

**2023 中国高校计算机大赛-人工智能创意**

**赛**

**复赛作品说明书**

所属赛区：\_\_\_\_\_\_\_\_\_\_华中赛区\_\_\_\_\_\_\_\_\_\_\_\_\_\_参赛单位：\_\_\_\_\_\_\_\_\_\_武汉理工大学\_\_\_\_\_\_\_\_\_\_团队名称：\_\_\_\_\_\_\_\_\_\_\_\_猪也想飞\_\_\_\_\_\_\_\_\_\_\_\_作品名称：基于PPDiffusers 的图像编辑集成平台联系人：\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_展一航\_\_\_\_\_\_\_\_\_\_\_\_\_联系电话：\_\_\_\_\_\_\_\_\_\_\_18932693179\_\_\_\_\_\_\_\_\_\_

组别：

□赋能组（□EasyDL□BML□AI □创新组（飞桨）  
□航天组（飞桨）

studio 大模型社区）

**2023 中国高校计算机大赛-人工智能创意赛**1

**基于PPDiffusers 的图像编辑平台**

**作者：展一航、吴乐恒、郭祥顺**

**（一）选题说明**

**1.1 设计背景**

当前市场上存在一些图像处理的产品或服务，但它们通常缺乏文本引导的功能和基于大模型的复杂任务处理能力。下面是一些市场上已有产品或服务的例子。

传统图像处理软件：例如Adobe Photoshop、GIMP 等，这些软件提供了广泛的图像处理功能，但用户需要熟悉复杂的界面和参数设置，并且缺乏文本引导的功能。

在线图像编辑工具：例如Canva、Pixlr 等，这些工具通常提供一些简单的图像编辑功能，但在处理复杂任务和个性化需求方面受限较多。

生成网络（GAN）应用：一些应用利用GAN 技术提供了图像风格迁移、图像变换等功能，但用户通常需要上传和选择样本图像进行处理，不能通过文本直接进行引导和个性化控制。

**1.2 目的及市场调研现状**

为了解决这些图像处理软件的不便之处，我们基于百度PaddlePaddle 框架开发了一个易于上手功能强大的图像编辑一体化平台。与这些已有产品或服务相比，我们的基于大模型的图像处理集成平台具有以下优势和区别，具体如下：

a. 文本引导功能：用户可以通过输入自定义的文本指令来引导图像处理过程，实现 个性化的图像变换、编辑和放大，从而提供更灵活、智能化的图像处理体验。

b. 基于大模型的强大能力：我们的平台整合了PPDiffusers 包，可以直接导入大模型和参数，从而提供更强大、高效的图像处理功能。用户可以通过平台快速应用最先进的图像生成和编辑技术。

c.突破传统限制：相比传统图像处理工具，我们的平台超越了界面复杂、参数繁琐的限制，通过结合文本引导和大模型的推理能力，使用户能够创造出更富有创意和个性化的图像效果。

2

虽然市场上存在一些图像处理产品或服务，但我们的基于大模型的图像处理集成平台通过文本引导和大模型处理能力的结合，提供了全新的图像处理方式和个性化控制的能力，与已有产品或服务形成明显的差异。PaddlePaddle 是百度自主研发的集深度学习核心框架、工具组件和服务平台为一体的技术领先、功能完备的开源深度学习平台，有全面的官方支持的工业级应用模型，涵盖自然语言处理、计算机视觉、推荐引擎等多个领域，并开放多个预训练中文模型。通过PaddlePaddle 我们可以方便快捷地利用已有的模型，开发新功能，并且结合不同模型的特点不断扩展功能。

**（二）作品内容**

**2.1 主要内容**

包含多个图像编辑任务的一体化平台，简单便捷，易于上手，可以快速导入模型，进行开发。文本引导的图像变换：用户可以通过自定义的文本引导指令，实现个性化的图像变换，包括风格迁移、颜色转换等。这种处理方式能够提供全新的图像编辑体验，并为用户带来更多创意和表达可能。

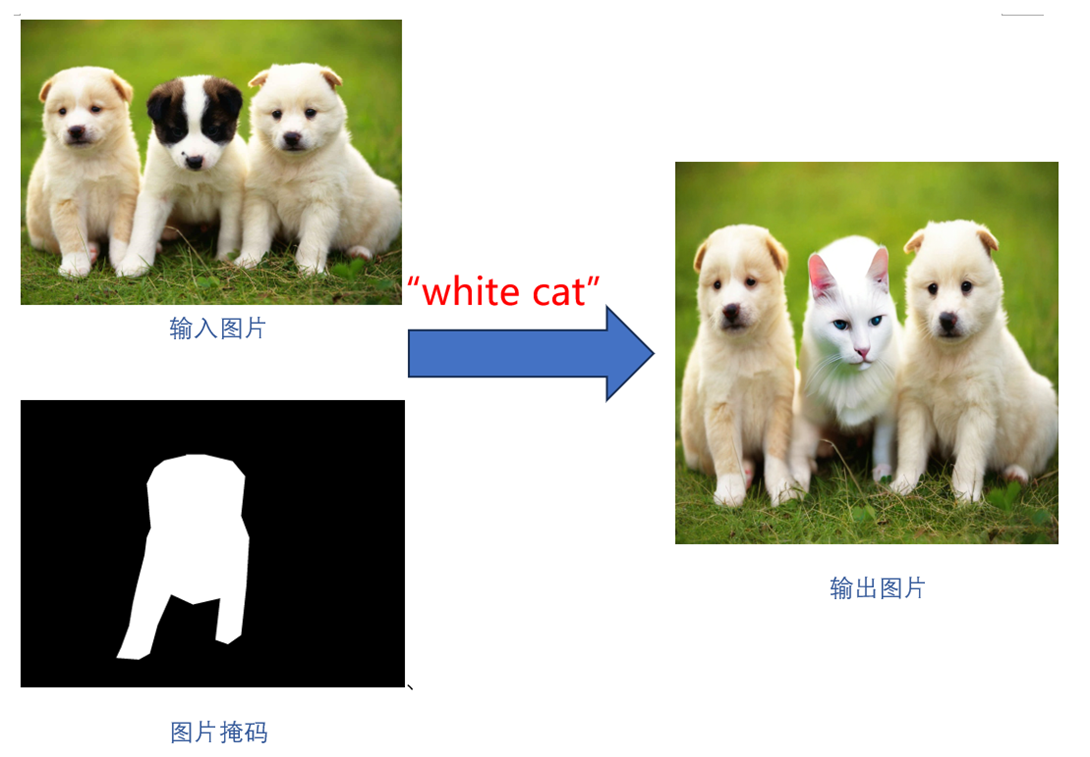
多引导的图像变化：用户可以输入图片与文字，对图片进行变换，并且在变换场景之后，图中的元素接近原图。如图1 所示，输入一辆汽车或者街道的图片，输入引导文本“in night”，得到输出图片：



**图1 文本引导的图像变换示例**

3

文本引导的图像编辑：相比于传统的图像编辑工具，我们的平台允许用户通过输入文本来引导图像的编辑过程，例如改变场景、调整细节等。这种基于文本的编辑方式具有强大的扩展性和创造力。如图2 所示，输入有三只小狗的图片与想要进行修改的位置的掩码，如果不加文字引导，可以将中间的小狗给去除，通过文字"white cat"，会在指定的位置生成一只白色的小猫。



**图2 文本引导的指定区域图像编辑示例**  
**2.2 创新点与项目特色**  
本项目的创新点和主要贡献为：  
基于国产的百度PaddlePaddle 平台开发，依托于PaddlePaddle 框架、Paddle 自然语言处理开发库以及PPDiffusers，借助现有的基础模型，根据需求功能调整模型的超参数，提升模型的精度和稳定性。同时添加了translate 功能，支持中文文本引导图像变换。

使用Gradio 将基于生成模型的图像编辑处理功能集成到一个平台上以供使用，极大简化用户的操作，包括可视化的strength 和guidance\_scale 参数的调整，并且由于PPDiffusers 的强大拓展能力，我们的平台可以根据用户需要，快速开发添加新功能。

本项目的特色主要为：  
 文本引导的图像生成是当前图像编辑在实际应用中一个非常有实用价值的研究方向，可以应用于广告设计、新闻宣传、学习教育、国防军事等重要领域。同时，当前对该技术的研究仍然存在生成精度与效率的问题，特别是在文本输入有限的情况下，模型的体量和

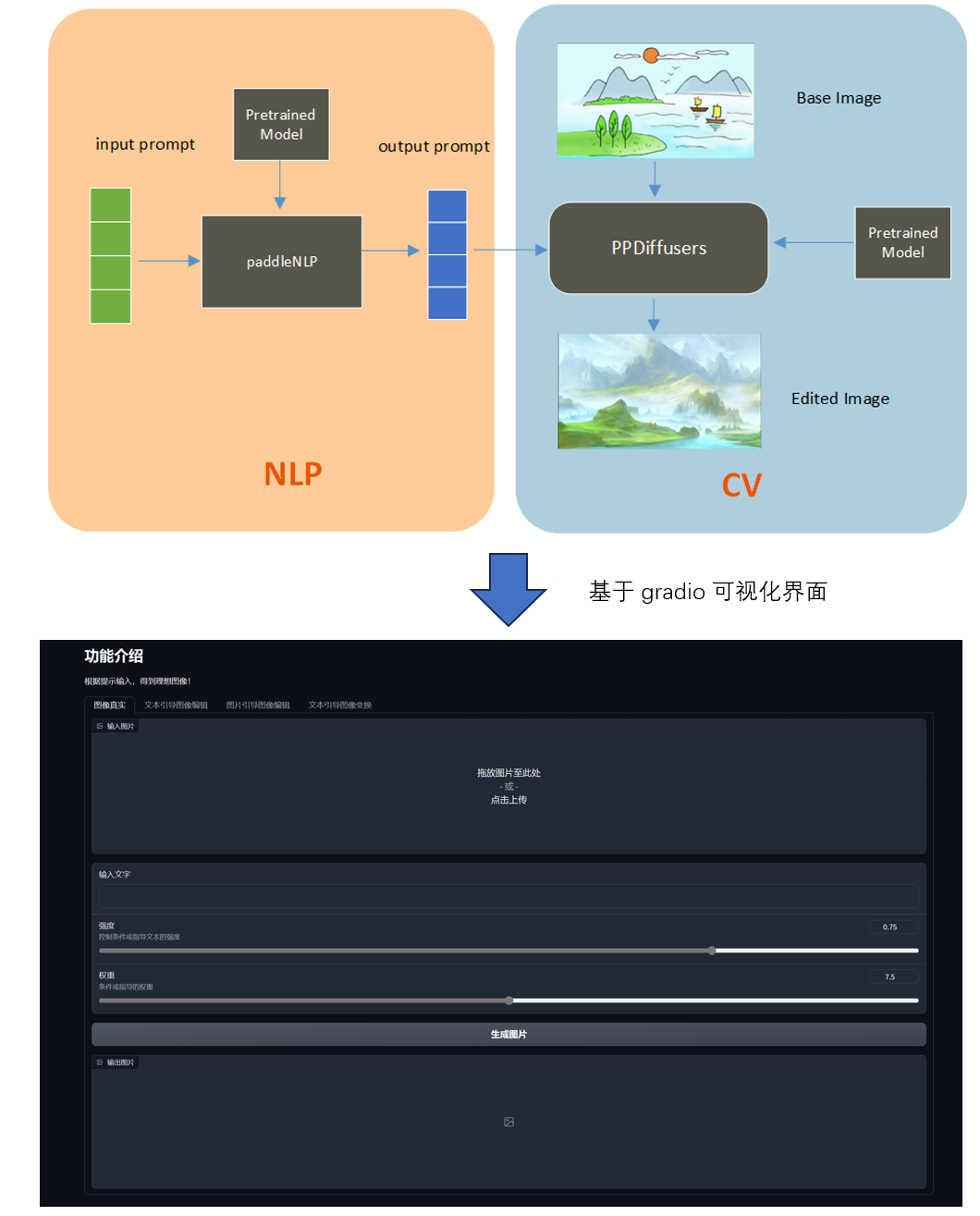
4

计算速度对于该技术的实用化十分关键。因此，本项目的研究具有较高的科学价值和实用价值。

**2.3 技术路线**

（1）总体框架

如图3 所示，我们采用paddleNLP 与PPDiffusers，根据不同任务的需要导入模型利用其中的模型功能完成任务，将先进的自然语言处理和计算机视觉模型借助gradio 等工具集成到我们的图像处理平台中。实现通过文本引导的图像处理功能，并提供更灵活、智能化的用户体验。



5

**图3 总体框架示意图**

以上述提到的通过文本引导图片变换风格为例，首先将文字通过paddleNLP 处理，得到特征向量然后将图片一起输入到PPDiffusers 中进行处理，最后得到变换风格后的图像。

（2）自然语言处理模块

通过**paddleNLP** 导入模型，为解决特定任务提供良好的文本处理能力。**paddleNLP** 有着强大的api 可以快速完成大部分任务，并且封装灵活，低耦合高内聚，保证用户可以通过继承和改写满足特定的数据处理需求。同样内置丰富的中文数据集，如中文诗歌古典文集数据、中文对联数据集等，可以针对比较特殊的中文处理任务，

（3）计算机视觉模块

通过PPDiffusers 导入模型，为解决特定任务提供良好的图像生成或编辑能力。对于图像风格变换任务来说，

我们使用StableDiffusion 模型来进行图像变换时，可以调整以下参数达到目的：

 扩散步数（num\_diffusion\_timesteps）：通过增加或减少步数，来控制扩散过 程的迭代次数。较多的步数可以生成更高质量的图像，但会增加计算时间和资源 消耗。较少的步数可能会导致生成结果缺乏细节。根据任务需求和资源限制，逐 步调整扩散步数以获得最佳效果。

 扩散温度（beta\_schedule）：扩散温度控制了噪声的强度，从而影响生成图像 的样式和多样性。较高的扩散温度会产生更随机多样化的图像，而较低的扩散温 度则会产生更加清晰和锐利的图像。

 图像尺寸和批次大小：调整输入图像的尺寸和批次大小来控制模型的计算和内存 消耗。较大的图像尺寸和较小的批次大小会产生更高质量的生成结果，但也会增 加计算时间和资源占用。根据任务要求资源限制，选择合适的图像尺寸和批次大 小。

同样对于特定的任务，可以采用自己的数据集进行微调，通过对提示词的调整，让用户经常输入的提示词可以被模型学习到。从而达到更好的效果，收集相关领域的对话数据，并将其作为微调数据集。然后，将这个数据集与基础模型结合，进行进一步训练。这样，模型就能够更好地理解和回答与与之相关的问题，并能够更准确地生成与相关的内容。

**（三）应用前景**

6

通过市场调研显示，大约有80%以上的人认为现有的Photoshop 等相关图像处理软件的操作过于复杂，难以在短时间内轻松上手，70%以上的人明确表示希望有一种通过输入自定义的文本指令来引导图像处理的功能来帮助人们提高工作效率。对照当前市场竞争，本项目可以很好地解决操作复杂问题，达到的效果相比现有的一些上市软件同样具有竞争力。

通过输入自定义的文本指令来引导图像处理可以增加其可普及性和可泛化程度，因为这种定制化的功能可以满足不同用户的需求，并扩展应用领域。用户可以根据自己的具体需求使用文本指令来指导图像处理任务，在设计、广告、医疗、安防等领域都可以应用该技术。

在社交媒体上，图片是吸引用户注意力和增加互动的重要元素。通过根据输入文本引导已有基础图片，可以生成更具吸引力和相关性的图片内容，提高社交媒体的曝光度和用户参与度。新闻和媒体行业需要频繁地报道和发布新闻内容，并配以相关图片。使用这种技术，可以根据新闻内容自动生成相关的图片，提高新闻报道的效率和准确性。在教育领域，使用图片来辅助教学和培训是常见的做法。通过根据已有文本适当修改图片，可以根据不同的教学内容和知识点生成相应的图片，提高学习效果和可理解度等等。以上需求，都可以通过该项目提供的平台快速且满意地完成。同时，国家对于国产化平台的开发也一直积极支持和鼓励创新，包括使用技术来改进图片的生成和修改过程，以促进相关领域的发展和应用，证实了项目进行的必要性。

7