智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

基于Cluster Contrast和ViT的东北虎个体重识别无监督学习框架研究

赵亚凤, 于继超, 孙骞, 康嘉璐, 王梓丞

引用本文:
赵亚凤, 于继超, 孙骞, 等. 基于Cluster Contrast和ViT的东北虎个体重识别无监督学习框架研究[J]. 智能系统学报, 2026, 21(2): 435-443.
ZHAO Yafeng, YU Jichao, SUN Qian, et al. Unsupervised framework for Amur tiger re-identification with Cluster Contrast and ViT[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(2): 435-443.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202507012

您可能感兴趣的其他文章

基于图嵌入的自适应多视降维方法
An adaptive multi-view dimensionality reduction method based on graph embedding
智能系统学报. 2021, 16(5): 963-970 https://dx.doi.org/10.11992/tis.202105021

基于二进制生成对抗网络的视觉回环检测研究
Visual loop closure detection based on binary generative adversarial network
智能系统学报. 2021, 16(4): 673-682 https://dx.doi.org/10.11992/tis.202007007

基于反馈注意力机制和上下文融合的非模式实例分割
Feedback attention mechanism and context fusion based amodal instance segmentation
智能系统学报. 2021, 16(4): 801-810 https://dx.doi.org/10.11992/tis.202007042

自步稀疏最优均值主成分分析
Sparse optimal mean principal component analysis based on self-paced learning
智能系统学报. 2021, 16(3): 416-424 https://dx.doi.org/10.11992/tis.201911028

高斯核函数卷积神经网络跟踪算法
Convolutional neutral network tracking algorithm accelerated by Gaussian kernel function
智能系统学报. 2018, 13(3): 388-394 https://dx.doi.org/10.11992/tis.201612040

行人重识别研究综述
Survey on pedestrian re-identification research
智能系统学报. 2017, 12(6): 770-780 https://dx.doi.org/10.11992/tis.201706084

DOI: 10.11992/tis.202507012
网络出版地址: https://link.cnki.net/urlid/23.1538.TP.20251224.1514.005

基于 Cluster Contrast 和 ViT 的东北虎个体重识别无监督学习框架研究

赵亚凤[1], 于继超[1], 孙骞[2], 康嘉璐[1], 王梓丞[1]
(1. 东北林业大学 计算机与控制工程学院, 黑龙江 哈尔滨 150040; 2. 哈尔滨工程大学 信息与通信工程学院, 黑龙江 哈尔滨 150001)

摘 要: 针对东北虎个体重识别中野外数据标注困难、样本失衡等挑战, 以野外东北虎 (Amur tiger re-identification in the wild, ATRW)数据集为基础提出一种“全局特征提取−空间位置强化−无监督均衡训练”的协同框架, 完成无监督重识别。使用 vision Transformer(ViT)自注意力机制捕捉东北虎条纹的全局长距离的依赖特征, 通过坐标注意力机制加强对条纹空间位置的精确解析, 解决传统卷积神经网络局部性导致的特征关联缺失问题。引入 Cluster Contrast 机制构建簇级内存字典, 通过动量更新平衡不同样本量东北虎的特征优化速率, 避免无监督学习中样本失衡导致的特征偏差。实验表明, 模型在 ATRW(r+i)数据集上平均精度的值为 86.4%, 高于原有的特征提取 ViT 和 Resnet50_ibn 模型, 对不同数据分布和数据量具有良好泛化能力, 适配野外可见光/红外多设备协同监测需求。本文所提方法为东北虎个体识别提供了兼具准确性与鲁棒性的技术方案。

关键词: 东北虎; 重识别; 深度学习; 无监督学习; ATRW 数据集; Cluster Contrast 机制; vision Transformer; 坐标注意力
中图分类号: TP391.4 **文献标志码**: A **文章编号**: 1673−4785(2026)02−0435−09

中文引用格式: 赵亚凤, 于继超, 孙骞, 等. 基于 Cluster Contrast 和 ViT 的东北虎个体重识别无监督学习框架研究 [J]. 智能系统学报, 2026, 21(2): 435–443.
英文引用格式: ZHAO Yafeng, YU Jichao, SUN Qian, et al. Unsupervised framework for Amur tiger re-identification with Cluster Contrast and ViT[J]. CAAI transactions on intelligent systems, 2026, 21(2): 435–443.

Unsupervised framework for Amur tiger re-identification with Cluster Contrast and ViT

ZHAO Yafeng[1], YU Jichao[1], SUN Qian[2], KANG Jialu[1], WANG Zicheng[1]
(1. College of Computer and Control Engineering, Northeast Forestry University, Harbin 150040, China; 2. College of Information And Communication Engineering, Harbin Engineering University, Harbin 150001, China)

Abstract: To address challenges like difficult annotation of wild data and sample imbalance in Amur tiger individual re-identification, a collaborative framework based on the ATRW dataset was developed for unsupervised re-identification. The ViT self-attention mechanism was used to capture global long-distance dependent features of stripes. Combined with the coordinate attention mechanism, it enhanced spatial position analysis of stripes, making up for missing feature correlations caused by the locality of traditional CNNs. The Cluster Contrast mechanism was introduced to build a cluster-level memory dictionary. Momentum update balanced feature optimization rates of individuals with different sample sizes, alleviating feature biases from sample imbalance in unsupervised learning. Experiments show the model achieves an mAP of86.4% on the ATRW (r+i) dataset, outperforming the original feature extraction models (ViT and Resnet50_ibn). It exhibits good generalization ability across different data distributions and data volumes, and is suitable for the demand of collaborative monitoring with visible light/infrared multi-devices in the wild. The method proposed in this study provides a technical solution with both accuracy and robustness for Amur tiger individual identification.

Keywords: Amur tiger; re-identification; deep learning; unsupervised learning; ATRW dataset; Cluster Contrast mechanism; vision Transformer; coordinate attention

东北虎又称西伯利亚虎, 不仅是现存体重最大的肉食性猫科动物, 在自然界的食物链中占据着顶端的位置, 更是我国一级重点保护野生动物[1],

收稿日期: 2025−07−12. **网络出版日期**: 2025−12−25.
基金项目: 国家自然科学基金项目 (32371864).
通信作者: 王梓丞. E-mail: wangzicheng1992@163.com.

©《智能系统学报》编辑部版权所有

其主要分布在俄罗斯远东地区、中国东北以及朝鲜边境等地[2]。由于栖息地破坏、非法捕猎等因素，东北虎的种群数量急剧减少，已被列为世界濒危物种[3]。为了制定科学有效的保护策略，需要实时监测野生东北虎，准确掌握其个体数量、分布范围以及活动规律等信息。传统的东北虎监测方法，如 DNA、足迹追踪、粪便分析和气味等[4-7]，普遍存在效率低、准确性差并且可能给研究人员以及野生动物带来难以预估的危险等问题[8]，难以满足现代保护工作的需求。随着计算机视觉与人工智能的普及，基于机器视觉的野生动物监测在多个物种中取得了显著进展，其首要任务是野生东北虎的个体识别。

多年来，个体重识别作为一个重要的研究领域引起了相当大的关注[9]，扩大了物体检测、跟踪、识别等任务的应用范围。它在智能监控、智慧城市和自然生态系统保护等领域具有重要的实际应用价值。近年来，重识别领域的研究，特别是涉及人员和车辆等主题的研究，得到了深刻的发展，并在传统环境下取得了显著的成功[10]，张国印等[11] 基于 FairMOT 框架，加入轻量化平衡模块、窗口注意力网络并设计遮挡恢复算法，相比原 FairMOT 高阶跟踪精度提升 1.5 百分点。王路遥等[12] 构建双流网络，结合多尺度特征互补模块与混淆学习策略，联合识别损失与异质中心三元组损失，在 SYSU-MM01 全搜索模式下 R-1 达 76.69%、平均精度值（mAP）达 72.45%，RegDB 可见光到红外模式下 R-1 达 94.62%、mAP 达 94.60%。在动物识别中，也得到广泛研究，在奶牛个体识别中，赵玲等[13] 基于改进的 Mask R-CNN（region based convolutional nerual network）算法，结合牛脸和驱赶综合信息在自建的数据集上进行奶牛个体识别，准确率高达 93.63%。在黑猩猩个体识别中，Freytag[14] 等使用 C-Zoo 和 C-Tai 黑猩猩（Pan troglodytes）数据集，训练 AlexNet 卷积神经网络，在 C-Zoo 数据集和 C-Tai 数据集上的识别准确率分别达到 75.66% 和 91.99%。在东北虎重识别领域，相关研究主要是根据提取东北虎体表极具标志性的条纹特征[15] 进行个体重识别，这些条纹如同人类指纹，具备唯一性和稳定性[16]。在早期的方法中，深度卷积神经网络（convolutional nerual network, CNN）凭借其强大的特征提取能力，在东北虎重识别任务中取得了显著进展，例如，Shi 等[17] 构建深度卷积神经网络实现东北虎个体自动识别，但该方法并未在复杂环境下应用；马光凯等[18] 首次利用 Transformer 方法对东北虎个体识别进行了成功的尝试，该方法与 CNN 相比，获得了更好的识别性能，但以上方法都需要依赖大量人工标注的东北虎图像数据，都是基于有监督的学习，而野生东北虎重识别不符合依赖密集标注数据集的研究方法，因此必须用无监督学习。无监督学习中，Chmiela 等[19] 针对动物重识别（re-identification, Re-ID），先用 ISNet 和 SAM 去除背景，并基于 DVE 无监督学习特征，在 YakReID-103 等数据集上，指标均表现优异，证明了其有效性，但存在一定的不足：一是未充分考虑东北虎长距离依赖的条纹特征；二是未考虑样本失衡导致的特征优化差异问题，当不同动物个体样本量差距较大时，容易导致特征学习偏向样本量多的个体。

在东北虎重识别领域，得到大量标注数据相当困难，而且野外环境下的红外触发相机布设密度较小，加之东北虎的活动具有较强的随机性，不易长时间连续观察获得大量样本，而通过动物园野生动物园拍摄到的 ATRW 数据集又无法很好地在野外场景下适用，因此使用无监督重识别的方法来做解决该问题较好的办法。

由于东北虎体表的条纹特征十分明显，加之无监督学习不需要人工标注的优势。利用 vision Transformer（ViT）整体特征获取的能力，利用坐标注意力机制获得的空间解析优势，采用 Cluster Contrast 样本均衡技术的优点，构建了一套全面且具有针对性的识别方法体系。该方法体系从数据特点和模型需求出发，通过全局与局部特征的协同提取、空间位置信息的精准强化，以及样本失衡问题的动态平衡，目的是弥补现有东北虎个体识别方法存在的长距离特征无法联系起来、无法分辨其细节特点、存在较强的非标定误差等缺点，从而提高识别的准确度、稳定性和野外使用性，对东北虎保护实践具有核心支撑价值。其一，可以通过精准识别个体可实现种群数量动态统计，避免重复统计误差；其二，基于个体识别的活动轨迹追踪，能分析东北虎的栖息地偏好及迁徙规律，为保护区功能分区规划提供数据支撑，为东北虎的保护和监测工作做准备。

1　数据与方法

1.1　数据获取

数据主要来自 ATRW[20] 数据集，该数据集由世界野生动物基金会对多个大型野生动物园中的东北虎进行采集并构建，是目前最大的东北虎重识别数据集，其包含 92 只东北虎，并根据每只东北虎的不同侧面构建了 182 个实体。数据集由 107

个实体的 1 887 张图像的训练集和 75 个实体的 1 762 张图像的测试集构成, 其中包含不同拍摄角度、不同光照条件、不同姿态以及有部分被遮挡的东北虎图片, 此外, 还有一部分数据为东北虎豹国家公园所提供的红外伪彩色数据, 构成 ATRW (wild) 用于野外场景测试, 部分样本如图 1 所示。

(a) ATRW 数据集的样例　　(b) 红外伪彩色数据样例

图 1　东北虎重识别数据集样例

Fig. 1　Sample of the Amur tiger re-identification dataset

1.2　东北虎个体识别方法

无监督东北虎个体重识别架构分为 3 个部分: 第 1 部分为特征提取, 选用 Transformer 架构下的 ViT 作为骨干网络, 并引入坐标注意力机制 (coordinate attention, CA) 从而提升模型对复杂数据特征的处理能力; 第 2 部分为聚类, 使用传统的聚类方法 DBSCAN, 依据提取特征进行分类, 并给每个类别分配一个伪标签, 在训练初期, 伪标签并不是很准确, 但随着训练过程的进行, 网络的精确度会逐步提高, 伪标签会逐渐接近真实标签; 第 3 部分是类别特征的存储和更新, 在网络训练的过程中, 根据网络参数的变化, 对每个类别的特征进行更新。具体的无监督东北虎个体重识别架构如图 2 所示。

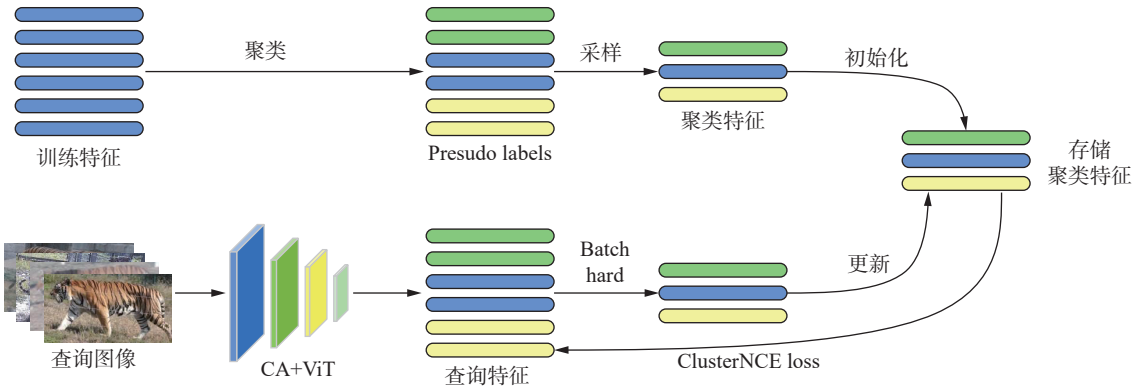

图 2　无监督东北虎个体重识别架构

Fig. 2　Unsupervised Amur tiger individual weight recognition architecture

图 2 中, 具有相同颜色的特征向量属于同一类。上部是内存初始化阶段, 下部是模型训练阶段。图片特征的存储和更新对于网络的训练影响很大, 在东北虎重识别数据集 ATRW 中, 不同的东北虎个体的样本量存在显著差异, 若采用传统“按样本量累计更新簇特征”的方式, 会导致样本量多的东北虎个体特征更新滞后, 进而阻碍网络整体优化。Cluster Contrast 机制的核心思想是无论一只东北虎有多少张图片, 对于训练网络来说, 它们都是一视同仁的, 都是用同一个速度去更新特征。选择该机制更新特征, 同时引入 Hard Example Mining 方法 (即优先选取与内存中已有簇特征最不相似的样本特征, 进行簇特征更新), 设计基于动量更新的簇特征迭代策略用于内存更新, 并使用 ClusterNCE 损失用于计算质量查询特征和聚类特征之间的损失。

1.2.1　全局特征提取

目前, Transformer[21] 在计算机视觉领域也取得了显著的成功, 用来解决计算机视觉领域的图像分类任务、目标检测任务、图像分割任务等。Transformer 最早是由 Google 在 2017 年针对自然语言处理任务 (natural language processing, NLP) 提出的神经网络架构, 通过自注意力机制能够有效地捕捉长距离依赖关系。

在东北虎个体重识别中, 模型实现个体精准识别的关键因素是能够获得全面且具有代表性的特征, 在东北虎图像中, 其条纹分布具有条纹从躯干延伸至四肢及尾部, 拥有跨区域特征的全局分布模式, 空间位置相隔较远的条纹区域之间存在的固有关联信息, 传统 CNN 通过固定尺寸卷积核逐步堆叠提取特征, 其感受野扩展效率低, 且在特征传递中易丢失远距离区域的关联细节, 导致无法充分学习到这类长距离依赖特征, 进而影响个体识别的准确性。但使用 ViT[22] 做骨干网络能够直接把图像划分成很多的小块 (patch), 并将每个 patch 映射为一个序列嵌入, 这样就可以直接从全局视角出发去提取东北虎图像整体特征。随后, 对提取的整体特征进行位置编码 (position

embedding)，即将每一个 patch 的空间位置信息加入进特征序列当中，使模型可以捕获到条纹间的相对位置关系，从而更好地理解图中不同区域间的相对位置和布局，有助于提高模型提取图像特征的能力。

引入多头自注意力机制(multi-head self-attention)后，其能够捕捉图像不同的位置之间的长距离依赖关系，不受卷积操作局限性影响，更好地掌握图像的全局结构并捕捉图像中多尺度的细节信息，一部分关注躯干条纹的大致走向，另一部分关注腿部条纹的大致趋势，最终获得全局结构加局部细节的复合特征，同时多头自注意力均能通过跨区域关联提取鲁棒特征，利用未遮挡的条纹推断被植被遮挡的条纹走向，这种设计适用于完整的东北虎全身图及不完整带有部分遮挡的东北虎图像。Vision Transformer 网络架构如图 3 所示。

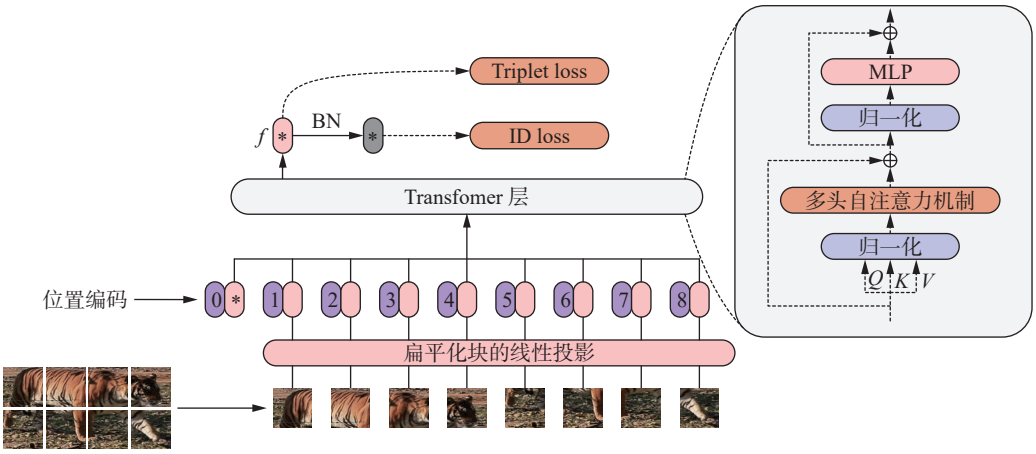

图 3　Vision Transformer 网络架构

Fig. 3　Vision Transformer network architecture

1.2.2　空间位置强化

在构建以 Transformer 为基础的特征提取网络时，引入坐标注意力机制[23]充分发挥坐标注意力机制在空间信息处理上的优势，优化基础模型的性能表现。这种引入可以进一步提升模型对复杂数据特征的理解与处理能力。

坐标注意力机制为了增强模型对空间信息的感知与利用能力，对输入特征图进行水平及垂直方向的处理，它先对输入特征图沿空间维度拆分为水平和垂直方向，通过水平和垂直方向的自适应平均池化，分别提取宽度与高度方向信息，将特征图转化为水平、垂直两个一维的特征描述；然后，把这两个方向池化结果拼接，经卷积、批归一化和激活操作编码，挖掘内在关系，经过编码后，特征被拆分为水平和垂直特征向量，各自卷积再通过 Sigmoid 函数转化为在 0~1 变化的注意力权重值，表示不同空间位置上各个方向的重要性，再采用其权重值对原始特征图加权求和，凸显重要的空间区域信息，并抑制无关的信息，最终得到一张加权后的特征图用于后续模型，助力提升模型在各类任务中的表现。坐标注意力机制结构如图 4 所示。

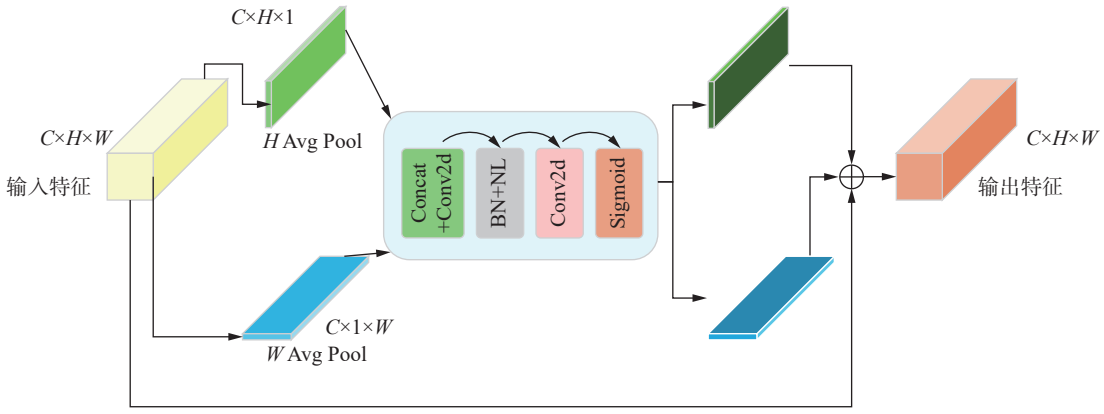

图 4　坐标注意力机制结构

Fig. 4　Structure of the coordinate attention mechanism

传统基于聚类与 Transformer 的模型用于东北虎图像会存在空间信息利用率不高的问题，主

要体现在东北虎个体的条纹分布、身体姿态等具有空间关系性特点的差异特征, 且以上差异特征会存在不同拍摄角度及不同环境下的变化, 所以传统注意力机制不易捕捉到细节、多变化性的空间关系, 在东北虎重识别过程较易产生误判, 不能达到保护工作中对东北虎高精度识别的实际需求。

在模型架构中, 图像分割模块把输入的东北虎图像转成 Transformer 适合的补丁嵌入形式, 同时将坐标注意力机制应用到图像分割模块里面, 经引入坐标注意力机制后的输入特征图经过水平和垂直的自适应平均池化变换后, 在水平的自适应平均池化中对特征图在高度方向上进行压缩来保留特征图在宽度方向上的关键信息, 其输出的特征图的高度将会变成 1, 宽度不变, 而在垂直的方向上的自适应平均池化则会保留特征图高度方向上的信息, 对宽度方向上的信息进行压缩, 压缩后其输出的特征图宽度也变为 1, 高度和原来一样不变。之后将这两个方向的池化结果拼接后通过一些卷积、批归一化、激活的操作得到对应的坐标注意力权重, 最后将经过卷积层提取得到的结果特征相乘得到加权后的特征, 突出关键空间信息, 为后续提取特征的处理提供更高质量的特征表示。

坐标注意力机制可以单独进行水平和垂直的信息编码并解码, 利于将空间的位置与方向关系信息更好地表示出来, 加入坐标注意力机制后, 理论上能够缓解 ViT 模型缺少水平、垂直方向空间细节的弊端, 其中, 水平方向主要关注腹部条纹的间隔宽窄问题, 垂直方向主要关注腿上条纹的稀疏程度, 从而在复杂的实际场景中实现更准确的个体识别, 为东北虎种群监测和保护策略制定提供有力支持。

1.2.3　无监督均衡训练

东北虎个体重识别由于不同个体间的样本量差别较大, 因此在无监督学习过程中会导致模型在特征优化时发生偏差, 为解决这一问题, 引入了 Cluster Contrast 机制, 该机制是 Dai 等[24] 针对无监督行人重识别任务提出的创新方法, 其核心在于构建了簇级内存字典和有针对性的对比损失函数, 实现不同样本量东北虎的特征优化均衡。

基于 DBSCAN 聚类算法, 以初始特征空间作为输入, 在将具有相似性的特征聚成一类后, 每个类别代表一类东北虎个体, 并用该类所有特征均值进行聚合得到初始簇特征向量, 构造内存字典, 将同一类东北虎的多样本特征压缩为统一表征, 消除大样本簇对特征空间的主导力, 字典将每类东北虎的簇特征向量全部保存, 并使用增量式更新策略来维护, 极大程度地减少了字典的存储复杂度。

动量更新机制公式为

$$\boldsymbol{\phi}_k^{(t)} = \alpha \cdot \boldsymbol{\phi}_k^{(t-1)} + (1-\alpha) \cdot \boldsymbol{q} \tag{1}$$

式中: $\boldsymbol{\phi}_k^{(t-1)}$为第 k 个簇在第 t 次迭代时的簇特征向量; α 为动量因子用于控制模型参数更新时对历史梯度依赖程度的超参数; $\boldsymbol{q}$ 为当前参与更新的查询特征, 通过融合历史簇特征与当前查询特征, 由编码器 f_θ从训练集中的实例提取所得到, 确保大样本簇与小样本簇的特征优化速率一致。式(1)为训练过程中的簇特征动态优化公式, 对于样本量大的簇, 高动量因子使历史特征占主导, 避免频繁更新导致的特征波动; 对于样本量小的簇, 当前特征的修正幅度相对更大, 确保少量样本特征不被稀释, 最终使簇特征收敛于所有历史特征的指数加权平均, 与样本量无关。

在计算对比损失时, Cluster Contrast 采用了簇级别的 ClusterNCE 损失, 该损失计算查询实例特征与所有簇特征之间的对比损失。ClusterNCE 损失是一种基于 K 路 Softmax 分类器的对数损失, 计算公式为

$$L_q = -\log \frac{\exp(\boldsymbol{q} \cdot \boldsymbol{\phi}_+ / \tau)}{\sum_{k=0}^{K} \exp(\boldsymbol{q} \cdot \boldsymbol{\phi}_k / \tau)} \tag{2}$$

式中: $\boldsymbol{q}$是查询实例特征向量; $\boldsymbol{\phi}_+$是与$\boldsymbol{q}$属于同一类别的簇的特征向量, 代表正样本簇特征; $\boldsymbol{\phi}_k$是第 k 个簇的特征向量; K 为簇的总数; τ 用于调整对比学习的难度和稳定性, 通常设置为一个较小的正值, 它控制了概率分布的平滑程度, 值越小, 分布越集中, 模型对相似特征的区分要求越高。

该损失函数的计算基于簇级。在训练过程中, 每次从训练集中采样一批实例, 将这些实例的特征作为查询特征 q。对于每个查询特征 q, 通过点积$q \cdot \phi_k$衡量计算它与内存字典中所有簇特征向量ϕ_k的相似度。分子$\exp(q \cdot \phi_+ / \tau)$表示查询特征 q 与正样本簇特征ϕ_+相似度的指数化结果, 分母$\exp(q \cdot \phi_k / \tau)$则是 q 与所有簇特征向量相似度的指数化结果之和。通过这种方式, 模型学习到的特征表示应使得查询特征与正样本簇特征的相似度在所有簇特征中脱颖而出, 从而最小化损失值。Cluster Contrast 机制对比损失计算及簇特征更新机制如图 5 所示。

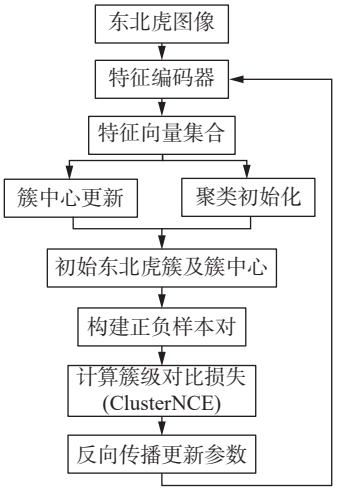

图 5　Cluster Contrast 机制对比损失计算及簇特征更新机制

Fig. 5　Contrastive loss calculation and cluster feature update mechanism of the Cluster Contrast

2　结果与分析

2.1　实验设置

实验选用 8 核 AMD EPYC 7601 CPU, 内存 32 GB, 主频 2.2 GHz, 显卡为 NVDIA 3060, 显存 12 GB。使用 Linux 系统搭建基于 Python3.8 和 PyTorch 1.11.0 的深度学习框架。使用 SGD 优化器, 初始学习率设置为 0.000 05, 批次大小设置为 32, 训练迭代次数为 100, 训练周期设置为 50。

2.2　评价指标

在重识别系统的评估体系中, 累计匹配特征(cumulative matching characteristic, CMC)[25] 与平均精度均值(mean average precision, mAP)[26] 是两种常用的评估指标。

CMC 曲线也称为 Rank-k 匹配准确率, 表示排名在前 k 的检索结果中存在正确的匹配结果的概率, 因为在评价时只考虑第一个出现的匹配结果, 其能对每一个查询图像仅有一个真实正确的图像的场景下进行正确评判。但在大规模摄像头采集图片时, 往往会有很多个真实的正确图像, 这就导致了 CMC 曲线无法较好地衡量模型在多摄像头间的正确匹配情况。mAP 也在图像检索领域被广泛应用, 主要用于衡量存在多个真实标注的复杂情况下模型的平均检索能力。

一个优秀有效的重识别系统, 目标不仅要被准确检索出来, 而且所有正确匹配结果的排名都应该靠前。考虑到检索目标在排名靠前的检索列表中, 且绝对不能被忽视, 尤其是在多摄像头环境下, 这样才能实现对目标的精准追踪。当目标在图库集中多个时间节点出现时, 最难匹配的正确结果的排名越靠后, 研究人员的工作量越大, 但当前广泛使用的累计匹配特性和平均精度均值指标, 无法有效地评估这一特性。如图 6 所示, 在相同的 CMC 情况下(均为 1), 列表 1 的平均精度明显优于列表 2, 但要找到所有正确匹配结果研究人员需要付出更多的精力与时间。

图 6　CMC、mAP、INP 特性及排名列表性能对比的关系

Fig. 6　Relationship of the performance comparison of the characteristics and ranking lists of CMC, mAP, and INP

为了解决这一问题, Ye 等[10] 提出了一种计算检索效率的指标, 即负惩罚(NP), 用于衡量找到最难匹配的正确结果所需付出的代价。

$$P_{\mathrm{N}i}=\frac{R_i^{\mathrm{hard}}-|G_i|}{R_i^{\mathrm{hard}}} \tag{3}$$

式中: 表示 R_i^{hard} 最难匹配结果的排名位置, $|G_i|$ 代表查询的正确匹配总数。显然, NP 值越小, 性能越好。为了与 CMC 和 mAP 保持一致, 指标更倾向于使用负惩罚倒数(INP), 它是 NP 的倒数运算。总体而言, 所有查询的平均 INP 表示为

$$P_{\mathrm{INm}}=\frac{1}{n}\sum_i(1-P_{\mathrm{N}i})=\frac{1}{n}\sum_i\frac{|G_i|}{R_i^{\mathrm{hard}}} \tag{4}$$

图 6 中列表 1 $P_{\mathrm{N}}=\dfrac{10-3}{10}=0.7$, $P_{\mathrm{IN}}=0.3$; 列表 2 $P_{\mathrm{N}}=\dfrac{5-3}{5}=0.4$, $P_{\mathrm{IN}}=0.6$。

P_{INm} 的计算效率相当高, 并且可以在 CMC 和 mAP 的计算过程中衔接, 在 CMC 和 mAP 评估中, 容易出现高分的匹配的结果占据主导的现象。P_{INm} 可以有效地避免该类问题; 但是与小图库比较, 大图库的 P_{INm} 值差异并不是很大, 所以这种方式依然能够反映出模型的一些信息, 并且能够辅助 CMC 和 mAP 指标。

2.3　结果与分析

依据文献 [9] 中提出的“实体独立性划分原则”, 确保训练集与测试集无重叠东北虎个体, 对原 ATRW 数据集测试集(test)进行切分, 并分别建立训练集(train)、测试集(gallery)、查询集(query), 模型训练中使用 train 集、测试时使用 gallery 集和 query 集; 由于原测试集的数据较少, 不利于模型特征的学习, 因此进一步将原来的

train 集按照文献的方法划分, 实验发现重新划分的 ATRW(train)相比原划分的 ATRW(test)有着更加优秀的 mAP 和 mINP 表现, 原因在于所划分的 ATRW(train)数量更多, 能够提供更多的特征供模型学习, 之后将 ATRW(train)与 ATRW(test)混合得到 ATRW(mix)数据集, 具体的划分方式和评估指标如表 1、2 所示。损失变化如图 7 所示。

表 1　数据集划分情况

Table 1　Division situation of the dataset

划分情况	ATRW(测试)		ATRW(训练)		ATRW(混合)	
	Ids	Images	Ids	Images	Ids	Images
训练集	19	741	74	1 336	93	2 077
测试集	9	280	34	420	43	718
查询集	9	42	34	131	43	155

表 2　数据集指标情况

Table 2　Indicator situation of the dataset　%

数据集	mAP	Rank-1	mINP
ATRW(测试)	64.1	98.3	33.0
ATRW(训练)	83.5	98.5	61.0
ATRW(混合)	77.1	96.1	50.6

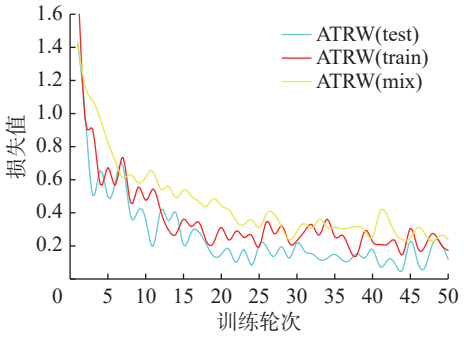

图 7　损失情况

Fig. 7　Circumstances of losses

表 1、2 中, ATRW(mix)的综合指标处于前两者之间, 进一步说明了 ATRW(train)与 ATRW(test)数据质量的差别。虽然 ATRW(train)所提供的训练数据量很大, 但是其中包含 ATRW(test)质量较差的数据会影响到模型的拟合情况, 从而使得模型无法达到和 ATRW(train)一样的效果, 不同的分布会导致模型在不同分布上的泛化性能不同的结果。

图 7 为不同数据集的 Cluster Contrast 损失变化曲线, 可见, ATRW(test)损失收敛最快, 在约第 20 轮后损失值趋于稳定至 0.2 左右, 这是因为该数据集样本少, 模型能快速学习到有效特征; ATRW(train)损失收敛速度次之, 在约第 30 轮后损失值稳定在 0.2~0.4, 其样本量充足, 模型需更多轮次迭代以学习全面特征; ATRW(mix)损失收敛最慢, 在约第 35 轮后才逐渐损失值稳定在 0.2~0.6, 这是由于该数据集混合了不同分布的样本, 增加了模型学习难度, 导致收敛过程更曲折。

由此得出结论, 数据集质量与模型性能呈正相关。因为 ATRW(train)有较多的样本量, 所以数据的质量较高, 能够学习到东北虎条纹特征的全局分布以及局部变化规律。在模型学习之后可对模型做更为准确的特征匹配, 但 ATRW(test)的样本量较少且不太具有代表性, 在模型检测的时候会对少数姿态或者少见环境下特征的判别能力较差, 因此结果指标不如 ATRW(train)。其次, ATRW(mix)中间状态的性能证明: 如果训练数据中含有分布不同的样本会干扰模型学习到最本质的关键特征, 因此该结果对于野外东北虎监测数据采集、模型优化具有参考意义。

为评估模型在红外数据集的性能, 在构建的 ATRW(wild)上进行了测试, 具体指标数据如表 3 所示。

表 3　不同模型在红外数据集上的性能对比

Table 3　Performance comparison of different models on the infrared dataset

模型	mAP/%	Rank-1/%	mINP/%	参数量/10^6
本文模型	96.3	98.3	86.7	21.9+1.688
ViT	95.3	98.3	85.3	21.9
Resnet50_ibn	94.9	98.3	84.6	25.6

本文模型在红外数据集上 mAP 达 96.3%、Rank-1 达 98.3%、mINP 达 86.7%, 高于 Resnet50_ibn、VIT, 证明 ViT+CA 的组合架构对红外低质量特征的适配性显著优于传统 CNN, 且其对野外红外场景中样本的识别能力更优, 进一步验证模型可有效适配野外实际监测需求。

由于红外数据集很少(仅有 12 个个体, 200 张图片), 导致模型容易过拟合, 因此为验证模型在可见光和红外数据综合的性能, 本研究把可见光数据和红外数据进行混合在数据集 ATRW(r+i)与可见光数据集开展对比实验, 并分析了 Resnet 与 ViT 特征提取网络的性能表现, 具体指标数据如表 4 所示。

可见, 本文模型与 ViT 模型在样本量更丰富的数据集上, 两者对条纹特征的捕捉更全面, 核心指标 mAP、mINP 显著优于样本量较少的数据集。而 Resnet50_ibn 表现特殊, 虽核心指标随样本量增加仍有提升, 但分类准确率 Rank-1 反而下降, 反映出传统 CNN 架构易受大规模数据中局部冗余特征如复杂背景、姿态差异的干扰, 分类稳定性不足, 凸显不同模型对数据集规模的适应性差异。

表 4　不同特征提取网络的数据集指标情况

Table 4　Dataset indicator situations of different feature extraction networks

网络	ATRW(测试)			ATRW(训练)			ATRW(r+i)			参数量/10^6
	mAP/%	R-1/%	mINP/%	mAP/%	R-1/%	mINP/%	mAP/%	R-1/%	mINP/%	
本文	66.2	98.3	34.3	84.9	98.5	63.7	86.4	97.1	69.8	21.9+1.688
ViT	64.1	98.3	33.0	83.5	98.5	61.0	85.2	96.2	67.0	21.9
Resnet50_ibn	71.0	98.3	36.4	81.6	96.2	57.6	84.4	95.7	67.2	25.6

在融合可见光与红外数据的 ATRW(r+i)数据集上，本文模型表现最优，其通过 ViT 全局注意力捕捉两种模态条纹特征的共性关联，结合 CA 机制强化空间细节解析，有效适配多模态特征差异，核心指标 mAP、mINP 均领先于 ViT 与 Resnet50_ibn，同时避免了单一红外数据量少可能导致的过拟合问题。证明其在不同数据分布、样本量场景下均具备强泛化能力，适配野外可见光/红外多设备协同监测需求。

3　结束语

本研究针对东北虎个体重识别中野外数据标注难、样本失衡及传统 CNN 局部性缺陷等问题，以 ATRW 数据集为基础，构建了融合 Vision Transformer(ViT)、坐标注意力与 Cluster Contrast 机制的无监督学习框架，实现了东北虎个体无监督重识别。

数据层面研究发现，数据集质量与模型性能存在显著关联，使用样本量更充足、分布更一致的数据集，能帮助模型更全面掌握东北虎条纹的全局分布与局部变化规律，进而提升识别效果，而训练数据中混入不同分布的样本时，会干扰模型对核心条纹特征的学习，导致性能下降。这一结果为野外东北虎监测数据的采集工作提供了实践参考，明确了数据分布一致性对模型泛化能力的重要性。

模型性能方面，本文模型在各类数据集上的综合表现均优于 ViT、Resnet50_ibn 等基线模型，其优势源于各组件的协同作用。ViT 的自注意力机能够有效捕捉东北虎条纹跨身体区域的长距离依赖特征，即便面对部分遮挡场景，也能通过未遮挡条纹的关联关系推断整体特征。坐标注意力机制进一步强化了对条纹空间位置的解析，针对不同拍摄角度、姿态下条纹的细微差异，通过水平与垂直方向的分别编码，精准突出关键空间信息，减少因视角变化导致的误判。Cluster Contrast 机制则通过构建簇级内存字典与动量更新策略，平衡了不同样本量东北虎的特征优化速率，避免了无监督学习中样本量多的个体主导特征学习、样本量少的个体特征被稀释的问题，确保模型对各类样本的特征学习更均衡。无论是在常规可见光场景、复杂野外红外场景，还是可见光与红外数据混合的跨模态场景中，该模型均能保持稳定优异的识别性能，体现出良好的环境适应性与鲁棒性。

综上所述，本文建立了一个通过多种技术手段相融合的方式提高东北虎个体重识别准确度及鲁棒性的无监督学习框架，进而为野外东北虎监测、保护工作提供一定的理论支撑，为其他濒危物种无监督重识别的相关研究提供了全新的思路——全局特征提取−空间位置强化−无监督均衡训练。

参考文献：

[1] 国家林业和草原局, 农业农村部. 《国家重点保护野生动物名录》(2021 年 2 月 1 日修订)[J]. 野生动物学报, 2021, 42(2): 605−640.
National Forestry and Grassland Bureau, Ministry of Agriculture and Rural Affairs. List of national key protected wild animals (revised on February 1, 2021)[J]. Chinese journal of wildlife, 2021, 42(2): 605−640.

[2] QI Jinzhe, HOLYOAK M, NING Yao, et al. Ecological thresholds and large carnivores conservation: implications for the Amur tiger and leopard in China[J]. Global ecology and conservation, 2020, 21: e00837.

[3] 王凤昆, 李艳, 姜广顺. 中俄东北虎自然保护地建设进展[J]. 自然保护地, 2024, 4(4): 38−52.
WANG Fengkun, LI Yan, JIANG Guangshun. Progress in the construction of Sino-Russia Amur Tiger(Panthera Tigris altaica) Protected Areas[J]. Natural protected areas, 2024, 4(4): 38−52.

[4] 吴峰, 温佩颖, 唐志珍, 等. 东北虎豹国家公园植物 DNA 微条形码数据库[J]. 北京师范大学学报 (自然科学版), 2023, 59(4): 623−628.
WU Feng, WEN Peiying, TANG Zhizhen, et al. A DNA mini-barcode reference library for environmental DNA study in the Northeast Tiger and Leopard National Park [J]. Journal of Beijing Normal University (natural science edition), 2023, 59(4): 623−628.

[5] 韦怡, 姜广顺. 虎豹及有蹄类猎物种群数量监测方法概述[J]. 生物多样性, 2022, 30(9): 129−146.
WEI Yi, JIANG Guangshun. Overview of monitoring methods for tigers, leopards and ungulate prey[J]. Biodiversity science, 2022, 30(9): 129−146.

[6] HE Fengping, LIU Dan, ZHANG Le, et al. Metagenomic analysis of captive Amur tiger faecal microbiome[J]. BMC veterinary research, 2018, 14(1): 379.

[7] KERLEY L L. Using dogs for tiger conservation and research[J]. Integrative zoology, 2010, 5(4): 390−396.

[8] TUIA D, KELLENBERGER B, BEERY S, et al. Perspectives in machine learning for wildlife conservation[J].

Nature communications, 2022, 13: 792.
[9] YE Mang, CHEN Shuoyi, LI Chenyue, et al. Transformer for object re-identification: a survey[J]. International journal of computer vision, 2025, 133(5): 2410−2440.
[10] YE Mang, SHEN Jianbing, LIN Gaojie, et al. Deep learning for person re-identification: a survey and outlook[J]. IEEE transactions on pattern analysis and machine intelligence, 2022, 44(6): 2872−2893.
[11] 张国印, 王传博, 高伟. 抗遮挡的行人多目标跟踪算法[J]. 智能系统学报, 2024, 19(5): 1248−1256.
ZHANG Guoyin, WANG Chuanbo, GAO Wei. Pedestrian multiobject tracking algorithm with anti-occlusion[J]. CAAI transactions on intelligent systems, 2024, 19(5): 1248−1256.
[12] 王路遥, 王凤随, 闫涛, 等. 结合多尺度特征与混淆学习的跨模态行人重识别[J]. 智能系统学报, 2024, 19(4): 898−908.
WANG Luyao, WANG Fengsui, YAN Tao, et al. Cross-modal person re-identification combining multi-scale features and confusion learning[J]. CAAI transactions on intelligent systems, 2024, 19(4): 898−908.
[13] 赵玲, 周桂红, 任力生. 基于牛脸和躯干综合信息的奶牛个体识别研究[J]. 河北农业大学学报, 2024, 47(2): 112−118.
ZHAO Ling, ZHOU Guihong, REN Lisheng. Individual identification of dairy cows based on comprehensive face and trunk information[J]. Journal of Agricultural University of Hebei, 2024, 47(2): 112−118.
[14] FREYTAG A, RODNER E, SIMON M, et al. Chimpanzee faces in the wild: log-euclidean CNNs for predicting identities and attributes of Primates[C]//Pattern Recognition. Cham: Springer, 2016: 51–63.
[15] 顾佳音, 刘辉, 姜广顺. 东北虎 (Panthera tigris altaica) 个体识别技术研究进展[J]. 野生动物, 2013, 34(4): 229−237, 248.
GU Jiayin, LIU Hui, JIANG Guangshun. A review of potential techniques for indentifying individual Amur tigers (Panthera tigris altaica)[J]. Chinese wildlife, 2013, 34(4): 229−237, 248.
[16] GLOVER J D, SUDDERICK Z R, SHIH B B, et al. The developmental basis of fingerprint pattern formation and variation[J]. Cell, 2023, 186(5): 940–956. e20.
[17] SHI Chunmei, LIU Dan, CUI Yonglu, et al. Amur tiger stripes: individual identification based on deep convolutional neural network[J]. Integrative zoology, 2020, 15(6): 461−470.
[18] 马光凯, 张静, 戴文锐, 等. 基于 Transformer 的东北虎体侧条纹个体识别[J]. 野生动物学报, 2024, 45(4): 734−743.
MA Guangkai, ZHANG Jing, DAI Wenrui, et al. Body stripes individual identification of Amur tigers based on Transformer[J]. Chinese journal of wildlife, 2024, 45(4): 734−743.
[19] CHMIELA S, SAUCEDA H E, MULLER K, et al. Addressing the elephant in the room: uncertainties in physical predictions from machine-learned force fields[J]. Bulletin of the American physical society, 2020.
[20] LI Shuyuan, LI Jianguo, TANG Hanlin, et al. ATRW: a benchmark for Amur tiger re-identification in the wild[C]// Proceedings of the 28th ACM International Conference on Multimedia. Seattle: ACM, 2020: 2590–2598.
[21] PARIKH A, TÄCKSTRÖM O, DAS D, et al. A decomposable attention model for natural language inference [C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2016: 2249–2255.
[22] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale[EB/OL]. (2020−10−22)[2025−08−03]. https://arxiv.org/abs/2010.11929.
[23] HOU Qibin, ZHOU Daquan, FENG Jiashi. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 13708–13717.
[24] DAI Zuozhuo, WANG Guangyuan, YUAN Weihao, et al. Cluster contrast forUnsupervised person re-identification [C]//Computer Vision–ACCV 2022. Cham: Springer, 2023: 319–337.
[25] WANG Xiaogang, DORETTO G, SEBASTIAN T, et al. Shape and appearance context modeling[C]//2007 IEEE 11th International Conference on Computer Vision. Rio de Janeiro: IEEE, 2007: 1–8.
[26] ZHENG Liang, SHEN Liyue, TIAN Lu, et al. Scalable person re-identification: a benchmark[C]//2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2016: 1116–1124.

作者简介:

赵亚凤, 副教授, 博士, 主要研究方向为计算机视觉与模式识别、人工智能与智能控制、无线传感器网络。主持和参与国家自然科学基金、黑龙江省自然科学基金等项目 10 余项, 获发明专利及实用新型专利授权 10 余项, 发表学术论文 20 余篇, 作为主编及副主编出版著作 2 部。E-mail: zyf@nefu.edu.cn。

于继超, 硕士研究生, 主要研究方向为计算机视觉。E-mail: jichaoyu@nefu.edu.cn。

王梓丞, 副教授, 博士后。黑龙江省光学学会会员, 国际期刊《Photonics》特邀编辑, 《半导体光电》编委。主要研究方向为高密度集成光学器件设计、智能光纤传感系统开发及多物理场耦合仿真技术。主持或参与省部级及横向项目 10 余项, 获黑龙江省科学技术发明二等奖, 指导学生获国家级、省部级竞赛奖项多项。获发明专利及软件著作授权 12 项。近 5 年第一作者/通信作者发表论文共 13 篇。E-mail: wangzicheng1992@163.com。

[责任编辑: 刘冰洁]