使用深度编码器-解码器模型进行溢油分割

Abhishek RamanathapuraSatyanarayana [a] 和 Maruf A. Dhali [b]人工智能系，*Bernoulli* 研究所，格罗宁根大学，*9747 AG Groningen*，荷兰{*a.ramanathapura.satyanarayana@student.rug.nl*, *m.a.dhali@rug.nl*

arXiv:2305.01386v2 [cs.CV] 9 Apr 2025

关键词: 溢油、语义分割、神经网络、深度学习。

摘要: 原油是世界经济和交通运输领域不可或缺的组成部分。随着原油因其广泛应用而需求不断增长，意外溢油虽令人遗憾，却难以避免。尽管溢油清理十分困难，但首要且最重要的挑战是对其进行检测。在本研究中，作者测试了深度编码器-解码器模型的可行性，这些模型能够经过有效训练，远程检测溢油。该研究考察并比较了多个分割模型在高维卫星合成孔径雷达（SAR）图像数据上的结果，为进一步深入研究铺平道路。实验采用了多种模型组合。表现最佳的模型采用 ResNet-50 编码器和 DeepLabV3+ 解码器。与此前的基准模型相比，该模型的平均交并比（IoU）达到 64.868%，并且“溢油”类别的类别交并比提高至 61.549%；此前的基准模型的平均 IoU 为 65.05%，“溢油”类别的类别 IoU 为 53.38%。

1 引言

海洋、 ocean 和沿海地区是环境、海洋生态系统以及人类活动的重要组成部分。水生生态系统是全球相当一部分人口获得经济稳定和维持生计的关键来源。然而，众多人类活动对海洋生态系统构成了重大威胁，其中溢油尤为突出。这些溢油可能源于多种因素，包括原油运输。原油是许多工业和制造业流程不可或缺的组成部分，广泛应用于汽油、柴油、航空燃料、润滑剂、纺织品、油漆、肥料、农药和药品等领域，是推动工业发展与扩张的关键力量。然而，全球运输原油的主要方式是使用油轮进行海运，这不可避免地会导致意外溢油事故。

除意外溢油外，其他原因还可能包括海上平台、钻井平台和油井事故、自然灾害、故意排放以及技术故障。这些事件可能造成灾难性的环境影响、损害人类健康，并带来重大的社会经济后果。此外，溢油还会严重影响野生动物，包括鸟类和海洋哺乳动物，扰乱生态系统并造成长期生态损害（Dunnet 等，1982）。溢油除了影响海洋生态系统，也会影响空气质量（Middlebrook 等，2010）。由于危害严重，溢油引发了公众和媒体的强烈批评，促使政治机构和政府部门采取行动，防止此类事件再次发生（Broekema，2016）。

溢油的清理可能需要很长时间，所需时长会因多种因素而显著不同，包括溢油规模、油品类型、环境条件以及应急响应工作的成效（Shigenaka，2009）。为及时采取围控和清理措施，自动、高效地检测溢油至关重要。一种有效的方法是将遥感与人工智能相结合，并使用专门训练用于识别溢油的监督式机器学习模型。为此，可以利用卫星图像进行遥感。近年来，人们对使用深度卷积神经网络（CNN）处理此类图像数据的兴趣日益浓厚。AlexNet 模型的问世表明，与传统的特征工程技术相比，其性能有了显著提升，尤其是在 ImageNet 目标识别竞赛中。（Deng 等人，2009；Russakovsky 等人，2015；Krizhevsky 等人，2012）。

早期研究使用卷积神经网络模型，利用合成孔径雷达（SAR）卫星图像检测溢油——

[a] https://orcid.org/0009-0003-1248-0988 [b] https://orcid.org/0000-0002-7548-3858

影像。Marios Krestenitis 等人在其研究中进行了一些改进，并训练了部分卷积神经网络模型来检测溢油（Krestenitis 等，2019）。他们将原始高维图像划分为较小的图像块，并以此训练各种模型。这一过程可能消耗大量内存，需要大量计算资源，并导致能耗增加。随着技术进步，直接在更高维度的图像上开发和训练模型，对于实现更好的性能至关重要。这有助于在一定程度上减少处理所需的内存，并降低计算能耗。本研究尝试在维度相对较高的图像上训练各种卷积神经网络模型，并研究其对模型整体性能的影响。

2 相关工作

随着人工神经网络（ANN）的出现，Yann LeCun 等人提出了 LeNet-5，这是首个应用于图像数字识别任务的卷积神经网络（CNN）（LeCun et al., 1989；LeCun et al., 1998）。后来，一个广受欢迎的数据集——MNIST 数据集成为图像数字识别任务的基准（Deng, 2012）。由于多种原因，卷积神经网络在计算机视觉和图像分析的各类任务中的应用曾长期滞后，其中包括计算能力不足以及缺乏大规模标注数据集。近年来，ImageNet 一直在大规模数据集上举办目标识别挑战赛（Deng 等人，2009；Russakovsky 等人, 2015）。该数据集包含 120 万张图像，分属 1000 个类别。2012 年，一个名为 AlexNet 的卷积神经网络模型赢得了 ImageNet 目标识别挑战赛，成绩远超当年的所有其他参赛者（Krizhevsky et al., 2012）。其他参赛者采用的则是非卷积神经网络方法，即将传统的手工设计特征与不同的机器学习技术相结合。

计算机视觉中的一个关键挑战是语义分割，即模型必须学习将图像中的每个像素分类为特定类别。2014 年，Jonathan Long 等人提出了全卷积网络（FCN），这是第一个仅包含卷积层和转置卷积层、不包含任何全连接（或稠密）层的卷积神经网络（Long 等人，2014）。随着 FCN 的出现，研究人员又提出并开发了其他几种最先进的语义分割模型。其中，较为流行的模型包括 UNet、LinkNet、PSPNet 和 DeepLabV3+（Ronneberger 等人，2015；Chaurasia 和 Culurciello，2017；Zhao 等人，2016；Chen 等人，2018）。这些模型有一个共同点，即它们都是编码器-解码器架构的变体。大多数模型都在 COCO 和 Cityscapes 等大型数据集上进行了基准测试（Lin 等人，2014；Cordts 等人，2016）。

Cityscapes 是一个用于语义分割任务基准测试的带标注数据集，包含 5000 张高质量标注图像和 20000 张弱标注图像。Marios Krestenitis 等人使用石油泄漏检测数据集对其中一些模型进行了基准测试。该数据集通过汇集从卫星合成孔径雷达（SAR）数据中提取的图像构建而成（Krestenitis 等人，2019）。该数据集仅包含 1002 张训练图像，数量远少于 Cityscapes 数据集。在他们的研究中，Marios Krestenitis 等人将尺寸为 1250×650 的原始图像切分为多个图像块，并将这些图像块作为模型的输入。针对不同模型，他们采用了不同尺寸的图像块。尽管这些模型的原始实现采用了不同的编码器，Marios Krestenitis 等人还是对模型进行了修改，使大多数解码器使用 ResNet-101 编码器；唯一的例外是 DeepLabV3+ 解码器搭配 MobileNetV2 编码器（He 等人，2015；Krestenitis 等人，2019；Chen 等人，2018；Sandler 等人，2018）。他们的研究发现，MobileNetV2 编码器搭配 DeepLabV3+ 解码器的平均交并比（m-IoU）最高，达到 65.06%。他们还发现，该模型在溢油类别上的类别交并比为 53.38%，排名第二。

3 方法

3.1 数据集

Marios Krestenitis 等人开发了本研究所使用的石油泄漏检测数据集（Krestenitis 等人，2019）。该数据集包含从卫星合成孔径雷达（SAR）数据中提取的图像，呈现溢油及其他相关语义类别，并附有对应的真实值掩码和标签（Krestenitis 等人，2019）。训练集中有 1002 张图像，测试集中有 110 张图像，均附有相应标签。数据集包含 5 个语义类别，分别代表以下类型：海面、溢油、溢油伪影、船舶和陆地。图1展示了训练集中的类别分布。可以观察到

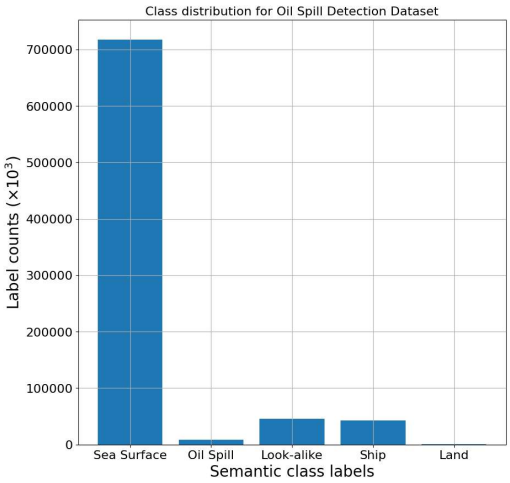


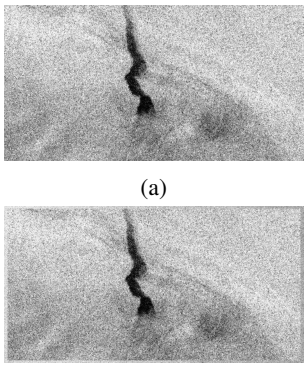
图1：石油泄漏检测数据集中语义类别分布图。

该数据集各语义类别的分布极不均衡。例如，属于“海面”类别的像素数量远多于数据集中其他已标注的语义类别。本研究关注的类别是“溢油”，其在数据集中的出现频率很低。本研究尝试训练经过优化、能够检测该类别的模型。

数据集中的图像尺寸为 1250×650 $W \times H$，其中W 和H 分别表示图像的宽度和高度。在本文的工作中，默认提供的训练数据集（1002张图像）被进一步随机划分为95%（951张图像）和5%（51张图像），分别用作训练集和验证集。这样做是为了使用随机验证集进行5折交叉验证。

3.2 图像预处理与数据增强

在他们的研究中，Marios Krestenitis 等人使用较小的图像块来训练各种模型（Krestenitis 等人，2019）。他们研究中使用的最小和最大图像块尺寸分别为 320×320和 336×336。在本文的研究中，原始图像的维度为 1250×650，通过在图像的四周填充图像块，生成尺寸为 1280×672的图像。训练集中的一个原始样本图像及其对应的填充图像分别如图 2(a) 和 2(b) 所示。因此，与 Marios Krestenitis 等人使用的图像相比（Krestenitis 等人，2019），本文将分辨率更高的图像作为模型的输入。图像块的填充方式确保选取的像素块中包含海面。随机选取一个训练集样本图像，使图像块填充任务更易于进行。



(a)

(b)

图2：（a）训练集中的原始图像示例。（b）训练集中的填充图像示例。

对训练集中的样本图像进行随机选取，以使图像块填充任务更加简便。应用数据增强来扩大训练数据集。仅对训练集应用随机水平和垂直翻转以进行数据增强。使用训练集的均值和标准差对图像进行归一化。随后，将归一化后的图像作为本研究所用各种模型的输入。

3.3 模型

3.3.1 残差网络

当卷积神经网络（CNNs）的深度超过某个程度后，由于梯度消失和梯度爆炸问题，模型的准确率会受到不利影响。为了解决这一问题，何恺明等人提出了残差网络（ResNet）（He 等人，2015）。ResNet 由残差块组成，在更深的网络中，梯度可能会消失，而这些残差块可以学习恒等映射。将残差块堆叠起来，有助于通过学习恒等函数缓解梯度消失问题。

设输入为x，从输入到输出的目标映射记为$H(x)$。输出 $H(x)$ 与输入 x 之间的残差记为$F(x)$，可通过方程（1）计算

$$F(x) = H(x) - x \tag{1}$$

因此，可以使用方程将原始映射重新表述为残差映射
(2)

$$H(x) = F(x) + x \tag{2}$$

要执行上述映射，一个约束是$F(x)$ 和 x 应具有完全相同的维度。不过

如果情况并非如此，则可以使用投影向量W_s来匹配维度。这在方程（3）中有所展示。

$$y = F(x,\{W_i\}) + W_s x \tag{3}$$

何恺明等人（He 等人，2015）表明，学习残差映射比学习原始映射更容易。

本文使用了多种预训练的残差网络模型变体——ResNet-18、ResNet-34、ResNet-50 和 ResNet-101 编码器，即移除了用于 ImageNet 数据集上的目标识别的最终分类层（Deng 等人，2009；Russakovsky 等人，2015；He 等人，2015）。

3.3.2 EfficientNet

Zoph 等人（Zoph 等人，2017）提出了神经架构搜索（NAS），用于优化图像分类的网络架构。为解决 EfficientNet 的训练瓶颈，谭明兴等人提出了 EfficientNetV2（Tan 等人，2019；Tan 和 Le，2021）。主要改动是在早期层中用谭明兴等人提出的新模块 FusedMBConv模块替换 EfficientNet 的 MBConv模块。在 MBConv模块中，3×3 深度卷积层后接一个 1×1 普通卷积层；而在 FusedMBConv模块中，这两层被一个融合的 3×3 卷积层取代。EfficientNetV2 的其他改动包括采用 3×3 卷积核尺寸而非较大的卷积核尺寸，同时增加层数，并减小 MBConv模块的扩展比，因为较小的扩展比以及移除最后一个步长为1的阶段，都是为减少内存访问开销而进行的优化（Tan 和 Le，2021）。

3.3.3 DeepLabV3 与 DeepLabV3+

编码器的输出被作为解码器的输入，即本研究考察的解码器之一——DeepLabV3 和 DeepLabV3+。

在 DeepLabV3 解码器中，使用了空洞空间金字塔池化（ASPP）模块（Chen 等人，2017；Chen 等人，2016）。这是对早期 DeepLab解码器版本的改进，早期版本使用空洞卷积层代替标准卷积层。空洞卷积或膨胀卷积会使用膨胀率，在卷积核作用于图像时扩大其覆盖的像素范围。在这个 ASPP 模块中，有四个空洞卷积层。ASPP 会对来自编码器模块的特征图应用平均池化层，以提供全局上下文信息。ASPP 模块这五个层的输出经过拼接后，再进行双线性上采样，从而生成维度与输入图像完全一致的特征图。

与 DeepLabV2 相比，DeepLabV3+ 解码器进行了一些细微改动（Chen 等人，2017；Chen 等人，2018）。ASPP 模块五个层的输出被拼接起来，并进行了 4 倍双线性上采样。对于编码器模块中对应的特征，使用了一个 1×1 卷积，以平衡骨干网络的低级特征与编码器模块压缩后的语义特征之间的重要性。随后，将得到的特征与上采样后的特征拼接，再通过 3 个卷积层对拼接后的特征进行细化。编码器模块的这条连接及拼接是 DeepLabV3+解码器的一项细微改动。最后，使用双线性上采样对得到的特征图进行上采样，使其维度与输入图像的维度完全一致。

3.4 模型训练

本研究使用以下编码器——ResNet-18、ResNet-34、ResNet-50、ResNet-101、Effi- cientNetV2S 和 EfficientNetV2M（He 等人，2015；Tan 和 Le，2021）。在使用迁移学习的编码器模型中，采用在 ImageNet 数据集上预训练的模型（Bengio，2012；Deng 等人，2009；Russakovsky 等人，2015）。本研究使用以下解码器——DeepLabV3 和 DeepLabV3+（Chen 等人，2017；Chen 等人，2018）。所有编码器-解码器模型均进行端到端训练，共训练 100 个训练轮次，超参数则通过经验确定。

由于数据集中包含 5 个语义类别，因此使用平均分类交叉熵损失。分类交叉熵的定义见公式 4，其中c_i表示编码后的类别，p_i表示模型对数据集中n 个类别中每个类别预测的概率。采用随机梯度下降（SGD）优化器，初始学习率为 1×10^{-2}，动量为 0.9，权重衰减为 1×10^{-4}。Liang-Chieh Chen 等人发现，将 SGD 与多项式学习率调度器结合时，分割模型的性能更高（Chen 等人，2016）。本研究将 SGD 优化器与多项式学习率调度器结合使用，使学习率按多项式方式衰减。多项式学习率调度器的定义见公式 5，其中lr_0为初始学习率，e 为当前训练轮次，T_e

表1：用于训练的不同模型的批量大小。

编码器	解码器	#参数（百万）	批量 size
ResNet-18	DeepLabV3+	12.34	32
ResNet-34	DeepLabV3+	22.45	24
ResNet-50	DeepLabV3+	25.07	8
ResNet-101	DeepLabV3+	44.06	8
EfficientNetV2S	DeepLabV3	21.42	8
EfficientNetV2M	DeepLabV3	54.11	4

表示训练轮次总数，*power* 控制学习率衰减。当某一训练轮次的学习率低于特定阈值并趋近于零时，学习过程可能受到影响。为避免这种情况，将最小学习率设为阈值；如果学习率低于该阈值，则采用该最小学习率。对于多项式学习率调度器，参数设置如下：lr_0 设为 1×10^{-2}，*power* 设为 0.9，最小学习率设为 1×10^{-6}。

$$CE(c,p) = -\sum_{i=1}^{n} c_i \log(p_i) \tag{4}$$

$$lr = lr_0 \times (1 - \frac{e}{T_e})^{power} \tag{5}$$

随机梯度下降优化器中使用权重衰减进行正则化，采用丢弃率为 10% 的丢弃层，并通过水平和垂直翻转进行数据增强。

由于不同模型的参数数量不同，所需的图形处理器（GPU）内存也各不相同，因此使用不同的批量大小训练不同模型。表 1 展示了训练不同模型时使用的批量大小。训练这些模型时使用了高性能计算集群上的 Nvidia V100 GPU。

3.5 迁移学习

迁移学习是一种机器学习方法，可简化从头开始训练深度神经网络的过程（Bengio，2012）。迁移学习使我们能够将先前开发的机器学习模型用于一项新的但相关的任务。这种方法在计算机视觉领域广受欢迎，主要是因为卷积神经网络能够将学到的低级特征提取能力迁移到各种任务中。

3.6 评估指标

为评估模型的性能，采用交并比（IoU）指标（ Krestenitis 等，2019）。平均交并比（m-IoU）是交并比的平均值

表 2：各种模型在随机验证集上进行五折验证的性能指标。

编码器	解码器	平均交并比 (%)
ResNet-18	DeepLabV3+	67.345 ± 1.407
ResNet-34	DeepLabV3+	67.522 ± 3.898
ResNet-50	DeepLabV3+	67.578 ± 3.631
ResNet-101	DeepLabV3+	68.152 ± 2.470
EfficientNetV2S	DeepLabV3	60.668 ± 2.796
EfficientNetV2M	DeepLabV3	58.997 ± 3.021

数据集中不同语义类别的交并比。类别交并比是针对单个语义类别计算的交并比。

4 结果与讨论

4.1 定量分析

我们进行了五折交叉验证，以评估模型在不同随机验证集划分上的性能偏差。表 2 展示了各模型在随机验证集上进行五折验证的性能指标。对于所有模型，五折验证中平均交并比（m-IoU）的偏差都很显著（大于 1%）。这是意料之中的，因为数据集高度不平衡，且被随机划分为训练集和验证集。由于样本划分及其语义类别分布不同，某些划分的平均交并比会高于其他划分。因此，验证集的平均交并比取决于交叉验证实验中的数据集划分。

表3展示了各模型中表现最佳的模型在包含 110幅图像的测试集上的性能指标。在测试集上，采用 ResNet-50 编码器和 DeepLabV3+解码器的模型取得了最高的平均交并比，为 64.868%。该模型的表现略低于 Mario Kresteni- tis 等人在先前研究中提出的表现最佳的模型；后者在测试集上的平均交并比为 65.06%（Krestenitis 等，2019）。表1 还展示了各种模型的参数数量。若比较采用 EfficientNetV2M 编码器和 DeepLabV3 解码器、参数量为 54.11 百万的模型，以及采用 ResNet-50 编码器和 DeepLabV3+ 解码器、参数量仅为 25.07 百万的模型，它们的表现分别为 55.504% 和 64.868%。这表明，增加模型的参数数量并不一定能提升模型在所有任务中的表现。

表 4 展示了测试集中 110

表3：在包含110张图像的测试集上，各模型中表现最佳模型的性能指标。

编码器	解码器	平均交并比 (%)
ResNet-18	DeepLabV3+	59.647
ResNet-34	DeepLabV3+	60.843
ResNet-50	DeepLabV3+	**64.868**
ResNet-101	DeepLabV3+	64.677
EfficientNetV2S	DeepLabV3	55.492
EfficientNetV2M	DeepLabV3	55.504

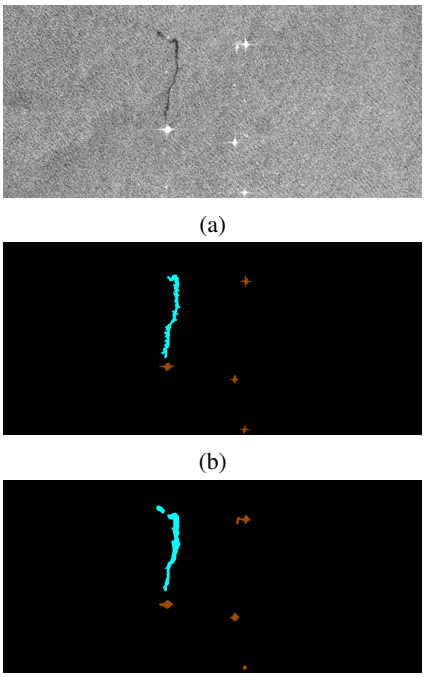
幅图像上性能最佳模型的分类性能指标，即本研究采用 ResNet-50 编码器和 DeepLabV3+解码器的模型，以及 Marios Krestenitis 等人研究中的最佳模型。Marios 等人的最佳模型在“溢油”和“溢油伪影”类别上的类别交并比分别为 53.38 % 和 55.40%（Krestenitis 等，2019）。另一方面，本研究中的最佳模型在“溢油”和“溢油伪影”类别上的类别交并比分别为 61.549% 和 40.773 %。尽管本研究中的模型在“溢油伪影”类别上的得分较低，但其在更受关注的“溢油”类别上的得分更高。另一个观察结果是，与 Marios Krestenitis 等人的最佳模型相比，本研究中的最佳模型在“船舶”检测方面表现更为突出，类别交并比分别为 33.378% 和 27.63%。对于其余类别，即“海面”和“陆地”，本研究与 Marios Krestenitis 等人的最佳模型性能相当。

表 4：在包含 110 张图像的测试集上，性能最佳模型的分类别性能指标，即类别交并比。

	类别交并比 (%)	
	我们的最佳模型	最佳模型 Marios 等人
语义类别		
海面	96.422	**96.43**
溢油	**61.549**	53.38
溢油伪影	40.773	**55.40**
Ship	**33.378**	27.63
Land	92.218	**92.44**
mean	64.868	**65.06**

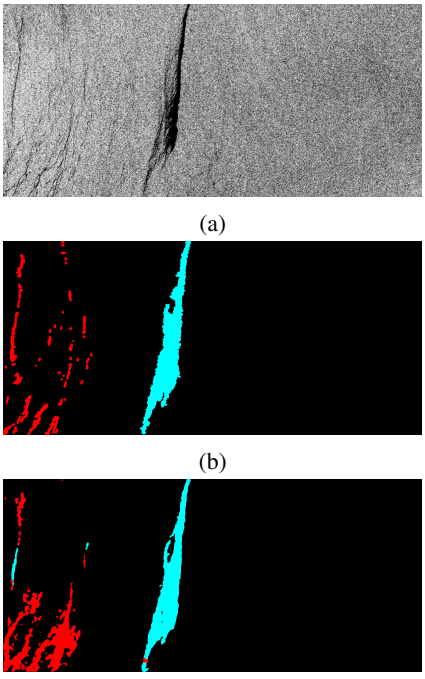
4.2 定性分析

从图3(a)所示的测试样本、图3(b)所示的真实标注掩膜及其预测结果



(a) (b) (c)

图3：(a) 测试样例图像 (b) 真实标注 (c) 预测掩膜。



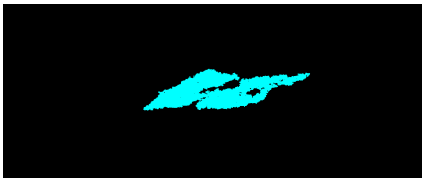
(a) (b) (c)

图4：(a) 测试样例图像 (b) 真实标注 (c) 预测掩膜。

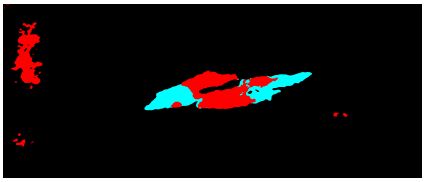
图3(c)所示的掩膜中，可以观察到，模型能够以合理的准确度检测出沿线相对较小的“溢油”（青色）区域



(a)



(b)



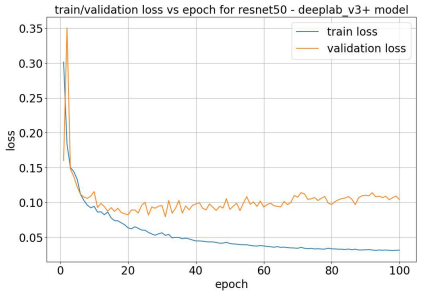
(c)

图 5：（a）测试样例图像（b）真实标注（c）预测的掩膜。

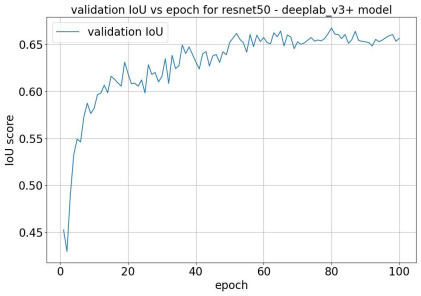
以及“船舶”（布朗色）。从图4(a)、4(b)和4(c)可以观察到，模型能够以合理的准确度检测出相对较大面积的“溢油”。在这个测试样本中，模型混淆了“溢油”和“溢油伪影”这两个类别，导致部分区域被同时预测为这两个类别。然而，从图5(a)、5(b)和5(c)可以观察到，模型**未能**以合理的准确度检测出相对较大面积的“溢油”。模型呈现出各种检测结果，既有高度准确的结果，也有一些不太可靠的结果。

4.3 学习曲线

图 6(a) 展示了采用 ResNet-50 编码器和 DeepLabV3+解码器的模型（即我们实验中表现最佳的模型）的训练集和验证集损失随训练轮次变化的曲线。图 6(b) 展示了采用 ResNet-50 编码器和 DeepLabV3+ 解码器的模型在验证集上的平均交并比随训练轮次变化的曲线。这两幅学习曲线图表明，验证指标已趋于收敛，也就是说，经过充分训练后，指标变化不大。不过，仍出现了一些过拟合：随着训练轮次数增加，训练损失与验证损失之间的差距逐渐扩大。



(a)



(b)

图 6：（a）使用 ResNet-50 编码器和 DeepLabV3+ 解码器的模型的训练损失和验证损失曲线。（b）使用 ResNet-50 编码器和 DeepLabV3+解码器的模型的验证平均交并比曲线。

5 结论

此前针对相同主题和数据集的研究将高维图像划分为较小的图像块，作为模型的输入。表现最佳的模型取得了 65.06% 的平均交并比（m-IoU）和 53.38% 的类别交并比。相比之下，本文直接将高维图像输入模型，而未将其划分为图像块。本研究中采用 ResNet-50 编码器和 DeepLabV3+ 解码器的最佳模型，在“溢油”类别上取得了 64.868% 的平均交并比和 61.549% 的类别交并比。这创造了新的基准结果，表明使用高维图像能够提升“溢油”检测性能。

然而，该数据集还包含“溢油伪影”类别的样本，这可能会使训练后的模型产生混淆，难以区分“溢油”和“溢油伪影”类别。未来可以尝试使用各种编码器和解码器，以取得更好的结果。此外，本研究使用的编码器和解码器还可以通过加入视觉自注意力模块加以改进。未来的研究可在本文分析的基础上探索这些改进方向。

参考文献

Bengio, Y. (2012). 面向无监督学习和迁移学习的表征深度学习。载于 Guyon, I., Dror, G., Lemaire, V., Taylor, G., and Silver, D.（编），*ICML* 无监督学习与迁移学习研讨会论文集，第 27 卷，机器学习研究论文集，第 17–36 页，美国华盛顿州贝尔维尤。PMLR。Broekema, W. (2016). 危机诱发的学习与议题政治化：Braer、Sea Empress、Erika 和 Prestige 溢油灾害。公共管理，94(2):381–398。Chaurasia, A. and Culurciello, E. (2017). Linknet：利用编码器表征实现高效语义分割。*CoRR*，abs/1707.03718。Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2016). Deeplab：基于深度卷积网络、空洞卷积和全连接条件随机场的语义图像分割。*CoRR*，abs/1606.00915。Chen, L., Papandreou, G., Schroff, F., and Adam, H. (2017). 重新思考用于语义图像分割的空洞卷积。*CoRR*，abs/1706.05587。Chen, L., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). 采用空洞可分离卷积的编码器-解码器，用于语义图像分割。*CoRR*，abs/1802.02611。Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., and Schiele, B. (2016). 用于语义城市场景理解的 Cityscapes 数据集。载于*IEEE* 计算机视觉与模式识别会议论文集（*CVPR*）。Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet：大规模分层图像数据库。载于*CVPR09*。Deng, L. (2012). 用于机器学习研究的手写数字图像 MNIST 数据库 [网络精选]。*IEEE* 信号处理杂志，29(6):141–142。Dunnet, G. M., Crisp, D. J., Conan, G., Bournaud, R., Cole, H. A., and Clark, R. (1982). 石油污染与海鸟种群。伦敦皇家学会哲学汇刊 *B* 辑：生物科学，297(1087):413–427。He, K., Zhang, X., Ren, S., and Sun, J. (2015). 用于图像识别的深度残差学习。*CoRR*，abs/1512.03385。Krestenitis, M., Orfanidis, G. A., Ioannidis, K., Avgerinakis, K., Vrochidis, S., and Kompatsiaris, Y. (2019). 利用深度神经网络从卫星图像中识别溢油。遥感，11:1762。Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). 使用深度卷积神经网络进行 ImageNet 分类。载于 Pereira, F., Burges, C., Bottou, L., and Weinberger, K.（编），神经信息处理系统进展，第 25 卷。Curran Associates, Inc. LeCun, Y., Boser, D., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). 反向应用于手写邮政编码识别的反向传播。《神经计算》，1(4):541–551。LeCun, Y.、Bottou, L.、Bengio, Y. 和 Haffner, P.（1998）。应用于文档识别的基于梯度的学习。《*IEEE*会刊》，86(11):2278–2324。Lin, T.、Maire, M.、Belongie, S. J.、Bourdev, L. D.、Girshick, R. B.、Hays, J.、Perona, P.、Ramanan, D.、Doll´ar, P. 和 Zitnick, C. L.（2014）。Microsoft COCO：上下文中的常见物体。*CoRR*，abs/1405.0312。Long, J.、Shelhamer, E. 和 Darrell, T.（2014）。用于语义分割的全卷积网络。*CoRR*，abs/1411.4038。Middlebrook, A.、Ahmadov, R.、Atlas, E.、Bahreini, R.、Blake, D.、Brioude, J.、Brock, C.、de Gouw, J.、Fahey, D.、Fehsenfeld, F.、Holloway, J.、Lueb, R.、McKeen, S.、Meagher, J.、Meinardi, S.、Murphy, D.、Parrish, D.、Peischl, J. 和 Watts, L.（2010）。深水地平线漏油事件对空气质量的影响（特邀）。美国地球物理联合会秋季会议摘要，第02页起。Ronneberger, O.、Fischer, P. 和 Brox, T.（2015）。U-net：用于生物医学图像分割的卷积网络。*CoRR*，abs/1505.04597。Russakovsky, O.、Deng, J.、Su, H.、Krause, J.、Satheesh, S.、Ma, S.、Huang, Z.、Karpathy, A.、Khosla, A.、Bernstein, M.、Berg, A. C. 和 Fei-Fei, L.（2015）。ImageNet大规模视觉识别挑战赛。《国际计算机视觉期刊》（*IJCV*），115(3):211–252。Sandler, M.、Howard, A. G.、Zhu, M.、Zhmoginov, A. 和 Chen, L.（2018）。倒置残差与线性瓶颈：用于分类、目标检测和分割的移动网络。*CoRR*，abs/1801.04381。Shigenaka, G.（2009）。回顾与展望：埃克森·瓦尔迪兹漏油事件20年后。美国：美国国家海洋和大气管理局，第18页。Tan, M. 和 Le, Q. V.（2021）。EfficientNetV2：更小的模型与更快的训练。*CoRR*，abs/2104.00298。Tan, M.、Pang, R. 和 Le, Q. V.（2019）。EfficientDet：可扩展且高效的目标检测。*CoRR*，abs/1911.09070。Zhao, H.、Shi, J.、Qi, X.、Wang, X. 和 Jia, J.（2016）。金字塔场景解析网络。*CoRR*，abs/1612.01105。Zoph, B.、Vasudevan, V.、Shlens, J. 和 Le, Q. V.（2017）。学习用于可扩展图像识别的可迁移架构。*CoRR*，abs/1707.07012。

致谢

我们感谢格罗宁根大学信息技术中心提供的支持，以及为我们提供使用 Peregrine 高性能计算集群的权限。我们还要感谢 Marios Krestenitis 和 Konstantinos Ioannidis 创建并提供数据集。