学号：雷伯涵

密级：

武汉大学本科毕业论文

一种图像-语义互译系统

院(系)名 称 ： 弘毅学堂

专 业 名 称 ： 计算机科学与技术

学 生 姓 名 ： 雷伯涵

指 导 教 师 ： 庄越挺 教 授

黄 浩 副教授

二〇二〇年五月

郑 重 声 明

本人呈交的学位论文，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、图片资料真实可靠。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确的方式标明。本学位论文的知识产权归属于培养单位。

本人签名: _____ 日期: _____

摘　　要

请使用中文分号“;”分割关键词!

摘要内容摘要内容摘要内容摘要内容摘要内容摘要内容摘要内容摘要内容摘要内容摘要内容摘要内容摘要内容摘要内容摘要内容摘要内容摘要内容

摘要内容

关键词：关键词 1；关键词 2；关键词 3

ABSTRACT

Please use English semicolon and space to separate key words.

This is abstract. This is abstract. This is abstract. This is abstract. This is abstract. This is abstract. This is abstract. This is abstract.

This is abstract. This is abstract. This is abstract. This is abstract. This is abstract. This is abstract. This is abstract. This is abstract. This is abstract. This is abstract.

Key words: Key1; Key2; Key3

目　　录

1 绪论

1.1 相关背景与需求

随着社会发展与技术进步，我国目前越来越重视高效工具的研发。从 2000 年伊始到 2010 年互联网普及，再到 2020 年的现在，个人电子终端的功能已经越来越强大，每个人都需要更新期、更有用的效率工具。近五年，图片标注技术与生成模型都有爆炸式的发展；近两年，视觉问答技术 (Visual Question Answering)[1]、视频标注技术与秒级的图片理解更是让失明群体有了“读懂光芒”的希望，也让图片的理解从学术界或产业界的科研层面有了走向应用、走向市场的可能。

对于看不见的人来说，读懂视野这一技术是他们改善生活质量的工具，画出语言这一技术是他们表达自我的窗口；对更多的普通人群体来说，一个简洁易用的新奇效率工具，更是一种生活方式的改变，让更多的人通过机器的“魔力”，换一种角度来看语言和图片。

1.2 技术应用意义与前景

1.2.1 研究意义

图像-语义的双向翻译在目前已经有了很大的需求面。

最基本的就是，图像作为视觉信号，无法被视觉失能人群感知，可以将语义转化为听觉信号，方便视觉失能人群感知世界。更进一步，对于大量的图像信号，靠肉眼处理起来需要很大的人力成本，如果使用图像标注技术，则可以从图像中提取重要信息，对语义信息进一步处理，形成简报构成参考，辅助决策。

语义的图像化更是意义非凡。基础地说，同样面对失能人群，没有视觉的人通过这一系统也有了“创作”的能力，将他们的思想从肉体的局限中有限地拓宽了出口；或者说，即使是有视觉能力的人，如果不擅长绘画，也可以通过这一技术直接表达自己思想中的画面。更进一步地说，想象力弱是很多成年人能力的局限，而一个语义向图像的转化，则可以使得一个人的表述清晰、直观地呈现在他人的视野里，可以促进交流；而表述之人也可以根据可视化的表达，发现自己表达中的问题，及时予以修正，而这是我们生活中每个人都需要的。

可以说，生成图像不仅仅是作为观赏，它可以切真实际地改变我们的生活方式。

1.2.2 研究前景

现在并没有简单易用的商业系统，可以提供语义与图像互译的简便功能，大部分系统都只能作为技术的样品，做单向的翻译工作，目前主要在各大展览会上起到展示企业技术实力之用。对于常常需要与图片、交流打交道的人来说，这样一套系统对工作效率提高很多；广泛地，对于任何一个人，这类系统都可以改变他的生活方式。

1.3 相关工作

1.3.1 循环神经网络

1.3.1.1 神经网络简介

神经网络 (Neural Networks) 是近年

1.3.1.2 循环神经网络

循环神经网络 (Recurrent Neural Networks, RNN) 被广泛地运用在图像处理和自然语言处理上。与一般的神经网络相同

1.3.1.3 长短期记忆

长短期记忆 (Long Short Term Memory, LSTM) 作为循环神经网络的一个分支，由在首先提出。长短期记忆的基本原理

1.3.2 对抗生成网络

1.3.2.1 背景介绍

对抗生成网络 (Generative Adversarial Networks, GAN) 由 Goodfellow et al. 在 2014 年首次提出[2]。目前，它已经发展成了生成神经网络最大的热点，其研究得到了长足的发展。短短几年之内，已经有了一百余种 GAN 网络的衍生模型，其应用范围囊括了包括自然语言、图像处理、计算机视觉在内的各个领域。

在提出 GANs 模型之前，也有其他的生成式模型存在。生成方法、判别方法分别时机器学习的两大分支方向，而生成式模型则是用生成方法来生成样本的一类模型。有一类生成式模型是从人类理解角度进行设计的，比如说最大似然估计法、近似法[3, 4] 与马尔可夫链法[5–7] 等，这一类方法对于机器来说各有限制。最大似然估计法的参数更新直接受限于数据样本，数据样本不够丰富会限制生成模型的结

果；近似法的目标函数太过复杂，算法只能逼近目标函数下界；马尔可夫链法的缺点便是复杂度过高。从机器理解角度设计的算法一般不直接进行拟合或者估计，而是通过采样数据样本调整模型，一般这种方法人类无法直接理解，但是生成样本是人类可以理解的。

1.3.2.2 基本原理

GANs 的基本结构即由一个生成生成模型 G 和一个判别模型 D 组成。生成模型 G 的目的是尽可最小化生成与真实训练样和生成样本的区别，而判别模型 D 的目的则是尽可能最大化地找出真实样本和生成样本之间的区别。

通过多轮迭代生成模型 G 和判别模型 D 的对抗，可以使两个模型都达到上述的目标效果。当训练判别模型 D 的时候，希望输入真实样本 x 可以使判别器对其的判断 $D(x)$ 尽量趋于 1，而生成样本 $G(x)$ 通过判别器 D 的时候可以使得 $D(G(x))$ 尽量趋于 0。在训练生成模型 G 的时候，输入噪声 z，希望生成的生成样本通过判别器 D 的时候尽量使得 $D(G(z))$ 趋于 0。

可以用简单的数学变换得到公式（1.1），来描述训练过程。

$$\min_G \max_D V_{G,D} = \mathbb{E}_{x\sim P_{data}(x)}[\lg D(x)] + \mathbb{E}_{z\sim P_G(z)}[\lg(1 - D(G(z)))] \tag{1.1}$$

其中 $P_{data}(x)$ 为真实图片集的分布。

当多轮博弈过后，极大极小问题达到最优解，即纳什均衡，当且仅当 $P_z = P_{data}$ 时[2]。这时 $\mathbb{E}D(G(z))$ 趋于 $\frac{1}{2}$，即相当于只能随机猜测 0 与 1，而生成模型 G 学会了真实样本的特性。

相对于传统的生成模型，我们可以发现 GANs 模型并不需要使用马尔可夫链，学习过程不需要近似推理，也不需要预先训练，自由度比较高，可以利用反向传播计算梯度，很好地利用了分段线性单元的优势，而可以回避近似计算的困难概率问题。同时，GANs 模型的缺点也相对应地明显。由于 GANs 模型自由度太高，在面对过于清晰的图片等训练样本时，收敛性表现较差，生成模型可能出现退化，重复生成相同样本点，导致判别器无法工作，进而导致模型崩溃。因此，在训练过程中，调整好两个模型网络的平衡与同步非常重要。

1.3.3 自然语言处理

1.3.3.1 背景介绍

自然语言处理 (Natural Language Processing, NLP) 是指计算机对自然语言的处理算法。

所谓自然语言，就是说人类在社会文明发展过程中为了交流而逐渐发展出的语言，例如中文、英文、日文，甚至包括手语，都是自然语言。

1.3.3.2 成熟自然语言处理平台 Watson

在 2011 年中，美国的一档综艺节目中，IBM 公司云服务器 bluemix 平台上的 Watson 板块 NLP 技术相关产品大放异彩，当时因出色的对话表现爆红。在 Watson 平台上，有多种语言、多种应用的 NLP 相关产品，包括语音识别、语音生成等，经过实际试用，发现目前制作水平比较完善，但是中文包的翻译效果比较差。经过调查与采访，我发现中文的翻译训练语料库主要是使用专业文件书籍构成，导致语言不够丰富。这也给了我们一个教训，在实际 NLP 训练中，一定要使用比较丰富、贴近日常生活的语料库。

在我的研究中，我以英语为基础，使用了完整、丰富的 MSCOCO 数据集。在这个数据集中，每一年份的数据都有八万多张图片作为训练集，另有四万多张图片作为测试集，且均带有相应的语义标注。这样完整的数据集，比企业为了宣传与展示技术实力所做的基本免费样本更为贴近一般语言环境，可以得到更好的效果。

1.3.3.3 基础依赖包

目前，常用的 NLP 算法在 python、JAVA 等语言中里面都有依赖包可以使用。

1.4 本文工作目标

1.4.1 设计大纲构思

本次设计需要实现如下功能，并达到要求：

1. 实现图像生成语义算法，并达到可用的效果；
2. 实现语义生成图像算法，并且图像对一般人清晰可识别；
3. 形成一个有机统一的系统，功能简洁易用。

1.4.2 本文篇章结构

本文共分为五个章节。

第一章，介绍了研究的背景，并通过阅读文献，整理了相关的工作，从中找出比较适合于本设计的工作，并加以总结。

第二章，详细介绍了本设计所使用技术的详情，并说明了选取技术原因。

第三章，设计了实验的方案，通过一些前人制作好的开源项目，加以修改、总结，设计系统。

第四章，详细记述了实验的结果，并通过测评手段进行测评，分析了实验结果。

第五章，阐述了设计的意义，并且展望了未来系统进一步扩展迭代的方向，指明了系统的局限性。

1.5 本章小结

本章介绍了设计相关领域的需求背景，也全面地介绍了这一设计的意义与前景。接下来介绍了涉及到的相关技术发展现状和相关领域的工作，也说明了其中最适合于本设计的技术及其原因。最后说明了本设计的基本要求和构思，另外设计了篇章结构，也设定了工作的目标。

2 相关技术理论

2.1 GAN 生成模型选取使用

2.1.1 衍生模型分类与特点

目前对抗生成网络的衍生模型众多，其优化方式大抵由两个大方向衍生而来。一个方向是由损失函数的改变来优化 GAN 模型的效果，另一个方向则是从模型使用的角度来优化 GAN 模型的效果。在表 2.1中，列举了一些最为常见的 GAN 模型衍生模型。

表 2.1

从损失函数角度提出的优化 GAN 模型[8]	从模型应用角度提出的优化 GAN 模型		
	网络构架角度[9]	编码器角度	其他角度改进
Least Square GANs, Loss-Sensitive GAN, Fisher GAN, WGAN, WGAN-GP, WGAN-LP, f-GANs[] DRAGAN 等	CGAN, DCGAN, InfoGAN, StackGAN, AL-GAN 等	BEGAN, VAE-GAN, tDCGAN, BiGAN, 文献中的算法[10–12] 等	LAPGAN, ESRGAN, SRGAN, 3D-GAN, MGAN 等

图像生成的模型与基于网络构架优化的 GAN 网络模型最为贴合，本文设计使用 StackGAN 作为基础，并针对相关实现进行优化。

2.1.2 StackGAN 的特性

2.2 神经网络技术

2.2.1 循环神经网络

3 方案设计

3.1 概要

3.2 系统模型

3.3 变量符号与定义

3.4 软件设计

3.4.1 图片标注方法

3.4.2 自然语言生成图片方法

3.5 本章小结

4 实验与分析

4.1 图片标注实验

4.2 自然语言生成图片实验

4.3 实验分析与总结

4.4 本章小结

5　设计意义阐述与展望

5.1　我的设计意义

5.2　制作了一个简洁易用系统

5.3　活用知识、实践了深度学习技术

5.4

6 公式插图表格

6.1 公式的使用

在文中引用公式可以这么写：$a^2 + b^2 = c^2$ 这是勾股定理，他还可以表示为 $c = \sqrt{a^2 + b^2}$，还可以让公式单独一段并且加上编号。注意，公式前请不要空行。

$$\sin^2\theta + \cos^2\theta = 1 \tag{6.1}$$

还可以通过添加标签在正文中引用公式，如式（6.1）。

我们还可以轻松打出一个漂亮的矩阵：

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 11 & 22 & 33 & 44 \end{bmatrix} \times \begin{bmatrix} 22 & 24 \\ 32 & 34 \\ 42 & 44 \\ 52 & 54 \end{bmatrix} \tag{6.2}$$

或者多行对齐的公式：

$$\begin{aligned} f_1(x) &= (x+y)^2 \\ &= x^2 + 2xy + y^2 \end{aligned} \tag{6.3}$$

6.2 插图的使用

LaTeX 环境下可以使用常见的图片格式：JPEG、PNG、PDF、EPS 等。当然也可以使用 LaTeX 直接绘制矢量图形，可以参考 pgf/tikz 等包中的相关内容。需要注意的是，无论采用什么方式绘制图形，首先考虑的是图片的清晰程度以及图片的可理解性，过于不清晰的图片将可能会浪费很多时间。

图示例如下：

图 6.1　插图示例

[htbp] 选项分别是此处、页顶、页底、独立一页。[width=\textwidth] 让图片占

满整行，或[width=2cm] 直接设置宽度。可以随时在文中进行引用，如图 6.1，建议缩放时保持图像的宽高比不变。

6.3　表格的使用

表格的输入可能会比较麻烦，可以使用在线的工具，如 Tables Generator 能便捷的创建表格，也可以使用离线的工具，如 Excel2LaTeX 支持从 Excel 表格转换成 LaTeX 表格。LaTeX/Tables 上及 Tables in LaTeX 也有更多的示例能够参考。

6.3.1　普通表格

下面是一些普通表格的示例：

表 6.1　简单表格

我是	一只	普通
的	表格	呀

表 6.2　一般三线表

姓名	学号	性别
张三	001	男
李四	002	女

6.3.2　跨页表格

跨页表格常用于附录（把正文懒得放下的实验数据统统放在附录的表中），以下是一个跨页表格的示例：

表 6.3　跨页表格示例

1	0	5	1	2	3	4	5	6
1	0	5	1	2	3	4	5	6
1	0	5	1	2	3	4	5	6
1	0	5	1	2	3	4	5	6
1	0	5	1	2	3	4	5	6
1	0	5	1	2	3	4	5	6
1	0	5	1	2	3	4	5	6
1	0	5	1	2	3	4	5	6
1	0	5	1	2	3	4	5	6
1	0	5	1	2	3	4	5	6

转下一页

接上一页

1	0	5	1	2	3	4	5	6
1	0	5	1	2	3	4	5	6
1	0	5	1	2	3	4	5	6
1	0	5	1	2	3	4	5	6

6.3.3 统计表格

要创建占满整个文字宽度的表格需要使用到 tabularx，如不需要，使用 tabular 就行。引用表格与其它引用一样，只需要：表 6.4，统计表格一般是三线表形式。

表 6.4 统计数据表格

序号	年龄	身高	体重
1	14	156	42
2	16	158	45
3	14	162	48
4	15	163	50
平均	15	159.75	46.25

6.4 列表的使用

下面演示了创建有序及无序列表，如需其它样式，LaTeX Lists 上有更多的示例。

6.4.1 有序列表

这是一个计数的列表

1. 第一项
   (a) 第一项中的第一项
   (b) 第一项中的第二项
2. 第二项
   (i) 第一项中的第一项
   (ii) 第一项中的第二项
3. 第三项

6.4.2 不计数列表

这是一个不计数的列表

- 第一项
  - 第一项中的第一项
  - 第一项中的第二项
- 第二项
- 第三项

7 其它格式

7.1 代码

7.1.1 原始代码

朴实的代码块：

使用 verbatim 可以得到原样的输出，如下：

```
print("Hello world!")
```

使用listings环境可以对代码进行进一步的格式化，如下：

```
import numpy as np

a = np.zeros((2,2))
print(a)
```

7.1.2 代码高亮

还可以对代码进行高亮，请参考 Code Highlighting with minted。请先到 cls 文件中启用 minted 库。注意使用 Minted 库时，需要系统默认 Python 有 Pygments 库，可以通过$ pip install Pygments 来进行安装。且需要在编译时加上--shell-escape 参数，否则会报错。

7.1.3 算法描述/伪代码

参考 Algorithms，下面是一个简单的示例：

Result: Write here the result

initialization;

while *While condition* **do**

    instructions;

    if *condition* **then**

        instructions1;

    else

        instructions3;

    end

end

算法 1: How to write algorithms

7.2 绘图

关于使用 LaTeX 绘图的更多例子，请参考 Pgfplots package 中的例子。一般建议使用如 Photoshop、PowerPoint 等制图，再转换成 PDF 等格式插入。

7.3 写在最后

工具不重要，对工具的合理运用才重要。希望本模板对大家的论文写作有所帮助。

参考文献

[1] Cai W, Qiu G. Visual question answering algorithm based on image caption[A]. 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)[C], 2019 : 2076 – 2079.

[2] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[A]. Advances in neural information processing systems[C], 2014 : 2672 – 2680.

[3] KINGMA D P, WELLING M. Auto-encoding variational bayes[J]. arXiv preprint arXiv:1312.6114, 2013.

[4] REZENDE D J, MOHAMED S, WIERSTRA D. Stochastic backpropagation and approximate inference in deep generative models[J]. arXiv preprint arXiv:1401.4082, 2014.

[5] HINTON G E, SEJNOWSKI T J, ACKLEY D H. Boltzmann machines: Constraint satisfaction networks that learn[M]. [S.l.] : Carnegie-Mellon University, Department of Computer Science Pittsburgh, 1984.

[6] ACKLEY D H, HINTON G E, SEJNOWSKI T J. A learning algorithm for Boltzmann machines[J]. Cognitive science, 1985, 9(1) : 147 – 169.

[7] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. science, 2006, 313(5786) : 504 – 507.

[8] NOWOZIN S, CSEKE B, TOMIOKA R. f-gan: Training generative neural samplers using variational divergence minimization[A]. Advances in neural information processing systems[C], 2016 : 271 – 279.

[9] MIRZA M, OSINDERO S. Conditional generative adversarial nets[J]. arXiv preprint arXiv:1411.1784, 2014.

[10] DONAHUE J, KRÄHENBÜHL P, DARRELL T. Adversarial feature learning[J]. arXiv preprint arXiv:1605.09782, 2016.

[11] LI Y, SWERSKY K, ZEMEL R. Generative moment matching networks[A]. International Conference on Machine Learning[C], 2015 : 1718 – 1727.

[12] PERARNAU G, VAN DE WEIJER J, RADUCANU B, et al. Invertible conditional gans for image editing[J]. arXiv preprint arXiv:1611.06355, 2016.

致谢

以简短的文字表达作者对完成论文和学业提供帮助的老师、同学、领导、同事及亲属的感激之情。

附录 A　数据

A.1　第一个测试

测试公式编号

$$1 + 1 = 2. \tag{A.1}$$

表格编号测试

表 A.1　测试表格

11	13	13	13	13
12	14	13	13	13