Voice-2-image：自然语言交互下的智能图像合成

吴丽佳[1]，徐　昆[2]

（1. 清华大学美术学院，北京　100084；
2. 清华大学信息科学技术学院，北京　100084）

摘要：基于自然语言处理以及智能图像处理算法，提出一种由语音自动生成真实感图像的方法。在实现中，由智能抠图得到图像素材，由语音识别、语意解析得到文本结构，以三分图计算、融合算法、图像语境匹配，自动生成真实感图像。相比于以往基于生成式对抗网络（generative adversarial network，GAN）进行文本合成图像、简笔画输入合成图像或用大量人工抠取素材合成图像的方法，本文所提方法更具趣味性、更能节约时间、达到更好的合成效果。该方法可以应用于多种类型的应用设计，如卡通图片合成、儿童学习型应用、照片合成等。

关键词：计算机应用；自然语言处理；图像处理；文本结构；真实感图像合成
中图分类号：TP399　　**文献标识码：**A　　**文章编号：**1674-2850(2020)04-0399-12

Voice-to-image: intelligent image synthesis based on natural language interaction

WU Lijia[1], XU Kun[2]

(1. *Academy of Arts & Design*, *Tsinghua University*, *Beijing*　100084, *China*;
2. *School of Information Science and Technology*, *Tsinghua University*, *Beijing*　100084, *China*)

Abstract: Based on natural language processing and intelligent image processing algorithms, we propose a method to synthesize realistic images automatically by voice interaction. In the preprocessing step, an image cutout database is built through image matting. At runtime, text semantic is reconstructed using voice recognition and text parsing, trimap is generated from mask, and then the realistic images are automatically synthesized by text and image context matching. Compared with the previous works that generating images from text based on generative adversarial network (GAN), sketches and labels, or a large amount of artificial material preparing, the proposed method is more interesting, time-saving and the synthesized images look more real. In addition, this method can be applied to different types of application designing, such as cartoon image synthesis, children learning applications, and photo synthesis.

Key words: computer applications; natural language processing; image processing; text structure; realistic image synthesis

0　引言

在计算机图形学和视觉领域，图像合成已被广泛且深入地研究。文本生成图像作为近几年的热门研究领域，其解决的问题是从描述性的文本生成与之对应的图片。

近年来最流行的解决方案主要以 GAN 思想为主干，经过 GAN[1~2]、GAWWN[3]（generative adversarial what-where network）、堆叠生成式对抗网络（stacked generative adversarial network，StackGAN[4]）、使用

基金项目：国家自然科学基金（61822204）
作者简介：吴丽佳（1992—），女，硕士研究生，主要研究方向：图像合成
通信联系人：徐昆，副教授，主要研究方向：计算机图形学. E-mail: xukun@tsinghua.edu.cn

堆叠生成式对抗网络生成真实感图像（realistic image synthesis with stacked generative adversarial network，StackGAN++[5]）、基于注意生成式对抗网络的细粒度文本图像生成（fine-grained text to image generation with attentional generative adversarial network，AttnGAN[6]）、文本条件辅助分类器生成式对抗网络（text conditioned auxiliary classifier generative adversarial network，TAC-GAN[7]）、单一流生成器结构[8]等模型的逐步优化和改进，从生成低分辨率 36×36 的图像到细节、逼真度上表现越来越好的 256×256 图像，展示了 GAN 的潜力。但此类解决方案主要针对单个类别物体生成单一物体图像，虽然生成的单个物体图像细节、逼真度上表现越来越好，但无法生成有多个交互对象的复杂场景。

针对有多个交互对象的复杂场景的图像生成，2018 年 JOHNSON 等[9]在以文本描述生成场景图（scene graph）的方法[10]的基础上，实现了用 scene graph 结合机器学习自动生成新图像的方法，但此方法合成图片尺寸受限为 64×64，且合成效果模糊。同年，HONG 等[11]提出了由文本描述生成多对象的 Bounding Box，再生成 Mask，得到 Layout，最后由 Layout 生成图像的方法，该方法可以生成语义丰富且有意义的图像，但是当图像中多个物体位置相近时，物体间的边界模糊不清，易出现多个物体糊在一起的效果。

除此类以像素级别生成图像的方法外，2009 年 CHEN 等[12]提出 Sketch2Photo，实现了以简笔画和文字标签合成图片的方法，该方法的缺点是需要人为画出简笔画，并且还要结合文字标签来选取合适的前景物体。2013 年 HU 等[13]提出 PatchNet，实现了以简笔画的形式，通过图片纹理对比选择合适的前景物体合成图片，但同样也需要人为画出简笔画。这两类方法合成的图像尺寸任意，且场景复杂，效果逼真。

基于上述研究背景，本文主要思考以下问题：1）希望合成有多个交互对象、内容丰富、场景复杂的图像；2）如何才能减少人工对素材的前期处理、图像合成时的干预；3）是否能用简单、便于推广应用的方法来实现；4）文本输入是否能用更自然、更有趣的方式代替。

鉴于此，本文提出语音合成图像的方法，以达到人工成本大量减少、交互方式有趣、合成效果较为真实、合成方法简单易推广的研究目的。

1 方法概述

本文主要使用语音交互输入文本的方式，利用现有的图像素材，通过相关图像处理技术来生成真实感图像：

1）在语音识别与语句分词的基础上，模拟日常说话习惯来设计文本句式，构建文本语义结构；

2）利用抠图、融合、羽化等图像处理算法，自动处理图片素材生成前景素材；

3）根据文本语义结构以及局部对比前景、背景图像语境，选取最合适的前景素材，合成真实感图像。

1.1 图片素材处理

前景素材处理：结合多种图像处理算法对图片素材进行处理，自动生成大量前景素材。本文尝试以下两种处理方法。

1）Mask→Trimap→Alpha：采用 CHENG 等[14]实现的 DenseCut 算法从前景素材中得到前景物体的 Mask；采用 GUPTA 等[15]实现的 Automatic Trimap Generation 方法由 Mask 得到 Trimap；采用 SHAHRIAN 等[16]提出的 Improving Image Matting 方法，结合 Trimap 及前景素材生成 Alpha.

2）Mask→边缘羽化 Mask：先采用 DenseCut 算法从前景素材中得到前景物体的 Mask，再采用 Median Blur 方法羽化 Mask 边缘。

背景素材处理：将背景图片 BG 按照图像语义不同分为不同的局部区域 BG_i，$BG=\{BG_1,BG_2,\cdots,BG_n\}$，

并计算每个区域的颜色均值。

1.2 文本语义结构构建

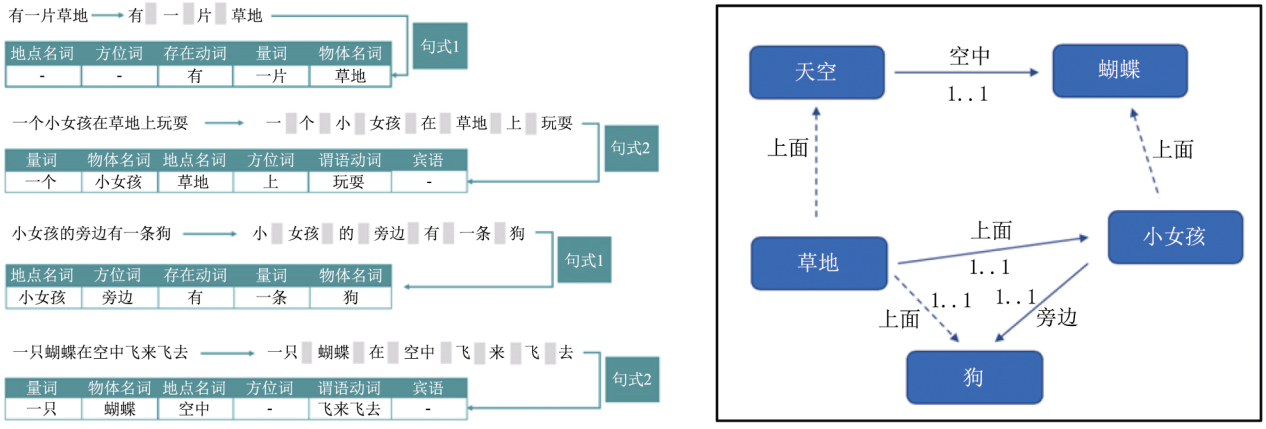
本文利用了百度语音识别技术，在语音输入后，可直接得到词语分解后的文本，例如，输入：*草地上有一棵树*。得到结果：*草地 上 有 一棵 树*。本文在此基础上，将词语分解后的文本按设计的句式构建文本语义结构。

本文设计的句式如下（这里暂只设计两种句式，可根据需求扩展至多种句式）：

句式 1 在 地点名词 方位词 存在动词 量词 物体名词（即：在某处有某物。）

句式 2 量词 物体名词 在 地点名词 方位词 谓语动词 宾语 （即：某物在某处做什么。）

通过以上句式，可以明确：背景 BgName、前景物体 FgName、前景物体的数量 FgNum、参照物，前景物体将要放置的方位。由此可得到前景物体之间的相对位置关系，并构建语义结构。将文本按句式分解、构建语义结构如图 1 所示。



a b

图 1 文本句式分解（a）与语义结构构建（b）

Fig. 1 Text parsing (a) and text semantic construction (b)

1.3 最合适前景物体检索及图像合成

首先，在前景素材库中检索根据文本句式得到的前景物体 FgName，得到前景物体候选项 F，$F=\{F_1,F_2,\cdots,F_n\}$，根据 BgName 在背景素材库随机选择一张作为背景图 BG；然后，由前景物体之间的相对位置关系可得到背景图中的目标位置区域 $\mathrm{BG_{dst}}$，计算目标位置区域 $\mathrm{BG_{dst}}$ 的颜色均值、前景物体候选项 F 的原有背景 F_{bg} 的颜色均值；最后，计算两者差距，选择差距最小的作为前景物体 FG：

$$\Delta C=\left\|\frac{\sum_0^{i\times j}\mathrm{BG_{dst}}}{i\times j},\frac{\sum_0^{N}F_{\mathrm{bg}}}{N}\right\|,\ \mathrm{FG}=\min\Delta C\ , \tag{1}$$

其中，i，j 为背景图中目标位置区域 $\mathrm{BG_{dst}}$ 的宽、高；N 为前景物体候选项原有背景 F_{bg} 的像素数量。

将得到的前景物体 FG，结合其 Alpha 或羽化后的 Mask，与背景图进行融合：

$$I=\alpha\mathrm{FG}+(1.0-\alpha)\mathrm{BG}\ , \tag{2}$$

其中，I 为合成的目标图像；α 为 Alpha 或羽化后 Mask 的 RGB 值。

1.4 相关算法

根据文本语义结构可知，前景物体之间的相对位置关系主要由参照物、方位决定，参照物可以是背

景的某个局部区域，也可以是某个已有的前景物体。判断此两种情况，可在背景图中寻找到绘制前景物体的目标位置（算法 1）。

算法 1：前景物体寻找目标位置的方法

1：语音文本按规定句式进行处理，得到前景物体候选项、参照物、方位；
2：**for** dstWhere ← BG_1 to BG_n **then**
3：　**if** 参照物 == dstWhere **then**
4：　　目标区域 BG_{dst} ← dstWhere
5：　　**for** FG ← F_1 to F_n **do**
6：　　计算 FG 的原背景 F_{bg} 与目标区域 BG_{dst} 颜色均值的差距 ΔC；
7：　　**end for**
8：　　选择最小 ΔC 代表的候选项作为前景物体 FG；
9：　　将前景物体 FG 随机放置在目标区域 BG_{dst} 内，不越界；
10：　　**return**；
11：　**end if**
12：**end for**
13：**for** dstWhere ← 已有的前景物体 **then**
14：　**if** 参照物 == dstWhere **then**
15：　　目标区域 BG_{dst} ← 参照物所在的背景局部区域
16：　　**for** FG ← F_1 to F_n **do**
17：　　计算 FG 的原背景 F_{bg} 与目标区域 BG_{dst} 颜色均值的差距 ΔC；
18：　　**end for**
19：　　选择最小 ΔC 代表的候选项作为前景物体 FG；
20：　　按照参照物、目标区域 BG_{dst}、方位选择目标位置 POS_{dst}；
21：　　在目标位置 POS_{dst} 绘制前景物体 FG；
22：　　**return**；
23：　**end if**
24：**end for**
25：**return null**；

以参照物为中心，绕此中心按九宫格的方式确定方位的具体位置 POS，如图 2 所示。但在此位置直

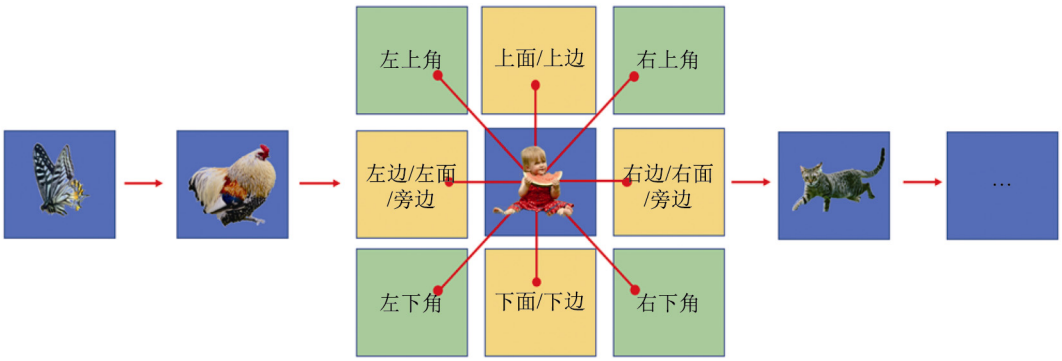


图 2　九宫格方法确定具体位置

Fig. 2　Determination of specific location by the method of nine grid

接绘制前景物体，可能会与已有的其他前景物体产生重叠。本文使用背景图占位标记的方法，调整目标位置，解决重叠问题（算法 2）。（在图像视觉领域，物体重叠是一个非常正常且合理的现象。但是在本文中，由于物体与物体的大小比例、物体与背景之间的融合问题，应尽量避开重叠以达到最优效果。）

算法 2：按照参照物、目标区域 BG_{dst}、方位选择目标位置 POS_{dst}

1：输入前景物体、参照物、目标区域、方位、占位图；
2：获得前景物体的大小 width1，height1；
3：按参照物、方位初始化目标位置 POS(dist*X*, dist*Y*)，不越界；
4：根据目标位置 POS、前景物体大小确定目标 BoundingBox；
5：temp ← 50；
6：**while** 占位图中，目标 BoundingBox 被占位/越界 **do**
7： **if** 目标 BoundingBox 左侧区域/右侧区域被占位/越界 **then**
8： POS(dist*X*, dist*Y*) ← (dist*X* +/− width1/2, dist*Y*)；
9： 更新目标 BoundingBox；
10： **if** 占位图中目标 BoundingBox 区域未被占位 **and** 不越界 **then**
11： POS_{dst} ← POS(dist*X*, dist*Y*)；
12： return POS_{dst}；
13： **else then**
14： POS(dist*X*, dist*Y*) ← (temp, dist*Y*)；
15： 更新目标 BoundingBox；
16： Temp ← temp + 100
17： **if** temp > 背景图宽度 **then**
18: temp ← random(0, 背景图宽度)
19： **end if**
20： **end if**
21： **end if**
22：**end while**
23：POS_{dst} ← POS(dist*X*, dist*Y*)；
24：**return** POS_{dst}；

以 RGB(0, 0, 0)初始化一张背景图大小的占位图，前景物体所在的位置以 RGB(255, 255, 255)占位。初始化的目标位置 POS 分两种情形：1）以 POS 为中心绘制前景物体，不越界，不与其他前景物体重叠，此 POS 即为 POS_{dst}；2）存在重叠或越界时，以前景物体的半个宽度向左或向右调整 POS 的位置，再次判断是否与其他前景物体发生重叠或越界，若否，则此 POS 即为 POS_{dst}；若是，则在目标区域 BG_{dst} 中横向选定某一位置，以 100（也可以是 30、50、80、150、200 等数值，为 100 时，调整的平均次数较少）的步长移动 POS，直到不重叠且不越界。

按占位图调整目标位置的方法如图 3 所示。

2 实验

2.1 实验平台

本文实验基于 Win10 64 位系统以及 Android 8.0，使用 Python、Java 编程语言完成。

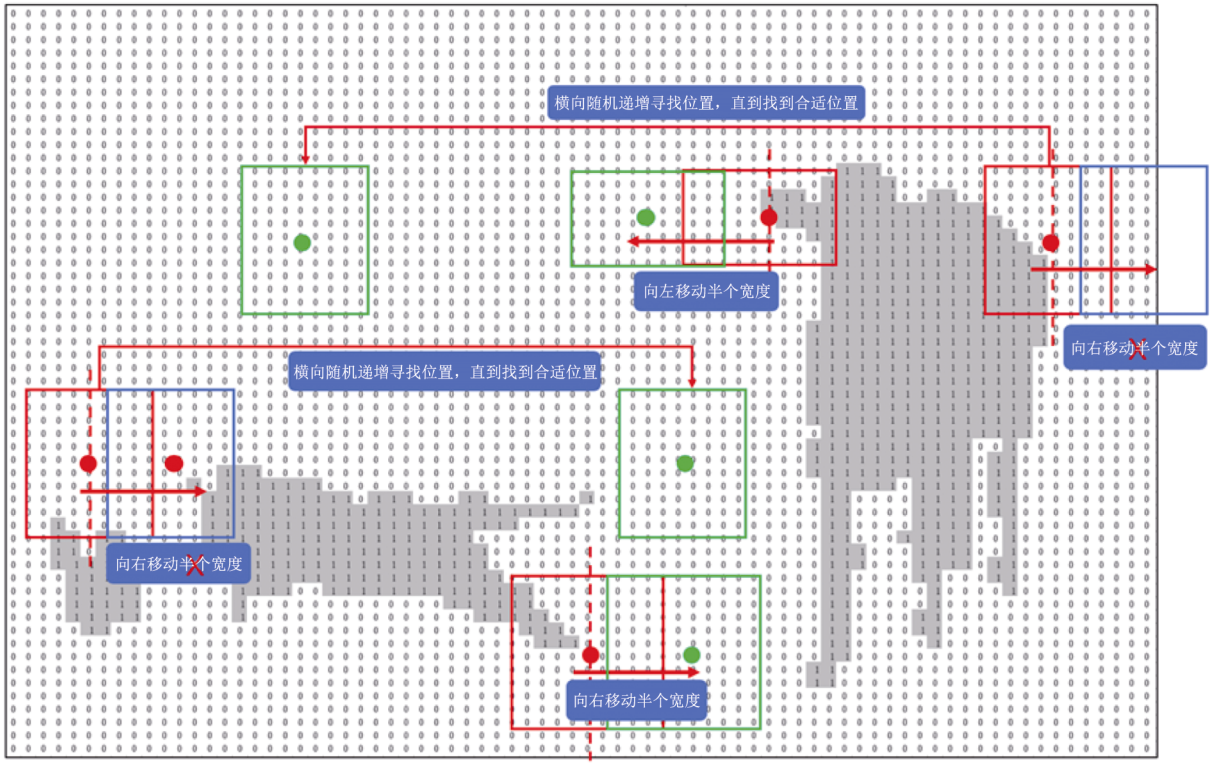


图 3　背景图标记占位法寻找合适位置

Fig. 3　Findation of the suitable location by marking the background image's pixels

2.2　图片素材收集及处理过程

本文的图像合成以户外风景图为主，主要涉及冰川、沙滩、草原、草地、河、海 6 个场景，总共 25 张 1 920×1 080 背景图片，194 张尺寸不限的前景素材图分为人、动物、飞机、船、木屋等类别，可根据需要进行扩充。部分素材如图 4 所示。

a　　b

图 4　背景素材（a）和前景素材（b）

Fig. 4　Background image material (a) and foreground image material (b)

本文尝试采用了两种方法对前景素材进行处理，图 5 和图 6 所示分别为以 Mask→Trimap→Alpha 方法和 Mask→边缘羽化 Mask 方法处理素材的过程。

背景素材处理：本文将每张背景图按天、地、水的类别从上至下分成 3 个局部区域（若想使前景、背景融合更自然，一定范围内局部区域分割越细越好）：$BG_{天}$={天空，空中，蓝天，…}，$BG_{地}$={沙滩，海

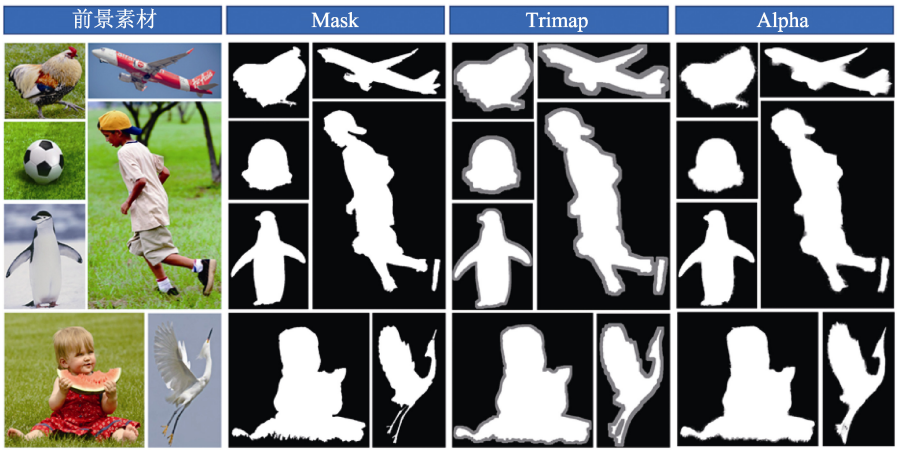


图 5 Mask→Trimap→Alpha 前景素材处理过程

Fig. 5 Foreground image material processing: Mask→Trimap→Alpha

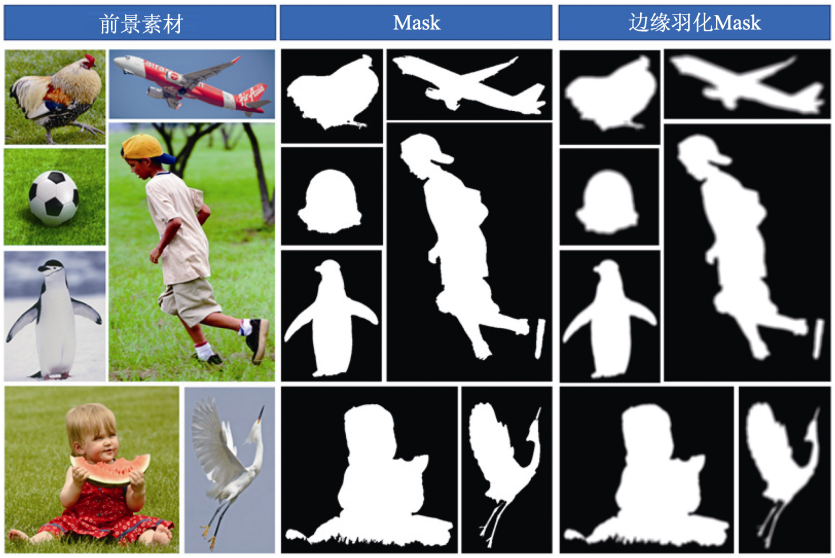


图 6 Mask→边缘羽化 Mask 前景素材处理过程（blur 的步长为 3）

Fig. 6 Foreground image material processing: Mask→Mask's edge feather (step=3)

边，河边，草地，…}，$BG_{水}$={海，河，海面，河面，…}，并计算每个区域的平均颜色值。图 7 展示了部分背景的局部区域分割。

图 7 背景图局部区域分割示意图

Fig. 7 Sketch map of local area segmentation for background image

2.3 实验结果

本文针对 25 张背景图，每张背景图添加 2～5 个前景物体，组合合成了新的图像。图 8 展示了 9 个

实验示例。

<table>
<tr><td rowspan="2">实验示例 1
语音输入：在草地上，一个小女孩在草地上玩耍，草地上有一只猫，小女孩的左边有一只公鸡，一架飞机在天空中飞过。</td><td>（1）在草地上</td><td>（2）一个小女孩在草地上玩耍</td></tr>
<tr><td> </td><td> </td></tr>
<tr><td>（3）草地上有一只猫</td><td>（4）小女孩的左边有一只公鸡</td><td>（5）一架飞机在天空中飞过</td></tr>
<tr><td> </td><td> </td><td> </td></tr>
<tr><td rowspan="2">实验示例 2
语音输入：在草原上，水面上有一只老鹰，草地上有一匹马，三只鸟在水面飞过。</td><td>（1）在草原上</td><td>（2）水面上有一只老鹰</td></tr>
<tr><td> </td><td> </td></tr>
<tr><td>（3）草地上有一匹马</td><td>（4）三只鸟在水面飞过</td></tr>
<tr><td> </td><td> </td></tr>
<tr><td rowspan="2">实验示例 3
语音输入：在河中，河面上有一艘小船，一只鸟在河面飞过，天空中有一架飞机。</td><td>（1）在河中</td><td>（2）河面上有一艘小船</td></tr>
<tr><td> </td><td> </td></tr>
<tr><td>（3）一只鸟在河面飞过</td><td>（4）天空中有一架飞机</td></tr>
<tr><td> </td><td> </td></tr>
<tr><td>实验示例 4
语音输入：在北极，一只北极熊在冰块上行走，北极熊的左边有一只海豹。</td><td>实验示例 5
语音输入：在草原上，一匹马在草地上吃草，一匹马在草地上奔跑，一只鸟在蓝天中飞翔。</td><td>实验示例 6
语音输入：在草原上，草地的左边有一只羚羊，草地的右边有一只长颈鹿，一只老鹰在空中飞。</td></tr>
<tr><td> </td><td> </td><td> </td></tr>
</table>

图 8 合成图像示例（Mask→Trimap→Alpha 处理）

Fig. 8 Examples of synthesized images (processed by Mask→Trimap→Alpha)

2.4 两种素材处理方法合成效果对比

图 9 所示两种处理方法对比结果表明，前景素材经 Mask→Trimap→Alpha 方法或 Mask→边缘羽化 Mask 方法处理后，均可合成效果较理想的图像，但在细节处稍有不同。在图 10 中给出了两种方法处理后的素材在细节处的比较，结果表明，在使用同一个 Mask 的前提下，经 Mask→Trimap→Alpha 方法处理过的素材，边缘更清晰，保留的细节更多。经 Mask→边缘羽化 Mask 方法处理过的素材，合成后易形成边缘虚化效果，影响真实感。

图 9 合成效果对比

Fig. 9 Comparison of synthetic effects

在背景图与前景物体的原背景颜色相近的情况下，使用 Mask→Trimap→Alpha 方法优于 Mask→边缘羽化 Mask 方法，合成的图像更逼真。在背景图与前景素材的原背景颜色相差较大的情况下，使用 Mask→边缘羽化 Mask 方法优于 Mask→Trimap→Alpha 方法，因为 Mask→Trimap→Alpha 方法处理后的素材所残留的背景会使合成的图像真实感大幅度下降。但背景残留的多少主要取决于获取 Mask 的抠图算法的优劣。

2.5 设计扩展至卡通图片合成

在图 11 中，以《小猪佩奇》为例，结合语音输入、文本语义结构构建、图片处理，合成卡通图片，图中展示了 3 个实验示例。卡通图片的合成更为简单，只需抠图，无需进行 Mask 后续处理以及前后背景语境对比。此方法还可扩展到任意的卡通动画。

图 10　两种方法处理后的素材细节对比

Fig. 10　Comparison of details in the image materials processed by these two methods

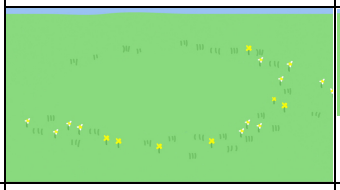
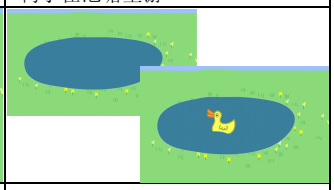
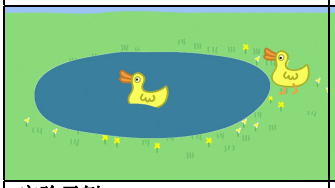
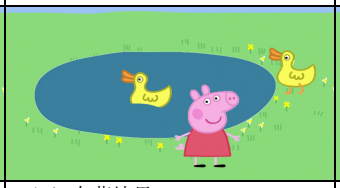
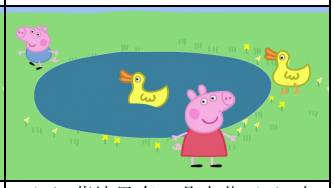
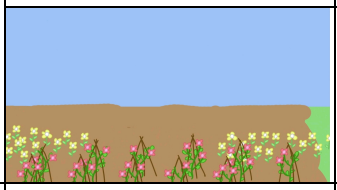
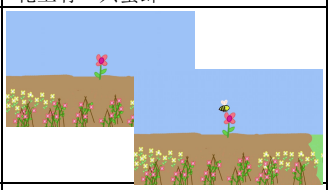
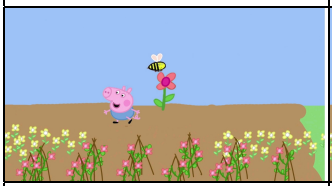
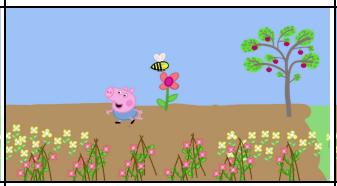
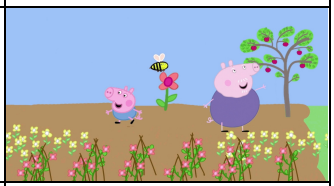
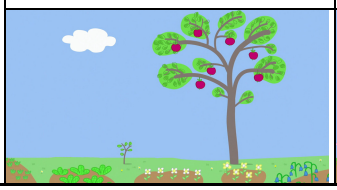
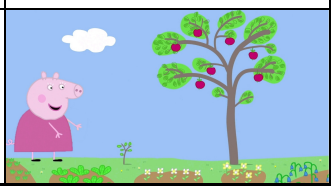
<table>
<tr><td rowspan="2">实验示例 1
语音输入：在草地上，草地上有一个池塘，一只鸭子在池塘里游，一只鸭子在草地上走，佩奇在池塘边玩，乔治在草地的左边跑来跑去。</td><td>（1）在草地上</td><td>（2）草地上有一个池塘（3）一只鸭子在池塘里游</td></tr>
<tr><td>  </td><td>  </td></tr>
<tr><td>（4）一只鸭子在草地上走</td><td>（5）佩奇在池塘边玩</td><td>（6）乔治在草地的左边跑来跑去</td></tr>
<tr><td>  </td><td>  </td><td>  </td></tr>
<tr><td rowspan="2">实验示例 2
语音输入：在菜地里，菜地里有一朵小花，小花上有一只蜜蜂，乔治在菜地里玩耍，菜地的右边有一棵苹果树，猪爷爷在苹果树下干活。</td><td>（1）在菜地里</td><td>（2）菜地里有一朵小花（3）小花上有一只蜜蜂</td></tr>
<tr><td>  </td><td>  </td></tr>
<tr><td>（4）乔治在菜地里玩耍</td><td>（5）菜地的右边有一棵苹果树</td><td>（6）猪爷爷在苹果树下干活</td></tr>
<tr><td>  </td><td>  </td><td>  </td></tr>
<tr><td rowspan="2">实验示例 3
语音输入：在猪奶奶的花园里，猪奶奶在花园里干活，佩奇在浇水，乔治在佩奇的旁边玩耍，一只蝴蝶在空中飞来飞去。</td><td>（1）在猪奶奶的花园里</td><td>（2）猪奶奶在花园里干活</td></tr>
<tr><td>  </td><td>  </td></tr>
</table>

图 11 《小猪佩奇》卡通图片合成

Fig. 11 Synthesis of cartoon images in *Peppa Pig*

3 结论

本文将自然语言处理与图像处理算法相结合，实现了全自动的智能语音图像合成，不仅在一定程度上节约了素材处理的大量人力成本，也使合成效果更为真实，图像内容更丰富，且交互方式更有趣，合成方法更简单。

但也存在一定缺陷：1）目前的文本结构构建适应句式有限，待进一步扩展；2）由于无法控制前景物体的相对尺寸大小，所以还需要人为进行图片素材的筛选；3）合成效果主要取决于前景物体抠取的准确性。

关于前景物体 Mask 的抠取，还可用 Mask R-CNN（区域卷积神经网络）实例分割前景物体的方法[17]来代替 DenseCut 方法。使用 Mask R-CNN，对素材库的选择几乎没有任何要求，且从同一张图中得到的前景物体包含了相对尺寸大小关系。

随着图像处理技术的成熟，本文实现的方法还能从不同的方向实现人力成本的节省，达到更优的合成效果。且本文实现的基于自然语言的无接触式交互合成图片的方法，还可以应用到更多的工程项目中，可在机器人行业、教育行业等得到扩展应用，具有广阔的市场前景。

[参考文献] (References)

[1] MANSIMOV E, PARISOTTO E, BA J, et al. Generating images from captions with attention[EB/OL]. New York: Cornell University[2016-02-29]. https://arxiv.org/abs/1511.02793.

[2] REED S, AKATA Z, YAN X C, et al. Generative adversarial text to image synthesis[EB/OL]. New York: Cornell University [2016-06-05]. https://arxiv.org/abs/1605.05396v2.

[3] REED S, AKATA Z, MOHAN S, et al. Learning what and where to draw[C]//Proceedings of the 29th Conference on Neural Information Processing Systems (NIPS'16). Red Hook: Curran Associates Inc., 2016: 217-225.

[4] ZHANG H, XU T, LI H S, et al. StackGAN: text to photo-realistic image synthesis with stacked generative adversarial networks[C]//2017 IEEE International Conference on Computer Vision (ICCV). Los Alamitos: IEEE Computer Society, 2017: 5908-5916.

[5] ZHANG H, XU T, LI H S, et al. StackGAN++: realistic image synthesis with stacked generative adversarial networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(8): 1947-1962.

[6] XU T, ZHANG P C, HUANG Q Y, et al. AttnGAN: fine-grained text to image generation with attentional generative adversarial networks[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 1316-1324.

[7] DASH A, GAMBOA J C B, AHMED S, et al. TAC-GAN-text conditioned auxiliary classifier generative adversarial network[EB/OL]. New York: Cornell University[2017-03-27]. https://arxiv.org/abs/1703.06412v2.

[8] ZHANG Z Z, XIE Y P, LIN Y. Photographic text-to-image synthesis with a hierarchically-nested adversarial

network[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 6199-6208.

[9] JOHNSON J, GUPTA A, LI F F. Image generation from scene graphs[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 1219-1228.

[10] SCHUSTER S, KRISHNA R, CHANG A, et al. Generating semantically precise scene graphs from textual descriptions for improved image retrieval[C]//Proceedings of the Fourth Workshop on Vision and Language. Lisbon: Association for Computational Linguistics, 2015: 70-80.

[11] HONG S, YANG D, CHOI J, et al. Inferring semantic layout for hierarchical text-to-image synthesis[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City: IEEE, 2018: 7986-7994.

[12] CHEN T, CHENG M M, TAN P, et al. Sketch2Photo: internet image montage[C]//ACM Transactions on Graphics, 2009, 28(5): 1-10.

[13] HU S M, ZHANG F L, WANG M, et al. PatchNet: a patch-based image representation for interactive library-driven image editing[J]. ACM Transactions on Graphics, 2013, 32(6): 196-1-196-12.

[14] CHENG M M, PRISACARIU V A, ZHENG S, et al. DenseCut: densely connected CRFs for realtime GrabCut[J]. Computer Graphics Forum, 2015, 34(7): 193-201.

[15] GUPTA V, RAMAN S. Automatic trimap generation for image matting[C]//2016 International Conference on Signal and Information Processing (IConSIP). Vishnupuri: IEEE, 2016: 1-5.

[16] SHAHRIAN E, RAJAN D, PRICE B, et al. Improving image matting using comprehensive sampling sets[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland: IEEE, 2013: 636-643.

[17] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 42(2): 386-397.

（责任编辑：张媛媛）