



(12) 发明专利

(10) 授权公告号 CN 111555781 B

(45) 授权公告日 2022.07.12

(21) 申请号 202010344573.1

G06K 9/62 (2022.01)

(22) 申请日 2020.04.27

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 111555781 A

(56) 对比文件

CN 110311718 A, 2019.10.08

CN 108390706 A, 2018.08.10

CN 110912598 A, 2020.03.24

(43) 申请公布日 2020.08.18

(73) 专利权人 天津大学

地址 300072 天津市南开区卫津路92号

Chao-Kai Wen 等. Deep Learning for Massive MIMO CSI Feedback. 《IEEE WIRELESS COMMUNICATIONS LETTERS》. 2018, 全文.

(72) 发明人 于小炯 白洋 贺以恒 郝子瀛

陈诗劼 吴华明

Tianqi Wang, 等. Deep Learning-Based CSI Feedback Approach for Time-Varying Massive MIMO Channels. 《IEEE WIRELESS COMMUNICATIONS LETTERS》. 2019, 全文.

(74) 专利代理机构 天津市三利专利商标代理有限公司 12107

专利代理师 张义

审查员 徐泉

(51) Int. Cl.

H04B 7/0413 (2017.01)

H04B 7/0456 (2017.01)

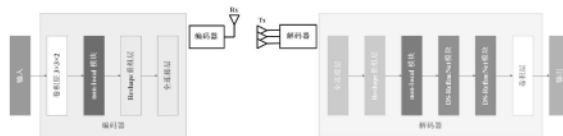
权利要求书2页 说明书5页 附图1页

(54) 发明名称

一种基于深度学习注意力机制的大规模MIMO信道状态信息压缩及重建方法

(57) 摘要

本发明公开了一种基于深度学习注意力机制的大规模MIMO信道状态信息压缩及重建方法,包括如下步骤:步骤一:对输入的信道矩阵做DFT变换;步骤二:构建DS-NLCsiNet模型;步骤三:对模型进行训练,得到模型各层参数;步骤四:对模型的输出进行逆DFT变换。本发明一种基于深度学习注意力机制的大规模MIMO信道状态信息压缩及重建方法,从而能够高效提取远距离信道信息的关联性,同时以更少的参数提取更深层次的特征,大大提高了信道矩阵特征提取的效率。



1. 一种基于深度学习注意力机制的大规模MIMO信道状态信息压缩及重建方法,其特征
在于,包括如下步骤:

步骤一:对输入的信道矩阵 \hat{H} 做DFT变换;

在用户端,对MIMO信道CSI在空频域的信道矩阵 \hat{H} 做DFT变换,得到在角延迟域稀疏的
信道矩阵H;将复数矩阵 $H \in \mathbb{C}^{32 \times 32}$ 的实部和虚部拆分为两个实数矩阵,作为模型的输入;

步骤二:构建DS-NLCsiNet模型;

DS-NLCsiNet模型包含编码器和译码器,其中编码器以信道矩阵H为输入,将其压缩编
码为更低维度的码字s;译码器从码字s中复原出信道矩阵重建值 \hat{H} ;

步骤三:对模型进行训练,得到模型各层参数;

A. 使用Adam优化算法和端到端的学习方式,联合训练编码器和译码器的参数,使损失
函数最小;

使用信道矩阵重建值 \hat{H} 与原始信道矩阵H的均方误差作为损失函数,损失函数表达式
为:

$$L = \frac{1}{T} \sum_{i=1}^T \|\hat{H}_i - H_i\|_2^2$$

其中,T为训练集样本数, $\|\cdot\|_2$ 为欧几里得范数;

B. 所述的模型参数主要包括全连接层的权重、偏置和卷积层的卷积核权重、偏置;

步骤四:对模型的输出 \hat{H} 进行逆DFT变换;

对通过DS-NLCsiNet得到的信道矩阵估计值 \hat{H} 进行逆DFT变换,获得原始信道矩阵重建
值 \hat{H} ;

在步骤二中:编码器位于用户端,包括一个卷积层、一个Non Local Block和一个全连
接层,随机初始化各层参数。

编码器第一层为卷积层,使用两通道 3×3 大小的卷积核与输入进行卷积,通过零填充、
激活函数LeakyReLU和BatchNormalization,得到两通道 32×32 大小的特征图;

编码器第二层为Non Local Block,其通用公式表示为:

$$y_i = \frac{1}{C(x)} \sum_{\forall j} f(x_i, x_j) g(x_j)$$

其中,x是输入信号,i,j表示输出位置,f函数计算i和j的相似度,g函数计算
featuremap在j位置的表示,C(x)表示响应因子,对结果进行标准化处理;

编码器第三层为Dense层,将第二层的输出重组为一个 2048×1 的向量,作为第三层的
输入,该层含有M个神经元,激活函数为线性激活函数,Dense层将 2048×1 的向量压缩为 $M \times 1$
的码字s,作为编码器的输出,传入基站端的解码器,假设传输过程无损失。

2. 根据权利要求1所述的一种基于深度学习注意力机制的大规模MIMO信道状态信息压
缩及重建方法,其特征不在于,

编码器第二层具体实施方法包括:

首先,使用两路两通道的 1×1 卷积核分别对输入的featuremap进行处理,并将两路卷积的输出进行点乘运算,即可得到上式中的 $f(x_i, x_j)$;然后使用两通道的 1×1 卷积核对输入的featuremap进行处理,得到上式中的 $g(x_j)$;再将得到 $f(x_i, x_j)$ 和 $g(x_j)$ 进行点乘运算,使用softmax激活函数;最后,将点乘的结果通过两通道的 1×1 卷积核,得到编码器第二层的输出。

3. 根据权利要求1所述的一种基于深度学习注意力机制的大规模MIMO信道状态信息压缩及重建方法,其特征在于,

基站端收到码字s后,使用位于基站端的解码器解码,解码器包含一个全连接层,两个DS-RefineNet单元和一个卷积层,随机初始化各层参数,码字在经过解码器各层处理后,输出信道矩阵H的估计值 \hat{H} ;

解码器Dense层以 $M \times 1$ 码字s为输入,将其复原为 2048×1 的一维向量,激活函数为线性激活函数,经过Reshape层,还原为两通道 32×32 大小的特征图,再通过non-local block,得到初步还原的特征图;

解码器的第三、四层为两个DS-RefineNet单元;

解码器的最后一层为一个两通道 3×3 卷积核,激活函数为Sigmoid。

4. 根据权利要求3所述的一种基于深度学习注意力机制的大规模MIMO信道状态信息压缩及重建方法,其特征在于,

解码器的第三、四层具体的实施方法,包括:

首先,将初步还原的特征图输入第一个DS-RefineNet,每个DS-RefineNet单元包含一个输入层和4个卷积层,前三个卷积层均使用 3×3 尺寸的卷积核,卷积核通道数分别为8、16、2,最后一个卷积层使用 1×1 尺寸的卷积核,对结果进行降维,每个卷积层都以输入与其之前所有卷积核的输出在channel维度上的拼接为其输入,其数学表达式为:

$$x_1 = H_1([x_0, x_1, \dots, x_{1-1}])$$

其中, x_1 表示第1层的输出;

最后一层卷积核的输出即为DS-RefineNet单元的输出,除最后一层外其余各层均采用LeakyReLU激活函数并使用批归一化。

一种基于深度学习注意力机制的大规模MIMO信道状态信息压缩及重建方法

技术领域

[0001] 本发明涉及大规模MIMO系统技术领域,尤其涉及一种基于深度学习注意力机制的大规模MIMO信道状态信息压缩及重建方法。

背景技术

[0002] 大规模MIMO (multiple-input multiple-output) 系统已被公认为是5G无线系统的关键技术之一,该技术通过在基站端配置大量天线,在空间域形成多个独立信道,从而大大增加无线通信系统的吞吐量。然而,在频分复用(frequency division duplexity,FDD) MIMO系统中,信道不具有互易性,在用户端获得的下行CSI需要通过反馈链路发送给基站端,反馈完整的CSI将带来巨大的资源开销。此外,随着天线数量大大增加,传统的减小反馈方案如基于量化和码本的有限反馈算法会在一定程度上损失信道状态信息,且反馈开销呈线性增长,因而不适用于大规模MIMO系统的信道反馈。

[0003] 随着5G技术的高速发展,未来的无线通信应用场景变得更加多样和复杂,对无线通信质量的要求提高。而无线通信智能化则是5G时代技术发展的一大趋势,基于深度学习的模型和算法可以促进无线网络分析和资源的有效管理,有利于应对移动应用通信量和计算量的爆炸式增长。部分深度学习网络技术已经应用于无线传输技术,为5G时代的到来铺平了道路。

[0004] 现有技术中公开了一种基于深度学习的大规模MIMO信道状态信息反馈方法。该方法的过程为:在用户端,对MIMO信道状态信息在空频域的信道矩阵 H 做二维傅里叶变换,获得在角延迟域稀疏的信道矩阵 H ;构建包括编码器和译码器的模型,其中编码器属于用户端,将信道矩阵 H 编码为更低维度的码字,译码器属于基站端,从码字重建出原信道矩阵估计值 \hat{H} ;对模型进行训练,使得 \hat{H} 和 H 尽可能接近,获得模型参数;对模型输出的重建信道矩阵 \hat{H} 进行二位逆DFT变换,恢复得到原始空频域的信道矩阵 H 重建值;将训练好的模型用于信道信息的压缩和重建。

[0005] 然而,传统的模型一般仅使用 3×3 或 5×5 大小的卷积核,特征提取仅局限于某一局部邻域,通过不断迭代来增大自己的感受野,但是这个迭代过程十分低效,对远程依赖提取效果较差,需要消耗巨大的计算资源,并且每一层的输出仅传入其上一层,所需参数较多,提取特征的效率较低。当层数很多时,会出现梯度弥散的情况,极大的降低了传输效率与精确度。

发明内容

[0006] 本发明提出一种基于深度学习注意力机制的大规模MIMO信道状态信息压缩及重建方法,从而能够高效提取远距离信道信息的关联性,同时以更少的参数提取更深层次的特征,大大提高了信道矩阵特征提取的效率。

[0007] 为实现本发明的目的,本发明提供的一种基于深度学习注意力机制的大规模MIMO

信道状态信息压缩及重建方法,包括如下步骤:

[0008] 步骤一:对输入的信道矩阵 \hat{H} 做DFT变换;

[0009] 在用户端,对MIMO信道CSI在空频域的信道矩阵 \hat{H} 做DFT变换,得到在角延迟域稀疏的信道矩阵 H ;将复数矩阵 $H \in \mathbb{C}^{32 \times 32}$ 的实部和虚部分为两个实数矩阵,作为模型的输入;

[0010] 步骤二:构建DS-NLCsiNet模型;

[0011] DS-NLCsiNet模型包含编码器和译码器,其中编码器以信道矩阵 H 为输入,将其压缩编码为更低维度的码字 s ;译码器从码字 s 中复原出信道矩阵重建值 \hat{H} ;

[0012] 步骤三:对模型进行训练,得到模型各层参数;

[0013] A.使用Adam优化算法和端到端的学习方式,联合训练编码器和译码器的参数,使损失函数最小;

[0014] 使用信道矩阵重建值 \hat{H} 与原始信道矩阵 H 的均方误差作为损失函数,损失函数表达式为:

$$[0015] \quad L = \frac{1}{T} \sum_{i=1}^T \|\hat{H}_i - H_i\|_2^2$$

[0016] 其中, T 为训练集样本数, $\|\cdot\|_2$ 为欧几里得范数;

[0017] B.所述的模型参数主要包括全连接层的权重、偏置和卷积层的卷积核权重、偏置;

[0018] 步骤四:对模型的输出 \hat{H} 进行逆DFT变换;

[0019] 对通过DS-NLCsiNet得到的信道矩阵估计值 \hat{H} 进行逆DFT变换,获得原始信道矩阵重建值 \hat{H} 。

[0020] 进一步地,

[0021] 在步骤二中:编码器位于用户端,包括一个卷积层、一个Non Local Block和一个全连接层,随机初始化各层参数。

[0022] 编码器第一层为卷积层,使用两通道 3×3 大小的卷积核与输入进行卷积,通过零填充、激活函数LeakyReLU和BatchNormalization,得到两通道 32×32 大小的特征图;

[0023] 编码器第二层为Non Local Block,其通用公式表示为:

$$[0024] \quad y_i = \frac{1}{C(x)} \sum_{\forall j} f(x_i, x_j) g(x_j)$$

[0025] 其中, x 是输入信号, i, j 表示输出位置, f 函数计算 i 和 j 的相似度, g 函数计算featuremap在 j 位置的表示, $C(x)$ 表示响应因子,对结果进行标准化处理;

[0026] 编码器第三层为Dense层,将第二层的输出重组为一个 2048×1 的向量,作为第三层的输入,该层含有 M 个神经元,激活函数为线性激活函数,Dense层将 2048×1 的向量压缩为 $M \times 1$ 的码字 s ,作为编码器的输出,传入基站端的解码器,假设传输过程无损失。

[0027] 进一步地,编码器第二层具体实施方法包括:

[0028] 首先,使用两路两通道的 1×1 卷积核分别对输入的featuremap进行处理,并将两路卷积的输出进行点乘运算,即可得到上式中的 $f(x_i, x_j)$;然后使用两通道的 1×1 卷积核对

输入的featuremap进行处理,得到上式中的 $g(x_j)$;再将得到 $f(x_i, x_j)$ 和 $g(x_j)$ 进行点乘运算,使用softmax激活函数;最后,将点乘的结果通过两通道的 1×1 卷积核,得到编码器第二层的输出。

[0029] 进一步地,基站端收到码字s后,使用位于基站端的解码器解码,解码器包含一个全连接层,两个DS-RefineNet单元和一个卷积层,随机初始化各层参数,码字在经过解码器各层处理后,输出信道矩阵H的估计值 \hat{H} ;

[0030] 解码器Dense层以 $M \times 1$ 码字s为输入,将其复原为 2048×1 的一维向量,激活函数为线性激活函数,经过Reshape层,还原为两通道 32×32 大小的特征图,再通过non-local block,得到初步还原的特征图;

[0031] 解码器的第三、四层为两个DS-RefineNet单元;

[0032] 解码器的最后一层为一个两通道 3×3 卷积核,激活函数为Sigmoid。

[0033] 进一步地,解码器的第三、四层具体的实施方法,包括:

[0034] 首先,将初步还原的特征图输入第一个DS-RefineNet,每个DS-RefineNet单元包含一个输入层和4个卷积层,前三个卷积层均使用 3×3 尺寸的卷积核,卷积核通道数分别为8、16、2,最后一个卷积层使用 1×1 尺寸的卷积核,对结果进行降维,每个卷积层都以输入与其之前所有卷积核的输出在channel维度上的拼接为其输入,其数学表达式为:

[0035] $x_1 = H_1([x_0, x_1, \dots, x_{1-1}])$

[0036] 其中, x_1 表示第1层的输出;

[0037] 最后一层卷积核的输出即为DS-RefineNet单元的输出,除最后一层外其余各层均采用LeakyReLU激活函数并使用批归一化。

[0038] 与现有技术相比,本发明的有益效果为,

[0039] (1) 本方案使用Non-local block,通过非局部操作直接计算两个位置之间的关系,即可快速捕获长范围依赖,让感受野很大而不是仅局限于在局部区域,且只需更少的堆叠层便可达到同等效果,极大地提高了非局部操作计算效率。

[0040] (2) 以往的结构中,每一层卷积核仅处理上一层卷积核提取的特征。本方案中每一层网络的输入包括前面所有层网络的输出,提升了信息和梯度在网络中的传输效率。每层都能直接得到输入信号,并且直接从损失函数拿到梯度,从而可以增强特征重用,减少参数,并缓解梯度消失问题。

附图说明

[0041] 图1所示为本申请中DS-NLCsiNet结构示意图;

[0042] 图2所示为本申请中non-local block结构示意图;

[0043] 图3所示为本申请中DS-RefineNet结构示意图。

具体实施方式

[0044] 需要说明的是,在不冲突的情况下,本申请中的实施例及实施例中的特征可以相互组合。

[0045] 以下结合附图和具体实施例对本发明作进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0046] 本发明所采用的技术方案主要包括以下步骤:

[0047] 步骤一:对输入的信道矩阵 \hat{H} 做DFT变换;

[0048] 在用户端,对MIMO信道CSI在空频域的信道矩阵 \hat{H} 做DFT变换,得到在角延迟域稀疏的信道矩阵H;将复数矩阵 $H \in \mathbb{C}^{32 \times 32}$ 的实部和虚部分为两个实数矩阵,作为模型的输入;

[0049] 步骤二:构建DS-NLCsiNet模型;

[0050] 如图1所示,DS-NLCsiNet模型包含编码器和译码器,其中编码器以信道矩阵H为输入,将其压缩编码为更低维度的码字s,译码器从码字s中复原出信道矩阵重建值 \hat{H} ;

[0051] 步骤三:对模型进行训练,得到模型各层参数;

[0052] A. 使用Adam优化算法和端到端的学习方式,联合训练编码器和译码器的参数,使损失函数最小。

[0053] 使用信道矩阵重建值 \hat{H} 与原始信道矩阵H的均方误差作为损失函数。损失函数表达式为:

$$[0054] \quad L = \frac{1}{T} \sum_{i=1}^T \|\hat{H}_i - H_i\|_2^2$$

[0055] 其中,T为训练集样本数, $\|\cdot\|_2$ 为欧几里得范数。

[0056] B. 所述的模型参数主要包括全连接层的权重、偏置和卷积层的卷积核权重、偏置。

[0057] 步骤四:对模型的输出 \hat{H} 进行逆DFT变换。

[0058] 对通过DS-NLCsiNet得到的信道矩阵估计值 \hat{H} 进行逆DFT变换,获得原始信道矩阵重建值 \hat{H} 。

[0059] 其中,上述步骤二包含以下步骤:

[0060] A. 编码器结构细节;

[0061] 编码器位于用户端,包括一个卷积层、一个Non Local Block和一个全连接层。随机初始化各层参数。

[0062] 编码器第一层为卷积层,使用两通道 3×3 大小的卷积核与输入进行卷积,通过零填充、激活函数LeakyReLU和BatchNormalization,得到两通道 32×32 大小的特征图。

[0063] 如图2所示,编码器第二层为Non Local Block,其通用公式表示为:

$$[0064] \quad y_i = \frac{1}{C(x)} \sum_{\forall j} f(x_i, x_j) g(x_j)$$

[0065] 其中,x是输入信号,i,j表示输出位置,f函数计算i和j的相似度,g函数计算featuremap在j位置的表示,C(x)表示响应因子,对结果进行标准化处理。

[0066] 具体实施方法为:首先,使用两路两通道的 1×1 卷积核分别对输入的featuremap进行处理,并将两路卷积的输出进行点乘运算,即可得到上式中的 $f(x_i, x_j)$ 。然后使用两通道的 1×1 卷积核对输入的featuremap进行处理,得到上式中的 $g(x_j)$ 。再将得到 $f(x_i, x_j)$ 和 $g(x_j)$ 进行点乘运算,使用softmax激活函数。最后,将点乘的结果通过两通道的 1×1 卷积核,得到编码器第二层的输出。

[0067] 编码器第三层为Dense层。将第二层的输出重组为一个 2048×1 的向量,作为第三层的输入。该层含有M个神经元,激活函数为线性激活函数。Dense层将 2048×1 的向量压缩为 $M \times 1$ 的码字s,作为编码器的输出,传入基站端的解码器。假设传输过程无损失。

[0068] 基站端收到码字s后,使用位于基站端的解码器解码,解码器包含一个全连接层,两个DS-RefineNet单元和一个卷积层,随机初始化各层参数。码字在经过解码器各层处理后,输出信道矩阵H的估计值 \hat{H} ;

[0069] 解码器Dense层以 $M \times 1$ 码字s为输入,将其复原为 2048×1 的一维向量,激活函数为线性激活函数。经过Reshape层,还原为两通道 32×32 大小的特征图。再通过non-local block,得到初步还原的特征图。

[0070] 解码器的第三、四层为两个DS-RefineNet单元。

[0071] 如图3所示为DS-RefineNet结构示意图

[0072] 首先将初步还原的特征图输入第一个DS-RefineNet。每个DS-RefineNet单元包含一个输入层和4个卷积层,前三个卷积层均使用 3×3 尺寸的卷积核,卷积核通道数分别为8,16,2,最后一个卷积层使用 1×1 尺寸的卷积核,对结果进行降维。每个卷积层都以输入与其之前所有卷积核的输出在channel维度上的拼接为其输入。其数学表达式为:

$$[0073] \quad x_1 = H_1([x_0, x_1, \dots, x_{1-1}])$$

[0074] 其中, x_1 表示第1层的输出。

[0075] 最后一层卷积核的输出即为DS-RefineNet单元的输出。除最后一层外其余各层均采用LeakyReLU激活函数并使用批归一化。

[0076] 解码器的最后一层为一个两通道 3×3 卷积核,激活函数为Sigmoid。

[0077] 本发明引入Non-Local Block模型,通过计算任意两个位置之间的交互直接捕捉远程依赖,而不用局限于相邻点,从而可以维持更多信息,提高了非局部操作计算效率,且只需要更少的堆叠层,具有很大的实用价值。

[0078] 本发明引入DenseNet模型,从以往每一层的输出仅传入其后的方式,转变为每一层的输出都传入其后每一层,鼓励了特征重用,从而可以以更少的参数提取更深层次的特征,同时减轻了梯度弥散的问题。

[0079] 以上所述仅是本发明的优选实施方式,应当指出的是,对于本技术领域的普通技术人员来说,在不脱离本发明原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也应视为本发明的保护范围。

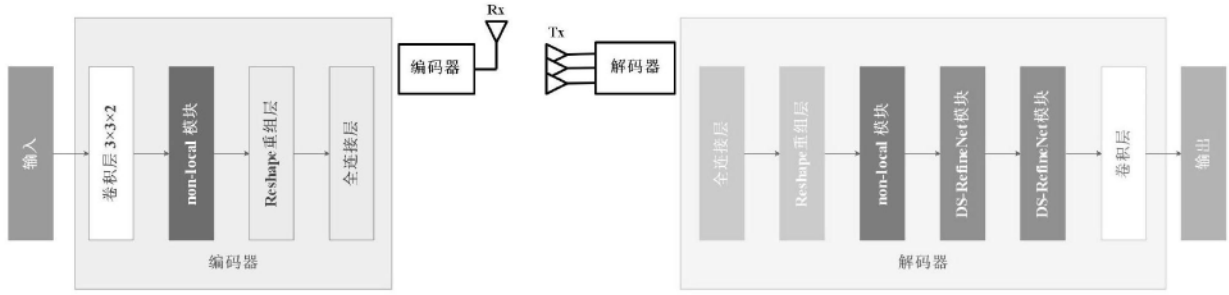


图1

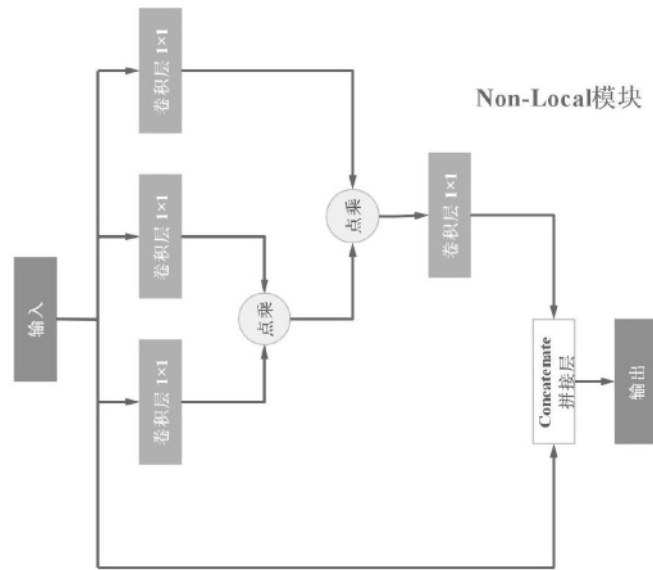


图2

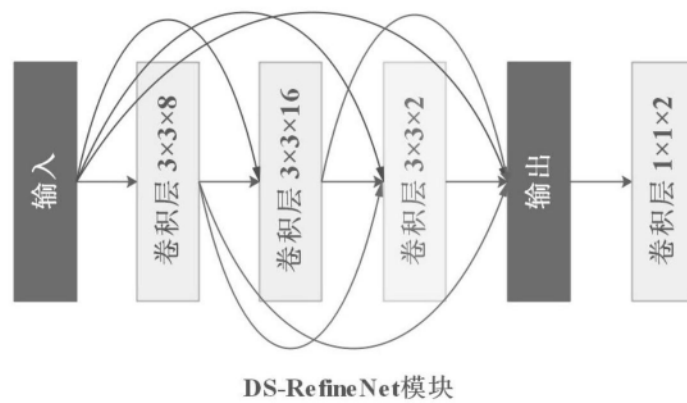


图3