第 50 卷 第 5 期：1813-1825 高电压技术 Vol.50, No.5: 1813-1825
2024 年 5 月 31 日 High Voltage Engineering May 31, 2024
DOI: 10.13336/j.1003-6520.hve.20232149

大模型时代：电力视觉技术新起点

赵振兵[1,2,3]，冯 烁[1]，席 悦[1]，张靖梁[1]，翟永杰[4]，赵文清[2,4]
（1. 华北电力大学电气与电子工程学院，保定 071003；2. 华北电力大学复杂能源系统智能计算教育部工程研究中心，保定 071003；3. 华北电力大学河北省电力物联网技术重点实验室，保定 071003；4. 华北电力大学控制与计算机工程学院，保定 071003）

摘 要：随着无人机、巡检机器人和远程监控系统在输电、变电、配电、安监等电力场景中的广泛应用，利用电力视觉技术完成对海量巡检图像的自动处理，能够进一步提升电力系统智能化运维水平，对我国源网荷储一体化进程的快速推进具有至关重要的作用。随着通用视觉大模型的兴起，电力视觉技术正处于从传统深度学习时代向大模型时代跨越的重要节点。该文首先综述了电力视觉技术和通用视觉大模型的最新研究进展，结合视觉大模型在多种公共场景的应用先例，探讨视觉大模型在电力视觉领域将面临的 3 重能力边界问题。从初步探索通用视觉大模型的潜力，到逐步构建电力视觉大模型的过程，提出 4 种模型应用范式以突破视觉大模型能力边界。最后分析了视觉大模型对电力视觉研究者的影响，并对大模型浪潮下电力视觉技术的发展方向进行了展望。
关键词：电力视觉；视觉大模型；目标检测；图像分割；深度学习；图像处理

The Era of Large Models: A New Starting Point for Electric Power Vision Technology

ZHAO Zhenbing[1,2,3], FENG Shuo[1], XI Yue[1], ZHANG Jingliang[1], ZHAI Yongjie[4], ZHAO Wenqing[2,4]
(1. School of Electrical and Electronic Engineering, North China Electric Power University, Baoding 071003, China; 2. Engineering Research Center of Intelligent Computing for Complex Energy Systems of Ministry of Education, North China Electric Power University, Baoding 071003, China; 3. Hebei Key Laboratory of Power Internet of Things Technology, North China Electric Power University, Baoding 071003, China; 4. School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China)

Abstract：With the widespread applications of drones, inspection robots, and remote monitoring systems in power scenarios such as transmission, transformation, distribution, and safety supervision, the power vision technology is utilized to automatically process massive inspection images, thus the intelligent operation and maintenance level of power systems can be further enhanced, playing a crucial role in the rapid advancement of China's source-grid-load-storage integration process. With the rise of general-purpose large vision models, the electric power vision technology is embarking on an important transition from the traditional deep learning era to the era of large models. In this paper, the latest research progress in electric power vision technology and general-purpose large vision models is summarized. Combined the precedents of large vision models' applications in various public scenarios, three main capability boundary issues that large vision models will face in the field of power vision are investigated. From initially attempting to apply general-purpose large vision models to establishing power-specific large vision models, the four model application paradigms to break through the capability boundaries of large vision models are proposed. Finally, the impact of large vision models on electric power vision researchers is analyzed, and prospects in the development direction of electric power vision technology under the wave of large models are put forward.
Key words：electric power vision; large vision models; object detection; image segmentation; deep learning; image processing

基金资助项目：国家自然科学基金(U21A20486；62373151；62371118；62303184)；河北省自然科学基金(F2021502008；F2021502013)；中央高校基本科研业务费专项资金(2023JC006)。
Project supported by National Natural Science Foundation of China (U21A20486, 62373151, 62371118, 62303184), Natural Science Foundation of Hebei Province (F2021502008, F2021502013), Fundamental Research Funds for the Central Universities (2023JC006).

0 引言

在“双碳”目标下新型电力系统[1]正在快速发展，新能源大规模接入使电网的复杂性、不确定性不断增加，电力系统智能化程度必须进一步提高[2]。电力视觉是一种利用图像处理、深度学习、模式识别等技术，结合电力专业领域知识，解决电力系统各环节视觉问题的电力人工智能技术[3]。作为提升巡检与安监工作智能化程度的核心技术之一，电力视觉技术在安全监测、设备检测、缺陷诊断等方面均发挥着重要作用[4]。电力视觉技术的持续进步，为无人机、智能监控系统和智能巡检机器人在输配电线路巡检和变电站运维场景中的部署提供了有力支持。

然而，随着电力行业对精细化、智能化需求的不断提升，现有电力视觉技术已难以满足现代电力行业的发展需求。目前，电力视觉技术面临诸多挑战和问题。首先，由于电力行业的场景和设备复杂多变，数据采集和标注难度较大，直接影响了模型的训练效果和准确性。其次，电力视觉技术主要依赖深度学习算法，但目前电力领域普遍使用的深度学习方法距离可信人工智能仍有差距。因此，基于电力视觉技术对电力设备进行状态监测和故障诊断的过程仍然需要人工辅助。

随着人工智能技术的持续发展，计算机视觉领域陆续出现了对比语言-图像模型(contrastive language-image pre-training，CLIP)[5]、分割一切模型(segment anything model，SAM) [6]和一次性分割一切模型(segment everything everywhere all at once，SEEM) [7]等视觉大模型，由于其强大的特征表达和零样本泛化能力，这些大模型在分类、检测和分割等视觉任务上普遍超越了此前的深度学习模型。目前普遍认为大模型的优异性能来自其“涌现”能力，即模型参数量与训练数据量的规模超越某个数量级后，模型会表现出极强的特征提取能力与逻辑推理能力，因此大模型普遍具有百亿甚至千亿级的参数量。视觉大模型技术的兴起[8]，为电力视觉技术的发展带来了新的机遇和挑战。

视觉大模型的出现打破了“数据标注、模型训练、模型改进、训练调优”的传统研发模式，在电力设备图像分类、检测、分割等基础视觉任务中，预训练后的视觉大模型即使不在特定场景标注数据上进行训练，也能够达到甚至超越全监督模型的预测精度。同时，电力领域的图像数据因保密要求不能公开流通，大多数研究者只拥有少量特定任务数据，导致对大模型进行训练时往往会受限于数据量和算力。探索大模型在电力视觉领域的应用场景与应用范式，在小规模电力图像和常规算力条件下最大程度发挥视觉大模型的图像处理、理解、推理能力，将成为大模型时代推动电力视觉技术持续发展的关键任务。

1 电力视觉技术发展现状

随着我国电力行业的高速发展，输电线路与变电站数量不断增加，传统的人工巡检方式已经不能满足电网的运行维护需求。因此，基于无人机、巡检机器人和智能监控系统的巡检方式因成本低、周期短和安全性高等优点，逐步取代了传统的人工巡检手段。部署在输变电场景下的巡检设备在运行过程中产生了大量的图像，依靠人工处理海量巡检图像并不现实[9]。基于深度学习的电力视觉技术能够显著提升巡检图像的处理效率，并且比传统图像处理方法更具鲁棒性与准确性，因此电力视觉技术被广泛应用于输、变电场景下的设备缺陷检测任务。随着电网智能化程度的提高，如图 1 所示，电力视觉技术的应用面也从早期的输、变电设备检测扩展到多种电力场景安全隐患检测。

1.1 输电设备及其缺陷检测

输电线路巡检过程中，电力视觉技术主要用于对金具和螺栓进行缺陷检测。金具主要用于支持、固定、连接裸导线、导体及绝缘子等[10]，而螺栓的作用是紧固连接各种金具，二者易因环境因素和受力影响出现损坏。因此，及时对巡检图像进行高精度智能分析是保证输电线路安全运行的重要环节。

针对航拍图像背景复杂、多种金具形状多样且存在目标遮挡等问题，Zhai 等根据输电线路中金具的装配规则，提出了多类别金具之间的共现知识、语义知识和空间知识，使用图神经网络实现知识推理，提升目标检测模型对于少样本和遮挡目标的检测精度[11]。Liu 等同样考虑到输电线路金具之间具有装配关系，在目标检测模型中加入了语义引导模块，以类别标签的文本信息辅助模型进行分类，以图卷积网络对输电线路金具的共现关系进行学习，以增强模型的检测精度[12]。针对输电线路图像数据中缺陷样本与正常样本数量不平衡的问题，赵振兵等提出了结构推理模块来改善模型对少样本的漏检

和误检情况，该模块将目标检测模型初步输出的类别与位置信息作为向量输入，先后经过双向堆叠门循环单元和自注意力层对预测结果进行修正，得到更为准确的检测结果[13]。

针对输电线路螺栓缺陷图像存在类内差异大，类间相似性高的问题，赵振兵等将螺栓结构中易丢失的销子、垫片和螺母划分为多个标签，将多个缺陷出现的关系作为领域知识来指导级联分类头对缺陷图像进行区分，同时考虑到同一类别的图像受拍摄角度影响较大，采用长短期记忆网络对划分为网格的特征图进行结构关系的学习[14]。针对部分螺栓缺陷存在视觉不可分的问题，赵振兵等利用螺栓朝向、螺栓种类和金具提取数据集中的标签共现知识，使用图卷积神经网络进行知识推理，完成对多种金具上的螺栓缺陷识别[15]。

1.2　变电设备及其缺陷检测

变电站场景下巡检人员利用无人机、智能巡检机器人和智能监控系统采集变电设备的图像信息，通过电力视觉技术对图像进行分析处理。由于变电设备出现部分故障时会导致设备温度异常上升，借助红外测温技术可以帮助巡检人员无接触地了解设备温度信息，辅助巡检人员判断设备运行状态，因此变电站场景下的无人机和智能巡检机器人普遍搭载了可见光与红外相机，智能监控系统也会采取可见光与红外双通道的运行方式[16]。根据处理图像的不同，变电场景下的电力视觉任务可分为基于可见光图像的设备外表面缺陷检测与基于红外图像的设备温度异常检测。

1.2.1　基于可见光图像的设备外观缺陷检测

变电站场景下，对可见光图像进行检测发现的外观缺陷包括多种电力设备外观机械性损坏和表面缺陷。其中机械性损坏主要包括仪表、端子箱等设备的破损；表面缺陷包括油枕、变压器等设备的锈蚀和渗漏油。

目前，变电站仪表存在的缺陷包括外壳或表盘破损、观察窗凝露或模糊和示数异常。针对变电站中各类仪表的外观缺陷图像特征变化多样与特征提取困难的问题，赵振兵等[17]首先将对比样本策略引入 YOLOX[18]模型的训练过程，通过缺陷样本与正常样本的对比学习，引导模型寻找具有缺陷判别能力的特征，并且利用全局上下文模块增强模型的特征提取能力，提升对比样本策略的效果，最后使用 SIoU 损失函数优化模型的训练过程，提升模型检测精度。考虑到仪表缺陷检测模型在移动边缘端设备

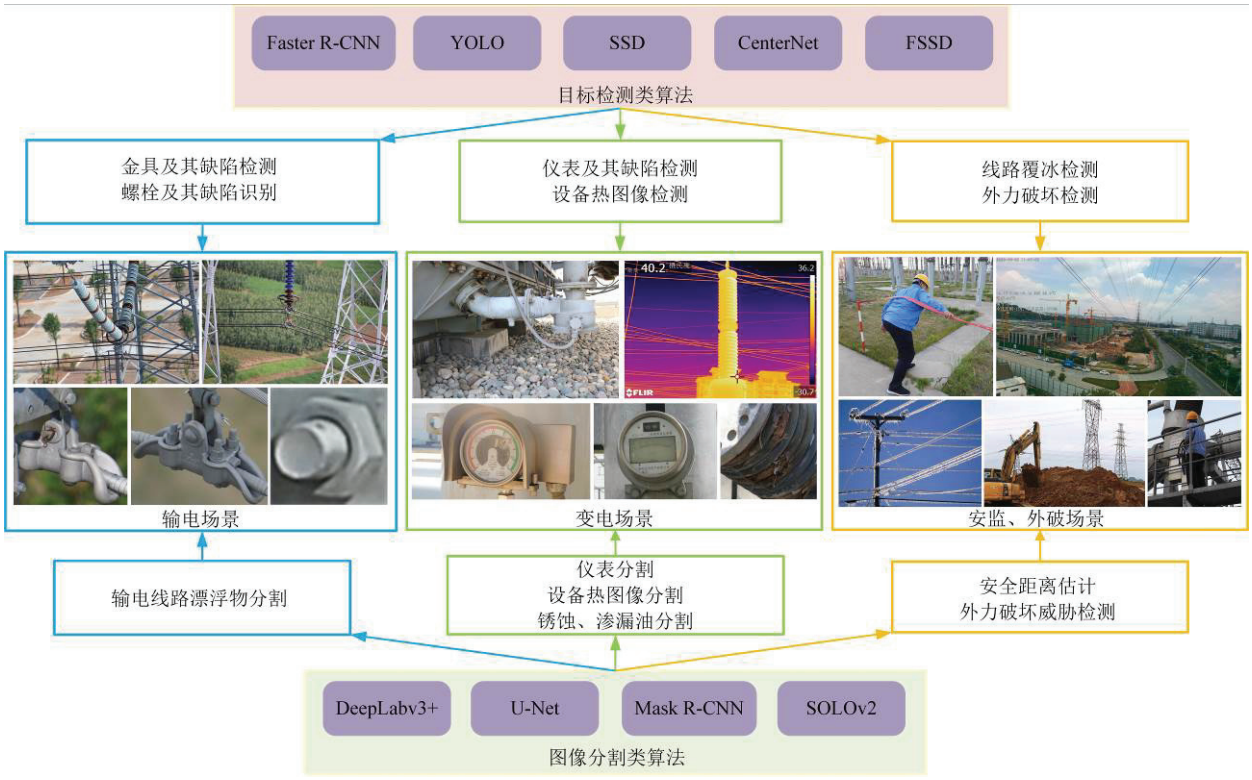


图 1　电力视觉应用场景与相关工作

Fig.1　Power vision application scenario and related work

上部署的需要，何敏等在 YOLOX 模型的骨干网络中引入了深度可分离卷积，并在特征融合网络中使用参数重组机制，提升模型的推理速度[19]，使用基于 Transformer[20]的池化模块对特征网络输出的深层特征进行细化来提升检测精度，该工作同样采取了缺陷样本与正常样本同时进行学习的对比样本策略。针对利用分割算法进行仪表精准检测时面临数据标注工作量过大的问题，滕国龙[21]借助 Faster R-CNN[22]模型从监控视频中提取仪表图像，采取半监督的训练方式，并利用交叉监督的双分支结构训练语义分割模型，实现了对指针刻度盘和表盘中心位置的分割，进而完成对仪表示数的提取。

锈蚀、渗漏油等变电设备表面缺陷具有外观极其不规则、特征多变且提取困难等特点，电力视觉技术中常使用的目标检测方法在该问题上监督强度过低，使得检测模型容易从学习缺陷特征偏移到更易拟合的设备特征学习。因此，基于像素级标注数据进行监督训练的语义分割方法在该问题上被广泛采用。针对变电设备渗漏油检测中存在的弱边缘与不规则分割难题，赵振兵等首先利用空间图推理支路建模像素间长距离依赖以增强边缘特征表征，并增设辅助损失提升网络边缘感知能力，其次利用语义图推理支路建模区域纹理的全局关系以增强纹理特征表征，从而实现渗漏油精准分割[23]。张迎晨等针对锈蚀与渗漏油图像均具有不规则外观，导致训练集和验证集分布差异过大的问题，提出了基于域适应模型的缺陷检测方法，首先将变电站场景下不同角度拍摄的图像分为源域和目标域数据，并为域适应模型添加纹理统计特征提取支路和提取局部形状特征的卷积块，二者作为辅助支路与原模型进行拼接，改善了模型对不规则区域纹理特征的提取能力[24]。赵文清等[25]在语义分割模型 DeepLabv3+[26]的基础上，将 Xception[27]作为骨干网络并在其尾端引入空洞金字塔模块来扩大模型感受野，使用通道与空间注意力增强模型的特征融合效果，提升基线模型对于变电设备渗漏油区域的分割精度。

1.2.2 基于红外图像的设备温度异常检测

利用红外图像对变电设备进行温度分析的前提是完成对设备的温度信息提取，以人工手动进行逐张图像的温度提取效率较低。现有工作普遍借助目标检测[28]和实例分割[29]来完成对设备区域的提取，借助设备的位置信息提取红外图像中的温度信息，用于后续分析设备温度状态。红外图像自身具有分辨率低、边缘信息模糊等特点，因此现有工作以精准定位设备为重点。Zheng 等[30]利用 Iresgroup 结构作为 CenterNet[31]算法的骨干网络，来提高算法的特征学习能力，并对超参数进行调整，实现了对变电设备红外图像的目标检测。Zheng 等[32]对单阶段检测模型 FSSD[33]的特征融合做了修改，利用骨干网络输出特征图重新组合了特征融合网络，在算法的浅层网络中加入特征增强模块，并借助聚类算法调整锚框长宽比，实现了对绝缘子和互感器红外图像的目标检测。

虽然利用目标检测算法对红外图像中变电设备进行检测时速度较快，但设备检测结果中含有大量背景信息，不利于后续对变电设备温度信息进行分析，而实例分割方法的预测结果更贴合变电设备边缘，可以为后续故障诊断提供更好的支持。Zhao 等首先提取多种变电设备的视觉显著特征，使用目标检测模型对视觉显著特征进行学习[29]，进而辅助 SOLOv2[34]实例分割模型完成对设备的分类和分割。Wang 等[35]使用 Mask R-CNN[36]算法对绝缘子热像进行实例分割，通过分割掩码来获取绝缘子的温度信息，最终完成对绝缘子的温度状态评估。

1.3 电力场景安全隐患检测

电力场景的安全隐患问题一直受到高度重视，近年来部分研究使用电力视觉技术完成对输电线路覆冰、输电走廊外破隐患和其他电力施工场景的安全隐患检测。马富齐等针对输电线路覆冰算法在边缘端设备上部署的问题[37]，利用轻量化模型 MobileNetv3[38]作为 SSD 模型[39]的骨干网络，提升基线模型检测速度，并将模型部署在边缘端设备完成对输电线路覆冰情况的检测。为提前发现输电通道中的安全隐患，杨知等通过空洞卷积池化金字塔模块和非局部特征提取模块[40]改进了 U-Net[41]语义分割模型，基于高分辨率卫星遥感影像完成对输电通道中易漂浮物的语义分割。郑含博[42]等针对输电线路在跨越铁路公路等场景下易受到外力破坏的问题，在 YOLOv6[43]模型的骨干网络中加入 CBAM[44]注意力模块，使用 softplus 激活函数替换了特征融合网络中的 ReLU 激活函数，提升基础模型对汽车、卡车和吊车的检测精度。Zheng 等[45]针对变压器与巡检人员之间的安全距离监测难以监测的问题，使用 Res2Net[46]和 BiFPN[47]分别作为 SOLOv2 的骨干网络和特征融合网络，改进后的 SOLOv2 被用来从监控图像中分割变压器和巡检人员，并通过深度估

计网络、深度重建和相机反投影变换实现对巡检人员和变压器的测距。

2　视觉大模型发展与应用现状

视觉是人类和动物获取外界信息的重要途径，而人工智能技术的核心目标之一就是设计视觉模型来模仿人类的视觉感知能力和视觉信号生成能力，与之对应，视觉模型可分为理解式模型和生成式模型，本文重点对理解式的视觉大模型进行介绍。

大模型又称基础模型，是在大量数据上训练的大型神经网络，旨在为各种下游任务提供一个统一的特征提取与推理框架。其核心思想在于通过多层次的神经网络结构，对大量复杂的视觉、语言数据进行处理和学习。大模型通过多层次的神经网络逐步学习图像中的特征，并逐渐提取更高级、更抽象的表征。这些特征表示使模型能够执行诸如图像分类、目标检测和语义分割等复杂的视觉任务。大模型与此前针对任务场景进行完整训练的深度学习模型不同，基础模型的特点是“大规模”与“预训练”，“大规模”意味着模型对数据的容纳与压缩能力更强，“预训练”强调模型可以直接应用或微调后用于下游任务。

2.1　视觉大模型研究进展

建立视觉大模型首先需要解决的问题是训练数据的来源和有效利用数据的训练方式。传统深度学习模型大多使用监督数据进行训练，但劳动密集型的数据标注方式难以应用在视觉大模型的训练任务中，因此现有研究大多探索如何使用来自公共领域的无标注或弱标注图像来训练视觉大模型。视觉大模型目前主要使用图像-文本对比监督、图像自监督和标签监督 3 种方式进行训练。

2.1.1　基于语言-图像对比学习的视觉大模型

基于语言-图像对比学习的视觉大模型大多使用来自互联网的图像与文本数据作为训练数据，将与图像匹配的文本作为监督信息，此类模型的训练方式兼具自监督学习和有监督学习的属性。此类模型的代表是 CLIP，其在包含 4 亿个图像-文本对的数据集上进行训练，通过图像编码器和文本编码器分别对输入数据进行特征编码，并使用相似度矩阵计算图像特征向量与文本特征向量之间的相似度，通过降低对比损失使模型学习到图像中目标与文本的相关性。CLIP 模型在图像分类任务上突破了闭集预测的限制，给模型提供了对新类别的文本描述即可实现对未知类别的图像分类。

2.1.2　基于图像自监督的视觉大模型

基于图像自监督的视觉大模型仅使用图像数据进行训练，这一方式称为掩码图像建模(masked ianguage modeling, MIM)，Bao 等首次在 BEiT[48]模型中使用 MIM 作为预训练方法。MIM 的设计受到 NLP 领域掩码语言建模(masked language modeling，MLM)方法的启发，MLM 方法基于文本数据自身具有监督信息这一特点，将文本中部分内容屏蔽后输入语言模型，训练模型预测被屏蔽内容的能力。MIM 的原理同样是“屏蔽后重建”，将图像的部分内容屏蔽并输入视觉模型，预测图像中缺失部分的信息或特征。He 等在 MAE(masked autoencoder)[49]中对输入图像分为切片(patch)后，随机采样选取少量 patch 输入视觉编码器，并将剩余 patch 屏蔽，解码器用一个可学习的特征向量和编码器输出的特征向量来预测被屏蔽部分的像素值，实现对图像的重建。MAE 将被屏蔽区域预测像素值与真实像素值之间的均方误差作为损失函数，并将输入图像归一化来优化训练效果。基于 MAE 完成预训练的视觉编码器可以直接与多种下游任务头进行结合，在监督数据集上进行微调即可完成图像分类、目标检测或语义分割等视觉任务。

2.1.3　基于标签监督的视觉大模型

基于标签监督的视觉大模型使用了与传统深度学习模型类似的标注数据进行训练，根据其采取的标注形式可分为目标检测模型和图像分割模型。GLIP(grounded language-image pre-training)模型[50]在 CLIP 的基础上引入物体位置信息来建立其区域-文本之间的联系，GLIP 首先在 1 个包含 300 万个精细标注的数据集上训练出 1 个教师模型，使用该模型对来自互联网的 2400 万个图文数据对进行预测，将预测结果作为伪标签再次对 GLIP 进行训练。GLIP 可利用文本提示进行目标检测，且具有较强的迁移学习能力。RegionCLIP(region-based language-image pretraining)[51]模型使用 CLIP 预训练的视觉编码器作为一个双阶段检测器的骨干网络，并使用训练好的区域建议网络对检测器输出特征进行定位，利用 CLIP 模型匹配图像与文本的能力对定位结果进行分类，将定位于分类的结果合并输出即可实现对图像的目标检测。该模型实现了利用图像文本对数据集训练目标检测模型，并且能够检测出与训练数据有相似概念的未知目标。

目前用于图像分割的视觉大模型普遍引入基于提示的交互式方法，通过外部提示来引导模型完成边缘检测、实例分割或开放词汇分割等任务。OpenSeeD(open-vocabulary segmentation and detection)模型[52]在Object365(365类，大约170万张图像)检测数据集和COCO2017分割数据集(133类，大约11万张图像)上进行训练，通过对文本和图像进行特征提取与特征对齐，实现对未知类别物体的检测与分割，该模型是第1个能够基于目标框提示进行交互分割的视觉大模型。为构建通用化的视觉大模型，Meta提出了分割一切模型SAM，该模型使用MAE预训练的ViT模型[53]作为视觉编码器，使用CLIP的文本编码器来处理输入的点、涂鸦或目标框等提示信息，使用卷积层来处理掩码输入，基于Transformer的解码器按照图像特征和输入提示完成图像分割，由于解码器只输出分割掩码，SAM并不能预测分割结果的类别。SAM模型在包含超过1000万张图像和10亿个像素标签的数据集上进行训练，对于公共领域的图像分割具有极强的零样本泛化能力，能够直接分割大量未知物体。Zou等提出的SEEM模型进一步提升了交互能力，SEEM能够接受点、目标框、涂鸦、掩码和文本形式的提示对图像进行分割，并且为每个预测掩码分配1个类别标签。SEEM模型在COCO2017数据集和RefCOCO+数据集(具有4万张图片，37万个图像表达式)上进行训练，其训练数据量远小于SAM模型。

2.2 视觉大模型在下游任务中的应用

现有视觉大模型已从CLIP等专用大模型发展到SAM等通用大模型，SAM等通用大模型不再聚集于解决图像分类等单一任务，而是通过多种外部提示实现图像分割，将预测结果应用于物体计数、目标跟踪和数据标注等任务。目前已有研究工作将SAM等视觉大模型应用到具体的下游任务中，根据应用大模型的方式可以将现有研究分为基于功能级联的大模型应用、基于知识蒸馏的大模型应用和大模型专业化3种类型。

2.2.1 基于功能级联的大模型应用

将具有特定功能的组件与视觉大模型进行级联是比较直接的应用方式，一般用于未标注图像数据量大但分割任务难度适中的场景。Giannakis等利用SAM模型完成对行星遥感图像的初步分割，将预测掩码输入形状筛选模块挑选出外形符合陨石坑的分割结果，实现对行星陨石坑的语义分割[54]。Williams等使用SAM完成对马铃薯叶片的图像分割，随后基于颜色阈值和形状过滤从分割掩码中筛选出对应绿色叶片的分割结果[55]。此类研究仅使用了SAM的零样本泛化能力，并未利用外部提示信息改善分割质量。Wang等利用眼动仪收集医生分析图像时的眼球活动，以此来确定图像被关注区域的大致位置，将位置信息转为点来引导SAM完成对关注区域的分割[56]。针对遥感图像分割难度较高的问题，Zhang等[57]将类别信息以文本的形式输入Grounding DINO[58]模型和CLIP Surgery[59]模型，两个模型分别输出与类别对应的检测框和分割区域，将分割区域采样为稀疏的点后和检测框作为提示信息引导SAM模型对遥感图像进行分割，最后将分割结果输入CLIP模型来评估分割区域和语义信息的相似度，通过多个视觉大模型的级联完成对遥感图像的语义分割。

2.2.2 基于知识蒸馏的大模型应用

现有基于通用视觉大模型的知识蒸馏一般采用基于响应的知识蒸馏，即训练学生模型直接模仿教师模型的最终预测结果。Chen等利用SAM模型对大规模未标注的红外图像进行预测，将置信度较高的预测掩码作为图像的标注结果[60]，该过程并未区分目标类别因此数据标注较为粗糙，利用伪标注数据来训练SegFormer[61]语义分割模型，最后使用人工标注数据对模型进行微调。Wang等利用公开遥感图像目标检测数据集中的目标框信息来提示SAM模型对遥感图像进行分割，进而构建了一个具有超过10万张图像和160万个标注实例的大规模语义分割数据集[62]。He等利用涂鸦提示SAM进行伪装目标分割，通过对每张图像进行旋转获得多个预测掩码进行集成来获得更高质量的分割结果，以此建立数据集来训练伪装目标分割模型[63]。

2.2.3 大模型专业化

将SAM为代表的基于提示的交互式分割模型直接应用在医学图像分割和玻璃检测等任务上表现并不理想，需要基于行业数据集来微调通用大模型或建立专业大模型。通用大模型的微调一般采取冻结原有模型的图像编码器等结构中的参数，对处理提示信息的编码器进行训练或者加入额外的网络层来获取更精准的信息。Hu等在皮肤癌图像分割数据集中的标注掩码上加入随机干扰并转为目标框，将其作为提示信息对SAM模型进行微调，超越了原始SAM的分割精度[64]。Julka等在SAM模型的图

像编码器后加入一个专用解码器，通过标注后的火星表面图像数据集对 SAM 模型进行微调，在训练过程中冻结图像编码器，仅对解码器进行训练以加快收敛，微调后的 SAM 模型在分割精度上超越了部分基于卷积神经网络的实例分割模型[65]。相比于微调通用大模型，建立专业大模型需在大量行业数据集上完整训练模型，对算力和数据规模要求较高。Wu 等参考提示分割模型提出了 PromptUNet 医学图像交互分割模型[66]，PromptUNet 不仅支持点、水平框等提示，也可以接受对正负样本的涂鸦来示意出需要分割的目标类别，该模型在超过十个医学图像数据集上进行训练并使用 U-Net 形式的解码器使得分割结果可以包含类别信息。

3　大模型在电力视觉中的能力边界与应用范式

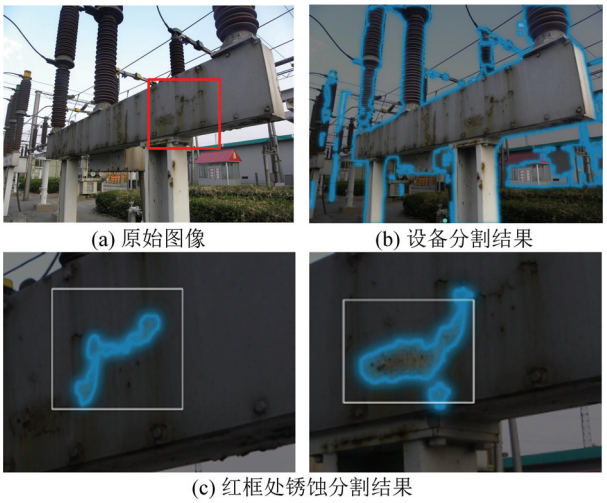
尽管视觉大模型在公共领域表现出极强的零样本泛化能力，但现有视觉大模型在电力缺陷检测、医学影像分割、伪装目标检测等专业知识门槛要求较高的场景中表现并不理想。直接将视觉大模型应用于电力视觉将面临以下问题：首先，大模型对高性能图形处理器(graphics processing unit，GPU)的依赖使得现有研究局限于直接使用或微调大模型；其次，完全使用大模型替代小模型在时延和成本方面并不合理，大小模型之间如何协作是电力视觉技术真正落地的技术前提之一；视觉大模型在电力行业的应用场景仍停留在小模型时代，拓展大模型的应用场景才能最大程度发挥其作用。目前直接开发电力行业大模型也面临较大挑战，相比于可在消费级 GPU 上完成全流程开发的传统深度学习模型，训练具有行业特性的视觉大模型对算力、能源的消耗较高，不仅如此，大多数电力视觉研究团队所拥有的数据集规模较小难以保证大模型对训练数据的需求。本文将从探索视觉大模型在电力视觉领域的能力边界和未来一段时间内大模型的应用范式两个角度分析电力视觉技术的发展方向。

3.1　大模型在电力视觉领域的能力边界

大模型的能力边界指针对大量特定问题训练大模型并不现实的情况下，将其直接应用于电力场景可能会出现预测结果不理想且难以对模型进行调整的情况。视觉模型的能力边界在以往面向特定电力视觉任务、基于传统深度学习模型的研究过程中已有所体现，但传统深度学习模型可塑性与灵活性较强，对于预训练的依赖性较低，使得研究者对应特定问题选取基线模型后，能够对模型进行专业化结构改进并通过重新训练来提升模型的性能，这一做法与强调“预训练”特性的大模型的应用过程存在较大分歧。相比于直接训练电力视觉大模型，在通用大模型的基础上进行能力拓展将更快且更易实现，因此全面、精准地定位视觉大模型的能力边界将是电力视觉技术重要的研究切入点。

如图 2 所示，在电力视觉场景下按照由低至高的专业性来应用视觉大模型时，本文认为通用视觉大模型将面临应用、数据和知识三重能力边界。从工程角度看，应用边界是利用大模型配合小模型处理具体任务时遇到的大模型能力边界；数据边界是利用大模型取代小模型来处理电力图像时遇到的能力边界；知识边界是利用大模型解决小模型能力范围之外的问题时遇到的能力边界。下面将结合实例进行具体阐述。

应用边界是指以不要求视觉大模型具备电力图像处理的专业性为前提，以通过视觉大模型直接增强现有电力视觉模型在多种任务上的性能为目的，偏重于直接利用视觉大模型在“预训练”与“通用性”方面的优势来辅助现有模型解决电力视觉任务，在此过程中视觉大模型表现出的应用能力上限。以变电站典型视觉缺陷分割为例，如图 2 所示，在变电设备图像上直接应用 SAM 模型时，可对变电设备取得较为理想的分割结果，但即使在矩形框提示的情况下，模型对于锈蚀缺陷的分割结果并不准确。



(a) 原始图像　(b) 设备分割结果

(c) 红框处锈蚀分割结果

图 2　视觉大模型的应用能力边界示例

Fig.2　Exemplars of visual large model practical limits in applications

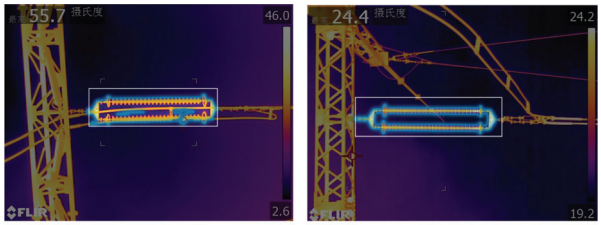
当传统深度学习模型因电力场景困难样本或稀缺样本出现能力边界时，补充训练数据、调整训练策略或改变模型结构等方法均可突破模型的应用能力边界。由于算力的限制，对现有视觉大模型在特定需求上进行训练并不现实，因此上述方法难以用于突破视觉大模型的应用边界。

数据边界指在通过微调或训练等方式将通用视觉大模型的图像分析能力对齐到电力视觉任务时，模型专业性因缺乏足够的优质数据这一限制进而难以展现，模型受电力数据规模的限制而展现出的性能上限。相较于小模型，大模型需要海量的数据才能发挥其超大规模参数的优势，大模型在电力场景表现不佳的原因往往是模型缺乏优质的训练数据而非模型性能不足。如图 3 所示，由于数据量较少，SAM 模型对绝缘子串组的红外图像分割效果并不稳定。

由于电力数据的保密需要，目前仍未出现适合训练视觉大模型的大规模电力数据集，因此突破数据边界的关键前提是完成原始电力数据积累。完成初步的数据积累后，面向大规模无标注数据探索不过度依赖人工的数据清洗与标注方法，推动数据预处理技术由劳动密集型向技术密集型发展，为大模型提供大规模优质训练数据，将是通用大模型专业化和建立电力视觉大模型过程中突破数据边界的关键技术。

知识边界指通用视觉大模型在处理电力图像时因不具备电力领域知识，缺乏电力业务逻辑，使得模型对图像中隐患或缺陷的识别能力受限，其预测结果易与专业知识有明显冲突。大模型的知识边界在安监、外破检测等电力图像理解任务中往往可以体现。以图 4 为例，在输电线路外破检测和变电站周边环境威胁分析等问题上，现阶段大小模型均难以自主完成潜在威胁分析任务，模型需要将电力业务逻辑与测距信息、点云数据等相结合才能完成较为抽象的电力图像分析任务。

电力视觉领域对于模型预测结果的可靠性要求极高，缺乏领域知识的大模型难免存在误检和漏检等情况，如果使用大模型取代小模型之后仍然需要人工频繁对大量模型输出结果进行复核校验，电力视觉技术的可信度将难以进一步提高，因此知识边界将是大模型较难突破但必须突破的难点，具有电力领域知识的视觉大模型有望进一步接近可信电力人工智能。



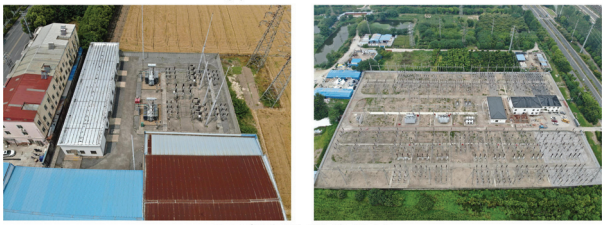
(a) 低质量分割结果 (b) 高质量分割结果

图 3 数据边界导致模型对同类设备的预测不稳定

Fig.3 Data boundaries cause unstable predictions for similar devices by the model



(a) 输电走廊场景



(b) 变电站周边场景

图 4 模型需依靠电力业务逻辑进行图像分析的场景

Fig.4 Scenarios where the model relies on the business logic of electricity for image analysis

3.2 大模型在电力视觉领域的应用范式

如图 5 所示，本文将电力视觉大模型的发展过程划分为大小模型协作、大模型专业化和电力视觉大模型 3 个阶段。大小模型协作指将大模型强大的特征提取和零样本泛化能力和小模型的专业信息处理能力灵活结合，由于电力视觉领域已积累大量专用视觉小模型，大小模型的高效协作能够快速拓展大模型的能力边界并加速大模型的实际落地，可以在短期内解决电力视觉大模型“从无到有”的问题。大模型专业化是指基于少量含有标注信息的电力数据对通用视觉大模型进行微调或结构改进使其适用于特定任务，类似做法已经被采用在医学影像分析领域[63-64]。由于电力数据大多因保密要求未公开，大多电力视觉研究团队目前的数据集储备规模较

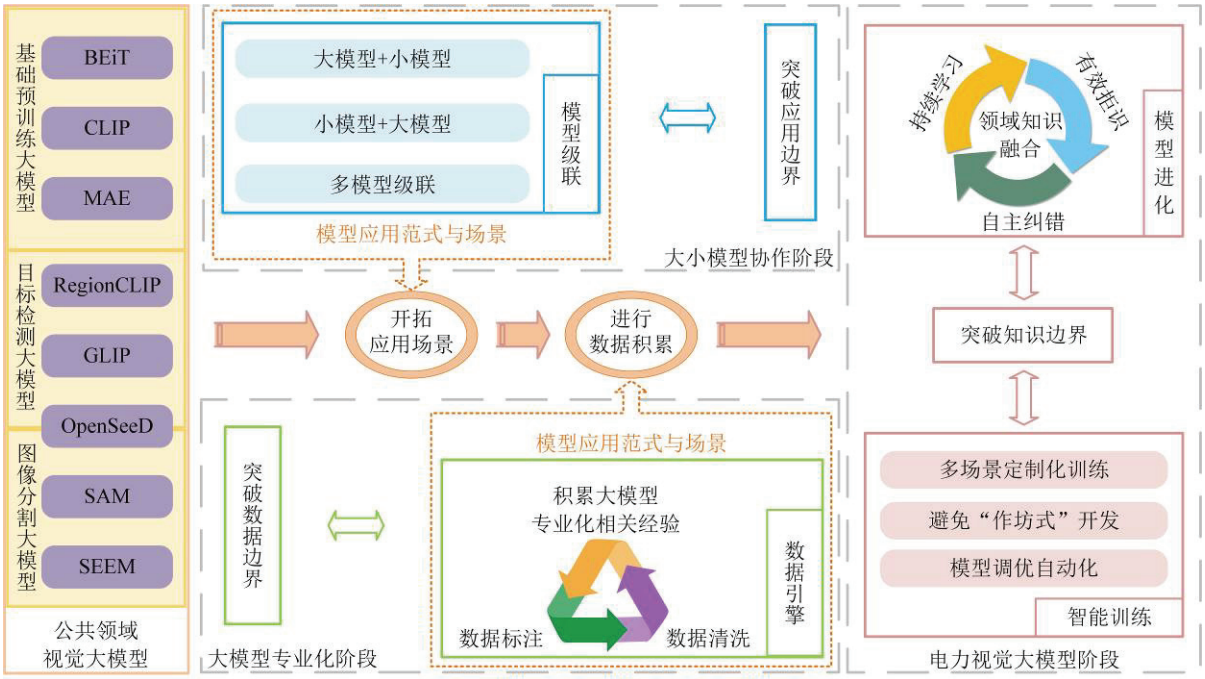


图 5　视觉大模型在电力视觉领域的应用范式

Fig.5　Application paradigm of visual large models in the field of power vision

小，大模型专业化过程将为建立电力视觉大模型逐步积累大规模训练数据与领域知识。目前产业界已有的电力大模型包括深圳供电局提出的祝融 2.0 预训练大模型和南方电网公司发布的电力大模型，学术界面向电力视觉大模型的研究也已起步。

本节将从模型级联、数据引擎、智能训练和模型进化 4 种可能的应用范式来分析电力视觉大模型在不同发展阶段的技术框架。

3.2.1　模型级联

模型级联指将大模型与小模型直接级联实现对大模型的能力拓展[53-55]，此范式在公共领域已有应用。模型级联具体包括“大模型+小模型”、“小模型+大模型”和多模型级联等实现方式。“大模型+小模型”指先由大模型完成对图像的初步处理以获得全局预测结果，而后利用小模型对预测结果进行筛选，例如在目标计数任务中，大模型完成对图像中所有目标的提取，小模型从目标中筛选符合特征的目标进行计数。“小模型+大模型”指先由小模型完成对图像的初步处理以获得特定目标的类别或位置信息作为提示，大模型将小模型预测结果作为提示完成对指定区域的分割等操作。例如在少、难样本计数和伪装目标分割场景中，当大模型对某些样本直接分割的效果较差时，可以采用小模型对样本区域进行初步预测并生成提示来辅助大模型进行有针对性的处理。以电力视觉领域较为经典的电力设备图像分割问题为例，如图 6 所示，SAM 在不依靠任何提示的情况下对图像中的多种变电设备分割取得了较为精准的结果。通过将 SAM 模型的分割结果输入到一个分类器进行设备类别预测(大模型+小模型)的方式，亦或是利用一个设备检测器的预测结果对 SAM 进行区域提示完成设备图像分割(小模型+大模型)的实现方式，均能高效完成电力设备的图像分割任务。由此可见，在训练数据处理较为繁琐的电力设备分割任务中，SAM 等大模型能够将分割数据的标注难度从像素级注释降维到目标或语义层面，在降低数据标注难度的同时提升了设备分割的精度。

多模型级联一般用于需要依靠电力领域知识来完成的视觉任务，例如在基于电力视觉的设备运行状态诊断过程中，可能需要使用两种大模型同时对文本和图像进行初步分析，并使用具有领域先验知识的小模型对处理结果进行后处理，最终输出具有专业性的处理结果。可见模型级联主要依靠的是大模型的零样本泛化能力，通过将针对性训练的小模型与预训练通用大模型进行组合来解决实际问题。

3.2.2　数据引擎

数据引擎的思想在 SAM 模型训练过程中被用于配合人工标注完成大规模分割数据集的标注任务。由于电力图像获取难度较大且标注成本较高，



(a) 原始图像 (b) SAM分割结果

图 6 视觉大模型分割结果

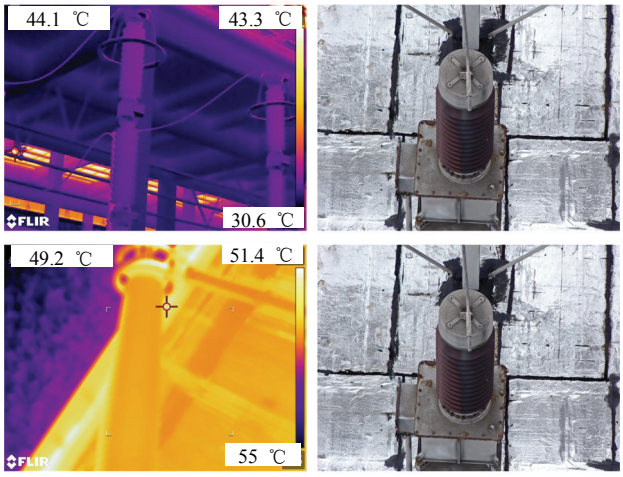
Fig.6 Segmentation results of visual large model

建立完善而强大的标注数据收集体系将是建立电力视觉大模型的前提。

数据引擎通过数据标注和数据清洗两部分来辅助专业人员完成数据积累。为提高数据初步标注的效率与准确性，可以采取“大帮小”的模型协作策略，即借助大模型的泛化能力对原始数据进行半自动或自动标注来提升数据标注速度，同时借助专业小模型对标注结果进行质量评估以保证标注信息可靠性，最后由专业人员对初步标注数据进行核验以保证标注质量。微调或训练电力视觉大模型需要海量的优质数据，而来自无人机和巡检机器人的图像数据良莠不齐且数量较大，依靠人工进行数据预处理效率较低，可以采用“小帮大”的模型协作策略，针对不同场景建立专用的小模型将高效、持续地为大模型进行数据清洗，保证大模型的训练效果。电力视觉任务中数据清洗主要包括低质量图像的剔除和去重两个方面，如图 7 所示，无人机和巡检机器人在巡检过程中难免产生模糊图像或重复图像，图像质量评估与图像去重技术或将称为数据清洗的主要技术手段。

3.2.3 智能训练与模型进化

此前为电力巡检现场开发专用视觉模型时需要挑选有限类别的标注数据对小模型进行训练和调优，当现场需求发生改变时往往需要重新设计模型，不难看出，作坊式的开发模式阻碍了具有实时监测优势的小模型在边缘端的部署进度。基于电力视觉大模型的智能训练范式将赋予一线巡检人员直接开发专用小模型的能力，可以降低一线人员开发针对性小模型的知识门槛并显著缩短提出需求与交付模型之间的技术空缺时间。智能训练这一应用范式允许一线巡检人员只将未标注的图像，以及具有提示性的语义信息输入电力视觉大模型或大小模型集成框架，通过大模型与开集检测模型等类型的专业小模型协作完成数据集标注，并自动完成对轻量化小



(a) 模糊图像 (b) 重复图像

图 7 低质量图像数据

Fig.7 Low-quality image data

模型的训练。将小模型部署后通过定期将小模型预测结果反馈给大模型实现对小模型的性能评估，同时挖掘小模型的困难样本对其进行加强训练，实现模型调优。

大模型面对复杂的外部环境将难免发生误检和漏检，如果仍然需要延续小模型时期的人工复核方式对错误结果进行弥补，大模型带来的效率和可靠性的提升将难以体现，因此有必要探索推动大模型进行持续学习的模型进化范式，最大程度缓解模型失效的风险。模型进化包括基于电力领域知识建立拒识机制和自主纠错机制两个方面。现有电力视觉模型擅长处理单一场景下的图像，普遍不具备拒识机制，而电力视觉大模型需要同时处理发电、输电、变电、配电和安监等多种任务场景下的巡检图像，模型如果对不相关或完全没有把握处理的输入图像强行进行预测，将会产生迷惑性较强且人工难以发现的错误处理结果，增加后续核查工作负担，因此研究如何在大模型框架中通过知识编码等方式融合领域知识并建立有效的拒识机制，让大模型将存疑的输入数据直接交由人工处理，可以有效降低模型错误率并推动模型的自我完善。自主纠错机制体现在大模型自我纠错和大模型对下游模型纠错两个方面。自我纠错过程依靠人工或专业性极强的小模型对大模型检测结果进行检查，目的是发现置信度较高但明显错误的结果，在此基础上对大模型进行针对性的微调或预训练以减轻模型出现幻觉的概率。基于多次自我纠错迭代后的大模型可以全面、自主地对部署在边缘设备上的小模型进行纠错并积

累高频出现的错检和漏检样本，并定期对其进行针对性训练。

4　大模型在电力视觉领域的局限性

视觉大模型赋予了电力视觉技术更高的预测精度和和更强的图像理解能力，但也有一些局限性。从能耗比角度看，由于模型复杂度高，计算和存储开销大，导致能耗较高。除此之外，大模型如何应对攻击也是一个关键问题，由于大模型基于大量的数据来训练，其预测的准确性在很大程度上取决于数据的完整性，相较于可以灵活调整训练样本的小模型，为确保大模型面对恶意攻击或噪声数据时能够做出准确的预测应对攻击方面，可能需要额外的防御机制来提高鲁棒性。在开发难度方面，开发面向电力场景的视觉模型需要大规模标注数据和更长的训练时间，其初期开发周期偏长。大模型的部署同样需要高性能硬件设备和大规模存储设备，部署和维护成本或将高于以往的轻量化视觉模型。因此，在电力视觉领域应用视觉大模型时，需要综合考虑这些因素，以充分发挥其优势并应对挑战。

5　大模型带给电力视觉研究者的思考

不可否认，SAM 等通用视觉大模型冲击了电力视觉技术部分研究方向，某些问题似乎已被大模型完美解决。大模型浪潮带来的影响不仅表现在技术进步上，还对电力视觉领域内的研究者提出了新的挑战和思考。部分研究者需要从“继续针对特定任务开发精致的小模型”与“多场景合并以开发通用的大模型”中做出选择，而对于那些在电力领域研究中资源有限的研究者，必须重新思考如何在这样的背景下进行有效的研究。电力视觉研究者需要认识到，数据和算力并不是决定性的因素，如何利用有限资源进行创新才是关键，例如对数据的深度理解和有效利用，对模型性能的精细调整和优化，以及对模型应用的综合性和适应性的追求，这些问题仍需研究者进行深入探索。总体来说，目前的通用视觉大模型在电力视觉领域仍处于“可以做很多事，但很多事做不好”的状态，如何理解和利用通用视觉大模型的优点和局限，如何将其应用到电力行业的特殊环境中，是研究者需要深入思考的问题。

6　结论与展望

1）本文回顾了电力视觉领域最新的研究进展与视觉大模型的发展现状，分析了通用视觉大模型在电力视觉技术中面临的应用、数据和知识三重能力边界，提出模型级联、数据引擎、智能训练和模型进化 4 种应用范式探索如何突破大模型的能力边界，基于大模型对现有研究方向的影响分析了电力视觉领域研究者需要思考的问题。

2）通用视觉大模型的崛起，既给电力视觉领域带来了挑战，也带来了机遇。对于研究者来说，这既是一个考验，也是一个成长的机会。未来，电力视觉大模型将会与物联网、云边协同等先进技术进行融合和应用。通过高度集成和智能化控制实现电力设备的全面监测和维护，为电力行业提供更加全面和精准的数据支持，提高电网的智能化水平。

参考文献　References

[1] 韩　笑，郭剑波，蒲天骄，等. 电力人工智能技术理论基础与发展展望(一)：假设分析与应用范式[J]. 中国电机工程学报，2023，43(8)：2877-2890.
HAN Xiao, GUO Jianbo, PU Tianjiao, et al. Theoretical foundation and directions of electric power artificial intelligence (I): hypothesis analysis and application paradigm[J]. Proceedings of the CSEE, 2023, 43(8): 2877-2890.

[2] 盛戈皞，钱　勇，罗林根，等. 面向新型电力系统的电力设备运行维护关键技术及其应用展望[J]. 高电压技术，2021，47(9)：3072-3084.
SHENG Gehao, QIAN Yong, LUO Lingen, et al. Key technologies and application prospects for operation and maintenance of power equipment in new type power system[J]. High Voltage Engineering, 2021, 47(9): 3072-3084.

[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)：1-8.
ZHAO Zhenbing, ZHANG Wei, ZHAI Yongjie, et al. Concept, research status and prospect of electric power vision technology[J]. Electric Power Science and Engineering, 2020, 36(1): 1-8.

[4] 齐冬莲，韩译锋，周自强，等. 基于视频图像的输变电设备外部缺陷检测技术及其应用现状[J]. 电子与信息学报，2022，44(11)：3709-3720.
QI Donglian, HAN Yifeng, ZHOU Ziqiang, et al. Review of defect detection technology of power equipment based on video images[J]. Journal of Electronics & Information Technology, 2022, 44(11): 3709-3720.

[5] ALEC R, JONG W K, CHRIS HRADFORD A, et al. Learning transferable visual models from natural language supervision[EB/OL]. (2021-02-26)[2023-11-30]. https://arxiv.org/abs/2103.00020.

[6] KIRILLOV A, MINTUN E, RAVI N, et al. Segment anything[EB/OL]. (2023-04-05)[2023-11-30]. https://arxiv.org/abs/2304.02643.

[7] ZOU X Y, YANG J W, ZHANG H, et al. Segment everything everywhere all at once[EB/OL]. (2023-04-13)[2023-11-30]. https: //arxiv.org/abs/2304.06718.

[8] ZHANG C H, LIU L, CUI Y W, et al. A comprehensive survey on segment anything model for vision and beyond[EB/OL]. (2023-05-19)[2023-11-30]. https: //arxiv. org/abs/2305.08196.

[9] 赵振兵，蒋志钢，李延旭，等. 输电线路部件视觉缺陷检测综述[J]. 中国图象图形学报，2021，26(11)：2545-2560.
ZHAO Zhenbing, JIANG Zhigang, LI Yanxu, et al. Overview of visual defect detection of transmission line components[J]. Journal of Image and Graphics, 2021, 26(11): 2545-2560.

[10] SADYKOVA D, PERNEBAYEVA D, BAGHERI M, et al. IN-YOLO: real-time detection of outdoor high voltage insulators using UAV imaging[J]. IEEE Transactions on Power Delivery, 2020, 35(3): 1599-1601.

[11] ZHAI Y J, WANG Q M, YANG X, et al. Multi-fitting detection on transmission line based on cascade reasoning graph network[J]. IEEE Transactions on Power Delivery, 2022, 37(6): 4858-4868.

[12] LIU X Y, MIAO X R, JIANG H, et al. Component detection for power line inspection using a graph-based relation guiding network[J]. IEEE Transactions on Industrial Informatics, 2023, 19(9): 9280-9290.

[13] 赵振兵，熊　静，徐厚东，等. 融合结构推理的深度模型输电线路金具及其缺陷检测[J]. 高电压技术，2023，49(8)：3346-3353.
ZHAO Zhenbing, XIONG Jing, XU Houdong, et al. Integrating structural reasoning for deep model transmission line fittings and their defects detection[J]. High Voltage Engineering, 2023, 49(8): 3346-3353.

[14] ZHAO Z B, WANG R, LI Y X, et al. A new multilabel recognition framework for transmission lines bolt defects based on the combination of semantic knowledge and structural knowledge[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-11.

[15] 赵振兵，王　睿，赵文清，等. 基于图知识推理的输电线路缺销螺栓识别方法[J]. 智能系统学报，2023，18(2)：372-380.
ZHAO Zhenbing, WANG Rui, ZHAO Wenqing, et al. Pin-missing bolts recognition method for transmission lines based on graph knowledge reasoning[J]. CAAI Transactions on Intelligent Systems, 2023, 18(2): 372-380.

[16] ZHENG H B, LIU Y, SUN Y H, et al. Arbitrary-oriented detection of insulators in thermal imagery via rotation region network[J]. IEEE Transactions on Industrial Informatics, 2022, 18(8): 5242-5252.

[17] 赵振兵，马迪雅，石　颖，等. 基于改进 YOLOX 的变电站仪表外观缺陷检测算法[J]. 图学学报，2023，44(5)：937-946.
ZHAO Zhenbing, MA Diya, SHI Ying, et al. Appearance defect detection algorithm of substation instrument based on improved YOLOX[J]. Journal of Graphics, 2023, 44(5): 937-946.

[18] GE Z, LIU S T, WANG F, et al. YOLOX: exceeding YOLO series in 2021[EB/OL]. (2021-07-18)[2023-11-30]. https://arxiv.org/abs/2107.08430.

[19] 何　敏，秦　亮，刘开培，等. 变电站电力仪表智能检测算法研究[J/OL]. 高电压技术，2023：1-15[2023-11-04]. https://doi.org/10.13336/j. 1003-6520. hve. 20230591.
HE Min, QIN Liang, LIU Kaipei, et al. Research on intelligent detection algorithms for substation power instruments[J/OL]. High Voltage Engineering, 2023: 1-15[2023-11-04]. https://doi.org/10.13336/j. 1003-6520. hve. 20230591.

[20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017: 6000-6010.

[21] 滕国龙，郑　易，闫云凤，等. 非对称交叉平均教师：一种半监督语义分割算法及其在变电站表计读数识别中的应用[J]. 中国电机工程学报，2023，43(8)：2979-2989.
TENG Guolong, ZHENG Yi, YAN Yunfeng, et al. Asymmetric cross mean teacher: a semi-supervised semantic segmentation algorithm and its application on substation meter reading recognition[J]. Proceedings of the CSEE, 2023, 43(8): 2979-2989.

[22] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C] // Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2015: 91-99.

[23] ZHAO Z B, LIU B, ZHAI Y J, et al. Dual graph reasoning network for oil leakage segmentation in substation equipment[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 3502415.

[24] 张迎晨，王　波，马富齐，等. 基于域适应网络的电力设备不规则外表面缺陷高精度检测方法[J]. 高电压技术，2022，48(11)：4516-4526.
ZHANG Yingchen, WANG Bo, MA Fuqi, et al. High-precision detection method of irregular outer surface defects of power equipment based on domain adaptation network[J]. High Voltage Engineering, 2022, 48(11): 4516-4526.

[25] 赵文清，刘　亮，胡嘉伟，等. 基于深度可分离空洞卷积金字塔的变压器渗漏油检测[J]. 智能系统学报，2023，18(5)：966-974.
ZHAO Wenqing, LIU Liang, HU Jiawei, et al. Detection of transformer oil leakage based on deep separable atrous convolution pyramid[J]. CAAI Transactions on Intelligent Systems, 2023, 18(5): 966-974.

[26] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]// Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer, 2018: 833-851.

[27] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C] // Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017: 1800-1807.

[28] 赵振兵，冯　烁，赵文清，等. 融合知识迁移和改进 YOLOv6 的变电设备热像检测方法[J]. 智能系统学报，2023，18(6)：1213-1222.
ZHAO Zhenbing, FENG Shuo, ZHAO Wenqing, et al. Thermd image detection method for substation equipment by incorporat-ing knowledge migration and improved YOLOv6[J]. CAAI Transactions on Intelligent Systems, 2023, 18(6): 1213-1222.

[29] ZHAO Z B, FENG S, ZHAI Y J, et al. Infrared thermal image instance segmentation method for power substation equipment based on visual feature reasoning[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 5029613.

[30] ZHENG H B, CUI Y H, YANG W Q, et al. An infrared image detection method of substation equipment combining iresgroup structure and centernet[J]. IEEE Transactions on Power Delivery, 2022, 37(6): 4757-4765.

[31] DUAN K W, BAI S, XIE L X, et al. CenterNet: keypoint triplets for object detection[C]// Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE, 2019: 6568-6577.

[32] ZHENG H B, SUN Y H, LIU X H, et al. Infrared image detection of substation insulators using an improved fusion single shot multibox detector[J]. IEEE Transactions on Power Delivery, 2021, 36(6): 3351-3359.

[33] LI Z X, YANG L, ZHOU F. FSSD: feature fusion single shot multibox detector[EB/OL]. (2017-12-17)[2023-11-30]. https://arxiv.org/abs/1712.00960.

[34] WANG X L, ZHANG R F, KONG T, et al. SOLOv2: dynamic and Fast instance segmentation[EB/OL]. (2020-03-23)[2023-11-30]. https://arxiv. org/abs/2003.10152.

[35] WANG B, DONG M, REN M, et al. Automatic fault diagnosis of infrared insulator images based on image instance segmentation and temperature analysis[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(8): 5345-5355.

[36] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 2980-2988.

[37] 马富齐，王　波，董旭柱，等. 面向输电线路覆冰厚度辨识的多感受野视觉边缘智能识别方法研究[J]. 电网技术，2021，45(6)：2161-2169.
MA Fuqi, WANG Bo, DONG Xuzhu, et al. Receptive field vision edge intelligent recognition for ice thickness identification of transmission line[J]. Power System Technology, 2021, 45(6): 2161-2169.

[38] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C]// 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea: IEEE, 2017: 1314-1324.

[39] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox

detector[C]// Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016: 21-37.

[40] 杨　知，江子航，张思航，等. 基于改进 UNet 的输电通道沿线易漂浮物遥感识别[J]. 高电压技术，2023，49(8)：3395-3404.
YANG Zhi, JIANG Zihang, ZHANG Sihang, et al. Identification of floating objects along transmission corridors with remote sensing images based on improved UNet[J]. High Voltage Engineering, 2023, 49(8): 3395-3404.

[41] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer, 2015: 234-241.

[42] 郑含博，胡思佳，梁炎燊，等. 基于 YOLO-2MCS 的输电线路走廊隐患目标检测方法[J/OL]. 电工技术学报，2023：1-12 [2023-11-05]. https://doi.org/10.19595/j.cnki.1000-6753.tces. 230666.
ZHENG Hanbo, HU Sijia, LIANG Yanshen, et al. A hidden danger object detection method for transmission line corridor based on YOLO-2MCS[J/OL]. Transactions of China Electrotechnical Society, 2023: 1-12[2023-11-05]. https: //doi. org/10.19595/j. cnki. 1000-6753. tces. 230666.

[43] LI C Y, LI L L, JIANG H L, et al. YOLOv6: a single-stage object detection framework for industrial applications[EB/OL]. (2022-09-07)[2023-11-30]. https://arxiv.org/abs/2209.02976.

[44] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer, 2018: 3-19.

[45] LI J H, SHUANG F, HUANG J J, et al. Safe distance monitoring of live equipment based upon instance segmentation and pseudo-LiDAR[J]. IEEE Transactions on Power Delivery, 2023, 38(4): 2953-2964.

[46] GAO S H, CHENG M M, ZHAO K, et al. Res2Net: a new multi-scale backbone architecture[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(2): 652-662.

[47] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection[C]// Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 10778-10787.

[48] BAO H B, DONG L, PIAO S H, et al. BEiT: BERT pre-training of image transformers[EB/OL]. (2021-06-15)[2023-11-30]. https: //arxiv. org/abs/2106.08254.

[49] HE K M, CHEN X L, XIE S N, et al. Masked autoencoders are scalable vision learners[C] // Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022: 15979-15988.

[50] LI L H, ZHANG P C, ZHANG H T, et al. Grounded language-image pre-training[C]// Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022: 10955-10965.

[51] ZHONG Y W, YANG J W, ZHANG P C, et al. RegionCLIP: region-based language-image pretraining[C]// Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022: 16772-16782.

[52] ZHANG H, LI F, ZOU X Y, et al. A simple framework for open-vocabulary segmentation and detection[C]// Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE, 2023: 1020-1031.

[53] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[EB/OL]. (2020-10-26)[2023-11-30]. https://arxiv.org/abs/2010.11929.

[54] GIANNAKIS I, BHARDWAJ A, SAM L, et al. Deep learning universal crater detection using segment anything model (SAM)[EB/OL]. (2023-04-16)[2023-11-30]. https: //arxiv. org/abs/2304.07764.

[55] WILLIAMS D, MACFARLANE F, BRITTEN A. Leaf only SAM: a segment anything pipeline for zero-shot automated leaf segmentation[EB/OL]. (2023-05-16)[2023-11-30]. https://arxiv.org/abs/2305.09418.

[56] WANG B, ABOAH A, ZHANG Z Y, et al. GazeSAM: What you see is what you segment[EB/OL]. (2023-04-26)[2023-11-30]. https://arxiv.org/abs/2304.13844.

[57] ZHANG J L, ZHOU Z L, MAI G C, et al. Text2Seg: remote sensing image semantic segmentation via text-guided visual foundation models[EB/OL]. (2023-04-20)[2023-11-30]. https://arxiv.org/abs/2304.10597.

[58] LIU S L, ZENG Z Y, REN T H, et al. Grounding DINO: marrying DINO with grounded pre-training for open-set object detection[EB/OL]. (2023-03-09)[2023-11-30]. https://arxiv.org/abs/2303.05499.

[59] LI Y, WANG H L, DUAN Y Q, et al. CLIP Surgery for better explainability with enhancement in open-vocabulary tasks[EB/OL]. (2023-04-12)[2023-11-30]. https://arxiv. org/abs/2304.05653.

[60] CHEN J Z, BAI X Z. Learning to “segment anything” in thermal infrared images through knowledge distillation with a large scale dataset SATIR[EB/OL]. (2023-04-17)[2023-11-30]. https://arxiv.org/abs/2304.07969.

[61] XIE E Z, WANG W H, YU Z D, et al. SegFormer: simple and efficient design for semantic segmentation with transformers[EB/OL]. (2021-05-31)[2023-11-30]. https://arxiv.org/abs/2105.15203.

[62] WANG D, ZHANG J, DU B, et al. SAMRS: Scaling-up remote sensing segmentation dataset with segment anything model[EB/OL]. (2023-05-03)[2023-11-30]. https://arxiv. org/abs/2305.02034.

[63] HE C M, LI K, ZHANG Y C, et al. Weakly-supervised concealed object segmentation with SAM-based pseudo labeling and multi-scale feature grouping[EB/OL]. (2021-05-18)[2023-11-30]. https://arxiv.org/abs/2305.11003.

[64] HU M Z, LI Y H, YANG X F. SkinSAM: empowering skin cancer segmentation with segment anything model[EB/OL]. (2023-04-27)[2023-11-30]. https://arxiv.org/abs/2304.13973.

[65] JULKA S, GRANITZER M. Knowledge distillation with segment anything (SAM) model for planetary geological mapping[C] // Proceedings of the 9th International Conference on Machine Learning, Optimization, and Data Science. Grasmere, UK: Springer, 2023: 68-77.

[66] WU J D, ZHU J Y, LIU Y P, et al. One-prompt to segment all medical images[EB/OL]. (2023-05-17)[2023-11-30]. https://arxiv.org/abs/2305.10300.



赵振兵(通信作者)
1979—，男，博士，教授，博导
主要研究方向为电力视觉。主持国家自然科学基金项目等纵向课题 10 项；获省科技进步一等奖 2 项；以第 1 完成人获得国家专利授权 16 项；以第 1 作者出版专著 2 部、发表学术论文 50 余篇
E-mail: zhaozhenbing@ncepu.edu.cn

ZHAO Zhenbing
Ph.D., Professor
Corresponding author

收稿日期　2023-12-25　修回日期　2024-02-18　　编辑　卫李静