细粒度空间-时间感知用于气体泄漏分割

Xinlong Zhao and Shan Du*

The University of British Columbia - Okanagan

ABSTRACT

气体泄漏对人类健康和环境构成重大风险。尽管长期存在担忧，但由于泄漏的隐蔽外观和随机形状，能够有效准确检测和分割泄漏的方法仍然有限。在本文中，我们提出了一种细粒度时空感知（FGSTP）算法用于气体泄漏分割。FGSTP 捕获关键运动线索并跨帧将它们与精细的对象特征集成在一个端到端的网络中。具体来说，我们首先构建一个相关体积来捕获连续帧之间的运动信息。然后，细粒度感知逐步使用先前输出优化对象级别的特征。最后，采用解码器进行边界分割优化。由于没有高度精确标记的数据集用于气体泄漏分割，我们手动标注了一个气体泄漏视频数据集 GasVid。在 GasVid 上的实验结果表明，我们的模型在分割非刚性物体如气体泄漏方面表现出色，并且与其它最先进的（SOTA）模型相比生成了最准确的掩模。

Index Terms— 气体泄漏分割、非刚性和模糊对象分割、视频监控系统

1. 介绍

气体泄漏是由于连接损坏、腐蚀或安装不当等原因意外发生的事件。无论是天然气、甲烷还是其他危险气体，它们逸入大气中可能导致爆炸、火灾以及长期健康问题。检测气体泄漏对于早期干预和减轻潜在灾害至关重要。传统的泄漏检测方法涉及检查人员手动使用仪器在疑似漏点附近评估管道状况。这些方法效率低下，并且对检查人员构成潜在危害。相比之下，视频监控系统可以提供准确的实时监测而无需人工操作。因此，设计一个通过分析运动模式和形状变化来识别泄漏的视频处理算法是进行气体泄漏分割的一个更好选择。此类算法能够实时处理连续视频流，与手动检查相比，可提供更快、更安全的检测。

气体泄漏具有与普通物体不同的特征。它们是无色、无味且人类感官无法察觉的。尽管红外相机可以可视化泄漏 [1]，但由于其非刚性形状和模糊纹理，通常仍然难以区分它们。在许多情况下，只有当泄漏处于运动状态时才可能进行分割。例如，Lu 等人 [2] 提出了基于高斯背景建模的自适应气体泄漏分割方法。然而，由于气体形态不规则以及背景的变化，这种方法在工业场景中面临显著限制。基于背景差分和光流 [3] 的常见运动提取方法因物体形状变化及严重噪声干扰而失效。

最近，深度学习技术在目标分割 [4] 方面表现出了鲁棒性。范等人。[5] 首先定位粗略隐藏的目标，然后在解码器中对其进行细化。王等人。[6] 基于 YOLOv7 [7]，展示了检测可见非刚体物体（如尘埃）的能力。对于 RGB 图像中的目标而言，仅依赖空间信息的方法可能是有效的。然而，如果模型忽略了运动信息，在红外相机中分割模糊物体变得具有挑战性。闫等人。[8] 结合光流来学习运动特征，杨等人。[3] 则使用无监督学习在卷积神经网络中生成像素级的运动标签。然而，这两种方法都容易受到噪声干扰的影响。基于记忆的分割方法 [9, 10, 11] 难以应对气体泄漏形状的多样性。此外，冗余的记忆显著影响了分割性能。程等人。[12] 集成了边界优化和光流以进行运动分析，显示了对气体泄漏分割的潜力。然而，它难以区分与泄漏纹理相似的对象，如云。据我们所知，很少有方法可以有效地在红外视频中分割模糊且非刚性的对象，比如气体泄漏。

在本文中，我们通过提出细粒度时空网络来解决非刚性、模糊物体在红外视频中的分割挑战。我们采

Thanks to SSHRC for funding support (NFRF-GR024801).

*Corresponding author: shan.du@ubc.ca

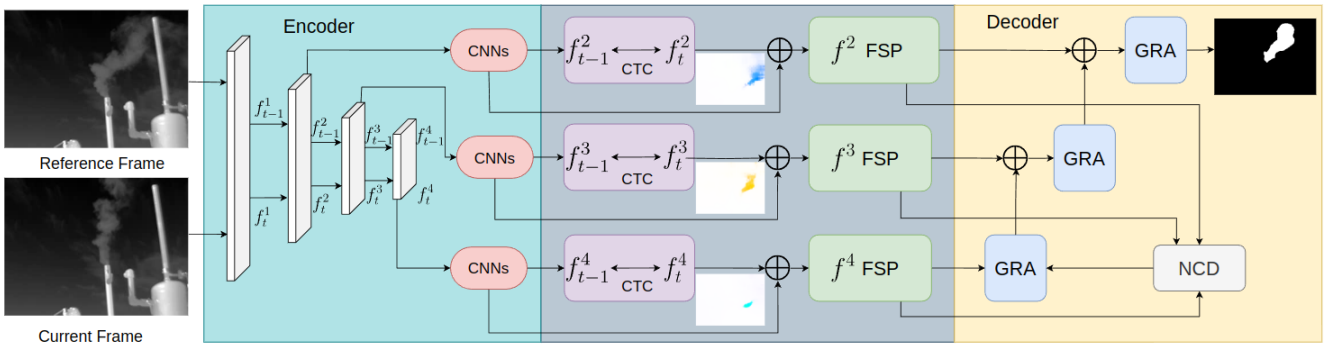

Fig. 1: FGSTP 的架构。该模型每次处理每一个当前帧以及一个相邻帧。每个当前帧需要与两个相邻帧分别进行两次处理。编码器对输入帧去噪并提取多尺度特征，而解码器通过 GRA 和 NCD 模块优化边界。f_i^j 表示第 i 帧和第 j 尺度的特征。连续时间相关（CTC）块捕获运动信息。f^j 表示第 j 尺度的 CTC 输出。细粒度空间感知（FSP）模块细化空间特征。

用多尺度基于变压器的 [13] 编码器以减轻视频噪声。为了提取气体泄漏的代表性运动特征，我们使用提出的连续时间相关模块（CTC）跨多个尺度细化相关的运动线索。我们还提出了细粒度空间感知模块（FSP），通过结合运动特征和编码器特征以及提炼泄漏的信息语义特征来捕捉空间特征。此外，边界确定对于多态物体来说至关重要但具有挑战性，因为它们缺乏清晰的边界使它们与背景融合。我们的解码器可以从粗略预测优化到准确的掩模。我们在 GasVid[1] 数据集上进行了实验，我们对每个帧的手动注释了气体泄漏区域以确保泄漏面积的准确表示和有效的模型训练。总之，我们的贡献包括：(1) 我们提出了一种专门用于分割红外视频中非刚性和模糊物体的端到端架构；(2) 我们设计了 CTC 和 FSP 模块以获得关键运动线索并从编码器输出感知形状或边界特征。

论文组织如下：第 2 节介绍了我们的 FGSTP 框架。第 3 节展示了 GasVid 数据集上的实验结果。第 4 节总结了我们的工作。

2. 方法

2.1. 框架概述

图 1显示了 FGSTP 的架构。该模型由四个组件组成：(1) 基于 Transformer 的编码器去噪并获取多尺度特征；(2) 连续时间相关性 (CTC) 捕捉关键动作；(3) 细粒度空间感知 (FSP) 通过编码器特征与 CTC 输出之间的残差连接来细化空间特征；(4) 解码器优化预测掩模边界。FGSTP 将推理帧与其相邻的两帧作为输入，以识别目标对象的位置和形状。模型输出是推理帧的二值掩模。

FGSTP 是一个单阶段模型，不需要额外的辅助输入。它设计用于分割模糊的非刚性物体，如气体泄漏。泄漏的形状在短时间内会发生显著变化。我们的研究发现，在这种视频对象分割中，两个相邻帧包含了足够的信息来进行推理帧的推断。因此，我们的模型每次仅计算 3 帧（两帧相邻和一帧推理）。此操作不仅减少了冗余计算，还有助于 FGSTP 更突出地捕捉帧间差异。

2.2. FGSTP 架构

(1) **编码器**。金字塔视觉转换器（PVT）[13] 是用于从三个输入帧中提取全局特征的模型主干。编码器有四层，生成不同尺度下的全局特征。每层特征的大小是 $C \times H/2^{i+1} \times W/2^{i+1}$，其中 $i \in \{1,2,3,4\}$,H,W,C 分别是高度、宽度和通道数。先前的研究发现，最后三层保留了更多的高层次语义特征，并且维度减少了 [12]。因此我们将 $2nd, 3rd$, 和 $4th$ 层的特征放入多尺度网络中提取局部特征 $f_t^i (i \in \{2,3,4\})$。

(2) **连续时间相关性**。连续时间相关性（CTC）旨在从连续帧中捕捉运动线索。与需要光流结果作为网络输入的参考 [3] 不同，CTC 是 FGSTP 中的一个嵌入组件。受 [14] 的启发，在每个尺度（$i \in \{2,3,4\}$）上，CTC 通过两帧间的特征一致性计算时间对应关系。CTC 的结构如图 2所示。给定编码器的两个帧的特征，$(f_t, f_{t-1}) \in R^{2C \times H' \times W'}$，可以按照以下方式计

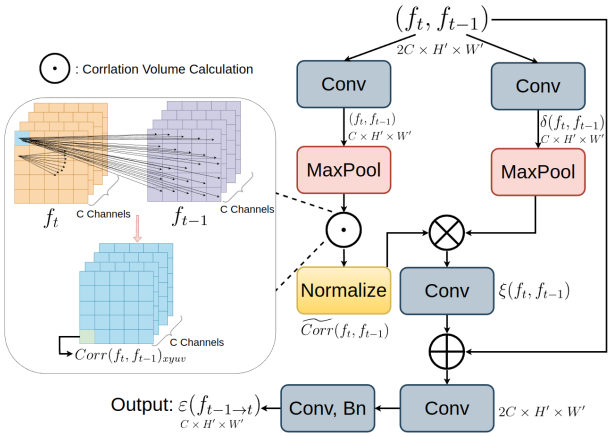

Fig. 2: CTC 的结构。它计算用于运动信息的 4D 相关体积 $Corr(f_t, f_{t-1})$，并整合通道级语义以实现精确的运动对齐。

算 4 维相关体积 $Corr(f_t, f_{t-1}) \in R^{H' \times W' \times H' \times W'}$：

$$Corr(f_t, f_{t-1})_{xyuv} = \exp\left(\sum_c (f_t)_{xyc} \cdot (f_{t-1})_{uvc}\right) \tag{1}$$

其中 c 是通道索引。当连续帧之间的所有特征相似性都计算完毕后，可以找到两帧之间最相关的部分。然后对相关体积 $Corr(f_t, f_{t-1})_{xyuv}$ 的后两个维度 (u, v) 进行归一化处理，以获得参考帧和当前帧之间的对应部分。归一化的形式如下：

$$\widetilde{Corr}(f_t, f_{t-1})_{xyuv} = \frac{Corr(f_t, f_{t-1})_{xyuv}}{\sum_u \sum_v Corr(f_t, f_{t-1})_{xyuv}} \tag{2}$$

通过每个通道计算相邻帧特征的归一化相关体积。应将其与通道特征 $\delta(f_t, f_{t-1})$ 合并以获得最终的时间相关性。因此，通道特征乘以归一化后的相关性，并经过一系列卷积层和残差连接，得到最终输出 $\varepsilon(f_{t-1\to t})$。通道特征与归一化相关体积的集成如下：

$$\xi(f_{t-1\to t}) = \delta(f_t, f_{t-1}) \cdot \widetilde{Corr}(f_t, f_{t-1}) \tag{3}$$

(3) **细粒度空间感知**。输出仅包含运动信息。仅依赖运动信息不足以进行气体泄漏分割，因为它忽略了形状和纹理等有价值细节。为解决此问题，受 [15] 启发，我们提出细粒度空间感知（FSP）以捕获空间特征。FSP 的输入 $\zeta(f_{t-1}, f_t)$ 通过残差连接结合了编码器中的 CTC $(\varepsilon(f^i_{t-1\to t}) \in R^{C \times H' \times W'})$ 和多尺度特征 $f^i_t \in R^{C \times H' \times W'}, i \in \{2, 3, 4\}$。图 3是 FSP 的结构。

在 FSP 中，首先增加 $\zeta(f_{t-1}, f_t)$ 的通道数，然后沿着通道维度将其分为 G 组，其中每组中的特征进一步分成三组。每组的第一组 $(\rho^1_j, j \in (1...G))$ 与下一组连接。然后，根据原始组中的位置，通过卷积和分裂操作将每组后两组的特征重新排列为两个新组（$\rho^2_1 \to \rho^2_G$ 和 $\rho^3_1 \to \rho^3_G$）。最后，两组混合特征（ρ^2 和 ρ^3）在卷积块中进行整合并生成输出 r_t。混合操作跨通道提取关键线索以实现稳健的特征表示。特征组迭代通过部分参数共享整合不同通道的特征。这些 FSP 过程提升了捕捉丰富线索的能力，增强了对象感知和预测细化。

(4) **解码器**。解码器旨在识别非刚性物体的边界。邻域连接解码器（NCD）[5] 从先前模块 FSP 的不同尺度中提取粗略掩模。粗略掩模是目标对象的大致位置。反转粗略掩模意味着切换前景和背景的值。然后在组反转注意力（GRA）[5] 模块中，通过连接和卷积使用反转后的掩模对粗略掩模进行插值。

$$q^k = \text{concat}(p^k_i : r^k) * W_{GRA} \tag{4}$$

其中，q^k 是帧 k 中每个尺度的集成掩码，p^k_i 是原始粗掩码之一，r^k 是反转掩码，而 W_{GRA} 是 GRA 的卷积核。这样，只有边界区域在粗掩码中被强调，因为前景和背景的值相互抵消。更多细节可见于 [5]。最后我们结合 GRA 和 FSP 的输出以获得气体泄漏分割的结果。

2.3. 损失函数

类似 [12]，我们的损失函数结合加权交叉熵损失 L^w_{ce} 和交并比（IoU）损失 L^w_{iou} 来实现稳健优化。这些组件平衡了预测的准确性和结构对齐。损失函数如下：

$$L_{hybrid} = L^w_{ce} + L^w_{iou} \tag{5}$$

加权交叉熵损失，L^w_{ce}，通过为这些区域分配权重确保模型更加关注困难的区域。IoU 损失，L^w_{iou}，通过强调整体预测形状和大小与真实值的一致性来补充交叉熵损失。

3. 实验

3.1. 数据集

我们的实验使用了公共数据集 GasVid[1]。该数据集是一个包含 31 个视频的气体泄漏视频集，每个视频时长为 24 分钟。更多详情请参阅 [1]。然而，该数据集

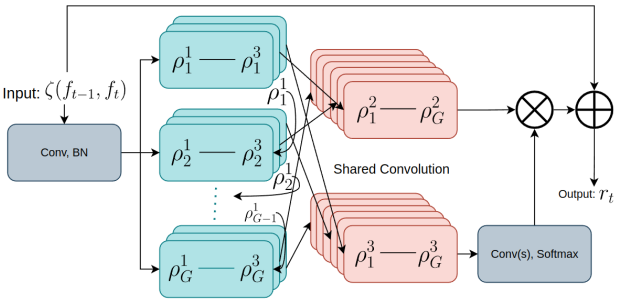

Fig. 3: FSP 的结构。它通过与编码器的残差连接增强了 CTC 中的运动特征，并计算语义表示。

不包括标注信息，因此我们手动标记了这些视频。在这个数据集中，有些视频与其他视频非常相似，而有些则过于困难以至于人类难以区分气体泄漏。我们将那些未使用的视频移除，并将剩余的视频分为以下几类：近距离（摄像头）且背景清晰（无云，易于分割）；中等距离且背景清晰（中等难度分割）；近距离但背景复杂（难以分割）；长距离且背景清晰（困难）；长距离且背景复杂（困难）。对于每一类别，我们分配一个视频用于训练和另一个用于测试。这种设置允许模型在训练过程中学习所有类型的视频，并有效地评估不同场景下的泄漏分割性能。总共选用了 14 个视频。为了加速训练，每个视频被分成多个片段，每段包含 150 帧。我们也删除了每个视频中的重复片段以减少不平衡现象。训练集由 9 个视频和 120 个片段组成，而测试集则有 5 个视频共 67 个片段。我们在 GasVid 数据集上的结果在第 3.3 节中展示。

3.2. 指标和配置

我们的指标包括以下指数：S-**测量** (S_α)：提出于[16]，结合了区域感知和对象感知的评估来衡量结构相似性。**加权 F 值** (F_β^ω)[17]：结合精度和召回率，通过平衡因子加权，以提供综合评估。**平均绝对误差**（MAE）：记为 M，通过比较预测掩码与真实值来量化像素级准确性。**增强对齐度量** (E_ϕ)[18]：评估像素级对应关系以及图像级别统计，使其非常适合于评价非刚性物体分割中的全局和局部准确性。**平均交并比**（mIoU）：计算预测与地面真实之间的平均重叠。**平均** Dice：计算平均相似系数，量化预测与地面真实之间的重叠，强调分割准确性。

不同于常规的目标分割方法，我们不使用任何预训练的骨干网络或模型，因为我们的数据集具有特定的特点。在 RGB 数据集上预训练的模型可能会对我们的实验产生负面影响。为了公平比较，我们从头开始训练我们的模型和基准模型。输入帧被调整为 352×352。我们在 60 个 epoch 中以学习率 1×10^{-4} 进行训练。所有模型都在 NVIDIA RTX 4090 GPU 上进行训练。

3.3. 结果

我们的实验结果呈现在表 1中。绿色的数值代表每个指标的最佳分数，而红色的数值表示第二高的分数。我们的方法优于所有基准模型。即使与第二好的方法 SLT-Net[12] 相比，我们的方法在大多数指标上也表现出更优的性能。此外，其他方法与我们模型相比显示出显著差异。

Table 1: GasVid 数据集上各种基准的定量比较。

Models	$S_\alpha \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$E_\phi \uparrow$	$mIOU \uparrow$	$mDice \uparrow$
MG [3]	-	-	-	-	-	-
SINet-V2 [5]	0.684	0.472	0.022	0.743	0.361	0.465
XMem++ [9]	0.675	0.424	0.036	0.675	0.361	0.459
RMem [10]	0.685	0.431	0.023	0.711	0.361	0.451
Zoomnext [15]	0.691	0.450	0.025	0.729	0.378	0.472
SLT-Net [12]	0.702	0.500	0.025	0.806	0.392	0.505
FGSTP（**我们的**）	0.705	0.507	0.022	0.797	0.399	0.509

表 1展示了不同模型的性能比较。MG[3] 在 GasVid 数据集上失败，产生了不完整或缺失的掩码，因为它仅依赖于光流而没有监督学习。这种方法对于分割模糊和非刚性对象（如气体泄漏）无效。SINet-V2[5] 是一种基于图像的方法，独立处理每个视频帧并忽略运动信息，这对其在分割泄漏方面的性能产生了显著影响。结果，其获得了最低的 mIoU（0.361）。XMem++ 和 RMem[9, 10] 是基于记忆的模型，需要将第一帧掩码作为输入来引导分割。虽然这些方法对于常规对象有效，但它们在处理非刚性和模糊对象时遇到困难，导致加权 F-值较低（分别为 0.424 和 0.431），相比之下 SINet-V2 的加权 F-值为 0.472。ZoomNext[15] 利用了强大的空间特征提取能力，在基于图像和基于记忆的模型中表现更好，达到了 0.378 mIoU。然而，其有限的运动提取能力阻碍了其在跟踪泄漏方面的性能。SLT-Net[12] 通过利用光流提供的运动信息增强了 SINet-V2 的性能，将 mIoU 提高到 0.392。然而，它

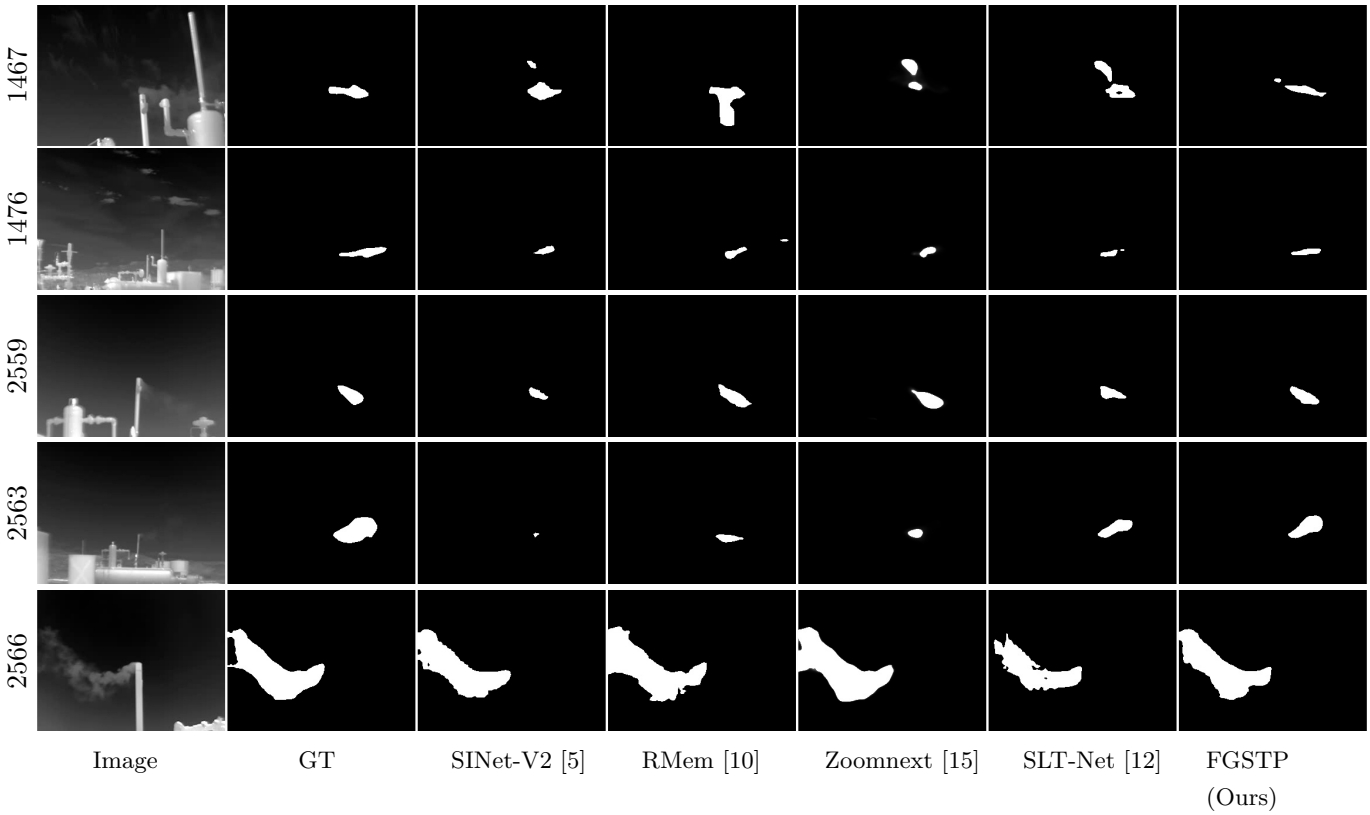

Fig. 4: GasVid 数据集中的可视化结果。我们的模型预测在不同情况下最为准确。*i.e* 近距离（相机）且背景复杂（云或鸟类干扰，1467），远距离且背景复杂（1476），中等距离且背景清晰（2559），远距离且背景清晰（2563），近距离且背景清晰（2566）

仍然缺乏对象级别的特征表示，限制了其分割能力。相比之下，我们的模型 FGSTP 将细粒度的空间感知与精确的运动细节相结合，使其能够有效找到泄漏，并达到最高的性能，即 0.399 mIoU。

我们还在图 4中提供了可视化结果。由于 MG 失效和空间限制，我们在所有测试视频中选择了最佳五个模型的结果。我们的模型在各种场景下始终生成最准确的掩码。对于 Video-1467（第一行），背景包含云和鸟，云的外观类似于气体泄漏。在这些具有挑战性的案例中，只有我们的模型正确地分割了泄漏。Video-1476 是最难的样本，因为它有较长的摄像距离并且背景被云覆盖。除了我们的模型外，所有模型都生成了形状和位置错误的掩码。Video-2559 也涉及较长的摄像距离但没有云。SLT-Net 在这段视频中正确地定位了泄漏，但与我们的模型不同的是，它未能准确匹配其形状。视频 2559 和 2566 相对容易，所有模型都能成功分割泄漏；然而，我们模型的预测包含更多细节，并且最接近地面真实值（GT）。这些结果表明，我们的模型不仅在简单的场景中捕获了更多的细节，而且在具有挑战性的条件下也表现出更加稳定和准确的表现。

3.4. 消融研究

我们在 GasVid 数据集上进行了消融研究，以评估两个核心模块：CTC 和 FSP 的有效性。结果呈现在表 2中。没有 CTC 和 FSP 的情况下，模型缺乏空间和时间信息，导致性能最差，只有 0.291 mIoU，甚至比许多基准模型还要差。这是因为解码器仅使用骨干特征生成准确的掩模非常困难。移除 FSP 而保留 CTC 会消除空间感知，使得性能下降至 0.383 mIoU。尽管逊色于完整模型，但仍然优于没有两个模块或没有 CTC 的情况。移除 CTC 并保持 FSP 会导致更大的性能下降到 0.313 mIoU，这表明运动特征非常重要。然而，空间感知也对模型性能有贡献。消融结果表明我们模型的优越性能来自于将 CTC 和 FSP 整合在一起，最

大限度地利用了时间和空间维度的优势。

Table 2: GasVid 数据集上 FGSTP 模型的 CTC 和 FSP 模块的消融研究结果

Backbone	CTC	FSP	$S_\alpha \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$E_\phi \uparrow$	$mIOU \uparrow$	$mDice \uparrow$
✓			0.644	0.419	0.027	0.741	0.291	0.408
✓	✓		0.697	0.488	0.024	0.782	0.383	0.488
✓		✓	0.655	0.430	0.027	0.761	0.313	0.428
✓	✓	✓	0.705	0.507	0.022	0.797	0.399	0.509

4. 结论

本文提出了一种创新的气体泄漏分割方法，实现了在各种场景下的更高准确性和适应性。我们的方法结合了用于降低噪声的编码器、用于细化运动线索的时间模块以及有效区分目标对象的空间感知能力。实验结果表明，我们的模型优于现有的最佳基准。此外，消融研究确认了我们两个核心模块对稳健性能的关键作用。最后，可视化结果显示，我们的模型不仅在较简单的情况下捕捉到更精细的细节，而且还能有效地处理各种具有挑战性的情况。

5. REFERENCES

[1] Jingfan Wang, Jingwei Ji, Arvind P Ravikumar, Silvio Savarese, and Adam R Brandt, "Videogasnet: Deep learning for natural gas methane leak classification using an infrared camera," Energy, vol. 238, pp. 121516, 2022.

[2] Quan Lu, Qin Li, Likun Hu, and Lifeng Huang, "An effective low-contrast SF_6 gas leakage detection method for infrared imaging," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–9, 2021.

[3] Charig Yang, Hala Lamdouar, Erika Lu, Andrew Zisserman, and Weidi Xie, "Self-supervised video object segmentation by motion grouping," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2021, pp. 7177–7188.

[4] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo, and Jieping Ye, "Object detection in 20 years: A survey," Proceedings of the IEEE, vol. 111, no. 3, pp. 257–276, 2023.

[5] Deng-Ping Fan, Ge-Peng Ji, Ming-Ming Cheng, and Ling Shao, "Concealed object detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 10, pp. 6024–6042, 2021.

[6] Mingpu Wang, Gang Yao, Yang Yang, Yujia Sun, Meng Yan, and Rui Deng, "Deep learning-based object detection for visible dust and prevention measures on construction sites," Developments in the Built Environment, vol. 16, pp. 100245, 2023.

[7] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in Proceedings of the

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2023, pp. 7464–7475.

[8] Pengxiang Yan, Guanbin Li, Yuan Xie, Zhen Li, Chuan Wang, Tianshui Chen, and Liang Lin, "Semi-supervised video salient object detection using pseudo-labels," in Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 7284–7293.

[9] Maksym Bekuzarov, Ariana Bermudez, Joon-Young Lee, and Hao Li, "Xmem++: Production-level video segmentation from few annotated frames," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), October 2023, pp. 635–644.

[10] Junbao Zhou, Ziqi Pang, and Yu-Xiong Wang, "Rmem: Restricted memory banks improve video object segmentation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2024, pp. 18602–18611.

[11] Ho Kei Cheng, Seoung Wug Oh, Brian Price, Joon-Young Lee, and Alexander Schwing, "Putting the object back into video object segmentation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2024, pp. 3151–3161.

[12] Xuelian Cheng, Huan Xiong, Deng-Ping Fan, Yiran Zhong, Mehrtash Harandi, Tom Drummond, and Zongyuan Ge, "Implicit motion handling for video camouflaged object detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2022, pp. 13864–13873.

[13] Wenhai Wang, Enze Xie, Xiang Li, Deng-Ping Fan, Kaitao Song, Ding Liang, Tong Lu, Ping Luo, and Ling Shao, "Pvt v2: Improved baselines with pyramid vision transformer," Computational Visual Media, vol. 8, no. 3, pp. 415–424, 2022.

[14] Dongxu Li, Chenchen Xu, Kaihao Zhang, Xin Yu, Yiran Zhong, Wenqi Ren, Hanna Suominen, and Hongdong Li, "Arvo: Learning all-range volumetric correspondence for video deblurring," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 7721–7731.

[15] Youwei Pang, Xiaoqi Zhao, Tian-Zhu Xiang, Lihe Zhang, and Huchuan Lu, "Zoomnext: A unified collaborative pyramid network for camouflaged object detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.

[16] Fan D, Cheng M, Liu Y, Li T, and Borji A, "A new way to evaluate foreground maps," in Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2017, vol. 245484557.

[17] Ran Margolin, Lihi Zelnik-Manor, and Ayellet Tal, "How to evaluate foreground maps?," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 248–255.

[18] Deng-Ping Fan, Cheng Gong, Yang Cao, Bo Ren, Ming-Ming Cheng, and Ali Borji, "Enhanced-alignment measure for binary foreground map evaluation," in Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18. 7 2018, pp. 698–704, International Joint Conferences on Artificial Intelligence Organization.