



(12) 发明专利

(10) 授权公告号 CN 110933429 B

(45) 授权公告日 2021.11.12

(21) 申请号 201911104216.1

G06N 3/08 (2006.01)

(22) 申请日 2019.11.13

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 106911930 A, 2017.06.30

申请公布号 CN 110933429 A

CN 106937125 A, 2017.07.07

CN 107155112 A, 2017.09.12

(43) 申请公布日 2020.03.27

A. Mousavi, et al..DeepCodec: Adaptive sensing and recovery via deep convolutional neural networks.《Proc. 55th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)》.2017,

(73) 专利权人 南京邮电大学

地址 210023 江苏省南京市栖霞区文苑路9号

Eric W. Tramel, et al..Video compressed sensing with multihypothesis.《Video compressed sensing with multihypothesis》.2011,

(72) 发明人 张登银 周超 陈灿 李俊江 刘子捷

(74) 专利代理机构 南京纵横知识产权代理有限公司 32224

代理人 朱远枫

万文韬.基于深度学习的图像压缩感知重建技术研究.《中国优秀硕士学位论文全文数据库》.2019,

(51) Int. Cl.

H04N 19/176 (2014.01)

H04N 19/50 (2014.01)

G06N 3/04 (2006.01)

审查员 何美伶

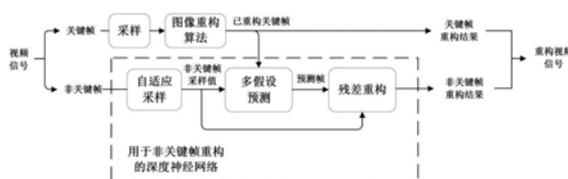
权利要求书4页 说明书11页 附图3页

(54) 发明名称

基于深度神经网络的视频压缩感知与重构方法和装置

(57) 摘要

本发明公开了一种基于深度神经网络的视频压缩感知与重构方法和装置。此方法将视频信号划分为关键帧与非关键帧。对于关键帧，采用已有的图像重构方法进行重构。对于非关键帧，此方法创新性的提出了一种特殊的深度神经网络来完成重构。此神经网络由自适应采样模块，多假设预测模块，残差重构模块组成，充分利用了视频信号的时空相关性来完成视频信号的采样与重构，在提高重构质量的同时，保证了算法的低时间复杂度。因此，本发明方法适用于采样端资源受限、对重构质量要求高、对实时性要求高的视频传感系统。



1. 一种基于深度神经网络的视频压缩感知与重构方法,其特征在于,该方法具体步骤如下:

构建用于对视频非关键帧进行重构的深度神经网络模块,包含构建自适应采样模块、多假设预测模块和残差重构模块构成的端到端的深度神经网络;

输入视频序列,将视频序列分为多个图片组,确定每一个图片组的关键帧和非关键帧;对关键帧进行重构得到重构后的关键帧;

利用所训练完成的深度神经网络对视频序列的非关键帧进行重构;

将重构的关键帧和深度神经网络输出的重构的非关键帧按顺序排列即为所得重构视频信号;

其中构建深度神经网络模块的具体方法包括:输入用于网络训练的视频序列;将视频信号序列分为多个图片组,确定每一个图片组的关键帧和非关键帧;将非关键帧分为多个互不重叠的图像块;

构建自适应采样模块:对非关键帧中每个图像块进行自适应采样得到图像块采样值;

构建多假设预测模块:确定每个图像块对应的假设块;对非关键帧中的每个图像块利用对应的假设块进行自适应采样得到假设块的采样值,根据假设块采样值和图像块采样值,计算每个图像块的预测值,得到多假设预测图像;

构建残差重构模块:对多假设预测图像进行自适应采样,构建多假设预测图像采样值与非关键帧采样值的残差值,根据残差值利用残差重构方法得到非关键帧重构;

将构建的自适应采样模块与多假设预测模块进行级联构建子网络,对子网络进行训练,将子网络预训练得到的参数作为初始值,对整个深度神经网络进行训练,得到训练完成的深度神经网络;

确定每个图像块对应的假设块的具体方法为:

对于非关键帧中每一个 $b \cdot b$ 图像块,以其位置为中心,在关键帧中框定一个大小为 $W \times W$ 的矩形搜索窗,将搜索窗中所有的重叠的图像块提取作为当前图像块的假设块;将非关键帧中每一个图像块的所有假设块整合在一起,可得一个 $h \times w \times (b \cdot b) \times n$ 的假设块张量,其中, n 代表了每一个图像块对应的假设块的个数, $h \times w$ 代表待预测的图像块的个数, $b \cdot b$ 为假设块的维度;

构建残差重构模块的具体方法为:

多假设预测图像进行自适应采样,得到多假设预测图像的采样值,维度大小为 $h \times w \times (SR \cdot b \cdot b)$;

将多假设采样值与基于图像块采样值得非关键帧采样值相减,得到残差采样值;

将残差采样值通过一个卷积层,转换成一个大小为 $h \times w \times (b \cdot b)$ 的特征图;此卷积层的卷积核大小为 1×1 ,输入信道数为 $SR \cdot b \cdot b$,输出信道数为 $b \cdot b$;

将输出特征图进行变形,得到一个 $(h \cdot b) \times (w \cdot b)$ 大小的特征图;

利用卷积层对得到的特征图进行卷积,得到残差图像的重构结果;

得到的残差重构结果与多假设预测图像相加,输出最后的非关键帧重构值。

2. 根据权利要求1所述的一种基于深度神经网络的视频压缩感知与重构方法,其特征在于,计算每个图像块预测值的计算公式如下:

$$P_i = \sum_{\forall j} \omega_{i,j} h_{i,j} = \frac{1}{\sum_{\forall j} e^{p(q_{i,j})^T p(y_i)}} \sum_{\forall j} e^{p(q_{i,j})^T p(y_i)} h_{i,j} \quad (2)$$

其中, P_i 为当前非关键帧中第 i 个图像块的预测结果, $h_{i,j}$ 为第 i 个图像块中第 j 个假设块, $\omega_{i,j}$ 为此假设块的权重, $q_{i,j}$ 为利用自适应采样对假设块进行采样的结果, y_i 为第 i 个图像块的自适应采样值, $p(\cdot)$ 为一个非线性映射函数, 用于将采样值从低维转换成高维。

3. 根据权利要求2所述的一种基于深度神经网络的视频压缩感知与重构方法, 其特征在于, 公式(2)具体实现方法为:

对假设块张量进行自适应采样, 得到一个包含所有图像块的各个假设块采样值 $q_{i,j}$ 的张量, 其维度大小为 $h \times w \times (SR \cdot b \cdot b) \times n$, 其中 $SR \cdot b \cdot b$ 为假设块采样值 $q_{i,j}$ 的维度;

利用三个卷积层来实现 $p(\cdot)$ 函数, 三个卷积层的具体参数为: 卷积核大小: 都为 1×1 , 输入信道数: 第一个卷积层为 $SR \cdot b \cdot b$, 其余为 $b \cdot b$, 输出信道数: 都为 $b \cdot b$;

利用 $p(\cdot)$ 函数, 将步骤(b)所得到的假设块的采样值张量转换成一个 $h \times w \times (b \cdot b) \times n$ 的假设块特征图;

利用 $p(\cdot)$ 函数, 将步骤(4)得到的非关键帧采样值张量进行卷积, 非关键帧采样值张量大小为 $h \times w \times (SR \cdot b \cdot b)$ 得到一个 $h \times w \times (b \cdot b)$ 大小的非关键帧特征图;

将得到的假设块特征图与得到的非关键帧特征图进行矩阵乘法, 实现公式(2)中 $p(q_{i,j})^T p(y_i)$, 得到一个 $h \times w \times n$ 大小的张量; 利用 softmax 函数对所得张量的最后一维进行

归一化实现公式(2)的 $\frac{1}{\sum_{\forall j} e^{p(q_{i,j})^T p(y_i)}} \sum_{\forall k} e^{p(q_{i,j})^T p(y_i)}$, 得到

大小为 $h \times w \times n$ 的系数张量;

将所得的系数张量与步骤(a)所得的假设块张量进行矩阵乘法, 得到大小为 $h \times w \times (b \cdot b)$ 的预测值张量;

将所得的预测值张量进行变形, 得到一个大小为 $(h \cdot b) \times (w \cdot b)$ 的当前非关键帧的多假设预测图像。

4. 根据权利要求2所述的一种基于深度神经网络的视频压缩感知与重构方法, 其特征在于, 各个假设块的权重由嵌入高斯函数决定, 具体如下式:

$$\omega_{i,j} = \frac{e^{p(q_{i,j})^T p(y_i)}}{\sum_{\forall j} e^{p(q_{i,j})^T p(y_i)}}。$$

5. 根据权利要求1所述的一种基于深度神经网络的视频压缩感知与重构方法, 其特征在于:

利用8个卷积层对得到的特征图进行卷积, 得到残差图像的重构结果; 8个卷积层的具体参数如下: 卷积核大小: 都为 3×3 , 输入信道数: 第一个卷积层为1, 其余为64, 输出信道

数:最后一个卷积层为1,其余都为64。

6. 根据权利要求1~5任意一条权利要求所述的一种基于深度神经网络的视频压缩感知与重构方法,其特征在于,自适应采样方法如下:对信号的采集如下式所示:

$$y = \Phi x \quad (1)$$

其中, y 为采样值, Φ 为采样矩阵, x 为原始信号,式(1)由一个卷积层来实现,此卷积层的卷积核大小为 $b \times b$,输入信道数为1,输出信道数为 $SR \cdot b \cdot b$,其中, SR 为非关键帧的采样率,此卷积层等同于对图像中每一个 $b \times b$ 大小的图像块进行压缩感知采样,输出为一个 $h \times w \times (SR \cdot b \cdot b)$ 的张量,其中, h 和 w 代表输入图像的高和宽两个维度上的块的数量,卷积层的权重为对应的采样矩阵。

7. 根据权利要求1~5任意一条权利要求所述的一种基于深度神经网络的视频压缩感知与重构方法,其特征在于,利用DNN的图像压缩感知重构方法对关键帧进行重构。

8. 一种基于深度神经网络的视频压缩感知与重构装置,其特征在于,包括:

视频序列输入模块、深度神经网络模块、非关键划分模块、关键帧重构模块和视频信号重构输出模块;

所述视频序列输入模块,用于输入视频序列;

所述非关键划分模块,用于将视频序列输入模块输入的视频信号序列分为多个图片组,确定每一个图片组的关键帧和非关键帧;

所述深度神经网络模块,用于对非关键划分模块划分的非关键帧进行重构得到重构的非关键帧;

所述关键帧重构模块,用于利用图像压缩感知重构方法对非关键划分模块划分得到的关键帧进行重构得到重构后的关键帧;

所述视频信号重构输出模块:将关键帧重构模块重构的关键帧和深度神经网络输出的重构的非关键帧按顺序排列即为所得重构视频信号;

所述深度神经网络模块包括图像块划分模块、自适应采样模块、多假设预测模块、残差重构模块、子网络训练模块和深度神经网络训练模块,其中自适应采样模块、多假设预测模块和残差重构模块构成端到端的深度神经网络;

所述视频序列输入模块,用于输入用于训练的视频序列;

所述非关键划分模块,用于将视频序列输入模块输入的训练视频序列分为多个图片组,确定每一个图片组的关键帧和非关键帧;

所述图像块划分模块,用于将非关键划分模块得到的训练视频序列非关键帧分为多个互不重叠的图像块;

所述自适应采样模块,用于对非关键帧中每个图像块进行自适应采样得到图像块采样值;

所述多假设预测模块,用于确定每个图像块对应的假设块;对非关键帧中的每个图像块利用对应的假设块进行自适应采样得到假设块的采样值,根据假设块采样值和图像块采样值,得到多假设预测图像;

所述残差重构模块,用于对多假设预测图像进行自适应采样,构建多假设预测图像采样值与非关键帧采样值的残差值,根据残差值利用残差重构方法得到非关键帧重构;

所述子网络训练模块,用于将构建的自适应采样模块与多假设预测模块进行级联构建

子网络,对子网络进行训练;

所述深度神经网络训练模块,用于将子网络预训练得到的参数作为初始值,对整个深度神经网络进行训练,得到训练完成的深度神经网络;

确定每个图像块对应的假设块的具体方法为:

对于非关键帧中每一个 $b \cdot b$ 图像块,以其位置为中心,在关键帧中框定一个大小为 $W \times W$ 的矩形搜索窗,将搜索窗中所有的重叠的图像块提取作为当前图像块的假设块;将非关键帧中每一个图像块的所有假设块整合在一起,可得一个 $h \times w \times (b \cdot b) \times n$ 的假设块张量,其中, n 代表了每一个图像块对应的假设块的个数, $h \times w$ 代表待预测的图像块的个数, $b \cdot b$ 为假设块的维度;

构建残差重构模块的具体方法为:

多假设预测图像进行自适应采样,得到多假设预测图像的采样值,维度大小为 $h \times w \times (SR \cdot b \cdot b)$;

将多假设采样值与基于图像块采样值得非关键帧采样值相减,得到残差采样值;

将残差采样值通过一个卷积层,转换成一个大小为 $h \times w \times (b \cdot b)$ 的特征图;此卷积层的卷积核大小为 1×1 ,输入信道数为 $SR \cdot b \cdot b$,输出信道数为 $b \cdot b$;

将输出特征图进行变形,得到一个 $(h \cdot b) \times (w \cdot b)$ 大小的特征图;

利用卷积层对得到的特征图进行卷积,得到残差图像的重构结果;

得到的残差重构结果与多假设预测图像相加,输出最后的非关键帧重构值。

基于深度神经网络的视频压缩感知与重构方法和装置

技术领域

[0001] 本发明涉及一种基于深度神经网络的视频压缩感知与重构方法和装置,属于视频图像处理技术领域。

背景技术

[0002] 目前,信息社会不断发展,对图像、视频等高维信号的采集与处理的需求日益增加。在传统的视频采集系统中,视频传感器往往会先根据奈奎斯特采样定律(Nyquist-Shannon sampling theorem)采集大量的、冗余的视频信号,然后利用视频压缩技术在编码端对信号进行压缩,以减少传输带宽和存储的压力。该做法造成了巨大的采样资源的浪费,并且,在一些编码端资源受限的应用中,如无线视频传感网(Wireless Video Sensing Networks, WVSNs),高复杂度的压缩编码往往难以在编码端应用。近年来,视频压缩感知(Compressive Video Sensing, CVS)技术被广泛应用于无线视频传感网当中。它在编码端利用单像素相机(single pixel camera)等设备,将采样与压缩融合为一个低复杂度的线性投影,得到采样值;在解码端,利用适当的重构算法对采样值进行重构,得到完整的视频信号。

[0003] 在已有的视频压缩感知系统中,视频信号的恢复往往通过一个优化算法来求解。具体的,此类优化算法一般利用视频信号的先验信息(例如:稀疏性,模型稀疏性,低秩性,非局部相似性)来设计一个凸优化问题。但是凸优化问题的求解往往是迭代的,高复杂度的,因而导致重构时间较长。一般的,利用此方法恢复一个分辨率为352*288的视频帧需要几秒甚至几分钟,此重构时间难以实现在解码端实时播放视频,因而无法满足无线传感网中实时监控的需求。另一方面,在此类优化方法当中使用的先验信息是人为设计的,只能大致描述信号的特征。自然中的视频信号通常更为复杂,因此基于优化的重构方法在重构质量仍有不足,往往导致重构视频模糊,重影,用户体验较差。

发明内容

[0004] 本发明目的在于解决目前视频重构方法存在的延迟较大,重构视频质量有待提高的问题,提出了一种基于深度神经网络(Deep Neural Networks, DNNs)的视频压缩感知与重构方法和装置,主要应用于编码端受限的无线视频传感网。

[0005] 为实现上述方案,本发明采用以下技术方案:

[0006] 一方面,本发明提供一种基于深度神经网络的视频压缩感知与重构方法,其特征在于,该方法具体步骤如下:

[0007] 构建用于对视频非关键帧进行重构的深度神经网络模块,包含构建自适应采样模块、多假设预测模块和残差重构模块构成的端到端的深度神经网络,具体方法包括:输入用于网络训练的视频序列;将视频信号序列分为多个图片组,确定每一个图片组的关键帧和非关键帧;对关键帧进行重构得到重构后的关键帧;将非关键帧分为多个互不重叠的图像块;

[0008] 构建自适应采样模块:对非关键帧中每个图像块进行自适应采样得到图像块采样值;

[0009] 构建多假设预测模块:确定每个图像块对应的假设块;对非关键帧中的每个图像块利用对应的假设块进行自适应采样得到假设块的采样值,根据假设块采样值和图像块采样值,计算每个图像块的预测值,得到多假设预测图像;

[0010] 构建残差重构模块:对多假设预测图像进行自适应采样,构建多假设预测图像采样值与非关键帧采样值的残差值,根据残差值利用残差重构方法得到非关键帧重构;

[0011] 将构建的自适应采样模块与多假设预测模块进行级联构建子网络,对子网络进行训练,将子网络预训练得到的参数作为初始值,对整个深度神经网络进行训练,得到训练完成的深度神经网络;

[0012] 输入视频序列,将视频序列分为多个图片组,确定每一个图片组的关键帧和非关键帧;对关键帧进行重构得到重构后的关键帧;

[0013] 利用所训练完成的深度神经网络对视频序列的非关键帧进行重构;

[0014] 将重构的关键帧和深度神经网络输出的重构的非关键帧按顺序排列即为所得重构视频信号。

[0015] 进一步地,确定每个图像块对应的假设块的具体方法为:

[0016] 对于非关键帧中每一个 $b \cdot b$ 图像块,以其位置为中心,在关键帧中框定一个大小为 $W \times W$ 的矩形搜索窗,将搜索窗中所有的重叠的图像块提取作为当前图像块的假设块。将非关键帧中每一个图像块的所有假设块整合在一起,可得一个 $h \times w \times (b \cdot b) \times n$ 的假设块张量,其中, n 代表了每一个图像块对应的假设块的个数, $h \times w$ 代表待预测的图像块的个数, $b \cdot b$ 为假设块的维度。

[0017] 进一步地,计算每个图像块的预测值的计算公式如下:

$$\begin{aligned}
 P_i &= \sum_{\forall j} \omega_{i,j} h_{i,j} \\
 &= \frac{1}{\sum_{\forall j} e^{p(q_{i,j})^T p(y_i)}} \sum_{\forall j} e^{p(q_{i,j})^T p(y_i)} h_{i,j} \quad (1),
 \end{aligned}$$

[0019] 其中, P_i 为当前非关键帧中第 i 个图像块块的预测结果, $h_{i,j}$ 为第 i 个图像块中第 j 个假设块, $\omega_{i,j}$ 为此假设块的权重, $q_{i,j}$ 为利用自适应采样对假设块进行采样的结果, y_i 为第 i 个图像块的自适应采样值, $p(\cdot)$ 为一个非线性映射函数,用于将采样值从低维转换成高维。

[0020] 进一步地,公式(1)具体实现方法为:

[0021] 对假设块张量进行自适应采样,得到一个包含所有图像块的各个假设块采样值 $q_{i,j}$ 的张量,其维度大小为 $h \times w \times (SR \cdot b \cdot b) \times n$,其中 $SR \cdot b \cdot b$ 为假设块采样值 $q_{i,j}$ 的维度;

[0022] 利用三个卷积层来实现 $p(\cdot)$ 函数,三个卷积层的具体参数为:卷积核大小:都为 1×1 ,输入信道数:第一个卷积层为 $SR \cdot b \cdot b$,其余为 $b \cdot b$,输出信道数:都为 $b \cdot b$ 。

[0023] 利用 $p(\cdot)$ 函数,将所得到的假设块的采样值张量转换成一个 $h \times w \times (b \cdot b) \times n$ 的

假设块特征图。

[0024] 利用 $p(\cdot)$ 函数,将得到的非关键帧采样值张量(大小为 $h \times w \times (SR \cdot b \cdot b)$)进行卷积,得到一个 $h \times w \times (b \cdot b)$ 大小的非关键帧特征图。

[0025] 将得到的假设块特征图与得到的非关键帧特征图进行矩阵乘法,实现公式(1)中 $p(q_{i,j})^T p(y_i)$,得到一个 $h \times w \times n$ 大小的张量;利用softmax函数对所得张量的最后一维进行

归一化实现公式(1)的 $\frac{1}{\sum_{\forall j} e^{p(q_{i,j})^T p(y_i)}} \sum_{\forall k} e^{p(q_{i,j})^T p(y_i)}$,得到大小为 $h \times w \times n$ 的系数张量;

[0026] 将所得的系数张量与所得的假设块张量进行矩阵乘法,得到大小为 $h \times w \times (b \cdot b)$ 的预测值张量;

[0027] 将所得的预测值张量进行变形,得到一个大小为 $(h \cdot b) \times (w \cdot b)$ 的当前非关键帧的多假设预测图像。

[0028] 进一步地,各个假设块的权重由嵌入高斯函数决定,具体如下:

$$[0029] \quad \omega_{i,j} = \frac{e^{p(q_{i,j})^T p(y_i)}}{\sum_{\forall j} e^{p(q_{i,j})^T p(y_i)}} \quad (2)。$$

[0030] 进一步地,构建残差重构模块的具体方法为:

[0031] 多假设预测图像进行自适应采样,得到多假设预测图像的采样值,维度大小为 $h \times w \times (SR \cdot b \cdot b)$;

[0032] 将多假设采样值与基于图像块采样值得非关键帧采样值相减,得到残差采样值;

[0033] 将残差采样值通过一个卷积层,转换成大小为 $h \times w \times (b \cdot b)$ 的特征图。此卷积层的卷积核大小为 1×1 ,输入信道数为 $SR \cdot b \cdot b$,输出信道数为 $b \cdot b$ 。

[0034] 将输出特征图进行变形,得到一个 $(h \cdot b) \times (w \cdot b)$ 大小的特征图;

[0035] 利用卷积层对得到的特征图进行卷积,得到残差图像的重构结果;

[0036] 得到的残差重构结果与多假设预测图像相加,输出最后的非关键帧重构值。

[0037] 在以上技术方案基础上优选地,为了使重构速度更快、图像质量更高,优选地,利用8个卷积层对得到的特征图进行卷积,得到残差图像的重构结果;8个卷积层的具体参数如下:卷积核大小:都为 3×3 ,输入信道数:第一个卷积层为1,其余为64,输出信道数:最后一个卷积层为1,其余都为64。

[0038] 作为以上技术方案的优选方案,自适应采样方法如下:对信号的采集如下式所示:

$$[0039] \quad y = \Phi x \quad (3),$$

[0040] 其中, y 为采样值, Φ 为采样矩阵, x 为原始信号,式(3)由一个卷积层来实现,此卷积层的卷积核大小为 $b \times b$,输入信道数为1,输出信道数为 $SR \cdot b \cdot b$,其中, SR 为非关键帧的采样率,此卷积层等同于对图像中每一个 $b \times b$ 大小的图像块进行压缩感知采样,输出为一个 $h \times w \times (SR \cdot b \cdot b)$ 的张量,其中, h 和 w 代表输入图像的高和宽两个维度上的块的数量,卷积层的权重为对应的采样矩阵。

[0041] 作为以上技术方案的优选方案,利用DNN的图像压缩感知重构方法对关键帧进行

重构。

[0042] 另一方面,本发明提供一种基于深度神经网络的视频压缩感知与重构装置,其特征在于,包括:

[0043] 视频序列输入模块、深度神经网络模块、非关键划分模块、关键帧重构模块和视频信号重构输出模块;

[0044] 所述视频序列输入模块,用于输入视频序列;

[0045] 所述非关键划分模块,用于将视频序列输入模块输入的视频信号序列分为多个图片组,确定每一个图片组的关键帧和非关键帧;

[0046] 所述深度神经网络模块,用于对非关键划分模块划分的非关键帧进行重构得到重构的非关键帧;

[0047] 所述关键帧重构模块,用于对非关键划分模块划分得到的关键帧利用图像压缩感知重构方法进行重构得到重构后的关键帧;

[0048] 所述视频信号重构输出模块:将关键帧重构模块重构的关键帧和深度神经网络输出的重构的非关键帧按顺序排列即为所得重构视频信号;

[0049] 所述深度神经网络模块包括图像块划分模块、自适应采样模块、多假设预测模块、残差重构模块、子网络训练模块和深度神经网络训练模块,其中自适应采样模块、多假设预测模块和残差重构模块构成的端到端的深度神经网络;

[0050] 所述视频序列输入模块,用于输入用于训练的视频序列;

[0051] 所述非关键划分模块,用于将视频序列输入模块输入的训练视频序列分为多个图片组,确定每一个图片组的关键帧和非关键帧;

[0052] 所述图像块划分模块,用于将非关键划分模块得到的训练视频序列非关键帧分为多个互不重叠的图像块;

[0053] 所述自适应采样模块,用于对非关键帧中每个图像块进行自适应采样得到图像块采样值;

[0054] 所述多假设预测模块,用于确定每个图像块对应的假设块;对非关键帧中的每个图像块利用对应的假设块进行自适应采样得到假设块的采样值,根据假设块采样值和图像块采样值,得到多假设预测图像;

[0055] 所述残差重构模块,用于对多假设预测图像进行自适应采样,构建多假设预测图像采样值与非关键帧采样值的残差值,根据残差值利用残差重构方法得到非关键帧重构;

[0056] 所述子网络训练模块,用于将构建的自适应采样模块与多假设预测模块进行级联构建子网络,对子网络进行训练;

[0057] 所述深度神经网络训练模块,用于将子网络预训练得到的参数作为初始值,对整个深度神经网络进行训练,得到训练完成的深度神经网络;

[0058] 与现有技术相比,本发明具有如下有益技术效果:

[0059] 本发明设计了一种特殊的深度神经网络,主要包含三部分:自适应采样模块,多假设预测模块,以及残差重构模块。由于深度神经网络的特性,此方法大大减少了视频信号的重构时间,使其达到实时性的要求。同时,此网络从训练数据中学习如何使用视频信号的时间相关性和空间相关性,以提高视频信号的重构质量;提出的深度神经网络,输入为已重构的关键帧和当前的非关键帧;输出为非关键帧的重构信号。对此网络的参数进行适当的训

练之后,将网络的自适应采样模块应用于无线传感网的编码节点,将多假设预测模块和残差重构模块应用于无线传感网的解码端,即可对视频信号进行低延迟,高质量的恢复;

[0060] 本发明采用一种自适应的块采样方法,提高采样过程中对有效信息的保留和对冗余信息的删减。同时,由于采样对象是块而非图像,因而减少对视频编码端的存储和运算的负担,得到的残差因为信号能量低,更易于重构;

[0061] 本发明充分利用深度神经网络来实现视频帧的多假设预测,其中各个假设的权重通过学习训练数据得到,使得预测结果准确率更高。

[0062] 本发明充分利用深度神经网络来实现残差重构,通过可学习的卷积层对整个残差图像卷积,减少了块效应,增加了可视野,从而提高了重构质量。

[0063] 本发明通过一个端到端的神经网络,实现了自适应采样,多假设预测和残差重构。由于深度神经网络低时间复杂度的特性,本发明提出的方法算法复杂度低,耗时短,能够满足实时性要求。

附图说明

[0064] 图1为本发明具体实施例一种基于深度神经网络的视频压缩感知与重构方法的框架图;

[0065] 图2为本发明具体实施例自适应采样模块的示意图;

[0066] 图3为本发明具体实施例多假设预测模块的流程图;

[0067] 图4为本发明具体实施例残差重构模块的流程图。

具体实施方式

[0068] 以下结合具体实施例,对本发明进行详细说明。

[0069] 如图1所示,该实施例提供了一种基于深度神经网络的视频压缩感知与重构方法,该方法具体步骤如下:

[0070] 参数设定:非关键帧采样率SR为0.01,0.04或者0.10;块的尺寸b为33;搜索窗大小W为 45×45

[0071] (1) 收集用于网络训练的高质量视频序列。

[0072] (2) 将用于训练的视频信号序列分为多个图片组(Group of Pictures,GOP)。每一个GOP的第一帧为关键帧,拥有较高采样率;其余为非关键帧,采样率低于关键帧。由于较高的采样率,关键帧拥有较好的重构质量,被用于指导非关键帧的重构。

[0073] (3) 对关键帧利用已有的图像压缩感知重构方法进行重构,具体实现方法是本领域的公知常识,不过多介绍。优选地,为了避免构整个视频的时候重构时间很长,使本发明重构更快,优选地,采用一种基于DNN的图像压缩感知重构方法进行重构,如 DR^2 -Net或ReconNet和 DR^2 -Net,其实现方法也是本领域的公知常识。

[0074] (4) 构建视频压缩感知网络的自适应采样模块,如图2所示。在压缩感知中,对信号的采集如下式所示:

[0075] $y = \Phi x$ (3),

[0076] 其中,y为采样值, Φ 为采样矩阵,x为原始信号。在此网络中,式(3)由一个卷积层来实现,并应用于非关键帧的采样。具体的,此卷积层的卷积核大小为 $b \times b$,输入信道数为

1, 输出信道数为 $SR \cdot b \cdot b$ 。其中, SR 为非关键帧的采样率。

[0077] 此卷积层等同于对图像中每一个 $b \times b$ 大小的图像块进行压缩感知采样, 输出为一个 $h \times w \times (SR \cdot b \cdot b)$ 的张量, 其中, h 和 w 代表非关键帧高和宽两个维度上的块的数量 (即用非关键帧的宽和高分别除以 b)。卷积层的权重即为对应的采样矩阵, 由于卷积层的参数是可训练的, 因此在此网络中的采样是自适应的。

[0078] (5) 构建视频压缩感知网络的多假设预测模块, 如图3所示。此模块将非关键帧分为多个互相不重叠的图像块, 大小为 $b \times b$, 对每一个图像块利用其对应假设块的线性组合来进行预测。具体的, 每一个图像块的预测值由下式计算得到:

$$\begin{aligned}
 P_i &= \sum_{\forall j} \omega_{i,j} h_{i,j} \\
 [0079] \quad &= \frac{1}{\sum_{\forall j} e^{p(q_{i,j})^T p(y_i)}} \sum_{\forall j} e^{p(q_{i,j})^T p(y_i)} h_{i,j}
 \end{aligned} \tag{1),$$

[0080] 其中, P_i 为当前非关键帧中第 i 个图像块的预测结果, $h_{i,j}$ 为第 i 个图像块中第 j 个假设块, $\omega_{i,j}$ 为此假设块的权重, 是一个关于图像块和假设块的采样值的函数, $q_{i,j}$ 为利用步骤 (4) 的采样方式对假设块进行采样的结果, y_i 为第 i 个图像块的采样值, $p(\cdot)$ 为一个非线性映射。

[0081] 具体实施例中假设块的权重 $\omega_{i,j}$ 可采用以下函数实现:

$$[0082] \quad \omega_{i,j} = \frac{f(q_{i,j}, y_i)}{\sum_{\forall j} f(q_{i,j}, y_i)},$$

[0083] 其中, $f(q_{i,j}, y_i)$ 是关于 $q_{i,j}$ 和 y_i 的函数。

[0084] 以上函数有多种实施方式, 优选地在具体实施例中, 各个假设块的权重 $\omega_{i,j}$ 由嵌入高斯函数决定, 具体如下式:

$$[0085] \quad \omega_{i,j} = \frac{e^{p(q_{i,j})^T p(y_i)}}{\sum_{\forall j} e^{p(q_{i,j})^T p(y_i)}} \tag{2)。$$

[0086] 在网络中, 公式 (1) 具体实现方法为:

[0087] (a) 从步骤 (3) 中已重构关键帧中提取假设块。具体方法为, 对于非关键帧中每一个图像块, 以其位置为中心, 在关键帧中框定一个大小为 $W \times W$ 的矩形搜索窗, 将搜索窗中所有的重叠的图像块提取作为当前图像块的假设块。将非关键帧中每一个图像块的所有假设块整合在一起, 可得一个 $h \times w \times (b \cdot b) \times n$ 的假设块张量, 其中, n 代表了每一个图像块对应的假设块的个数, $h \times w$ 代表待预测的图像块的个数, $b \cdot b$ 为假设块的维度。

[0088] (b) 利用步骤 (4) 中的自适应采样方法对步骤 (a) 得到的假设块张量进行采样, 得到一个包含所有图像块的各个假设块采样值的张量, 其维度大小为: $h \times w \times (SR \cdot b \cdot b) \times n$, 其中 $SR \cdot b \cdot b$ 为假设块采样值 $q(i, j)$ 的维度。。

[0089] (c) 利用三个卷积层来实现 $p(\cdot)$ 函数,三个卷积层的具体参数为:卷积核大小:都为 1×1 ,输入信道数:第一个卷积层为 $SR \cdot b \cdot b$,其余为 $b \cdot b$,输出信道数:都为 $b \cdot b$ 。

[0090] (d) 利用 $p(\cdot)$ 函数,将步骤(b)所得到的假设块的采样值张

[0091] 量转换成 $h \times w \times (b \cdot b) \times n$ 的特征图。

[0092] (e) 利用 $p(\cdot)$ 函数,对步骤(4)得到的非关键帧采样值张量(大小为 $h \times w \times (SR \cdot b \cdot b)$)进行卷积,得到一个 $h \times w \times (b \cdot b)$ 大小的特征图。

[0093] (f) 将步骤(d)得到的特征图与步骤(e)得到的特征图进行矩阵乘法,即可实现公式(1)中 $p(q_{i,j})^T p(y_i)$,得到一个 $h \times w \times n$ 大小的张量。

[0094] (g) 对步骤(e)所得张量的最后一维利用softmax函数进行归一化,即可实现公式

(1) 中的
$$\frac{1}{\sum_{vj} e^{p(q_{i,j})^T p(y_i)}} \sum_{vj} e^{p(q_{i,j})^T p(y_i)} h_{ij}$$
 得到大小为 $h \times w \times n$ 的系数张量。

[0095] (h) 将步骤(g)所得的系数张量与步骤(a)所得的假设块张量进行矩阵乘法,得到大小为 $h \times w \times (b \cdot b)$ 的预测值张量。

[0096] (i) 将步骤(h)所得的预测值张量进行变形,即可得到一个大小为 $(h \cdot b) \times (w \cdot b)$ 的当前非关键帧的多假设预测图像。

[0097] (6) 构建视频压缩感知网络的残差重构模块,如图4所示。此模块用于重构出原始非关键帧与步骤(5)得到的多假设预测图像的残差值。由于残差信号相比于图像信号能量更低,所以残差信号的重构更为容易。此残差重构模块的具体的实现方法为:

[0098] (a) 对步骤(5)得到的多假设预测图像采用步骤(4)的方法进行采样,得到多假设预测图像的采样值,维度大小为 $h \times w \times (SR \cdot b \cdot b)$ 。

[0099] (b) 将步骤(a)的多假设采样值与步骤(4)得到的非关键帧的采样值相减,得到残差采样值。

[0100] (c) 将残差采样值通过一个卷积层,转换成 $h \times w \times (b \cdot b)$ 的特征图。此卷积层的卷积核大小为 1×1 ,输入信道数为 $SR \cdot b \cdot b$,输出信道数为 $b \cdot b$ 。

[0101] (d) 将步骤(c)的输出特征图进行变形,得到一个 $(h \cdot b) \times (w \cdot b)$ 大小的特征图。

[0102] (e) 优选地,利用8个卷积层对步骤(d)得到的特征图进行卷积,得到残差图像的重构结果。8个卷积层的具体参数如下:卷积核大小:都为 3×3 ,输入信道数:第一个卷积层为1,其余为64,输出信道数:最后一个卷积层为1,其余都为64。

[0103] (f) 将步骤(e)得到的残差重构结果与步骤(5)的多假设预测结果相加,输出最后的非关键帧重构值。

[0104] (7) 将步骤(4)中的自适应采样模块与步骤(5)的多假设预测模块级联,组成一个子网络,此子网络输入为原始非关键帧以及步骤(3)重构的关键帧,输出为多假设预测图像。将此子网络各层的初始参数设定为随机值,进行预训练。训练中,损失函数为均方误差,标签为真实的视频帧的图像。

[0105] (8) 将步骤(7)预训练得到的参数作为初始值,对整个网络进行训练,设计损失函数,使其在不断的训练过程中,降低损失值,从而得到训练好的网络参数。训练中,损失函数为均方误差。

[0106] (9) 将步骤(8)训练得到的网络带入实际无线视频传感网系统中,对视频信号进行重构,具体实现方法为:

[0107] (a) 利用步骤(2)相同的方法对视频序列的关键帧进行独立重构。

[0108] (b) 利用步骤(8)训练得到的网络对视频序列的非关键帧进行重构。

[0109] (c) 将步骤(a)重构的关键帧和步骤(b)重构的非关键帧按顺序排列即为所得重构视频信号。

[0110] 下面结对本发明方法的效果做进一步说明。

[0111] 表1本发明实施例结果与现有技术的非关键帧重构质量对比表

方法	采样率		
	0.01	0.04	0.10
D-AMP	6.20	13.85	26.68
ReconNet	21.20	24.27	27.45
DR2-Net	21.67	25.79	29.64
MH-BCS-SPL	24.36	28.60	31.31
本发明所提方法	28.36	32.82	36.09

[0112] 衡量标准:(峰值信噪比)PSNR,测试对象:100个GOP大小为7的视频序列,非关键帧采样率:0.01、0.04和0.10。

[0114] 由表1可以看出,本发明方法相比于已有方法都有着明显的效果提升。对比于已有最好方法(MH-BCS-SPL),本发明PSNR在0.01采样率时提升了4dB,在0.04采样率时提升了4.22dB,在0.10采样率时提升了4.78dB。

[0115] 表2本发明实施例与现有技术的非关键帧单帧平均重构时间对比表

方法	采样率		
	0.01	0.04	0.10
D-AMP	57.9816	58.9956	50.0234
ReconNet	0.0101	0.0101	0.0101
DR2-Net	0.0225	0.0226	0.0233
MH-BCS-SPL	4.7552	4.8278	4.7085
本发明所提方法	0.0241	0.0251	0.025

[0117] 时间单位:秒(s),测试对象:100个GOP大小为7的视频序列,非关键帧采样率:0.01、0.04和0.10。

[0118] 由表2可以看出,本发明方法与同是基于深度神经网络的ReconNet方法和DR²-Net方法重构时间在同一数量级,比MH-BCS-SPL方法快两个数量级,比D-AMP快三个数量级。因此,本发明方法拥有着目前领先的重构速度,适用于实时的视频传感系统。

[0119] 本发明方法将视频信号划分为关键帧与非关键帧;对于关键帧,采用已有的图像重构方法进行重构。对于非关键帧,此方法创新性的提出了一种特殊的深度神经网络来完成重构。此神经网络由自适应采样模块,多假设预测模块,残差重构模块组成,充分利用了视频信号的时空相关性来完成视频信号的采样与重构,在提高重构质量的同时,保证了算法的低时间复杂度。因此,本发明方法适用于采样端资源受限、对重构质量要求高、对实时性要求高的视频传感系统。

[0120] 实施例:一种基于深度神经网络的视频压缩感知与重构装置,包括:

[0121] 视频序列输入模块、深度神经网络模块、非关键划分模块、关键帧重构模块和视频信号重构输出模块;

[0122] 所述视频序列输入模块,用于输入视频序列;

[0123] 所述非关键划分模块,用于将视频序列输入模块输入的视频信号序列分为多个图片组,确定每一个图片组的关键帧和非关键帧;

[0124] 所述深度神经网络模块,用于对非关键划分模块划分的非关键帧进行重构得到重构的非关键帧;

[0125] 所述关键帧重构模块,用于对非关键划分模块划分得到的关键帧利用图像压缩感知重构方法进行重构得到重构后的关键帧;

[0126] 所述视频信号重构输出模块:将关键帧重构模块重构的关键帧和深度神经网络输出的重构的非关键帧按顺序排列即为所得重构视频信号;

[0127] 所述深度神经网络模块包括图像块划分模块、自适应采样模块、多假设预测模块、残差重构模块、子网络训练模块和深度神经网络训练模块,其中自适应采样模块、多假设预测模块和残差重构模块构成的端到端的深度神经网络;

[0128] 所述视频序列输入模块,用于输入用于训练的视频序列;

[0129] 所述非关键划分模块,用于将视频序列输入模块输入的训练视频序列分为多个图片组,确定每一个图片组的关键帧和非关键帧;

[0130] 所述图像块划分模块,用于将非关键划分模块得到的训练视频序列非关键帧分为多个互不重叠的图像块;

[0131] 所述自适应采样模块,用于对非关键帧中每个图像块进行自适应采样得到图像块采样值;

[0132] 所述多假设预测模块,用于确定每个图像块对应的假设块;对非关键帧中的每个图像块利用对应的假设块进行自适应采样得到假设块的采样值,根据假设块采样值和图像块采样值,得到多假设预测图像;

[0133] 所述残差重构模块,用于对多假设预测图像进行自适应采样,构建多假设预测图像采样值与非关键帧采样值的残差值,根据残差值利用残差重构方法得到非关键帧重构;

[0134] 所述子网络训练模块,用于将构建的自适应采样模块与多假设预测模块进行级联构建子网络,对子网络进行训练;

[0135] 所述神经网络训练模块,用于将子网络预训练得到的参数作为初始值,对整个神经网络进行训练,得到训练完成的神经网络。

[0136] 该实施例中每个模块的具体实现方式如上述一种基于深度神经网络的视频压缩感知与重构方法的技术方案中描述,不再赘述。

[0137] 本发明方法基于神经网络设计了自适应采样模块,多假设预测模块,残差重构模块。此三个模块充分利用了视频信号的时空相关性来完成非关键帧的采样与重构,并保证了算法的低时间复杂度。因此,本发明方法适用于采样端资源受限,对重构质量要求高,对实时性要求高的视频传感系统。

[0138] 本领域内的技术人员应明白,本申请的实施例可提供为方法、系统、或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0139] 本申请是参照根据本申请实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0140] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0141] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或

其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0142] 以上结合附图对本发明的实施例进行了描述,但是本发明并不局限于上述的具体实施方式,上述的具体实施方式仅仅是示意性的,而不是限制性的,本领域的普通技术人员在本发明的启示下,在不脱离本发明宗旨和权利要求所保护的范围情况下,还可做出很多形式,这些均属于本发明的保护之内。

[0143] 在本实用新型的描述中,需要理解的是,术语“内侧”、“外侧”、“上层”、“顶”、“下层”、“左”、“右”、“垂直”、“水平”、“平行”、“底”、“内”、“外”等指示的方位或位置关系为基于附图所示的方位或位置关系,仅是为了便于描述本实用新型和简化描述,而不是指示或暗示所指的装置或元件必须具有特定的方位、以特定的方位构造和操作,因此不能理解为对本实用新型保护范围的限制。

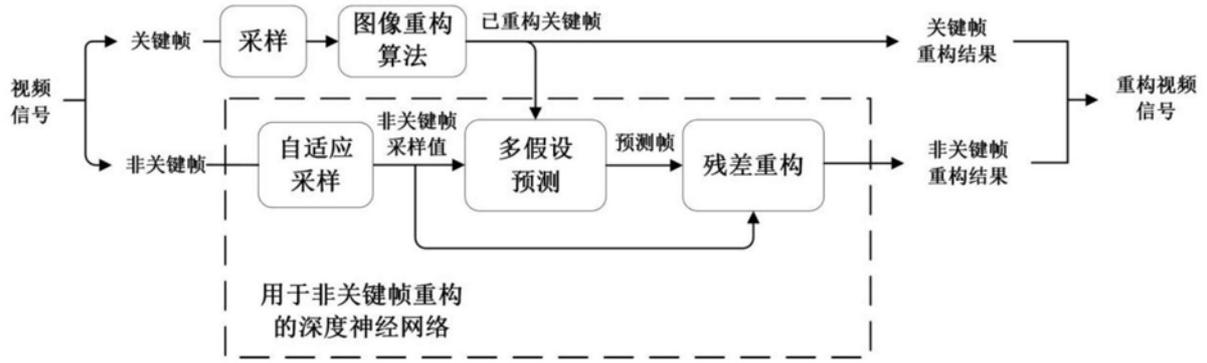


图1

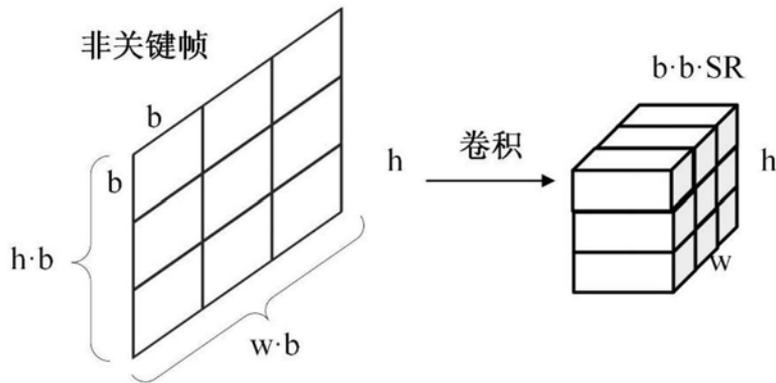


图2

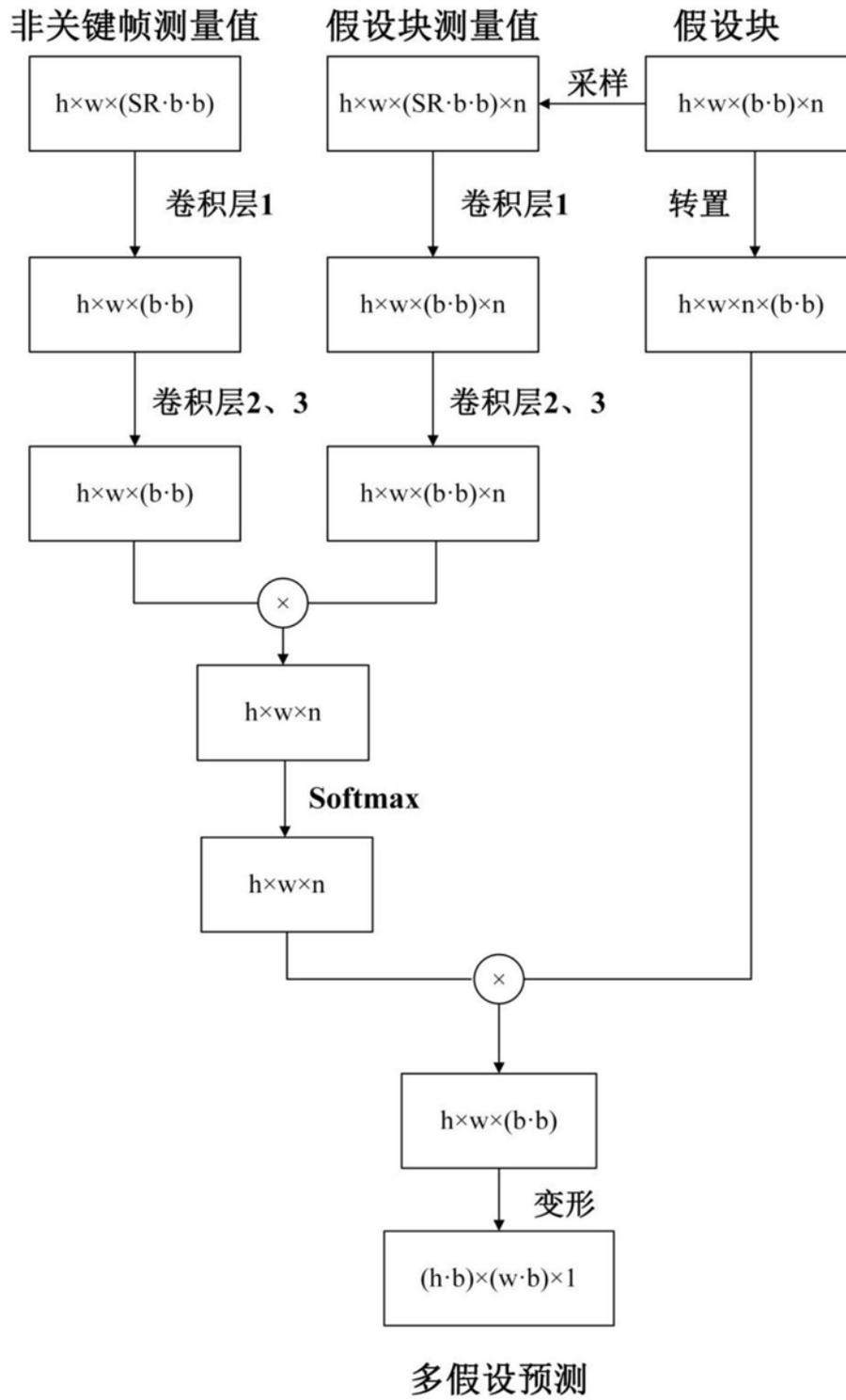


图3

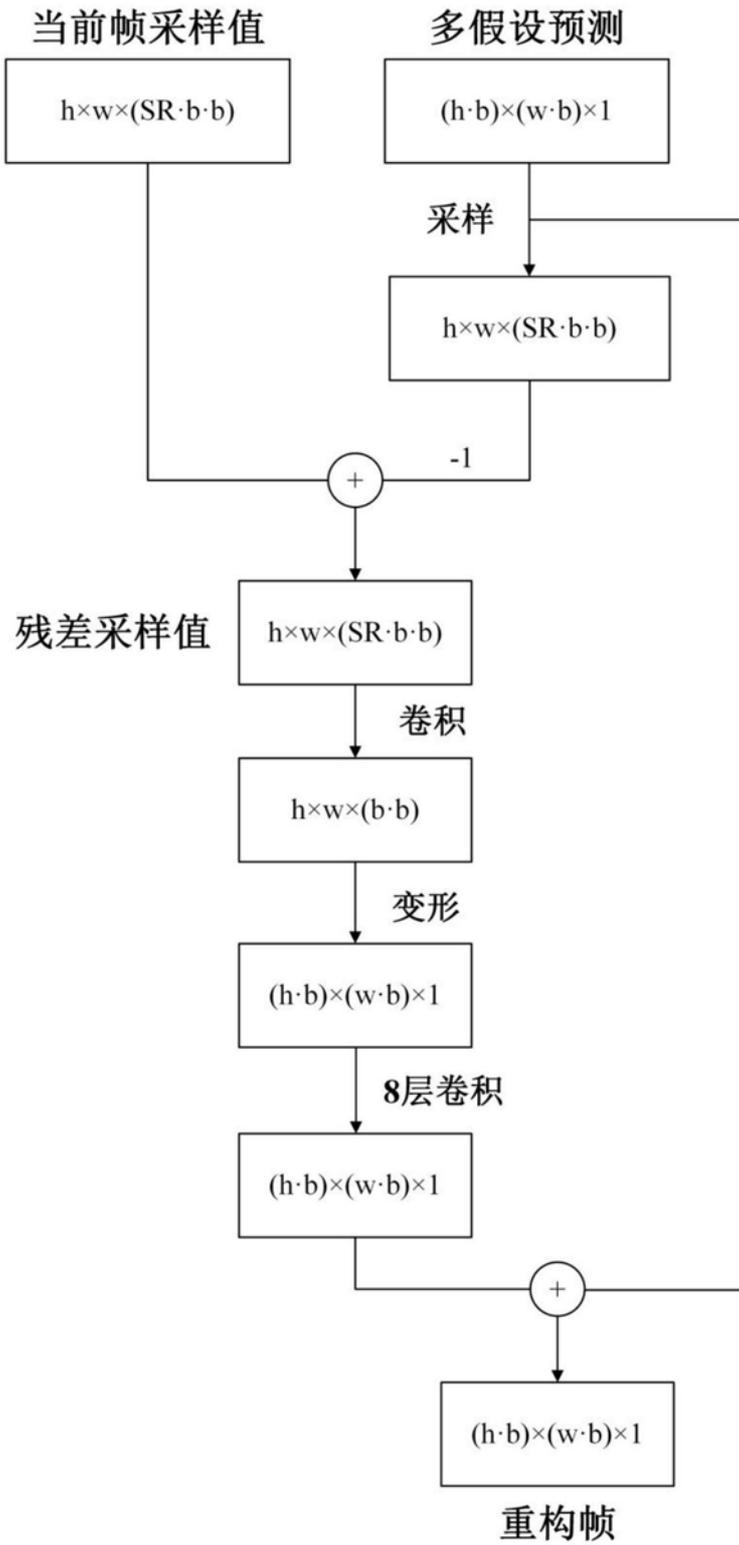


图4