCN 解决了 Faster R-CNN 检测速度慢的问题。2017 年 Lin 等[55]提出了特征金字塔网络 FPN，重点解决 Faster R-CNN 多尺度目标检测问题。2018 年 Bharat 等[56]提出一种新的图像金字塔尺度归一化，它借鉴多尺度训练思想，提出新的多尺度训练方案，解决了小目标检测问题。2017 年 He 等[57]提出 Mask R-CNN 算法对 Faster R-CNN 的功能进一步延伸，增加了额外的 Mask 分支，使得位置检测与实例分割同步进行。

随着对 R-CNN 系列方法的不断改进，在检测精度上毋庸置疑，但由于两阶段方法自身结构特点，其在实时性方面仍存在难以克服的缺陷。同时，其他的研究学者也试图去探索其它的检测思路，克服两阶段模型准确率高但速度慢的弊端。

2.3.2 YoLo

YoLo 是另一种类型基于回归的检测算法，该类算法移除了两阶段方法（如 R-CNN，Faster R-CNN 等）中提取候选区域的步骤，将目标检测视为一个回归问题，输入图像经过一次推理，直接得到物体的位置信息和所属类别，在维持准确率（63.4% mAP,VOC2007）的前提下，达到 45FPS 的检测速度。

YoLo 整体框架只使用一个网络，同时完成目标位置的预测和分类，实现了真正意义上端到端的训练和预测，网络结果如图 2.4 所示。特别的，YoLo 将输入图像划分为 $S\times S$ 的网格，每个格子负责落入该格子的物体（如某个物体的中心位置坐标落入到某个格子，那么这个格子就负责检测该物体），输出 B 个边界框信息以及属于某种类别的置信度，该置信度包含了两方面的信息：当前矩形是目标的概率，以及该矩形是目标矩形的精确度。置信度定义为：

$$p(\text{object})\times IOU_{\text{pred}}^{\text{truth}} \tag{2.6}$$

其中，上式前半部分为矩形是一个目标的概率，后半部分是矩形的精确度预测值，即和预测出的矩形框与真实目标矩形框的重合度。

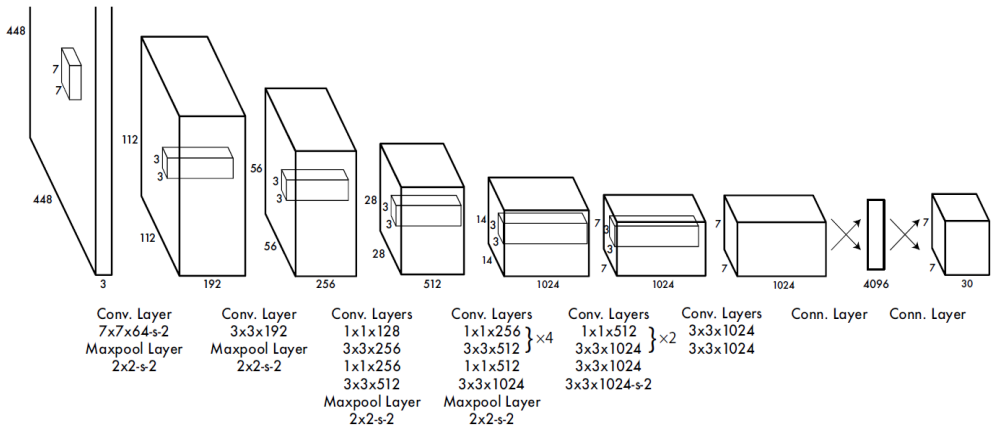


图 2.4 YoLo 网络结构

由于对整张图像进行处理，相较于基于滑动窗口、候选框的方法，YoLo 可以利用更大范围的图像语义信息，从而提高检测精度。然由于网络设计的粗糙以及使用格子作为中心点，导致算法具有识别物体位置精确性差、召回率低等问题。后续，YoLov2[58]算法对 YoLo 进行了进一步改进，它借鉴了 Faster R-CNN 中的锚点机制，使用锚点框取代边界框的预测并通过引入 BN（Batch Normalization）、训练高分辨率分类器、使用细粒化特征等手段，在保持原有速度的同时精度得到大幅的提升。YoLov3[59]在 YoLov2 的基础上继续改进，它借鉴了残差网络中的跳跃连接，形成更深的网络层次以及利用多尺度特征，提升了对小目标物体的检测效果。

2. 3. 3 SSD

在基于”区域+分类”的目标检测方法中，R-CNN 系列（如 R-CNN、SPPnet[27]、Fast R-CNN 以及 Faster R-CNN 等）取得了非常好的效果，但是在检测方面离实时效果还比较远。在提高 mAP(Mean Average Precision)的同时兼顾速度，逐渐成为神经网络目标检测领域未来的趋势。YOLO 检测算法虽然能够达到实时的效果，而且 mAP 与相比于 R-CNN 系列有很大的提升,但也存在一些缺陷：（1）每个网格只能预测一个物体，容易造成漏检；（2）对于物体的尺度相对比较敏感；（3）面对尺度变化较大的物体时泛化能力较差。SSD 网络在这两方面都有所改

进，同时兼顾了 mAP 和实时性的要求。

SSD 结合 YoLo 中的回归思想和 Faster R-CNN 中的锚点机制，使用全图各个位置的多尺度区域进行回归，既保持了 YoLo 的速度优势，也兼顾了优于 Faster R-CNN 的检测准确性，SSD 的整体结构如图 2.5 所示。整体卷积网络以 VGG[60]网络为基础，并在卷积层后面添加一些额外的卷积层，用于目标检测多尺度特征提取。SSD 继承了 YoLo 将检测转成回归的思路，使用卷积一次完成定位目标与分类，借鉴 Faster R-CNN 中锚点框的理念在得到的特征图中为每个单元设置不同宽高比的先验框，减少了训练难度，最后通过 Softmax 分类和边界回归的方式获得真实目标边界框。

尽管 SSD 算法将不同层次的特征图进行融合，提升了检测速度和精度，但对小目标仍具有鲁棒性。针对该问题，DSSD[46]使用 ResNet101[61]作为特征提取网络，提取更深层次的特征；提出基于自上而下的网络设计，采用反卷积代替传统的双线性插值上采样；最后，在预测阶段引入残差单元，优化候选框回归和分类。RefineNet[13]提出了一种使用 Focal Loss[62]的全新结构，降低了大量简单负样本在训练中所占的权重，从而大幅提升单阶段检测算法的精度

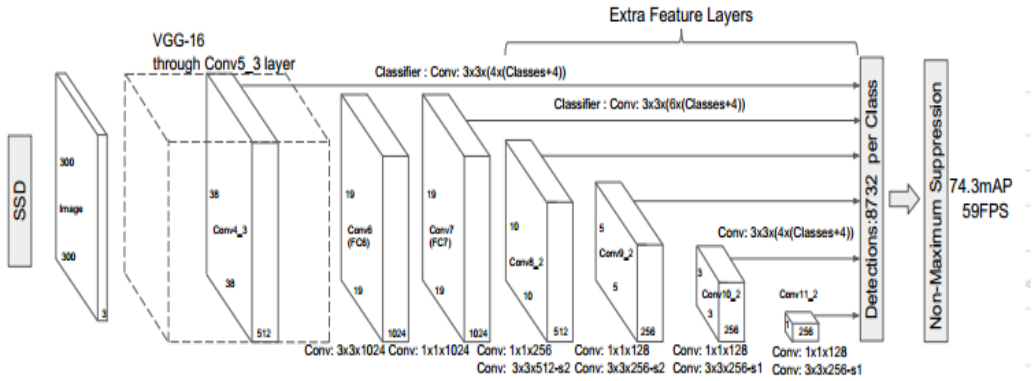


图 2.5 SSD 网络结构图示

2.4 本章小结

本章节主要概述了论文设计的相关技术，首先介绍了 2D 图像处理基础—卷积神经网络，然后介绍了当前图像生成领域常的数据生成模型—生成对抗网络，最后，对一些常用的目标检测模型简要概述，为下面甲骨文字检测算法的提出奠定理论基础。

3 甲骨文字检测

3.1 引言

甲骨文字检测是甲骨学领域基础性研究任务之一。作为字符破译的前提工作，目标是从甲骨片上定位甲骨字符的位置并确定相应的字符类别。早些时期，该工作需要甲骨学专家结合专业知识以及经验积累得以完成，不仅占用了昂贵的专家资源，而且效率低下，因此，探索甲骨文字自动化检测技术很有价值，对加快甲骨学的研究与推广具有重要的意义。

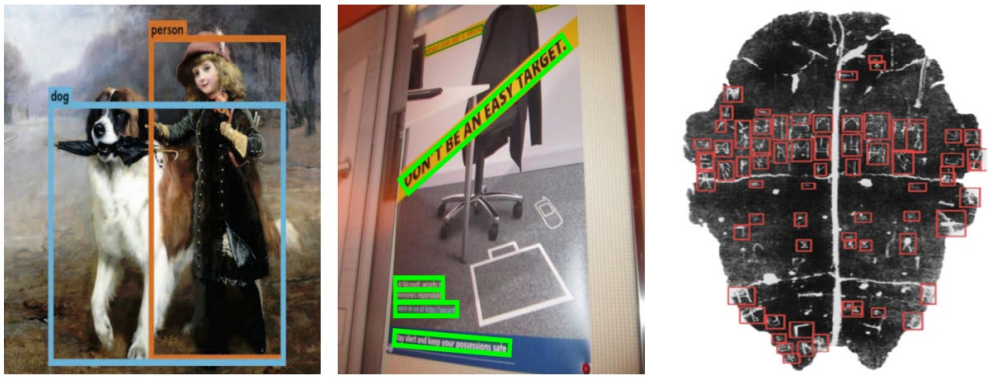


图 3.1 由左至右依次是通用自然场景图像，自然场景文本图像，甲骨拓片图像

由于自身的特殊性，甲骨学一直停留在少数甲骨学专家的学术研究中。目前仅仅存在少数的方法用于解决和甲骨拓片图像相关的计算机视觉问题。而针对甲骨文字检测问题，目前几乎是一片空白。其次，甲骨拓片作为甲骨学研究的第一手资料，其数据本身的特殊性也为甲骨文字检测工作带来不小的挑战。不同于一般的自然场景或文本图像，甲骨拓片图像也面临一些特有的难题如图 3.1 所示。首先拓片上的甲骨文字分布任意，大小不易，尤其是一些被遮挡或尺度极小的目标，检测难度较大；其次，甲骨拓片图像表面退化严重，上面的文字模糊不清，并充斥大量的背景噪声；此外，由于长期掩埋以及私掘等缘由，出土的甲骨片存在破裂形成裂痕，这些裂痕在纹理上与甲骨字符非常相似难以区分；最后，拓片上异体甲骨字出现频率严重失衡，并且这些字符风格不易，差异极大；尽管如此，总体上，甲骨文字可以视为一种特殊的多尺度小目标检测问题，但这个问

题需要克服复杂背景环境带来的文本定位困难以及字体外观多样性等挑战。目标检测或场景文本检测是与本论文研究的甲骨字符检测较为相近的视觉任务。理论上，这些领域中的方法能够很大程度上为甲骨字符的检测问题带来更多的受益和启发。

近年来，受益深度学习在各个视觉领域的流行，当前存在少数的工作用于字符检测任务的研究和技术探索。但这些方法大多是将一些经典的目标检测方法经过略微修改后直接应用在甲骨文检测数据集，因此，在检测效率、准确度上仍存在一定的局限性。由于甲骨文检测数据集缺少字符水平的类别标记，仅仅回归字符的位置坐标难以捕获充足的字符语义特征，导致一些特殊的字符误检，例如，由多个子部件组成的复合字容易被误认为多个字符，如图 3.2(左)所示。相似的，多个密集分布的字符也有可能误认为同一复合字符，如图 3.2(右)所示。另一方面，当前多数用于检测甲骨文的方法基于锚点框的设计，这涉及复杂的网络设计和大量锚点框的需要。例如 DSSD[63]中锚点框的个数高达 40k，RetinaNet[64]中的锚点框数量高达 100k，这在一定程度上降低了模型的检测效率。

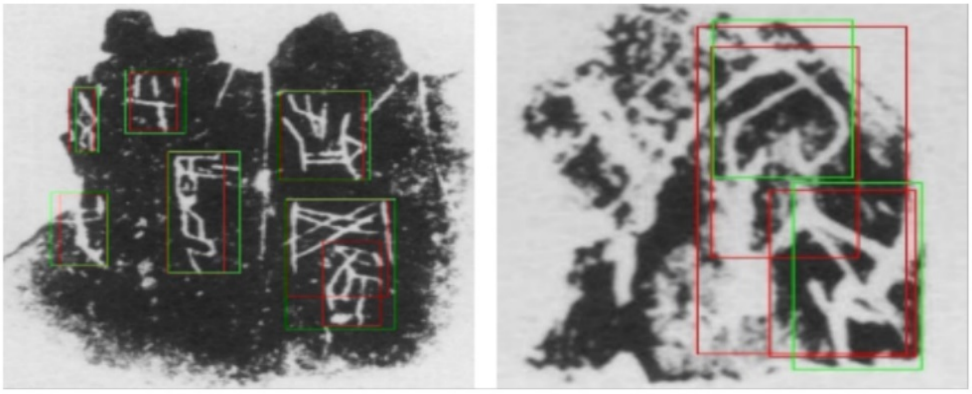


图 3.2 Faster R-CNN 误检样例展示，其中红色框和绿色框分别表示预测的字符边界框和真实的位置标记

本章节注重甲骨文字检测效率和准确率的提升，提供了一种更为简单且有效的甲骨字符检测器，该检测器采用非锚点框的计划，使用多尺度高斯核表示甲骨文字区域，将字符检测任务转换为对应字符高斯图的预测。实验表明，该甲骨字符检测器实现了 83.2%的 F-Measure，大幅超越一些主流的目标检测器。

3.2 高斯核甲骨字符检测器

3.2.1 甲骨字符检测器简介

本章工作主要受到近来提出的 CRAF[65]的启发。该工作使用自适应形状的高斯核表示字符的空间区域，将文本实例的检测任务转换为对应字符高斯图的预测，因此，不仅避免了锚点框的需要，同时使得检测模型学习字符的空间区域信息成为可能。我们遵循 CRAFT 的字符空间区域表示方式，使用自适应形状的高斯核表示甲骨字符，直接输出关于甲骨字符区域预测，如图 3.3 所示。然而，实验发现，该高斯核表示仅仅在处理稀疏分布的字符区域时，展现优异的效果，而对于一些紧密分布的字符，容易出现区域重叠问题，如图 3.4 所示。为解决区域重叠导致的误检问题，我们引入了多尺度高斯表示策略，表示单个甲骨字符同时以多个不同的尺度，其中尺度越小，字符间的间距就越大。然后基于这些不同尺度的高斯核预测，采用一种渐进性行尺度延伸策略，获取精确的字符边界框。实验表明，该字符检测器在甲骨文检测数据集上是有效的并取得比一些主流的目标检测模型更优的检测效果。

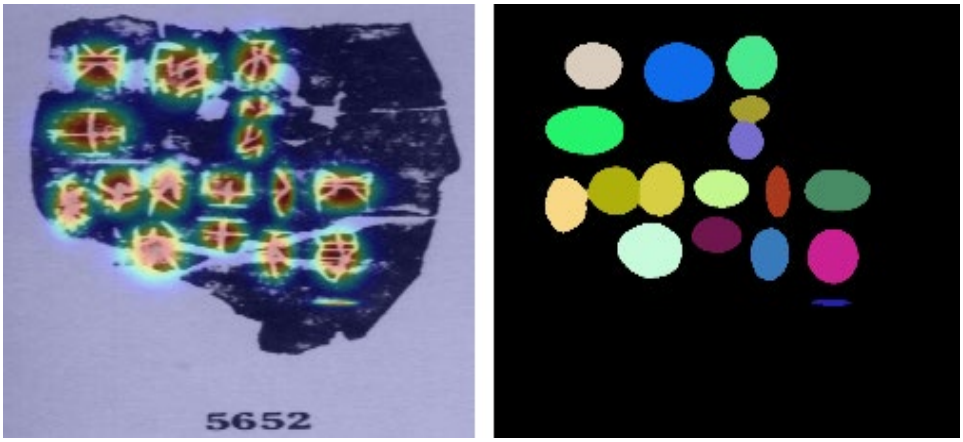


图 3.3 基于高斯核表示的字符检测可视化. 左：高斯核检测器字符区域热图输出；右：字符热图分割结果

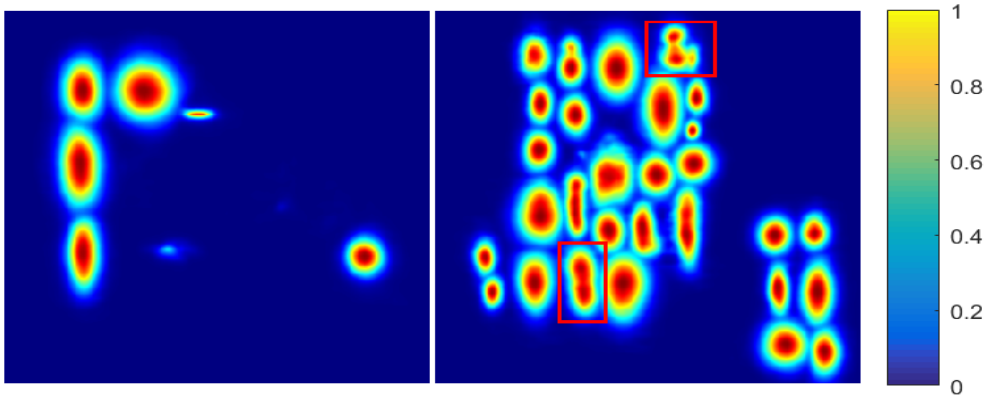


图 3.4 单尺度高斯核条件下，字符检测模型的高斯热图输出

针对甲骨文字检测任务，该检测器将每一个甲骨字符视为一个特殊的关键点，使用高斯核函数对字符空间位置进行编码，并通过训练一个编码译码网络，预测出甲骨字符在拓片图像中对应的特征编码。整体的数据流程如图 3.5 所示，首先，甲骨拓片图像 I_O 输入到一个卷积神经网络经前向传播后生成一个融合多层上下文语义信息的中间特征图 $I_F \in R^{H\times W\times C}$，然后 I_F 通过区域预测模块映射到 n 个分支，产生 n 个不同尺度的区域得分图 $S_1, S_2, ..., S_n$，其中每一个 S_i 分别代表一种尺度大小的字符区域得分图，S_1 表示最小尺度的字符区域预测，而 S_n 表示最大尺度的字符区域预测，基于这些不同尺度的字符区域预测，采用一种渐近性尺度延伸策略从极小核朝最大核方向逐渐延伸，以获取完整且相互分离的字符区域。最后，通过一系列简单的后处理操作获取最终的字符边界框。

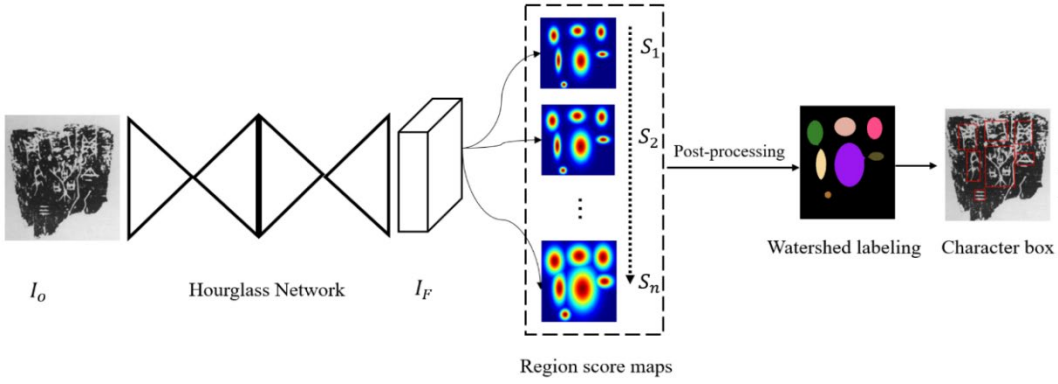


图 3.5 基于多尺度高斯核表示的甲骨字符检测器网络结构

3.2.2 Hourglass Network

为了捕获丰富的字符细节特征，本甲骨字符检测器使用沙漏网络(Hourglass Network)[66]作为基本骨架。顾名思义，沙漏网络因为在结构上像一个沙漏而得名，由于其便于处理多尺度图像信息，尤其在捕获局部信息的巨大优势，作为通用框架广泛应用于人体姿态估计任务中。沙漏网络是一种级联结构的全卷积神经网络，由一个或多个沙漏模块组成。沙漏模块是在结构上类似 FCN[67], 先通过一系列卷积和最大池化层对输入特征进行下采样，然后再通过一系列上采样和卷积层将其恢复到原始分辨率. 不同的是，沙漏模块没有使用 uppooling 操作或反卷积层，而是使用了最简单的最近邻上采样和跳跃连接完成上采样操作，此外多个沙漏块可以若干个堆叠起来，能够满足处理各个尺度信息的需要。

3.2.3 目标函数

在训练过程中，甲骨字符检测器损失计算包含两个部分：单个完整尺度高斯区域损失 $L_{FullMap}$ 和多个不同收缩尺度后的高斯区域损失 $L_{ZoomMap}$，其数学表达如下：

$$L = \lambda L_{FullMap} + (1-\lambda) L_{ZoomMap} \tag{3.1}$$

λ 用来平衡 $L_{FullMap}$ 和 $L_{ZoomMap}$ 之间的权重。

$$L_{FullMap} = L_{Pix}(S(p), S^*(p)) \tag{3.2}$$

$$L_{Pix}(T(p), T^*(p)) = \sum_p \| T(p) - T^*(p) \|_2^2 \tag{3.3}$$

p 表示单个像素点的位置坐标，$S(p)$ 表示字符区域预测得分，$S^*(p)$ 表示真实的字符区域得分。

3.2.4 训练标记生成

高斯核又称高斯核函数，一种基于高斯距离的编码方式，由于其天然的数据优势如旋转不变性、单值函数、可分离性等，在诸多图像处理任务中广泛被采用。在人体姿态估计任务中普遍被用来编码人体关键的空间区域，并在建立关键点联系、指导网络学习、转变输出形式等方面展现出较强的灵活性，其具体数学表达如下：

$$f(x,y)=A\exp(-(\frac{(x-x_0)^2}{2\sigma_x^2}+\frac{(y-y_0)^2}{2\sigma_y^2})) \tag{3.4}$$

其中 A 为幅度值，x_0,y_0 为中心点坐标，σ_x,σ_y 为方差。

本部分通过一系列的图像 Shrinking 操作生成多尺度字符高斯核，根据甲骨文检测数据集提供字符水平边界框标记，先生成完整尺度的字符热力图，然后按照收缩的比例沿着字符边界框内部逐渐收缩得到 n 个不同尺度的字符热力图。具体的，遵循文献[65]中的流程，主要步骤如下：1)根据甲骨文检测数据集提供的字符水平边界框标记，按照文献[68]中的参数设置，定义 n 个收缩间距 $D=\{d_1,d_2,...,d_n\}$；2)预先准备一个二维的、各向同性的 2D 高斯图；(3)计算高斯图与每一个字符框的透视变换矩阵；(4)将高斯图扭曲到对应的盒子区域；该操作的可视化如图 3.6 所示。

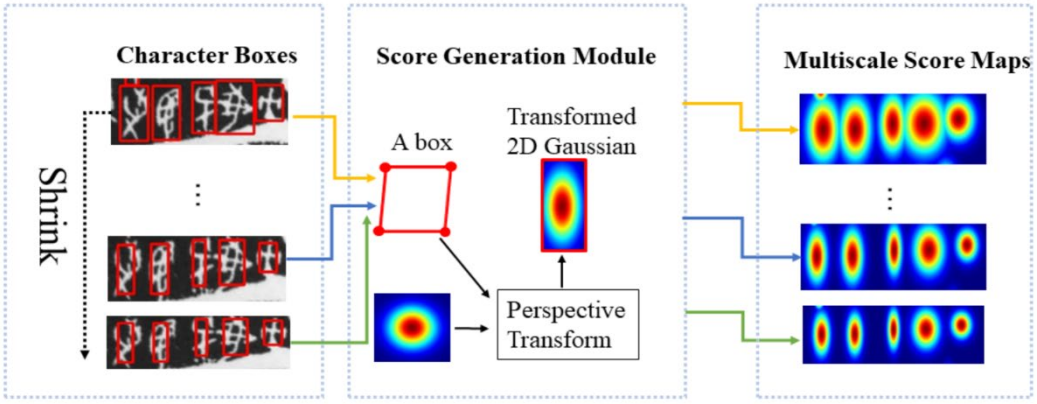


图 3.6 多尺度字符高斯标记生成流程

3.2.5 字符框后处理

获取不同尺度的字符区域得分估计后，通过一系列简单图像后处理操作生成字符水平边界框。该后处理管道中的关键源自于文献[68]中的尺度延伸算法。其本质上是深度优先算法的一种应用扩展，基于层次遍历的规则，渐进性的检测密集的物体如密集的场景文本。在本字符检测器中，主要利用不同尺度 mask map 间的相邻关系，按照深度优先的遍历方式，将最小核的文本区域朝着完整形状的最大核区域逐渐扩展。具体步骤如下：首先对原始的多尺度高斯图预测执行简单的预处理，然后通过一些形态学运算来降低高斯图中的噪声。其次，对于尺尺度延伸算法获得相互分离字符区域 K, 计算了它们的连接分量 C 并分配了不同的标

签 Label。最后，基于这些分配的标签，计算每个连接组件的最小外接矩形以获得最终的准确边界框。Opencv 提供的诸如 *connectedComponents*，*morphologyEx* 和 *minAreaRect* 之类的功能可以用于此目的。边界框获取详情如表 3.1 所示。

表 3.1 高斯字符检测模型后处理

算法 1 高斯字符检测模型后处理
输入：高斯核预测结果 $Z=\{Z_1,Z_2,...,Z_n\}$ 输出：字符边界框集合 L 方法名 Prediction(Z) 1. 初始化默认值为 0 的矩阵集合 $M=\{M_1,M_2,...,M_n\}$ 2. 循环 $i=1$ to n do 3. 如果 $Z_i(p)>\delta$ 然后 $M_i(p)=True$ // δ 是默认值为 0.35 的阈值 4. $M \leftarrow morpho\log yEx(M)$ //形态学开操作 5. $K \leftarrow scaleExpanded(M)$ // 尺度延伸操作 6. $C, Lable \leftarrow connectedComponents(K)$ // 计算连通分量 7. $L \leftarrow \min Area\,\mathrm{Re}ctByLabel(C, Label)$ //求最小外界矩形 8. 返回 L

3.3 实验验证

3.3.1 甲骨文检测数据集

本章节中的所有实验均基于安阳师范学院甲骨文信息处理实验室提供的甲骨文检测数据集。该数据集专注于甲骨文检测的任务，主要包括两部分：使用高分辨率扫描仪从甲骨文文献收藏中收集的甲骨拓片图像和手工制作的字符水平的边界框。不同于一般自然场景图像，甲骨拓片图像主要具有以下特征：

高噪声：甲骨片，甲骨文的主要载体，长期掩埋在安阳的废墟中，直到 120 年前才被发现。因此，拓片表面不可避免地存在一定的退化，其中最明显的是存在大量噪声。这些噪声具有不同的规则，并且密集地分布在整张图像上，给甲骨文检测任务带来了极大的挑战。

裂痕：由于埋葬环境和私人发掘等缘故，许多出土的甲骨拓片出现破裂，形成各种各样的裂痕。这些裂痕在纹理上和字符特征非常相似，容易被误认为甲骨字符

特征。

分布：同一甲骨拓片图像上的字符具有不同的尺度、方向且分布不一。此外，在56,743 个甲骨片中，包含 1,425 个字符。其中，共有 366 个常见字符，不常用字 500 个，罕见字 559 个。 本甲骨文检测数据包含 9500 对甲骨拓片图像记录，其中训练集、验证集和测试集分别包含 8287、436 和 411 对数据记录。

3.3.2 评估指标

本章节主要从效率和准确性的角度 评估字符检测模型的整体性能。网络权重参数，浮点计算和推理速度的三个指标用于评估模型的整体检测效率。主流对象检测方法中常用的测量指标：精确度（P），召回率（R）和 F-Measure （F）用于测量模型的检测精度。这些指标的计算规则如下：

$$P = \frac{TP}{TP + FP} \tag{3.4}$$

$$R = \frac{TP}{TP + FN} \tag{3.5}$$

$$F = \frac{2 * P * R}{P + R} \tag{3.6}$$

其中，TP, FP, FN 分别表示真正，假正，假负的样例的个数。

3.3.3 实验环境

本章节所有程序代码均基于 Pytorch 深度学习框架实现，使用四块 Nvidia TITAN X GPU 显卡进行模型训练，具体参数配置如表 3.2 所示。由于甲骨文字检测数据集不包含字符的类别信息，本实验将所有的甲骨字符视为单一的目标，赋予一个相同的类别标记。在网络训练过程中，原始的甲骨拓片图像被放缩到 512x512 分辨率，使用 Adam 优化器对参数进行更新优化，根据经验优化器的学习率和衰减系数分别设置为 0.0001 和 0.0005。

表 3.2 深度学习机配置

操作系统	Ubuntu 16.04
GPU 驱动	NVIDIA UNIX 440.100
CUDA	10.0
CUDNN	7.4
CPU	Inter(R)Xeon(R)E5-2650 v3@2.30GHz
RAM	64GB
$GPU_{1\text{-}4}$	Nvidia Titian Pascal Xp @12GB

3.3.4 消融研究

3.3.4.1 高斯核有效性验证

除了高斯核可以用来表示字符区域之外，二进制掩码也是另一种选择。为了比较两种表示之间的差异，本实验简单的将字符检测模型（仅使用单一尺度的高斯核）与语义分割模型进行比较。模型的输出结果的可视化如图 3.6 所示。显然，二进制掩码使用离散值无区别地表示字符区域，并且获得的预测结果具有更多的区域重叠。相反，高斯核基于与中心像素的距离关系对字符区域进行编码，获得的字符区域在边界上更加清晰。

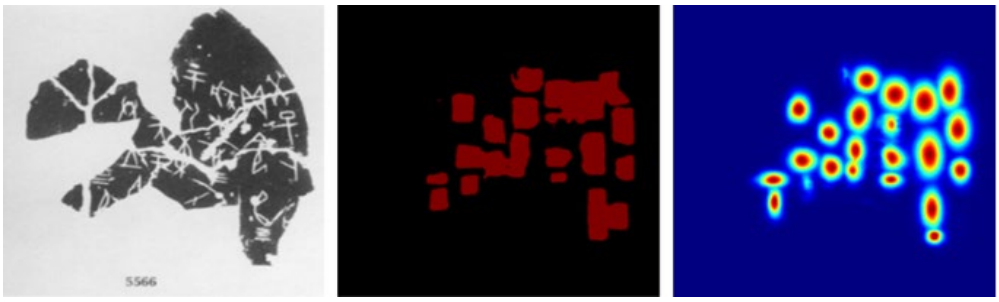


图 3.6 二进制与高斯核表示比较；从左到右依次为甲骨片输入，DeepLabv3 二进制掩膜输出，高斯核检测器输出

表 3.3 二进制掩膜与高斯核表示量化结果

模型	精确度(P)	召回率(R)	F-Measure(F)
DeepLabv3	0.626	0.638	0.632
Gaussian(ours)	0.776	0.646	0.705

表 3.3 的量化结果显示，在精确度(P)，召回率(R)和 F-Measure(F)指标上，

基于高斯核表示的方法明显高于二进制掩码表示，这再一次表明高斯核在表示字符空间区域上更具有优势，尤其是紧密排列的甲骨字符区域。

3.3.4.2 多尺度高斯核必要性

为了回答这个问题，本部分有选择的改变多尺度高斯核的个数，对该字符检测模型重新训练并评估检测效果。随着尺度个数的变化，字符检测效果变化如图3.7所示。从F值变化上看，仅仅当n小于6时，F值随着多尺度高斯核个数的增加而增加，显然并非尺度个数越多，字符检测模型的检测效果就越好。然而，尽管当n大于6时，F值有所下降，但相对于使用单一尺度高斯核，多尺度高斯核检测效果明显更好，这在一定程度上表明同时使用多尺度高斯核表示不同大小的字符区域对于检测效果是有效且有必要的。

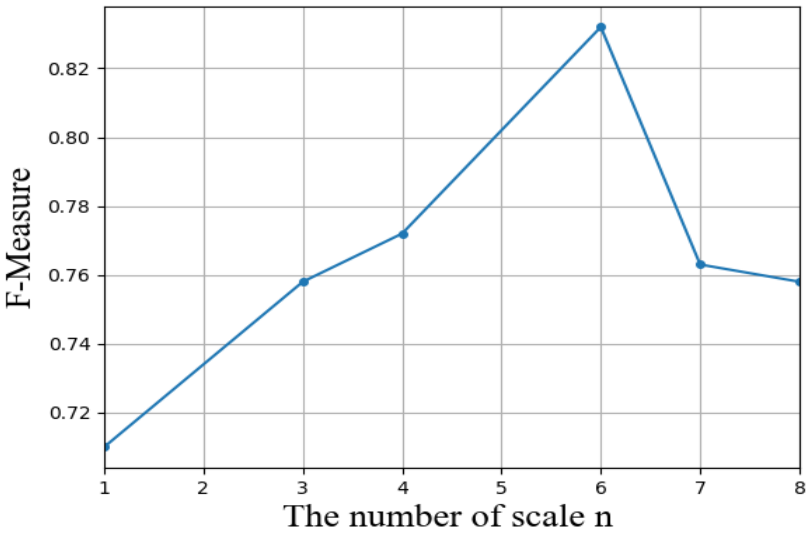


图3.7 高斯核尺度个数的消融

3.3.5 对比评估

3.3.5.1 精确度比较

为评估本章节高斯字符检测的检测效果，将其与几个一流的通用目标检测模型进行比较，这些模型既包含了具有精确度优势的两阶段模型如 Faster R-CNN[29]，也有具有速度优势的单阶段检测模型如 YoLov3[59].

表3.4展示与一流检测模型的精确度量化结果，就准确率而言，本文的字符

检测模型取得最高 89.7%的分数并超于次优模型 YoLov3 12%的差距。然而，从召回率上看，本文字符检测模型的表现相对较差，几乎处于垫底。针对该问题，本文认为可能的原因在于非最大抑制(NMS)的非严格使用。通常在基于锚点框方法的检测路线中，NMS 操作使用人工预先定义的阈值去过滤掉一些无效或冗余的候选框，可能存在遗漏，进而导致召回率的值偏高。为更好的评估字符模型的检测效果， 本文继续比较精确率和召回率的均衡指标 F-Measure，同样本文的字符检测依据取得最高的 F-Measure 得分，超越次优模型 5%。综述所述，多尺度高斯核甲骨字符检测器在检测精度上具有很好的优势。此外也不难想象，使用高斯核对甲骨字符的空间位置进行编码能够逼迫检测网络学习去捕获更多关于字符的区域信息，使得模型具有字符区域意识，因此能够得到更准确的检测结果。

表 3.4 与一流检测模型的精确度量化结果

模型	精确度(P)	召回率(R)	F-Measure(F)
Faster R-CNN	0.754	0.778	0.766
SSD	0.748	0.758	0.753
RefineDet[69]	0.752	0.805	0.778
RBFNet[70]	0.761	0.789	0.775
YoLov3	0.776	0.784	0.780
Ours	0.897	0.775	0.832

3.3.5.2 效率比较

本小节继续评估字符检测模型的检测效率，统计其推理速度、权重参数、浮点运算量并与一流目标检测模型进行比较。

表 3.5 展示了与一流检测模型的效率比较结果。在推理速度上，本文的模型取得了最快的推理速度 23FPS，并高于检测速度占有的 YoLov3 5FPS。 在权重参数量上，本文模型需要更少的参数，仅仅占用 12.73M 的内存空间，远低于 SSD 的 26.29M. 在浮点运算量上，本文模型仅仅比 YoLov3 稍弱一些，取得了第二名的成绩并远比其它检测模型更优。综合得知，本文模型能够在兼顾网络复杂度的同时取得更快的推理速度。

表 3.5 与一流检测模型的效率比较结果

模型	速度(FPS)	参数量(M)	浮点量(GMac)
Faster R-CNN	3	41.37	129.27
SSD	9	26.29	90.4
RefineDet[69]	14	34.44	97.94
RBFNet	15	36.64	103.65
YoLov3	17	61.92	50.06
Ours	23	12.73	57.34

3.6 本章小结

本章节主要介绍甲骨文字检测方面的研究工作，首先指出了当前现有的方法存在的缺点，并针对这些缺点，提出了一种基于多尺度高斯核的甲骨字符检测器，提升检测效率的同时，改善检测精度。最后，为证明论文方法的有效性和优异性，对方法进行消融研究和对比评估，实验结果表明，该甲骨字符检测器实现了 83.2%的 F-Measure，大幅超越一些主流的目标检测器。

4 甲骨文字提取

4.1 引言

甲骨片，作为甲骨文字的重要载体，由于某些历史原因，长久的掩埋在安阳的废墟中，直到 120 年前才被发现。因此在甲骨拓片表明，不可避免的存在一定的退化如噪声、裂痕等。这些不同程度的退化严重干扰了甲骨文字的可视性，为甲骨文字检测与识别等工作带来极大的阻碍。考虑到甲骨字符是甲骨学研究的第一手资料，从甲骨拓片图像中自动提取甲骨字符将有助于甲骨学研究的开展，并对甲骨文活化与利用产生重大帮助。

甲骨文字提取是一种特殊的关键信息增强技术，该任务通过移除复杂的甲骨背景，仅保留相应的甲骨文信息，间接增强甲骨文信息的可视性。显然，退化严重的甲骨拓片图像经过背景移除后，然后再进行文字检测与识别工作，将有助于提升检测与识别的效率。此外，这些得到的仅包含甲骨文的图像还有助于改善其它甲骨文工作效率低下的困境，如甲骨文临摹工作，通过计算机技术自动化提取甲骨片上的文字，可避免耗时耗力的人工临摹工作。另外，这些干净的字符图像还可以作为甲骨文创作品的基础素材，通过一系列加工处理，将将枯燥、乏味的甲骨字符转换普通大众易于理解的形式，从而有助于甲骨文的推广工作。

同甲骨字符检测任务一样，针对甲骨文字提取问题，当前几乎是一片空白。同样甲骨拓片数据自身的特殊性是该字符提取工作的重要阻碍。不同于一般的自然场景图像和场景文本图像，甲骨拓片图像表面存在着严重的退化、污染问题如图 4.1 所示，精确提取拓片图像中的甲骨字符是一项极具挑战性的图像处理问题，其主要表现在：（1）甲骨拓片表面包含大量不规则的噪声，这些噪声密集的分布在拓片图像表面，不仅干扰字符特征的识别，还容易增加字符提取模型过拟合风险。（2）甲骨拓片表面存在各种样式的裂痕干扰，这些裂痕具有不同的尺度和形状并且在外观上和甲骨字符十分相似，严重干扰甲骨字符的识别。（3）甲骨字符在拓片图像中的位置信息、几何先验等是未知的，这为字符特征的判别及约束字符在空间上的完整性上，带来了极大的阻碍。

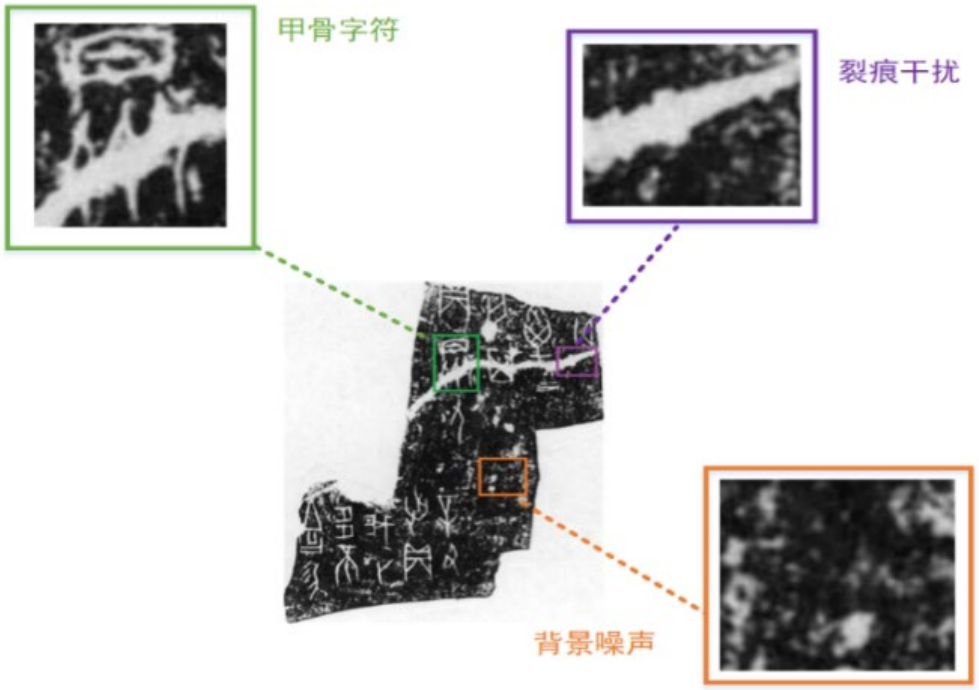


图 4.1 甲骨拓片图像局部特征展示

近年来，随着深度学习在诸多视觉领域的成功应用，出现了一些在理论上能够直接或间接的用于提取拓片图像中甲骨字符的方法。这些方法大致分为两大类：基于图像生成的方法和基于图像分割的方法。图像生成的方法（如 Pix2Pix[36]）将甲骨字符的提取视为一个图像到图像转换任务，通过训练一个端到端的神经网络，学习拓片图像与相应字符图像间的映射。基于图像分割的方法（如 U-Net[39]）将甲骨字符提取视为像素分类任务，通过对拓片图像进行逐像素分类，预测出字符在拓片图像中的所在区域。然而，由于这些方法在设计时并没有考虑高噪声、强干扰等因素，因此在实际字符提取效果上存在一定的局限性。通过本文的实验发现，基于分割的方法具有较强的区分拓片图像背景和甲骨字符的能力，但得到的字符图像往往比较粗糙，存在字符笔画粘连、模糊等问题，如图 4.2（中）所示；而基于生成的方法具有较强的结构信息描述能力，生成的甲骨字符在局部笔画细节上更为清晰，但往往会受背景噪声和裂痕的干扰，如图 4.2（右）所示。

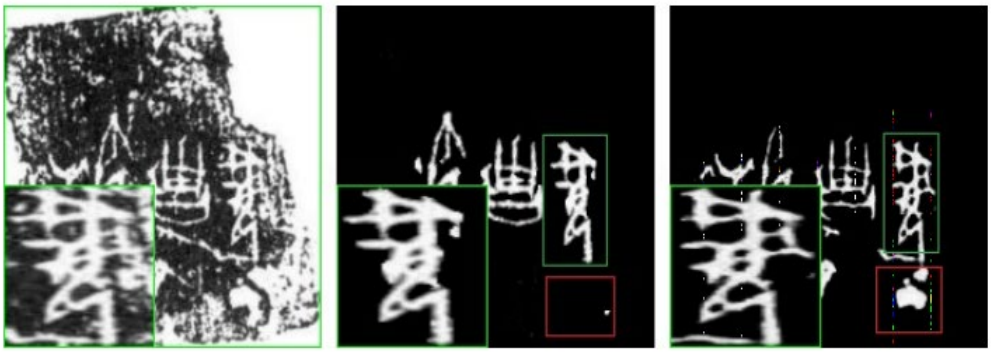


图 4.2 基于分割和生成方法的甲骨字符提取结果示例；从左至右依次是甲骨拓片输入，SegNet[71]的分割结果，Pix2Pix 的生成结果

鉴于上述的实验观察，本章介绍一种专门的甲骨字符提取模型，该模型将甲骨字符提取任务视为图像到图像转换任务，以生成网络为模型的基础骨架，将分割网络嵌入编码器网络以消除拓片背景噪声的影响，以期建立更为准确的拓片图像与对应甲骨字符图像间的映射关系。实验结果表明，相比于一些主流的图像生成和分割方法，该模型能够生成更加清晰、锐利的甲骨字符图像。

4.2 甲骨字符提取网络

为便于后续甲骨文字的检测与识别等工作，本章构建了一个专门的甲骨字符提取网络。该网络注重字符空间上下文信息的学习，并同时兼顾对复杂甲骨背景特征的判别。在网络结构上，该模型以生成对抗网络为基本骨架，由嵌入学习分支、字符生成分支、空间注意融合模块以及结构判别模块构成。下面将会对该字符提取模型的关键模块组成进行介绍。

4.2.1 网络模型概述

如 4.1 小节所述，基于分割的方法具有良好的背景噪声去除能力，而基于生成方法拥有更好的结构信息描述能力，本章节将两种方法相结合，构建了一个全新的甲骨字符提取模型。该模型将甲骨字符提取任务视为图像到图像转换任务，以生成网络为模型的基础骨架，将分割网络嵌入编码器网络以消除拓片背景噪声的影响，以期建立更为准确的拓片图像与对应甲骨字符图像间的映射关系。具体来说，（1）为了缓解拓片图像中背景噪声和甲骨裂痕的干扰，该模型的生成网

络包含一个嵌入学习分支（Embedding Learning Stream）以实现特征嵌入空间中甲骨背景和甲骨字符的可判别特征表示学习；（2）为适应拓片图像中甲骨字符大小的变化并生成清晰完整的甲骨字符图像，该模型使用残差模块（Residual Block）和多尺度特征通道连接，在生成网络中构建了一个字符生成分支（3）为了在降低甲骨噪声和甲骨裂痕干扰的同时保证字符在空间结构上的完整性，生成网络利用空间注意力模型（Spatial Attention Model, SAM）对两个分支的结果进行融合；（4）为保证生成的甲骨字符图像整体完整且细节清晰，该模型使用 GAN 为结构约束模型，分别从全局和局部角度评估生成的甲骨字符图像的一致性，络的整体结构如图 4.3 所示。

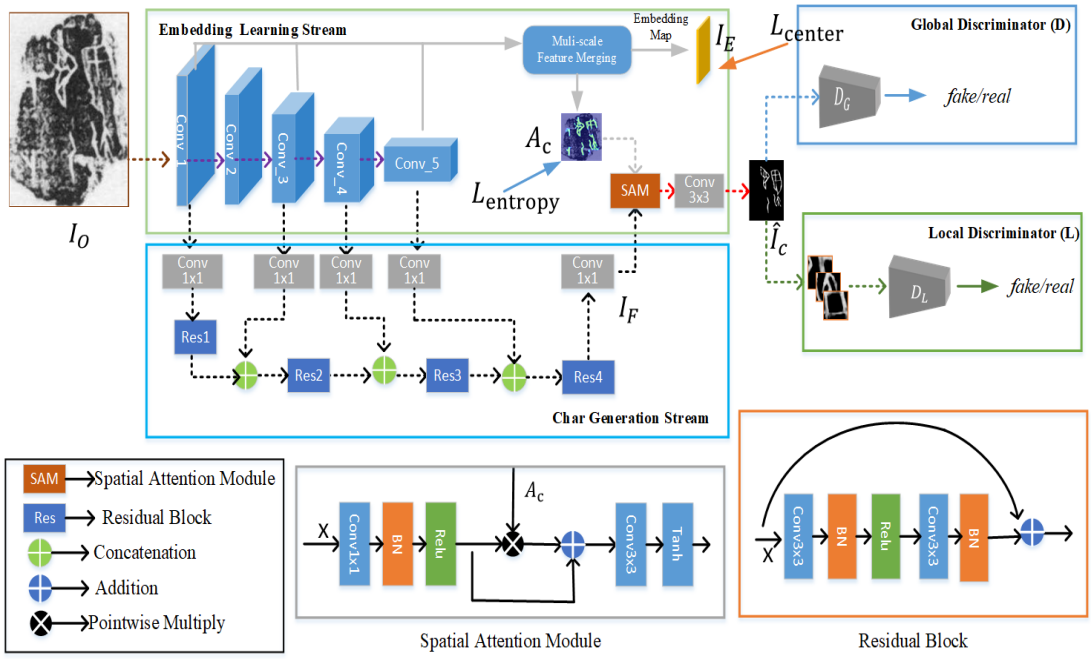


图 4.3 甲骨字符提取模型网络结构

4.2.2 嵌入学习分支

嵌入学习分支将甲骨背景和甲骨字符视为不同的类别实例，尝试在嵌入空间中学习它们的可判别嵌入特征，以缓解背景噪声、裂痕对字符提取的干扰。近来，基于聚类思想的学习方法在嵌入空间判别嵌入特征学习方面展现出较好的性能。如 DeepCluster[72]对分类网络的预测进行聚类，利用聚类结果更新深度卷积网络参数，用于无监督视觉特征学习；张等人将任意形状的场景文本视为不同的实例，并鼓励属于相同实例的像素朝着相同的中心靠近，反之相反。然而，这

些方法大多针对特定的应用场景，没有考虑到目标实例的视觉特征属性，不能直接应用到字符提取任务。

嵌入学习分支通过提高背景特征和字符特征的“类内一致性”来学习可判别特征表示。首先，利用分割网络提高模型像素级类别特征的判别能力，对拓片图像进行逐像素分类，达到学习甲骨背景和甲骨字符的可判别特征的目的。然后，基于 CenterLoss[73]自适应性地为不同的像素类别学习特定的特征中心，并鼓励属于同一类别的嵌入特征朝着相应的特征中心靠近。这样，嵌入学习分支不仅能够学习甲骨背景和甲骨字符的可判别特征，同时还能保留它们的视觉属性。具体的语义分割损失 $L_{entropy}$ 和中心损失 L_{center} 的数学表达如下：

$$L_{entropy} = -\frac{1}{N}\sum_{i=1}^{N} I_M^i \times \log(A_C^i) + (1 - I_M^i) \times \log(1 - A_C^i) \tag{4.1}$$

其中，I_M 表示真实字符图像 I_C 的二进制掩膜，i 表示 I_M 中第 i 的索引，N 表示 I_M 的像素总数。$A_C \in (0,1)$ 表示嵌入分支预测出的字符区域概率图。

$$L_{center} = \frac{1}{2}\sum_{i=1}^{N} \| I_{E_i} - C_{y_i} \|_2^2 \tag{4.2}$$

其中，I_{E_i} 代表嵌入学习分支中嵌入图 I_E 的第 i 个特征向量，C_{y_i} 表示 I_{E_i} 所属类别 y_i 的特征中心向量。

在网络结构上，嵌入学习分支相对简单，仅仅由一个编码器和四个卷积层组成。其中，编码器包含了一系列卷积层，并且卷积层的个数及内部组成和 VGG16 完全一致。在训练过程中，编码器首先对原始拓片图片输入进行特征编码，以获取多个尺度的特征图。紧接着对来自于(Conv_1、Conv_3、Conv_5)的特征图依次经过上采样、通道连接等操作进行特征合并。随后，合并后的特征图经过两个连续的卷积层进行上下文融合。最后，融合后的特征图分别经过两个并行的 3x3 卷积，得到最终的特征嵌入图 I_E 和字符区域得分图 A_C，具体结构如图 4.3 绿色框所示。

4.2.3 图像间映射学习

图像生成角度上，甲骨字符提取任务可视为从甲骨拓片图像到甲骨字符图像间的转换。和大多数图像到图像转换算法一样如 Pix2Pix、CycleGAN[74]，该字符提取网络中的字符生成分支通过训练一个特征编码器和解码器学习拓片图像

与对应字符图像之间的映射。为更好的将图像到图像转化模型应用充满大量噪声和裂痕干扰的甲骨片数据，本文对其做进一步改进。

具体的，在字符生成分支的基础上引入共享相同特征编码的嵌入学习分支，嵌入学习分支通过对拓片图像逐像素分类在嵌入特征空间学习甲骨背景和字符的可判别特征，间接提高字符生成分支对不同像素特征的区分能力，以缓解拓片图像中复杂的噪声和裂痕对字符生成的干扰。最后，在生成网络的末尾，本文又嵌入了一个空间注意模块（SAM）对两个不同分支的结果进行融入。SAM 利用来自于嵌入学习分支中字符区域注意 A_C，指导字符生成分支注重融合特征图中的字符区域，具体结构如图 4.3 蓝色框所示。

4.2.4 空间结构约束

甲骨文字形状多样、结构复杂，为确保生成的甲骨文字在空间结构上完整性，本文使用 GAN 作为结构模型，用以融入字符的空间结构先验。像一些图像修补方法一样如文献[75, 76]，使用全局和局部判别器评估生成的字符图像全局和局部特征的一致性。其中，全局判别器以完整的字符图像的作为输入，比较其于真实字符图像的全局一致性，检查其是否引入额外的噪声、裂痕干扰；而局部判别器以局部字符块作为输入，比较其与真实局部块的局部一致性，检查局部字符在笔画细节上是否完整。为充分挖掘高噪声背景下的甲骨字符局部特征，生成的字符图像在送入局部判别器之前，首先将其裁剪为若干个局部块，然后选择其中和真实字符块误差相对较大的若干个局部块，并从通道维度进行连接作为局部判别器的输入，以强迫生成网络发现并提取复杂甲骨背景中的字符特征。众所周知，GAN 在训练过程中不稳定容易模式塌陷，本文使用了近来提出比较稳定、收敛速度更快的 LSGAN[77]。生成网络 G，全局和局部判别器 D_G，D_L 的损失函数表达如下：

$$L_{global}(G,D_G)=E_{I_{C\sim P_{data}(I_C)}}[(D_G(I_C)-1)^2]+E_{I_{O\sim P_{data}(I_O)}}[D_G(G(I_O)^2] \tag{4.3}$$

$$L_{local}(G,D_L)=E_{I_{C\sim P_{data}(I_C)}}[(D_L(T(I_C))-1)^2]+E_{I_{O\sim P_{data}(I_O)}}[D_L(T(G(I_O))^2] \tag{4.4}$$

其中，P_{data} 是训练数据的经验分布，I_O 表示原始拓片图像输入，I_C 表示与拓片图像对应的真实字符图像，T 表示裁剪和连接操作。

判别器在网络结构设计上，遵循 PatchGAN 的设计原则，通过预测一个 N×N

的评估矩阵，用于捕获更加清晰、细致的字符局部细节。全局和局部判别器具体的结构和参数设置如表 4.1、表 4.2 所示。

表 4.1 全局判别器结构详情

Type	Kernel	Stride	OutPuts
Conv	5x5	1x1	32
Conv	5x5	2x2	64
Conv	5x5	2x2	128
Conv	5x5	1x1	64
Conv	3x3	1x1	1

表 4.2 局部判别器结构详情

Type	Kernel	Stride	OutPuts
Conv	3x3	1x1	32
Conv	3x3	2x2	64
Conv	3x3	2x2	128
Conv	3x3	1x1	1

4.3 实验验证

4.3.1 甲骨文提取数据集

目前，在计算机视觉领域，几乎没有任何公开可达的甲骨文研究数据集，本章节实验用到的数据集源自于论文作者手工构建。甲骨文作为重要的文化遗产，主要保存在甲骨拓片或收录在由图片构成的著录中，本章节实验中用到的甲骨片数据主要是利用扫描仪器从甲骨学典藏中扫描而来。从视觉上看，这些扫描而来的甲骨拓片图像表面存在严重的退化，但是在构成上相对简单，仅由甲骨背景（包括背景噪声和甲骨裂痕）和甲骨字符构成，其中甲骨背景、字符特征在外观上相对单一，字符特征在亮度值上偏亮，而背景特征偏暗。基于这样的一个观察，实验中，本文仅仅选择了少量具有代表性且退化严重的拓片图像进行训练和测试，包括 405 对训练样例（甲骨拓片图像和对应的甲骨字符图像）、35 对验证样例和 300 张测试样例。

为了确保模型能够学习准确的特征表示，根据上述的少量拓片图像训练样例，对样本进行简单扩充。扩充主要涉及以下操作：（1）线性变换：缩放、裁剪、平移、操作；（2）仿射变换：随机旋转、翻转、变形操作；（3）颜色变换：模糊、对比度提升、高斯滤波等操作；（4）拓片图像与字符图像重新组合。首先，利用

工具软件从拓片图像中裁剪甲骨字符，构成 甲骨字符字典； 然后，选取若干张背景复杂的拓片图像并移除其中的甲骨字符，得到候选甲骨背景；最后根据字符字典和甲骨背景进行重新组合，具体如图 4.4 所示。

最终，得到一个包含（405+2825）对训练样例、（35+165）对验证样例、（300+200）测试样例的混合甲骨拓片数据集。

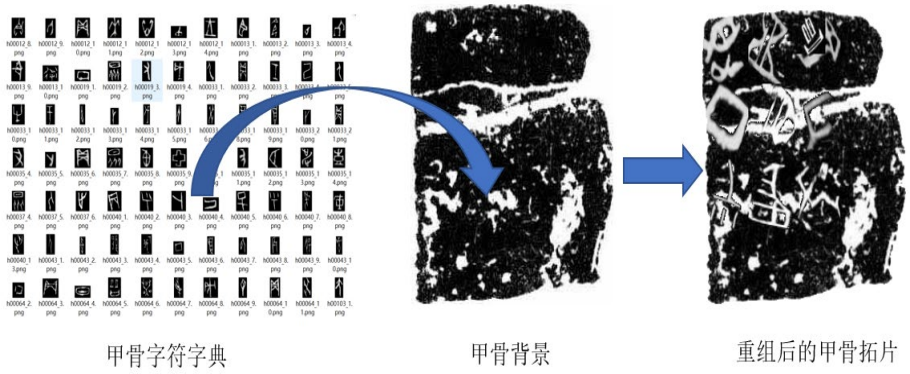


图 4.4 甲骨拓片图像与甲骨字符图像重新组合图示

4.3.2 评估指标

本章节从两个角度对提出的字符提取模型的性能进行评估：图像生成角度和图像分割角度。

从图像生成角度，使用峰值信噪比(PSNR)和结构相似性（SSIM）指标来测量预测值和真实值之间的差距。PSNR 和 SSIM 是一种常见评估图像成质量的客观标准。PSNR 和 SSIM 的值越高，表明生成的字符图像质量越高，越接近真实值。PSNR、SSIM 的计算如下:

$$MSE = \frac{1}{h*w}\sum_{i=1}^{h}\sum_{j=1}^{w}\| F(i,j) - G(i,j) \|^2 \tag{4.5}$$

$$PSNR = 20*\log_{10}(\frac{MAX_1}{MSE}) \tag{4.6}$$

其中，MSE 为生成图像与对应真实图像的均方误差。MAX_1表示图像颜色的最大值。

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{4.7}$$

其中，x、y为测量的成对图像，μ_x、μ_y为x、y的平均值，μ_x^2、μ_y^2为x、y的方差，σ_{xy}为x、y的协方差。c_1、c_2分别为常数，用以避免分母为0。

从图像分割的角度，由于大多数甲骨字符的像素值（归一化后）趋向于 1（字符边缘或者一些特殊字符除外），可近似的将生成的字符图像视为一种特殊的图像分割（二分类）。和图像分割模型的评估相似，使用平均交并比 mIoU 和单个字符类别的交并比 IoU(char),分别测量生成的字符图像与真实值之间的相关程度以及局部字符与对应真实值的相关程度。其中，mIoU 或 IoU 的值越高，说明像素被正确分类的比例就越高，生成的字符图像接近真实值的概率就越大。此外，由于生成的甲骨字符图像的非字符区域像素值接近于 0，但不为 0，对于字符图像的 IoU 计算可能存在一定的误差。为了获得更加准确的 IoU 值，在 IoU 计算之前，需要对生成的字符图像进行阈值选择处理。具体的阈值根据经验设定，本实验中，该阈值设置为 0.2，IoU 的计算表达如下:

$$IoU = \frac{TP}{TP + FP + FN} \tag{4.8}$$

其中，TP、FP、FN表示分类结果为真正、假正、假负的像素个数。

此外，为验证模型抑制裂痕干扰的能力，本实验对生成的字符图像上的裂痕数量进行了统计。对于生成的字符图像，假设其仅仅由背景噪声、甲骨字符和裂痕构成，这些背景噪声相对稀少，可通过简单的形态学开运算进行滤除，而裂痕干扰则可以使用对应的字符真实值选取，最后求解裂痕干扰中连通分量并统计其个数。具体包括以下 5 个步骤：

步骤 1 使用真实字符图像标记I_{GT}按照公式(4.9)中的计算原则，滤除生成的 字符图像I_P上的字符特征，得到粗糙的裂痕背景。

$$\mathrm{Mask} = I_P \times (1 - I_{GT}) \tag{4.9}$$

步骤 2 使用形态学开运算对粗糙的裂痕背景进行膨胀和腐蚀操作，去除其中的背景噪声,得到纯净的裂痕。

步骤 3 计算纯净裂痕中的连通分量，并去除关于背景的连通分量。

步骤 4 遍历每个连通分量，并移除小于 30 个像素大小的连通区域。

步骤 5 对现有的连通分量进行统计，得到每一张字符图像上的裂痕总数。

本实验中，随机从 190 条验证集记录中抽取 50 条作为统计样本，相应的统计的结果如表 4.3 所示。

表 4.3 裂痕个数统计

Type	Models	Counts
(a)	BiclyGAN[78]	304
	Pix2Pix	272
	CycleGAN[74]	268
(b)	PSPNet	1196
	SegNet	190
	U-Net	174
(c)	Ours	18

4.3.3 消融实验

4.3.3.1 判别损失函数

本章节提出的甲骨字符提取模型联合交叉熵损失L_E和中心损失 L_C共同约束嵌入学习分支的甲骨背景和甲骨字符的可判别嵌入特征学习。为验证该联合损失的有效性，将其与单独的使用交叉熵损失L_E、区别损失 *DiscLoss* [79] 进行对比。在实验设置上，除了损失函数的不同之外，整个甲骨字符生成模型的结构及超参数设置均是相同的。表 4.4 展示了在不同损失函数下的评估结果。

从表 4.4 中可以看出，区别损失 L_D 在各项指标上都是最差的。其原因可能是在鼓励同簇特征向中心靠拢过程中，丢失了某些视觉属性（例如，极端情况下，嵌入特征朝向零向量方向靠近）。相比于区别损失，交叉熵损失L_E的表现（在$mIoU$、IoU、PSNR, SSIM 指标上，分别提升了 0.63、1.18、0.51、0.43 个百分点）。最关键的是，在联合损失（$L_E + L_C$）的监督下，甲骨字符提取模型的表现最佳，在各项指标均是最优的。这表明联合交叉熵损失和中心损失能够更有利于字符可判别嵌入特征的学习和甲骨字符图像的生成。

表 4.4 不同可判别损失的比较结果

Loss	mIoU	IoU	PSNR	SSIM
L_E	87.03%	76.35%	23.25	94.73%
L_D	86.4%	75.17%	22.74	94.3%
$\boldsymbol{L_E + L_C}$	88.07%	78.28%	23.83	95.17%

4.3.3.2 嵌入学习分支

为缓解拓片图像中噪声、裂痕对字符提取的影响，字符提取模型引入了额外的嵌入学习分支。为了验证嵌入学习分支的有效性，将嵌入学习分支从字符提取模型中移除。移除后的评估结果如表 4.5(a)所示（为了便于描述，字符提取模型的关键组成使用字母符号表示，符号含义如表 4.6 所示。）

从表 4.5 结果显示，移除嵌入分支后，字符提取模型的性能显著下降（如表 4.5 中，（a）和（b）（c）（d）的各项指标比较）。这充分表明嵌入学习分支的存在对甲骨字符提取模型的提取效果有显著的提升。

表 4.5 字符生成模型不同模块组合的评估结果

ID	组合	mIoU	IoU	PSNR	SSIM
(a)	CGL	82.60%	68.81%	19.46	89.04%
(b)	ECGL	87.7%	77.6%	23.11	94.44%
(c)	ECGA	87.63%	77.46%	23.52	95.00%
(d)	ECGLA	88.07%	78.28%	23.83	95.17%

表 4.6 字符提取模型关键模块的符号表示

符号	含义
E	嵌入学习分支
C	字符生成分支
G	全局判别器
L	局部判别器
A	空间注意模块

4.3.3.3 空间注意模块

给出一张甲骨拓片图像，甲骨字符提取模型的目标是生成对应的甲骨字符图像。该过程中,甲骨字符在拓片图像中的位置信息是未知的。为此，在生成网络的末尾，引入了空间注意力模型（SAM）。SAM 利用来自于嵌入学习分支中字

符区域信息，指导字符生成分支注重特征图的字符区域.为了证明使用 SAM 的有效性，本实验对甲骨字符提取模型中的 SAM 模块进行移除,移除后的评估结果如表 4.5(b)所示。

通过表 4.5(b)和表 4.5(d)的比较可以看出，移除字符空间注意模块后，字符提取模型的性能出现小幅下降。相比于使用 SAM,模型在 $mIoU$ 、IoU、PSNR、SSIM 指标上,分别降低了 0.37、0.68、0.72 和 0.73 个百分点。这在一定程度上表明，在生成网络的末尾，使用 SAM 对字符提取模型的性能是有益的。

4.3.3.4 局部判别器

甲骨字符形状多样、结构复杂且随机的分布在拓片上的任意位置。为约束生成的字符在空间结构上的完整性，使用额外的局部判别器评估字符特征的局部一致性。为验证局部判别器空间约束的有效性，在训练期间，将局部判别器移除，移除后的评估结果如表 4.5（c）所示。

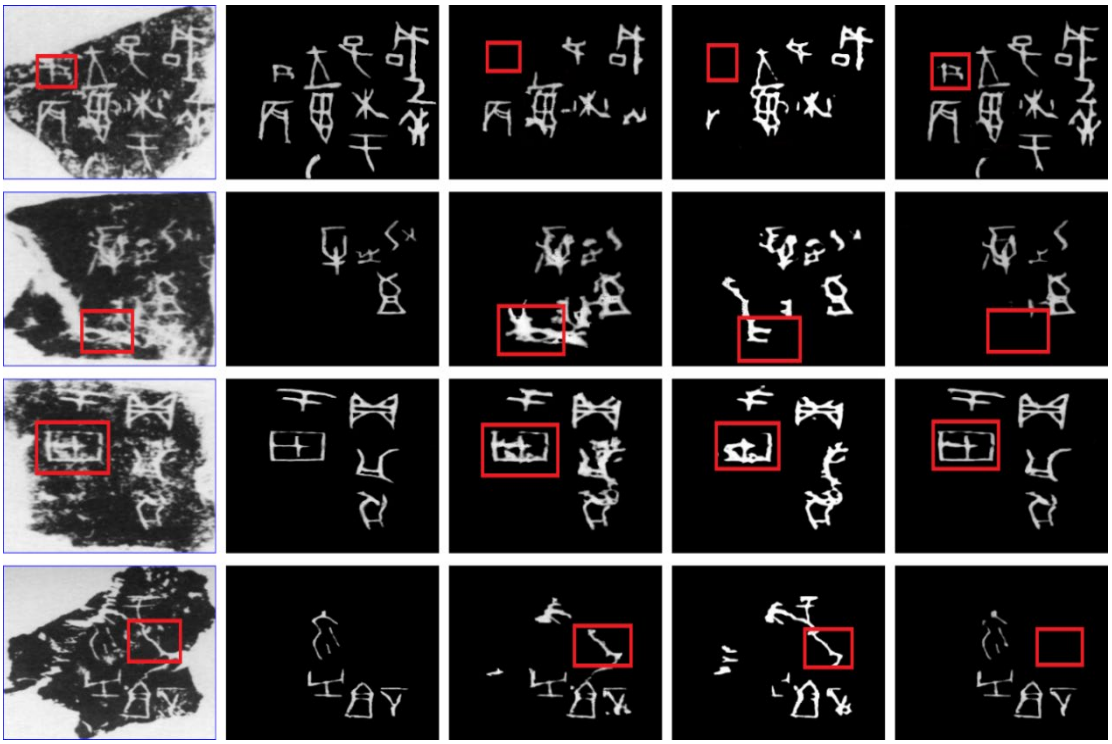
通过表 4.5（c）和表 4.5（d）的比较可以看出，移除局部判别器后，字符提取模型的性能出现一定的下降。相比于使用局部判别器，移除后模型在 mIoU、IoU、PSNR、SSIM 指标上分别降低了 0.44、0.82、0.31 和 0.17 个百分点。这意味着，使用局部判别器约束字符的局部细节的完整性是有效的。

4.3.4 对比评估

4.3.4.1 与经典图像生成模型的比较

本部分，将字符提取模型与一些经典的图像到图像转换模型（Pix2Pix, CycleGAN, BicyCleGAN）进行比较。为公平起见，直接使用了这些模型的官方代码和默认的超参数设置。相应的定量评估、定性评估以及裂痕统计结果如图 4.5、表 4.7、表 4.3 所示。

如图 4.4 所示，一流的图像到图像转换模型一定程度上也可以提取拓片图像中的字符信息，并能保留清晰的局部细节。然而，对于一些尺度较小、不太显著的字符有可能被忽略（如图 4.5 第一行所示）。其次，在生成的字符图像上引入大量和字符特征比较相似的噪声或裂痕干扰（如图 4.5 第二、四行所示）。相反，由本文字符生成模型生成的字符图像几乎将拓片上的字符信息完全保留，并没有引入过多的噪声和裂痕干扰（如图 4.5 第五列）。主观上看，本文提出的字符提取模型生成的甲骨字符图像局部笔画清晰，并引入较少的噪声和裂痕干扰。



(a) Input (b) GT (c)Pix2Pix (c)CycleGAN (d)Ours

图 4.5 甲骨拓片图像和经典的图像生成模型的字符提取结果

表 4.3(a)展示不同的生成模型输出的字符图像中裂痕连通分量的个数，整体上这三个模型的输出中都引入了较多的裂痕，其中 Pix2Pix 模型引入的最少，但也高达 268 个。表 4.3(c)展示了字符提取模型的统计结果，仅仅包含 18 个，远低于其它三个模型，这表明，相比于这些经典的图像到图像转模型，该提取模型对裂痕干扰的抑制是有效的。

表 4.7 展示了不同的生成模型输出的字符图像在 PSNR 和 SSIM 指标上的测量结果。很显然，本文提出的字符提取模型在这两个指标上均是最佳的，并大

幅超越次优结果 Pix2Pix（分别超越 5.27%, 5.93%）. 这表明，相比于这些经典的图像到图像转换模型，本论文提出字符提取生成模型生成的字符图像中引入更少的噪声和裂痕干扰并捕获更多的字符局部细节。

综上所述，无论是在裂痕引入量上还是 PSNR，SSIM 上，本论文提出的字符提取模型均取得较优的效果，因此上述的主观结论是正确的，相比于这些经典的图像到图像转换模型，本章节的字符生成模型能够生成更加清晰、更加完整的字符图像。

表 4.7 和经典图像生成模型的量化比较结果

Models	PSNR	SSIM
BicyleGAN	17.90	86.40%
CycleGAN	18.03	88.25%
Pix2Pix	18.56	89.24%
Ours	23.83	95.17%

4.3.4.2 与经典分割方法的比较

大多数甲骨字符特征的像素值（归一化后）趋向于 1，因此，可近似的将生成的字符图像视为一种特殊的图像分割（二分类）。本部分将字符提取模型与一流的图像分割模型（FCN16、ERFNet[80]、U-Net、SegNet）进行比较。 此外，由于拓片图像中字符像素和背景像素在比例存在严重的失衡，在模型训练期间，默认为每个分割模型使用相同的类别平衡策略，以获得更加的字符分割效果。类别平衡策略的具体表示如下:

$$W^{(c)} = \begin{cases} 1 & N_c = 0 \\ 2 - \frac{N_c}{N} & otherwise \end{cases} \tag{4.9}$$

其中，$W^{(c)}$ 代表不同类别实例的权重系数，N_c 和 N分别代表类别c的像素个数和拓片图像中总的像素个数。

图 4.6 展示了字符提取模型和分割模型的字符提取效果。从视觉上看，分割模型几乎将所有的字符区域都预测出来，并且引入了较少的噪声或裂痕干扰。然而，通过分割的方式得到的字符图像，在字符的局部细节上往往比较模糊、粗糙，

甚至存在部分笔画粘连的问题（如图 4.6 一、三、四列所示）。 其次，由于分割的方法仅仅预测出字符在拓片图像上的区域信息，并没有对字符特征进行重建，一些字符笔画存在与真实字符风格不一致的问题（如图 4.6 第二行所示）。相反，本文的字符提取模型对拓片图像进行重建，生成的字符图像在结构上更为清晰、风格更为统一（如图 4.6 第五列所示）。

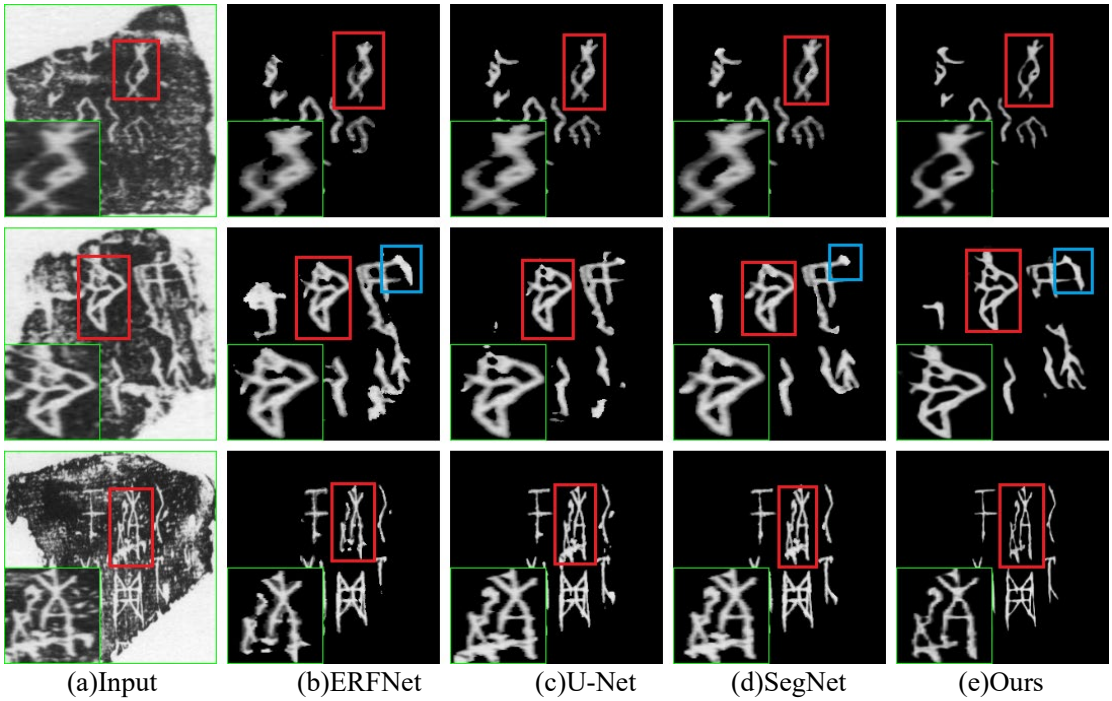


图 4.6 甲骨拓片图像和经典的图像分割模型的字符提取结果

表 4.3(b)展示了不同分割模型输出的字符图像中裂痕连通分量的个数，其中 SegNet、U-Net 引入了相对较少的裂痕，远低于表 4.3(a)中的图像生成模型。但相对于本论文提出的字符提取模型，仍有一定的差距，这也表明，即使相较这些经典的分割模型，该字符提取模型仍然具有抑制裂痕干扰的优势

表 4.8 和主流分割模型的量化比较结果

Models	mIoU	IoU(char)
ERFNet	80.9%	64.5%
FCN16	83.3%	68.7%
U-Net	87.48%	76.48%
SegNet	88.6%	78.5%
Ours	88.07%	78.28%

4.4 本章小结

本章节构建了一个全新的甲骨字符提取网络，该网络以 GAN 作为基本骨架，在原有图像生成网络 Pix2Pix 的基础上，引入了额外的嵌入学习分支，以缓解噪声、裂痕等干扰对字符提取的影响。新模型拥有分割网络区分不同类别像素特征的能力的同时还兼备生成网络描述空间结构的特性，能够生成高质量的甲骨字符图像。在实验阶段，分别和较为经典的图像生成模型如 Pix2Pix、CycleGAN 和图像分割模型如 SegNet、U-Net 进行了比较，证明所提出的模型在性能上以及生成质量上均具有明显的提升。

5 总结与展望

5.1 总结

甲骨文是迄今为止中国已发现的最古老、体系最完整的文字之一，被认为是现代汉字的早期形式，其记录着 3600 年前殷商时期先民的生活、思想状态、经济生产以及社会生活等方方面面，对于了解中国以及世界的过去具有非常重要的意义。作为甲骨文字字符破译的基础，甲骨文字检测是甲骨学研究领域一项重要的研究内容，然而当前字符检测工作需要甲骨学专家的参与，对甲骨学知识水平要求较高，且效率低下。此外，经过长期的自然腐蚀，甲骨片表面退化严重，上面的甲骨字符模糊不清，严重阻碍甲骨文字的可视性，不利于甲骨文字的检测和识别、推广等一系列工作。针对这些问题，本文主要独创性工作包括：

1. 在甲骨文字检测方面，提出一种更为简单且有效的甲骨字符检测算法，使用多尺度高斯核函数多字符区域进行编码，改善了字符区域之间容易区域重叠问题并使得模型学习字符区域意识成为可能，有利于检测准确率的提升。其次，高斯核的使用避免了复杂的网络设计以及大量锚点框的需要，极大的提高检测效率。在甲骨文检测数据集上，取得 F-Measure 高达 83.3%的成绩，远超一些较为主流的通用目标检测算法。
2. 在甲骨文字提取方面，首先构建了一个像素集甲骨文提取数据集，为后续甲骨文检测和提取工作提供数据基础。接着论文提出了一种全新的甲骨字符提取算法框架，在原有图像生成框架 Pix2Pix 的基础上，引入了额外的嵌入学习分支以消除拓片背景噪声的影响，以期建立更为准确的拓片图像与甲骨字符图像间的映射关系。最后，为获取内容完整 且细节清晰的生成结果，该模型结合使用全局判别器和局部判别器对生成的甲骨字符图像进行一致性判别。在甲骨字符提取数据集上，相比于较为主流的图像生成算法和图像分割算法，能够生成更加清晰、完整的甲骨字符图像。

5.2 展望

本文主要探索深度学习技术在甲骨文研究数据上的运用，提出使用多尺度

高斯核表示密集分布的甲骨字符区域，将字符边界框回归转换为对应高斯核的预测，从而提高字符检测的效率。为增强甲骨片图像的可视性，分别探索图像分割方法和图像生成方法在甲骨片数据上的应用，最后融合图像分割网络和图像生成网络各自的特性，构建了一个专门的甲骨字符提取算法，用以生成更加清晰、细致的甲骨字符图像。但是对于甲骨文检测以及字符提取工作仍然处于早期发展阶段，存在很多问题有待解决，未来进一步改进可以有以下几个方面：

(1) 截至到目前，甲骨学领域缺少统一的甲骨文检测数据集，大多数现有的甲骨文检测方法都是使用自建的数据集进行模型训练以及方法验证，这显然不利于不同甲骨文检测方法的比较，因此，构建一个统一的甲骨文检测基准是后续亟待解决的一个重要任务。

(2) 众所周知，深度学习模型的训练依赖大量的标记的数据集，而构建像素水平和字符水平的甲骨文数据集是一个耗时、耗力的工程，再加上甲骨片上存在很多模糊不清的字符，甚至存在残缺，依赖人工主观判断，进一步增加了数据集制作的难度。除此之外，甲骨片的字符信息是类未知的，网络不能捕获足够的字符语义信息，不利于字符提取任务或检测任务精度上的提升，因此，探索在弱监督以及类未知条件下的甲骨字符检测与提取任务是未来研究的一个重要方向。

(3) 甲骨片经过长期的自然腐蚀，表面充斥着大量的噪声和各种样式的裂痕干扰，这些干扰对与甲骨文相关的视觉任务带来极大的影响。尽管本文中使用分割网络来学习字符和甲骨背景的可判别特征，但实际去噪效果依然有限，因此，探索一种有效的、简单的方法避免噪声和裂痕的干扰，对于后续甲骨文相关技术的研究至关重要。

参考文献

[1] 陈梦家.解放后甲骨的新资料和整理研究[J].文物参考资料, 1954, 5: 3-8.
[2] 刘鹗. 铁云藏龟[M]. 抱残守缺斋（石印本）, 1903.
[3] 陈梦家. 殷墟卜辞综述[M]. 北京:中华书局，1988.
[4] 李学勤，彭裕商. 殷墟甲骨分期研究[M].上海:上海古籍出版社. 1996
[5] Cheung C. The Chinese History That Is Written in Bone [EB/OL]. https://www.sapiens.org/archaeology/chinese-oracle-bones-history/. 2018.1.23
[6] 刘永革，栗青生.可视化甲骨文输入法的设计与实现[J].计算机工程应用(17):139-140.
[7] 顾绍通, 马小虎, 杨亦鸣. 基于字形拓扑结构的甲骨文输入编码研究[J]. 中文信息学报, 2008, 22(4):123-128.
[8] 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]. 科学技术工程, 2014(30):67-70.
[9] 顾绍通.基于拓扑配准的甲骨文字形识别方法[J]. 计算机与数字工程, 2016, 44(10):2001-2006.
[10] 毛建军.甲骨文献全文数据库的建设与思考[J]. 图书馆学研究,2010(12):37-3.
[11] 李志勇,高峰.基于知网的甲骨文可拓模型建模技术[J].计算机与现代化, 2015(5):30-34.
[12] 袁冬, 熊晶, 刘永革. 面向甲骨文的实例机器翻译技术研究[J]. 数据分析与知识发现,2012, 28(5):48-54.
[13] 熊晶,高峰,吴琴霞.甲骨文大规模基础数据的语义挖掘研究[J]. 数据分析与知识发现, 2015, 31(2):7-14.
[14] 陈婷珠. 殷商甲骨文字形系统再研究[J]. 上海: 华东师范大学, 2007.
[15] LAKSHMI T R V，REDDY C V K. Object Classification Using SIFT Algorithm and Transformation Techniques[M]. Singapore：Springer，2019：403-408.
[16] LIENHART R，MAYDT J. An extended set of Haar-like features for rapid object detection[C]. International Conference on Image Processing. IEEE，2002：900-903.
[17] AL JAROUF Y A，KURDY M B. A hybrid method to detect and verify vehiclecrash with haar-like features and SVM over the web[C]. International Conference on Computer and Applications. IEEE，2018：177-182.
[18] DALAL N，TRIGGS B. Histograms of oriented gradients for human detection[C]. IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE，2005：886-893.
[19] APOOR R，GUPTA R，JHA S，et al. Detection of power quality event using histogram of oriented gradients and support vector machine[J]. Measurement，2018，120：52-75.
[20] SUBAS A，DAMMAS D H，ALGHAMDI R D，et al. Sensor based human activity recognition using adaboost ensemble classifier[J]. Procedia Computer Science，2018，140：104-111.
[21] FARIS H，HASSONAH M A，ALA'M A Z，et al. A multi-verse optimizer approach for

feature selection and optimizing SVM parameters based on a robust system architecture[J]. Neural Computing and Applications，2018，30（8）：2355-2369.

[22] VIOLA P，JONES M. Robust real-time object detection[J]. International Journal of Computer Vision，2004，57（2）：137-154.

[23] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection[C]. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,2005, Diego, CA, USA,886-893.

[24 R. Girshick, J. Donahue, T. Darrell, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, Columbus, OH, 2014, 580-587.

[25] J. Redmon, S. Divvala, R. Girshick, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, Las Vegas, 779-788.

[26] J. R. Uijlings, K. E. Van De Sande, T. Gevers, et al. Selective search for object recognition[J]. International journal of computer vision, 2013, 104(2): 154-171.

[27] K. He, X. Zhang, S. Ren, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[C]. European Conference on Computer Vision, 2014, 346-361.

[28] R. Girshick. Fast r-cnn[C]. Proceedings of the IEEE international conference on computer vision, 2015, 1440-1448.

[29] S. Ren, K. He, R. Girshick, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6): 1137-1149.

[30] Everingham M, Van Gool L, Williams C K I, et al. The PASCAL visual object classes challenge 2007 (VOC2007) results[J]. 2007.

[31] J. Redmon, A. Farhadi. YOLO9000: better, faster, stronger[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, 2017, 6517-6525.

[32] Meng, L. Two-stage recognition for oracle bone inscriptions[C]. Image Analysis and Processing - ICIAP , 2017, 10485:672-682.

[33] 王浩彬. 基于深度学习的甲骨文检测与识别研究[D]. [硕士学位论文]. 广州：华南理工大学, 2019.

[34] J. Xing, G. Liu, and J. Xiong. Oracle bone inscription detection: A survey of oracle bone inscription detection based on deep learning algorithm. Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing, Sanya china, December 2019, 1-8.

[35] G. Liu, J. Xing, and J. Xiong. Spatial Pyramid Block for Oracle Bone Inscription Detection. ICSCA 2020: 2020 9th International Conference on Software and Computer Applications, February 2020, 133-140.

[36] Wang X, Yan H, Huo C et.al. Enhancing Pix2Pix for Remote Sensing Image Classification[C]，

2018 IEEE International Conference on Pattern Recognition (ICPR). Beijing: IEEE, 2018: 2332–2336.

[37] Goodfellow I, M Xu, B, et.al. Generative adversarial nets[C]. Advances in Neural Information Processing Systems (NIPS), 2014: 2672–2680.

[38] Alec Radford,Luke Metz, Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks[EB/ OL]. https: / / arxiv. org / abs/ 1511. 06434,2016-01.

[39] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation [C]. Medical Image Computing and Computer-Assisted Intervention (MICCAI). Munich, Germany : Springer, Cham ,2015: 234–241.

[40] Wang X, Yan H, Huo C et.al. Enhancing Pix2Pix for Remote Sensing Image Classification[C]. IEEE International Conference on Pattern Recognition (ICPR). Beijing: IEEE, 2018: 2332–2336.

[41] C. Wang, C. Xu, C. Wang et al. Perceptual adversarial networks for image-to-image transformation[J]. IEEE Transactions on Image Processing,2018,27(8):4066-4079.

[42] C Wang, H J Zheng, Z B Yu,et al. Discriminative region proposal adversarial networks for high-quality image-to-image translation[C].Proceedings of the 15th European Conference on Computer Vision(ECCV),Munich,Germany,2018:796-812.

[43] W Chen, J Hays. SketchyGAN:towards diverse and realistic sketch to image synthesis [C] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Salt Lake City,UT,USA,2018:9416-9425.

[44] M Liu, Y Ding, M Xia, et al. STGAN:a unified selective transfer network for arbitrary image attribute editing[C] Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR),Long Beach,CA,USA,2019:3668-3677.

[45] T Park, M Liu, T Wang et al. Semantic Image Synthesis With Spatially-Adaptive Normalization [C] Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, 2337-2346.

[46] J Long, E Shelhamer T Darrell et al. Fully convolutional networks for semantic segmentation [C].2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, 431-3440..

[47] L C Chen , G Papandreou , I Kokkinos , et al. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs[J]. arXiv, 2014.

[48] L. Chen, G. Papandreou, I. Kokkinos et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs[C]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017.40(4):834-848.

[49] L Chen. L C et al. Rethinking Atrous Convolution for Semantic Image Segmentation[J].arXiv.2017.

[50] H Zhao，J Shi，X Qi，et al. Pyramid scene parsing network[C] Proceedings of the IEEE

Conference on Computer Vision and Pattern Recognition，2017：2881-289

[51] H Li, P Xiong, J An et al. Pyramid Attention Network for Semantic Segmentation[J]. arXiv.2018.

[52] Fu J，Liu J，Tian H，et al. Dual attention network for scene segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2019, 3146-3154.

[53] Liu W, Cheng F, Dragomir A, et al. SSD: Single Shot MultiBox Detector[C]. ECCV2016, 2016, 9905: 21-37.

[54] J. Dai, Y. Li, K. He, et al. R-fcn: Object detection via region-based fully convolutional networks[C]. Advances in neural information processing systems, Barcelona, 2016, 379-387.

[55] Lin T, P Dollar, He K, et al. Feature Pyramid Networks for Object Detection[C], Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, 2117-2125.

[56] Bharat S, Larry S. Davis. An Analysis of Scale Invariance in Object Detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018,3578-3587.

[57] He K, Gkioxari G, Dollar P, et al.Mask R-CNN[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2020,42(2): 386－397.

[58] Redmon, A. Farhadi. YOLO9000: Better, Faster, Stronger[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, 6517-6525.

[59] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[60] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[61] 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 (CVPR), 2016, 770-778.

[62] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]. Proceedings of the IEEE international conference on computer vision, 2017,2980-2988.

[63] Fu C Y, Liu W, Ranga A, et al. Dssd: Deconvolutional single shot detector[J]. arXiv preprint arXiv:1701.06659, 2017.

[64] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]. Proceedings of the IEEE international conference on computer vision, 2017,2980-2988.

[65] Baek, Y., Lee, B., Han, D, et al.Character Region Awareness for Text Detection[C],IEEE/CVF Computer Society Conference on Computer Vision and Pattern Recognition, Long Beach, 15-20 June 2019, 9357-9366.

[66] Newell, A., Yang, K. and Deng, J. Stacked Hourglass Networks for Human Pose Estimation. Computer Vision—ECCV 2016, Springer, Cham, 483-499.

[67] Jonathan L, Evan S, Trevor D. Fully Convolutional Networks for Semantic Segmentation[C].Proceedings of the IEEE Conference on Computer Vision and Pattern

Recognition (CVPR), 2015, pp. 3431-3440.

[68] Wang, W., Li, X,Liu, T. Shape Robust Text Detection with Progressive Scale Expansion Network. IEEE/CVF Computer Society Conference on Computer Vision and Pattern Recognition, Long Beach, 15-20 June 2019, 9336-9345.

[69] Zhang, S., Wen, L., Bian, X., et al. Single-Shot Refinement Neural Network for Object Detection. IEEE Transactions on Circuits and Systems for Video Technology, 31, 674-687.

[70] Liu, S., Huang, D. and Wang, Y. Receptive Field Block Net for Accurate and Fast Object Detection. Computer Vision. Computer Vision—ECCV 2018, Springer, Cham, 404-419.

[71] Badrinarayanan V, Kendall A, Cipolla R, et al. SegNet : A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481- 2495.

[72] Caron M, Bojanowski P, Joulin A, et al. Deep clustering for unsupervised learning of visual features[C]. 2018 European Conference on Computer Vision(ECCV). Munich, Germany: Springer, Cham 2018:1139–156.

[73] Soetens S, Sarris A, Vansteenhuyse K. A Discriminative Feature Learning Approach for Deep Face Recognition [C]. 2016 European Conference on Computer Vision (ECCV). Amsterdam, The Netherlands: Springer, Cham,2016:181–184.

[74] Park T, Isola P, Efros A.et al. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks[C]. 2017 IEEE International Conference on Computer Vision (CVPR). Venice, Italy: IEEE,2017:2242– 2251.

[75] P Isola, J Zhu, T Zhou, et al. Image-to-image translation with conditional adversarial networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR),Honolulu,HI,USA,2017:1125-1134.

[76] Lu Y, Wu S, Tai T, et al. Image Generation from Sketch Constraint Using Contextual GAN .Proceedings of the European Conference on Computer Vision (ECCV), 2018, 205-220.

[77] Mao X, Li Q, Xie H,et al. Least squares generative adversarial networks [C]. Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017: 2794-2802.

[78] Zhu J, Zhang R, Pathak D.et al.Toward multimodal image-to-image translation[J].ariXiv,2017.

[79] B.D Brabandere, D. Neven L. Gool. Segmentation with a Discriminative Loss Function [EB/OL].[2017].http://ArXiv.org/abs/1708.02551

致谢

时如白驹过隙,不知不觉中已经来到硕士生涯的最后一年。在这三年既漫长又短暂的研究生生活中，自己受益多，其间所取得的每一点进步都离不开老师的谆谆教诲，同学、朋友的热心帮助和亲人的默默支持，谢谢你们!

本文是在导师刘国英教授精心指导下完成的，在论文的选题、试验方案的确定、理论分析、数据处理直至论文的撰写和定稿，刘老师给予我悉心的教诲和无私的帮助，论文完成的整个过程中渗透着他们的心血和汗水。

在两年的研究生生涯里，在刘老师的教导下，我的理论知识得以升华，懂得了怎样将自己的知识应用于实际工程，从实际出发重新认识以前学过的知识。刘老师对学生的学习、生活和为人处事等各个方面都给予了无微不至的关心与指导。从；刘老师的身上所学到的不仅是专业技能和知识，还有对待困难的从容和对待工作的执着。刘老师开阔的学术思维、严谨的治学态度和勤奋踏实的工作作风，将成为我终生学习的榜样。

论文的撰写即将结束，学生生涯也即将完结，在此特向刘老师表示深深的谢意!其次，谢谢我身边的同龄伙伴!亲人，尤其是勤劳的父母，二十多年来毫无怨言、默默地奉献和着意地培养，才使我走到今天。其实这点点滴滴的成绩都是你们心血与汗水的凝结，无论走到哪里，您们的教诲将永远铭刻在我心。

个人简历、在校期间发表的学术论文与研究成果

个人简介

陈双浩，男，1993 年 8 月生，河南周口人。2014 年 9 月 2018 年 6 月就读安阳师范学院计算机与信息工程学院物联网专业；2018 年 9 月至今就读于郑州大学信息工程学院计算机技术专业。

已发表的学术论文

[1] 一种甲骨拓片图像甲骨字符提取网络[J]. 厦门大学学报自然科学版（已录用）

[2] An Oracle Bone Inscription Detector Based on Multi-Scale Gaussian Kernels[J]. Applied Mathematics,2021, 12:224-2312

研究成果

[1] 一种割合分割网络和生成网络的甲骨拓片图像字符提取方法(实质审查中，专利号：20210300152.3)

[2] 基于生成对抗网络的甲骨拓片文本提取系统 v1.(软件著作）

科研项目

[1] 自然科学基金项目：基于深度学习的甲骨文字检测与识别研究(项目编号：U1804153)