

**智图设计——基于深度学习的图像处理平台**

**目 录**

[1 作品简介 2](#_Toc86219952)

[2 应用场景 2](#_Toc86219953)

[3 作品亮点 3](#_Toc86219954)

[3.1 技术创新 3](#_Toc86219955)

[3.1.1 TWO STAGE结构设计 3](#_Toc86219956)

[3.1.2 多尺度特征信息的使用 4](#_Toc86219957)

[3.1.3 前后端框架搭建 4](#_Toc86219958)

[4 工作原理 4](#_Toc86219959)

[5 技术方案 5](#_Toc86219960)

[5.1 技术路线 6](#_Toc86219961)

[5.2 模型设计 7](#_Toc86219962)

[5.3 部署设计 7](#_Toc86219963)

[6 解决的实际问题 7](#_Toc86219964)

[7 团队分工 8](#_Toc86219965)

[8 后续工作 9](#_Toc86219966)

**1 作品简介**

本项目的方向是人工智能图像处理，为了满足自媒体工作者的需求，以及方便普通用户，简化图像处理任务。优点和期望如下：用途广泛：为自媒体行业及个人用户提供便携服务，无需下载软件，在线使用即可；操作简单，无需使用PS、美图秀秀等软件复杂的操作，AI自动识别主体，通过上传图片就能完成作图；功能易用，上传图片后无需标注前景背景，自动识别抠图区域，快速获得带透明背景PNG。

**2 应用场景**

在日常工作生活中，抠图技术被广泛使用，例如替换证件照背景，图层遮罩甚至于视频会议时的背景实时替换。抠图技术的强大效果促进了互联网媒体的发展，但它也有着不小的局限性，对于设备层面，需要使用PS这类专业的软件去实现，而且这类软件对于设备性能也有一定的需求；在使用层面，这些专业软件的使用方法较为繁琐，软件内置的、工具功能繁多，对于没有使用过图像处理软件的人群学习成本较高，而且想要获得好的抠图效果需要进行精细的描边工作。因此很有必要设计一款软件来简化抠图操作，帮助更多的人实现轻松便捷的抠图工作。

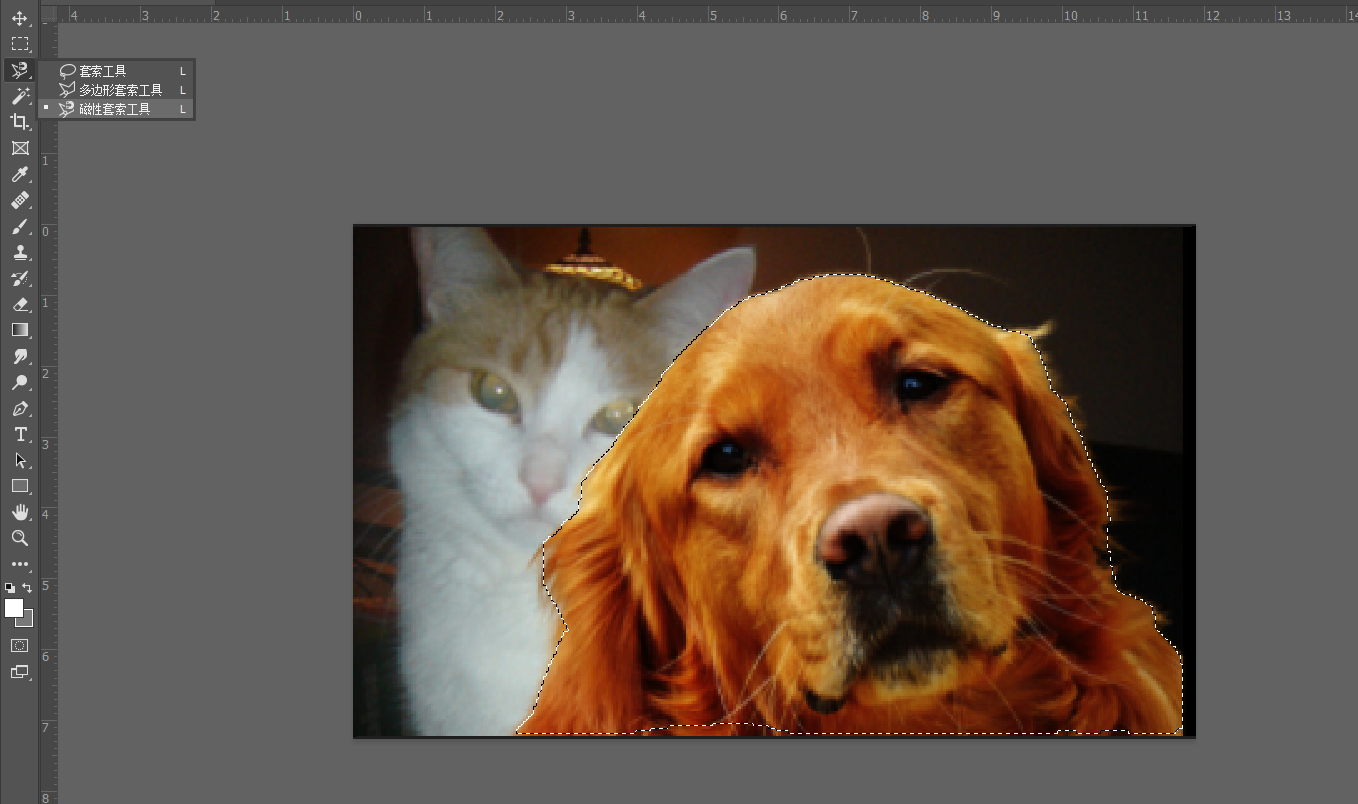


图 1 PS抠图实例

上图是使用专业软件Photoshop来抠图的实例，抠图的方法有很多，如魔棒工具或者套索工具等等，对于专业人士处理起来得心应手，但是对于没有使用经验的人群来说抠图的效果会适得其反，这是专业图像处理软件所面临的问题。

因为抠图技术注重于对发丝级的细节处理，图片的清晰度在处理过程中发挥了至关重要的作用。但对于实际环境中，使用相机获取的照片，因为拍摄时的环境因素，如光线不足或过强，或者网络上的图片在传播过程中出现图片大小压缩的情况，都会导致图片不清晰，后续处理丢失细节信息。因此，引入清晰度修复功能对待处理图像修复，恢复细节信息，能有效提高抠图等任务的准确率。对于图像的修复，通常需要极为专业的技术，所以需要设计算法来对模糊的图像进行像素级的预测与修复，并通过部署项目到互联网中来实现智能化清晰处理工作。

**3 作品亮点**

**3.1 技术创新**

以下将从模型结构设计、多尺度特征信息的使用以及前后端框架搭建这三个方面分别介绍本项目在实现过程中的技术创新

**3.1.1 TWO STAGE结构设计**

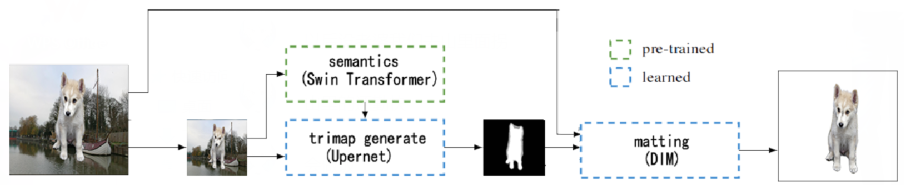


图 2 TWO STAGE结构

抠图模型中，由于E2E的模型构造无法获取足够的特征信息，效果达不到需求，因此我们基于三分图(trimap)的理念设计了二阶段模型，通过三分图作为中间输出，将SwinTransformer和DIM(Deep Image Matting)结合使用。并且我们改进了一阶段模型中的Upernet编码头结构，使得它可以用于三分图生成任务。

**3.1.2 多尺度特征信息的使用**

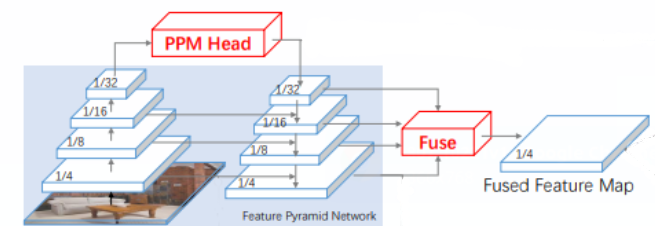


图 3 多尺度特征信息

单纯的使用单尺度图像无法满足我们对精细抠图的需要，如发丝的处理使得主体的边缘变得难以分辨。因此我们采用了swin-transformer与upernet相结合的方式，获得多尺度信息进行模型的构建，并且我们改进了一阶段模型中的Upernet编码头结构，使得它可以用于三分图生成任务。

**3.1.3 前后端框架搭建**

开发过程中，我们利用Vue.js前端框架的MVVM模式降低了前端代码耦合程度，提高了可重用性，使得前后端交互相应更加迅速。

我们使用Flask框架开发后端，其主要特点是小而轻，同时Flask自带的开发服务器使我们在调试程序时无须再安装其他任何网络服务器，便于我们在开发与测试时快速搭建web服务.

**4 工作原理**

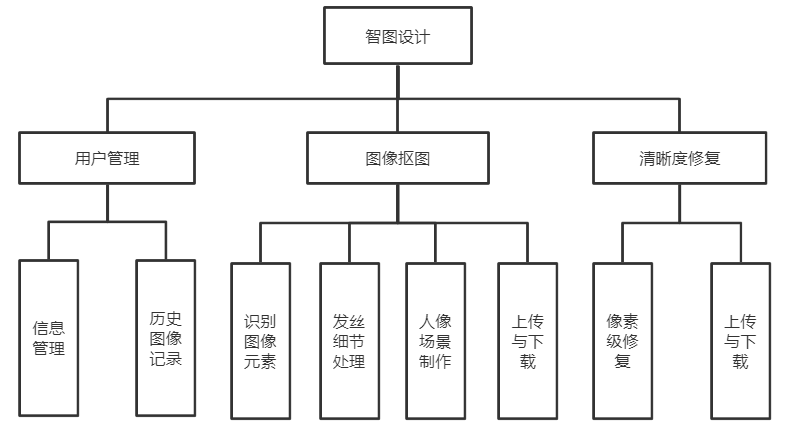


图 4 系统架构图

如图所示，该项目主要由三个模块组成，分别为用户管理模块，图像抠图模块，清晰度修复模块。

用户管理模块主要实现功能有信息管理、历史图像记录两点，用于记录用户的历史信息，并使用户可以在个人中心内查询到历史的抠图和清晰度修复图像记录，便于在不同设备上共享数据。

图像抠图模块主要有图示中四点功能，在用户上传图片之后，Web端将会把输入图片传至后台模型，经过模型的判断，识别出图像中的元素，并根据图像中前景的实例类别，进行针对性的优化，实现发丝精度的抠图处理。在获得扣取出的前景之后，用户可以选择为图像添加新的背景，从而实现新的场景制作。

清晰度修复模块在抠图技术的基础上，实现了对用户上传的模糊图像，经过模型识别图中实例，对特定类别进行针对性的修复，获得清晰图像供用户使用。

**5 技术方案**

为了实现该项目所需的功能，该技术方案将从技术路线、模型设计、项目开发角度给出相应的解决方案。

**5.1 技术路线**

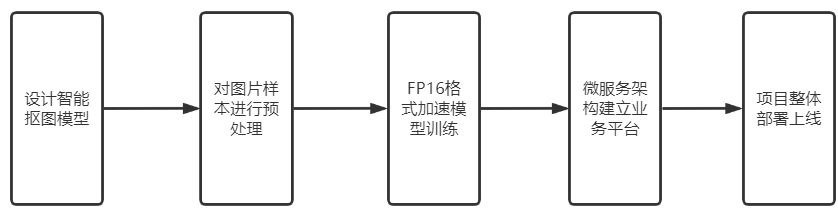


图 5 技术路线图

该项目的实现主要分为五个步骤，如上图所示，分别为设计智能抠图模型；对图片样本进行预处理；采用FP16格式加速模型训练；使用Flask微服务框架和vue框架，构建业务平台以及前端项目；项目整体部署上线，在网页中投入实际使用。

第一步设计智能抠图模型，我们采用基于python语言的pytorch框架作为技术栈，对抠图技术中所需完成的生成trimap任务和精细抠图任务分别设计对应模型。

第二步对图片样本的预处理，我们采用opencv作为技术栈，对数据集中的无用数据进行了清洗，基于voc、coco等数据集进行数据增广，以满足模型训练的需要。

第三步加速模型训练中，我们采用由nvidia开发的，基于pytorch的混合精度训练加速器Apex来减少训练过程中现存的占用，加快训练和推断的计算，从而实现模型训练的加速。

第四步微服务架构建立业务平台，我们采用Flask以及vue作为技术栈，用于实现前后端之间，不同服务之间的分离解耦，利于开发人员间的同步开发以及后期维护。

第五步项目整体部署上线，我们采用Devops工具，实现开发运维一体化，将构建好的各项微服务按照既定规则进行合成发布，降低了部署与运维的复杂性。

**5.2 模型设计**

抠图的精度通常要求达到发丝级处理精度，当前的基准方法需要用户输入额外的语义信息，通常是以三分图Trimap(前景、背景、未知区域)或草图的形式去约束输入图像从而生成alpha通道图像。用户输入的Trimap和草图在该任务中起到了至关重要的作用，但这种user-in-the-loop的形式带来的交互成本极大，使得该技术难以应用于大规模数据的处理。

为了解决传统抠图技术中交互成本高、鲁棒性较差的问题，我们提出了一个无需用户显式输入约束便可以很好的获得alpha matte从而完成抠图的方法。该方法对传统的语义分割任务进行了改进，与现有的基于显着性检测或基于语义分割的方法不同，我们遵循前景区域是离成像源最近的语义实体，将语义分割任务转变为trimap生成任务，并设计了对应的损失函数。

**5.3 部署设计**

在项目开发中，我们采用了前后端分离的设计思路。

对于前端部分的设计，我们采用vue.js前端框架，实现了视图、数据、结构分离，前端工程化。同时作为单页面应用，使页面局部刷新，不用每次跳转页面都要请求所有数据和dom，这样大大加快了访问速度和提升用户体验。并且开发过程中使用了第三方ui库，节省了开发时间。

对于后端部分，为了维护开发过程中的易拓展，易维护性质，我们采用Flask微服务架构，将整体项目拆分成一套小且互相关联的服务，易于使用和拓展，并且配置灵活。该框架所使用的语言与模型算法部分相同，均为python语言编写，因此我们可以很轻松的将Web服务与模型部署连接到一起，降低开发时间。

**6 解决的实际问题**

随着自媒体行业高度发展，越来越多的人选择自媒体工作。在制作视频时，需要将图片添加到视频中作为素材等，快速去掉图片的背景则成了难题。对于一部分工作者，在进行图像专业处理时有一定的难度，所以快速简单的图像处理也成了他们的需求之一。

为了解决抠图技术在实际应用中所带来的不便问题，本项目采用云服务+人工智能技术，使用户脱离传统的机器设备与繁琐的操作，通过简单的上传下载便可得到抠取得到的需求图像，大幅度提高自媒体工作者与普通用户的效率，降低使用门槛。

本项目无需使用PS、美图秀秀等专业图像处理软件进行复杂的操作，AI自动识别主体，通过上传图片就能完成图像处理。直接在网页上上传图片，无需标注前景背景，省时省力，就可以快速获得带透明背景PNG图片。

对于证件照拍摄时没有保存多种背景时，或者用户想要自己足不出户完成证件照的简单拍摄时，处理证件照背景便成了一大问题，而抠图技术正提供了捷径。在线根据需求更改证件照背景颜色，处理照片更方便，省时省力省钱。

综上，我们的希望是开发一款面向大众的、功能齐全的智能抠图软件，功能易用，无需学习成本，涵盖更广的问题域，为用户提供更优质的服务。

**7 项目展示**

本项目一共由三个模块组成，第一个为起始导航模块，该模块的对项目功能进行了简单介绍，可以通过左上角切换智能抠图或者清晰度处理两个功能，并且通过点击上传图片或者拖动一张图片到指定位置后完成上传图片，等待后台模型的处理。

第二个为智能抠图模块，该模块从前端接收到图片后，送入后台的抠图模型，在模型处理好之后，对比页面会将原图和处理好的图片放在一起对比，并且保存按钮可以实现将图片一键保存到本地。

第三个为清晰处理模块，在接收到图片后，后台模型会分析图中的实体轮廓，进行针对性的修复，对比页面会将原图和处理好的图片放在一起对比，拖动中间的横条可以达到实时对比的效果，并且同样可以一键保存。

**8 团队分工**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 姓名 | 学号 | 所在学院 | 年级 | 专业 | 项目中的分工 |
| 段续 | 1925121008 | 计算机科学与技术学院 | 2019级 | 软件工程 | 设计构建前端项目，设计网页端初步原型，设计页面。后期材料整合、PPT制作与视频录制。 |
| 谢佳宸 | 1925121013 | 计算机科学与技术学院 | 2019级 | 软件工程 | 设计智能抠图模型，完成模型的训练，设计后端测试模型。完成文档的技术部分，调整文档结构，技术解答顾问。 |
| 林恭瑜 | 1915103013 | 计算机科学与技术学院 | 2019级 | 计算机科学与技术（人工智能） | 设计构建后端模型，调整前后端接口，完成前后端的连接。 |
| 张志豪 | 1925121035 | 计算机科学与技术学院 | 2019级 | 软件工程 | 测试项目流程，保证项目功能对需求的满足。及时发现项目的Bug并反馈，以做修改。 |
| 潘恩楦 | 1925121018 | 计算机科学与技术学院 | 2019级 | 软件工程 | 对项目后期各个阶段的运营与维护，及时接收用户反馈，以便后期升级迭代。 |

**9 后续工作**

首先，在项目运行方面，未来还需调整系统架构，促进微服务转型，通过运用云计算技术，从而改进性能，降低软件成本。在用户使用层面，利用人工智能技术，识别更多种类图像主体，提供用户个性化的选择，在此基础上我们未来希望添加图像编辑功能，可以对处理好的图片进行细微调整，使得抠图系统更人性化。与此同时，希望在未来集成更多图像增强处理模块，如图像无损放大、拉伸图像恢复等给用户带来全方位便捷的图像技术服务，覆盖更广的业务范围。

希望我们的项目可以帮到更多的人，在之后面对社会公众开放，让更多的人轻松使用网页工具来体验智能抠图便捷之处，或者是快速修复受损图片，减轻工作量的同时使创作更智能，感谢AI的发展，可以替我们处理这些乏味单调的工作，留下了更多发挥创意的空间。