通过权重共享聚合利用隐式刚性约束进行点云场景流估计

Yun Wang †
wangyun@hust.edu.cn

ChengChi †
chengchi.hust@gmail.com

Xin Yang ✉
xinyang2014@hust.edu.cn

arXiv:2303.02454v2 [cs.CV] 1 Apr 2023

摘要

场景流估计通过点云预测场景点的三维运动，是自动驾驶和许多其他三维视觉应用中的核心任务。现有方法要么因忽略刚性运动一致性而遭受结构失真，要么需要显式的位姿估计和三维目标分割。估计位姿和分割目标中的误差会导致刚性约束不准确，进而误导场景流估计。本文提出一种新颖的权重共享聚合（*WSA*）方法，用于特征和场景流上采样。*WSA* 不依赖估计位姿和分割目标，能够隐式施加刚性约束，从而避免场景流估计中的结构失真。为进一步利用几何信息并保留局部结构，我们设计了一个变形程度模块，旨在保持局部区域不变性。我们对 *PointPWC-Net* 进行改进，并将所提出的 *WSA* 和变形程度模块集成到增强后的 *PointPWC-Net* 中，构建了一个端到端场景流估计网络，称为 *WSAFlowNet*。在 *FlyingThings3D [19]* 和 *KITTI [21]*数据集上的大量实验结果表明，*WSAFlowNet* 达到了最先进的性能，显著优于以往方法。我们将在 *https://github.com/wangyunlhr/WSAFlowNet.git* 发布源代码

1. 引言

从连续两帧中估计场景点的三维运动，称为场景流估计，对于许多三维应用至关重要，包括自动驾驶 [2,23,38] 和点云分割 [6]。传统方法通常从 RGB 或 RGB-D 图像中估计场景流 [4, 20, 22, 27, 33]。近年来，随着激光雷达等三维传感器的应用日益广泛，直接从三维点云中估计场景流引起了广泛关注。

当前最先进的场景流估计方法 [24, 34,36]主要基于三维点卷积神经网络（CNN），例如 PointNet [25] 和 PointNet++[26]。例如，FlowNet3D [17] 基于 PointNet++ 设计了一个端到端场景流估计网络，并引入流嵌入层来编码源点云与目标点云之间的三维运动。PointPWC-Net [36]采用 PointConv [35]作为卷积操作，并引入由粗到细策略来提高精度，尤其是针对大位移点。其流程如下：首先在最粗分辨率下估计场景流，然后逐层细化，直至达到最高分辨率。在相邻层之间，通过上采样层将估计信息从较粗层传递到较细层。为兼顾精度与计算效率，一些方法 [3, 29, 30] 在场景流估计任务中采用由粗到精结构。尽管这些方法取得了成功，但它们忽略了这样一个事实：同一刚性物体上的各点，其场景流应符合相同的几何变换，即刚性运动一致性。因此，这些方法在场景流估计中可能会出现刚性结构畸变。

为解决上述问题，我们提出了一种基于由粗到细架构、利用隐式刚性约束的网络。我们关注点的邻域 $N(p_i)$ 内部的关系。首先，我们提出了一种权重共享聚合方法，该方法应用于上采样层，并通过隐式施加刚性约束，避免场景流估计中的结构失真。重要的是，我们的方法不需要显式的刚性物体聚类/分割或位姿估计，有助于避免组合多项任务时可能产生的误差干扰。权重共享聚合的核心思想基于我们的数学证明：对于任意一个三维点 p_i 及其邻近点 $N(p_i)$，使用相同的权重聚合 $N(p_i)$ 的点坐标，以及它们的特征 $N(p_i)$ 和场景流 $N(p_i)$，能够对 $N(p_i)$ 中的点隐式地施加一致运动约束。此外，我们提出了一个形变程度模块，用于衡量参考点之间的局部结构变形程度——

† 同等贡献。 ✉ 通讯作者。

参考点 $N(p_i)$ 与变形后的点 $N(p_i^w)$ 之间的局部结构。该模块使用源点云和预测场景流构建，以进一步增强局部结构一致性。形变程度模块作为估计器的额外输入，为其提供补充几何信息。我们构建了一个场景流估计网络，将权重共享聚合和形变程度模块集成到增强版 PointPWC-Net [36] 中，并将其命名为 WSAFlowNet。大量实验结果证明了我们方法的有效性和泛化能力。我们的方法超越了当前最先进的方法 Bi-PointFlowNet [3]。根据 EPE3D 指标，在 FlyingThings3D [19] 数据集上，我们的性能比 Bi-PointFlowNet 高出 14.6%；在 KITTI [21] 数据集上高出 7.6%。

总而言之，本文有三项贡献：

1. 一种新颖的上采样权重共享聚合方法，通过使用相同的权重来聚合点坐标、场景流和特征，从而隐式地施加刚性约束。我们通过数学分析和实验验证了该方法的可行性。此外，该方法无需显式进行位姿估计和/或三维目标分割。

2. 一种变形程度模块，用于捕捉源点云与变形点云之间局部结构的差异。该模块有助于在场景流估计过程中保留刚体的局部结构。

3. 一种名为 WSAFlowNet 的有效场景流估计网络，它整合了上述两种方法，在 FlyingThings3D 和 KITTI 基准上实现了当前最先进的性能。

2. 相关工作

场景流估计。近年来，多种方法在场景流估计任务 [1,11,31,32,36]上取得了令人瞩目的性能。PointPWC-Net [36]引入了一种由粗到细网络架构来估计场景流。它无需扩大搜索范围即可捕捉大幅运动。此外，它提出了一种新颖的块间代价体，以有效编码点运动。另一方面，RMS-FlowNet[1], 设计了一种块到膨胀块流嵌入模块，结合随机采样后，能够在大规模点云上运行。PV-RAFT [34] 构建点-体素相关场以捕捉所有点对关系。针对点云的不规则性，RCP [10] 提出了一种方法，先对三维流进行逐点优化，然后将正则信息整合到循环网络中，以实现全局正则化。上述方法仅使用单向特征，这可能导致信息不足。为解决这一限制，Bi-PointFlowNet[3] 引入双向流嵌入层，以提取前向和后向的相关性。3DFlow [29]提出了一种全对全流嵌入层，能够捕捉远距离点，并结合反向可靠性验证。然而，这种逐点场景流估计方法在运动过程中未能充分利用结构信息。因此，它们在一些具有挑战性的场景中可能会遇到结构失真，这通常由点云的稀疏性引起。

刚性约束。防止结构失真的一种直接方法，是显式估计场景中刚体的几何变换，并对估计出的场景流施加刚性运动一致性约束。为此，Gojcic 等人 [9] 将场景点划分为前景簇（由多个刚体组成）和背景簇。随后，他们通过 [5, 14, 37]计算源点云与目标点云之间的自运动及每个物体的几何变换。接着，他们约束同一物体的估计场景流，使其符合相同的几何变换。Dong 等人 [7]也采用了类似方法：他们利用预训练的分割网络 [9]和 DBSCAN 聚类算法 [8]，为每个输入源点云生成抽象掩码。随后，他们使用 [14]估计每个物体的位姿。之后，根据抽象掩码和位姿计算直接的多体刚性约束，并将这些约束整合到循环神经网络中，以减轻场景流估计中的结构失真。为缓解三维目标分割中不可避免的误差，HCRF-Flow [16]将空间上相邻的点视为一个刚体，并采用条件随机场（CRF [28]）来施加局部平滑性和刚性运动一致性约束。然而，以上所有方法都需要显式进行位姿估计，而这本身也极具挑战性。位姿估计中的误差不可避免，可能导致约束不准确，进而误导场景流估计。

3. 方法

图 1 展示了我们提出的 WSAFlowNet 的总体概览（§ 3.1）。该网络由两个重要组件构成：权重共享聚合模块和变形度模块，它们在场景流细化过程中发挥着重要作用。接下来，我们首先详细介绍所提出的权重共享聚合模块（§ 3.2），然后描述变形度模块（§ 3.3）的构建。最后，我们讨论如何修改网络架构，并介绍所提出模块的实现细节（§ 3.4）。

3.1. 概述

我们的网络输入是两个点云集合$P = \{p_i \in \mathbb{R}^3\}_{i=1}^{N_1}$，分别对应时间戳 t 和 $Q = \{q_i \in \mathbb{R}^3\}_{i=1}^{N_2}$时间戳 t+1。我们的目标是估计场景流

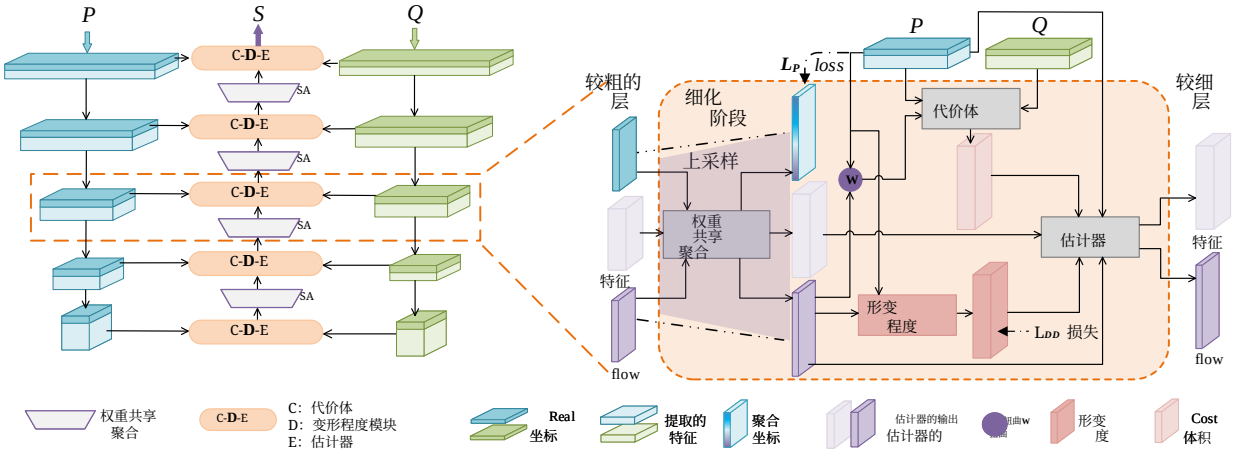

图 1. **WSAFlowNet 概览**。 左侧展示了由粗到细网络的整体流程。网络的输入是两组连续的点云 P, Q，输出对应于 P 的场景流。整个流程包含三个主要步骤：特征抽象、代价体构建和场景流细化。右侧重点展示**细化过程**，其中我们设计了两个重要模块，即**权重共享聚合**（图 2）和**变形程度模块**（图 3）。权重共享聚合实现上采样操作。变形程度模块用于捕捉源点云与变形点云之间局部结构的差异。它还作为估计器的输入，以提供充足的几何信息。

$S = \{s_i \in \mathbb{R}^3\}_{i=1}^{N_1}$ 对于每个 $p_i \in P$。由于点云数据具有稀疏性，P和 Q 并非完全对应，因此 P、 Q和 S 之间的关系只能近似表示，即 $P+S \approx Q$。我们还假设，大多数局部区域都满足 HCRF-Flow [16], [18] 中描述的近似刚性假设。局部区域内各点的运动遵循相同的变换，该变换定义为 $R \in \mathrm{SO(3)}$ 和$t \in \mathbb{R}^3$。

我们采用由粗到细结构 PointPWC- Net [36] 作为基线。首先，我们为点云 P和 Q构建特征金字塔。接下来，基于提取的特征，我们在最粗层进行场景流估计，然后逐层细化至最细层。在相邻层之间，我们采用基于权重共享聚合的上采样层。上采样后的场景流将通过变形传播至较细层，从而调整匹配的搜索中心。上采样后的特征将输入场景流估计器，以提供运动信息。在每一层中，我们依次执行变形层、代价体构建、变形程度模块和场景流估计器这四项操作。

3.2. 权重共享聚合模块

在场景流估计任务中，由于点云数据的稀疏性，连续两帧的点云之间并不存在一一对应关系。这会导致匹配误差，并使变形点云发生结构形变。为缓解不匹配问题，我们

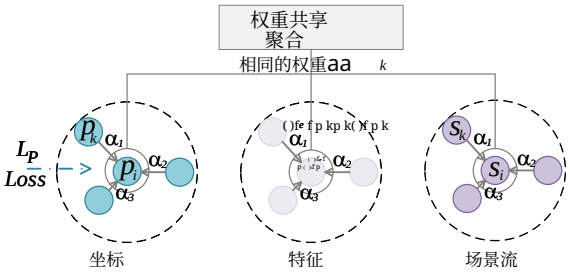

图 2. **权重共享聚合**。 在上采样过程中，点坐标、特征和场景流之间的权重保持一致。 $f_e(p_i)$ 和 s_i 是与 p_i 对应的特征和场景流。

提出了一种权重共享聚合模块，在上采样过程中利用隐式刚性约束。

在上采样过程中，权重聚合是一种常见的实现方式，它通过为较粗层中邻近点的信息分配不同的权重，来计算较细层中中心点的信息。本文设计了一种权重共享聚合模块，通过考虑坐标、特征和场景流之间的关系，隐式地施加刚性约束。

根据刚性约束（同一物体上的点具有相同的运动（R，t））以及场景流的定义，每个点对应的场景流聚合过程应与该点的坐标聚合过程保持一致，即采用权重共享，如公式（1-6）所示。此外，特征是与场景流和点坐标密切相关的中间量，因此我们将权重共享扩展到特征层级，即公式（7）。具体实现中，WSA模块应用于细化阶段的上采样过程。在权重共享之外，还加入了坐标损失公式（18）来约束聚合权重。这将权重方程求解问题转化为回归问题，从而确保插值坐标与真实值保持一致，进而使对应的场景流接近目标值。

量与场景流和点坐标密切相关，因此我们将权重共享扩展到特征层级，即公式（7）。具体实现中，WSA模块应用于细化阶段的上采样过程。除了权重共享之外，还加入了坐标损失公式（18）来约束聚合权重。这将权重方程求解问题转化为回归问题，从而确保插值坐标与真实值保持一致，进而使对应的场景流接近目标值。

通过引入点坐标的聚合过程，权重共享聚合有助于约束特征的聚合过程。此外，它包含这样的条件：同一物体上的点遵循相同的运动，并隐式地施加刚性约束，从而减轻不匹配的影响。大多数现有方法 [29, 30] 不涉及点坐标上采样，而是分别对特征和场景流执行独立的上采样过程。

权重共享聚合约束。 在点具有相同运动的局部刚性区域内（R, t），如果聚合权重 α_k 同时满足公式 (1) 和公式 (2)，则可以推导出与场景流聚合相关的公式 (3)。

条件：

$$\sum_{k=1}^{K} \alpha_k = 1 \tag{1}$$

$$\sum_{k=1}^{K} \alpha_k p_k = p_i \tag{2}$$

结论：

$$\sum_{k=1}^{K} \alpha_k s_k = s_i \tag{3}$$

其中， $p_k \in N(p_i)$、 $N(p_i)$定义为由其与p_i之间的距离确定的 p_i的K个最近邻点组成的集合。常数 K 可根据需要设为任意值，表示邻域的大小。 s_i 是点 p_i的预测场景流。

证明：
利用 R、 t 表示 p_k 的运动，如公式所示。(4).

$$T(p_k) = Rp_k + \mathrm{t} \tag{4}$$

根据场景流的定义， S的聚合可明确表示为：

$$s_k = T(p_k) - p_k = (R - I)p_k + \mathrm{t} \tag{5}$$

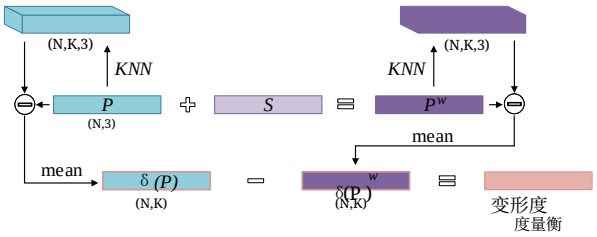

图3。**变形程度模块**。上半部分展示了点与邻域点之间邻域关系的计算过程。随后，我们比较源点云的 $\delta(P)$与变形点云的 $\delta(P^w)$之间的邻域关系。

$$\begin{aligned}\sum_{k=1}^{K} \alpha_k s_k &= \sum_{k=1}^{K} \alpha_k \{(R - I)p_k + \mathrm{t}\} \\ &= (R - I)\sum_{k=1}^{K} \alpha_k p_k + t\sum_{k=1}^{K} \alpha_k \\ &= (R - I)p_i + t \\ &= s_i\end{aligned} \tag{6}$$

我们推断，作为点云与场景流之间中间变量的特征，应具有匹配一致性关系。

对应公式如下：

$$\sum_{k=1}^{K} \alpha_k f_e(p_k) = f_e(p_i) \tag{7}$$

其中， $f_e(p_i)$表示场景流估计器针对点 p_i输出的特征。

3.3. 变形度模块

除了根据场景流的定义使用（R， t）来施加刚性约束外，另一种实现刚性约束的方法是在场景流估计任务中保持局部结构不变性。拉普拉斯正则化 [15,16,36]利用拉普拉斯坐标向量表示局部结构，该向量通过计算中心点与邻域点之间方向向量的平均值来得到。随后，通过添加损失，约束变形后的点 P^w 与 Q具有相同的局部结构。

我们提出了一种名为变形程度模块的新型模块，用于保持局部刚性结构不变性，从而提供更强的损失约束。首先，我们构造变量 δ 来表示局部结构，计算中心点与局部刚性区域中 K 个最近邻域点之间的距离。不同于拉普拉斯正则化，我们保持 K 的维度不变，以强制中心点与其邻域点之间建立一一对应的结构关系。其次，我们计算 δ_{DD} 与 δ（P）和 δ（P^w）之间的差异。其中存在

这样一种先验知识：即使场景中的物体发生运动，局部刚性区域内任意两点之间的距离也应保持不变。因此，我们施加损失，使差异 δ_{DD} 收敛至 0，以保持局部结构；同时将 δ_{DD} 输入场景流估计器，以补充几何信息。通过这种方式，我们利用更强的先验信息来减少刚体的局部形变。

$$p_i^w = p_i + s_i \tag{8}$$

$$\delta(P) = \left\{ \frac{1}{C} |p_k - p_i| \mid p_k \in N(p_i) \right\} \tag{9}$$

$$\delta(P^w) = \left\{ \frac{1}{C} |p_k^w - p_i^w| \mid p_k^w \in N(p_i^w) \right\} \tag{10}$$

$$\delta_{DD} = |\delta(P) - \delta(P^w)| \tag{11}$$

其中，p_i^w表示变形后的点。$N(p_i)$、$N(p_i^w)$分别表示以 p_i 和 p_i^w 为中心的邻域。C 表示通道数。

3.4. 网络架构

我们采用 PointPWC-Net [36]作为基线，并为其配备我们提出的模块，同时修改网络结构。整个流程采用由粗到精结构：首先在最低分辨率下估计场景流，然后逐步细化至最高分辨率。主要处理步骤包括基于权重共享聚合的上采样层、变形层、代价体、变形程度模块和场景流估计器。具体实现细节如下。

特征金字塔。我们采用 FlowNet3D [17] 提出的 $set\ conv$ 层来编码特征，并通过最远点采样将 $l-1$层下采样至 l 层。我们的金字塔结构包含五个层级 $\{l_0 - l_4\}$，通过增加一个用于在最粗分辨率下进行估计的层级（输入尺度的$1/128$），避免造成过多的内存消耗。对于l_3层，增加一个更粗的层级以为代价体构建提供场景流初始值，比直接在该层进行操作更好。

上采样层中的 WSA。上采样层是将坐标、特征和场景流从粗层插值到较细层的过程。我们的 WSA模块对这三个组成部分均采用一致的加权聚合方法，以下以特征为例进行说明。对于较细层级 $l-1$ 中的每个点 p_i^{l-1}，我们选择其在粗层级 l 中的 K 个最近邻 p_k^l，记为 $p_k^l \in N(p_i^{l-1})$。特征的上采样过程可表述如下。

$$\alpha_k = \mathrm{softmax}\left(\mathrm{mean}\left(\mathrm{MLP}\left(\left[p_k^l - p_i^{l-1}, f_e\left(p_k^l\right)\right]\right)\right)\right) \tag{12}$$

$$f_e\left(p_i^{l-1}\right) = \sum_{k=1}^{K} \alpha_k f_e\left(p_k^l\right) \tag{13}$$

其中，$[\cdot,\cdot]$ 表示拼接操作。mean () 表示在通道维度上执行平均操作。$f_e\left(p_i^l\right)$ 表示点 p_i^l的场景流估计器输出特征。

随后，我们复用 α_i 对坐标和场景流进行上采样，并加入坐标损失。其他方法不涉及点坐标上采样，且通常将特征和场景流的上采样过程分开执行。**变形层**。我们使用来自较粗层的上采样场景流，对较细层中的点云 p_i 进行变形。

$$\left(p_i^w\right)^{l-1} = p_i^{l-1} + \mathrm{up}\left(s_i^l\right) \tag{14}$$

其中，up () 表示 WSA 上采样过程。

代价体。我们在变形后的源点 P^w 与目标点 Q之间构建代价体，从而缩小搜索区域。我们采用块到膨胀块代价体来扩大感受野，遵循 RMS-FlowNet [1]。

变形程度模块。首先，计算两点之间的距离，以表示P和 P^w的局部结构。其次，测量 $\delta(P)$ 与变形后的 $\delta(P^w)$ 之间的差异，即变形程度模块。需要注意的是，δ_{DD} 不包含可学习参数，并将作为几何信息的补充输入估计器。

场景流估计器。场景流估计器的输入由五个部分组成：点云 p_i 的特征 $f^{l-1}(p_i)$、代价体 CV_i、变形度模块 δ_{DD_i}、粗层估计器的上采样输出特征 up $\left(f_e^l\right)$，以及上采样场景流 up $\left(s^l\right)$。受 DenseNet [13]，启发，我们在场景流估计器中采用特征复用和旁路设置。估计器的输出特征 $f_e^{l-1}(p_i)$ 表示流运动信息，预测场景流如下。

$$\begin{aligned} f_e^{l-1}(p_i) = \mathrm{MLP}\Big(\Big[&f^{l-1}(p_i), CV_i, \delta_{DD_i}, \\ &\mathrm{up}\left(f_e^l(p_i)\right), \mathrm{up}\left(s_i^l\right)\Big]\Big) \end{aligned} \tag{15}$$

$$s_i^{l-1} = FC\left(f_e^{l-1}(p_i)\right) \tag{16}$$

其中，$f_e^{l-1}(p_i)$ 表示估计器的输出特征。s_i^{l-1} 表示预测场景流。$[\cdot,\cdot]$ 表示拼接操作。up() 表示 WSA 上采样过程。FC 表示全连接层。

数据集	方法	三维端点误差(m)↓	严格三维准确率↑	宽松三维准确率↑	三维离群点↓	二维端点误差↓	二维准确率↑
FlyingThings3D [19]	PointPWC-Net [36]	0.0588	0.7379	0.9276	0.3424	3.2390	0.7994
	RMS-FlowNet [1]	0.0560	0.7920	0.9550	0.3240	-	-
	HCRF-Flow [16]	0.0488	0.8337	0.9507	0.2614	2.5652	0.8704
	PV-RAFT [34]	0.0461	0.8169	0.9574	0.2924	-	-
	FlowStep3D [15]	0.0455	0.8162	0.9614	0.2165	-	-
	RCP [10]	0.0403	0.8567	0.9635	0.1976	-	-
	Bi-PointFlowNet [3]	0.0280	0.9180	0.9780	0.1430	1.5820	0.9290
	3DFlow [29]	0.0281	0.9290	0.9817	0.1458	1.5229	0.9279
	Ours	**0.0239**	**0.9391**	**0.9821**	**0.1103**	**1.3703**	**0.9358**
KITTI [21]	PointPWC-Net [36]	0.0694	0.7281	0.8884	0.2648	3.0062	0.7673
	RMS-FlowNet [1]	0.0530	0.8180	0.9380	0.2030	-	-
	HCRF-Flow [16]	0.0531	0.8631	0.9444	0.1797	2.0700	0.8656
	PV-RAFT [34]	0.0560	0.8226	0.9372	0.2163	-	-
	FlowStep3D [15]	0.0546	0.8051	0.9254	0.1492	-	-
	RCP [10]	0.0481	0.8491	0.9448	**0.1228**	-	-
	Bi-PointFlowNet [3]	0.0300	0.9200	0.9600	0.1410	1.0560	0.9490
	3DFlow [29]	0.0309	0.9047	0.9580	0.1612	1.1285	0.9451
	Ours	**0.0277**	**0.9209**	**0.9613**	0.1350	**0.9773**	**0.9574**

表 1. **FlyingThings3D 和 KITTI Scene Flow 2015 数据集上的定量结果**。 所有列出的方法均以全监督方式仅在 FlyingThings3D 数据集上训练。最佳结果以粗体标出。

3.5. 损失

整个系统采用全监督方式，使用多尺度损失 $\mathcal{L}_S$，与 [1, 3, 36] 相同。 $\mathcal{L}_S$ 可表示为:

$$\mathcal{L}_S = \sum_{l=0}^{L} \gamma^l \sum_i \left\| \widehat{s}^l - s^l_{gt} \right\|_2 \tag{17}$$

其中， γ^l 表示每个金字塔层 l的权重。预测场景流记为 $\widehat{s}^l$。而 $||*||_2$指 L_2-范数。

为了施加权重共享聚合约束，我们使用 $\mathcal{L}_P$ 损失，通过使聚合坐标收敛到真实坐标来间接约束权重。

$$\mathcal{L}_P = \sum_{l=0}^{L} \gamma^l \sum_i \left\| \widehat{p^l} - p^l_{gt} \right\|_2 \tag{18}$$

其中 p^l 表示聚合坐标。

此外，我们引入变形程度损失 $\mathcal{L}_{DD}$，以在点云移动过程中尽可能保持几何结构。 $\mathcal{L}_{DD}$ 可表示为:

$$\mathcal{L}_{DD} = \sum_{l=0}^{L} \gamma^l \sum_i \left\| \delta^l_{DD} \right\|_2 \tag{19}$$

其中 δ^l_{DD} 表示变形程度模块。

总损失 $\mathcal{L}_{all}$ 由 $\mathcal{L}_S$、 $\mathcal{L}_P$和 $\mathcal{L}_{DD}$组成。

$$\mathcal{L}_{all} = \alpha_S \mathcal{L}_S + \alpha_P \mathcal{L}_P + \alpha_{DD} \mathcal{L}_{DD} \tag{20}$$

其中 α_S、 α_P 和 α_{DD} 是各项的权重。

4. 实验

本节首先介绍数据集和评估指标（§ 4.1）。接着，展示实验设置（§ 4.2）。然后，将我们的方法与其他网络进行比较（§ 4.3），以验证其有效性和泛化能力。最后，在消融研究（§ 4.4）中评估并分析我们方法的设计。

4.1. 数据集和评估指标

数据集。 为验证我们方法的有效性，我们使用合成数据集 FlyingThings3D [19] 和真实场景数据集 KITTI Scene Flow 2015 [21]，与以往方法相同。

FlyingThings3D 数据集包含训练集中的 19,640 对样本和测试集中的 3,824 对样本。我们遵循 HPLFlowNet [12] 对数据进行预处理。利用视差和光流生成点云数据，并移除深度大于 35m 的点。KITTI 数据集包含训练集中的 200 对样本和测试集中的 200 对样本。我们也按照 HPLFlowNet [12]对数据集进行预处理，并在训练集中生成 142 对点云，因为测试集的视差不可用。地面点（高度<0.3m）会被移除。

评估指标。 为进行公平比较，我们采用与 [1, 3, 36]相同的指标评估场景流。

- 三维端点误差: $\left\| \widehat{S}^l - S^l_{gt} \right\|_2$每个点的平均端点误差，单位为米。
- 三维严格准确率：三维端点误差 $< 0.05m$或相对误差 < 5%以内的点所占的百分比。
- 三维宽松准确率：误差满足三维端点误差 $< 0.1m$的点所占的百分比

MN	DD	WSA	三维端点误差(m)↓	严格三维准确率↑	宽松三维准确率↑	三维离群点↓	二维端点误差↓	二维准确率↑
✓			0.0282	0.9308	0.9799	0.1317	1.7321	0.9273
✓	✓		0.0279	0.9264	0.9786	0.1410	1.6202	0.9223
✓		✓	0.0243	0.9387	0.9819	0.1162	1.3868	**0.9358**
✓	✓	✓	**0.0239**	**0.9391**	**0.9821**	**0.1103**	**1.3703**	**0.9358**

表2. **FlyingThings3D数据集上的消融实验**。 我们列出了由所提出的不同模块组成的各种设计。“MN”表示改进网络。“DD”表示变形程度模块。“WSA”表示权重共享聚合模块。“✓”表示使用该模块。最佳结果以粗体标出。

方法	参数 (M)	浮点运算次数 (G)	运行时间 (ms)
PointPWC-Net [36]	7.7	14.4	77.5
Bi-PointFlowNet [3]	7.9	13.3	82.5
3DFlow [29]	4.8	24.1	53.8
3DFlow [29](更改)	4.8	25.6	127.8
Ours	3.7	15.5	86.4

表 3. **计算资源消耗和运行时间**。结果是在单个 NVIDIA RTX 3090 图形处理器上评估的。

或相对误差 $< 10\%$。

• 三维离群点：具有三维端点误差$> 0.3m$或相对误差 $> 10\%$ 的点所占百分比。

• 二维端点误差：通过反向投影到图像平面得到的二维平均端点误差。

• 二维准确率：满足二维端点误差 $< 3px$ 或相对误差小于 $< 5\%$ 的点所占的百分比。

4.2. 实验设置

我们在 NVIDIA RTX 3090 图形处理器上进行实验。我们在 FlyingThings3D 训练数据集上进行训练。我们的网络输入仅为连续两帧的点坐标，并通过随机采样将输入大小设为 8192。为加快训练速度，训练过程分为两个阶段。首先，我们在训练集的四分之一（4910 对）上训练模型，然后在完整训练集上进行微调。在第一阶段，学习率设为 0.001，每 20 个训练轮次的衰减率设为 0.7。预训练进行 80 个训练轮次。在第二阶段，学习率设为 0.000343，衰减率与之前相同。加载预训练模型后，进行 160 个训练轮次的微调。Adam优化器的参数设为 $\beta_1 = 0.9$、 $\beta_2 = 0.99$、$weight\ decay = 0.0001$。损失的权重 γ^l 设为 $\gamma^0 = 0.02$、 $\gamma^1 = 0.04$、 $\gamma^2 = 0.08$、 $\gamma^3 = 0.16$ 和$\gamma^4 = 0.16$。我们在 FlyingThings3D 测试数据集上进行评估，以证明该方法的有效性。此外，为验证我们方法的泛化能力，我们在未经微调的情况下于 KITTI 数据集上测试模型。

4.3. 主要结果

我们在 FlyingThings3D [19] 和 KITTI 场景流 2015 [21]数据集上与已发表的最先进方法进行比较，定量结果如表 1 所示。在所列方法中，HCRF-Flow [16] 仅在 FlyingThings3D 训练数据集的四分之一上进行训练，其他方法则在完整训练数据集上进行训练。RMS-FlowNet [1]对多层采用随机采样机制，其他方法则采用最远点采样机制。此外，与其他方法评估 8192 个点的场景流结果不同，3DFlow [29]仅评估 2048 个点。

在 FlyingThings3D 数据集上，我们的方法在所有评估指标上均优于以往的最先进方法，证明了我们方法的有效性。在 EPE3D 指标上，我们的方法比当前最先进方法 Bi-PointFlowNet [3]高出 14.6%。在 EPE3D 指标上，我们的方法比采用直接刚性约束的 HCRF-Flow [16] 高出 51.0%。

在 KITTI Scene Flow 2015 数据集上，我们在不进行微调的情况下评估模型，以验证其泛化能力。在 EPE3D 指标上，我们的方法比以往最先进方法 Bi-PointFlowNet [3]高出 7.6%。

图4中的可视化结果表明，在 FlyingThings3D 和 KITTI Scene Flow 2015 数据集上，我们的方法比近期最先进方法 [3, 29]具有更高的准确率。我们还展示了局部细节，以便于观察。在一些具有挑战性的区域，我们仍取得了良好的结果（红色点越少，表示错误越少）。

4.4. 消融实验

表2展示了不同设计的结果。下面我们将详细分析所提出的各个模块的有效性。

改进网络。 如表2第一行所示，我们提出的改进网络（修改后的网络）MN 的表现大幅优于基线网络 PointPWC-Net [36] 和 RMS-FlowNet [1] 。

形变程度模块。我们比较了使用和不使用形变程度模块时的结果。在改进的网络结构中加入该模块可以提升性能。即使在出色的设计（MN +WSA）中，使用形

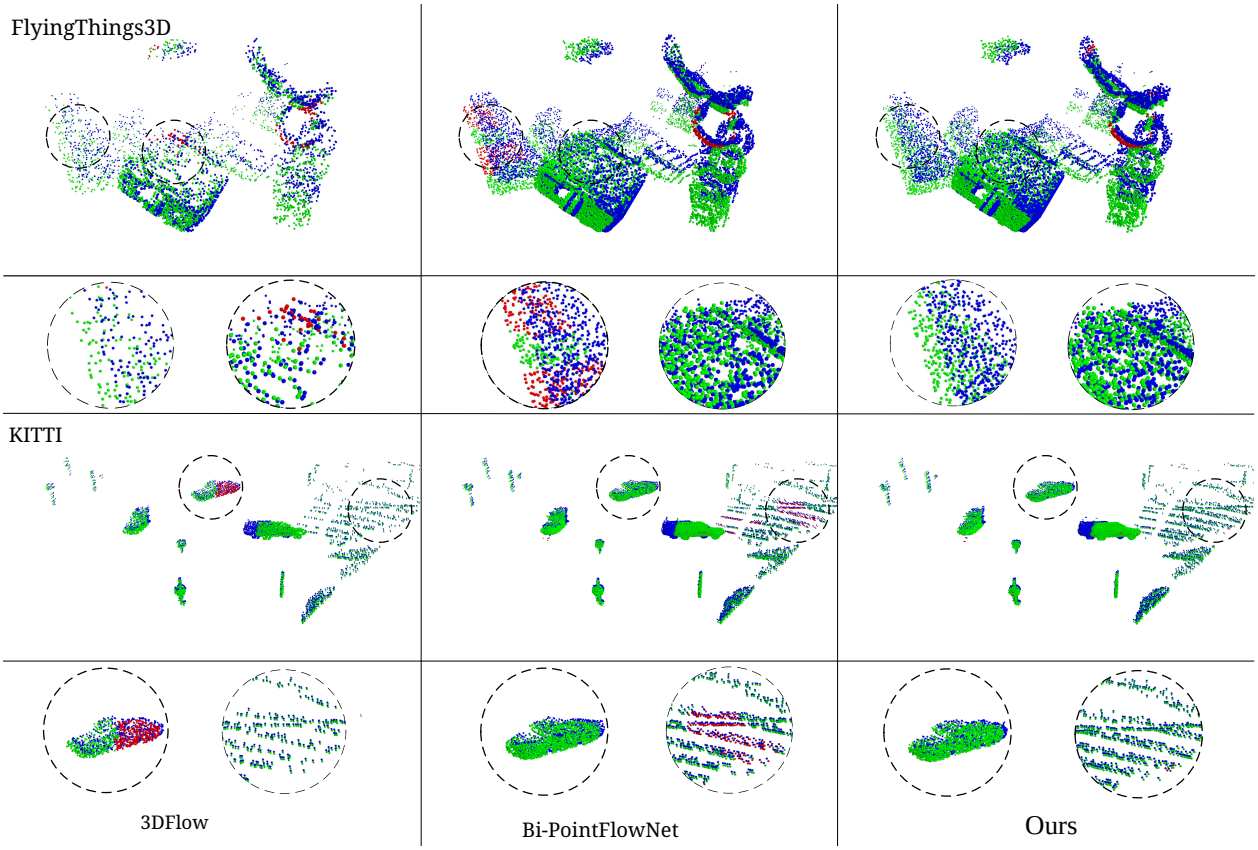

图4。**可视化结果**。从左到右依次为：3DFlow、Bi-PointFlowNet 和我们的方法。前两行是 FlyingThings3D 数据集上的实验结果，后两行是 KITTI 数据集上的实验结果。蓝色：源点。绿色：根据三维严格准确率指标，通过预测场景流得到的正确变形后的源点。红色：根据三维严格准确率指标，通过预测场景流得到的错误变形后的源点。

变程度模块（MN + DD + WSA）仍能进一步提升性能。因此，保留局部结构可以获得更准确的场景流。

权重共享聚合。对比第一行（MN）和第三行（MN + WSA）的结果可以看出，加入权重共享聚合模块后，EPE3D 指标的准确率提高了 13.8%。这证明了权重共享聚合约束的有效性，该约束使用相同的权重对点坐标、场景流和特征进行聚合。这与之前的公式推导结论（§ 3.2）一致。

根据上述不同设计的实验结果分析，我们提出的每个模块都有效。最佳方法设计由三个模块组成（修改后的网络 + 形变程度模块 + WSA）。

4.5. 计算资源消耗和运行时间

我们将计算资源消耗和运行时间与此前的最先进方法进行了比较。需要注意的是，3DFlow [29] 仅评估了 2048 个点的实验结果，而其他方法则针对每帧的 8192 个点进行测试。为确保公平比较，我们对 3DFlow 的网络进行了调整，使其能够像其他方法一样预测 8192 个点的场景流，该版本记为 3DFlow(change)。结果表明，我们的方法参数量最少。在浮点运算次数和运行时间指标上，我们取得了与其他方法相当的结果。

5. 结论

本文提出权重共享聚合约束，以间接利用刚性约束，从而避免与其他三维任务结合所导致的误差。权重共享聚合约束使特征和场景流的聚合权重与点坐标的聚合权重保持一致。我们通过公式推导和实验验证了这些约束的有效性。此外，我们还通过构建变形程度模块来保持局部几何结构不变性，该模块表示源域与目标域之间局部区域的结构差异。

该模块还为后续估计器提供几何信息，帮助其获得更精确的场景流。我们改进了由粗到细网络，并为其配备了上述模块。在FlyingThings3D和KITTI场景流数据集上开展的实验表明，我们的WSAFlowNet具有良好的有效性和泛化能力。

参考文献

[1] Ramy Battrawy、Ren´e Schuster、Mohammad–Ali Nikouei Mahani 和 Didier Stricker。Rms-flownet：一种用于大规模点云的高效、稳健多尺度场景流估计方法。发表于*2022*年国际机器人与自动化会议（*ICRA*），第883–889页，2022年。[2] Aseem Behl、Despoina Paschalidou、Simon Donne 和 Andreas Geiger。Pointflownet：学习用于从点云估计刚性运动的表征。发表于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，2019年6月。[3] Wencan Cheng 和 Jong Hwan Ko。Bi-pointflownet：基于点云的场景流估计的双向学习。发表于欧洲计算机视觉会议，第108–124页。Springer，2022年。[4] Cheng Chi、Qingjie Wang、Tianyu Hao、Peng Guo 和 Xin Yang。特征级协作：光流、立体深度和相机运动的联合无监督学习。发表于*IEEE/CVF*计算机视觉与模式识别会议论文集，第2463–2473页，2021年。[5] Marco Cuturi。Sinkhorn距离：以闪电般的速度计算最优传输。神经信息处理系统进展，第26卷，2013年。[6] David Deng 和 Avideh Zakhor。Rsf：无需标签优化三维点云中的刚性场景流。发表于*IEEE/CVF*冬季计算机视觉应用会议论文集，第1277–1286页，2023年。[7] Guanting Dong、Yueyi Zhang、Hanlin Li、Xiaoyan Sun 和 Zhiwei Xiong。利用刚性约束进行激光雷达场景流估计。发表于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，第12776–12785页，2022年6月。[8] Martin Ester、Hans-Peter Kriegel、J¨org Sander、Xiaowei Xu 等。基于密度的算法：在含噪声的大型空间数据库中发现聚类。发表于*kdd*，第96卷，第226–231页，1996年。[9] Zan Gojcic、Or Litany、Andreas Wieser、Leonidas J. Guibas 和 Tolga Birdal。刚性三维场景流的弱监督学习。发表于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，第5692–5703页，2021年6月。[10] Xiaodong Gu、Chengzhou Tang、Weihao Yuan、Zuozhuo Dai、Siyu Zhu 和 Ping Tan。Rcp：用于点云的循环最近点算法。发表于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，第8216–8226页，2022年6月。[11] Xiuye Gu、Yijie Wang、Chongruo Wu、Yong Jae Lee 和 Panqu Wang。Hplflownet：用于大规模点云场景流估计的分层置换格流网络。发表于*IEEE/CVF*计算机视觉与模式识别会议论文集，第3254–3263页，2019年。[12] Xiuye Gu、Yijie Wang、Chongruo Wu、Yong Jae Lee和Panqu Wang。HPLFlowNet：用于大规模点云场景流估计的分层置换格流网络。发表于*2019*年*IEEE*国际计算机视觉与模式识别会议（*CVPR*），2019年。[13] Gao Huang、Zhuang Liu、Laurens Van Der Maaten和Kil-ian Q Weinberger。密集连接的卷积网络。发表于*IEEE*计算机视觉与模式识别会议论文集，第4700–4708页，2017年。[14] Wolfgang Kabsch。求解两个向量集之间最佳旋转的方案。《晶体学报》*A*辑：晶体物理学、衍射、理论与综合晶体学，32(5):922–923，1976年。[15] Yair Kittenplon、Yonina C. Eldar和Dan Raviv。Flow-step3d：用于自监督场景流估计的模型展开。发表于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，第4114–4123页，2021年6月。[16] Ruibo Li、Guosheng Lin、Tong He、Fayao Liu和Chunhua Shen。HCRF-Flow：利用连续高阶条件随机场和位置感知流嵌入，从点云估计场景流。发表于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，第364–373页，2021年6月。[17] Xingyu Liu、Charles R. Qi和Leonidas J. Guibas。FlowNet3D：在三维点云中学习场景流。发表于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，2019年6月。[18] D Man和A Vision。对人类表征及视觉信息处理的计算研究。*WH San Francisco*：*Freeman and Company*，旧金山，1982年。[19] Nikolaus Mayer、Eddy Ilg、Philip Hausser、Philipp Fischer、Daniel Cremers、Alexey Dosovitskiy和Thomas Brox。用于训练卷积网络以估计视差、光流和场景流的大型数据集。发表于*IEEE*计算机视觉与模式识别会议（*CVPR*）论文集，2016年6月。[20] Moritz Menze和Andreas Geiger。用于自动驾驶车辆的物体场景流。发表于*IEEE*计算机视觉与模式识别会议（*CVPR*）论文集，2015年6月。[21] Moritz Menze、Christian Heipke和Andreas Geiger。物体场景流。《摄影测量与遥感国际学报》，140:60–76，2018年。[22] Jean-Philippe Pons、Renaud Keriven和Olivier Faugeras。利用全局图像匹配分数进行多视图立体重建和场景流估计。《国际计算机视觉杂志》，72(2):179–193，2007年。[23] Jhony Kaesemodel Pontes、James Hays和Simon Lucey。利用或不利用学习，从点云估计场景流。发表于*2020*年国际三维视觉会议（*3DV*），第261–270页，2020年。[24] Gilles Puy、Alexandre Boulch和Renaud Marlet。FLOT：由最优传输引导的点云场景流。发表于

欧洲计算机视觉会议，第527–544页。Springer，2020年。[25] Charles R Qi、Hao Su、Kaichun Mo和Leonidas J Guibas。PointNet：用于三维分类和分割的点集深度学习。发表于*IEEE*计算机视觉与模式识别会议论文集，第652 660页，2017 – 。[26] Charles Ruizhongtai Qi、Li Yi、Hao Su和Leonidas J Guibas。PointNet++：在度量空间中的点集上进行深度层次特征学习。发表于神经信息处理系统进展，第30卷，2017年。[27] Julian Quiroga、Thomas Brox、Fr´ed´eric Devernay和James Crowley。基于RGB-D图像的密集半刚性场景流估计。发表于欧洲计算机视觉会议，第567–582页。Springer，2014年。[28] Huihsin Tseng、Pi-Chuan Chang、Galen Andrew、Dan Juraf-sky和Christopher D Manning。用于SIGHAN bakeoff 2005的条件随机场分词器。发表于第四届*SIGHAN*中文处理研讨会论文集，2005年。[29] Guangming Wang、Yunzhe Hu、Zhe Liu、Yiyang Zhou、Masayoshi Tomizuka、Wei Zhan和Hesheng Wang。三维场景流网络中哪些因素至关重要。发表于欧洲计算机视觉会议，第38–55页。Springer，2022年。[30] Guangming Wang、Xinrui Wu、Zhe Liu和Hesheng Wang。在三维点云中进行场景流的层次注意力学习。*IEEE*图像处理汇刊，30:5168–5181，2021年。[31] Haiyan Wang、Jiahao Pang、Muhammad A. Lodhi、Yingli Tian和Dong Tian。FESTA：通过时空注意力估计场景点云的流。发表于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，第14173–14182页，2021年6月。[32] Zirui Wang、Shuda Li、Henry Howard-Jenkins、Victor Prisacariu和Min Chen。FlowNet3D++：用于深度场景流估计的几何损失。发表于*IEEE/CVF*冬季计算机视觉应用会议，第91–98页，2020年。[33] Andreas Wedel、Thomas Brox、Tobi Vaudrey、Clemens Rabe、Uwe Franke和Daniel Cremers。用于三维运动理解的立体场景流计算。《国际计算机视觉杂志》，95(1):29–51，2011年。[34] Yi Wei、Ziyi Wang、Yongming Rao、Jiwen Lu和Jie Zhou。PV-RAFT：用于点云场景流估计的点-体素相关场。发表于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，第6954–6963页，2021年6月。[35] Wenxuan Wu、Zhongang Qi和Li Fuxin。PointConv：在三维点云上的深度卷积网络。发表于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，2019年6月。[36] Wenxuan Wu、Zhi Yuan Wang、Zhuwen Li、Wei Liu和Li Fuxin。PointPWC-Net：用于（自）监督场景流估计的点云代价体。发表于欧洲计算机视觉会议，第88–107页。Springer，2020年。[37] Zi Jian Yew和Gim Hee Lee。RPM-Net：利用学习特征进行鲁棒点匹配。发表于论文集*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*），2020年6月。[38] 翟光耀、孔鑫、崔锦浩、刘勇和杨振。Flowmot：通过场景流关联进行三维多目标跟踪。*arXiv*预印本 *arXiv:2012.07541*，2020年。