

基于 D-SegNeXt 的露天矿高分辨率遥感图像道路提取

崔朋志^{1,2}, 孟祥福^{1,3}, 张雯慧²

1 辽宁工程技术大学 测绘学院 中国 辽宁阜新 123000

2 应急管理部 信息研究所 中国 北京 100029

3 辽宁工程技术大学 电子与信息工程学院 中国 辽宁葫芦岛 125015

摘要: 露天矿的高精度三维路网在生产计划、卡车调度和无人驾驶中起着至关重要的作用。与城市路网相比, 露天矿道路边界模糊, 宽度不一。这些道路的颜色与周围的环境融为一体, 并经历着迅速的变化。然而, 准确、高效、及时地获取矿山路网仍然面临着许多挑战。随着无人机技术的发展和普及, 现在可以获得实时的空间图像数据。我们提出了一种混合扩张的多尺度卷积注意单元, 并设计了 D-SegNeXt 网络。这个方法有几个优点。首先, 它通过混合扩张卷积降低了计算复杂度, 扩大了感受野。其次, 残差网络和多尺度卷积可以提取局部、远距离、长距离和窄距离的特征, 从而增强网络捕获长距离依赖关系的能力。此外, 我们还构建了一个真实的露天矿道路数据集, 并对模型进行了测试。实验结果表明, 该模型在图像分类和道路提取方面均优于多个基准网络。数据集和源代码可在 [D-SegNeXt · GitHub](#) 上获得。

索引词: 道路提取、注意力机制、语义分割、遥感图像

I. 介绍

在中国, 随着智能矿山的建设, 露天矿自动化水平正在稳步提高。随着矿山自动化的提高, 需要精确和最新的路线图来满足露天矿生产计划、卡车调度以及载人和无人驾驶的需求。因此, 迫切需要一种高度可访问且提供较短更新周期的路网提取方法。

目前, 露天矿路网的建设可以根据输入数据分为不同的方法。这些方法包括由专业测量人员进行实地测量、GPS 轨迹聚类提取、航空图像分割和三维激光雷达分割。对于前者, 手动更新路线图是一个繁琐、费力、容易出错的过程^[1], 不适合进行大规模的地图更新。对于后者, 虽然矿山中的 GPS 设备比城市地区的消费级 GPS 设备具有更高的精度, 可以快速生成露天矿的主要道路^[1], 但由于车辆分布不均匀无法覆盖整个矿山^[2]。此外, 准确描述坡度、道路宽度和其他重要特征可能具有挑战性^[3]。在后两种方法中, 随着无人机技术的发展和广泛应用, 获得高分辨率的实时空间图像和点云成为可能。基于这些数据, 路网提取的任务转向语义分割, 这引起了众多学者的关注。

近年来, 越来越多的学者将深度学习应用于遥感图像中的道路提取。这些深度学习方法可以分为基于 CNN、基于 U-Net、基于 LinkNet 和基于 Transformer 的方法^[4]。与城市路网相比, 露天矿道路边界模糊, 宽度不一。这些道路的颜色与周围的环境融为一体, 并经历着迅速的变化。Xiao 等人^[5]设计了一种结合残差连接、注意力机制和 U-Net 的网络 RATT-UNet, 用于从露天矿的航空图像中提取道路。Guo 等人^[6-7]提出用 DeepLabv3+C 网络构建露天矿路网三维模型。Gu 等人^[8]介绍了用于露天矿道路提取的 MD-LinkNeSt (Multi Dilation-LinkNeSt) 网络。

尽管在过去的十年中进行了广泛的研究, 但由于其独特的特点, 利用高分辨率航空图像提取露天矿道路的任务仍然是一项具有挑战性的工作。首先, 输入图像的分辨率不断提高, 需要具有更大感受野的神经网络来完成这项任务。第二, 与城市路网相比, 露天矿路网具有上述不同的特征。露天矿的道路通常蜿蜒曲折, 而且更加复杂。因此, 为了应对这些挑战, 我们设计了一个为露天矿路网量身定制的模型, 称为 D-SegNeXt, 它基于 HDMSCA (混合扩张多尺度卷积注意力)。本文的贡献总结如下:

- (i) 我们设计了一个新的注意单元 HDMSCA，它可以捕获局部、长距离、窄距离特征，并融合多尺度的上下文信息。
- (ii) 我们设计了 D-SegNeXt 神经网络，它本质上是一个语义分割网络，用于从高分辨率遥感图像中提取露天矿道路。
- (iii) 我们使用 ImageNet-1k 对模型进行预训练。随后，该模型在小数据集上快速收敛，并取得了良好的效果。

II. 相关工作

A. 语义分割

语义分割是计算机视觉中的一项基本任务，其目的是将给定的图像划分为不同的区域和对象。卷积神经网络（CNNs）和全卷积网络（FCNs）利用分层特征表示在各种分割任务^[9]中取得了巨大成功。近年来，随着 Transformers 在自然语言处理(NLP)领域的显著进步，研究人员开始将其应用到计算机视觉任务中。Transformer 及其变体，如 Segformer^[10]和 Swin^[11]，利用其固有的全局自注意力机制，在语义分割任务中取得了较好的性能。SegNeXt^[9]设计了一个具有注意力机制的类似 CNN 的主干。到目前为止，SegNeXt 在几个常见的数据集上都取得了最好的结果。

B. 扩张卷积和 HDC

扩张卷积通过在连续元素之间插入孔洞来扩展卷积核（输入），如图 1 所示。它是 CNNs 中使用的一种卷积运算，使网络能够在不增加参数数量的情况下具有更大的感受野。这使得 CNNs 在结合更大的上下文的同时能够保持分辨率，从而提高语义分割等任务的性能，其中局部和全局信息都很关键。扩张卷积已成为现代 CNN 架构的关键技术，应用于需要多尺度处理的任务^[13-15]。但也可能会忽略局部较少但有用的信息，甚至导致网格问题^[16]，如图 1（b）所示。

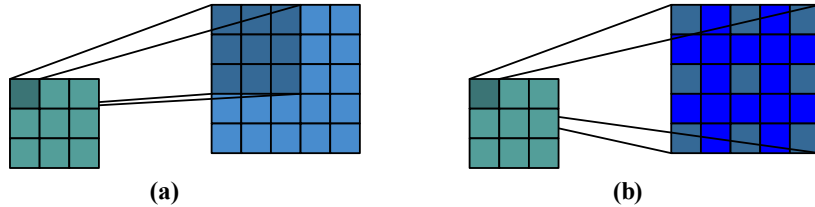


图 1 标准和扩张卷积。用蓝色标记的像素表示网格。

目前提出了一种由不同扩张速率组成的混合扩张卷积方法用于语义分割。在他们的工作中，扩张速率的分配遵循锯齿波状的启发式方法^[16]。根据这一策略， $[1, 2, 3]$ ^[16]， $[1, 2, 4, 8, 16]$ ^[14]和 $[1, 1, 2, 3, 5, 8, 13, 21, 34, 55]$ ^[17]的速率集是很好的组合。

C. 多尺度注意力

在计算机视觉（CV）领域，多尺度推理和注意力机制在众多的视觉任务中取得了巨大的成功^[13, 17-20]。多尺度推理广泛用于密集预测任务，通过利用来自多个尺度的信息来增强模型性能。注意力机制可被视为一个动态选择的过程^[18]，使模型能够专注于显著区域并忽略输入特征中的不相关噪声。

SegNeXt^[9]提出了一种多尺度卷积注意力（MSCA）方法，以提高大核注意力（LKA）。虽然它可以得到很好的分类结果，但它需要大量的计算和较长的时间。为了克服上述问题，本文提出了 HDMSCA 单元，它可以利用少量参数构建通道和空间注意力。

III. 提出的方法

在本节中，我们将详细介绍我们模型的结构，包括新的注意力机制 HDMSCA 和基于 HDMSCA 的 DSegNeXt 的网络架构。

A. 混合扩张多尺度卷积注意力

对于我们编码器中的构建块，我们采用与 SegNeXt^[9]类似的结构。然而，不同之处是

我们不仅使用多尺度卷积机制，而且还设计了一个称为 HDMSCA 的新单元。如图 2 所示，HDMSCA 由两个组件组成：一组核尺寸为 5，扩张速率为 1, 2, 5, 9 的卷积以及一个 1×1 的卷积来捕捉不同通道之间的关系。这些卷积的核尺寸为 5×5 ，在水平方向上串联，在垂直方向上形成三个分支。这种设计允许捕获到局部、远距离、长距离、窄距离和多尺度上下文，同时还可以通过小型核卷积加速计算。当前卷积层的感受野可以表示为：

$$RF_n = RF_{n-1} + ((K_n - 1) * D_n + 1) \prod_{i=1}^{n-1} S_i \quad (1)$$

其中 RF_n 是卷积序列中第 n 次混合扩张卷积的感受野， K_n 是核尺寸（高度或宽度，假设这里它们相等）， D_n 是扩张速率， S_i 是步幅。

在数学上，我们的 HDMSCA 可以表示为：

$$DW_{Conv_n} = DW_{Conv}(DW_{Conv_{n-1}}) \quad (2)$$

$$Attention = Conv_{1 \times 1} \left(\sum_{i=1}^4 DW_{Conv_i} \right) \quad (3)$$

$$Output = Attention \otimes F \quad (4)$$

其中 DW_{Conv} 是一个递归函数， $DW_{Conv_n}, n \in (1, 2, 3, 4)$ 是图 2 中的第 n 个分支， $DW_{Conv_0} = F$ ， F 是输入特征， $Attention$ 是注意力图， $Output$ 是输出， \otimes 表示逐元素乘法， $Conv_{1 \times 1}$ 是指一个 1×1 的卷积。

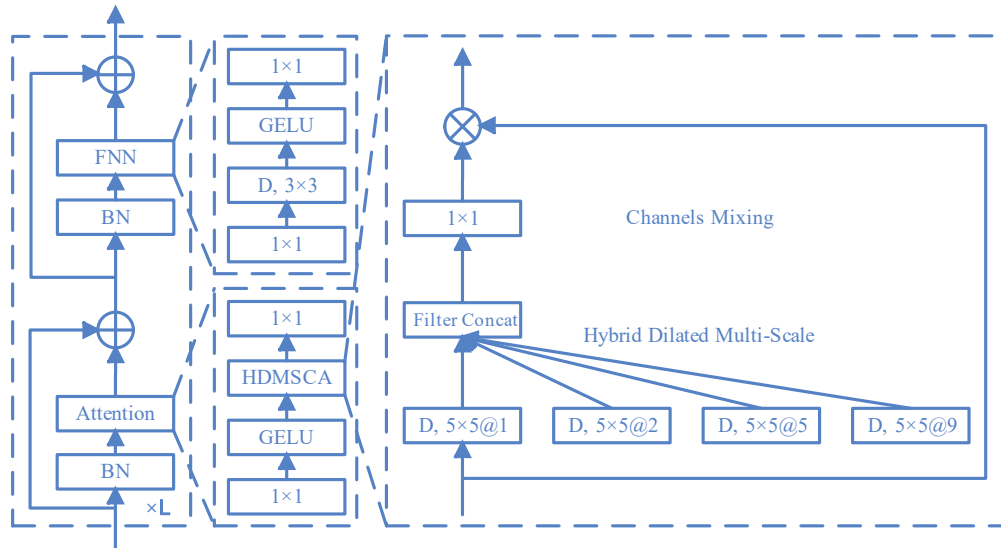


图 2 D-SegNeXt 的一个阶段和所提出的 HDMSCA 模块的细节。

这里 $D, k \times k @ r$ 表示一个核尺寸为 $k \times k$ ，扩张速率为 r 的深度卷积。

我们选择核尺寸为 1、2、5 和 9 作为混合扩张卷积有两个原因。首先，我们希望避免因扩张速率过大而导致的网格状伪影现象。其次，通过实验我们发现这些卷积核尺寸在我们的应用中能够在感受野大小和计算效率之间取得良好的平衡。较小的卷积核（尺寸为 1 和 2）可以在局部范围内聚合信息，而较大的卷积核（尺寸为 5 和 9）则扩大了感受野，覆盖更大的图像上下文。三个卷积层的感受野分别为 13、33 和 69。如表 I 所示，我们的 HDMSCA 具有更大的感受野，而且参数数量相同或更少。此外，为了保留张量的个别信息并灵活地将它们连接起来，我们使用“concat”操作替代了“add”操作。拼接操作允许我们自由地组合不同的张量，而不会丢失它们固有的结构，这在构建复杂的深度神经网络时非常有用。

表 I 不同注意力机制的参数数量。11/15/25 表示有三个分支，
每个分支的感受野大小根据公式 1 计算得到分别为 11、15 和 25。

注意力机制（通道数=256）	参数量（千）	感受野
LKA	85.25	21
MSCA	93.85	11/15/25
HDMSCA	92.42	13/33/69

B. 网络架构

像大多数竞争性的语义模型一样，我们的 D-SegNeXt 网络也含有一个编码器-解码器结构,并且是一个包含四个阶段的分层网络,以降低空间分辨率,如图 3 所示。在编码器中，与 SegNeXt 网络类似，每个阶段都包含一个下采样块和一堆构建块。不同的是，我们用 HDMSCA 取代了 MSCAN。给定一幅 $H \times W \times 3$ 的图像，经过四个阶段，我们将得到原始图像分辨率的 $\{1/4, 1/8, 1/16$ 和 $1/32\}$ 处的多层次特征，当然，每一层次的通道数也会相应增加。在解码器中，我们使用名为 LightHamburger 的分类头来聚合来自 SegNeXt 网络解码器之后最后三个阶段的特征。在我们的实验部分，我们将展示我们的 D-SegNeXt 网络相比于最近的先进模型 D-LinkNet^[15]、DUnet^[25]、Segformer^[10]和 SegNeXt^[9]表现更好。

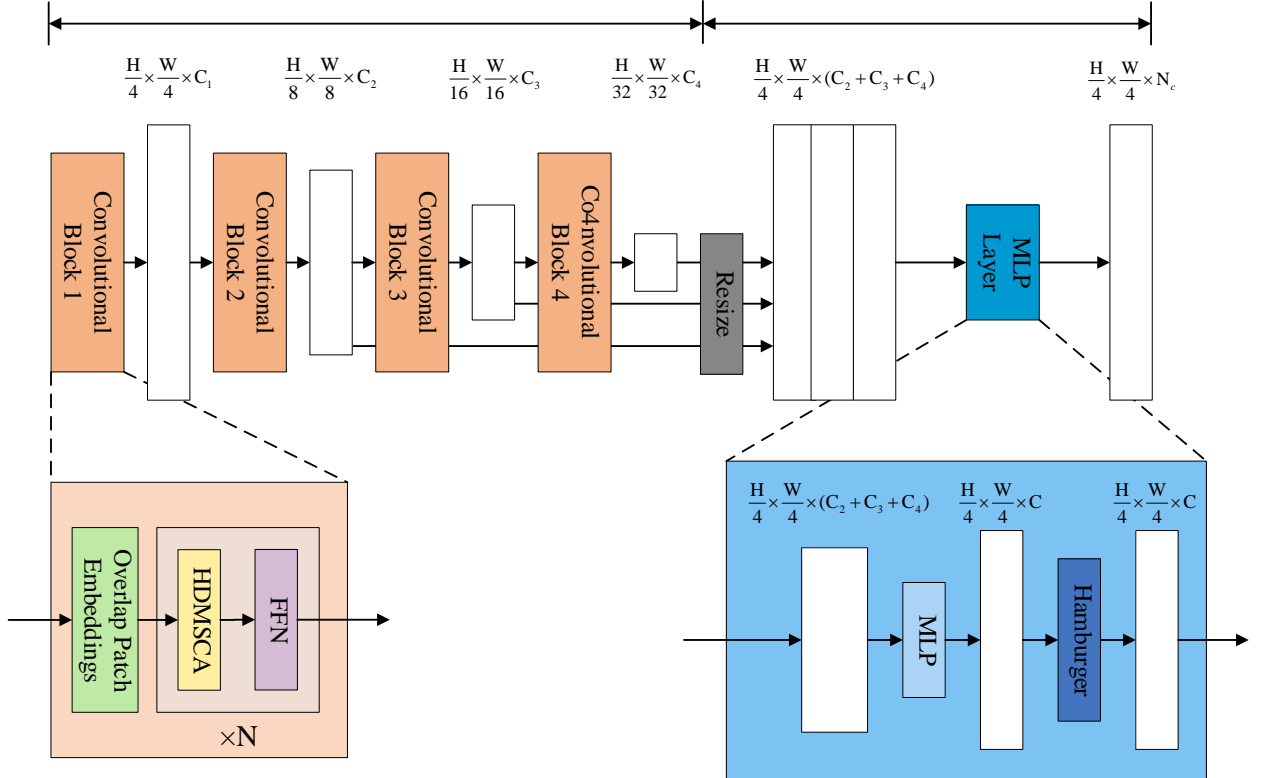


图 3 D-SegNeXt 的网络架构包含两个模块：分层编码器提取特征，
轻量级解码器直接融合这些多层次特征并预测语义分割掩码。

根据 MSCA 和 SegNext 的配置，用 HDMSCA 替代 MSCA，设计了四种编码器模型：HDMSCA-T, HDMSCA-S, HDMSCA-B 和 HDMSCA-L。此外，我们还开发了四个相应的分割模型：D-SegNeXt-T、D-SegNeXt-S、D-SegNeXt-B 和 D-SegNeXt-L。详细的网络设置如表 II 所示。在这张表中，“e. r.”表示前馈网络中的扩展比例，“C”和“L”分别是通道数和构建块的数量。

表 II 提出的 D-SEGNEXT 网络的不同尺寸的详细设置。

阶段	输出大小	前馈网络扩展比例	极小	小	基础	大
1	$\frac{H}{4} \times \frac{W}{4} \times C$	8	C=32 L=3	C=64 L=2	C=64 L=3	C=64 L=3
2	$\frac{H}{8} \times \frac{W}{8} \times C$	8	C=64 L=3	C=128 L=2	C=128 L=3	C=128 L=5
3	$\frac{H}{16} \times \frac{W}{16} \times C$	4	C=160 L=5	C=320 L=4	C=320 L=12	C=320 L=27
4	$\frac{H}{32} \times \frac{W}{32} \times C$	4	C=256 L=2	C=512 L=2	C=512 L=3	C=512 L=3
参数 (百万)			5.3	17.32	34.27	48.77
浮点运算数量 (十亿)			6.82	17.76	37.56	64.56

C. 预训练编码器

迁移学习是计算机视觉中的一种有效方法，特别是在训练图像数量有限的情况下。与从头开始训练整个模型相比，迁移学习可以帮助模型在小数据集上快速收敛，并取得良好的效果。采用迁移学习的策略，可以大大提高计算机视觉任务的训练效率。根据最新的研究[9]、[10]和[12]，我们使用 ImageNet-1k 对我们的模型进行预训练，该数据集包含了来自 1000 个不同类别的 128 万张训练图像和 5 万张验证图像。在我们的实验中，我们发现迁移学习可以提高我们的神经网络的效率，取得比从头开始训练更好的性能。值得一提的是，我们的模型在 ImageNet-1K 数据集上也取得了最先进（SOTA）水平的结果。在表格 III 中展示了不包括 HDMSCA-Large¹的结果。

表 III 与在 ImageNet-1K 验证集上的方法进行比较。

'ACC.'代表 Top-1 准确率。

方法	参数 (百万)	Top-1 准确率 (%)
VAN-Tiny	4.1	75.4
MSCA-Tiny	4.2	75.9
HDMSCA-Tiny	5.3	76.1
VAN-Small	13.9	81.1
MSCA-Small	14.0	81.2
HDMSCA-Small	17.3	81.3
VAN-Base	26.6	82.8
MSCA-Base	26.8	83.0
HDMSCA-Base	34.2	82.8

IV. 实验结果与讨论

A. 数据集

高分辨率和高质量的数据集对于验证模型性能至关重要。随着无人机倾斜摄影技术在露天矿的普及，无人机拍摄的航空图像相对于卫星遥感具有几个优势。这些优势包括更高的精度、更高的分辨率、更低的成本、更短的采集周期和更少的天气限制。在本文中，我们提供了一个包含露天矿道路图像的大型数据集。该航空图像是由一架无人机在中国内蒙古的一个矿区拍摄的。该航空图像的尺寸为 81848×45383 像素，空间分辨率为 0.11 米。我们将该航空图像切割成了 1313 个小图像，每个小图像的尺寸为 1024×1024 像素。最后，

¹ 如果训练 HDMSCA-Large 模型将花费超过 155 个小时，对我们来说太长了。因此，我们决定不对模型进行预训练。

将数据集正式划分为三个分区。样本的 60%用于训练，20%用于验证，其余 20%用于测试。

B. 评价指标

在道路分割任务中，有一些广泛使用的评价指标，包括 precision、F1-score、Accuracy、recall 和 IoU。根据先前的工作^[7]，我们采用交并比（IoU）和准确率（Accuracy）来评估道路提取的性能。

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (5)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (6)$$

其中 TP 和 FP 分别表示正确预测为正类别和错误预测为正类别的像素数量。类似地，TN 和 FN 分别表示正确预测为负类别和错误预测为负类别的像素数量。

C. 实现细节

我们的模型是用 VAN-Classification^[12]和 mmsegmentation^[21]来实现分类和分割的。我们的模型编码器是在 ImageNet-1K 数据集上进行预训练的，使用 8 个 RTX 3090 GPUs, 64GB 内存，遵循与 SegNeXt 相同的设置。我们采用 Top-1 准确率作为分类的评价指标。

在相同超参数设置条件下，在 2 块 120GB RTX 5000 GPUs 上进行了道路分割实验。用 BCE（二元交叉熵）作为损失函数。我们采用了多种数据增强技术，包括随机尺寸调整比例为 0.5 到 2.0、随机水平翻转和随机裁剪到 1024×1024 大小。为了评估提出的 D-SegNeXt 网络在露天矿道路提取中的性能，在相同的数据集和超参数上使用 LinkNet^[22]、D-LinkNet^[15]、UNet^[23]、NLLinknet^[24]和 DUnet^[25]、SegFormer^[10]和 SegNeXt^[9]网络进行测试。除了 DUnet 和 D-SegNeXt_base 之外，它们的批处理大小为 2，其他模型的批处理大小为 4。我们使用了 AdamW 优化器，其中 β 值在 0.9 到 0.999 之间，权重衰减为 0.01。初始学习率为 0.00006，衰减策略为多项式学习率策略，幂值为 1.0。我们将模型迭代训练了 20000 次。对于批量较小的模型，我们相应地增加迭代的次数，以确保模型使用相似数量的图像进行训练。

D. 结果

表 IV 不同模型验证集的结果

模型	参数（百万）	交并比（%）	准确率（%）
LinkNet	21.64	71.45	85.19
D-LinkNet	31.1	71.75	86.1
Unet	39.5	59.29	75.63
DUnet	19.2	72.53	83.35
NL-Linknet	21.82	71.69	85.69
Segformer-b0	3.72	72.24	83.47
SegNeXt-T	4.23	71.51	86.54
D-SegNeXt-T w/o Pretrain	5.3	64.14	76.94
D-SegNeXt-T w Pretrain	5.3	73.32	86.66
SegFormer-b1	13.68	74.02	84.29
SegNeXt-S	13.89	74.24	87.16
D-SegNeXt-S w/o Pretrain	17.32	68.36	80.58
D-SegNeXt-S w Pretrain	17.32	74.61	87.39
Segformer-b2	24.72	75.42	86.06
SegNeXt-B	27.56	75.41	85.98
D-SegNeXt-B w/o Pretrain	34.27	72.12	85.13
D-SegNeXt-B w Pretrain	34.27	75.59	87.31

在相同的超参数设置下，这些模型的交并比（IoU）和准确率（Accuracy）的结果总结在表 IV 中，最佳性能用粗体标注。从结果中可以看出，Segformer-b0、SegNeXt-T 和 D-SegNeXt-T 模型在交并比（IoU）和准确率（Accuracy）方面实现了与 D-LinkNet、DUnet 和 NL-LinkNet 模型几乎相同的性能，但参数量仅为其四分之一甚至更少。随着通道数和构建块数量的增加，模型变得越来越大，从而具有更强的分割能力。在同一水平上，D-SegNeXt 的表现优于 SegNeXt。图 4 展示了不同模型预测的道路情况。线路 1、2、3、4 和 8 是在 D-LinkNet、DUnet 和 Segformer_b2 中显示边缘不平滑问题的示例。线路 2、3 和 6 是在 D-LinkNet、DUnet、Segformer_b2 和 SegNeXt 中显示道路连接性问题的示例。对于 DUnet、Segformer_b2 和 SegNeXt_base 模型，线路 2、5 和 8 是显示空洞问题的示例。线路 1、3 和 6 是展示 D-SegNeXt_base 模型识别窄道路能力的示例。线路 4 和 7 是 SegNeXt_base 模型在低误识别率和漏报率方面的示例。



图 4 不同模型预测结果的比较

V. 结论

在本文中，我们提出了一个高质量的露天矿道路数据集，并引入了 HDMSCA 单元，以及基于 HDMSCA 的 D-SegNeXt 网络，用于从无人机卫星图像中提取露天矿道路信息。通过扩大感受野并融合多尺度特征，HDMSCA 操作可以有效地捕捉局部、长距离和窄距离的特征，并融合多尺度的上下文信息。基于 HDMSCA 的编码器在 ImageNet-1k 上取得了最先进的性能，其在 Top-1 分类中高达 82.8% 的准确率。我们的 D-SegNeXt 模型在露天矿

道路数据集上取得了 75.59% 的 IOU 分数，超过了其他任何关于露天矿道路数据集的出版物。我们计划进一步优化网络架构，减少参数，并在额外的分割数据集上进行测试，以增强模型的泛化能力。

参考文献

- [1] Seiler, Konstantin M, “Haul Road Mapping from GPS Traces,” 2022, arXiv:2206.13936.
- [2] X.-Y. Sun, F.-L. Tian, H. Zhang, and Z. Li, “Automatic extraction of road network in open-pit mine based on GPS data,” *J. China Coal Soc.*, vol. 42, no. 11, pp. 3059–3064, 2017.
- [3] F.-L. Tian, X.-Y. Sun, X.-W. Gu, F.-Y. Xin, and L. Ma, “Updating road information in open-pit mines using truck trajectories,” *Math. Probl. Eng.*, pp. 1–8, Jun, 2019.
- [4] Chen. Z, Deng. L, Luo. Y, et al, “Road extraction in remote sensing data: A survey,” *Int J Appl Earth Obs.*, Aug, 2022, 112: 102833.
- [5] D. Xiao, L. Yin and Y. Fu, “Open-pit mine road extraction from highresolution remote sensing images using RATT-UNet,” *IEEE Geosci Remote S.*, vol. 22, no.19, pp. 1-5, Mar, 2022.
- [6] M.-L. Guo, S.-L. Ruan, C.-W. Lu, and Q.-H. Gu, “Road Extraction Method of Open-Pit Mine Based on Improved DeepLabv3+ Network,” *Laser & Optoelectronics Progress.*, vol. 58, no 22, pp, 2228005-1 – 2228005-10, Nov, 2022.
- [7] Q.-H. Gu, B.-Q. Xue, J.-S. Song, X.-X. Li, Q. Wang, “A high-precision road network construction method based on deep learning for unmanned vehicle in open pit,” *Minining, Metallurgy & Exploration.*, vol. 39, no 2, pp, 397-411, Apr, 2022.
- [8] Q.-H. Gu, B.-Q. Xue, S.-L. Ruan and X.-X. Li, “A road extraction method for intelligent dispatching based on MD-LinkNet network in open-pit mine,” *Int. J. Min., Reclam. Environ.*, vol. 35, no. 9, pp. 656669, Oct. 2021.
- [9] M.-H. Guo, C.-Z. Lu, Q. Hou, et al, “SegNeXt: Rethinking convolutional attention design for semantic segmentation,” *NIPS.*, vol. 35, 1140-1156, Dec, 2022.
- [10] E.-Z. Xie, W.-H. Wang, Z.-D. Yu, Anandkumar. A, Alvarez. J.M, P. Luo, “Segformer: Simple and efficient design for semantic segmentation with transformers,” *Adv. Neural Inform. Process. Syst.*, vol. 34, pp. 12077-12090, 2021.
- [11] Z. Liu, Y.-T. Lin, Y. Cao, H. Hu, Y.-X. Wei, Z. Zhang, S. Lin, B.-N. Guo, “Swin transformer: Hierarchical vision transformer using shifted windows,” In: *Int. Conf. Comput. Vis.*, pp.10012-10022, 2021.
- [12] M.-H. Guo, C.-Z. Lu, Z.-N. Liu, M.-M. Cheng, S.-M. Hu, “Visual attention network,” *Comput Vis Media.*, vol. 9, no. 4, pp. 733-752, 2022.
- [13] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” *IEEE Trans Patt Ana Mach Inte.*, vol. 40, no. 4, pp. 834-848, 2017.
- [14] F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” 2015, arXiv:1511.07122.
- [15] L.-C. Zhou, C. Zhang, M. Wu, “D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction,” In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp, 182–186.

- [16] P.-Q. Wang, P.-F. Chen, Y. Yuan, D. Liu, Z.-H. Huang, X.-D Hou, and Garrison Cottrell, "Understanding convolution for semantic segmentation," In WACV, Lake Tahoe, NV, USA, 2018, pp. 1451-1460.
- [17] M. Anthimopoulos, S. Christodoulidis, L. Ebner, T. Geiser, A. Christe, S.Mougiakakou, "Semantic Segmentation of Pathological Lung with Dilated Fully Convolutional Networks," IEEE J. Biomed Health., vol. 23, no. 2, pp. 714-722, 2018.
- [18] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, Ralph R Martin, M.-M. Cheng, and S.-M. Hu, "Attention mechanisms in computer vision: A survey," Comput Vis Media., vol. 8, no. 3, pp. 311-368, 2022.
- [19] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y Zhang, M.-H Yang, Torr. P, " Res2net: A new multi-scale backbone architecture," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 2, pp. 652–662, 2021.
- [20] T.-Y. Lin, Piotr Dollar, Ross Girshick, K.-M. He, Bharath Hariharan, and Serge Belongie, "Feature pyramid networks for object detection," in Proc. IEEE-CVPR, Honolulu, HI, USA, 2017, pp. 2117–2125.
- [21] MMSegmentation. Contributors, 2020, "MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark (version 0.30.0).".
- [22] Abhishek Chaurasia and Eugenio Culurciello., "Linknet: Exploiting encoder representations for efficient semantic segmentation," in Proc. VCIP, St. Petersburg, Florida, USA, 2017, pp. 1-4.
- [23] Ronneberger, O., Fischer, P., & Brox, "Unet: Convolutional networks for biomedical image segmentation," in Proc. MICCAI, Munich, Germany, 2015, pp. 234–241.
- [24] Y. Wang, J. Seo, T. Jeon, "NL-LinkNet: Toward lighter but more accurate road extraction with nonlocal operations," IEEE Geosci. Remote Sens., vol. 19, pp. 1-5, 2021.
- [25] S.-Y. Piao, and J.-M. Liu, "Accuracy improvement of UNet based on dilated convolution," J. Phys.: Conf. Ser., vol. 1345, no. 5, 2019.