专题 黄土高原生态环境

2024 年中国西北地区光伏板阵列空间数据集

魏亚宁[1]，姚晓军[1*]

1. 西北师范大学，地理与环境科学学院，兰州 730070

摘要：我国西北地区光伏板阵列的规模化开发对高精度空间信息提取和管理提出了迫切需求。本研究基于 2024 年 Sentinel-2 MSI 影像，通过融合光谱与空间特征，构建多尺度分割对象集，对比分析最近邻法、基于样本的面向对象法和 BP（Back Propagation）人工神经网络法提取光伏板阵列的精度。基于独立验证样本的混淆矩阵精度评价，优选 BP 人工神经网络法为最佳模型，形成了空间分辨率为 10 m 的西北地区光伏板阵列分布数据集，其总体分类精度（OA）为 91.6%，Kappa 系数为 0.89，显示出数据集的高精度和可靠性。结合“2020 年全国 10 m 分辨率地面光伏分类数据集”与历史 Wayback 影像，通过跨年度时序对比与典型区域空间分布分析，从数据时间更新程度和空间识别准确性两方面验证了本数据集的可靠性与优势。结果显示，2024 年西北地区光伏板阵列总面积为 1676.13 km^2，其中新疆（605.80 km^2）面积最大，其次是青海（484.27 km^2）、甘肃（272.98 km^2）、宁夏（174.65 km^2）和陕西（138.43 km^2），主要集中于塔克拉玛干沙漠边缘、天山北坡、青海海西与海南藏族自治州、甘肃河西走廊、宁夏银川与吴忠市以及陕西榆林市等地区。本数据集为“双碳”目标下西北地区能源规划、生态环境评估与可持续发展提供了可靠的数据支撑与科学依据。

关键词：光伏板阵列；遥感信息提取；多尺度分割；空间分布；中国西北地区

数据库（集）基本信息简介

数据库（集）名称	2024 年中国西北地区光伏板阵列空间数据集
数据通信作者	姚晓军（yaoxj_nwnu@163.com）
数据作者	魏亚宁、姚晓军
数据时间范围	2024年
地理区域	中国西北地区（32°N–50°N，73°E–112°E，包括甘肃省、陕西省、宁夏回族自治区、青海省和新疆维吾尔自治区）
空间分辨率	10 m
数据量	21.5 MB
数据格式	ESRI Shapefile
数据服务系统网址	https://doi.org/10.12262/lsdc.17040.2541 https://www.scidb.cn/doi/10.12262/lsdc.17040.2541
基金项目	西北师范大学绿洲科学科研成果突破行动计划（NWNU-LZKX-202301）
数据库（集）组成	本数据集共包括1个数据文件，为中国西北地区五省（自治区）2024年光伏板阵列空间数据。

www.sciengine.com/CSD
ISSN 2096-2223
CN 11-6035/N

文献 CSTR：
32001.14.11-6035.csd.2025.0157.zh

文献 DOI：
10.11922/11-6035.csd.2025.0157.zh

数据 DOI：
10.12262/lsdc.17040.2541

文献分类：地球科学

收稿日期：2025-08-29
录用日期：2026-03-12
发表日期：2026-03-18

* 论文通信作者
姚晓军：yaoxj_nwnu@163.com

引　言

在“碳达峰、碳中和”国家战略背景下，光伏发电作为我国重点发展的绿色可再生能源之一，正持续发展[1]。我国西北地区具有得天独厚的日照资源优势，目前光伏装机容量已占全国总量的32%[2]。随着该区域光伏板阵列面积的显著增加，对高精度空间信息提取和管理的需求日益迫切，亟需建立涵盖光伏板阵列的地理位置、占地面积等信息的空间数据库，以支撑光伏板阵列的科学规划、智能运维、环境影响评估及政策实施效果评估。

近年来，遥感技术在光伏板阵列识别与制图领域的应用研究不断增多，数据类型和相关方法持续丰富。在数据源方面，研究已从使用单一的 Landsat 或 Sentinel-2 多光谱影像，扩展至融合 Sentinel-1 SAR 数据、高分辨率影像以及地形和夜光等多源数据，显著提升了复杂地表环境下光伏设施与其他地物的区分能力[3]。在方法上，从传统的随机森林等机器学习算法逐步发展到基于 CNN、U-Net 等深度学习模型，实现了全球和国家尺度上的光伏设施自动化识别[4]。在公开数据集方面，已发布“中国 2020 年 10 m 分辨率地面光伏分布数据”[5]和“全球 20 m 年度光伏面板数据集（2019–2022）”[6]，为能源政策制定与相关研究提供了重要数据支撑。然而，现有研究仍存在一些不足：一是数据时间更新程度有限，多数公开数据集与研究成果仍停留在 2020–2022 年[5-6]，在一定程度上难以准确反映近年来光伏设施快速建设背景下的最新空间分布状态；二是在复杂背景条件下分类精度与稳健性不足。在西北地区广泛分布的荒漠、戈壁及盐碱地等复杂地表类型中，光伏板阵列虽然在多光谱影像中具有可区分的光谱和纹理特征，但传统识别方法仍可能出现误分与漏分，主要由于光伏板阵列面积较小、边界像元受分辨率影响以及复杂地表背景干扰[7]。此外，现有光伏板阵列数据集在影像质量、标注尺度与标注标准方面存在差异，造成模型在跨区域应用时泛化能力下降[8]。针对上述问题，仍需构建具有更高时效性、稳定性与区域适应性的光伏板阵列识别数据基础，为大范围光伏设施监测提供支持。

本数据集基于 2024 年我国西北地区 Sentinel-2 MSI 影像，通过融合光谱特征与空间特征构建多尺度分割对象集，采用最近邻法、基于样本的面向对象法及 BP 人工神经网络法三种典型分类方法进行光伏板阵列提取，并利用混淆矩阵交叉验证对比分析各类方法的识别精度，最终优选最佳分类模型，实现研究区全域光伏板阵列的精细化识别。在此基础上，结合“2020 年全国 10 m 分辨率地面光伏分类数据集”及 Wayback 影像开展时序对比验证分析，深入探究光伏板阵列的空间分布格局及其阶段化变化特征。本数据集可为我国西北地区光伏板阵列的空间监测提供时间更新明确、空间精度可验证的数据支持，并为区域新能源发展规划与管理决策提供科学依据。

1　数据采集和处理方法

1.1　研究区概况

本数据集的研究区域为我国西北地区（图 1），包括陕西省、甘肃省、宁夏回族自治区、青海省以及新疆维吾尔自治区，总面积 304×10^4 km^2，约占我国国土总面积的 32%。该区域深处欧亚大陆腹地，远离海洋，加之高原与山脉对湿润气流的阻挡作用，形成了以干旱、半干旱为主的气候特征，具有降水量少、昼夜温差大、日照时数长等特点[9]。其中，黄土高原区域广泛分布于陕西、甘肃和宁夏等省区，是西北地区生态环境问题最为集中、土地利用变化最为剧烈的典型区域之一。作

为我国重要的新能源战略基地，西北地区太阳能资源储量居全国首位，尤其是在青海、新疆、甘肃等省区，光伏发电潜力巨大，目前已建成青海塔拉滩光伏产业园、新疆哈密光伏基地等多个大型光伏集聚区[10]。光伏设施的快速建设在为区域能源结构转型提供支撑的同时，也对土地利用格局与生态环境产生了深远影响，使得该区域成为开展光伏板阵列空间分布识别及其生态环境效应研究的典型区域。

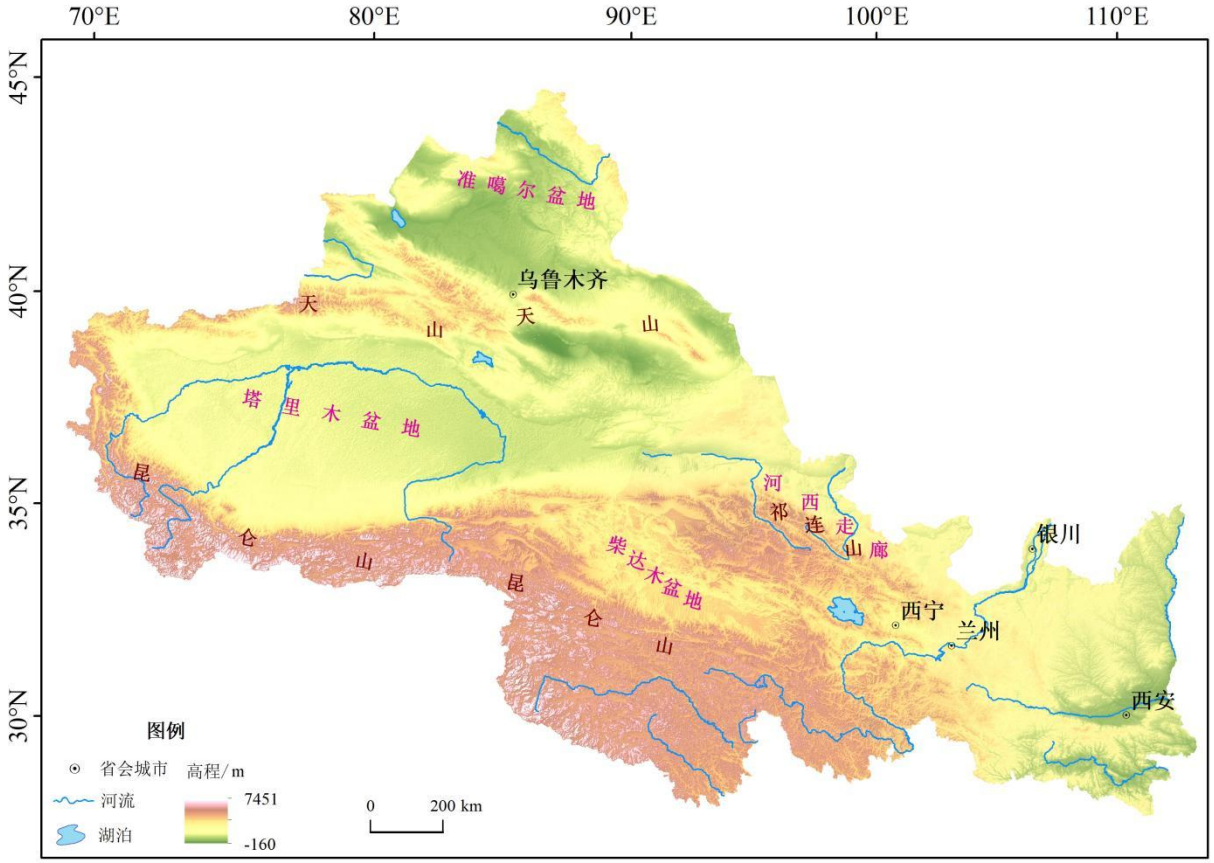

图 1　研究区域

Figure 1　The study area

1.2　数据来源

1.2.1　基础地理数据与卫星遥感影像

本数据集所使用的基础地理数据包括西北地区省级与市级行政区划数据，由国家基础地理信息中心提供。卫星遥感影像数据包括 Sentinel-2 MSI 影像和 World Imagery Wayback 影像，其中 Sentinel-2 MSI 影像从哥白尼数据中心网站下载获得，World Imagery Wayback 影像通过 ArcGIS Pro 在线地图服务调用。此外，还参考了 Feng 等[5]发布的“2020 年全国 10 m 分辨率地面光伏分类数据集”，该数据获取自科学数据银行网站。各类数据的详细时间及来源信息见表 1。

根据研究区范围及影像质量要求，共选取 2024 年 7–8 月期间的 Sentinel-2 MSI 影像 304 景，总数据量约 324 GB。按影像数量统计（图 2），新疆维吾尔自治区最多（67 景），其后依次是甘肃省（64 景）、宁夏回族自治区（61 景）、青海省（60 景）和陕西省最少（52 景）。

表 1　本研究所用的基础地理数据信息

Table 1　Information of the basic geographical data used in this study

数据名称	数据时间	数据来源
省级行政区划数据	2024 年	国家基础地理信息中心（https://cloudcenter.tianditu.gov.cn/administrativeDivision）
市级行政区划数据	2024 年	国家基础地理信息中心（https://cloudcenter.tianditu.gov.cn/administrativeDivision）
Sentinel-2 MSI 影像	2024 年	哥白尼数据中心（https://dataspace.copernicus.eu/）
World Imagery Wayback 影像	2024 年 11 月 18 日	ArcGIS Pro 在线地图（https://livingatlas.arcgis.com/wayback）
2020 年全国 10 m 分辨率地面光伏分类数据集	2020 年	科学数据银行（https://doi.org/10.57760/sciencedb.o00121.00001）

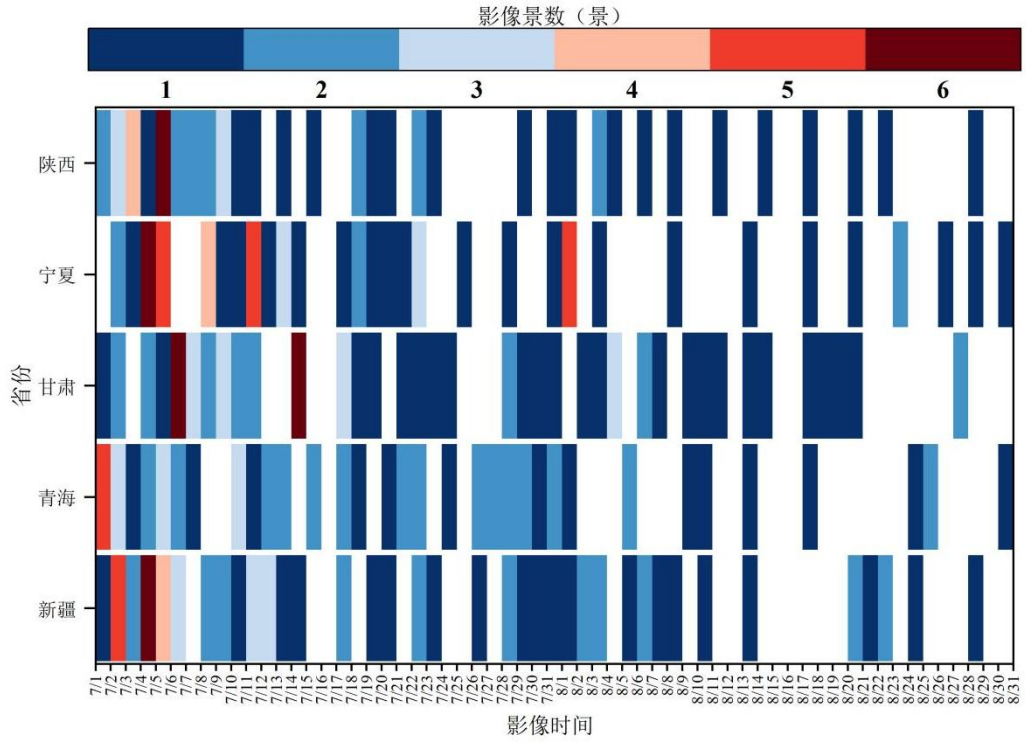

图 2　2024 年西北地区各省（自治区）Sentinel-2 MSI 影像数量统计

Figure 2　Statistics of Sentinel-2 MSI images of each province (autonomous region) in Northwestern China in 2024

1.2.2　样本数据

本研究共选取 5 类地物样本，包括光伏板阵列、植被、水体、建设用地和裸地。其中，光伏板阵列样本数量为 1000 个，非光伏样本共 700 个。样本的制作主要基于 2024 年获取的 Sentinel-2 MSI 影像，并结合高分辨率 Google Earth 影像与已有光伏板阵列分布数据进行辅助判读与验证。首先，通过人工目视解译在影像中标注真实存在的光伏板阵列区域，确保样本多样性，涵盖不同建设规模、不同背景环境（如沙漠边缘、戈壁滩、草地和耕地周边）以及不同光照和地形条件下的光伏板阵列。其次，非光伏样本按照地类典型性和分布代表性原则进行选取，确保涵盖研究区内主要的混淆地物

类型，以提高模型的判别能力。

光伏板阵列训练样本的空间分布如图 3 所示，样本在西北地区各省份均有布设，以增强样本集的区域代表性。例如，新疆和青海的光伏板阵列多分布于荒漠与戈壁地带，由于其通常建设在大面积裸地背景上，易在分割阶段出现对象边界不清或混合像素现象；甘肃和宁夏的光伏板阵列部分邻近农田和城镇建设用地，易与植被或建筑物混淆；陕西的光伏板阵列则多分布于黄土高原边缘，地形复杂，阴影效应明显。通过在不同地理环境与背景下选取光伏板阵列样本，能够有效提升分类模型的鲁棒性与普适性。

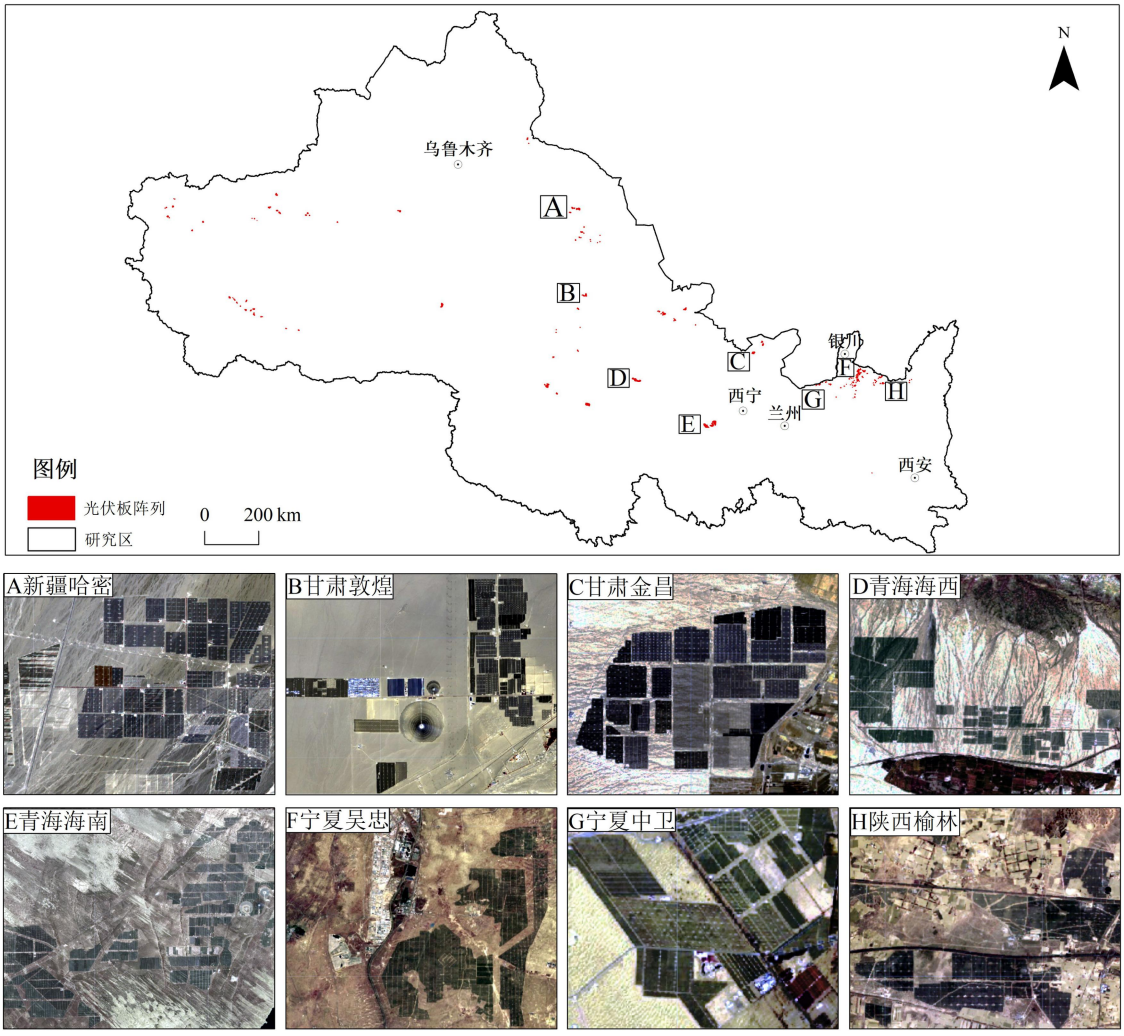

图 3　光伏板样本分布及典型光伏板阵列（子图 A-H 为 Sentinel-2 MSI 真彩色影像）

Figure 3　Distribution of photovoltaic samples and typical photovoltaic panel arrays (subfigure A-H: Sentinel-2 MSI true-color imagery)

1.3　数据处理过程

图 4 为本研究所采用的技术流程图。首先，在数据预处理阶段，对获取的 Sentinel-2 MSI 影像采用 ESA 官方提供的 Sen2Cor 工具进行辐射定标和大气校正，随后对处理后的影像进行波段合成、裁剪与镶嵌，形成覆盖研究区的 2024 年空间分辨率为 10 m 的多波段影像集[11-12]。其次，在数据提取阶段，通过不同分割参数的试验确定最佳分割参数，并结合部分光谱特征与空间特征指标构建多特征图像，分别采用最近邻法、基于样本的面向对象法及 BP 人工神经网络法进行光伏板阵列信息提

取，利用训练样本和验证样本评估不同方法的提取效果。最后，在精度评价阶段，基于总体精度、用户精度、生产者精度及 Kappa 系数对三种方法的精度进行对比分析，以确定最优方法并结合历史影像开展时序验证，进一步分析其在西北地区的空间分布特征。

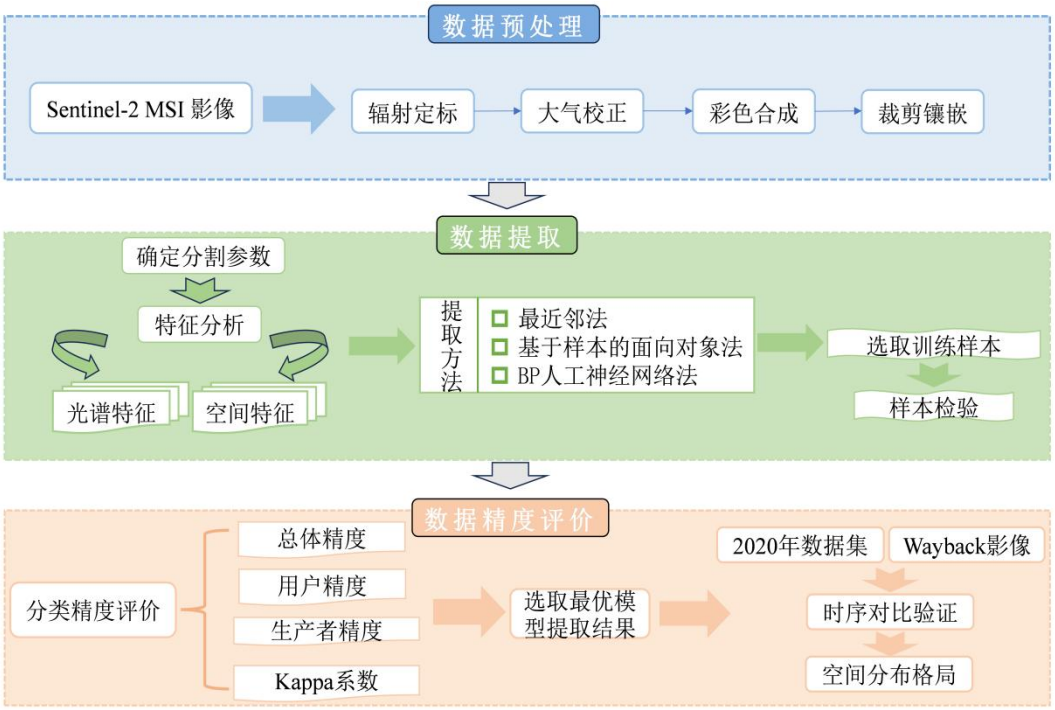

图 4　技术路线图

Figure 4　Technical workflow

1.3.1　影像多尺度分割

多尺度分割是通过迭代合并光谱、形状与纹理相似的像元，生成具有地物意义的对象单元[13]。首先从较小的比例开始分割，以获得更多碎片化的对象；然后逐渐增大分割比例，将具有相似属性的相邻对象组合成更大的分割对象。分割边缘越来越接近对象的实际边界，直到对象之间的异质性达到阈值[14]。已有研究表明，多尺度方法能够有效缓解单尺度分割在“过分割”与“欠分割”之间的权衡问题，并能更准确地刻画复杂地物边界[15]。本研究采用 eCognition 软件对多源遥感影像应用多尺度分割，经过反复实验对比，得到较优分割参数，即设定分割尺度 100，形状因子和紧致度因子分别设为 0.2 和 0.5。

1.3.2　光谱特征

光谱特征是一组描述影像中对象相似性的数值，通过计算每个波段的亮度值获得，能够反映不同地物材料在多波段遥感影像中的物理差异，是分类的重要基础特征[16]。常用的光谱特征包括平均值、亮度值和标准差等[17]。此外，本研究还引入了归一化植被指数（NDVI）与归一化差异水体指数（NDWI），计算公式为：

$$NDVI = \frac{NIR-R}{NIR+R} \quad (1)$$

$$NDWI = \frac{G-NIR}{G+NIR} \quad (2)$$

式中，NIR 为近红外波段的反射率，R 为红波段的反射率，G 为绿波段的反射率。

1.3.3 空间特征

在遥感影像分析中，由于光谱混合、传感器噪声和阴影等因素的影响，仅依赖光谱特征往往难以实现对地物的稳定区分。因此，在对象层面引入空间结构信息有助于提升分类精度[18]。在遥感图像解译与信息提取过程中，融入空间特征可显著增强对具有规则几何形态地物的辨识能力，例如区分城市道路网络、农田布局等具有明显空间结构的地表特征[19]。鉴于光伏板阵列在形状上通常呈现较高规则性，本研究选用面积与长宽比两类空间特征，以有效降低道路误识别对光伏板阵列提取结果的干扰。

1.3.4 基于样本的面向对象法

基于样本的面向对象法（Object-based Image Analysis, OBIA）是一种充分利用影像光谱信息、空间结构与对象间上下文关系的遥感分类技术。其核心流程一般包括两个阶段：首先，通过多尺度分割将遥感影像划分为具有语义意义的同质性对象，使影像单位从单一像素提升至结构化对象，从而有效减少噪声影响并增强地物边界表达。其次，在特征空间中，基于对象的光谱特征、几何形态以及邻域关系，利用“样本对象”构建类别特征模型，并将未分类的对象按照其与类别模型的相似度匹配到最合适的类别中[20]。该方法以像素聚合形成的图像对象为识别基础，避免了对单一像素的孤立评估，从而有效弥补传统像素级分类的不足，并基于对象的综合特征实现更具一致性的分类。

1.3.5 最近邻法

最近邻法（K-Nearest Neighbors, KNN）是一种非参数提取方法，在图像识别领域得到广泛应用[21]。其基本思想是依据训练集中已知类别样本在特征空间中的分布，通过距离度量来确定未知样本的最近邻，从而完成类别标签的判定[22]。在实际应用中，KNN 是通过计算待提取物体与特征空间中待提取类别之间的最短距离，选取最近的邻域信息以实现类别归属，其距离计算公式如下：

$$D = \sqrt{\sum_f \left[\frac{v_f^s - v_f^o}{\sigma_f}\right]^2} \tag{3}$$

式中，D表示欧几里得距离，v_f^s为样本对于特征f的特征值，v_f^o为待提取对象对于特征f的特征值，σ_f为特征值f的标准值。

1.3.6 BP 人工神经网络

BP（Back Propagation）人工神经网络是多层前馈神经网络模型，通过反向传播算法对权值进行迭代更新，从而实现对非线性特征空间的建模与分类[23]。该方法已被广泛应用于多光谱遥感影像的土地覆盖与目标识别研究中，具有实现稳定、参数可控、结果可复现等优点。利用 BP 网络对遥感图像进行分类主要包括两个部分：学习和模式识别。前者是通过输入经过标注的样本集，使网络不断调整权值以最小化损失函数；后者则是利用训练完成的网络对未知像元或对象进行类别判定[24]。BP 人工神经网络分类法要求属于同一类别的训练样本必须是同质的，不能包含其他类别，也不能是其他类别的边界或混合像素[25]。在分类之前，需要规划网络的结构。研究表明，三层前馈网络可以解决大多数图像分类问题，过深的结构不仅显著增加计算成本，还可能导致梯度消失或过拟合，从而降低分类性能[26]。基于上述原理，本研究采用 ENVI 5.3 软件平台内置的 BP 人工神经网络分类模

块，建立分类训练样本库，并采用三层 BP 网络进行学习。将分类样本输入网络中进行 500 次迭代训练后收敛，经反复试验，将初始权重值、学习速率、动量因子和训练误差分别设置为 0.2、0.8、0.9、0.01，训练完成后的网络模型用于 2024 年西北地区光伏板阵列空间数据集的构建。

1.3.7 精度评价方法

本研究的精度评价基于混淆矩阵评价方法，以独立的人工目视解译样本作为地表真实参考数据，并利用选取的验证样本对不同方法所得到的结果进行精度检验。验证样本的获取基于 2024 年同期 Sentinel-2 MSI 影像，并结合高分辨率 Google Earth 影像进行人工目视解译，且与模型训练样本相互独立，以保证评价精度的客观性与可靠性。采用的精度评价指标有总体精度（OA）、生产者精度（PA）、用户精度（UA）、Kappa 系数、漏分误差（OE）与错分误差（CE）[27-28]，如式（4）－式（9）所示。

$$OA = \sum_{i=1}^{k} N_{ii}/N \quad (4)$$

$$PA = N_{jj}/\sum_{i=1}^{k} N_{ij} \quad (5)$$

$$UA = N_{ii}/\sum_{j=1}^{k} N_{ij} \quad (6)$$

$$kappa = \frac{p_o - p_e}{1 - p_e} \quad (7)$$

$$OE = 1 - PA \quad (8)$$

$$CE = 1 - UA \quad (9)$$

式中，k 为总的分类类别数目，N 为总样本数，N_{ij}为测试样本中本应分为 i 类而在实际结果中却被分为 j 类的样本数目。p_o是每一类正确分类的样本数量之和除以总样本数，p_e指期望准确度[29]。

2 数据样本描述

2.1 数据集基本信息

本数据集包含 2024 年西北地区各省（自治区）的光伏板阵列空间分布数据，总数据量 21.5 MB，属性表中各字段信息见表 2。

表 2 数据属性说明

Table 2 Description of data attributes

字段名	数据类型	描述
Lat	Float	光伏板阵列质心纬度坐标，单位为度
Lon	Float	光伏板阵列质心经度坐标，单位为度
Perimeter	Float	光伏板阵列周长，单位为 m
Area	Float	光伏板阵列面积，单位为 km^2
Province	String	光伏板阵列所属省（自治区）名称
City	String	光伏板阵列所属市名称

2.2 数据统计分析

2024 年，西北地区 5 省（自治区）共有 43 个市（自治州/地区）建有光伏板阵列（表 3），全区光伏板阵列总面积为 1676.13 km^2。其中，新疆维吾尔自治区光伏板阵列面积最大（605.80 km^2），其次是青海省（484.27 km^2）和甘肃省（272.98 km^2），三者占全区光伏板阵列总面积的 81.32%；宁夏回族自治区的光伏板阵列面积（174.65 km^2）略高于陕西省（138.43 km^2）的光伏板阵列面积，分别位居第四和第五位。

从各省（自治区）所辖市（自治州/地区）来看，新疆维吾尔自治区除乌鲁木齐市外，其余 13 个地级市（自治州/地区）都建有光伏板阵列。其中，昌吉、哈密、阿克苏、和田、巴音郭楞和喀什 6 个市（自治州/地区）的光伏板阵列总面积为 480.21 km^2，占全疆的 79.27%。青海省的光伏板阵列集中分布在海南藏族自治州和海西蒙古族藏族自治州，二者光伏板阵列面积（475.09 km^2）占全省的 98.10%。甘肃省光伏板阵列主要分布在河西地区的酒泉、张掖、金昌和武威 4 市，总面积 243.72 km^2，占全省光伏板阵列总面积的 89.28%。宁夏回族自治区以银川市光伏板阵列面积最大（91.96 km^2），其次是吴忠市（43.01 km^2）和中卫市（30.18 km^2），三者合计占全区的 94.56%；陕西省的光伏板阵列则显著集中于榆林市（116.17 km^2，83.92%），位居第 2 位的渭南市光伏板阵列面积仅 14.62 km^2。

表 3　2024 年中国西北地区各省（自治区）光伏板阵列面积统计

Table 3　Area statistics of photovoltaic panel arrays in provinces (autonomous regions) in Northwestern China in 2024

省（自治区）	市（自治州/地区）	光伏板阵列面积及占比		光伏板阵列面积合计及占比	
		面积（km^2）	占比（%）	面积（km^2）	占比（%）
新疆维吾尔自治区	克拉玛依市	2.26	0.37	605.80	36.14
	吐鲁番市	29.67	4.90		
	哈密市	92.42	15.26		
	昌吉回族自治州	125.96	20.79		
	博尔塔拉蒙古自治州	18.25	3.01		
	巴音郭楞蒙古自治州	57.23	9.45		
	阿克苏地区	83.66	13.81		
	克孜勒苏柯尔克孜自治州	17.27	2.85		
	喀什地区	50.57	8.35		
	和田地区	70.37	11.62		
	伊犁哈萨克自治州	8.05	1.33		
	塔城地区	8.17	1.35		
	阿勒泰地区	7.18	1.19		
	阿拉尔市*	15.97	2.64		
	图木舒克市*	4.40	0.73		
	五家渠市*	4.03	0.67		
	铁门关市*	1.64	0.27		
	可克达拉市*	3.15	0.52		
	昆玉市*	1.56	0.26		

省（自治区）	市（自治州/地区）	光伏板阵列面积及占比		光伏板阵列面积合计及占比	
		面积（km²）	占比（%）	面积（km²）	占比（%）
	新星市*	3.99	0.66		
青海省	西宁市	1.27	0.26	484.27	28.89
	海北藏族自治州	4.65	0.96		
	黄南藏族自治州	1.00	0.21		
	海南藏族自治州	253.29	52.30		
	玉树藏族自治州	2.26	0.47		
	海西蒙古族藏族自治州	221.80	45.80		
甘肃省	兰州市	4.40	1.61	272.98	16.29
	嘉峪关市	12.68	4.65		
	酒泉市	106.21	38.91		
	张掖市	39.11	14.33		
	金昌市	55.26	20.24		
	武威市	43.14	15.80		
	白银市	12.18	4.46		
宁夏回族自治区	银川市	91.96	52.65	174.65	10.42
	石嘴山市	9.50	5.44		
	吴忠市	43.01	24.63		
	中卫市	30.18	17.28		
陕西省	铜川市	1.72	1.24	138.43	8.26
	宝鸡市	3.04	2.20		
	咸阳市	2.18	1.57		
	渭南市	14.62	10.56		
	汉中市	0.70	0.51		
	榆林市	116.17	83.92		

注：*表示直辖县级市。

由图 5 可知，西北地区光伏板阵列在空间分布上整体呈现 4 大集聚区格局：环塔里木盆地、准噶尔盆地与柴达木盆地外缘的“环盆集聚区”；沿东疆戈壁与河西走廊荒漠戈壁分布的“戈壁走廊集聚区”；位于宁夏中北部平原至陕北黄土高原一带的“荒漠-草原过渡带集聚区”；以及龙羊峡周边以水光互补为特色的“龙羊峡集聚区”。这些集聚区的形成，本质上是自然条件基础优势与社会政策驱动力量协同作用的结果。一方面，这些集聚区均处于我国太阳能资源最丰富的地带，年日照时数和年太阳总辐射量普遍超过 2800 h 和 5800 MJ/m²，尤以环盆区和戈壁走廊区突出[30]，为光伏发电提供了优越的自然条件。同时，塔里木盆地、准噶尔盆地、柴达木盆地及河西走廊地区地势平坦开阔，分布有大面积荒漠、戈壁等未利用土地，显著降低了土地开发成本。另一方面，政策驱动与基础设施配套发挥了关键作用。在“十四五”期间建设的九座大型清洁能源基地中，新疆、河西走廊、黄河上游和黄河几字湾四大清洁能源基地均布局于西北地区，有力引导光伏项目集中建设。此

外，“哈密-郑州”“酒泉-湖南”“青海-河南”等特高压输电工程建成投运，有效解决了东疆、河西走廊及龙羊峡等地的光伏电力外送与消纳问题。在荒漠-草原过渡带推广的“光伏+治沙”“光伏+牧业”等创新模式，则实现了生态修复与清洁能源生产的协同发展。

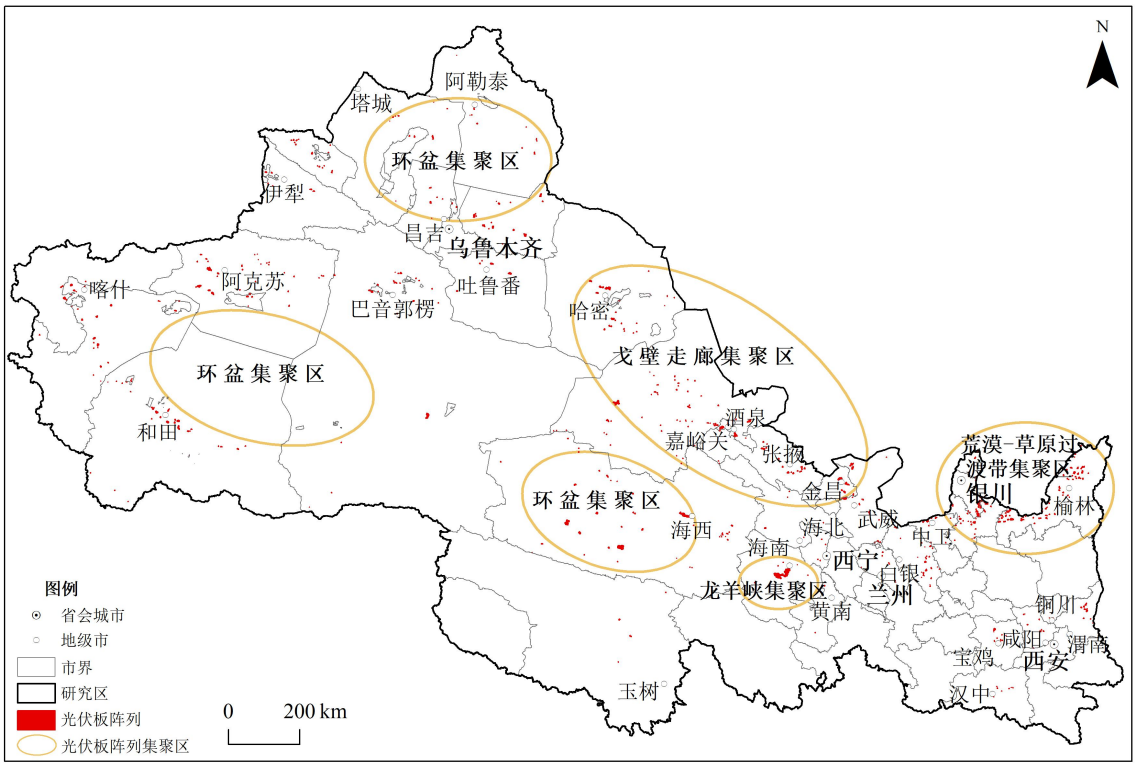

图 5　2024 年中国西北地区光伏板阵列空间分布

Figure 5　Spatial distribution of photovoltaic panel arrays in Northwestern China in 2024

3　数据质量控制和评估

3.1　质量评估

3.1.1　多种方法结果对比

基于 Sentinel-2 MSI 影像，本研究对新疆哈密、青海海西、甘肃金昌、宁夏中卫及陕西榆林五个典型光伏板阵列分布区域，分别采用最近邻法、基于样本的面向对象法以及 BP 人工神经网络法三种方法进行光伏板阵列信息提取，并对提取效果进行了对比分析。由图 6 可知，BP 人工神经网络法在光伏板阵列的信息识别与边界界定中表现最佳，能够有效且清晰地识别光伏板阵列边界；最近邻法提取效果最差，其识别出的光伏板阵列面积显著偏大，并存在明显的提取错误，即错分问题突出，特别是对于道路、建筑用地及植被与光伏板阵列的区别不明显；基于样本的面向对象方法提取精度介于前两者之间，整体表现尚可，仅出现少量漏分和错分，能够较好地区分光伏板阵列与水体、植被与裸地，但对于光伏板阵列与建设用地的区分上仍存在一定不足。综合分析可知，BP 人工神经网络法具有最低的漏分率和最小的错分程度，其提取结果更为精确，是当前识别光伏板阵列信息的最优方法。

3.1.2　与 Wayback 影像对比

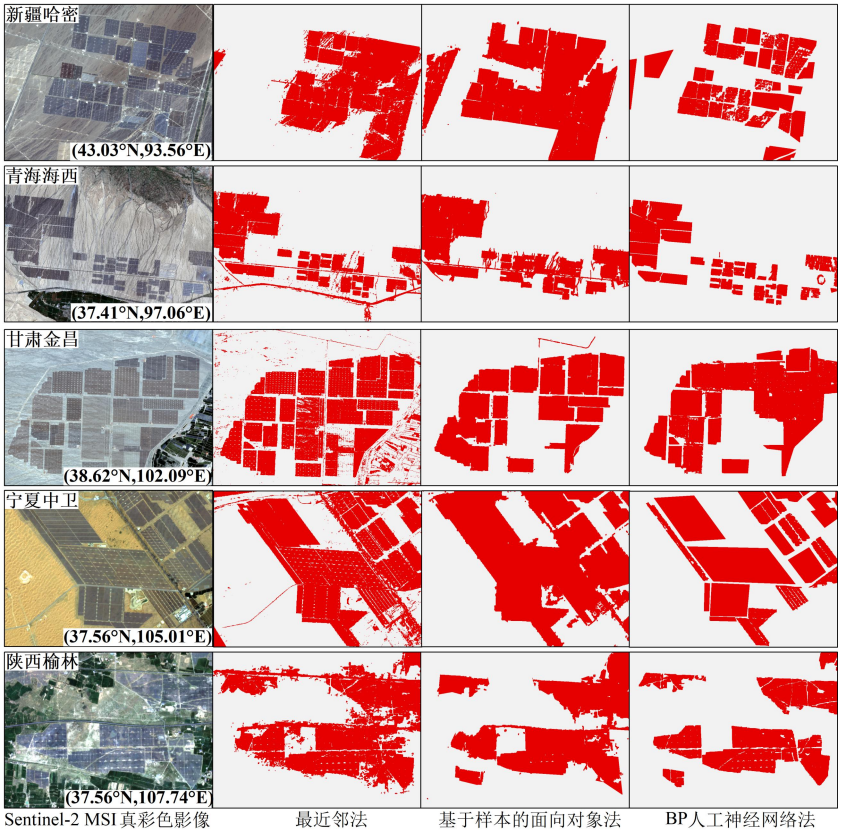

图 6　基于不同方法的典型光伏板阵列结果对比

Figure 6　Comparison of typical photovoltaic panel array extraction results using different methods

为进一步验证本数据集的精度与可靠性，利用 2024 年部分空间分辨率为 0.3 m 的 Wayback 历史遥感影像对提取的光伏板阵列面积进行了对比分析（图 7）。叠加比对与人工目视解译结果表明，本数据集中标注的光伏板阵列位置与 Wayback 影像中实际建成区域高度吻合；在 ArcGIS 中判读的光伏板阵列面积与本数据集提取结果近似一致。由于光伏板阵列之间存在细小间隔，判读结果与提取结果间存在一定差异，但总体一致性较好，证明了本数据集在识别西北地区光伏板阵列方面具有较高的时序敏感性与判读准确性。

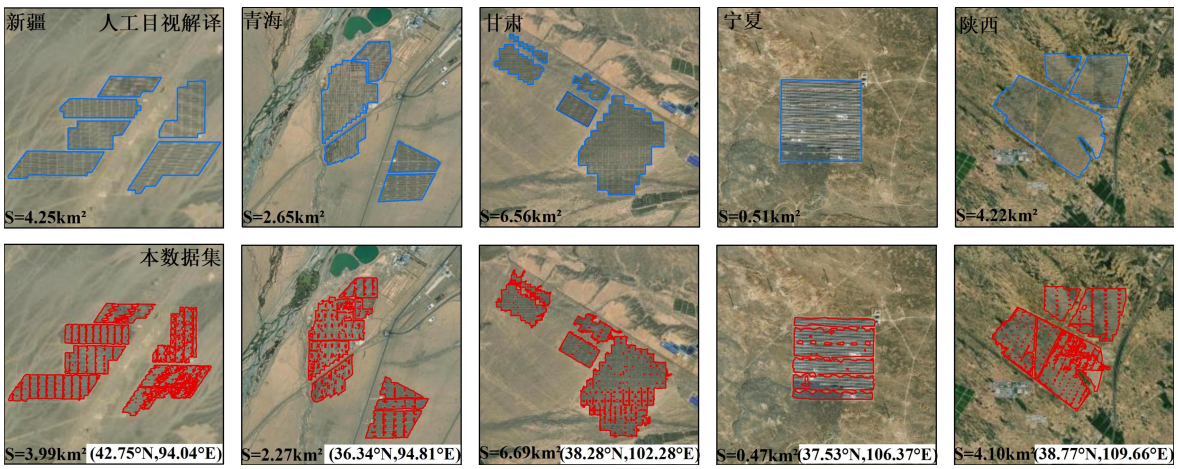

图 7　本数据集与 Wayback 影像对比结果（底图为 Wayback 真彩色影像）

Figure 7　Comparison between this dataset and Wayback imagery (basemap: Wayback true color imagery)

此外，通过对不同地理区域和建设阶段的样本开展交叉验证，结果显示本数据集在空间识别精度上具有良好的稳定性，位置误差控制在影像分辨率允许范围内。总体而言，基于 Wayback 影像的验证结果充分表明，本数据集在时间判定和空间定位两方面均具备较强的适用性和可靠性，可为光伏开发评估、新能源布局分析及遥感建模研究提供坚实的数据支撑。

3.1.3 与 2020 年数据集结果对比

为验证本研究构建的光伏板阵列数据集的外部一致性与时间更新能力，将其与已有的“2020 年全国光伏分类数据集”[5]进行了对比分析（图 8）。统计结果显示，2020 年西北地区光伏板阵列总占地面积为 986.88 km^2，其中新疆（272.54 km^2）、青海（258.06 km^2）、甘肃（188.47 km^2）、宁夏（134.02 km^2）、陕西（133.79 km^2）。至 2024 年，该区域光伏板阵列面积净增长了 689.25 km^2，呈现显著的年际扩张趋势。其中，新疆增长面积最大（333.26 km^2），青海次之（226.21 km^2），陕西最少（4.64 km^2）。通过对图 8 中 A、B、C、D 四组结果的对比分析可知，本数据集在既有光伏板阵列区域与 2020 年数据集保持了较高的空间一致性，且能够完整识别 2020 年后新增的光伏板阵列区域，体现了良好的数据更新能力。同时，图 8 中的 E、F 两组结果表明，2020 年全国光伏板阵列数据集在部分区域将裸地、植被与道路误识别为光伏板阵列，存在明显的错分现象。而本数据集在相同区域有效减少了此类误分，并对光伏板阵列边界实现了更完整的提取。综上，该对比分析从跨年度一致性、新增目标识别能力以及复杂背景下的误分控制等方面，验证了本研究数据集在空间识别准确性和数据更新能力方面的可靠性。

3.1.4 外部一致性验证

为进一步验证本数据集在区域尺度上的合理性与可靠性，本研究引入截至 2024 年底各省累计并网装机容量的官方统计数据[31]，构建装机容量与遥感提取面积之间的外部一致性检验框架。需要说明的是，本研究基于中高分辨率遥感影像提取的光伏板阵列主要表现为大尺度连续分布的地面电站形态，难以识别分散布局的屋顶分布式光伏。因此，在外部一致性验证过程中，本研究采用集中式累计并网装机容量作为统计基准，以保证容量口径与遥感识别对象在空间形态上的可比性。该处理方式有助于避免分布式光伏对区域面积匹配关系产生干扰，提高一致性检验的科学性。大型地面光伏电站的装机容量与占地面积之间通常存在相对稳定的对应关系，可用“单位面积装机容量（MW/km^2）”作为表征光伏阵列占地容量强度的核心指标，为区域尺度遥感提取结果的合理性评估提供可量化的参照基准。值得指出的是，工程规划与用地评估研究中普遍引用的装机密度范围（35–60 MW/km^2）多针对项目用地尺度，其面积统计通常包含道路、升压站、运维区、场区空地等非组件区域[32]。本研究基于 Sentinel-2 MSI 影像提取的面积为光伏板阵列尺度，更接近组件阵列的空间覆盖范围，因此其对应的单位面积装机密度通常显著高于项目用地尺度。为避免尺度口径不一致带来的偏差，本研究采用各省（自治区）“累计装机容量—阵列提取面积”的反推结果，得到西北地区地面光伏阵列的单位面积装机密度整体介于 70–180 MW/km^2 范围内，并将此范围作为理论估算区间进行后续合理性判断。

本研究以新疆、青海、甘肃、宁夏和陕西五省（自治区）为验证基准，分别采用装机密度上限（180 MW/km^2）与下限（70 MW/km^2）计算装机容量所对应的光伏板阵列理论的面积区间，计算如下：

$$A = \frac{C}{D} \quad (10)$$

式中，A 为光伏板阵列理论面积（km²）；C 为累积并网装机容量（MW）；D 为装机密度（MW/km²）。

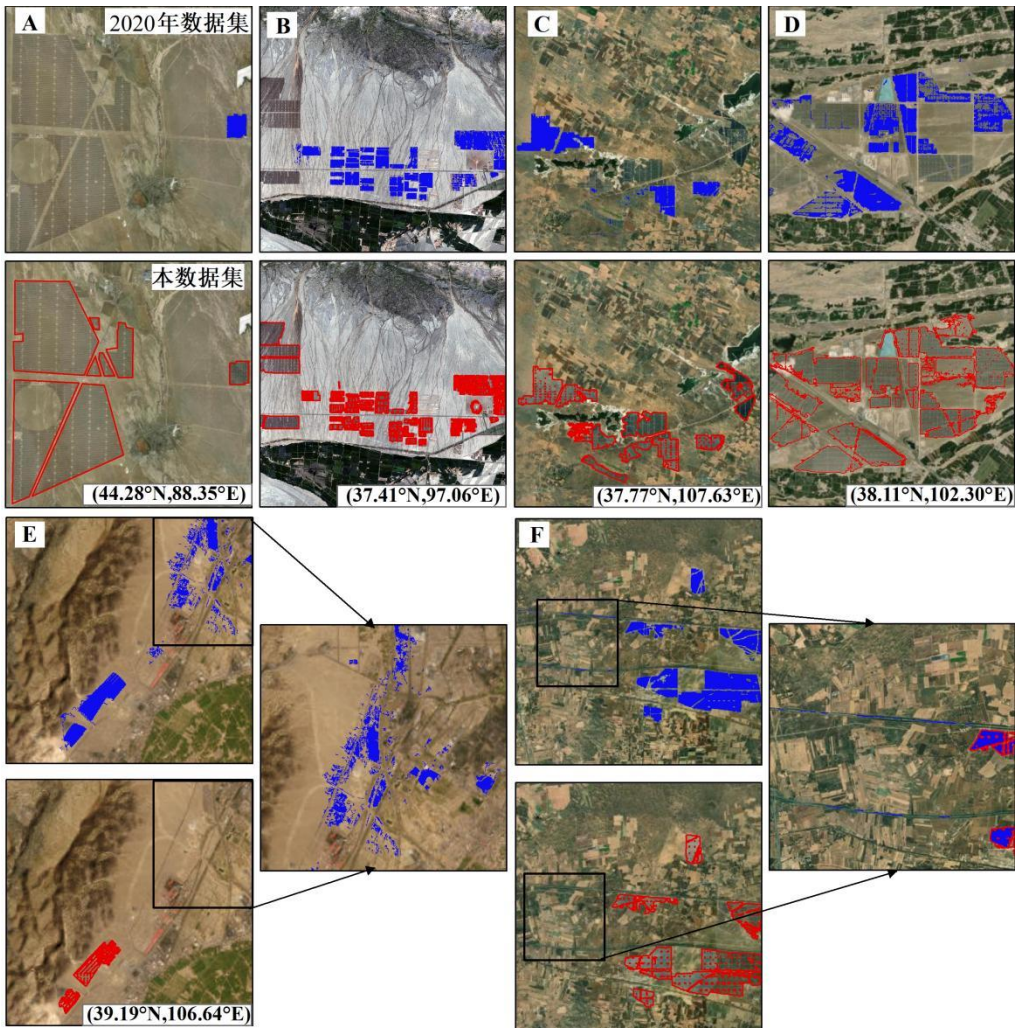

图 8　本数据集与 2020 年全国光伏板阵列数据集结果对比（A-D 为数据更新示例，E、F 为 2020 年数据集错判示例；底图为 Sentinel-2 MSI 真彩色影像）

Figure 8　Comparison between this dataset and the 2020 national photovoltaic panel array dataset (A-D: examples of dataset updates; E-F: misclassification cases in the 2020 dataset; basemaps: Sentinel-2 MSI true-color imagery)

基于上述方法，可将各省集中式累计并网装机容量换算为对应的理论光伏阵列面积区间。其中，新疆 53187 MW 的装机容量对应的理论阵列面积约为 295–760 km²，青海 36043 MW 对应约 200–515 km²，甘肃 29549 MW 对应约 164–422 km²，宁夏 24493 MW 对应约 136–350 km²，陕西 25239 MW 对应约 140–361 km²。将上述理论面积区间与遥感提取的光伏板阵列面积进行对比（图 9），可以发现各省提取面积均落在或接近理论估算区间范围内，整体数量级一致性较好。该外部一致性检验进一步验证了本数据集在表征区域装机规模与支持面积统计分析方面的可靠性，为其在后续定量研究与区域规划评估中的应用提供了坚实的数据支撑。

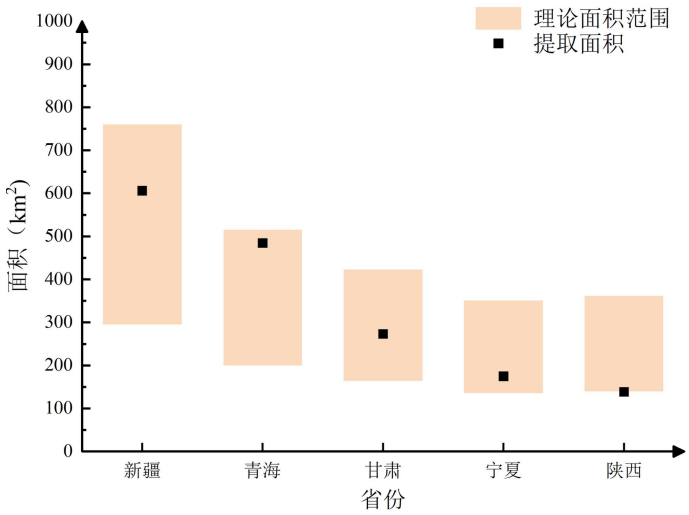

图 9　2024 年西北地区各省（自治区）光伏板阵列理论面积区间与遥感提取面积对比

Figure 9　Comparison between theoretical and remotely sensed areas of photovoltaic panel arrays across provinces (autonomous regions) in Northwestern China in 2024

3.2　精度评价

为定量评估三种提取方法的分类性能，本研究采用混淆矩阵方法，将分类结果与人工目视解译得到的验证样本数据进行对比分析。混淆矩阵通过比较地表真实数据与分类结果来反映分类器的判别能力[33]，在本研究中，所有精度指标均基于全部 304 景影像的验证样本合并后生成的总体混淆矩阵计算所得。该方法避免了因单景影像样本数量不均可能导致的结果偏差，能更稳健地反映分类器在整个研究区的总体性能[34]。不同方法得到的各类精度指标如表 4 所示。由表中结果可知，在使用同一时期、同一数据源影像的情况下，BP 人工神经网络法对光伏板阵列的识别结果最接近实际情况，提取精度最高；基于样本的面向对象法次之，最近邻法的精度最低。

表 4　基于 Sentinel-2 MSI 影像不同方法提取的光伏板阵列精度评价

Table 4　Accuracy assessment of photovoltaic panel array extraction using different methods based on Sentinel-2 MSI imagery

方法	用户精度（%）	错分误差（%）	生产者精度（%）	漏分误差（%）	总体精度（%）	Kappa 系数
最近邻法	64.51	35.49	98.61	1.39	85.06	0.81
基于样本的面向对象法	91.34	8.66	96.14	3.86	87.93	0.85
BP 人工神经网络法	76.46	23.54	99.38	0.62	91.56	0.89

4　数据价值

本数据集基于 Sentinel-2 MSI 遥感影像，利用 BP 人工神经网络模型实现了中国西北地区光伏板阵列空间数据集的制作，包含位置、面积等关键信息，具有较高的科研与应用价值。作为区域尺度的光伏板阵列基础空间数据，本数据集可直接用于新能源资源调查与统计分析，为研究西北地区光伏板阵列分布格局、建设规模及其空间差异提供可靠的数据支撑。同时，本数据集可作为土地利用/

覆盖变化研究中的专题图层，用于分析光伏用地扩展特征及其与其他地类之间的空间关系。此外，本数据集可为光伏板阵列识别方法研究提供统一的验证对象与对比基准，有助于在不同算法与数据源条件下进行提取结果的横向比较与精度评估。结合历史数据或其他公开数据集，还可支持光伏建设变化的基础统计分析与空间演变研究。

综上所述，本数据集可用于支撑区域尺度光伏板阵列空间分布制图、新能源建设规模统计、光伏用地扩展分析以及相关遥感识别方法的对比验证，为我国西北地区在“双碳”战略目标下开展光伏发展相关研究提供基础空间数据支撑与科学决策依据。

5 AIGC 使用声明

在撰写本手稿期间，作者仅使用 ChatGPT-3.5 进行语法检查、语言润色和语句的流畅度优化。所有生成的内容均由作者进行修改和确认，作者对论文的整体内容、数据真实性及学术诚信负全部责任。

作者分工职责

魏亚宁（2002—），女，甘肃省兰州市人，硕士研究生，研究方向为环境遥感。主要承担工作：数据收集、处理及论文撰写。

姚晓军（1980—），男，山西夏县人，博士，教授，研究方向为地理信息技术与冰冻圈变化。主要承担工作：总体方案设计、论文过程指导和文稿撰写。

参考文献

[1] 张弦, 王明培, 史伟. 碳中和目标下鄂尔多斯地区的光伏发电路径情景分析[J]. 太阳能, 2024, 4: 20-28. [ZHANG X, WANG M P, SHI W. Scenario analysis of photovoltaic power generation pathways in the Ordos region under the carbon neutral target[J]. Solar Energy, 2024, 4: 20-28.]

[2] 屈准, 杨肃昌, 肖建华. 双碳背景下中国西北地区光伏电站建设现状与潜力分析[J]. 干旱区资源与环境, 2024, 38(2): 20-26. [QU Z, YANG S C, XIAO J H. Analysis of the current situation and potential of photovoltaic power plant construction in Northwest China under the dual-carbon background[J]. Arid Zone Resources and Environment, 2024, 38(2): 20-26.]

[3] MALOOF J M, BRADBURY K, COLLINS L M, et al. Automatic detection of solar photovoltaic arrays in high resolution aerial imagery[J]. Applied Energy, 2016, 183: 229-240.

[4] CHENG G, XIE X, HAN J, et al. Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 3735-3756.

[5] FENG Q L, NIU B, REN Y, et al. A 10-m national-scale map of ground-mounted photovoltaic power stations in China of 2020[DS/OL]. V3. Science Data Bank, 2023. https://doi.org/10.57760/sciencedb.o00121.00001. DOI:10.57760/sciencedb.o00121.00001.

[6] LI A, LIU L, LI S, et al. Global photovoltaic solar panel dataset from 2019 to 2022[J]. Scientific Data, 2025, 12(1): 637.

[7] JI C, BACHMANN M, ESCH T, et al. Solar photovoltaic module detection using laboratory and airborne imaging spectroscopy data[J]. Remote Sensing of Environment, 2021, 266: 112692.

[8] ZHU H, ZHANG C, GAO L, et al. Multi-Source Collaborative Domain Generalization for Cross-Scene Remote Sensing Image Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5535815.

[9] 李学武, 张京朋, 赵天保, 等. 我国西北地区干湿变化特征及其未来预估[J]. 气候与环境研究, 2024, 29(1): 59-74. [LI X W, ZHANG J P, ZHAO T B, et al. Characteristics of wet and dry changes in Northwest China and their future prediction[J]. Climate and Environment Research, 2024, 29(1): 59-74.]

[10] 薛晨, 程林, 任景. 清洁电能出西北 坚强电网送四方[N]. 国家电网报, 2023.02.07. [XUE C, CHENG L, REN J. Clean power from Northwest China to send strong power grid to all directions[N]. State Grid News, 2023.02.07.]

[11] DRUSCH M, DEL BELLO U, CARLIER S, et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services[J]. Remote Sensing of Environment, 2012, 120: 25-36.

[12] MAIN-KNORN M, PFLUG B, LOUIS J, et al. Sen2Cor—L2A processor for Sentinel-2: software description and user information[EB/OL]. (2017-03-20) [2025-12-07]. https://step.esa.int/main/third-party-plugins-2/sen2cor/.

[13] HOSSAIN M D, CHEN D. Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 150: 115-134.

[14] 丘鸣语, 甘淑. 面向对象的多层次规则分类地物遥感信息提取方法试验分析研究[J]. 贵州大学学报(自然科学版), 2023, 40(2): 67-73. [QIU M Y, GAN S. An experimental and analytical study on object-oriented multi-level rule classification method for remote sensing information extraction of ground features[J]. Journal of Guizhou University (Natural Science Edition), 2023, 40(2): 67-73.]

[15] QIN R, LIU T. A review of landcover classification with very-high resolution remotely sensed optical images—Analysis unit, model scalability and transferability[J]. Remote Sensing, 2022, 14(3): 646.

[16] CHEN G, WENG Q, HAY G J, et al. Geographic object-based image analysis (GEOBIA): Emerging trends and future opportunities[J]. GIScience & Remote Sensing, 2018, 55(2): 159-182.

[17] MONTERO D, AYBAR C, MAHECHA M D, et al. A standardized catalogue of spectral indices to advance the use of remote sensing in Earth system research[J]. Scientific Data, 2023, 10(1): 197.

[18] MA L, LIU Y, ZHANG X, et al. Deep learning in remote sensing applications: A meta-analysis and review[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 152: 166-177.

[19] YANG R, QI Y, ZHANG H, et al. A study on the object-based high-resolution remote sensing image classification of crop planting structures in the loess plateau of eastern Gansu province[J]. Remote Sensing, 2024, 16(13): 2479.

[20] YILMAZ E O, KAYZOGL T. An In-depth Investigation of OBIA Classification with High-Resolution Imagery: Unravelling the Explanations Behind Deep Learning and Machine Learning[J]. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2025, 48: 317-324.

[21] THANH NOI P, KAPPAS M. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery[J]. Sensors, 2017, 18(1): 18.

[22] AKBULUT Y, SENGUR A, GUO Y, et al. NS-k-NN: Neutrosophic set-based k-nearest neighbors classifier[J]. Symmetry, 2017, 9(9): 179.

[23] RICHARDS J A, JIA X. Remote sensing digital image analysis: an introduction[M]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.

[24] SONI P K, RAJPAL N, MEHTA R, et al. Urban land cover and land use classification using multispectral sentinal-2 imagery[J]. Multimedia Tools and Applications, 2022, 81(26): 36853-36867.

[25] 赵英时. 遥感应用分析原理与方法[M]. 北京: 科学出版社, 2003: 421-425. [ZHAO Y S. Principles and Methods of Remote Sensing Application Analysis [M]. Beijing: Science Press, 2003: 421-425.]

[26] REZAEE M, MAHDIANPARI M, ZHANG Y, et al. Deep convolutional neural network for complex wetland classification using optical remote sensing imagery[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11(9): 3030-3039.

[27] FOODY G M. Status of land cover classification accuracy assessment[J]. Remote Sensing of Environment, 2002, 80(1): 185-201.

[28] CONGALTON R G. A review of assessing the accuracy of classifications of remotely sensed data[J]. Remote Sensing of Environment, 1991, 37(1): 35-46.

[29] 白方铄, 马骁, 杨永均, 等. 基于三参数阈值分割的露天矿山土地遥感分类方法[J]. 测绘与空间地理信息, 2024, 47(5): 27-30. [BAI F S, MA X, YANG Y J, et al. Remote sensing classification of open pit mine land based on three-parameter threshold segmentation[J]. Mapping and Spatial Geographic Information, 2024, 47(5): 27-30.]

[30] HE Y, WANG K, YANG K, et al. Homogenized daily sunshine duration over China from 1961 to 2022[J]. Earth System Science Data, 2025, 17(4): 1595-1611.

[31] 国家能源局. 2024 年光伏发电建设情况[EB/OL]. (2025-01-27) [2026-02-07]. https://www.nea.gov.cn/20250221/f04452701c914d51a89d0c0ea6f4acd1/c.html. [National Energy Administration (NEA). China's PV power development in 2024 [EB/OL]. (2025-01-27) [2026-02-07]. https://www.nea.gov.cn/20250221/f04452701c914d51a89d0c0ea6f4acd1/c.html.]

[32] HERNANDEZ R R, HOFFACKER M K, FIELD C B. Land-use efficiency of big solar[J]. Environmental Science & Technology, 2014, 48(2): 1315-1323.

[33] 韩德祥, 李成鹏. 基于 Landsat8 OLI 遥感影像冰湖边界分类叠加提取[J]. 北京测绘, 2020, 34(5): 645-650. [HAN D X, LI C P. Classification overlay extraction of glacial lake boundary based on Landsat8 OLI remote sensing images[J]. Beijing Surveying and Mapping, 2020, 34(5): 645-650.]

[34] CONGALTON R G, GREEN K. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices[M]. Boca Raton: CRC Press, 2008.

论文引用格式

魏亚宁, 姚晓军. 2024 年中国西北地区光伏板阵列空间数据集[J/OL]. 中国科学数据, 2026, 11(1). (2026-03-18). DOI: 10.11922/11-6035.csd.2025.0157.zh.

数据引用格式

魏亚宁, 姚晓军. 2024 年中国西北地区光伏板阵列空间数据集[DS/OL]. 中国科学院地球环境研究所黄土科学数据中心, 2026. (2026-03-13). DOI: 10.12262/lsdc.17040.2541.

A spatial dataset of photovoltaic panel arrays in Northwestern China in 2024

WEI Yaning[1], YAO Xiaojun[1*]

1. College of Geography and Environment Sciences, Northwest Normal University, Lanzhou 730070, P.R. China

*Email: yaoxj_nwnu@163.com (YAO Xiaojun)

Abstract: The large-scale development of photovoltaic (PV) panel arrays in Northwestern China has proposed an created an urgent demand for high-precision spatial information extraction and management. In this study, Sentinel-2 MSI imagery in 2024 was used to establish a dataset of PV panel arrays. By integrating spectral and spatial features, a multi-scale segmentation approach was applied, and three classification methods—nearest neighbor, sample-based object-oriented, and BP neural network—were comprehensively evaluated. Accuracy assessment based on confusion matrices with independent validation samples identified the BP artificial neural network as the optimal model, resulting in a 10 m spatial resolution dataset of photovoltaic panel arrays in Northwestern China, with an overall accuracy (OA) of 91.6% and a Kappa coefficient of 0.89, demonstrating high precision and reliability. Temporal comparison and typical-area spatial distribution analysis using the "10 m National Ground-Mounted PV Dataset of 2020" and historical Wayback imagery, further confirmed the dataset's reliability and advantages in temporal updating and spatial identification accuracy. The results indicate that by 2024, the total area covered by PV panel arrays in Northwestern China reached 1,676.13 km^2, with the largest areas in Xinjiang (605.80 km^2), Qinghai (484.27 km^2), Gansu (272.98 km^2), Ningxia (174.65 km^2), and Shaanxi (138.43 km^2). These installations are mainly concentrated along the periphery of the Taklamakan Desert, the northern slopes of the Tianshan Mountains, Haixi and Hainan Prefectures in Qinghai, the Hexi Corridor in Gansu, Yinchuan and Wuzhong in Ningxia, and Yulin in Shaanxi. This dataset provides a reliable basis for energy planning, ecological assessment, and sustainable development under China's "dual-carbon" strategy.

Keywords: photovoltaic panel arrays; remote sensing information extraction; multi-scale segmentation; spatial distribution; Northwestern China

Dataset Profile

Title	A dataset of photovoltaic panel arrays in Northwestern China in 2024
Data corresponding author	YAO Xiaojun (yaoxj_nwnu@163.com)
Data authors	WEI Yaning, YAO Xiaojun
Time range	2024

Geographical scope	Northwestern China (32°N–50°N and 73°E–112°E), including Gansu Province, Shaanxi Province, Ningxia Hui Autonomous Region, Qinghai Province, and Xinjiang Uygur Autonomous Region
Spatial resolution	10 m
Data volume	21.5 MB
Data format	ESRI Shapefile
Data service system	https://doi.org/10.12262/lsdc.17040.2541 https://www.scidb.cn/doi/10.12262/lsdc.17040.2541
Source of funding	Oasis Science Research Achievement Breakthrough Action Plan of Northwest Normal University (NWNU-LZKX-202301)
Dataset composition	The dataset consists of one data file, i.e., the spatial distribution data of photovoltaic panel arrays covering the five provinces (autonomous regions) in Northwestern China in 2024.