

面向 AI-Ready 的路面状态标准化图像数据集

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



文献 CSTR:
32001.14.11-6035.ncdc.2025.0043.zh



文献 DOI:
10.11922/11-6035.ncdc.2025.0043.zh

数据 DOI:
10.12072/ncdc.dprsc.db6686.2024

文献分类: 地球科学

收稿日期: 2025-03-26

录用日期: 2025-06-16

发表日期: 2025-12-17

刘景琦^{1,2,3}, 张耀南^{1,2*}, 康建芳^{1,2}, 刘杰^{4,5}, 王斌^{4,5}

1. 中国科学院西北生态环境资源研究院, 兰州 730030
2. 国家冰川冻土沙漠科学数据中心, 兰州 730030
3. 中国科学院大学, 北京 100049
4. 新疆交通规划勘察设计研究院有限公司, 乌鲁木齐 830006
5. 新疆高寒高海拔山区交通基础设施安全与健康重点实验室, 乌鲁木齐 830006

摘要: 针对当前国际公开道路表面气象条件标准化数据集的稀缺性, 尤其是极端天气下路面状态图像资源的不足, 本研究构建了面向 AI-Ready 的 6 类路面状态标准化图像数据集, 旨在为深度学习模型提供高可靠性的训练基础。基于极端天气对交通影响的统计特征, 数据集将冰雪灾害细化为路面结冰、强降雪和风吹雪三大类, 通过整合公路监控、移动设备及网络开源数据, 最终形成包含干燥、积雪、结冰、风吹雪、融雪和潮湿 6 种典型路面状态的标准化图像数据集。为避免数据增强带来的子集间相关性干扰, 采用先划分后增强的策略, 将原始数据独立分割为训练集 (60%)、验证集 (20%) 和测试集 (20%), 再对每个子集分别实施翻转、旋转、平移及高斯噪声添加等增强操作, 最终扩展至 9000 张图像规模。数据集提供深度学习训练预处理参数, 其 RGB 三通道均值与标准差分别为[0.550, 0.565, 0.568]和[0.082, 0.082, 0.085], 通过零均值单位方差归一化可有效提升模型训练效率。本数据集系统性覆盖冰雪灾害多相态路面特征, 为智能交通系统中的路面状态识别算法开发与评估建立了标准化基准。

关键词: AI-Ready; 图像识别; 深度学习; 路面状态

数据库 (集) 基本信息简介

数据库 (集) 名称	面向 AI-Ready 的路面状态标准化图像数据集 (2023–2024 年)
数据作者	刘景琦
数据通信作者	刘景琦 (liujingqi@nieer.ac.cn)
数据时间范围	2023年10至2024年10月
数据量	335 MB
数据格式	*.jpg
数据服务系统网址	http://doi.org/10.12072/ncdc.dprsc.db6686.2024 http://scidb.cn/doi/10.12072/ncdc.dprsc.db6686.2024
基金项目	国家重点研发计划 (2022YFF0711704); 2022年度新疆交通运输行业科技项目 (2022-ZD-006); 新疆交投2021年揭榜挂帅科技项目 (ZKX FWCG2022060004); 新疆交通设计院公司科研基金 (KY2022041101)。

* 论文通信作者

张耀南: yaonan@lzb.ac.cn

数据库（集）组成

数据集由4个数据文件夹、1个json文件、2个python脚本组成，具体包括：（1）data文件夹，包含6个子集，分别存放6类标准化典型路面状态数据；（2）train、val、test文件夹，分别存放已划分的训练集，验证集和测试集；（3）class_indices.json 文件，提供标准化标签索引；（4）split_data.py 用于训练集、验证集和测试集的划分；（5）calculate_mean_variance.py用于计算数据集的均值和方差。

引言

作为现代交通网络的核心枢纽，高速公路不仅是保障公众高效出行的重要基础设施，更是区域经济发展的战略支撑，其安全性和畅通性直接关系到民生福祉，并深刻影响着区域间要素资源的优化配置与产业协同发展^[1]。特别是在构建新发展格局的背景下，高速公路网络对促进区域经济一体化、提升产业链供应链现代化水平具有显著的全局性带动作用。然而，在高寒山区，特殊气候环境使得这一交通命脉面临着严峻的考验。以独库公路为例，由于其沿线横跨多种地质构造、地形地貌、气候类型及植被土壤带，导致风吹雪、雪崩、崩塌、泥石流、滑坡等地质灾害广泛分布。尤其是每年长达半年的冰雪灾害高发期（10月至次年3月），极易形成强降雪、风吹雪和路面结冰等复合型灾害^[2-4]。这些灾害的链式反应不仅威胁行车安全，也会对区域经济运行产生系统性冲击。如何破解冰雪困局，已成为高寒地区交通治理与经济发展的关键命题。

针对高寒山区复杂气候与道路灾害频发的特征，构建高速公路路面状态智能化监测系统具有紧迫性和实践价值。具体而言，该系统通过实时风险感知与分级响应机制，显著提升行车安全性。同时，依托自动化除冰设备与动态交通调度技术，提升路网运营效率，实施精准干预，从而形成从风险感知到处置响应的全链条保障机制。然而，技术落地所面临的核心挑战在于数据基座的不完善。世界范围内可用的标准化路面状态数据集较少，特别是极端气象场景下的路面图像样本尤为匮乏。这类数据不仅是深度学习模型的基础要素，更是制约模型性能的关键因素。现有数据集难以覆盖风吹雪、道路结冰等极端气象过程，导致深度学习模型无法有效捕捉多维度气象作用下的路面特征规律，直接影响结冰预警的时空精度。

为了解决这一关键问题，本研究通过整合高速公路摄像头、移动镜头及在线资源，构建了涵盖6种路面状况的数据集。高速公路摄像头提供了连续、实时的路面监测图像；移动镜头在不同的地点和角度获取路面信息，增加了数据的多样性；在线资源则补充了特殊场景或极端天气下的图像数据，进一步完善了数据集。在数据处理环节，针对深度学习模型的应用需求，实施了光照校正与标准化处理。通过光照校正，消除光照差异的影响，使图像中的路面特征更加清晰和一致。标准化处理则可以统一数据的格式和特征，提高模型的训练效率和准确性。

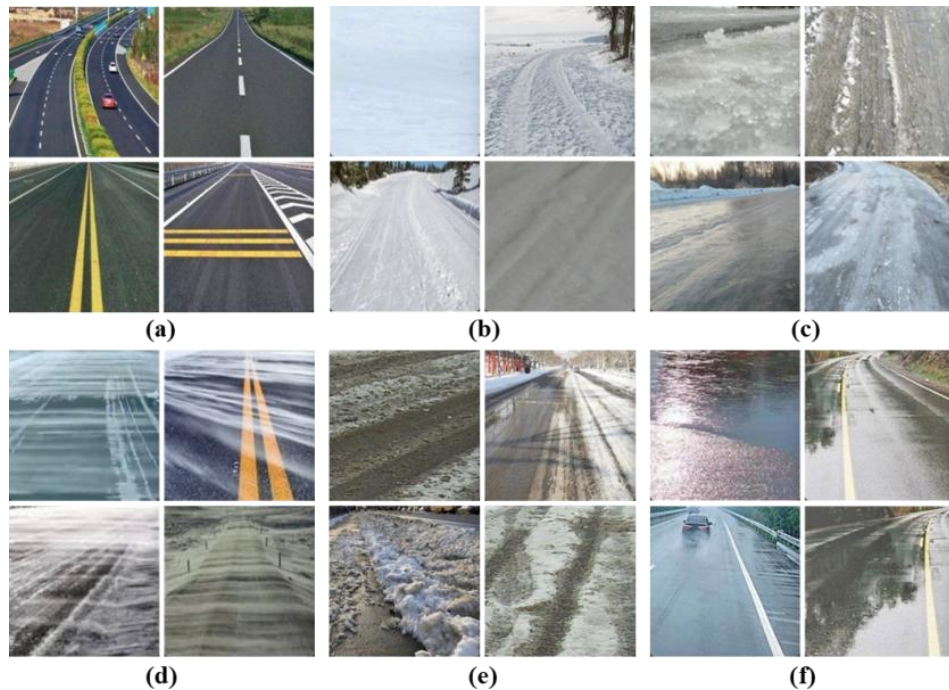
通过上述实践，本研究为提升路面状态识别模型的准确性提供了关键的数据支撑，使得深度学习模型能够更有效地学习和识别不同路面状态的特征，从而提高高速公路实时监测与管理的技术可靠性，最终保障高寒山区高速公路的安全畅通和当地经济的稳定发展。

1 数据采集和处理方法

1.1 数据采集方法

基于对天山高速公路在极端天气条件下交通管制的统计记录及沿线气象数据的系统分析（数据

来源：<https://cxfw.jtyst.xinjiang.gov.cn/roadCondition>），本研究结合不同类型的冰雪灾害对道路通行能力的影响，将天山高速公路冰雪灾害主要划分为三类：道路结冰、强降雪和风吹雪。基于此，本研究通过整合天山高速公路沿线布设的监控摄像系统（官方网址：<https://cxfw.jtyst.xinjiang.gov.cn/home/index>）、移动设备及互联网开源数据，系统性地构建了一套路面状态数据集，采样周期覆盖 2023 年 10 月至 2024 年 10 月。其中，监控摄像系统受限于数据开放政策，无法获取摄像头具体参数。由于道路结冰和风吹雪属于较为特殊的极端天气现象，其自然发生频率较低，导致原始数据样本数量相对不足。为提升数据集的鲁棒性，确保模型对不同类别特征的均衡学习，本研究采用数据增强技术对所有类别的图像样本进行了扩充，使数据集总量达到 9000 张，有效降低了因数据不平衡引发的模型偏差问题。本数据集涵盖 6 种典型路面状态，全面表征了天山高速公路在不同气象条件下的通行环境。具体包括：（1）干燥路面；（2）积雪路面；（3）结冰路面；（4）风吹雪路面；（5）融雪路面；（6）潮湿路面。表 1 展示了各类别样本的分布情况，图 1 展示了相应类别的示例图像。



(a) 干燥路面；(b) 积雪路面；(c) 结冰路面；(d) 风吹雪路面；(e) 融雪路面；(f) 潮湿路面 (a) dry road surface; (b) snow-covered road surface; (c) icy road surface; (d) blowing-snow road surface; (e) snow-melting road surface; (f) wet road surface

图 1 数据样本

Figure 1 Dataset samples.

表 1 数据分布

Table 1 Dataset distribution

路面状态分类	干燥	积雪	结冰	风吹雪	融雪	潮湿
原始图像	898	499	275	92	336	402
增强图像	1500	1500	1500	1500	1500	1500
样本大小 (MB)	28.6	17.2	36.0	19.8	38.1	27.6

1.2 数据处理

在深度学习任务中,数据预处理对模型的泛化能力和性能稳定性具有至关重要的作用。本研究采用了一套系统化的数据预处理方法,该流程不仅涵盖了常规的数据增强操作,还包括数据集划分、光照校正以及归一化参数计算等关键步骤。通过这一系列预处理操作,本研究旨在优化数据质量,最大限度地减少潜在的数据偏差,从而为后续的模式训练与评估提供可靠的数据基础。数据预处理的完整流程如图 2 所示,其设计充分考虑了数据特征的多样性与模型需求的匹配性,以确保预处理结果的科学性与实用性。值得注意的是,数据集的划分与增强顺序对模型性能评估具有重要影响,在数据集划分之前进行数据增强操作,可能会在训练集、验证集和测试集之间引入潜在的关联性,这种关联性主要体现在数据分布特征的相似性上,可能破坏验证集和测试集的独立性,进而导致模型性能评估的准确性和可靠性降低。为有效解决这一问题,本研究首先将原始数据集划分为训练集、验证集和测试集 3 个独立的子集,再分别对每个子集进行独立的数据增强操作。这种处理方式能够确保各个子集之间的独立性,从而为模型性能评估提供更为可靠的基准。此外,与目标检测、语义分割等计算机视觉任务不同,图像分类任务的样本标注流程相对简单。通过定义类别索引的方式,仅需单一维度的数字化编码即可建立样本与标签的映射关系,详见第 2.2 小节。

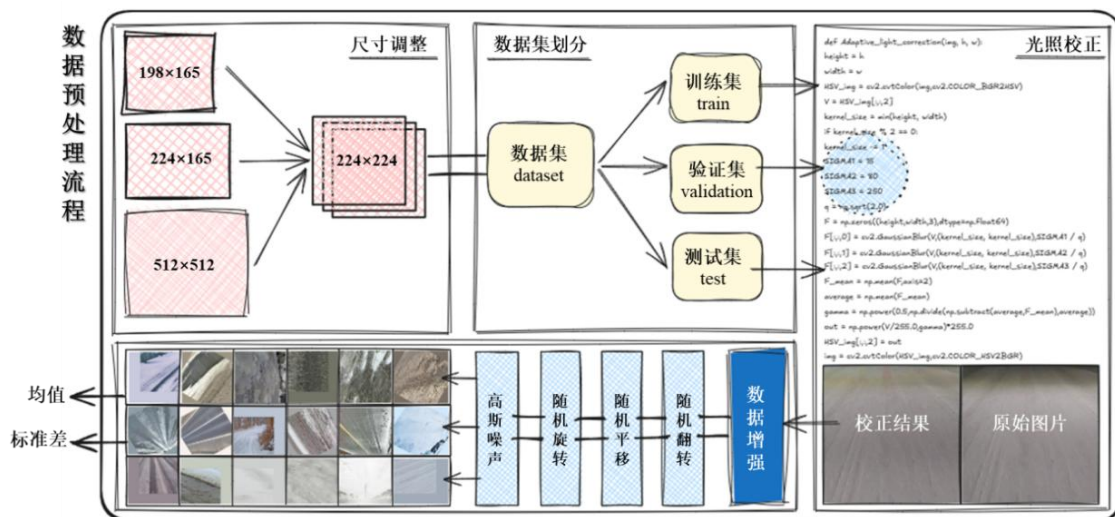


图 2 数据预处理流程

Figure 2 Data preprocessing workflow

1.2.1 数据尺寸调整

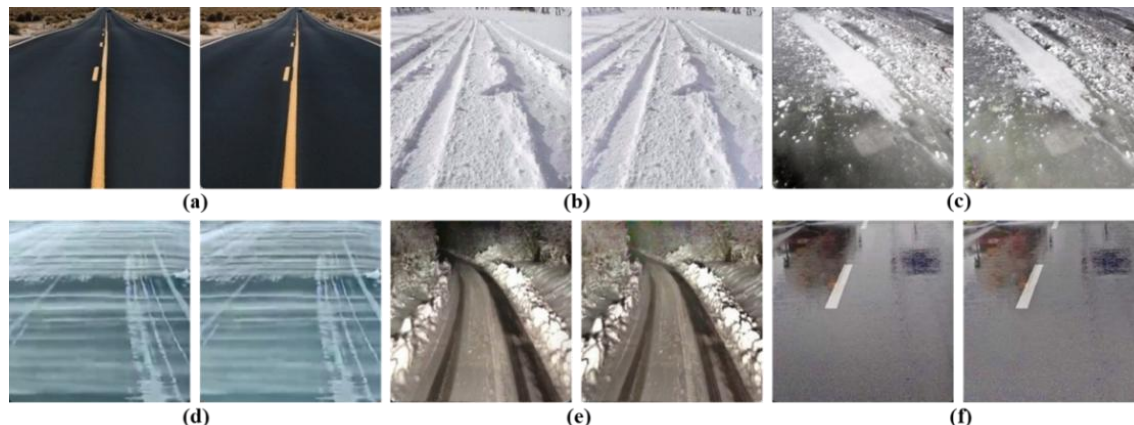
针对多源数据图像尺寸不一致的问题,采用标准化处理策略,将图像统一调整为 224×224 像素。 224×224 像素作为深度学习领域的标准输入尺寸,在计算资源消耗与模型性能优化之间实现较好的平衡,既能有效降低计算复杂度,又能保留足够的图像信息,为模型的特征提取过程提供支撑。同时,该像素尺寸与基于 ImageNet 等大规模数据集的预训练模型保持一致,如 VGG、ResNet 等,确保了数据与主流深度学习模型的兼容性。

1.2.2 数据集划分

为确保模型训练、验证和测试的独立性,本研究将数据集随机划分为训练集、验证集和测试集,比例分别为 60%、20%和 20%。

1.2.3 光照校正

针对复杂路面条件下可能出现的物体遮挡和光照不均匀问题,本研究采用基于二维伽马函数的自适应校正算法对图像亮度进行调整。该算法能够有效解决图像中过亮或过暗区域的问题,从而提升关键细节的可见性,并降低不同类型路面外观相似性带来的识别难度^[9]。校正结果如图 3 所示。



(a) 干燥路面; (b) 积雪路面; (c) 结冰路面; (d) 风吹雪路面; (e) 融雪路面; (f) 潮湿路面 (a) dry road surface; (b) snowy-covered road surface; (c) icy road surface; (d) blowing-snow road surface; (e) snow-melting road surface; (f) wet road surface

图 3 数据集样本的亮度调整

Figure 3 Brightness adjustment of dataset samples

1.2.4 数据增强

为解决数据集中类别不平衡问题,本研究通过数据增强技术扩充了数据规模。数据增强方法包括翻转、平移、旋转及加噪等,具体参数配置如下。翻转操作基于离散参数控制,随机选择垂直、水平或双向翻转模式;平移在水平与垂直方向分别设定 ± 50 像素的偏移幅度,边界区域采用随机像素值填充;旋转以图像中心为原点,从预设角度 $\pm 45^\circ$, $\pm 60^\circ$, $\pm 90^\circ$, $\pm 210^\circ$, $\pm 240^\circ$ 中随机选择,缩放比例固定为 1,边界填充策略与平移一致;加噪主要为高斯噪声,方差值限定在 0.001 至 0.003 区间。所有增强过程的参数均采用动态随机化机制,避免固定模式对数据集的影响。通过上述变换,图像数量从原始数据集扩充至 9000 张,大幅度提升了数据集的多样性和代表性。

1.2.5 数据归一化

在深度学习模型的训练过程中,数据归一化是提升模型收敛速度和训练稳定性的关键步骤。通过对图像数据进行归一化处理,使输入数据的均值为零且标准差为单位值,可以显著降低数据分布的差异性,从而为模型训练提供更加稳定和高效的输入条件。为实现这一目标,本研究在数据集制备完成后,计算了深度学习模型训练所需的归一化参数,即均值与标准差。具体而言,数据集的均值为[0.550, 0.565, 0.568],红色、绿色和蓝色通道的标准差分别为[0.082, 0.082, 0.085]。这一步骤可以有效减少数据分布的差异性,进一步优化了模型训练的输入条件。

2 数据样本描述

2.1 数据层级结构

主文件夹：data 为结构化的完整数据集，包含 6 类典型道路场景的图像数据，分别为干燥路面（dry road）、积雪路面（fully snowy road）、结冰路面（icy road）、风吹雪路面（snow-blowing road）、融雪路面（snow-melting road）及潮湿路面（wet road）。每类场景数据均由原始图像和增强图像构成。原始图像编号规则：干燥路面 000.jpg-897.jpg、积雪路面 000.jpg-709.jpg、结冰路面 000.jpg-274.jpg、风吹雪路面 000.jpg-091.jpg、融雪路面 000.jpg-335.jpg、潮湿路面 000.jpg-401.jpg。增强图像编号规则：干燥路面 0898.jpg-1499.jpg、积雪路面 0710.jpg-1499.jpg、结冰路面 0275.jpg-1499.jpg、风吹雪路面 0092.jpg-1499.jpg、融雪路面 0336.jpg-1499.jpg、潮湿路面 0402.jpg-1499.jpg。每类场景图像总计 1500 张，全局数据集规模为 9000 张。

预划分子集：train、val、test 为标准化分割的训练集、验证集与测试集。使用提供的数据集分割脚本（split_data.py）对原始数据集进行划分，采用固定随机种子（seed=2024）确保实验可重复性。各子集均包含与 data 相同的 6 类场景，训练集每类样本量为 900 张（编号 000.jpg-899.jpg），验证集和测试集每类样本量均为 300（编号 000.jpg-299.jpg）。训练集、验证集与测试集的总样本量分别为 5400 张、1800 张及 1800 张（按 6:2:2 比例划分）。

2.2 标签索引规范

类别标签通过 class_indices.json 文件实现数字化映射，该文件为图像分类任务提供标准化标签索引，支持跨平台数据加载与模型训练，其键值对定义如表 2 所示。

表 2 类别索引

Table 2 Class indices

索引	道路场景类别
0	干燥路面
1	积雪路面
2	结冰路面
3	风吹雪路面
4	融雪路面
5	潮湿路面

2.3 数据特性与适用场景

本数据集基本涵盖高寒地区极端天气条件下的道路通行环境特征。针对风吹雪等低频灾害类型，采用数据增强策略（将样本数量扩充至每类 1500 张），有效缓解了自然发生频率差异所导致的类别不均衡问题，从而提升模型在不同灾害场景下的泛化能力。此外，数据集采用标准化格式，并提供预定义类别的 JSON 标签，可直接用于计算机视觉领域的经典深度神经网络训练，如 ResNet、DenseNet 等模型。

本数据集适用于计算机视觉领域的多种研究任务，包括不同天气条件下的路面状况识别、基于深度学习的智能交通监测及自动驾驶环境感知等。其标准化的数据处理流程有助于不同模型的训练、验证和测试，进一步推动智能交通管控及冰雪灾害预警系统的优化。数据集的具体情况如图 4 所示。

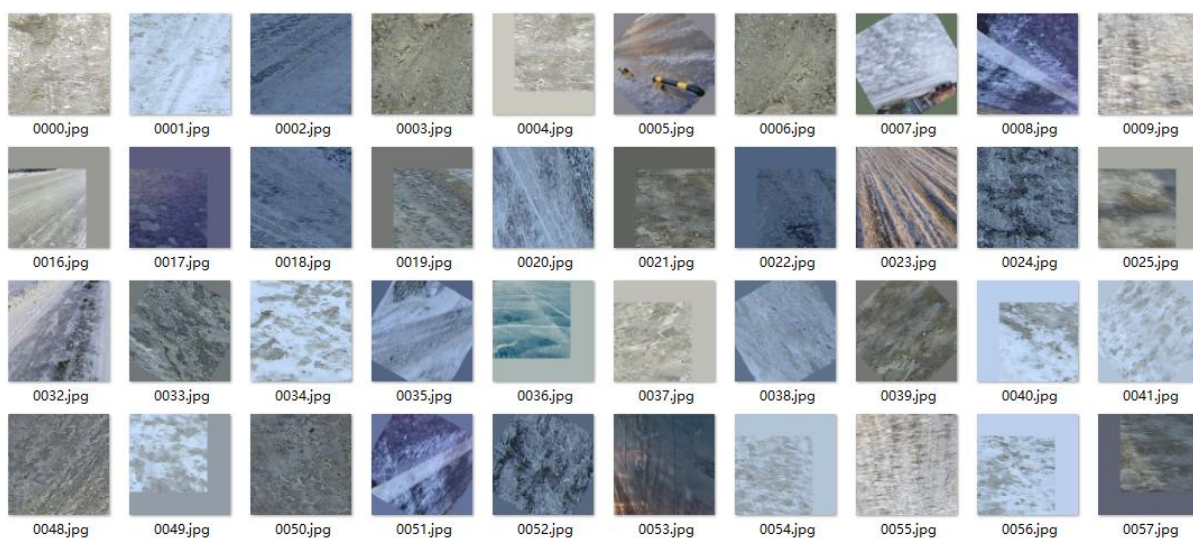


图 4 数据集示例（道路结冰为例）

Figure 4 Examples of the dataset (taking road icing as an example)

3 数据质量控制和评估

3.1 极端噪声剔除

在数据集构建过程中，部分图像因特征高度相似而难以准确分类，这一问题在实验过程中极易导致过拟合现象。为确保数据集的精确性与可靠性，对初始图像数据进行了严格筛选。首先，剔除那些在视觉感知上存在显著模糊性、难以辨识的图像样本，例如图 5a 至图 5c 所示的干燥路面与潮湿路面、积雪路面的混淆案例，以及图 5d 中因大面积条状阴影影响而难以区分融雪路面与轻度风吹雪路面。这类极端案例在工程实践中较为罕见，不仅无助于提升模型的实用性，反而增加了不必要的噪声。其次，删除因光照异常或后期彩色变换导致特征模糊的图像，如图 5e、图 5f，此类图像不仅无法为模型训练提供有效信息，反而会影响模型的性能。移除视觉特征不明确、分类边界模糊的图像，使数据集更贴近实际应用场景，为模型的工程化部署奠定基础。

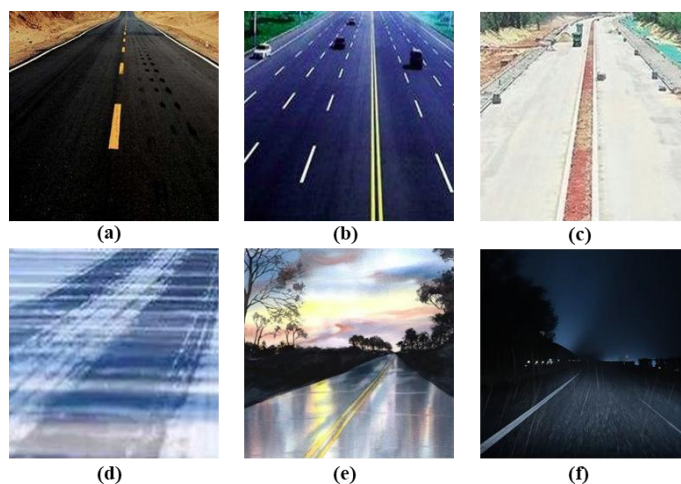


图 5 极端噪声数据示例

Figure 5 Examples of extreme noise data

3.2 深度学习模型应用评估

在数据集构建过程中,本研究通过多轮迭代更新与优化,大幅度提高了数据质量和代表性,使其能够更准确地反映实际应用场景的复杂性和多样性。每轮迭代均采用系统性的实验设计与方法探索,涵盖数据清洗、特征工程、样本平衡及异常值处理等关键环节,以确保数据的完整性与有效性。为验证数据集的适用性,针对十二种经典神经网络模型开展了系统性的对比分析,包括 VGG、ResNet、Inception 系列等主流神经网络模型,具体信息见表 3。

图 6 展示了深度学习模型的实验结果,表明优化后的数据集在所有测试模型中均展现出稳定的收敛轨迹,从算法层面进一步证实了本数据集的有效性与可靠性。实验结果不仅验证了数据集的实用价值,也为相关领域的研究与工程应用提供了重要的参考依据。

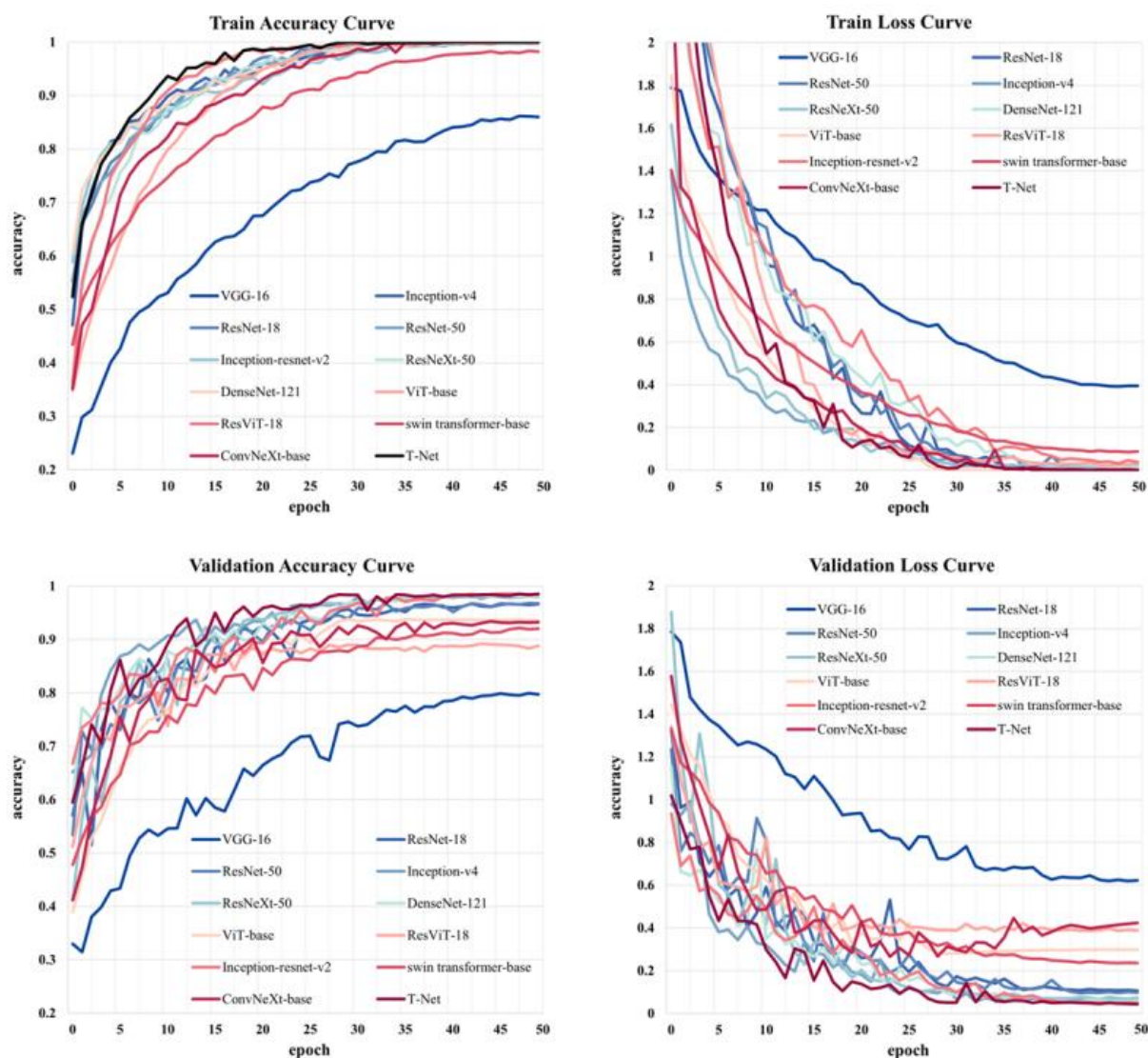


图 6 深度学习拟合曲线

Figure 6 Deep learning fitting curves

表 3 十二种神经网络结构特征

Table 3 Structural features of twelve neural networks

神经网络模型	关键特征/设计方法
VGG-16 ^[6]	顺序架构
ResNet-18 ^[7]	残差连接的顺序架构
ResNet-50 ^[7]	残差连接的顺序架构
Inception-V4 ^[8]	多分支架构
Inception-ResNet-V2 ^[8]	残差连接的多分支架构
ResNeXt-50 ^[9]	残差连接的多分支架构
DenseNet-121 ^[10]	稠密连接的顺序架构
ViT-base ^[11]	Transformer 架构
ResViT-18 ^[11]	残差连接与 Transformer 混合架构
Swin transformer-base ^[12]	Transformer 架构
ConvNeXt-base ^[13]	基于 Swin Transformer 结构设计的残差连接顺序架构
T-Net ^[14]	多分支与 Transformer 混合架构

4 数据价值

本研究构建的面向 AI-Ready 的 6 类路面状态标准化图像数据集，具有以下数据价值：

（1）填补数据稀缺性空白：当前国际公开道路表面气象条件标准化数据集，尤其是极端天气下的路面状态图像资源较为匮乏。本数据集通过系统采集和标准化处理，可以为相关研究提供数据支持。

（2）提升深度学习模型可靠性与场景覆盖能力：深度学习模型的性能高度依赖于训练数据的质量和多样性，尤其在极端天气条件下，模型的识别和预测能力更需依赖高可靠性的数据基础。本数据集融合了高速公路摄像头、移动镜头及在线资源，形成了覆盖六种路面状态数据集，进一步增强了场景覆盖能力，为模型训练提供了全面且高质量的数据支持。

（3）标准化处理提升模型泛化能力：针对光照干扰这一核心问题，本研究基于二维伽马函数的自适应校正算法对图像亮度进行了调整，消除过曝、阴影等噪声对特征提取的影响。配合标准化处理形成的规范数据格式，使模型能够快速拟合。

5 数据使用方法和建议

本数据集中所有图像数据均以.jpg 格式存储，并提供了格式转换脚本，便于用户在深度学习训练代码中直接导入和使用。需要注意的是，不同神经网络模型对输入图像的尺寸要求各异。例如 InceptionV4 等模型的输入尺寸为 299×299 像素，因此在训练过程中需确保图像尺寸的适当缩放。此外，为满足不同用户的需求，本研究还提供了数据集分割脚本以及均值和标准差计算脚本。用户可根据自身需求对数据集进行切分，并分别计算训练集、验证集和测试集的均值和标准差，有助于用户在模型训练过程中实现数据归一化，从而提升模型的收敛效率与性能表现。

数据作者分工职责

刘景琦（1997—），男，甘肃平凉人，博士，主要研究方向为数据工程防灾减灾。主要承担工作：数据采集和论文撰写。

张耀南（1966—），男，甘肃天水人，研究员，博士生导师，主要研究方向为地学数据工程及数据工程防灾减灾、基于高性能计算环境的地学模型模拟、遥感图像处理及多源数据融合。主要承担工作：论文修改。

康建芳（1981—），女，甘肃天水人，高级工程师，主要研究方向为数据工程防灾减灾。主要承担工作：数据集标准化处理。

刘杰（1986—），男，甘肃酒泉人，博士，博士生导师，主要研究方向为公路工程防灾减灾技术。主要研究方向为人工智能。主要承担工作：数据审核。

王斌（1997—），男，甘肃武威人，博士，主要研究方向为公路工程防灾减灾技术。主要承担工作：数据汇总。

参考文献

- [1] 吴彦, 陈春艳, 路光辉. 沿天山高速公路冰雪灾害分析及其对交通安全的影响[J]. 沙漠与绿洲气象, 2013, 7(5): 66-70. DOI:10.3969/j.issn.1002-0799.2013.05.013. [WU Y, CHEN C Y, LU G H. Ice and snow disasters along the Tianshan expressway and its impact on traffic safety[J]. Desert and Oasis Meteorology, 2013, 7(5): 66-70. DOI:10.3969/j.issn.1002-0799.2013.05.013.]
- [2] 魏学利, 陈宝成, 李宾, 等. 地质灾害对旅游公路高质量发展的影响及适应性分析——以新疆独库公路为例 [J]. 中外公路, 2020, 40(S2): 279 - 284. DOI:10.14048/j.issn.1671-2579.2020.S2.057. [WEI X L, CHEN B C, LI B, et al. Impact of Geological Hazards on High-Quality Development of Tourist Highways and Adaptive Analysis: A Case Study of the Duku Highway in Xinjiang [J]. Journal of China & Foreign Highway, 2020, 40(S2): 279 - 284. DOI:10.14048/j.issn.1671-2579.2020.S2.057.]
- [3] 时环生. 独库公路独山子至那拉提段地质灾害分析与防治[J]. 国防交通工程与技术, 2008, (3): 53-56. DOI:10.13219/j.gjgyat.2008.03.012. [SHI H S. Analysis Prevention and Treatment of the Geological Hazards to the Dushanz i Lalati Section of the Duku Highway [J]. Traffic Engineering and Technology for National Defence, 2008, (3): 53-56. DOI:10.13219/j.gjgyat.2008.03.012.]
- [4] 沈永平, 苏宏超, 王国亚, 等. 新疆冰川、积雪对气候变化的响应(II): 灾害效应[J]. 冰川冻土, 2013, 35(6): 1355-1370. [SHEN Y P, SU H C, WANG G Y, et al. The Responses of Glaciers and Snow Cover to Climate Change in Xinjiang (II) : Hazards Effects [J]. Journal of Glaciology and Geocryology, 2013, 35(6): 1355-1370.]
- [5] 刘志成, 王殿伟, 刘颖, 等. 基于二维伽马函数的光照不均匀图像自适应校正算法[J]. 北京理工大学学报, 2016, 36(2): 191-196, 214. DOI:10.15918/j.tbit1001-0645.2016.02.016. [LIU Z C, WANG D W, LIU Y, et al. Adaptive Adjustment Algorithm for Non-Uniform Illumination Images Based on 2D Gamma Function [J]. Transactions of Beijing Institute of Technology, 2016, 36(2): 191-196, 214. DOI:10.15918/j.tbit1001-0645.2016.02.016.]

- [6] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. 2014: arXiv: 1409.1556. <https://arxiv.org/abs/1409.1556>.
- [7] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. IEEE, 2016: 770 – 778. DOI:10.1109/CVPR.2016.90.
- [8] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. IEEE, 2016: 2818-2826.. DOI:10.1109/cvpr.2016.308.
- [9] XIE S N, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. IEEE, 2017: 5987 – 5995. DOI:10.1109/CVPR.2017.634.
- [10] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. IEEE, 2017: 2261 – 2269. DOI:10.1109/CVPR.2017.243.
- [11] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale [C]. In Proceedings of the 9th International Conference on Learning Representations, Vienna, Austria, May 04 2021.
- [12] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada. IEEE, 2022: 9992 – 10002. DOI:10.1109/ICCV48922.2021.00986..
- [13] LIU Z, MAO H Z, WU C Y, et al. A convnet for the 2020s [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA. IEEE, 2022: 11966 – 11976. DOI:10.1109/CVPR52688.2022.01167.
- [14] LIU J Q, ZHANG Y N, LIU J, et al. Automated recognition of snow-covered and icy road surfaces based on t-net of mount Tianshan [J]. Remote Sensing, 2024, 16(19): 3727. DOI:10.3390/rs16193727.

论文引用格式

刘景琦, 张耀南, 康建芳, 等. 面向 AI-Ready 的路面状态标准化图像数据集[J/OL]. 中国科学数据, 2025, 10(4). (2025-12-17). DOI: 10.11922/11-6035.ncdc.2025.0043.zh.

数据引用格式

刘景琦. 面向 AI-Ready 的路面状态标准化图像数据集 (2023–2024 年) [DS/OL]. 国家冰川冻土沙漠科学数据中心(<http://www.ncdc.ac.cn>), 2024. (2024-12-17). DOI: 10.12072/ncdc.dprsc.db6686.2024.

A dataset of AI-Ready standardized images for road surface conditions

LIU Jingqi^{1,2,3}, ZHANG Yaonan^{1,2*}, KANG Jianfang^{1,2}, LIU Jie^{4,5}, WANG Bin^{4,5}

1. Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730030, P.R. China
2. National Cryosphere Desert Data Center, Lanzhou 730030, P.R. China
3. University of Chinese Academy of Sciences, Beijing 100049, P.R. China
4. Xinjiang Transport Planning Survey and Design Institute Co. Ltd, Urumqi 830006, P.R. China
5. Xinjiang Key Laboratory for Safety and Health of Transportation Infrastructure in Alpine and High-altitude Mountainous Areas, Urumqi 830006, P.R. China

*Email: yaonan@lzb.ac.cn (ZHANG Yaonan)

Abstract: In response to the global scarcity of standardized datasets characterizing road weather conditions (RSCs), especially the lack of RSC images in extreme weather, this study develops an AI-ready dataset covering six categories of RSCs, designed to provide a high-quality and reliable training foundation for deep learning models. Based on statistical patterns of traffic impacts under extreme weather, the dataset categorizes snow and ice hazards into three subtypes: road ice formation, heavy snowfall, and snow blowing. Through systematic integration of highway surveillance footage, mobile device recordings, and open-source web data, we developed a standardized image repository covering six typical RSCs: dry road, fully snowy road, icy road, snow-blowing road, snow-melting road, and wet road. To mitigate subset correlation interference caused by data augmentation, we implemented a "split-then-augment" strategy. The original data was independently partitioned into training (60%), validation (20%), and test sets (20%), followed by separate augmentation operations for each subset, including random flipping, rotation, translation, and Gaussian noise addition. This methodology ultimately expanded the dataset to 9,000 images. The dataset provides preprocessed parameters optimized for deep learning training, with RGB channel means and standard deviations measured at [0.550, 0.565, 0.568] and [0.082, 0.082, 0.085] respectively. Zero-mean unit-variance normalization has been demonstrated to significantly enhance model training efficiency. This comprehensive dataset systematically covers multiphase characteristics of snow-ice affected road surfaces, establishing a standardized benchmark for development and evaluation of road state recognition algorithms in intelligent transportation systems.

Keywords: AI-Ready; image recognition; deep learning; road surface condition

Dataset Profile

Title	A dataset of AI-Ready standardized images for road surface conditions (2023–2024)
Data authors	LIU Jingqi
Data corresponding author	LIU Jingqi (liujingqi@nieer.ac.cn)

Time range	October 2023 to October 2024
Data volume	335 MB
Data format	*. jpg
Data service system	http://dx.doi.org/10.12072/ncdc.dprsc.db6686.2024 http://scidb.cn/doi/10.12072/ncdc.dprsc.db6686.2024
Sources of funding	The National Key R&D Program of China (2022YFF0711704); the 2020 Xinjiang Transportation Industry Science and Technology Project (2022-ZD-006); the Xinjiang R&D Project (ZKXFWCG2022060004); and the Research Fund of the Xinjiang Transportation Design Institute (KY2022041101).
Dataset composition	The dataset consists of four data folders, one JSON file, and two Python scripts, including: (1) the data folder, containing six subsets, each storing standardized data for six typical RSCs; (2) the train, val, test folders, storing the partitioned training set, validation set, and test set, respectively, each containing the six corresponding subsets of RSCs; (3) the class_indices.json, providing standardized label indices; (4) split_data.py, a script to divide the dataset into training, validation, and test sets; (5) calculate_mean_variance.py, a script to compute the mean and variance of the dataset.