



# 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：

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。

本文探讨了使用原始的传感器数据（即 RAW 图像）直接执行高级视觉任务的可行性。通过这种方法，我们可以避开数十年来在相机拍摄于图像处理中涉及到的图像信号处理过程（ISP）。

由于缺乏用以训练 RAW 域视觉模型的大规模 RAW 图像数据集，我们开发了一种生成对抗网络（GAN），可以将大量的压缩成 JPEG 格式的 RGB 图像转换为对应的 RAW 域图像以此获得大规模 RAW 图像数据集。我们还收集并标记了专用的小规模 RAW 图像数据集用于高级检测和分割任务。

经实验证明，与使用 RGB 图像进行视觉处理相比，RAW 域的图像处理在对象检测和语义分割方面具有明显优势。另外一个优点是，使用 RAW 图像执行视觉任务消除了对图像信号处理过程（ISP）的需求，优化了功耗和时延方面的问题。

我们希望通过相机传感器拍摄的原始图像（RAW）进行视觉分析和优化处理，来最大化地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；RAW 图像数据集；图像处理；计算机视觉；人工智能

# 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:

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

This paper mainly discusses the way and possibility to directly execute high-level vision tasks using the raw sensor data. By performing CV tasks in this way, we can completely ignore the Image Signal Processor (ISP) used in digital cameras recently.

Due to the lack of RAW image datasets for training robust RAW-domain vision models, our team have developed a generative adversarial network (GAN) to transform existing abundant JPEG compressed RGB images to RAW-domain images. We have also collected and labelled a small-scale RAW-domain image dataset for high-level object detection and segmentation tasks.

Our experiments showed the fact that RAW-domain images have competitive efficiency in object detection and semantic segmentation over the conventional RGB images. Meanwhile, performing vision tasks using RAW images removes the latency and power caused by ISP.

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, RAW Image Dataset, Image Processing, Computer Vision, Artificial Intelligence

# 目 录

<b>1</b>	<b>绪论</b> .....	<b>1</b>
1.1	研究背景 .....	1
1.2	本文工作和贡献 .....	1
1.3	本文结构 .....	1
<b>2</b>	<b>相关工作</b> .....	<b>3</b>
2.1	传统 ISP.....	3
2.2	RAW 图像处理 .....	4
<b>3</b>	<b>研究方法</b> .....	<b>7</b>
3.1	问题公式化 .....	7
3.2	通过机器学习实现的 invISP .....	7
3.3	RAW 域视觉任务 .....	7
3.3.1	准确率 .....	7
3.4	损失函数 Loss Function .....	7
<b>4</b>	<b>实验</b> .....	<b>9</b>
4.1	数据集 .....	9
4.1.1	iPhone RAW Scape 1K .....	9
4.1.2	BDD.....	9
4.2	训练细节 .....	9
4.2.1	基于学习的 invISP 网络 .....	9
4.2.2	物体探测网络 .....	9
4.2.3	实例分割网络 .....	9
4.3	结果 .....	9
4.3.1	基于学习的 invISP .....	9
4.3.2	目标检测结果 .....	9
4.3.3	实例分割结果 .....	9
<b>5</b>	<b>总结与讨论</b> .....	<b>11</b>
	参考文献.....	<b>13</b>
	致 谢.....	<b>15</b>



## 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<sup>[2]</sup>。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<sup>[2]</sup>。不同的文本有不同的书写风格，如图 1-1。

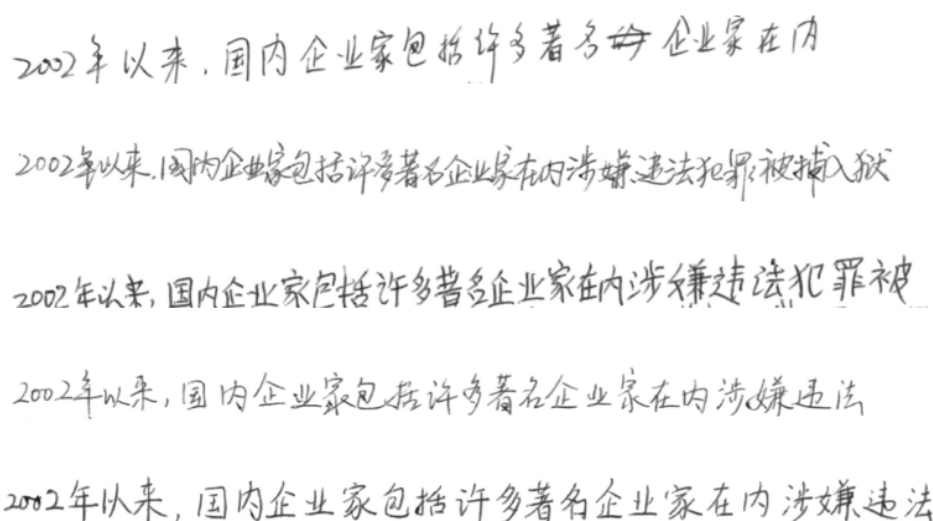


图 1-1: 不同的书写风格。对于同一句话，有不同的书写风格：倾斜，写错字，工整，潦草等。

### 1.2 本文工作和贡献

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<sup>[2,3]</sup>。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 本文结构

本文的各章节组织结构如下：

第一章：绪论。简要说明计算机视觉领域 RAW 域与 RGB 域图像研究背景并概括地描述了这篇文章的工作，总结了本文结构。

第二章：相关工作。概括介绍了传统 ISP 和 RAW 图像处理的相关工作。

第三章：研究方法。详细介绍本课题进行研究所使用的方法和原理。

第四章：实验。详细介绍本课题在 RAW 域进行视觉任务研究的使用工具和实验过程。

第五章：总结和讨论。



## 第二章 相关工作

如何使用这个模板

### 2.1 传统 ISP

一行一图，如图 2-1

2002年以来，国内企业信用评级者为企业家内幕舞弊犯罪被  
狱问题不断报道，此方面的报道也是屡见不鲜，不是明升暗降  
哪？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  
内幕新闻。  
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  
判决是正在监狱，最后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

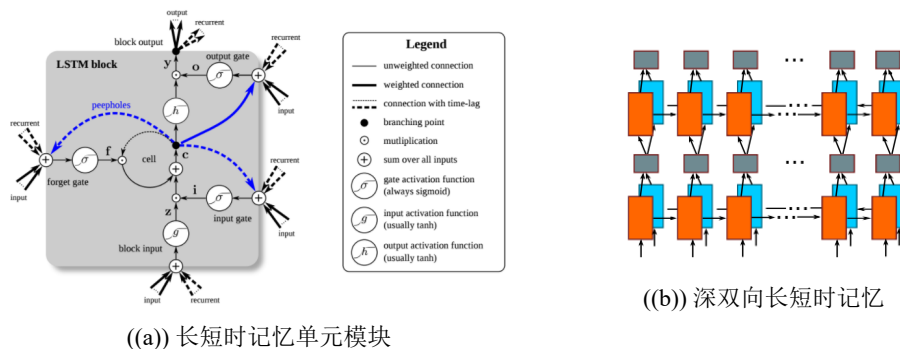


图 2-2: (a) 一个长短时记忆单元模块。(b) 深度双向长短时记忆的结构。

多行多图

2002年以来,国内企业包括许多著名企业家在内涉嫌违法犯罪被  
 狱的人数不断增多,这方面的报道也是屡见不鲜,不过,明什么  
 哪什么被判了,或者是个案子开庭了,那个案子判定了。总之,几乎 明部有  
 样的新闻。  
 企业家落马、判刑、入狱,甚至犯罪被枪毙,明什么,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  
 表各种意见的,不是经济学家、管理学家,而是有法律背景来参与讨论,这  
 中不正常的现象。企业家不是在经营、管理上有什么问题,而是  
 为早是在进监狱,最后的结果如果是由法院判决有罪,那  
 重要的应该是法律问题!

## ((a)) 全局损失切割

2002年以来,国内企业包括许多著名企业家在内涉嫌违法犯罪被  
 狱的人数不断增多,这方面的报道也是屡见不鲜,不过,明什么  
 哪什么被判了,或者是个案子开庭了,那个案子判定了。总之,几乎 明部有  
 样的新闻。  
 企业家落马、判刑、入狱,甚至犯罪被枪毙,明什么,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  
 表各种意见的,不是经济学家、管理学家,而是有法律背景来参与讨论,这  
 中不正常的现象。企业家不是在经营、管理上有什么问题,而是  
 为早是在进监狱,最后的结果如果是由法院判决有罪,那  
 重要的应该是法律问题!

## ((b)) 局部损失切割

图 2-3: 分行结果比较。(a) 全局损失切割; (b) 局部损失切割; (c) 局部水平投影切割; (d) 投影损失切割

## 2.2 RAW 图像处理

$$\frac{\partial L}{\partial \alpha_k^t} = d(s)^2 (y_k^t - \frac{\sum_{lab(l,k)} \alpha_l(s) \beta_l(s)}{y_k^t}) \quad (2-1)$$

$$d_{0j} = \sum_{k=1}^j w_{ins}(a_k), \quad \text{for } 1 \leq j \leq n$$

$$d_{ij} = \begin{cases} d_{i-1,j-1} & \text{for } a_j = b_i \\ \min \begin{cases} d_{i-1,j} + w_{del}(b_i) \\ d_{i,j-1} + w_{ins}(a_j) \\ d_{i-1,j-1} + w_{sub}(a_j, b_i) \end{cases} & \text{for } a_j \neq b_i \end{cases} \quad \text{for } 1 \leq i \leq m, 1 \leq j \leq n. \quad (2-2)$$

$$\begin{aligned} \beta_T(|l'|) &= y_b^T \\ \beta_T(|l'| - 1) &= y_{l|l}^T \\ \beta_T(s) &= 0, \forall s < |l'| - 1 \end{aligned} \quad (2-3)$$

递归公式

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



## 第三章 研究方法

### 3.1 问题公式化

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 通过机器学习实现的 invISP

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 RAW 域视觉任务

#### 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<sup>[2]</sup> 和 Mishra 等人<sup>[2]</sup> 的方法, 通过调用百度的文字识别系统<sup>[2]</sup>, 进行对比实验得到以下结果。

方法	HWDB1	HWDB2	HWDB3
Wang 等人 <sup>[2]</sup>	74.0	60.0	68.0
Mishra 等人 <sup>[2]</sup>	80.8	63.6	73.5
百度通用文字识别 <sup>[2]</sup>	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	<b>81.8</b>	<b>67.8</b>	<b>73.9</b>

表 3-1: 识别准确率

#### 3.3.1.1 测试

1234

### 3.4 损失函数 Loss Function

test



## 第四章 实验

### 4.1 数据集

#### 4.1.1 iPhone RAW Scape 1K

#### 4.1.2 BDD

### 4.2 训练细节

#### 4.2.1 基于学习的 invISP 网络

#### 4.2.2 物体探测网络

#### 4.2.3 实例分割网络

### 4.3 结果

#### 4.3.1 基于学习的 invISP

#### 4.3.2 目标检测结果

#### 4.3.3 实例分割结果





## 第五章 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。



## 参考文献

- [1] 郭莉莉, 白国君, 尹泽成, 魏惠芳. “互联网+”背景下沈阳智慧交通系统发展对策建议 [A]. 中共沈阳市委、沈阳市人民政府. 第十七届沈阳科学学术年会论文集 [C]. 中共沈阳市委、沈阳市人民政府: 沈阳市科学技术协会, 2020: 4.
- [2] 陈香敏, 魏伟, 吴莹. “文化+人工智能”视阈下文化创意产业融合发展实践及路径研究 [A]. 中共沈阳市委、沈阳市人民政府. 第十七届沈阳科学学术年会论文集 [C]. 中共沈阳市委、沈阳市人民政府: 沈阳市科学技术协会, 2020: 4.
- [3] 田晓曦, 刘振鹏, 彭宝权. 地方高校开展教育人工智能深度融合的路径探究 [A]. 中共沈阳市委、沈阳市人民政府. 第十七届沈阳科学学术年会论文集 [C]. 中共沈阳市委、沈阳市人民政府: 沈阳市科学技术协会, 2020: 5.
- [4] 柏卓君, 潘勇, 李仲余. 彩色多普勒超声在早期胚胎停育诊断中的应用 [J]. 影像研究与医学应用, 2020, 4(18): 129-131.
- [5] 杨芸. 我院 2018 年人血白蛋白临床应用调查与分析 [J]. 上海医药, 2020, 41(17): 34-35+74.



## 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。