



南京大學

本科畢業論文

院 系 電子科學與工程學院

專 業 集成電路設計與集成系統

題 目 面向自動駕駛的邊緣視覺任務研究

年 級 2017 級 學 號 171180566

學生姓名 馬晨昊

指導老師 馬展 職 稱 教授

提交日期 2021 年 5 月 20 日

南京大学本科生毕业论文(设计、作品)中文摘要

题目：面向自动驾驶的边缘视觉任务研究

院系：电子科学与工程学院

专业：集成电路设计与集成系统

本科生姓名：马晨昊

指导老师（姓名、职称）：马展教授

摘要：

计算机视觉作为人工智能发展不可或缺的一大领域，是一门专注研究如何使机器拥有类似于人眼功能的科学。研究者们用摄影机和电脑代替人眼对目标进行识别、跟踪和测量等任务，并进一步做图形处理，使电脑处理成为更适合人眼观察或传送给仪器检测的图像。

作为一个科学学科，计算机视觉研究相关的理论和技术，试图建立能够从图像或者多维数据中获取“信息”的人工智能系统。

本文专注于研究自动驾驶中的视觉任务，通过对相机传感器拍摄的原始图像（RAW）进行视觉分析和优化处理，以此来最大化的发挥机器视觉在自动驾驶领域的作用，推动自动驾驶更快更好地发展，为人们服务。

关键词： 自动驾驶；模拟 RAW 图；图像处理；计算机视觉

南京大学本科生毕业论文(设计、作品)英文摘要

THESIS: Thesis paper template

DEPARTMENT: School of Electronic Science and Engineering

SPECIALIZATION: Integrated Circuit Design

UNDERGRADUATE: Ma Chenhao

MENTOR: Professor Ma Zhan

ABSTRACT:

Nowadays computer vision(CV) is a hot issue in the artificial intelligence(AI) world. It is a major which focuses on how to make machines see just like human beings. Researchers use cameras and computers instead of human eyes to do tasks like segmentation, classification and so on for future CV use.

As a scientific study, computer vision tries to set up the way between image or multi-dimensional data and artificial intelligence.

The main contents of this paper is about the computer vision work used in autonomous driving. By analyzing the RAW images simulated from powerful simulating softwares in vision aspect, the work tries to optimize the process of image processing in autonomous driving.

This paper will show how well RAW image will play in image processing and how well computer vision can do for autonomous driving and for everyone.

KEY WORDS: Autonomous Driving, Simulated RAW Image, Image Processing, Computer Vision

目 录

1	绪论	1
	1.1 研究背景	1
	1.2 相关工作	1
	1.3 本文主要工作	1
	1.4 本文结构	2
2	系统	3
	2.1 图片	3
	2.2 公式	4
	2.3 表格	5
	2.4 算法	5
3	实验	7
	3.1 实现细节	7
	3.2 文本分行结果	7
	3.3 识别结果	7
	3.3.1 准确率	7
4	总结与讨论	9
致	谢.....	11

第一章 绪论

1.1 研究背景

在过去的 40 年中，手写中文文本识别（HCTR）的研究获得了很大的进展，效果得到了很大的提升^[2]。但是，由于手写中文文本的多样性，它依然是一个具有研究意义和挑战性的问题^[2]。不同的文本有不同的书写风格，如图 1-1。

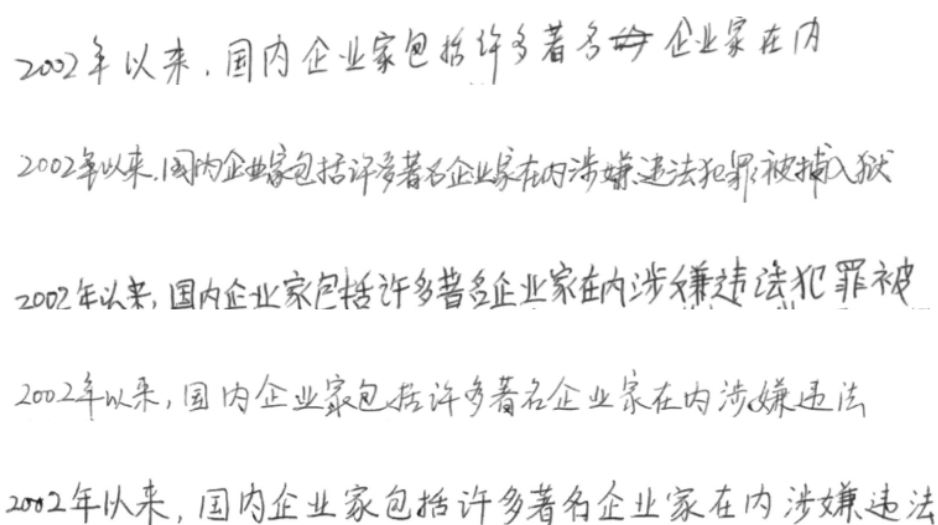


图 1-1: 不同的书写风格。对于同一句话，有不同的书写风格：倾斜，写错字，工整，潦草等。

1.2 相关工作

作为深度神经网络中处理序列的一个重要模型，循环神经网络，如图，在训练和测试过程中不需要知道视觉序列对象中每个元素的位置。但是，对于循环神经网络，非常重要的一点是将输入图片通过图片预处理转化为一串图片特征^[2,3]。但是通常的基于循环卷积神经网络的网络，因为预处理不在系统训练流程之内，所以无法用从头到尾的方式进行训练，不是很方便。

1.3 本文主要工作

本文旨在对图片级手写中文文本做出识别分类。主要工作如下：

1. 在标准公开数据集里获得了一定的识别准确度。
2. 在标准公开数据集上击败了一些相关工作的结果。

3. 建立分本分行规划表格，很好地处理了文本的分行，降低了训练开销。

1.4 本文结构

本文的各章节组织结构如下：

第一章：绪论。简要说明了手写中文文本识别的研究来由及相关工作。并概括地描述了这篇文章的工作，总结了本文结构。

第二章：识别系统

第三章：实验。介绍了实验进行的配置环境，文中使用的测度，文本分行的结果，识别的结果并分析了得到这种结果背后的缘由。

第四章：总结与讨论。总结全文工作，讨论存在的问题和今后可以继续研究的方向。

第二章 系统

如何使用这个模板

2.1 图片

一行一图，如图 2-1

2002年以来，国内企业也频频成为企业家因涉嫌违法犯罪被
狱的话题不断涌现，从方丽的报道也是屡见不鲜，不是明升暗降
哪升被判了，或者是这个案子开庭了，那个案子判决了。总之，几乎 100% 都有
将的所谓。
企业家落马、判刑、入狱，甚至死无葬身之地，媒体关注的
往往不是法律问题，而更多的企业家经营和管理上的问题，在媒体上
是各种各样的，不是经济学家、管理专家，而是从法律定论来写讨论，这
中不正之风的现象。企业家不管在经营、管理上有什么问题，最终的
判决是在法律框架内，最终的判决如果是经由法院判决有罪，那
重要的应该是法律问题！

图 2-1: 待分行文本

一行两个图

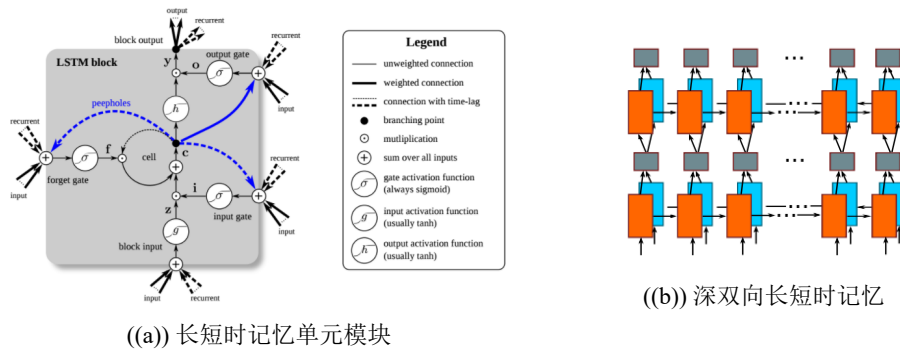


图 2-2: (a) 一个长短时记忆单元模块。(b) 深度双向长短时记忆的结构。

多行多图

递归公式

$$\beta_t(s) = \begin{cases} (\beta_{t+1}(s)d(s) + \beta_{t+1}(s+1))d(s+1)y_{l'_s}^t, & \text{if } l'_s = b \text{ or } l_{s+2}' = l'_s \\ (\beta_{t+1}(s)d(s) + \beta_{t+1}(s+1)d(s+1) + \beta_{t+1}(s+2)d(s+2))y_{l'_s}^t, & \text{otherwise} \end{cases} \quad (2-4)$$

2.3 表格

		国	内	企	业	包	括	许	多
	0	1	2	3	4	5	6	7	8
国	1	0	1	2	3	4	5	6	7
著	2	1	1	2	2	3	4	5	6

表 2-1: 编辑距离（乐文斯汀距离计算过程示例表格。字符串“国内企业包括许多”与“国著名括许多”乐文斯汀距离是 3。

2.4 算法

算法 2.1 Beam Search

- 1: 将初始节点插入到集束中。
- 2: **while** 遍历未结束 **do**
- 3: 遍历集束中所有节点的后续节点。
- 4: **if** 该节点是目标节点 **then**
- 5: 算法结束。
- 6: **else**
- 7: 扩展该节点，取集束宽度的节点入堆。
- 8: **end if**
- 9: **end while**

集束宽度可以在搜索过程中保持为一个定值，也可以根据搜索的进行而变化。搜索算法可以根据搜索的结果进行调整，比如，当以一个小的集束宽度搜索解却无法找到适合解的时候，可以增大集束宽度重新进行一次搜索。

第三章 实验

3.1 实现细节

我们在 Tensorflow 框架上实现了我们的网络系统。实验在一个搭载 2.40GHz 英特尔志强 Xeon E5-2673 CPU, 32GB RAM 和一块英伟达 1080Ti 12GB 显存的服务器电脑上运行。网络系统使用 Adam 训练算法。

3.2 文本分行结果

尽管如此,在局部损失切割和局部水平投影切割之后,每一个竖直段的分行结果的对应关系却很难处理。在一些特殊情况下,无法做到每一竖直段分行关系的对应。所以这两个方法不适用。

3.3 识别结果

3.3.1 准确率

我们根据数据集中人的笔迹将数据集分为了 HWDB1-HWDB3,并实现了 Wang 等人^[2]和 Mishra 等人^[2]的方法,通过调用百度的文字识别系统^[2],进行对比实验得到以下结果。

方法	HWDB1	HWDB2	HWDB3
Wang 等人 ^[2]	74.0	60.0	68.0
Mishra 等人 ^[2]	80.8	63.6	73.5
百度通用文字识别 ^[2]	64.8	36.8	60.8
我们的方法(没有字典信息)	81.5	67.5	73.6
我们的方法	81.8	67.8	73.9

表 3-1: 识别准确率

3.3.1.1 测试

1234

第四章 总结与讨论

在本文中，我们使用预处理层-卷积层-循环卷积层-转录层网络来处理手写中文文本识别的问题。这种网络很好地结合了卷积网络和循环网络各自的优势。

致 谢

感谢在实验室度过的两年时光，老师无论在学术还是人生的指导上都对我起到了很大的帮助；师兄师姐小伙伴们的鼓励支持和陪伴是我坚持下去的动力。