

基于多任务学习的输电线路小金具 缺失推理加速算法

程 绳¹, 葛 雄¹, 肖 非², 朱传刚¹, 吴 军¹, 肖海涛¹,
李 嗣¹, 楚江平¹, 袁雨薇¹

(1. 国网湖北省电力有限公司超高压公司, 武汉 430050; 2. 浙江大学滨江研究院, 杭州 310000)

摘要: 针对输电线路小金具缺失的检测问题, 对小金具缺失算法的加速推理进行研究, 采用多任务头的学习方法, 将小金具缺失检测任务使用一个 Swin Transformer 网络结构^[12]和多个由多层感知机组成的任务头组合的方式进行多任务学习推理, 并进行单任务和多任务学习的推理精度和推理性能对比实验, 最后还通过实验得到在多任务学习中插拔式扩展任务, 实验结果表明在输电线路小金具缺失推理算法中多任务学习比单任务学习的推理性能提升了 2 倍多, 同时显存占用降低了 22% 以上; 通过插拔式扩展任务实验, 验证了扩展任务的效果, 可灵活扩展配置任务。

关键词: 多任务头学习; 加速推理; 输电线路; 小金具缺失; 扩展任务学习

Multi-task Learning Based Acceleration Algorithm for Missing Small Fittings in Transmission Line

CHENG Sheng¹, GE Xiong¹, XIAO Fei², ZHU Chuangang¹, WU Jun¹, XIAO Haitao¹,
LI Si¹, CHU Jiangping¹, YUAN Yuwei¹

(1. State Grid Hubei Extra High Voltage Company, Wuhan 430050, China;

2. Binjiang Institute of Zhejiang University, Hangzhou 310000, China)

Abstract: Aiming at the detection problem of missing small fittings in transmission lines, the acceleration of the inference speed of the missing small fittings algorithm is studied. The multi-task learning method is adopted, and the missing small fittings detection task is connected by a Swin Transformer backbone network and multiple MLP task heads. We propose a new approach of doing multi-task learning as well as multi-task inference, and compare the accuracy against single-task learning in different scenarios. We then verified that new tasks can be seamlessly added to the multi-task learning framework, without any performance degradation. The experimental results show that the speed of proposed multi-task learning is more than 2 times higher than that of single-task learning, and the memory usage is reduced by more than 22%. These experimental results verified the effectiveness of the proposed multi-task learning framework and the flexibility of the task settings.

Keywords: multi-task heads learning; inference acceleration; transmission line; missing small fittings; extended task learning

0 引言

输电线路是我国电力系统的重要组成部分, 输电线路的稳定是我国电力系统的稳定基础。螺栓小金具是输电线路的最小单元, 用于连接和紧固输电线路中各种金具部件和各种结构件, 保证输电线路系统的稳定运行^[1]。但螺栓小金具长期在户外环境经过震动、摩擦、温度变化等原因, 存在松动甚至缺失的情况。一旦螺栓小金具出现缺失有可能导致漏电, 线路毁坏, 甚至线路断电, 将严重影响用电安全, 造成重大事故^[2-3]。深度学习在输电线路中应用越来越多^[19-22], 因此深度学习对螺栓小金具缺失现象的识别和判断, 有助于即时修复小金具, 对输电线路的安全运行有

着重要意义。

螺栓小金具主要包括单螺母、双螺母、开口销、垫片、螺杆等。在输电线路中不同作用的螺栓由不同的小金具组成, 比如, 塔身垂直螺栓一般在螺栓杆轴向受力较大, 故由螺栓杆和双螺母进行紧固和防松。活动件螺栓一般在轴向受力较小, 故由螺栓杆、单螺母和开口销组合进行连接。线夹螺栓、跳线螺栓、均压环安装螺栓, 一般采用螺栓杆、平垫片、弹簧垫片单螺母进行紧固和放松。按照螺栓的功能作用进行分类, 输电线路中存在多种螺栓类型。通过归纳总结, 按照功能作用分为以下 11 种螺栓类型: 活动件螺栓、塔身垂直螺栓、塔身 U 型螺栓、塔身非垂直螺栓、线

收稿日期: 2022-11-08; 修回日期: 2022-12-05。

作者简介: 程 绳(1988-), 男, 湖北仙桃人, 硕士, 高级工程师, 主要从事输电运维与管理方向的研究。

通讯作者: 肖 非(1985-), 男, 浙江杭州人, 学士, 高级工程师, 主要从事人工智能技术与应用方向的研究。

引用格式: 程 绳, 葛 雄, 肖 非, 等. 基于多任务学习的输电线路小金具缺失推理加速算法[J]. 计算机测量与控制, 2023, 31(7): 251-257.

夹螺栓、线夹无螺母螺栓、均压环螺栓、跳线螺栓、绝缘子盘螺栓、重锤螺栓、绝缘子槽碗螺栓。

输电线路中小金具缺失主要是单螺母、双螺母或开口销的缺失,需要根据螺栓类型判断是否缺失某种小金具。目前主要通过无人机技术、图像处理技术、神经网络算法等技术对输电线路小金具缺失进行检测。主要步骤为人工或自主巡检无人机对输电线路中的小金具部位进行拍照,再通过人工检查或图像处理技术对获得的图像进行缺失判断,从而得出螺栓是否存在小金具缺失或者异常的现象^[4]。近几年,随着深度学习技术和无人机技术的快速发展,输电线路小金具智能化检测的研究越来越多。从输电线路小金具数据生成到小金具缺失、异常检测的研究越来越多。通过深度卷积生成式对抗神经网络方法对小金具图像进行增强和生成,为输电线路小金具缺失数据难以获得的问题提供了一种解决方法和思路^[5]。通过 SSD^[10]和 YOLOv3^[11]对小金具目标进行分级检测,使得输电线路在多层级进行检测。具体步骤为,首先对输电线路的关键挂点位置进行识别,然后在对其小金具缺陷目标进行检测。同时,小金具的目标检测准确率通过数据增强的方式进行提升。同时,采用了两次特征金字塔和空间的聚类算法,针对高置信螺栓缺失开口销的目标区域和螺栓缺失开口销的目标区域目标进行增强和检测。根据螺栓和螺母的组合特征,提出栓螺母对的概念,利用知识图谱的方法分别对缺垫片、缺开口销、缺垫片+缺开口销和完整 4 类进行检测,缺垫片、视觉可见缺开口销、视觉不可见缺开口销、缺垫片+视觉可见缺开口销、缺垫片+视觉不可见缺开口销和完整 6 类进行检测^[8]。

综上,当前输电线路中小金具缺失检测的算法都采用单任务学习,缺垫片、缺开口销、缺螺母等作为单一任务进行需学习。由于多种小金具的缺失算法学习的螺栓特征不同,训练模型的参数时,特征提取变得困难,导致算法的识别效果变差。单任务学习每个模型需要单独训练,无法在原有的特征提取模型上进行算法扩展,对于新的检测任务,需要进行新的设计和训练算法。

对于上述存在的问题,本文采用多任务学习的输电线路小金具缺失算法,同时使用加速推理的技术实现多个任务共享主干网络的并行推理算法,从而提升推理性能。按照学习任务的类型将输电线路小金具缺失拆分成多个任务,分别为单螺母缺失任务、开口销缺失任务、螺栓类型识别任务。为了提升小金具的缺陷目标检测的效果及推理加速,我们采用模型参数共享的方式进行推理,具体为多个任务采用同一个主干网络进行关键特征的提取。同时该算法对输电线路小金具缺失检测任务具有插拔式的扩展功能。如图 1 所示,给定一个指定输入图像,不同的任务采用共享网络主干进行特征提取。不同的检测任务采用并行计算的方式得到不同任务的识别结果。此外,由于该算法结构中不同检测任务的的任务头相互独立互不影响,算法任务可以进行可插拔式的扩展,大大提升模型的灵活度和推理效率。

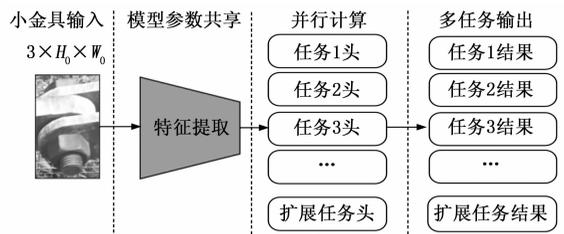


图 1 小金具多任务加速算法结构图

1 研究思路

本文依据输电线路小金具缺失的特点,对算法识别准确率、算法推理性能和任务可扩展性三个角度设计输电线路小金具缺失加速推理算法。首先提出共享主干的多任务头学习算法,提高小金具缺失算法的检测准确率。将输电线路小金具缺失检测分为多个不同任务,小金具的缺失检测,可以拆分为不同的子任务。通过该方法,可以将较为复杂的识别任务分解为 N 个较为简单的识别检测任务。为了保障多个任务推理的性能,本文使用多任务共享主干网络结构进行特征提取的方式,从而在保障模型推理速度的基础上,提升单个任务准确率,进而提升小金具缺失检测的整体效果。对不同的子任务,我们采取 triton 的并行推理技术,使得不同的子任务进行解耦,提升识别子任务的可插拔性。通过实验,在算法模块训练完成的情况下,新增额外的识别子任务,不会对原任务产生影响。从而可以实现新任务的算法扩展。

将输电线路小金具缺失识别任务分解为螺栓类型识别任务、双螺母缺失任务、开口销缺失任务。判断小金具的缺失,首先需要确定待识别物体的螺栓类型。根据其在输电线路中的位置和功能,螺栓可以被分为多种类型,其包含的小金具类型也不同。比如,塔身垂直安装的螺栓一般需要安装双螺母进行紧固和防松,活动件螺栓一定要有开口销进行防松。因此判断螺栓小金具缺失需要识别螺栓类型和双螺母、开口销存在状态,通过螺栓类型识别任务、双螺母缺失任务、开口销缺失任务获得螺栓类型和小金具存在状态,最后可根据螺栓类型和小金具的存在状态判断小金具是否缺失。本文对输电线路小金具缺失模型加速推理算法进行研究,通过任务拆解及子任务解耦的方式,将小金具缺失的任务拆分为独立的子任务,并结合模型参数共享和 Triton 框架达到模型并行计算结果组合的目的,从而提高输电线路小金具缺失模型加速推理算法推理性能。该算法具有以下优点,可以原算法基础上进行插拔式任务扩展,将独立的学习任务,融合到多任务学习中。本文通过添加单螺母缺失检测任务,实现插拔式任务扩展。

本文的总体研究思路为基于多任务学习的输电线路小金具缺失模型加速推理算法,在算法准确率、推理性能方面得到提升,同时具有任务插拔式的可扩展性。

2 算法研究

输电线路小金具缺失方面的工作仍存在缺陷目标检测

识别准确率不足的问题。此外, 由于目标检测任务的复杂性, 其在推理速度上效果有待提升。因此, 本研究的目的是针对输电电路缺陷目标检测在推理性能和效果两方面进行提升。通过设计一种新颖的模型框架方法和推理引擎优化, 达到目标识别优化的效果。在模型设计方法方面, 对以前针对每一类小金具单独进行缺陷分类, 优化为将多种金具统一输入我们设计的算法模型框架, 即多任务学习框架进行多个子任务的任务头训练, 最终通过共享模型参数和可插拔可扩展的子任务并行计算方式输出每个任务的结果。在推理引擎方法方面, 本文在本文设置了全面的实验, 在不同推理模式下, 对推理速度进行了对比。通过实验, 共享模型参数结合多任务头学习的方式, 在 Triton 的框架下, 比相同参数的单任务推理方法显著的提升了推理性能。

2.1 模型结构

如图 2 所示, 本研究的多任务学习模型主要由通用的主干网络和用于分类的多任务头组成。其中, 多任务可以分为原始任务和扩展任务。原始任务是在模型初始训练时包含的任务, 扩展任务是可插拔的动态任务。

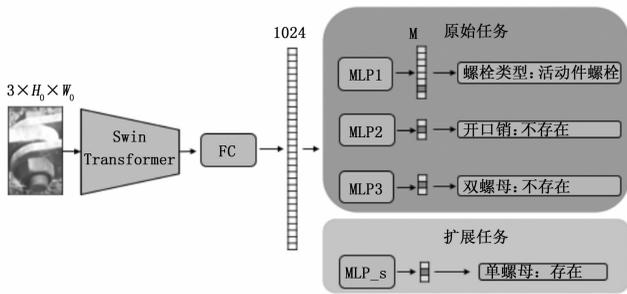


图 2 多任务学习模型结构图

2.1.1 主干网络模块

输电线路小金具的输入数据是 3 通道的 RGB 图像, 用 $i \in R^3 \times H_0 \times W_0$ 表示。将原始图像 i 输入到主干网络, 通过主干网络的视觉编码器生成图像 i 的视觉特征向量, 用 $f_i \in R^{1024}$ 表示。本研究的主干网络采用 Swin Transformer^[12], 一个将 Transformer 从自然语言处理领域应用到图像识别领域的一个通用的循环神经网络结构。

本研究的主干网络采用 Swin Transformer^[12], 是计算机视觉领域最先进的网络结构之一, 借鉴 Transformer 在自然语言处理的效果, 在图像识别领域得到了优秀的语义识别效果。该网络对输入图像进行视觉编码, 产生不同尺度的图像特征图, 为下游任务提供多维度的图像特征。其中, 共享参数的主干网络的流程为: 原始图像 i 裁剪成多个不重叠的小块。其中, 被裁剪的小块局部图像大小为 4×4 , 每

个小块拥有 3 个通道, 所以每个小块图像具有 $4 \times 4 \times 3 = 48$ 的特征维度。

然后对 48 维度特征的每个小块图像组成的序列经四个 Swin Transformer Block 模块依次处理。第一次处理,

首先, 使用线性位置编码器, 对图像中的各小块进行位置编码。然后, 各小块经过 Swin Transformer Block 得到语义特征图, 其维度为 C 。第二阶段至第四阶段同理, 对小块的特征进行融合, 然后使用 Swin Transformer Block 模块, 对图像特征进行 2 倍上采样, 同时特征维度 C 依次扩展为 2 倍。最后, 图像特征经过全连接线性层 $fc^{[13]}$ 映射为特征向量。

$$f_i = fc(\text{swint}(i)) \quad (1)$$

其中: swint 表示 Swin Transformer 网络。

2.1.2 多任务集

通过任务拆分的方式, 输电线路小金具缺失的数据采用子任务标注的方式, 对每个任务进行标记, 并通过多任务学习的方式进行训练。我们将小金具缺失拆分成三个学习任务分别为双螺母缺失任务、开口销缺失任务、螺栓类型任务, 三个任务的数据集被选择成为实验数据集。双螺母缺失任务的任务数据集 task_d_dataset 由输电线路螺栓图像和图像标签对组成, 每个图像是 RGB 的三通道图像, 标签有三个类别: 存在、不存在、未知, 分别表示对应的图像存在双螺母、不存在双螺母、由于视角或模糊原因导致的无法判断是否存在双螺母。



图 4 双螺母缺失任务数据及标签

开口销缺失任务的任务数据集 task_p_dataset 由输电线路螺栓图像和图像标签对组成, 每个图像是 RGB 的三通道图像, 标签有三个类别: 存在、不存在、未知, 分别表示对应的图像存在开口销、不存在开口销、由于视角或模糊原因导致的无法判断是否存在开口销。



图 5 开口销缺失任务数据及标签

螺栓类型任务的任务数据集 task_bt_dataset 由输电

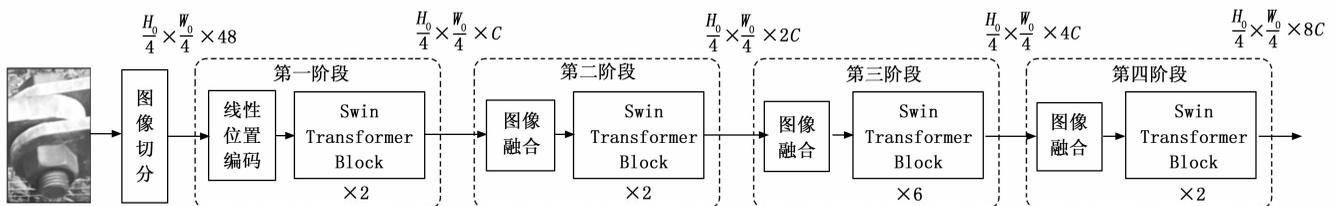


图 3 Swin Transformer 主干网络结构图

线路螺栓图像和图像标签对组成，每个图像是 RGB 的三通图像，标签为 11 种类别的螺栓分别表示螺栓所属种类。

2.1.3 多任务学习分类模块

主干网络对原始图像进行编码得到图像特征，将图像特征输入到多个多层感知机网络头，每个感知机网络头用于学习一个小金具缺失的识别任务。本研究采用的多层感知器 MLP^[15-16] (Multi-Layer Perception) 如图 6，作为小金具检测任务头网络。

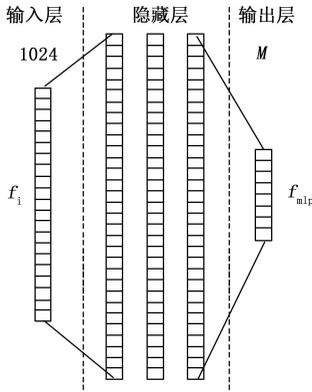


图 6 MLP 结构图

网络可表示如下：

$$f_{mlp1} = MLP1(f_i), f_{mlp1} \in R^N * M_1 \quad (2)$$

$$f_{mlp2} = MLP2(f_i), f_{mlp2} \in R^N * M_2 \quad (3)$$

$$f_{mlp3} = MLP3(f_i), f_{mlp3} \in R^N * M_3 \quad (4)$$

式中， f_{mlpi} 为小金具相关任务头网络输出的识别特征， M_i 为小金具相关任务对应的输出标签数量。

使用 softmax 函数作为激活函数，将每个任务头输出的 f_{mlpi} 转换为对应任务的分类概率 p_i 。

2.1.4 损失函数

在模型训练中，针对不同的子任务分别计算其网络损失。本研究采用 CE (Cross Entropy)^[17-18] 损失函数进行计算。计算方法为，在得到每一类的小金具预测标签后，将其与标注的正确标签计算 CE 损失。计算过程如下：

$$L_{task} = - \sum_i y_i \log(p_i) \quad (5)$$

$$L = \sum L_{task} \quad (6)$$

式中， y_i 为 task 小金具任务数据集的标注标签的索引， p_i 为小金具任务头网络预测的识别类别概率。

2.2 加速推理

为了提升模型的推理效率，我们采用 NVIDIA 开发的 Triton Inference Server 的方式进行模型的多任务推理。通过推理请求的批处理，Triton 可以将算法请求传输给不同任务的后端。后端通过批处理集中请求成批处理的方式进行推理，并返回相对应的输出。如图 1 所示，模型的结构可以拆分为共享参数的主干网络和各自的多任务头。各模块分别在 Triton 模型库中注册。主干网络接受推理图片作为输入，输出图片的特征向量 f_i 。每个任务头以 f_i 作为输入，并返回推理的类别及置信度。多个并行的任务头与主

干模型通过后连接的方式，形成新的推理模型。不同人物之间采用异步推理的方式进行推理，新模型负责控制整个服务的输入和输出。

2.3 小金具缺失的扩展算法任务

2.3.1 单螺母缺失的扩展任务数据集

本研究提出的共享参数多任务学习模型脚本单任务与其他任务解耦的能力，可以做到在训练完成的任务集中单独配置训练新人物。由于多任务之间只有主干网络共享，而子任务之间相互独立，因此在已完成的主干网络和原始任务基础上，增加了对单螺母缺失任务的支持。首先准备单螺母缺失任务的数据集，准备方法与其他任务相同。单螺母缺失任务的任务数据集 task_s_dataset 由输电线路螺栓图像和图像标签对组成，每个图像是 RGB 的三通道图像，标签有三个类别：存在、不存在、未知，分别表示对应的图像存在单螺母、不存在单螺母、由于视角或模糊原因导致的无法判断是否存在单螺母。



图 7 单螺母缺失任务数据及标签

2.3.2 单螺母缺失扩展任务学习

单螺母缺失的扩展任务，添加一个单螺母的任务头 MLP，

$$f_i = fc(swint(i)) \quad (7)$$

$$f_{mlps} = MLP_s(f_i) \quad (8)$$

式中， MLP_s 的参数是可学习的参数，而其他部分的参数则采用原模型的参数。通过使用损失函数独立计算小金具缺失扩展任务损失，对 MLP_s 的参数进行优化更新，从而实现小金具缺失扩展任务的学习。

2.4 模型算法评价指标

2.4.1 精度评价指标

本研究使用与学术界其他目标检测相同的评估方式，即识别精度、召回、以及识别准确率。其中，识别精度是正确的数量在被识别的总数中所占的比例，召回率则是识别正确的数量占该类别的比例。其中，置信度设置为 0.25，置信度是超参数，用于表示模型识别结果的概率。我们将结果概率低于 0.15 的类别过滤。准确率则是预测正确的占所有样本的比率。

$$P = \frac{TP}{TP + FP} \quad (9)$$

$$R = \frac{TP}{TP + FN} \quad (10)$$

$$acc = \frac{TP + TN}{TP + FN + FP + TN} \quad (11)$$

式中， TP 表示正确检测到的数量， FN 表示未检测到的数

量, FP 表示错误检测到数量, TN 表示正确检测到的负类样本数量。

2.4.2 性能评价指标

推理性能在吞吐量^[23-24]和显存^[25]占用量上得以表现。在本研究中, 我们采用了这两个指标来评价推理性能。具体而言, 我们使用每秒处理图像数量的 FPS 进行评估, FPS 越高, 表示模型的推理速度越快, 效率越高。此外, 模型推理时的显存占用量也是模型的评估指标之一, 用来衡量对 GPU 的要求。显存占用量于晓, 表示模型对推理资源的利用效率越高。因此, 我们综合使用 FPS 和显存占用量来综合评估模型的推理性能。

3 实验结果与分析

3.1 实验设备及参数

小金具缺失算法实验在 GPU 服务器 Linux 操作系统上实现。采用 Python 和 PyTorch 完成实验代码。

表 1 实验环境及参数配置

设备名称	型号
CPU 芯片	64 Intel(R) Xeon(R) Silver 4314 CPU @ 2.40 GHz
GPU 显卡 训练/推理	1 * NVIDIA A100 /A2
内存大小	固态 377 G
GPU 显存训练/推理	40 G/16 G
OS	CentOS 7.6

3.2 数据集准备

输电线路小金具缺失数据集采用无人机巡检数据和模拟无人机巡检的人工采集数据。由于无人机在真实输电线路巡检数据中小金具缺失的样本量极少, 为了平衡训练的数据, 人工模拟无人机巡检采集小金具缺失的数据样本。

表 2 螺栓类型任务数据分布

螺栓类型	训练集	验证集
塔身垂直螺栓	1 328	148
塔身非垂直螺栓	2 381	265
塔身 U 型螺栓	296	33
活动件螺栓	1 022	114
线夹螺栓	980	109
线夹无螺母螺栓	202	22
重锤螺栓	264	29
均压环螺栓	235	26
跳线螺栓	426	47
绝缘子盘螺栓	212	24
绝缘子槽碗螺栓	406	45

对准备好的输电线路原始数据进行多任务学习标注, 双螺母缺失任务、开口销缺失任务、螺栓类型任务以及扩展的单螺母缺失任务。多任务学习数据集采用 OmVision Studio 的数据集模块对小金具缺失的每个人物数据集进行标注。

在人工标注结束后, 我们使用 json 格式保存所有的图片和标注信息。数据处理时通过读取和处理标注框坐标,



图 8 OmVision Studio 标注界面

得到多任务学习的监督数据。

表 3 单螺母、双螺母、开口销缺失任务数据分布

任务	类别	训练集	验证集
单螺母缺失	存在	5 925	658
	不存在	2 097	233
	未知	636	70
双螺母缺失	存在	1 458	162
	不存在	5 961	662
	未知	627	69
开口销缺失	存在	2 933	325
	不存在	3 932	436
	未知	1 205	133

3.3 实验结果

3.3.1 单任务学习和多任务学习结果分析

在实验中, 我们采用单任务学习和多任务学习两种方式的精确率、召回率以及准确率来评估验证多任务学习在输电线路小金具这个应用场景缺失检测模型的效果。具体来说, 我们使用了 Swin Transformer (Tiny) 作为模型的主干网络, 并设置输入图像大小为 224×224 , 图像分块大小为 4, 输入通道数为 3, 图像分块的特征维度为 256。对于每一个子任务, 本研究使用 MLP 分类器模块来完成单个子任务的训练。此外, 单任务和多任务重相同的任务采用了相同的 MLP 模型结构。模型的训练在一台 Nvidia A100 GPU 的服务器进行训练。在小金具缺失的单任务学习中, 设置批处理大小 (Batch-Size) 为 64, 并将相应任务的训练集数据进行 100 个 epoch 的迭代训练, 学习率设置为 0.001。在小金具缺失的多任务学习, 设置与单任务相同。在验证阶段, 输入图片调整为 224×224 大小。图 9 展示了训练过程中 loss 的变化情况。

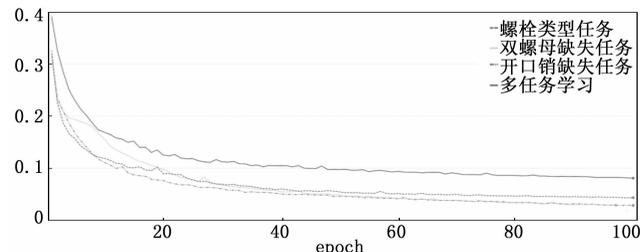


图 9 训练过程中多任务学习与单任务学习的 loss

我们设置模型训练 100 个 epoch 并测试不同任务和不同学习方式的精度，结果见表 4。

表 4 不同任务和不同学习方式的精度

学习方式	任务类型	精确率 P	召回率 R	准确率 Acc
单任务学习	螺栓类型	62.6	73.5	78.5
	双螺母缺失	66.3	69.8	88.2
	开口销缺失	75.4	73.5	85.3
多任务学习	螺栓类型	60.9	75.4	74.6
	双螺母缺失	64.2	63.4	84.5
	开口销缺失	75.6	71.2	83.6

通过结果对比可知，多任务学习方式螺栓类型、双螺母缺失、开口销缺失的精确率比单任务学习方式降低了 2% 左右、螺栓类型任务的召回率升高了 1.9%。多任务学习的螺栓类型任务、双螺母缺失任务、开口销缺失任务的准确率较单任务学习分别降低了 3.9%，3.7%，1.7%，说明输电线路小金具缺失多任务学习相对于单任务学习的准确率降低，这是由于多任务共享同一个主干网络的参数，主干网络参数对特定的任务提取的特征能力有所下降。由于主干网络的参数量大，约占网络模型参数的 80%，通过减少整体模型参数节约大量的训练时间和计算资源，仅牺牲了少量准确率。

3.3.2 单任务推理和多任务推理性能结果分析

本文通过对比实验分析，对单任务推理和多任务推理的速度和显存占用情况进行评估。其中，一台安装 Nvidia A2 GPU 的 linux 服务器设备用来作为推理设备。使用单任务推理时，为了提高模型的推理效率，我们将每个单独的子任务模型转为 onnx 模型文件，并配置到 Triton 型库中，使用 Triton 推理服务的 onnx 推理后端进行推理。对于多任务的推理，我们将共享参数的主干网络和每个任务对应的 MLP 模块任务头分别转换为 onnx 模型文件，并对其进行配置。此外，Triton 模型配置文件同样通过将主干网络和 MLP 模块组合的方式进行推理。体重，子任务头包括螺栓类型任务头、双螺母缺失任务头以及开口销缺失任务头。在推理过程中，图片输入大小为 224 * 224，推理的 batch size 设置为 4。

表 5 不同推理方式的推理性能

推理方式	任务类型	FPS	显存占用
单任务推理	螺栓类型	6.4	3 980
	双螺母缺失	6.8	3 720
	开口销缺失	7.2	3 720
	串联总和	2.2	11 420
多任务推理	螺栓类型	5.1	8 903
	双螺母缺失		
	开口销缺失		

通过对单任务推理和多任务推理性能的实验结果进行分析，可以得出以下结论：在单任务推理中，螺栓类型任务的 FPS 为 6.4，双螺母缺失任务的 FPS 为 6.8，开口销缺

失任务的 FPS 为 7.2，三个任务完成一张图片所需的时间约为 442.1 毫秒，FPS 为 2.2。而在小金具缺失的多任务推理中，总的 FPS 为 5.1，一张图片推理时间约 196 ms。可以看出，多任务推理相较于单任务推理，推理速度提升了 2.25 倍。由此可见，本文提出的小金具缺失多任务学习的加速推理算法的性能得到显著提升。小金具缺失推理时多任务比单任务显存占用降低了 22%。

3.3.3 扩展任务结果分析

3.3.3.1 扩展任务及原任务精度结果分析

为了验证多任务学习的任务可扩展性，本研究进行了多任务并行学习和单任务学习的对比实验。在多任务学习中，我们使用了小金具螺栓缺失扩展任务数据集，并按照 3.3.1 节中相同的参数进行训练。与此同时，我们使用相同参数进行了单任务学习作为对比。在单任务学习中，采用了与前述单任务学习相同的训练方式。而在扩展任务学习中，我们冻结了主干网络。

表 6 不同学习方式下的小金具扩展任务精度

学习方式	任务类型	精确率 P	召回率 R	准确率 Acc
扩展任务	单螺母缺失	60.9	71.2	73.5
单任务学习	单螺母缺失	62.3	73.6	78.1

根据实验结果可以得出结论，当在冻结主干网络参数的条件下进行扩展任务学习时，其精确、召回和准确率稍微小于单任务学习。但下小金具缺失的扩展任务学习的优势在于仅训练 MLP 任务头。

然而，多任务学习的优势为在对任务进行扩展，以及对新的任务进行冷启动时，无需训练主干网络，只需要训练一个 MLP 任务头，即可灵活新增子任务，并无缝添加到现有的推理模块。

3.3.3.2 扩展任务及原任务推理性能结果分析

在上一节中，验证了扩展任务学习的精度。本节将对扩展任务的推理性能进行测试。将单螺母缺失扩展任务的头转换为 onnx 模型文件，并将模型文件注册和配置到 Triton 的模型库和配置文件中。添加单螺母缺失扩展任务头，并进行了相应的测试。

表 7 在三个原任务基础上不同推理方式的扩展新增单螺母缺失任务后的推理性能

扩展方式	扩展前 FPS	扩展后 FPS	显存占用
单任务扩展	2.2	1.7	15 140
多任务扩展	5.1	4.6	10 009

通过扩展任务的实验结果，观察到在多任务推理中增加了单螺母缺失任务后，总体 FPS 为 4.6，而单独进行单螺母缺失任务的 FPS 为 6.9。根据 3.3.2 的实验结果，可以计算出单任务推理完成螺栓类型、双螺母缺失、开口销缺失和单螺母缺失这四个任务的推理时间为 587.0 毫秒，相应的 FPS 为 1.7。而多任务推理完成全部任务的时间为 217.3 毫秒，速度提高了 2.7 倍。从显存占用量的角度来看，单任

务推理共占用了 15 140 MB 的显存, 而多任务推理仅占用了 10 009 MB 的显存, 显存占用降低了 34%。

5 结束语

本文针对输电线路小金具缺失检测问题提出了一个可插拔的多任务学习加速推理模型。该模型能够通过主干网络共享模型参数的方式, 显著提升模型参数的使用效率, 提升模型训练的灵活性, 提升推理速度, 同时降低显存占用。

此外, 本研究验证了多任务学习的方式用来有效提升算法任务的训练部署效率。在不明显降低模型效果的情况下, 本研究提出的算法模型能够解耦不同任务的训练推理, 使得输电线路的缺陷检测任务配置更加灵活。然而, 本算法牺牲了一定的推理精度, 对于对推理精度要求较高的任务, 是一个挑战。因此, 在后续的研究中, 我们将致力于减少多任务学习对推理精度的损失。

参考文献:

[1] 武学良, 戚银城. 基于深度学习的输电线路螺栓缺陷分类与检测方法 [J]. 集成电路应用, 2021, 38 (4): 64-65.

[2] 刘云鹏, 裴少通, 武建华, 等. 基于深度学习的输变电设备异常发热红外图片目标检测方法 [J]. 南方电网技术, 2019, 13 (2): 7.

[3] 赵振兵, 张 薇, 翟永杰, 等. 电力视觉技术的概念, 研究现状与展望 [J]. 电力科学与工程, 2020, 36 (1): 8.

[4] 胡 毅, 刘 凯, 吴 田, 等. 输电线路运行安全影响因素分析及防治措施 [J]. 高电压技术, 2014, 40 (11): 9.

[5] 戚银城, 郎静宜, 赵振兵, 等. 结合注意力机制的相对 GAN 螺栓图像生成 [J]. 电测与仪表, 2019, 56 (19): 6.

[6] 张 妹, 王昊天, 董晓翀, 等. 基于深度学习的输电线路螺栓检测技术. 电网技术, 2021, 45 (7): 2821-2828.

[7] 赵振兵, 张 帅, 蒋 炜, 等. 基于 DBSCAN-FPN 的输电线路螺栓缺陷检测方法 [J]. 中国电力, 2021, 54 (3): 45-54.

[8] 赵振兵, 段记坤, 孔英会, 等. 基于门控图神经网络的栓母对知识图谱构建与应用 [J]. 电网技术, 2021, 45 (1): 98-106.

[9] 欧阳宁, 马玉涛, 林乐平. 基于多任务学习的多姿态人脸重建与识别 [J]. 计算机应用, 2017, 37 (3): 5.

[10] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C] //European conference on computer vision, Springer, Cham, 2016: 21-37.

[11] REDMON J, FARHADI A. Yolov3: An incremental improvement [J]. arXiv preprint arXiv: 1804.02767, 2018.

[12] KOAY H V, CHUAH J H, CHOW C O. Shifted-Window Hierarchical Vision Transformer for Distracted Driver Detection [C] //2021 IEEE Region 10 Symposium (TENSYP), IEEE, 2021: 1-7.

[13] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60 (6): 84-90.

[14] ELFADEL I M, WYATT J L. The Softmax Nonlinearity: Derivation Using Statistical Mechanics and Useful Properties as a Multiterminal Analog Circuit Element [J]. advances in neural information processing systems, 1994: 882-887.

[15] WANG Z, YANG J, DAI M, et al. A Method of Detecting Webshell Based on Multi-layer Perception [J]. Communications Technology, 2019, 2 (1): 81-91.

[16] TOLSTIKHIN I O, HOULSBY N, KOLESNIKOV A, et al. Mlp-mixer: An all-mlp architecture for vision [J]. Advances in Neural Information Processing Systems, 2021, 34: 24261-24272.

[17] DE BOER P T, KROESE D P, MANNOR S, et al. A tutorial on the cross-entropy method [J]. Annals of operations research, 2005, 134 (1): 19-67.

[18] SHORE J, JOHNSON R. Properties of cross-entropy minimization [J]. IEEE Transactions on Information Theory, 1981, 27 (4): 472-482.

[19] 张智坚, 曹雪虹, 焦良葆, 等. 基于改进 YoloX 的输电通道工程车辆检测识别 [J]. 计算机测量与控制, 2022, 30 (9): 7.

[20] 李 飞, 王 超, 浦 东, 等. 基于卷积注意力的输电线路防震锤检测识别 [J]. 计算机测量与控制, 2022, 30 (3): 48-53, 105.

[21] 毛 元, 张 斌. 基于神经网络的高压输电线路单端测距方法 [J]. 计算机测量与控制, 2015, 23 (10): 3316-3318.

[22] 林礼健, 陈 灵, 沈明松, 等. 基于无人机智能视觉的输电线路全息全景重建 [J]. 计算机测量与控制, 2018, 26 (1): 5.

[23] 周立君, 刘 宇, 白 璐, 等. 使用 TensorRT 进行深度学习推理 [J]. 应用光学, 2020, 41 (2): 337-341.

[24] 白首华, 胡天彤, 李 响. 用于车联网的深度学习离线推理负载均衡方法: CN110874270A [P]. 2020.

[25] 孟 伟, 袁丽雅, 韩炳涛, 等. 深度学习推理侧模型优化架构探索 [J]. 信息通信技术与政策, 2020 (9): 42-47.

[13] 吕镇邦, 孙 倩, 李 鹏. PHM 知识工程与数据建模 [J]. 航空计算技术, 2021, 51 (1): 125-129.

[14] 许 晴, 孟 上, 许爱强. 基于 WLAN 的分布式混合总线 ATS 设计与实现 [J]. 计算机测量与控制, 2015, 23 (4): 1137-1140.

[15] 张 雷, 李志栋, 孙 波. 航天器健康管理平台设计技术研究 [J]. 计算机测量与控制, 2019, 27 (10): 18-22.

[16] 王建军, 向永清, 何正文. 航天器试验大数据系统实现与典型应用 [J]. 计算机测量与控制, 2019, 27 (4): 254-257.

[17] 陈 强, 陈 双, 吴立金, 等. 分布式复杂系统软件测试建

ment [J]. arXiv preprint arXiv: 1804.02767, 2018.

[12] KOAY H V, CHUAH J H, CHOW C O. Shifted-Window Hierarchical Vision Transformer for Distracted Driver Detection [C] //2021 IEEE Region 10 Symposium (TENSYP), IEEE, 2021: 1-7.

[13] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60 (6): 84-90.

[14] ELFADEL I M, WYATT J L. The Softmax Nonlinearity: Derivation Using Statistical Mechanics and Useful Properties as a Multiterminal Analog Circuit Element [J]. advances in neural information processing systems, 1994: 882-887.

[15] WANG Z, YANG J, DAI M, et al. A Method of Detecting Webshell Based on Multi-layer Perception [J]. Communications Technology, 2019, 2 (1): 81-91.

[16] TOLSTIKHIN I O, HOULSBY N, KOLESNIKOV A, et al. Mlp-mixer: An all-mlp architecture for vision [J]. Advances in Neural Information Processing Systems, 2021, 34: 24261-24272.

[17] DE BOER P T, KROESE D P, MANNOR S, et al. A tutorial on the cross-entropy method [J]. Annals of operations research, 2005, 134 (1): 19-67.

[18] SHORE J, JOHNSON R. Properties of cross-entropy minimization [J]. IEEE Transactions on Information Theory, 1981, 27 (4): 472-482.

[19] 张智坚, 曹雪虹, 焦良葆, 等. 基于改进 YoloX 的输电通道工程车辆检测识别 [J]. 计算机测量与控制, 2022, 30 (9): 7.

[20] 李 飞, 王 超, 浦 东, 等. 基于卷积注意力的输电线路防震锤检测识别 [J]. 计算机测量与控制, 2022, 30 (3): 48-53, 105.

[21] 毛 元, 张 斌. 基于神经网络的高压输电线路单端测距方法 [J]. 计算机测量与控制, 2015, 23 (10): 3316-3318.

[22] 林礼健, 陈 灵, 沈明松, 等. 基于无人机智能视觉的输电线路全息全景重建 [J]. 计算机测量与控制, 2018, 26 (1): 5.

[23] 周立君, 刘 宇, 白 璐, 等. 使用 TensorRT 进行深度学习推理 [J]. 应用光学, 2020, 41 (2): 337-341.

[24] 白首华, 胡天彤, 李 响. 用于车联网的深度学习离线推理负载均衡方法: CN110874270A [P]. 2020.

[25] 孟 伟, 袁丽雅, 韩炳涛, 等. 深度学习推理侧模型优化架构探索 [J]. 信息通信技术与政策, 2020 (9): 42-47.

[18] 阎 露, 杜承烈, 魏学杭. 基于自标识的航空电子系统 ICD 定义方法 [J]. 航空计算技术, 2014, 44 (4): 109-113.

[19] 邹会荣, 吕镇邦, 沈新刚. 机载维护仿真平台的设计与验证 [J]. 计算机测量与控制, 2019, 27 (7): 214-217.

[20] 张耀鸿, 罗雪山, 余 滨. 基于组件的分布离散事件仿真环境 [J]. 系统仿真学报, 2002 (8): 1019-1021.

[21] 王连浩, 王占选, 齐非凡, 等. 基于 Simulink 的数模混合电路仿真平台设计与实现 [J]. 计算机测量与控制, 2022, 30 (9): 148-154.