

基于生成模型的三维波束形成图像压缩方法

赵昀杰 贺岩松* 张志飞 徐中明

重庆大学机械与运载工程学院,重庆,400044

摘要:针对通道压缩方法在高压缩率下导致 DenseNet 模型定位性能显著降低的问题,提出一种基于改进向量量化变分自编码器(VQ-VAE-2)模型的三维波束形成图像压缩(3D-BFMC)方法。先利用 VQ-VAE-2 模型的层级编码器将三维波束形成图压缩为向量化局部特征矩阵,再将该矩阵输入 DenseNet 模型实现三维定位。仿真结果表明,使用 3D-BFMC 方法压缩数据训练的 DenseNet 模型在定位精度、频率泛化性能、噪声鲁棒性上均优于通道压缩方法。单声源试验验证了 3D-BFMC 方法在真实环境中的有效性和可行性。

关键词:波束形成;数据压缩;深度学习;改进向量量化变分自编码器;三维空间

中图分类号:TP391;TN912

DOI:10.3969/j.issn.1004-132X.2025.07.015

开放科学(资源服务)标识码(OSID):



3D Beamforming Map Compression Method Based on Generative Model

ZHAO Yunjie HE Yansong* ZHANG Zhifei XU Zhongming

College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing, 400044

Abstract: To address the significant degradation in localization performance of DenseNet model under high compression ratios caused by channel compression method, a 3D beamforming map compression(3D-BFMC) method was proposed based on VQ-VAE-2 model. The hierarchical encoder of the VQ-VAE-2 model was used to compress 3D beamforming maps into vectorized local feature matrices, and then the matrices were input into the DenseNet model to perform 3D localization. Simulation results show that DenseNet models trained on compressed data by the 3D-BFMC method have better localization accuracy, frequency generalization and noise robustness than those of outperform channel compression approaches. A single-source experiment confirms the effectiveness and feasibility of 3D-BFMC in real-world environments.

Key words: beamforming; data compression; deep learning; vector quantized-variational autoencoder-2(VQ-VAE-2); three-dimensional space

0 引言

波束形成作为一种基于传感器阵列的信号处理技术,被广泛用于无线电波^[1]、地震波^[2]、声波^[3]领域。声源定位场景中,传统波束形成(conventional beamforming, CBF)通过对麦克风阵列信号的滤波计算感兴趣区域(region of interest, ROI)内的声源分布,通常的输出为二维声学图像^[4]。然而,一些实际应用需借助成像手段获取声源的三维表示^[5-6]。因此,探索在三维空间中计算波束形成结果的可能性引起学者的广泛关注。

为提高成像分辨率,有学者提出了替代 CBF 的反卷积波束形成(deconvolution beamforming, DBF)。作为 DBF 的一种方法,反卷积声源成像方法(deconvolution approach for the mapping of acoustic sources, DAMAS)应用范围从二维拓展至三维^[7]。BROOKS 等^[7]发现,尽管大尺寸阵列

具有更好的纵向聚焦能力,但纵向波束在波束形成结果中仍比横向波束宽,且 DAMAS 受限于较低的计算效率。为进一步提高三维波束形成图像的分辨率,XENAKI 等^[8]在改进反卷积声源成像方法(deconvolution approach for the mapping of acoustic sources 2, DAMAS2)的基础上引入坐标系变换,增强了点扩散函数的平移不变性,并结合替代扫描技术改善了近场条件下的三维成像结果,但该方法仍存在计算成本较高的问题。为降低三维 DBF 的计算成本,ZAN 等^[9]提出变尺度计算网格法,通过压缩网格降低数据维度,并针对关键区域实施网格加密,在一定程度上提高了定位精度。上述研究表明,尽管 DBF 具有提供高分辨率波束形成结果的潜力,但在三维成像中仍存在纵向分辨率不足、计算效率较低的问题。

近年来,结合深度学习技术进行声源定位逐渐成为研究热点。MA 等^[10]为获取兼顾精度和速度的声源定位算法,设计了以二维声源分布图

像为输出的卷积神经网络(convolutional neural network, CNN),并在高频条件下实现了高精度定位。KUJAWSKI等^[11]提出一种使用残差神经网络(residual neural network, ResNet)的无网格方法,并将其用于直接估计声源坐标和强度。XU等^[12]通过密集卷积神经网络(densely connected convolutional networks, DenseNet)模型,在特定频率下获得了超过瑞利分辨率的成像结果。LEE等^[13]构建一个目标函数来表示目标图中多个声源的位置和强度,并使用具有编码器-解码器结构的全卷积神经网络(fully convolutional network, FCN)定位平面中的多个声源。该方法不仅具有较高的分辨率,而且显著缩短了计算时间。上述研究表明,基于深度学习的声源定位方法在提高定位精度和计算效率方面具有较大潜力,但现有方法只能提供声源位置的二维表示。

为克服传统三维 DBF 的局限性,获取具有高精度和计算效率的三维声源定位算法,提出一种基于深度学习的三维无网格(three-dimensional grid-free, 3DGF)方法^[14]。然而,随着波束形成图从二维拓展至三维,图像的参数成倍增加,导致模型在训练时占用的内存显著增多。因此,在3DGF的基础上引入一种通道压缩方法,将三维图像的多个通道分组求和,可有效减少内存使用量,但不同通道之间的线性叠加会丢失大量特征信息,严重的信息损失会导致3DGF方法中DenseNet模型的定位性能显著下降。

近年来,深度生成模型(deep generative model)在数据处理领域取得显著进展^[15-16],其中,变分自编码器(variational autoencoder, VAE)^[17]在图像数据压缩中表现突出。受此启发,本研究提出一种基于改进向量量化变分自编码器(vector quantized-variational autoencoder-2, VQ-VAE-2)模型的三维波束形成图像压缩(three-dimensional beamforming map compression, 3D-BFMC)方法来应对通道压缩中信息损失的问题,提高以压缩数据为输入的DenseNet模型的声源定位性能。该方法首先训练VQ-VAE-2模型对三维波束形成图进行编码压缩,再将编码结果作为新特征输入训练DenseNet-121模型,实现三维空间中声源的高效定位。需要注意的是,与传统图像压缩不同,3D-BFMC方法关注的是压缩特征的可学习性而非视觉质量。

1 理论方法

1.1 三维波束形成

声源定位过程中,波束形成首先对ROI进行

离散化处理,形成一系列间隔为 ΔL 的网格点。随后利用相控麦克风阵列(phased microphone array, PMA)采集的声压信号对各个离散网格点进行反向聚焦,获得ROI中的声源位置。三维波束形成的实现原理如图1所示。图1包含1个拥有 I 个通道的平面PMA和1个含有 S 个($S=l \times l \times n$,其中, l 为三维ROI在 X,Y 方向的网格点数, n 为三维ROI在 Z 方向的网格点数)计算网格的三维ROI。由一系列平行于PMA平面的横截面构成的三维ROI中,网格点和PMA平面的距离记为 $z(z_a \leq z \leq z_b$,其中, z_a, z_b 分别为ROI内所有网格点到PMA平面的最小距离和最大距离)。三维波束形成的关键参数及其数值如表1所示。

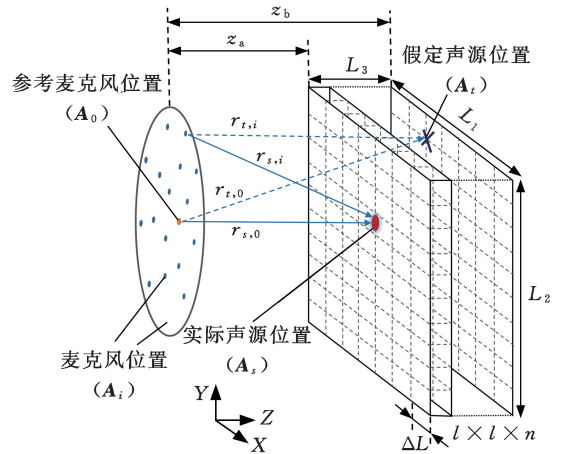


图1 三维波束形成示意图

Fig.1 The schematic of 3D beamforming

表1 三维波束形成参数

Tab.1 The parameters of 3D beamforming

参数名称	符号	数值
距离	z_a/m	0.2
	z_b/m	0.5
ROI尺寸	L_1/m	0.8
	L_2/m	0.8
	L_3/m	0.3
ROI网格点数量	S	26 896(41 × 41 × 16)
ROI网格间距	$\Delta L/m$	0.02
麦克风数量	I	18
参考麦克风位置	A_0/mm	(46, -3, 0)

假设声源为单极子源,其实际位置为 A_s ($s = 1, 2, \dots, S$),强度为 q_s 。PMA中的第 i 个麦克风在 A_i ($i = 1, 2, \dots, I$)处测量得到的声压

$$p_i = q_s h(A_i, A_0, A_s) \quad (1)$$

$$A_0 = \sum_{i=1}^I A_i \quad (2)$$

式中: $h(A_i, A_0, A_s)$ 为声音信号从 A_s 到 A_i 的传递函数; A_0 为参考麦克风位置。

$h(A_i, A_0, A_s)$ 描述了声音信号在传播过程中受到的各种影响。将式(1)写成矢量形式:

$$p = Gq \quad (3)$$

式中: G 为传递矩阵, 又称格林函数 (Green's function); q 为波束形成在 ROI 中输出的声源强度。

波束形成将 ROI 中的网格点视为假定声源位置 $A_i (i=1, 2, \dots, S)$, 则波束形成在 A_i 处的输出为

$$b(A_i) = \mathbf{v}^H(A_i) \mathbf{C} \mathbf{v}(A_i) / I^2 \quad (4)$$

其中, $\mathbf{v}(A_i)$ 为转向向量, 用于将 PMA 采集的声压信号反向聚焦至 A_i 处; \mathbf{C} 为互谱矩阵, $\mathbf{C} = \mathbf{p} \mathbf{p}^H$, 上标 H 表示共轭转置。文献[18]详细分析了 4 种不同转向向量的性能, 发现 Formulation I 和 Formulation IV 更适合三维波束形成的实际应用。本研究强调声源定位的精确性, 所以将 Formulation IV 作为转向向量。Formulation IV 的数学表达式为

$$\mathbf{v}_i(A_i) = \frac{\exp(-jk_{BF}(r_{i,i} - r_{i,0}))}{r_{i,i} \sqrt{I \sum_{i'=1}^I \frac{1}{r_{i,i'}^2}}} \quad (5)$$

式中: $\mathbf{v}_i(A_i)$ 为第 i 个麦克风对应的转向向量; $r_{i,i}$ 、 $r_{i,0}$ 分别为 A_i 到 A_i 和 A_0 的欧氏距离, $r_{i,i} = |A_i - A_i|$, $r_{i,0} = |A_i - A_0|$; k_{BF} 表示波束, $k_{BF} = 2\pi f/c$; f 为分析频率; c 为声速, 一般取 340 m/s。

根据式(4) 计算波束形成在所有网格点上的输出, 即得到声源分布图像。本研究将 CBF 方法获得的声源分布图像称为传统波束形成图 (conventional beamforming map, CB Map)。

1.2 3DGF 方法

作为一种基于深度学习的声源定位方法, 3DGF 的关键在于利用 DenseNet 模型^[19], 以预处理后的三维 CB Map 为输入, 在不依赖网格的情况下给出声源位置的三维表示。相较于 CNN, DenseNet 通过密集连接 (dense connection) 机制有效增强了特征信息的跨层级流动, 从而缓解深层神经网络中的梯度消失问题^[20]。假设密集块第 k 层 (k 表示层索引) 生成的特征图为 \mathbf{x}_k , 则密集连接可表示为

$$\mathbf{x}_k = \mathbf{F}([\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{k-1}]) \quad (6)$$

其中, $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{k-1}$ 为密集块中第 0、1、 \dots 、 $k-1$ 层生成的特征图; $[\ast]$ 表示将各个特征图串联。

复合函数 $\mathbf{F}(\ast)$ 表示依次执行批量归一化 (batch normalization, BN)、修正线性单元 (ReLU) 激活和 3×3 卷积 (Conv) 操作。表 2 描述了用于三维定位的 DenseNet 模型的具体结构, 其中, 增长率 $\kappa = 24$, Conv* 表示 BN-ReLU-Conv 操作。该模型包含 4 个密集块和 3 个过渡层, 并使用步长为 1 的 3×3 卷积核对输入图像进行预处理。预处理层的输入通道数应根据输入 CB Map 的通道数调

整。回归层中, 自适应平均池化层将特征图的尺寸压缩为 1×1 , 然后通过全连接层输出神经网络的预测值。

表 2 DenseNet-121 模型的层结构

Tab.2 The layer structure of the DenseNet-121

网络层类型	详细结构
预处理层	3×3 Conv
密集块(1)	$[1 \times 1 \text{ Conv}^*, 3 \times 3 \text{ Conv}^*] \times 6, \kappa$
过渡层(1)	$1 \times 1 \text{ Conv}^*, 2 \times 2$ 平均池化
密集块(2)	$[1 \times 1 \text{ Conv}^*, 3 \times 3 \text{ Conv}^*] \times 12, \kappa$
过渡层(2)	$1 \times 1 \text{ Conv}^*, 2 \times 2$ 平均池化
密集块(3)	$[1 \times 1 \text{ Conv}^*, 3 \times 3 \text{ Conv}^*] \times 24, \kappa$
过渡层(3)	$1 \times 1 \text{ Conv}^*, 2 \times 2$ 平均池化
密集块(4)	$[1 \times 1 \text{ Conv}^*, 3 \times 3 \text{ Conv}^*] \times 16, \kappa$
回归层	自适应平均池化层
	全连接层

3DGF 方法的主要框架如图 2 所示。首先对三维 ROI 进行网格划分 ($l \times l \times n$), CBF 方法利用 PMA 采集到的声压信号, 获得含有 n (由三维 ROI 在 Z 向的切片数决定) 个通道的三维 CB Map 矩阵 $\mathbf{M}_{l \times l \times n}$ 。压缩 $\mathbf{M}_{l \times l \times n}$, 并对压缩后的矩阵应用最大-最小值归一化预处理, 得到输入数据 \mathbf{N}_{in} 。最后将 \mathbf{N}_{in} 输入已完成训练的 DenseNet 模型, 获得声源在笛卡儿坐标系的位置坐标 $(\hat{x}, \hat{y}, \hat{z})$ 。训练采用 Pytorch^[21] 中的随机梯度下降 (stochastic gradient descent, SGD) 优化器更新网络参数, 并以均方误差 (mean squared error, MSE) 作为损失函数。

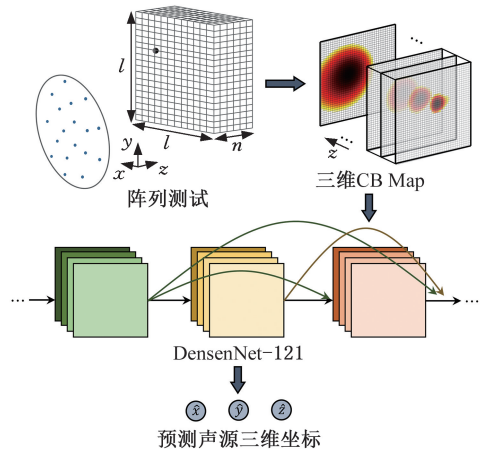


图 2 3DGF 方法框架

Fig.2 The framework of 3DGF

神经网络的训练依赖规模庞大的数据集。为高效获取高质量的训练数据, 使用声学软件 Acoular^[22] 生成指定位置处声源的时域信号, 并结合 AcouPipe^[23] 工具箱和数据预处理过程, 构建包括训练数据、验证数据和测试数据的完整数据集。表 3 展示了数据集构建的关键参数, 其中, 原始 CB Map 的尺寸为 $41 \times 41 \times 16$, 其中, 41 为

CB Map 在 X-Y 平面的高度和宽度,16 为通道数(深度)。所有样本数据中,训练数据、验证数据和测试数据的比例为 8:1:1。这 3 个数据集各自独立生成,不存在包含关系。后续针对 VQ-VAE-2 模型的训练数据生成方式与 3DGF 一致,但训练数据规模减少为 50 000 个样本(每个频率 10 000 个样本,共 5 个频率),以控制 VQ-VAE-2 模型在训练过程中所需的计算资源。

表 3 数据集构建参数

Tab.3 The parameters of the dataset construction

数据集属性	相关设置
三维 ROI 范围	$x \in [-0.4 \text{ m}, 0.4 \text{ m}]$ $y \in [-0.4 \text{ m}, 0.4 \text{ m}]$ $z \in [0.2 \text{ m}, 0.5 \text{ m}]$
采样频率 f_s/Hz	32 768
采样时间 t_s/s	5
训练频率 f/kHz	1,2,3,4,6
阵列形式	18 通道扇形阵列
原始 CB Map 尺寸	$41 \times 41 \times 16$
声源信号类型	不相关白噪声
声源位置分布	均匀分布
训练集样本数目	2×10^5 (每个训练频率 40 000 个样本)

1.3 通道压缩方法

作为 3DGF 方法中的辅助技术,通道压缩方法有效减小模型在训练过程中的内存占用量^[14]。矩阵 $\mathbf{M}_{l \times l \times n}$ 可表示为

$$\mathbf{M}_{l \times l \times n} = [\mathbf{M}_{l \times l}^{(1)} \quad \mathbf{M}_{l \times l}^{(2)} \quad \cdots \quad \mathbf{M}_{l \times l}^{(n)}] \quad (7)$$

其中, $\mathbf{M}_{l \times l}^{(i)}$ 代表矩阵 $\mathbf{M}_{l \times l \times n}$ 的第 i ($i=1,2,\dots,n$) 个通道。将矩阵 $\mathbf{M}_{l \times l \times n}$ 的 n 个通道按位置顺序划分为 m 个小组,每个小组包含 ϵ 个通道,则 $n = \epsilon m$ 。对每个小组内的 ϵ 个通道执行叠加运算,可将通道数从 n 压缩至 m 。因此,将 ϵ 定义为压缩率, ϵ 越大表示压缩程度越高。压缩后的矩阵 $\mathbf{M}_{l \times l \times m}$ 可表示为

$$\mathbf{M}_{l \times l \times m} = [\mathbf{M}_{l \times l}^{(1)} \quad \mathbf{M}_{l \times l}^{(2)} \quad \cdots \quad \mathbf{M}_{l \times l}^{(m)}] \quad (8)$$

$$\mathbf{M}_{l \times l}^{(K)} = \sum_{i=(K-1)\epsilon+1}^{K\epsilon} \mathbf{M}_{l \times l}^{(i)} \quad K = 1,2,\dots,m$$

$\epsilon=1$ 时, $\mathbf{M}_{l \times l \times m} = \mathbf{M}_{l \times l \times n}$, 原始 CB Map 矩阵没有压缩。

1.4 3D-BFMC 方法

VQ-VAE 是一种基于离散隐变量的生成模型,可无监督捕获数据的关键特征^[24]。VQ-VAE 能将图像转换为离散的潜在向量(latent vector),具有强大的特征提取能力。VQ-VAE-2 作为 VQ-VAE 的改进形式,采用多尺度层级结构进一步增强模型的表达能力^[25]。因此,将 VQ-VAE-2 模型作为 3D-BFMC 方法的基础,对原始 CB Map 样本 $\mathbf{M}_{l \times l \times n}$ 进行特征提取,以获取具有更少元素

的矩阵 \mathbf{N}_{vq} ,从而实现数据压缩。

本研究将 VQ-VAE-2 拆分为层级编码 E_{vq} 和解码 D_{vq} 两个部分。 E_{vq} 首先将输入的 CB Map 矩阵 $\mathbf{M}_{l \times l \times n}$ 映射到局部特征 \mathbf{b}_m 和全局特征 \mathbf{t}_m 上。随后,通过 2 个向量量化层(每个向量量化层的码本含有 2048 个原型向量)将 \mathbf{b}_m 、 \mathbf{t}_m 量化为离散特征 $\bar{\mathbf{b}}_m$ 和 $\bar{\mathbf{t}}_m$ 。设置向量维度为 1 的目的是将特征提取后的矩阵通道数减少至 1。最后, D_{vq} 根据离散特征 $\bar{\mathbf{b}}_m$ 和 $\bar{\mathbf{t}}_m$ 重构 CB Map 矩阵 $\tilde{\mathbf{M}}_{l \times l \times n}$ 。基于重构矩阵 $\tilde{\mathbf{M}}_{l \times l \times n}$ 和输入矩阵 $\mathbf{M}_{l \times l \times n}$, 定义 VQ-VAE-2 的重构损失:

$$L_{l2} = \|\tilde{\mathbf{M}}_{l \times l \times n} - \mathbf{M}_{l \times l \times n}\|_2^2 \quad (9)$$

L_{l2} 能衡量 $\tilde{\mathbf{M}}_{l \times l \times n}$ 与 $\mathbf{M}_{l \times l \times n}$ 的差异,评估模型的重构性能。使用编码器输出的指数移动平均(exponential moving average, EMA)更新码本中的原型向量,并引入局部特征的承诺损失 L_{bc} 和全局特征的承诺损失 L_{tc} :

$$L_{bc} = \|\mathbf{b}_m - sg[\bar{\mathbf{b}}_m]\|_2^2 \quad (10)$$

$$L_{tc} = \|\mathbf{t}_m - sg[\bar{\mathbf{t}}_m]\|_2^2 \quad (11)$$

其中, $sg(\mathbf{W})$ 代表 stop-gradient 运算,即在向前传播时保持 \mathbf{W} 不变,在反向传播时将 \mathbf{W} 的梯度强制归零,从而阻止梯度通过 \mathbf{W} 反向传播^[24]。

使用承诺损失的目的是约束编码器的输出,使其趋近于所选码本中的原型向量,减少量化过程中的信息丢失,提高模型在训练过程中的稳定性。

综上,VQ-VAE-2 的总损失定义为

$$L_{vq} = \beta_{l2} L_{l2} + \beta_c L_c \quad (12)$$

$$L_c = L_{bc} + L_{tc}$$

其中, β_{l2} 、 β_c 为重构损失 L_{l2} 和承诺损失 L_c 的权重,本研究设定 $\beta_{l2} = \beta_c = 1$ 。

3D-BFMC 方法的主要框架如图 3 所示。首先将 16 通道的原始 CB Map 作为输入数据训练 VQ-VAE-2 模型。为确保 VQ-VAE-2 模型的正常运行,在输入前必须裁剪尺寸 $41 \times 41 \times 16$ 的原始 CB Map,将其边长从 41 调整为偶数,裁剪后的尺寸为 $40 \times 40 \times 16$ 。训练完成后,提取 VQ-VAE-2 模型的层级编码部分 E_{vq} ,使用 E_{vq} 将 20 万个原始 CB Map 样本 $\mathbf{M}_{41 \times 41 \times 16}$ 全部提取为向量化局部特征 $\bar{\mathbf{b}}_m$,并作为新的输入数据 \mathbf{N}_{vq} (通道数为 1)训练 DenseNet 模型。最后依次利用 VQ-VAE-2 模型的 E_{vq} 和 DenseNet 模型,从原始 CB Map 中获取声源位置 $(\hat{x}, \hat{y}, \hat{z})$ 。输入数据 \mathbf{N}_{vq} 中的元素显著少于原始 CB Map,减少了模型在训练过程中的内存消耗,节省了数据存储空间,提高了数据的便携性和传输效率。因此,将 3D-

BFMC 称为三维图像数据压缩方法。值得注意的是,3D-BFMC 的图像压缩质量取决于下游

DenseNet 模型的定位性能,而非重构损失 L_{12} 的大小。

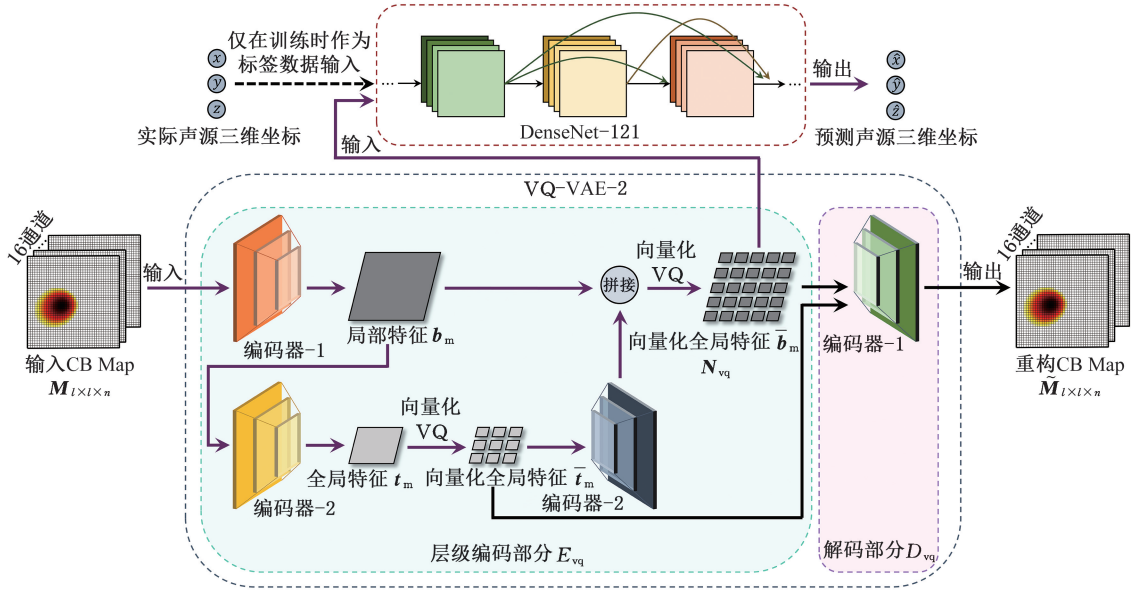


图 3 3D-BFMC 方法框架

Fig.3 The framework of 3D-BFMC

为研究不同压缩程度的输入数据 N_{vq} 对 DenseNet 模型性能的影响,选取具有不同尺寸向量化局部特征 \bar{b}_m 的 VQ-VAE-2 模型 VQ- α 和 VQ- β 进行训练。VQ- α 的编码器 1 采用 1×1 卷积将输出特征图的通道数压缩为 1,该特征图的尺寸为 40×40 ,与输入 CB Map 在 X-Y 平面的投影相同。VQ- β 的编码器 1 采用一个转置卷积,输出特征图的尺寸扩大为 80×80 。将 VQ- α 、VQ- β 的层级编码部分记为 $E_{vq-\alpha}$ 和 $E_{vq-\beta}$,所提取的向量化局部特征记为 \bar{b}_m-40 和 \bar{b}_m-80 (对应尺寸分别为 40×40 和 80×80)。与原始 CB Map 相比, \bar{b}_m-40 、 \bar{b}_m-80 中的元素数量分别减小 94% 和 76%。

VQ-VAE-2 模型的训练采用 Pytorch^[21] 中的 Adam^[26] 优化器更新网络参数,设置初始学习率为 0.001,批量大小为 16,以节省内存使用量。为确保模型充分收敛,设置 epoch 为 200。使用余弦退火学习率调度策略^[27],使学习率在 epoch 范围内按照余弦曲线逐步衰减至 0。下游 DenseNet 模型的网络结构和参数设置与 3DGF 方法相同。

2 仿真结果与对比分析

上述研究表明,3D-BFMC 方法和通道压缩方法均能显著减少训练样本的参数。为展示 3D-BFMC 方法在数据压缩的优越性,仿真分别使用 3D-BFMC 和通道压缩方法进行数据压缩,并对

比下游 DenseNet 模型的定位误差和噪声鲁棒性。数据压缩的首要目的是降低训练内存消耗,因此仿真首先分析上游 VQ-VAE-2 模型的计算资源使用情况,再对比下游 DenseNet 模型的资源使用情况,最后选择内存占用相似的 DenseNet 模型进一步比较三维定位性能。

2.1 计算资源

训练时,VQ-VAE-2 与 DenseNet 模型没有关联,因此,需分别评估 VQ-VAE-2 和 DenseNet 模型在训练过程中的计算资源消耗。本研究的所有工作均在配备 Intel i7-14700KF 和 GeForce 4070 Ti 显卡的计算机上完成。表 4 展示了 VQ- α 和 VQ- β 在训练期间所占用的内存 M_C 和显存 M_G 。模型在显卡上训练,结果表明,VQ- α 和 VQ- β 的内存占用量与显存占用量都在可接受范围内。VQ- β 拥有更大尺寸的特征,导致模型参数比 VQ- α 多,因此显存消耗更多,但二者的内存使用量均为 3.2 GB。这是因为除编码器 1 的第一层和解码器 1 的最后一层存在差别外,其他结构完全一致,且采用相同的训练数据和参数设置,这使得二者在内存消耗上并无显著差异。

表 4 VQ- α 和 VQ- β 在训练期间的内存和显存消耗

Tab.4 The CPU memory and GPU memory usage of VQ- α and VQ- β during training

计算资源	模型名称	占用量/GB
M_C	VQ- α	3.2
	VQ- β	3.2
M_G	VQ- α	1.5
	VQ- β	4.7

将由 \bar{b}_m-40 、 \bar{b}_m-80 组成的训练数据集分别记为 $T-40$ 和 $T-80$ ，使用 $T-40$ 、 $T-80$ 训练的 DenseNet 模型分别记为 BFMC-40 和 BFMC-80。通道压缩方法中， $\epsilon=1$ 时，CB Map 未压缩，训练数据为原始 CB Map； $\epsilon=4$ 时，压缩后的矩阵元素数量为 6724，与 \bar{b}_m-80 的元素数量相当； $\epsilon=16$ 时，压缩后的矩阵元素数量为 1681，与 \bar{b}_m-40 的元素数量相当。因此，选取 $\epsilon=1, 4, 16$ 的数据训练 DenseNet 模型。将 $\epsilon=1, 4, 16$ 训练得到的模型分别记为 3DGFM-1、3DGFM-4 和 3DGFM-16。3DGFM-1 在训练时使用未压缩数据，因此其性能用作参照。BFMC-80 的训练数据尺寸为 80×80 ，导致训练模型所需的计算资源增多，训练效率显著下降，因此将 BFMC-80 的预处理层替换为步长为 2、填充为 3 的 7×7 卷积核，从而将 80×80 的输入数据缩小至 40×40 。除此之外，模型其余部分结构完全相同，仅输入数据存在差异。

表 5 展示了 5 种 DenseNet 模型在训练期间所占用的 M_C 和 M_G 。3D-BFMC 和通道压缩方法均能有效减少模型训练时的内存占用量。BFMC-40 和 3DGFM-16 的内存占用量基本一致，约为 3DGFM-1 的 19%；BFMC-80 和 3DGFM-4 的内存占用量基本一致，约为 3DGFM-1 的 35%。训练采用小批量数据，且不同样本之间的数据大小相差不到 100 kB，因此含有不同数量参数的样本在训练时并不会对显存占用量产生较大影响。这 5 种模型的参数数量非常接近，最终的显存占用量大致相同。后续讨论中，对比具有相似内存占用量的 BFMC-40 和 3DGFM-16 模型，以及 BFMC-80 和 3DGFM-4 模型的性能，以进一步分析 3D-BFMC 的优越性。

表 5 五种 DenseNet 模型在训练期间的内存和显存消耗
Tab.5 The CPU memory and GPU memory usage of five DenseNet models during training

计算资源	模型名称	占用量/GB
M_C	BFMC-40	5.4
	BFMC-80	10.0
	3DGFM-1	28.8
	3DGFM-4	10.2
	3DGFM-16	5.5
M_G	BFMC-40	4.9
	BFMC-80	4.9
	3DGFM-1	4.9
	3DGFM-4	4.9
	3DGFM-16	4.9

2.2 定位误差

本研究采用均方根误差

$$R_{MSE} =$$

$$\sqrt{\frac{1}{N} \sum_{i=1}^N ((\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2 + (\hat{z}_i - z_i)^2)}$$

式中： N 为测试数据集中的样本总数； $(\hat{x}_i, \hat{y}_i, \hat{z}_i)$ 为模型对第 i 个样本的预测结果； (x_i, y_i, z_i) 为第 i 个样本的实际位置坐标。

作为衡量 DenseNet 模型定位精度的指标。 R_{MSE} 越小，模型的定位精度越高。在评价时， R_{MSE} 四舍五入保留至小数点后四位。

定义“已知”频率为模型训练使用的频率，“未知”频率为其他频率。表 6 展示了 5 种 DenseNet 模型在“已知”频率上的定位误差。BFMC-40 的平均定位误差比 3DGFM-16 小 7.3%。BFMC-80 的平均定位误差比 3DGFM-4 仅小 2.5%，但经接近 3DGFM-1，二者在 5 个“已知”频率上的平均差值仅为 0.1 mm。

表 6 五种 DenseNet 模型在“已知”频率上的定位误差
Tab.6 Localization errors of five DenseNet models at “known” frequencies

频率/kHz	R_{MSE}/mm				
	BFMC-40	BFMC-80	3DGFM-1	3DGFM-4	3DGFM-16
1	4.1	3.9	3.8	3.9	4.2
2	4.8	3.7	3.5	3.8	5.3
3	4.6	3.9	3.8	4.2	4.9
4	4.1	3.9	3.8	4.0	4.6
6	4.0	3.9	3.9	3.9	4.3

为进一步评估模型在“未知”频率上的泛化性能，在 1~8 kHz 范围内随机选取 10 个“未知”频率作为测试频率，其中，6.4、7.0、8.0 kHz 不在训练频率范围(1~6 kHz)内。5 种 DenseNet 模型在这些“未知”频率上的定位误差如图 4 所示。在频率泛化性能方面，相较于 3DGFM-16，BFMC-40 在 10 个“未知”频率上的平均定位误差减小 35%；相较于 3DGFM-4，BFMC-80 的平均定位误差减小 3.1%。

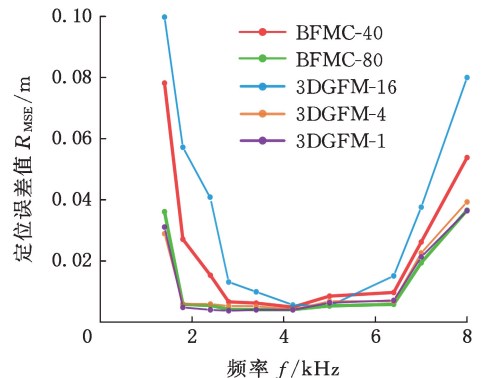


图 4 五种 DenseNet 模型在“未知”频率上的定位误差
Fig.4 Localization errors of five DenseNet models at “unknown” frequencies

综上所述,无论是“已知”频率还是“未知”频率,BFMC-40 和 BFMC-80 均表现出更高的定位精度。BFMC-40 的提升效果更显著;BFMC-80 的定位性能更接近使用原始数据为输入的 DenseNet 模型 3DGMF-1。因此,利用 3D-BFMC 压缩得到的数据能在较低维度上保留更多原始数据的特征信息,这更利于下游 DenseNet 模型的学习,获得更高的定位精度。使用 3D-BFMC 和通道压缩方法对原始数据进行压缩,训练得到的 DenseNet 模型在预测结果上的 R_{MSE} 随频率变化的趋势基本一致,即使用不同压缩方法获得的数据训练模型,不会改变模型误差在不同频率下的变化趋势,这说明影响 DenseNet 模型精度的主要因素仍是声源频率。频率相同时,数据压缩程度的增加导致 DenseNet 模型的定位精度显著下降。

2.3 噪声鲁棒性

为评估外部噪声干扰下,压缩方法对下游 DenseNet 模型定位精度的影响,在每个麦克风采集的声压信号中添加不同信噪比 S_{NR} 的随机噪声。图 5 所示为不同 DenseNet 模型的噪声鲁棒性表现。表 7 展示了 5 种 DenseNet 模型在不同 S_{NR} 的 $\bar{\delta}_{SNR}$ (特定 S_{NR} 下,模型在 5 个测试频率上的 R_{MSE} 均值)。结合图 5a、图 5b 和表 7 可知,在抵抗外部噪声干扰方面,BFMC-40 优于 3DGMF-16,BFMC-80 优于 3DGMF-4; S_{NR} 在 5~15 dB 范围内,BFMC-40、BFMC-80 的 $\bar{\delta}_{SNR}$ 分别减小 5.2% 和 18.4%。这是因为 VQ-VAE-2 模型是在无噪声的“干净”数据上训练得到的,因此在对含噪声的数据进行特征提取时,VQ-VAE-2 的编码器倾向于抑制噪声干扰,提取更接近原始“干净”数据的潜在特征。这使得 3D-BFMC 更能为下游 DenseNet 模型提供“干净”的输入数据,减小 DenseNet 模型的定位误差。BFMC-80 的抗噪性能比 BFMC-40 更优,这表明在 3D-BFMC 方法中,拥有大尺寸(80×80)特征信息的 VQ- β 模型在处理含噪声数据时比特征尺寸较小(40×40)的 VQ- α 模型更具优势。

由图 5c 可知,1~4 kHz 内, S_{NR} 较小(5~15 dB)时,BFMC-80 的定位误差明显小于 3DGMF-1。这是因为,尽管压缩会损失原始数据中的部分特征信息,但噪声对模型精度的影响“掩盖”了信息损失的影响。VQ-VAE-2 在数据压缩时能过滤部分噪声,使下游 DenseNet 模型具备更高的精度。随着 S_{NR} 的增大,噪声对模型精度的影响逐渐减弱,信息损失导致的额外误差逐渐凸显。由于模型的抗噪性能随声源频率的升高逐渐增

强,因此频率达到 6kHz 时,数据压缩产生的误差占主导地位,导致此时 BFMC-80 的定位误差略大于 3DGMF-1。

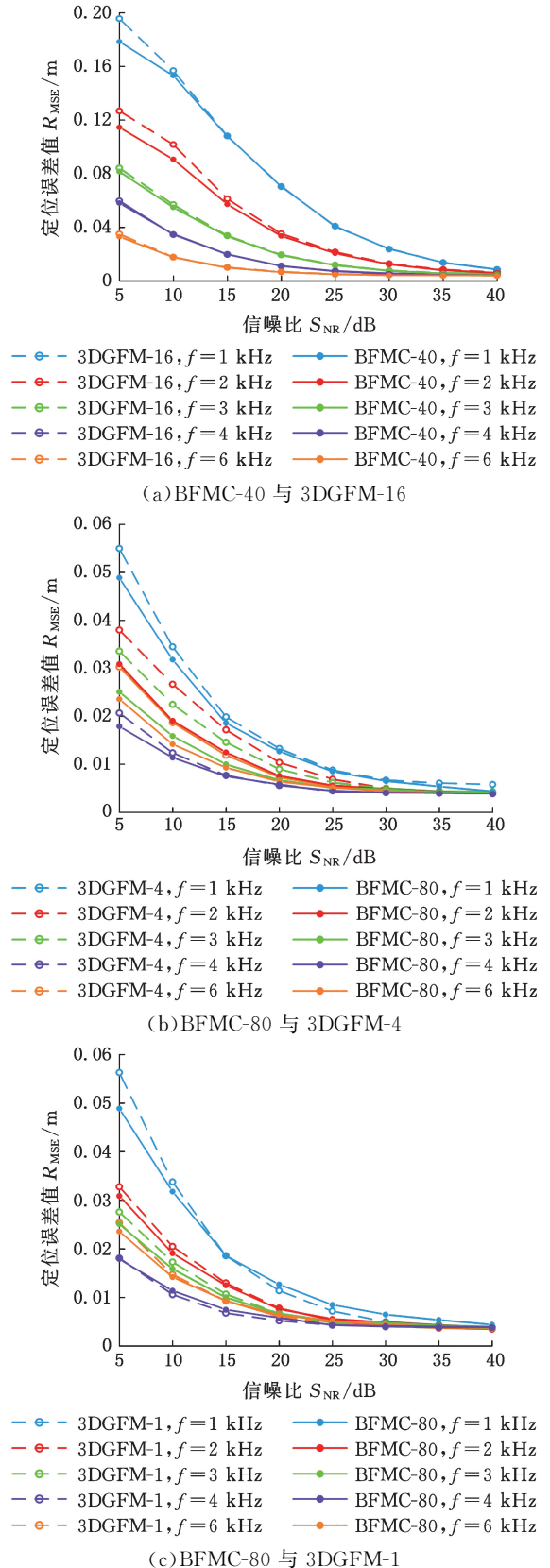


图 5 不同 S_{NR} 下 DenseNet 模型的噪声鲁棒性
Fig.5 Noise robustness of DenseNet models under different S_{NR}

表 7 5 种 DenseNet 模型的 $\bar{\delta}_{SNR}$

Tab.7 The $\bar{\delta}_{SNR}$ of 5 DenseNet models

S_{NR}/dB	$\bar{\delta}_{SNR}/mm$				
	BFMC-40	BFMC-80	3DGM-1	3DGM-4	3DGM-16
5	93.2	29.3	32.0	35.5	100.3
10	70.3	18.5	19.4	23.0	73.6
15	45.6	11.6	11.7	14.3	46.6
20	28.3	7.8	7.4	9.1	28.6
25	17.0	5.7	5.2	6.4	17.4
30	10.7	4.9	4.2	5.1	10.9
35	7.2	4.4	3.9	4.6	7.4
40	5.5	4.0	3.8	4.4	5.8

综上所述,与通道压缩方法相比,使用 3D-BFMC 压缩数据训练的 DenseNet 模型能在较小 S_{NR} 下有效提高模型的定位精度;在内存和显存消耗可接受的范围内,选择具有较大特征尺寸的 VQ-VAE-2 模型更利于抵抗外部噪声干扰,减弱噪声对下游 DenseNet 模型精度的影响。

3 试验验证

为进一步验证 3D-BFMC 方法的有效性和可行性,在普通室内环境中进行试验。试验设备包括 B&K 公司的 18 通道扇形轮形麦克风阵列

WA-1558-W-021 和 LAN-XI 数据采集器。以阵列面中心为坐标原点,扬声器布置于 (0.2, -0.1, 0.4)m 处。声源类型为随机白噪声,总采样时间为 5 s,采样频率为 32 768 Hz。试验布局如图 6 所示。



图 6 普通室内环境中的试验布局

Fig.6 The experimental layout in an indoor environment

选取“已知”频率 1 kHz、6 kHz 和“未知”频率 2.4 kHz、4.2 kHz 作为分析频率。使用 3D-BFMC 方法对 CB Map 进行压缩,并将不同压缩程度的数据分别提交给 BFMC-40 和 BFMC-80,定位试验环境中的声源。定位结果如图 7 所示,其中,云图为使用 CBF 方法计算得到的 CB Map,

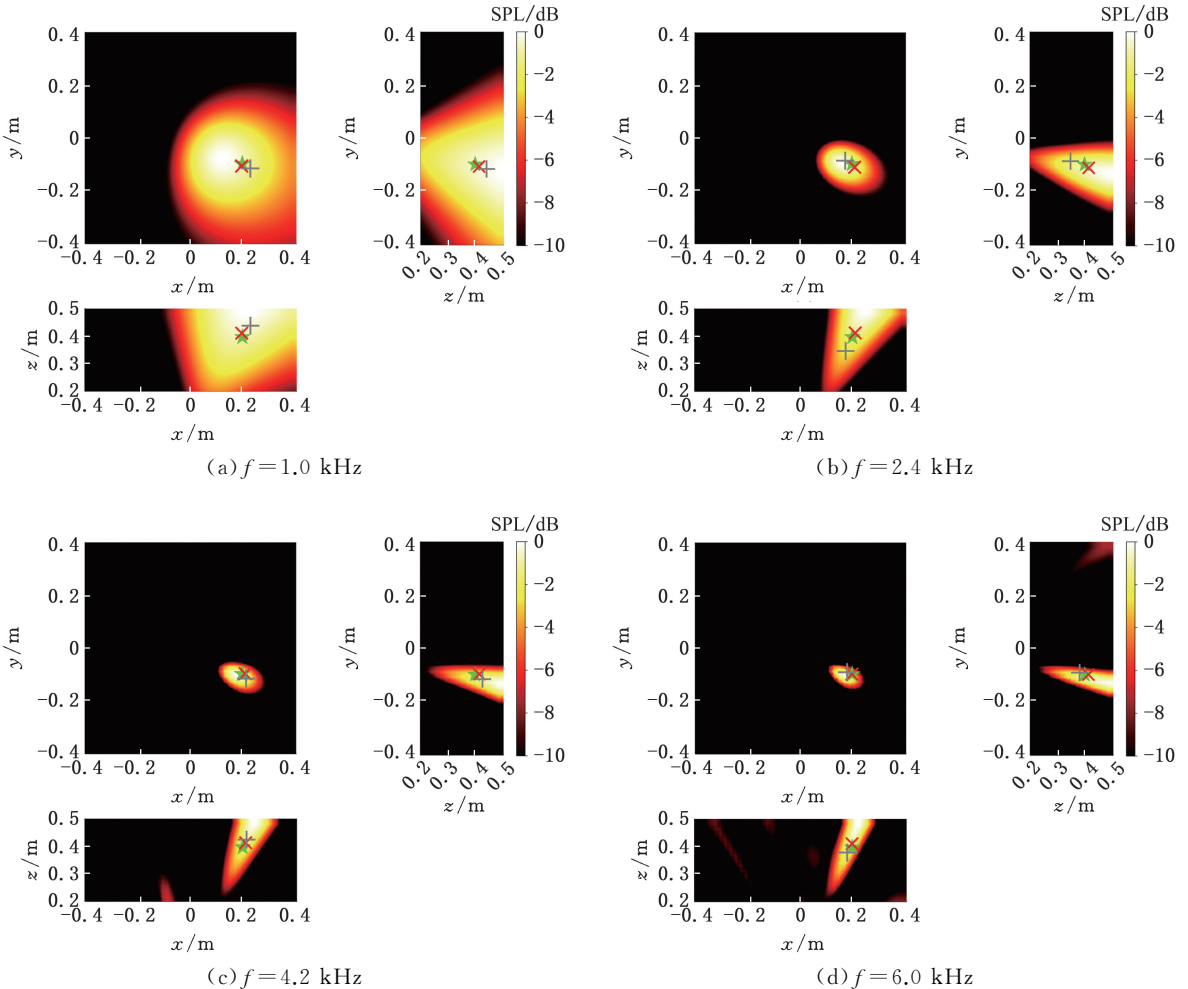


图 7 BFMC-80 和 BFMC-40 在试验中的定位结果

Fig.7 The localization results of BFMC-80 and BFMC-40 in the experiment

绿色“★”代表声源的实际位置,红色“×”代表 BFMC-80 的定位结果,灰色“+”代表 BFMC-40 的定位结果。每个分图中,左上、右上和左下的子图分别对应波束形成结果在 X-Y、Z-Y、X-Z 平面上的投影。模型在试验环境中的预测结果四舍五入保留至小数点后三位。

从图 7 中可以看出,BFMC-80、BFMC-40 均能较准确地预测出声源的坐标,但 BFMC-80 的定位精度明显高于 BFMC-40。尽管 BFMC-40 在“未知”频率 2.4 kHz 上出现较大的纵向误差,但仍在可接受范围内。为进一步对比 3D-BFMC 和通道压缩方法在试验环境中的性能,表 8 展示了 5 种 DenseNet 模型在全部试验频率下的平均定位误差 \bar{R}_e 。与 3DGFM-4 相比,BFMC-80 的定位精度显著提升, \bar{R}_e 减小 14%;与 3DGFM-16 相比,BFMC-40 的 \bar{R}_e 减小 22%,同样展现出更优的定位性能。此外,BFMC-80 的定位精度相当接近使用未压缩数据训练的模型 3DGFM-1。

表 8 5 种 DenseNet 模型的平均定位误差

Tab.8 Average localization error of 5 DenseNet models
mm

模型	平均定位误差
BFMC-40	46
BFMC-80	19
3DGFM-1	20
3DGFM-4	22
3DGFM-16	59

3D-BFMC 使用 VQ-VAE-2 模型特征提取会产生额外的计算时间,为评估不同 VQ-VAE-2 模型的计算效率,分别测试了 VQ- α 、VQ- β 在试验数据上的特征提取时间。VQ- α 、VQ- β 在 4 个试验频率上的平均特征提取时间分别约为 15 ms 和 52 ms,都具有较高的计算效率。因此,使用 3D-BFMC 方法压缩数据不会显著影响整个定位流程的计算效率。

综上所述,3D-BFMC 方法能在试验环境中对 CB Map 进行有效压缩,并拥有较高的计算效率,在实际应用中具有一定可行性。与通道压缩方法相比,使用 3D-BFMC 压缩数据训练的 DenseNet 模型能获得更高的定位精度,这与仿真结果一致。

4 结论

1) 提出的 3D-BFMC 方法能显著减少下游 DenseNet 模型在训练时占用的内存。相较于原始 CB Map,将 BFMC-40、BFMC-80 压缩的数据作为输入进行训练时,DenseNet 模型的内存消耗

分别减少 81% 和 65%。上游 VQ-VAE-2 模型的训练仅需消耗较少的计算资源,占用的内存和显存均在 4.7 GB 以内。

2) 在 3D-BFMC 方法中,与 VQ- α 相比,具有更大特征尺寸(80×80)的 VQ- β 在训练时会占用更多显存。输入数据尺寸较大的 DenseNet 模型 BFMC-80 在训练时的内存消耗大于 BFMC-40,但 BFMC-80 在定位精度、频率泛化性能、噪声鲁棒性上显著优于 BFMC-40。因此,在实际应用中,应根据任务需求选择最合适的压缩数据尺寸。

3) 3D-BFMC 方法具有较高的计算效率。压缩单个 CB Map 样本时,VQ- α 、VQ- β 平均处理时间仅为 15 ms 和 52 ms。仿真和试验验证了 3D-BFMC 方法的有效性和可行性,有助于进一步提升 3DGF 在实际场景中的应用潜力。

参考文献:

- [1] ABDELREHEEM A, MOHAMED E M, ESMAIL H. Location-based Millimeter Wave Multi-level Beamforming Using Compressive Sensing[J]. IEEE Communications Letters, 2018, 22(1):185-188.
- [2] van den ENDE M P A, AMPUERO J P. Evaluating Seismic Beamforming Capabilities of Distributed Acoustic Sensing Arrays[J]. Solid Earth, 2021, 12(4):915-934.
- [3] MALGOEZAR A, SNELLEN M, SIJTSMA P, et al. Improving Beamforming by Optimization of Acoustic Array Microphone Positions[C]//6th Berlin Beamforming Conference. Berlin, 2016:BeBeC-2016-S5.
- [4] CHIARIOTTI P, MARTARELLI M, CASTELLINI P. Acoustic Beamforming for Noise Source Localization—Reviews, Methodology and Applications [J]. Mechanical Systems and Signal Processing, 2019, 120:422-448.
- [5] DÖBLER D, OCKER J, PUHLE C. On 3D-Beamforming in The Wind Tunnel[C]//6th Berlin Beamforming Conference. Berlin, 2016:BeBeC-2016-S10.
- [6] MERINO-MARTINEZ R, VON DEN HOFF B, MORATA D, et al. Three-Dimensional Acoustic Imaging Using Asynchronous Microphone Array Measurements[C]//9th Berlin Beamforming Conference. Berlin, 2022:S8.
- [7] BROOKS T, HUMPHREYS W. Three-dimensional Applications of DAMAS Methodology for Aeroacoustic Noise Source Definition [C] // 11th AIAA/CEAS Aeroacoustics Conference. Reston, 2005:2960.
- [8] XENAKI A, JACOBSEN F, FERNANDEZ-

- GRANDE E. Improving the Resolution of Three-dimensional Acoustic Imaging with Planar Phased Arrays[J]. Journal of Sound and Vibration, 2012, 331(8):1939-1950.
- [9] ZAN Ming, XU Zhongming, TANG Zhonghua, et al. Three-dimensional Deconvolution Beamforming Based on the Variable-scale Compressed Computing Grid[J]. Measurement, 2022, 205:112211.
- [10] MA Wei, LIU Xun. Phased Microphone Array for Sound Source Localization with Deep Learning[J]. Aerospace Systems, 2019, 2(2):71-81.
- [11] KUJAWSKI A, HEROLD G, SARRADJ E. A Deep Learning Method for Grid-free Localization and Quantification of Sound Sources[J]. The Journal of the Acoustical Society of America, 2019, 146(3):EL225.
- [12] XU Pengwei, ARCONDOULIS E J G, LIU Yu. Acoustic Source Imaging Using Densely Connected Convolutional Networks[J]. Mechanical Systems and Signal Processing, 2021, 151:107370.
- [13] LEE S Y, CHANG J, LEE S. Deep Learning-based Method for Multiple Sound Source Localization with High Resolution and Accuracy[J]. Mechanical Systems and Signal Processing, 2021, 161:107959.
- [14] ZHAO Yunjie, HE Yansong, CHEN Hao, et al. Three-dimensional Grid-free Sound Source Localization Method Based on Deep Learning[J]. Applied Acoustics, 2025, 227:110261.
- [15] 陈昱, 项薇, 林文文, 等. 基于半监督学习双模型结构的注塑产品异常检测[J]. 中国机械工程, 2025, 36(3):576-583.
- CHEN Yu, XIANG Wei, LIN Wenwen, et al. Abnormal Detection of Injection Molding Products Based on Semi-supervised Learning Dual-model Structure [J]. China Mechanical Engineering, 2025, 36(3):576-583.
- [16] 杨光友, 刘浪, 习晨博. 自适应辅助分类器生成式对抗网络样本生成模型及轴承故障诊断[J]. 中国机械工程, 2022, 33(13):1613-1621.
- YANG Guangyou, LIU Lang, XI Chenbo. Bearing Fault Diagnosis Based on SA-ACGAN Data Generation Model[J]. China Mechanical Engineering, 2022, 33(13):1613-1621.
- [17] KINGMA D P, WELLING M. Auto-Encoding Variational Bayes[J]. arXiv Preprint arXiv, 2013: 1312.6114.
- [18] SARRADJ E. Three-dimensional Acoustic Source Mapping with Different Beamforming Steering Vector Formulations[J]. Advances in Acoustics and Vibration, 2012, 2012(1):292695.
- [19] HUANG Gao, LIU Zhuang, van der MAATEN L, et al. Densely Connected Convolutional Networks[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, 2017:2261-2269.
- [20] ALZUBAIDI L, ZHANGJinglan, HUMAIDI A J, et al. Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions[J]. Journal of Big Data, 2021, 8(1): 53.
- [21] PASZKE A, GROSS S, MASSA F, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library[J]. Advances in Neural Information Processing Systems, 2019, 32.
- [21] PASZKE A, GROSS S, MASSA F, et al. PyTorch: an Imperative Style, High-performance Deep Learning Library [J]. arXiv, 2019: 1912.01703v1.
- [22] SARRADJ E, HEROLD G. A Python Framework for Microphone Array Data Processing[J]. Applied Acoustics, 2017, 116:50-58.
- [23] KUJAWSKI A, PELLING A J R, JEKOSCH S, et al. A Framework for Generating Large-scale Microphone Array Data for Machine Learning [J]. Multimedia Tools and Applications, 2024, 83(11):31211-31231.
- [24] VAN DEN OORD A, VINYALS O. Neural Discrete Representation Learning [J]. arXiv, 2017: 1711.00937.
- [25] RAZAVI A, VAN DEN OORD A, VINYALS O. Generating Diverse High-Fidelity Images with VQ-VAE-2[J]. arXiv, 2019:1906.00446.
- [26] KINGMA D P, BA J. Adam: a Method for Stochastic Optimization[J]. arXiv, 2014:1412.6980.
- [27] LOSHCHELOV I, HUTTER F. SGDR: Stochastic Gradient Descent with Warm Restarts[J]. arXiv, 2016:1608.03983.

(编辑 张 洋)

作者简介:赵昀杰,男,1999年生,硕士研究生。研究方向为振动与噪声控制。发表论文1篇。E-mail:202207131152t@stu.cqu.edu.cn。贺岩松*(通信作者),男,1968年生,教授、博士研究生导师。研究方向为汽车振动与噪声控制、车辆系统动力学与智能控制、车辆舒适性主/客观评价等。发表论文80余篇。E-mail:hys68@cqu.edu.cn。

本文引用格式:

赵昀杰,贺岩松,张志飞,等.基于生成模型的三维波束形成图像压缩方法[J].中国机械工程,2025,36(7):1520-1529.

ZHAO Yunjie, HE Yansong, ZHANG Zhifei, et al. 3D Beamforming Map Compression Method Based on Generative Model [J]. China Mechanical Engineering, 2025, 36(7):1520-1529.