

# 基于机器视觉的数码印花工厂自动报工系统

陆雨轩<sup>1</sup>, 周于钧<sup>1</sup>, 陈晋音<sup>1</sup>, 朱威<sup>1,2</sup>

(1. 浙江工业大学信息工程学院, 杭州 310023; 2. 浙江省嵌入式系统联合重点实验室, 杭州 310023)

**摘要:** 针对纺织行业中现有人工布匹生产订单报工管理困难的问题, 设计并实现了一个基于机器视觉的数码印花工厂自动报工系统, 将正在生产的布匹图像与订单库中的花型图案进行智能匹配, 从而对生产订单进行快速准确的报工管理; 首先系统客户端使用工业相机实时采集生产线上的印花布匹图像, 并对图像进行滤波、归一化等预处理; 然后将图像通过 HTTP 协议发送至系统服务端, 利用基于 Vision Transformer 的神经网络模型进行特征提取和多层特征融合, 再与订单库中的花型图案进行匹配, 找出待报工订单; 最后对匹配结果进行数据库管理, 并将匹配结果返回客户端用于报工结果展示与确认操作; 实验结果表明, 该自动报工系统的订单匹配 TOP-3 命中率达到 90.4%, 满足工厂布匹生产需求。

**关键词:** 数码印花工厂; 自动报工; 机器视觉; 视觉变换器; 图案匹配

## Automatic Work Reporting System for Digital Printing Factory Based on Machine Vision

LU Yuxuan<sup>1</sup>, ZHOU Yujun<sup>1</sup>, CHEN Jinyin<sup>1</sup>, ZHU Wei<sup>1,2</sup>

(1. School of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China;

2. Zhejiang Province Joint Key Laboratory of Embedded Systems, Hangzhou 310023, China)

**Abstract:** Aimed at the challenges of exiting manual reporting management in textile production, a machine vision-based automatic reporting system was developed for a digital textile printing factory. This system intelligently matches the patterns of the fabrics currently in production with the design patterns in the order database, enabling rapid and accurate management of production orders. Firstly, the system client uses an industrial camera to capture the real-time images of printed fabrics on the production line, and the images is filtered and normalized for the pre-processing. Then, these images are transmitted to the server system via HTTP protocol, and based on the neural network model of Vision Transformer, the feature extraction and multi-layer feature fusion are carried out. And then, the images are matched with the pattern design in the order database to identify the orders needed to be reported. Finally, it displays and confirms that the matching results are managed in the database and returned to the client for, the reporting results. Experimental results demonstrate that this automatic reporting system achieves a TOP-3 match rate of 90.4%, meeting the needs of factory fabric production.

**Keywords:** digital printing factory; automatic reporting; machine vision; vision transformer; pattern matching

## 0 引言

全球经济的低迷以及逆全球化的趋势给纺织行业带来了前所未有的挑战<sup>[1]</sup>。如何在这样的大环境下实现纺织产业的持续健康发展, 是目前纺织行业面临的重大课题。在纺织生产过程中, 订单报工环节承担着监控和管理各个生产工序的重要职责, 对于提升生产效率、保证产品质量以及优化资源配置具有至关重要的作用。报工不仅涉及生产进度的实时反馈, 还包含了对生产资源如原材料、机器设备以及人力资源的有效管理, 可以确保生产活动按照既定计划顺利进行, 同时也为生产过程中可能出现的问题提供了及时的预警和干预机制。

然而, 纺织行业传统的人工报工管理方式需要工人依靠随着布料一起流转的纸质记录单来跟踪和控制生产的各个阶段, 不仅效率低, 容易发生错误, 而且难以迅速适应市场的变化需求。特别是在数码印花布匹生产线上, 每个布匹的生产工序(如热转印、蒸化、水洗、定型等)都需要人工精准地登记和实时监控, 以确保生产订单的顺利完成。这一过程不仅消耗大量的人力资源, 还难以控制生产效率和成本。

随着人工智能视觉技术的快速发展, 针对纺织品生产, 已涌现了一些机器视觉技术应用方法, 主要集中在布匹检验环节。Liu 等人<sup>[2]</sup>基于改进的 SSD 网络对布匹瑕疵进行检

收稿日期: 2024-04-30; 修回日期: 2024-05-15。

基金项目: 国家自然科学基金项目(62072406)

作者简介: 陆雨轩(1998-), 男, 硕士。

陈晋音(1982-), 女, 博士, 教授。

通讯作者: 朱威(1982-), 男, 博士, 副教授。

引用格式: 陆雨轩, 周于钧, 陈晋音, 等. 基于机器视觉的数码印花工厂自动报工系统[J]. 计算机测量与控制, 2024, 32(10): 146-153.

测, 提高布匹质检的效率。Zhang 等人<sup>[3]</sup>使用在自然语言处理中应用较多的 Swin Transformer<sup>[4]</sup> 模块进行布匹瑕疵检测。本系统面向布匹生产的报工环节, 考虑到生产订单库中的订单印花图案与生产流水线中布匹印花图案是一一对应关系, 使得采用智能图案匹配技术实现自动化的布匹生产订单报工成为可能。因此, 本文针对数码印花工厂, 设计了一种基于机器视觉的印花图案自动生产报工系统, 以实现生产过程订单的实时跟踪和管理, 使工厂管理人员和订单客户可以及时了解生产状况, 具有良好的实际应用价值。所设计的系统结合了人工智能视觉技术, 能够实现对生产流水线上的布匹花型准确匹配。此外, 该系统在带图像处理器 (GPU) 的服务器平台进行匹配算法的加速部署, 并开发了基于 C/S 架构<sup>[5]</sup> 的客户端与服务端软件。

### 1 系统总体设计

本系统采用了面阵相机作为图像采集工具, 具体选用了海康威视的工业面阵相机 MV-CA020-10GC, 其分辨率达到 1 624 × 1 240, 最高支持帧率为 60 fps。考虑到生产线的实际速度约为 1 m/s, 布匹生产单的长度通常超过 100 m, 系统设定相机的采集帧率为 30 帧/秒。为了能得到稳定的图像数据, 系统选择 LED 灯作为相机的照明光源, 轨道 LED 灯能够提供稳定且高质量的照明效果。另外, 系统使用无锡华建电子仪器研究所的 CC-5D-99CRS 计米器来记录生产报工中所需要的布匹米数。

在客户端硬件平台方面, 为了确保客户端设备既能满足实际工业生产环境的稳定运行需求, 又能实现成本的最优化, 系统选择了工业一体机作为客户端开发平台。该工业一体机采用 X86 架构, 搭载了酷睿 i5 CPU 和集成显卡, 配备 2 GB 内存, 运行 Windows 7 操作系统。

在服务端硬件平台方面, 为了确保自动化检测系统的稳定性和实时性, 系统选择了搭载 NVIDIA RTX3090 GPU 的 Linux 服务器作为开发平台。

系统设计充分考虑工厂实际使用情况和成本控制, 采用分离客户端与计算资源的策略, 以达到高效性和经济性的平衡。客户端负责图像采集和用户交互, 采用低成本工业一体机, 减轻了设备采购和维护成本。服务端在高性能服务器上集中处理图像数据, 执行各个报工环节复杂的图像识别和匹配任务, 确保了处理效率和系统稳定性。自动报工软件总体框架如图 1 所示。

硬件系统各部分的连接如图 2 所示, 相机与 LED 灯悬挂于布匹生产流水线之上协作拍摄布匹图像, 计米器置于布匹上滚动测量生产米数, 而工业一体机收集图像与计米器数据, 并通过通信协议将数据发送到服务器中进行算法匹配与报工操作。

### 2 基于 C/S 结构的自动报工应用软件开发

所开发的自动报工应用软件, 基于 Qt 框架<sup>[6]</sup> 与 Linux 操作系统, 软件具体流程如图 3 所示。

本软件采用了分布式模型推理结构<sup>[7]</sup>, 该结构通过将计算密集型的推理任务从客户端迁移到服务端执行, 不仅

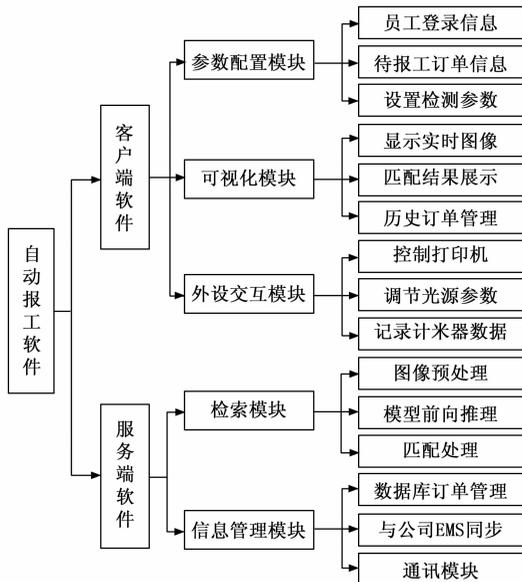


图 1 自动报工软件框架

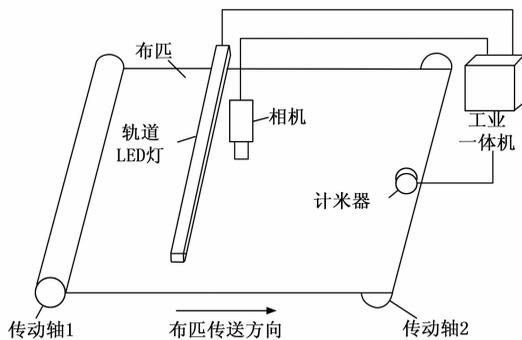


图 2 图像采集装置结构

显著降低了客户端设备的计算负担, 而且充分利用了服务端的强大计算资源来处理复杂的图像处理和模式匹配任务。客户端设备仅负责图像的采集和结果的展示, 同时将采集到的图像数据通过网络传输至服务端, 服务端接收到图像数据后, 执行深度学习模型的推理计算, 并将匹配结果反馈给客户端进行展示。

#### 2.1 客户端软件模块设计与实现

##### 2.1.1 客户端主界面

客户端的主界面是用户与系统交互的核心, 因此, 其设计不仅要满足功能需求, 还要考虑到用户体验的优化, 如图 4 所示。客户端的主界面是工作人员日常操作的核心, 包括了图像预览、实时匹配结果展示、历史匹配结果记录和控制面板等关键功能。

1) 图像预览模块: 该模块利用 Qt 的图形视图框架 (QGraphicsView) 实时展示流水线上的布匹图像。为了实现图像的即时显示和高效管理, 采用了信号与槽机制来处理相机捕获到的图像数据。此外, 通过调用 Qt 的多线程处理 (QThread) 接口, 确保了图像采集和展示的流畅性, 同

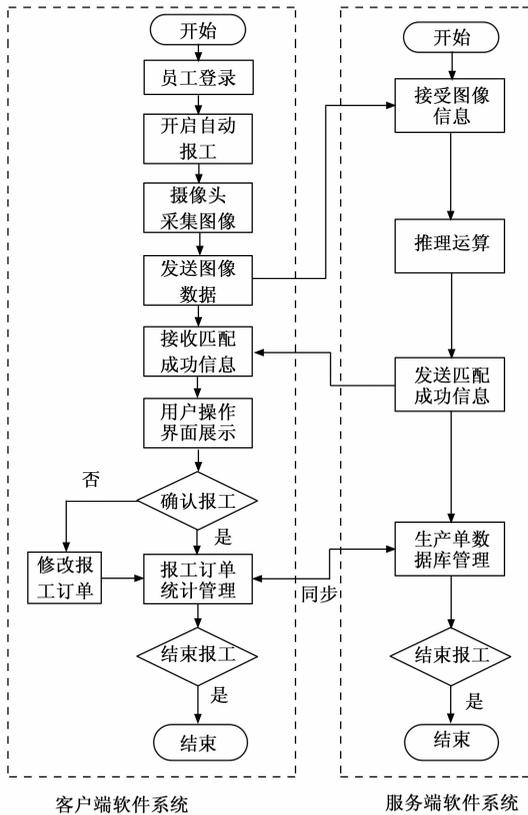


图 3 自动报工应用软件流程图

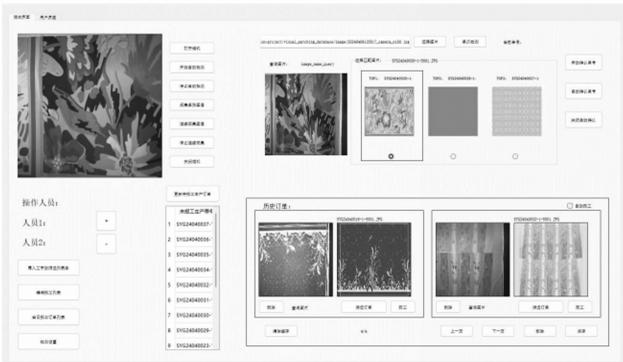


图 4 自动报工软件主界面

时避免了主界面在连续采集模式下的响应延迟

2) 实时匹配结果：直接显示当前拍摄的布匹图像与匹配成功的生产订单花型图像。此模块使用 Qt 的模型视图编程来实现匹配结果的动态更新和展示。通过自定义的列表模型 (QListModel) 和视图 (QListView)，实现了匹配结果的即时反馈。为了支持用户对匹配结果的操作，如自动确认和手动选择，引入了 Qt 的事件处理机制来响应用户的交互行为。

3) 历史匹配结果展示模块：记录并展示了所有成功匹配的订单，为生产管理和数据分析提供支持。该模块提供了数据的持久化存储功能，利用 SQLite 数据库通过 Qt 的

SQL 模块 (QSqlModule) 进行实现，保证了数据的稳定性和安全性。为了提升用户操作的便捷性，集成了简单直观的数据筛选、查看和删除功能。

4) 控制面板模块：作为系统管理的核心，涵盖了从流水线工作人员信息管理到操作人员添加、手动筛选列表等功能。利用 Qt 的布局管理实现了一套清晰的界面布局。此外，采用了 Qt 的打印机接口 (QPrinter) 来实现生产报告的打印功能，满足了生产记录和管理的需求。

### 2.1.2 图像采集模块

图像采集模块是自动订单匹配应用软件的高效运作的核心。本模块通过图像采集装置，捕捉经过热转印、蒸发、水洗与定型等多阶段处理的布匹图像。在此过程中，计米器滚轮的同步运转与布匹的移动紧密协同，通过向工业一体机 PC 发送脉冲信号，实时传递米数信息，确保了数据的准确性与实时性。

### 2.1.3 计米器模块

计米器模块负责实现工业工控机平台与计米器之间的有效信息交流。为了确保系统的高实时性和稳定性，本模块通过 UART 通讯串口实现双方的数据交互，数据协议为本系统定义的带有校验位、起止位与数据位的通讯协议。

计米器采用了与布面直接接触的编码器，其滚轮随布面移动而转动，产生脉冲信号。计米器控制器通过正交解码技术捕获这些脉冲信号，并计算出布匹的实时米数和运动速度发送回工控机平台，为后续的生产报工提供了精准的生产米数支持。

## 2.2 服务端软件模块设计与实现

在自动订单报工系统中，服务端承担了算法推理的核心任务，是实现自动化报工的关键环节。服务端软件模块的设计与实现，围绕着通讯收发图像与报工信息、推理计算以及数据库管理三大功能展开，以保证整个系统的高效性和高准确性运行。

### 2.2.1 通讯模块

在自动订单匹配系统中，通讯模块是实现服务端与客户端之间高效、稳定信息交换的关键。具体来说，客户端通过 HTTP 通信协议将捕获的图像数据安全地传输至服务端。服务端收到图像后，立即执行算法推理，这包括深度学习模型的计算，目的是精确地识别和匹配对应的生产订单。推理结束后，服务端将处理好的结果反馈给客户端，客户端接收到信息后，迅速在界面上展示，包括匹配的订单详情等，使得操作人员能够方便地获取所需信息。

服务端基于 C++ 开发的 HTTP 服务器，处理来自客户端的图像数据，并利用神经网络匹配算法进行图像识别与匹配，快速准确地输出匹配结果。考虑到自动订单匹配系统部署在工厂流水线时，其负责的工序数量众多，接收待匹配图片流量大，对处理速度和并发能力要求十分高，服务器设计上采用了高效的 epoll-ET 模式和非阻塞 IO，通过实现主从 Reactor 模式，大幅提升了服务器的事务处理速度和并发处理能力。

### 2.2.2 推理计算模块

服务端的算法推理计算是整个软件中计算资源消耗最为集中的部分。为确保系统的高效运行和实时性, 采用了生产者消费者模式进行多线程设计, 优化了图像处理和推理计算的流程。此模式下的图像通讯接收线程充当生产者角色, 负责将从客户端接收并预处理的图像放入缓冲队列中。相对地, GPU 上运行的推理线程则作为消费者角色, 负责从队列中取出图像并进行深度学习模型的推理计算。

在软件实现中, 使用了 C++17 的 `std::thread` 类创建并管理线程。此外, 条件变量用于实现生产者和消费者之间的同步机制。推理计算部分则采用了 TensorRT<sup>[8]</sup> 框架。为了将这些框架整合进 C++ 基于 HTTP 的服务器中, 需要在系统环境下编译生成相应的动态链接库, 然后通过 CMake 将这些依赖库加入项目中。

### 2.2.3 数据库管理模块

服务端数据库不仅作为数据存储和管理的中心, 还负责将汇总后的报工信息通过特定的通讯协议上报给工厂的生产执行管理系统 (MES)。这一集成过程提高了生产管理的自动化程度, 使得生产状态和订单进度能够实时反馈给企业管理层, 支持了更加精准和高效的生产决策。

数据库的字段设计包括生产单号、生产子单号、报工状态、花型下载地址、花型名称、花型 ID 以及布匹米数等。在数据库设计时进行了分表设计, 分为 4 个子表: 等待报工生产订单数据表、已报工历史生产订单数据表、等待报工生产子单数据表、已报工历史生产子单数据表。

数据库优化方面, 为了能在大量生产订单数据中快速地检索到本系统所需要报工管理的生产订单, 在数据库字段设计时, 选择合适的字段建立索引, 在多字段排序或者查询时, 也遵循最左前缀法则。

## 3 基于 Vision Transformer 的图像匹配方法

使用卷积神经网络 (CNN) 对图像特征进行提取可以实现图像检索, 已在医学图像等领域得到实际应用<sup>[9]</sup>。在本文的自动报工系统中, 图案匹配算法可以采用比 CNN 更为先进的 Vision Transformer (ViT)<sup>[10]</sup> 网络提取特征。因此, 本节设计了一种基于 ViT 模型的图像匹配方法。首先, 为了减少噪声干扰, 对实拍图像做滤波、归一化等预处理。接着, 将预处理后的图像输入到 ViT 网络模型进行特征提取, 并进行多层特征融合, 以更好地捕捉布匹图像的细节和纹理特征。最后将实拍花型图像与订单花型图像进行特征向量余弦相似度计算, 得到最高匹配度的图库花型进行生产报工。此外, 引入了 MoCo V3<sup>[11]</sup> 的自监督训练策略, 利用对比学习机制<sup>[12]</sup>, 使 ViT 模型能够在没有大量标注数据的情况下学习到有效的图像特征表示。

### 3.1 图像预处理

#### 3.1.1 图像采集

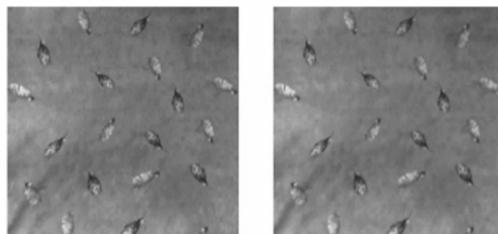
为了解决面阵相机高预设帧率 (30 Hz) 与实际生产线速度不匹配的问题, 通过降低相机的采集帧率, 有效减少了在生产速度较慢时拍摄到的高度重复的布匹花型图片。

这一策略不仅显著降低了对算力资源的需求, 而且提高了系统处理的效率, 确保了在限定时间内完成高质量的图像处理与匹配任务。

#### 3.1.2 图像滤波

在系统中, 选用了双边滤波<sup>[13]</sup>作为主要的降噪方法, 因为它在去除噪声的同时能保持图像边缘的锐度, 适合于系统需要保留布匹图案细节的任务。

在实际应用中, 本系统将双边滤波的参数空间域的标准差设定为 3 个像素单位, 这个值能够兼顾保持图像细节和去除噪声。对于强度域的标准差通常取决于图像强度分布的范围, 这里设置为 0.1 倍的可能像素强度范围 (在一张图像中像素值可以达到的最小和最大值之间的区间)。这个比例是基于典型数码印花图像的特性, 旨在去除细小的噪点, 同时保留图案的清晰边缘, 如图 5 所示。



(a) 原图

(b) 双边滤波

图 5 双边滤波器的滤波效果

#### 3.1.3 光照均衡化

在整理和分析现场收集的数据时, 观察到大多数采集到的布匹花型图像存在光照不均匀的问题。造成这种现象的原因可能包括: 摄像头镜头畸变、环境光照不均匀、光源亮度不均匀等。这种光照变化对于依赖于全局图像信息进行特征提取的神经网络算法而言, 会影响其性能。为此, 本系统采用了一种简单而高效的光照均衡化处理方法, 以改善图像的光照条件并提升算法的识别准确性。

光照均衡化的目标是减少图像内的光照不均匀性, 增强图像的细节可见性。处理流程如下:

步骤一: 颜色空间转换。

对于图像中的每个像素, 进行 RGB 颜色空间到 HSV 颜色空间的转换:

$$\text{HSV} = \text{converFromRGB}(\text{RGB}) \quad (1)$$

步骤二: 全局平均亮度计算。

在 HSV 空间中, 对 V (亮度) 通道计算全局平均亮度, 这反映了图像整体的亮度水平。

$$\alpha = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N V(x, y) \quad (2)$$

其中:  $V(x+y)$  是在位置  $(x, y)$  的亮度值,  $M$  和  $N$  分别是图像的宽度和高度。

步骤三: 子块平均亮度与调整。

类似于前述方法, 将 V 通道分为  $32 \times 32$  的子块, 计算每个子块的平均亮度, 并生成亮度矩阵  $\mathbf{A}$ 。然后得到亮度调整矩阵  $\mathbf{B}$  通过减去全局平均亮度:

$$\mathbf{B} = \mathbf{A} - \alpha \quad (3)$$

步骤四：上采样与亮度调整。

对矩阵  $\mathbf{B}$  进行上采样，并从 V 通道的每个像素值中减去对应的调整值，进行亮度均衡化。

步骤五：颜色空间还原。

在完成 V 通道的亮度均衡化后，将调整后的 HSV 图像转换回 RGB 颜色空间，以便进行后续处理：

$$\text{RGB}_{\text{equalized}} = \text{convertToRGB}(\text{HSV}_{\text{adjusted}}) \quad (4)$$

通过这种处理方式，可以有效地解决彩色图像的光照不均匀问题，而不会对图像的颜色饱和度和色调产生负面影响。这一步骤确保了在进行特征提取和图像匹配前，每个图像都具有相对均衡的亮度分布，效果如图 6 所示。



(a) 光照均衡化前 (b) 光照均衡化后

图 6 光照均衡化效果图

### 3.2 基于 ViT 的多层特征融合

#### 3.2.1 ViT 基础结构

系统中采用的是 ViT 的基础版本，即 ViT-Base，配合  $16 \times 16$  大小的图像块切分策略。这种选择考虑到了模型的复杂度与计算效率的平衡，同时能够充分捕获图案的细节特征。本系统选择的 ViT 模型具备 12 个编码层，这一配置在处理中等规模的图像数据集时表现出了良好的性能和效率平衡。

ViT 的核心思想是将输入图像分割为若干固定大小的图像块 (patches)，并将这些图像块通过线性映射转换为高维空间的嵌入向量。接着通过添加位置编码以融入空间信息，这些图像块的嵌入向量被送入一个标准的 Transformer 结构<sup>[14]</sup>中进行处理。Transformer 结构主要由自注意力 (Self-Attention) 层<sup>[15]</sup>和前馈神经网络 (FFN) 层交替堆叠而成，其中自注意力机制的引入使得模型能够在处理每个图像块时，考虑到其他所有图像块的信息，从而有效捕获图像的全局上下文关系。

#### 3.2.2 多层特征融合设计

考虑到 ViT 不同层输出的特征向量代表了图像的不同层次信息，下面重点讨论如何优化 ViT 的结构以获得更适合花型图案匹配的特征向量，特别是通过特征融合策略<sup>[16]</sup>来增强模型的表现。

ViT 模型通过多个编码层处理图像，每个编码层都能捕获不同层次的视觉特征。编码层的输出通常包含了丰富的局部和全局信息，而最后一层的输出则更倾向于整合这些信息，为分类任务提供了高层次的特征表示。对于花型图案匹配任务而言：

编码层输出：提供了从浅层到深层的逐渐抽象化的特征表示，能够捕获图像的细节和结构信息。

最后一层输出：通过整合前面所有编码层的信息，提供了一个全局化和高层次的特征表示。

为了充分利用 ViT 模型提供的不同层次的特征，并提高花型图案匹配的准确性和鲁棒性，在进行 ViT 模型深入研究的过程中，本系统针对自动布匹订单匹配任务，对 ViT 模型进行了实验分析，特别是对模型各个编码层的输出进行了细致的探索。目标是找到最优的特征表示，以提高图像匹配任务的性能。经过实验验证，发现 ViT 模型的第 4、5、6 编码层输出，相比于其他层及最终输出层，能够为布匹图像匹配任务提供更加丰富和有效的特征表示。

本系统设计了一组实验，分别利用 ViT 模型的各个编码层及最终输出层的特征进行图像匹配任务，并计算了在不同特征表示下的平均精确率 Mean Average Precision (MAP) @1、3、5。实验结果如表 1 所示。

表 1 ViT 各个编码层及最终输出层的性能对比

特征来源	MAP@1	MAP@3	MAP@5
编码层 1	0.157	0.171	0.172
编码层 2	0.334	0.355	0.344
编码层 3	0.445	0.478	0.467
编码层 4	0.54	0.582	0.548
编码层 5	0.662	0.689	0.635
编码层 6	0.651	0.669	0.628
编码层 7	0.557	0.523	0.485
编码层 8	0.534	0.499	0.451
编码层 9	0.56	0.529	0.478
编码层 10	0.548	0.525	0.477
编码层 11	0.582	0.539	0.483
最终层	0.571	0.562	0.516

通过对比不同编码层及最终输出层的 MAP 值，发现第 4、5、6 编码层的输出在图像匹配任务中具有更优的性能。基于这一发现，本系统认为第 4、5、6 编码层能够提供多尺度的视觉特征，既包含了图像的低级特征，如纹理、边缘等，也融合了一定程度的高级特征，如图案的形状和布局等。因此，为了最大化这些编码层特征的利用效率，对第 4、5、6 编码层输出的特征张量进行拼接融合，从而生成一个优化和加强了的特征表示。

#### 3.2.3 自监督预训练策略

为了在自动布匹订单匹配任务中有效提升模型性能，本系统采用了自监督预训练方法，具体应用动态对比学习框架 MoCo V3。该框架通过构建和维护一个大规模的动态负样本队列，极大地丰富了训练过程中的负样本多样性，从而促进了模型在正负样本间建立更加精确的区分能力。

负样本队列的更新策略是 MoCo V3 训练的核心之一。在每次训练迭代中，最新生成的特征向量将会被加入队列的末端，而队列的头部最旧的特征向量将会被移除，以保持队列的固定大小。这个过程可以通过以下公式描述：

$$Q_{new} = \text{Enqueue}(Q_{old}, f_{new}) - \text{Dequeue}(Q_{old}) \quad (5)$$

其中:  $Q_{new}$  和  $Q_{old}$  分别表示更新后和更新前的负样本队列,  $f_{new}$  是新生成的特征向量。Enqueue 操作负责将新的特征向量添加到队列末尾, 而 Dequeue 操作负责从队列头部移除最旧的特征向量。通过这种负样本队列的构建与利用方式, 本系统的模型可以有效地进行大规模、高质量的对比学习。

在自监督预训练过程中, 采用了 InfoNCE 损失函数<sup>[17]</sup>, 这是对比学习中的一种关键损失函数, 用于优化模型的特征表示能力。InfoNCE 损失旨在拉近正样本对之间的距离, 同时推远负样本对之间的距离, 从而增强模型对正负样本的区分能力。InfoNCE 损失的数学定义如下:

$$L_{\text{InfoNCE}} = -\log \frac{\exp(\text{sim}(f_q, f_k)/\tau)}{\sum_{i=0}^N \exp(\text{sim}(f_q, f_m)/\tau)} \quad (6)$$

其中:  $f_q$  和  $f_k$  分别代表查询样本和正样本的特征表示,  $f_m$  代表负样本的特征表示,  $\text{sim}(\cdot)$  表示特征之间的相似度计算函数, 本系统使用余弦相似度,  $\tau$  表示温度参数, 用于控制相似度分数的尺度,  $N$  是负样本的数量。

在实际应用中, 正确选择温度参数  $\tau$  对于模型性能有重要影响。较小的  $\tau$  值会使得相似度分数的分布更加尖锐, 强化了正负样本对间差异的学习; 而较大的  $\tau$  值则会使得分布更加平滑, 可能减弱模型的判别力。理论上,  $\tau$  的值取决于具体的数据集和任务。对于印花布匹图像特征的学习,  $\tau$  值在 0.1~0.5 之间时模型性能较好。

## 4 实验结果与分析

实验对几种图像匹配技术进行分析和比较, 包括 VGG<sup>[18]</sup>、ResNet<sup>[19]</sup>, 以及原始的 ViT 模型, 还有本系统设计的多层特征融合机制和自监督训练的 ViT 模型。

### 4.1 实验平台与训练参数

实验的硬件平台为高性能服务器, 其中操作系统为 Ubuntu20.04, 搭载具有 24 GB 显存的 RTX3090 显卡, CUDA 版本为 11.1, Pytorch 版本为 1.8.2, 具体服务器配置参数表 2 所示。

表 2 服务器配置参数

名称	参数
系统	Ubuntu20.04
CPU	Intel i9-12 900 k
GPU	RTX3090
内存	64 G
Pytorch 版本	1.8.2
CUDA 版本	11.1

本系统构建的实验数据集汇集了来自多个生产批次的高分辨率布匹图像, 总计 1 000 张, 覆盖 200 种独特的图案与颜色, 通过工业相机在统一的光照环境下精确采集。为贴近实际应用场景, 对数据集执行了一系列预处理步骤, 包括随机裁剪、旋转和添加噪声等数据增强方法<sup>[20]</sup>。

评估模型性能所采用的关键指标是 MAP。MAP 综合

考量了匹配结果的排序质量, 通过在不同的匹配深度 (值设定为 1、3、5) 下计算 MAP, 本系统全面评估了以图搜图系统的整体性能。MAP 作为评价指标, 尤其在实际应用中对于快速获得最关联信息的需求背景下显示出其重要性, 它反映了系统在前个最相关图像匹配中的准确性。在评估过程中, 订单花型图像采用缩放、镜像、裁剪等方式进行扩充。

在网络训练方面, 选择 Adam 优化器, 初始学习率设定为  $1.5 \times 10^{-4}$ , 批处理大小 (batchsize) 为 32, 整个训练过程迭代 200 个周期 (epoch)。实拍图像正负样本数据增强同样采用随机裁剪、随机旋转及水平翻转等操作, 旨在增强模型对于图像变化的适应性和泛化能力。另外, 网络输入图像的分辨率统一设置为  $224 \times 224$  像素。

### 4.2 本文方法与现有特征提取方法的对比

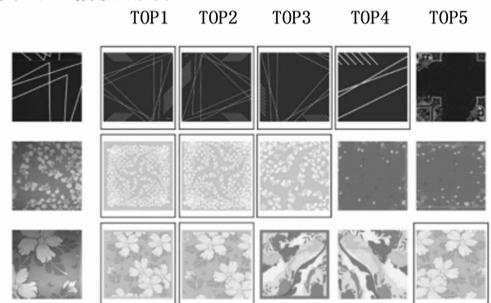
对比参数如表 3 所示。

表 3 多种特征提取网络的性能对比

特征来源	MAP1	MAP3	MAP5
VGG	0.454	0.464	0.432
ResNet	0.552	0.559	0.508
ViT	0.571	0.562	0.516
Ours	0.706	0.719	0.678

作为深度学习中的经典模型, VGG 和 ResNet 在图像识别任务上表现突出。尤其是 ResNet, 其深层次的残差连接有助于捕获更复杂的图像特征。然而, 这两种模型仍然受限于固定的架构, 难以动态地适应图像特征的多样性。ViT 作为一种 Transformer 模型能够通过自注意力机制有效地捕获了全局图像特征, 但在未采用多层特征融合的情况下, 其对于细节特征的捕获仍有限。本系统的模型通过结合多层特征融合机制和 MoCo V3 自监督训练策略, 实现了对布匹图像特征的深层次和细粒度的捕获。该模型不仅提高了对正样本的敏感度, 同时通过对比学习损失函数减小了与负样本的相似度, 从而在 MAP1、MAP3 和 MAP5 上均取得了最优的性能。

如图 7 所示, 展示了本系统设计的图像匹配网络在数据集上的测试结果。从图中可以看出本章设计网络能够有效地识别出目标订单图片, 并且所匹配到的订单图像往往具有较高的置信度排名。



(a) 查询图片

(b) 匹配图片

图 7 可视化匹配结果

### 4.3 基于实际生产环境评价指标与实验结果

#### 4.3.1 实际生产环境中的评价指标

传统上, MAP 被广泛用于衡量匹配系统的效能, 它反映了数据库中所有正确图片在排序列表前端的平均程度。然而, 尽管 MAP 为研究提供了一个全面的性能评估, 它可能不完全符合实际生产环境中的需求。在本系统的自动报工软件实际生产环境应用场景中, 工人更关注于是否能在前三的匹配结果中找到对应的订单图片, 因此, 需要一种更贴近实际应用场景的评价指标来准确评估匹配系统的性能。

针对上述需求, 设计了一种新的评价指标: TOP-3 命中率。该指标定义为: 在所有匹配任务中, 匹配结果的前 3 名里至少包含一个正确结果的比例。具体而言, 如果匹配系统在前 3 个返回的图像中至少返回一个用户寻找的订单图片, 则认为该匹配任务成功。该指标直接对应于实际生产环境中用户的使用场景, 更加注重匹配结果的即时效用。

#### 4.3.2 实验设计与结果分析

为了验证自动报工系统的性能, 在真实的生产线环境中部署了该系统。测试集包括不同类型、不同图案的布匹图像, 以及对应的生产订单数据。通过系统对这些数据的处理与匹配, 计算了 TOP-3 命中率作为评价指标。

表 4 本文设计方法与现有特征提取网络的性能对比 %

模型	TOP-1 命中率	TOP-3 命中率
VGG	51.3	65.2
ResNet	56.3	75.6
ViT	57.1	85.36
Ours	70.6	90.14

实验结果表明, 本文自动报工系统的方法在实际生产环境中展现出了良好的性能。经过测试, 本文设计方法的 TOP-1 命中率达到 70.6%, TOP-3 命中率达到 90.14%, 明显高于现有特征提取方法。这意味着在绝大多数情况下, 系统能够在前 3 个匹配结果中找到正确的订单图像。

为了进一步提升系统的可靠性, 系统引入了后续的人工核实环节。每日整个工厂所有自动报工流程结束后, 专业工人会对系统的匹配结果进行核实。这一步骤虽然引入了人工参与, 但由于系统已经大幅度缩小了核实范围 (仅限于 TOP-3 的结果), 因此大大减轻了工人的工作量, 并且最小化了误判率。如图 8 所示, 为印染工厂中自动报工的匹配结果。从图中可以看出, 该检测算法对于摄像头拍摄的布匹花型图片能够进行有效的匹配, 并且匹配软件可以直观地显示出匹配结果。

## 5 结束语

本文设计并实现了一个基于机器视觉的数码印花工厂自动报工系统。首先, 基于 C/S 架构设计了计算资源与用户资源分离的软件结构。客户端的海康工业面阵相机首先采集工业流水线上的布匹印花图像后对图像进行滤波、归一化等预处理。然后利用 HTTP 协议将图片与生产单信息发送至服务端进行神经网络推理运算, 利用特征融合改进



图 8 自动报工软件在实际生产环境中部署与测试

的 ViT 神经网络进行布匹图像特征提取, 并将之与订单库中生产单花型图案所提取的特征向量进行匹配, 找出待报工订单。最后将匹配结果进行服务端数据库管理后报工, 将结果返回客户端进行用户展示与确认操作。实验结果表明, 该订单匹配系统 TOP-3 命中率为 90.4%, 满足工业生产需求。

#### 参考文献:

- [1] 路虹. 世经弱势, 新兴经济体更显成长性 [J]. 国际商报, 2024: 4-4.
- [2] LIU Z, LIU S, LI C, et al. Fabric defects detection based on SSD [C] //International Conference on Graphics and Signal Processing (ICGSP). ACM, 2018: 74-78.
- [3] ZHANG H, TAN Q, LU S, et al. Yarn-dyed fabric defect detection using U-shaped de-noising convolutional auto-encoder [C] //Data Driven Control and Learning Systems Conference (DDCLS). IEEE, 2020: 18-24.
- [4] LIU Z, NING J, CAO Y, et al. Video swin transformer [C] //International Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022: 3202-3211.
- [5] ORFALI R, HARKEY D, EDWARDS J. Client/server survival guide [M]. New York: John Wiley & Sons, 2007.
- [6] BLANCHETTE J, SUMMERFIELD M. C++ GUI programming with Qt 4 [M]. Upper Saddle River, NJ: Prentice Hall in Association with Trolltech Press, 2008.
- [7] ROSENBLATT J K. DAMN: A distributed architecture for mobile navigation [J]. Journal of Experimental & Theoretical Artificial Intelligence, 1997, 9 (2/3): 339-360.
- [8] JEONG E J, KIM J, HA S. Tensorrt-based framework and optimization methodology for deep learning inference on jets on boards [J]. ACM Transactions on Embedded Computing Systems, 2022, 21 (5): 1-26.
- [9] 毛亚青, 王亮, 胡俊峰. 基于 Flink 的海量医学图像检索系统设计与实现 [J]. 计算机测量与控制, 2020, 28 (9): 212-217.
- [10] KHAN S, NASEER M, HAYAT M, et al. Transformers in vision: a survey [J]. ACM Computing Surveys, 2022, 54 (10): 1-41.
- [11] CHEN X, XIE S, HE K. An empirical study of training self-supervised vision transformers [C] //Proceedings of the

