# 行人重识别展示系统

用户手册

## 1 引言

行人重识别（Pedestrian Re-identification）也称行人再识别，简称 Re-ID，是利用计算机视觉技术判断图像或视频序列中是否存在特定行人的技术，被广泛认为是一个图像检索的子问题。给定一个待查询的对象，Re-ID 的目标是检索跨设备下的该行人的图像。待查询对象可以用图像、视频序列、文本描述等方式来表示。

在监控视频中，要找到某一个人，我们首先想到的是人脸识别技术，而人脸识别技术限定了图像的最低分辨率，然而由于相机分辨率和拍摄角度的缘故，通常无法得到质量非常高的行人人脸图片，并且这种情况在日常生活中是占绝大多数的。在人脸识别失效的情况下，Re-ID 就成为了一个非常重要的替代品技术。它可以弥补目前固定摄像头的视觉局限，并可与行人检测、行人跟踪等技术相结合，应用于视频监控、智能安防等领域。

本系统旨在基于行人重识别系统，实现一个操作简单、便于观察的行人重识别任务的展示系统。

### 1.1 编写目的

（1）本手册将为使用行人重识别展示系统的用户提供使用功能参考；

（2）本系统基于 ResNet50 和 DeiT 的深度学习网络模型、Pytorch 开发框架和 Django web 网络框架，完成了行人重识别展示系统的开发和实现。用户进入系统后，需从系统提供的 model、dataset 以及 gallery_num 参数中进行选择，然后系统会根据提交的参数值，从 dataset 中进行随机采样，并其中待检索行人图片（query）默认采样一张，行人图像库图片（gallery）采样的数量则根据 gallery_num 参数选择。最后深度学习网络模型会计算 query 图片与 gallery 图片的相似度并将其展示在对应的板块中。系统操作简单，界面简洁，用户可以随时随地进行操作。

（3）行人重识别展示系统使用深度学习中的 CNN 及 Transformer 网络框架构建模型，使得用户只需要关注采样所需要的参数以及检索图像本身，而不用担心识别过程，促使原本繁琐的工作简单化。

### 1.2 背景

（1）用户手册所描述的软件系统的名称：行人重识别展示系统

（2）项目的任务提出者：北京林业大学；

（3）项目的开发者：北京林业大学；

（4）项目的用户（或首批用户）：北京林业大学信息学院；

（5）项目著作权人：北京林业大学。

### 1.3 开发工具及技术

Python：Python 提供了高效的高级数据结构，还能简单有效地面向对象编程。Python 语法和动态类型，以及解释型语言的本质，使它成为多数平台上写脚本和

快速开发应用的编程语言，随着版本的不断更新和语言新功能的添加，逐渐被用于独立的、大型项目的开发。Python 是一种效率极高的语言，其语法有助于创建整洁的代码，相比其他语言，使用 Python 编写的代码更容易阅读、调试和扩展。Python 有很多方面的应用：游戏、web 应用程序、解决商业问题，而且在科学领域被大量用于学术研究和应用研究。

Pytorch：一个开源的 Python 机器学习库，由 Facebook 人工智能研究院（FAIR）基于 Torch 推出了 PyTorch。它是一个基于 Python 的可续计算包，提供两个高级功能：①具有强大的 GPU 加速的张量计算；②包含自动求导系统的深度神经网络。Pytorch 可高效的构建深度神经网络，被用于人工智能前沿技术及算法的研究。

Django：Django 是一个高级的 Python 网络框架，可以快速开发安全和可维护的网站，并进一步开发出全功能的 Web 服务。其有以下优点：①完备性：遵循“功能完备”的理念，提供开发人员可能想要“开箱即用”的几乎所有功能；②通用性：可以（并已经）用于构建几乎任何类型的网站；③安全性：通过提供一个被设计为“做正确的事情”自动保护网站的框架来避免许多常见的安全错误；④可扩展：Django 使用基于组件的“无共享”架构；⑤可维护：Django 代码编写是遵照设计原则和模式，鼓励创建可维护和可重复使用的代码等。

ResNet50：随着 CNN 的不断发展，为了获取深层次的特征，卷积的层数也越来越多，网络层数到达一定的深度之后，再增加网络层数，那么网络就会出现随机梯度消失的问题，也会导致网络的准确率下降。为了解决这一问题，传统的方法是采用数据初始化和正则化的方法，这解决了梯度消失的问题，但是网络准确率的问题并没有改善。ResNet 网络的关键就在于其结构中的残差单元，它不仅可以解决梯度问题，而且网络层数的增加也使其表达的特征也更好，相应的检测或分类的性能更强，再加上残差中使用了 1×1 的卷积，这样可以减少参数量，也能在一定程度上减少计算量。Resnet50 网络就是包含了 49 个卷积层和一个全连接层的 ResNet 网络。

DeiT：DeiT 是一个全 Transformer 的网络架构。其核心是针对 ViT 训练数据巨大，超参数难设置导致训练效果不好的问题提出了教师-学生蒸馏训练策略，使得 DeiT 能在大幅减少训练所需的数据集和训练时长的情况下依旧能够取得很不错的性能。除此之外还提出了 token-baseddistillation 方法，允许模型从教师网络的输出中学习，就像在常规的蒸馏中一样，同时也作为一种对 classtoken 的补充，使得 Transformer 在视觉领域训练得又快又好。

### 1.4 参考资料

[1] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[2] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[3] Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image

recognition at scale." arXiv preprint arXiv:2010.11929 (2020).
[4] Touvron, Hugo, et al. "Training data-efficient image transformers & distillation through attention." International conference on machine learning. PMLR, 2021.
[5] He, Shuting, et al. "Transreid: Transformer-based object re-identification." Proceedings of the IEEE/CVF international conference on computer vision. 2021.
[6] Zhang, Yundong, Huiye Liu, and Qiang Hu. "Transfuse: Fusing transformers and cnns for medical image segmentation." Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. Springer International Publishing, 2021.
[7] Cheng, De, et al. "Person re-identification by multi-channel parts-based cnn with improved triplet loss function." Proceedings of the iEEE conference on computer vision and pattern recognition. 2016.
[8] Li, Yulin, et al. "Diverse part discovery: Occluded person re-identification with part-aware transformer." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
[9] Liao, Shengcai, and Ling Shao. "Transmatcher: Deep image matching through transformers for generalizable person re-identification." Advances in Neural Information Processing Systems 34 (2021): 1992-2003.
[10] Yu, Rui, et al. "Cascade transformers for end-to-end person search." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
[11] Cao, Jiale, et al. "Pstr: End-to-end one-step person search with transformers." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
[12] Han, Kai, et al. "Vision gnn: An image is worth graph of nodes." arXiv preprint arXiv:2206.00272 (2022).
[13] Jin, Xin, et al. "Cloth-changing person re-identification from a single image with gait prediction and regularization." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
[14] Zheng, Anlin, et al. "Progressive end-to-end object detection in crowded scenes." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
[15] Li J, Wang Z, Pan Z, et al. Looking at Boundary: Siamese Densely Cooperative Fusion for Salient Object Detection[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021.
[16] Guo, Jianyuan, et al. "Distilling object detectors via decoupled features." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[17] VS, Vibashan, et al. "Mega-cda: Memory guided attention for category-aware unsupervised domain adaptive object detection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
[18] Garnot, Vivien Sainte Fare, and Loic Landrieu. "Panoptic segmentation of satellite image time series with convolutional temporal attention networks." Proceedings of the IEEE/CVF International Conference on Computer Vision . 2021.
[19] Li, Dangwei, et al. "Learning deep context-aware features over body and latent parts for person re-identification." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
[20] Su, Chi, et al. "Pose-driven deep convolutional model for person re-identification." Proceedings of the IEEE international conference on computer vision. 2017.
[21] Suh, Yumin, et al. "Part-aligned bilinear representations for person re-identification." Proceedings of the European conference on computer vision (ECCV). 2018.
[22] Sun, Yifan, et al. "Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)." Proceedings of the European conference on computer vision (ECCV). 2018.

# 2 应用概述

## 2.1 网络模型

本系统首先设计了一个行人重识别网络模型，该模型可以对行人图像进行特征提取以及相似度度量，并将此模型集成于行人重识别展示系统中，用于行人重识别任务的展示。

除了最基础的 ResNet50 和 DeiT 模型之外，本系统将两者进行融合，其中 CNN 分支提取从局部到全局的特征，Transformer 分支则从全局自注意力开始，逐步恢复局部细节。然后将两个并行的分支中提取的具有相同分辨率的特征通过如像素相加等方式进行融合，作为最终的特征表示。

## 2.2 系统集成

我们使用 Django 框架设计了一个交互式系统，将上述网络模型集成于 Web 系统中，可以通过选择不同的 model、dataset 以及 gallery_num 参数来干预模型的采样过程，从而得到不同行人之间的相似度。

# 3 运行环境

## 3.1 硬件环境

CPU：2.8GHz
GPU 显存：12G

## 3.2 软件环境

操作系统：Ubuntu18.04 以上或 Windows
Python: 3.8.0

Cuda: 11.3
Pytorch: 2.0.1
Django: 2.2

# 4、操作说明

进入网页系统后，界面由两部分组成，参数选择模块和展示模块。

## 4.1 参数选择模块

如图 1 所示，参数选择模块中设置了 model 参数的选择、dataset 参数的选择以及 gallery_num 参数的选择，用于对采样及识别过程进行干预。

其中，model 参数包含三种模型，如图 2 所示，分别是 ResNet50、Deit 和 FFusion，其中 FFusion 为我们自己设计的深度学习模型。该模型作为行人重识别过程中的深度学习网络框架，用于对图像进行特征提取以及相似度度量。

dataset 参数包含两个数据集，如图 3 所示，分别是 Market1501 和 DukeMTMC-reID，这两个数据集是行人重识别领域的常用数据集，用于为模型提供行人图像数据。

gallery_num 参数包含 5 和 10 两个选择，如图 4 所示，该参数表示选择多少张行人图像组成 gallery 集，并与 dataset 参数一起作为采样过程的参考，上传参数后，系统会从选择的 dataset 数据中随机采样 gallery_num 张图像组成 gallery 集，随机采样一张图片作为 query 待查询图像。

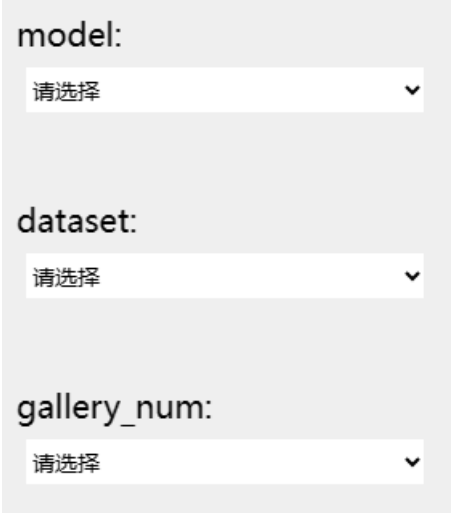


图 1 参数选择模块


图 2 model 参数选择

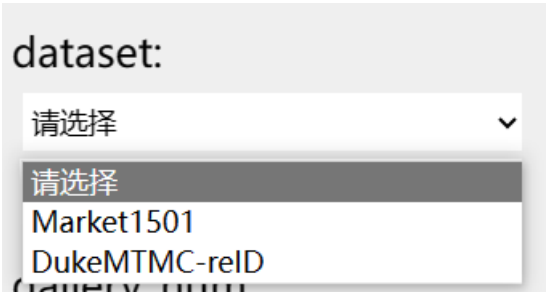


图 3 dataset 参数选择

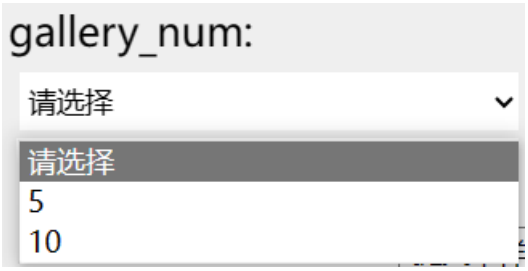


图 4 gallery_num 参数选择

本系统将 model 参数、dataset 参数、gallery_num 参数均设置为必选项，若某一项没有选择，则会出现错误提示，如图 5 所示。

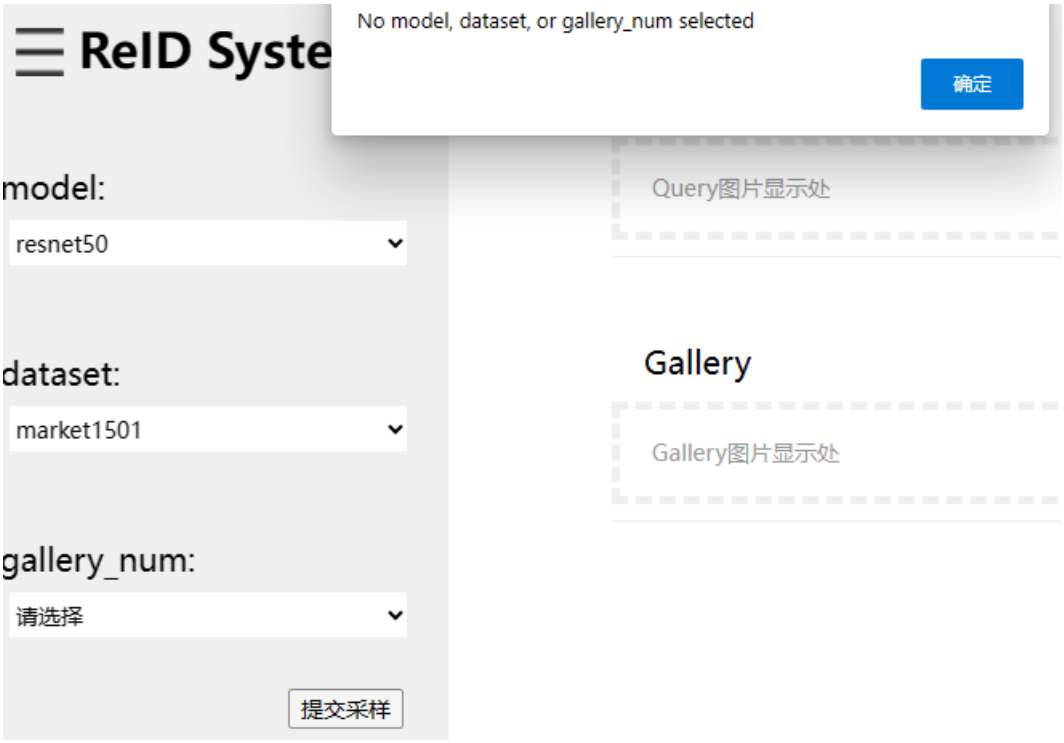


图 5 参数未选择错误提示

## 4. 2 展示模块

系统的展示模块如图 6 所示。其中，Query 部分用来展示采样的 query 图片，Gallery 部分用来展示采样的 gallery 集中的图片，并在每一张 gallery 图片的下方展示其与 query 图片的相似度参数，数值越小，则表示越相似。

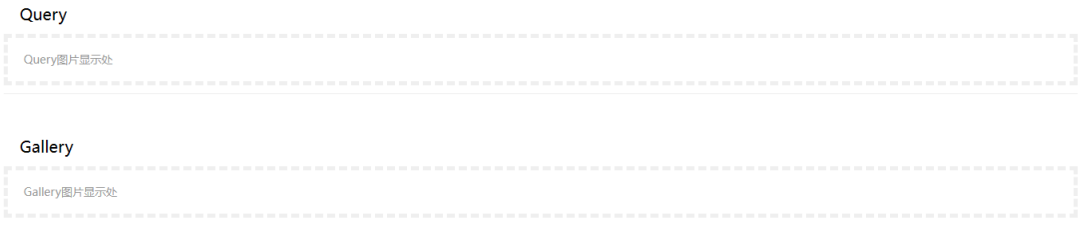


图 6 展示模块

选择 model 参数、dataset 参数、gallery_num 参数后，点击“提交采样”按钮，系统将自动从选择的 dataset 数据集中进行随机采样，其中，待检索行人图片（query）默认采样一张，行人图像库图片（gallery）采样的数量则根据 gallery_num 参数选择，并采样的 gallery_num 张图像组成 gallery 集。采样的 query 图片以及 gallery 集图片将存放在指定的文件夹中，等待使用。保存本次采样的图片之前，为了方便展示，系统会自动将上一次的采样结果删除。

采样成功后，系统会使用选择的 model 对采样图片进行特征提取以及相似度计算。首先，系统对 query 图像进行特征提取，然后，对 gallery 集中的每一张图像进行特征提取，并计算其与 query 图像的相似度，系统采用欧式距离作为相似度度量，距离越小，则表示越相似。最后，系统会将采样的 query 图和 gallery 图展示在对应的区域中。

例如，我们选择 Resnet50、Market1501 以及 5 分别作为 model、dataset 以及 gallery_num 的参数值，点击“提交采样”按钮后，系统会将采样的 query 图和 gallery 图，以及每一张 gallery 图与 query 图的相似度分别展示在对应的区域中，如图 7 所示。

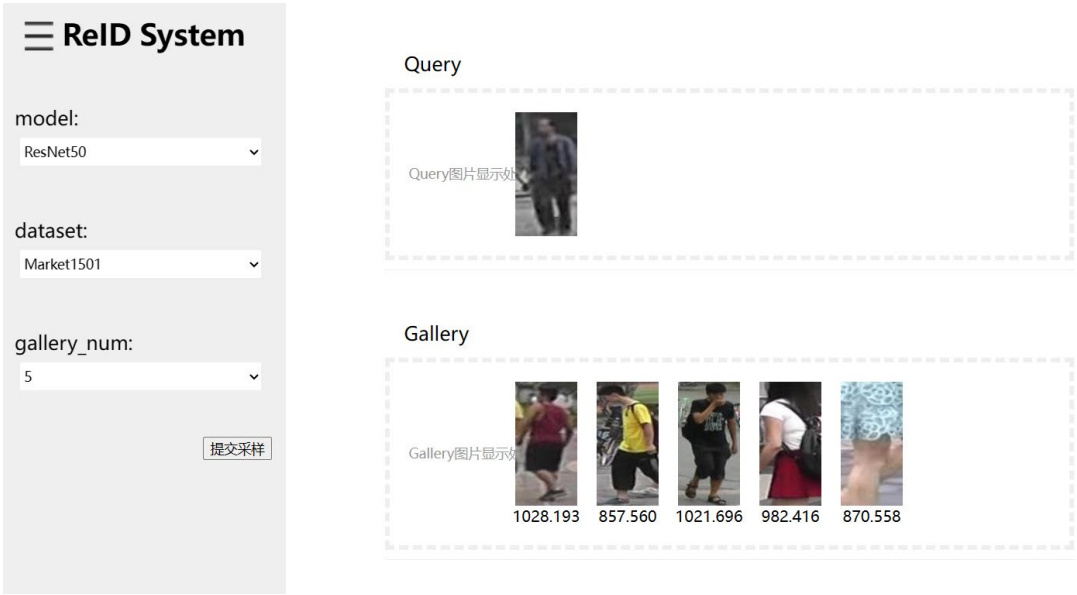


图 7 采样结果展示