

一种解码、编码方法、装置及其设备

技术领域

[01]本申请涉及编解码技术领域，尤其是涉及一种解码、编码方法、装置及其设备。

背景技术

[02]为了达到节约空间的目的，视频图像都是经过编码后才传输的，完整的视频编码可以包括预测、变换、量化、熵编码、滤波等过程。针对预测过程，预测可以包括帧内预测和帧间预测，帧间预测是指利用视频时间域的相关性，使用邻近已编码图像的像素预测当前像素，以达到有效去除视频时域冗余的目的。帧内预测是指利用视频空间域的相关性，使用当前帧图像的已编码块的像素预测当前像素，以达到去除视频空域冗余的目的。

[03]随着深度学习的迅速发展，深度学习在许多高层次的计算机视觉问题上取得成功，如图像分类、目标检测等，深度学习也逐渐在编解码领域开始应用，即可以采用神经网络对图像进行编码和解码。虽然基于神经网络的编解码方法展现出巨大性能潜力，但是，基于神经网络的编解码方法，仍然存在解码性能较差和复杂度较高等问题。

发明内容

[04]本申请提供一种解码、编码方法、装置及其设备，能够提高解码性能，降低复杂度。

[05]本申请提供一种解码方法，应用于解码端，所述方法包括：对当前图像块对应的一个码流进行解码，得到所述当前图像块对应的系数超参特征；基于所述系数超参特征确定概率分布参数，基于所述概率分布参数对所述当前图像块对应的另一码流进行解码，得到所述当前图像块对应的重建特征；将所述重建特征输入合成变换网络，得到所述当前图像块对应的重建图像块；其中，所述合成变换网络至少包括注意力模块，所述注意力模块为级联混合注意力模块；其中，所述级联混合注意力模块包括第一注意力子模块和第二注意力子模块，且所述第一注意力子模块和所述第二注意力子模块为串行的两个子模块。

[06]本申请提供一种编码方法，应用于编码端，所述方法包括：对当前图像块对应的一个码流进行解码，得到所述当前图像块对应的系数超参特征；基于所述系数超参特征确定概率分布参数，基于所述概率分布参数对所述当前图像块对应的另一码流进行解码，得到所述当前图像块对应的重建特征；将所述重建特征输入合成变换网络，得到所述当前图像块对应的重建图像块；其中，所述合成变换网络至少包括注意力模块，所述注意力模块为级联混合注意力模块；其中，所述级联混合注意力模块包括第一注意力子模块和第二注意力子模块，且所述第一注意力子模块和所述第二注意力子模块为串行的两个子模块。

[07]本申请提供一种解码装置，应用于解码端，所述装置包括：解码模块，用于对当前图像块对应的一个码流进行解码，得到所述当前图像块对应的系数超参特征；基于所述系数超参特征确定概率分布参数，基于所述概率分布参数对所述当前图像块对应的另一码流进行解码，得到所述当前图像块对应的重建特征；处理模块，用于将所述重建特征输入合成变换网络，得到所述当前图像块对应的重建图像块；其中，所述合成变换网络至少包括注意力模块，所述注意力模块为级联混合注意力模块；其中，所述级联混合注意力模块包括第一注意力子模块和第二注意力子模块，且所述第一注意力子模块和所述第二注意力子模块为串行的两个子模块。

[08]本申请提供一种编码装置，应用于编码端，所述装置包括：解码模块，用于对当前图像块对应的一个码流进行解码，得到所述当前图像块对应的系数超参特征；基于所述系数超参特征确定概率分布参数，基于所述概率分布参数对所述当前图像块对应的另一

码流进行解码,得到所述当前图像块对应的重建特征;处理模块,用于将所述重建特征输入合成变换网络,得到所述当前图像块对应的重建图像块;其中,所述合成变换网络至少包括注意力模块,所述注意力模块为级联混合注意力模块;其中,所述级联混合注意力模块包括第一注意力子模块和第二注意力子模块,且所述第一注意力子模块和所述第二注意力子模块为串行的两个子模块。

[09]本申请提供一种解码端设备,包括:处理器和机器可读存储介质,所述机器可读存储介质存储有能够被所述处理器执行的机器可执行指令;所述处理器用于执行机器可执行指令,以实现上述的解码方法。

[10]本申请提供一种编码端设备,包括:处理器和机器可读存储介质,所述机器可读存储介质存储有能够被所述处理器执行的机器可执行指令;所述处理器用于执行机器可执行指令,以实现上述的编码方法。

[11]本申请提供一种电子设备,包括:处理器和机器可读存储介质,所述机器可读存储介质存储有能够被所述处理器执行的机器可执行指令;所述处理器用于执行机器可执行指令,以实现上述的解码方法或者编码方法。

[12]本申请提供一种机器可读存储介质,所述机器可读存储介质上存储有若干计算机指令,所述计算机指令被处理器执行时,用于实现上述的解码方法或者编码方法。

[13]本申请提供一种计算机应用程序,所述计算机应用程序令被处理器执行时,实现上述的解码方法或者编码方法。

[14]由以上技术方案可见,本申请实施例中,针对基于神经网络的编解码技术,提出一种注意力机制的合成变换网络,该合成变换网络包括注意力模块,且注意力模块为级联混合注意力模块,通过级联混合注意力模块实现合成变换网络时,在保证合成图像质量的同时,有效降低网络复杂度和计算复杂度,提高解码性能,在降低复杂度的同时,有效保证重建图像块的质量,提升编码性能和解码性能。

附图说明

[15]图1是本申请一种实施方式中的三维特征矩阵的示意图;

[16]图2是本申请一种实施方式中的像素重组的示意图;

[17]图3是本申请一种实施方式中的解码方法的流程图;

[18]图4是本申请一种实施方式中的编码端的处理过程的示意图;

[19]图5是本申请一种实施方式中的解码端的处理过程的示意图;

[20]图6A-图6D是本申请一种实施方式中的合成变换网络的结构示意图;

[21]图7A-图7B是本申请一种实施方式中的合成变换网络的结构示意图;

[22]图8A-图8H是本申请一种实施方式中的合成变换网络的结构示意图;

[23]图9A是本申请一种实施方式中的解码端设备的硬件结构图;

[24]图9B是本申请一种实施方式中的编码端设备的硬件结构图。

具体实施方式

[25]在本申请实施例使用的术语仅仅是出于描述特定实施例的目的,而非用于限制本申请。本申请实施例和权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式,除非上下文清楚地表示其它含义。还应当理解,本文中使用的术语“和/或”是指包含一个或多个相关联的列出项目的任何或所有可能组合。应当理解,尽管在本申请实施例可能采用术语第一、第二、第三等来描述各种信息,但是,这些信息不应限于

这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如，在不脱离本申请实施例范围的情况下，第一信息也可以被称为第二信息，类似地，第二信息也可以被称为第一信息，取决于语境。此外，所使用的词语“如果”可以被解释成为“在……时”，或“当……时”，或“响应于确定”。

[26]本申请实施例中提出一种解码方法，该解码方法可以涉及如下概念：

[27]熵编码 (Entropy Encoding)：熵编码即编码过程中按照熵原理不丢失任何信息的编码，信息熵为信源的平均信息量（不确定性的度量）。熵编码的编码方式可以包括但不限于：香农(Shannon)编码、哈夫曼(Huffman)编码和算术编码(arithmetic coding)。

[28]神经网络 (Neural Network, NN)：神经网络是指人工神经网络，神经网络是一种运算模型，由大量节点（或称为神经元）之间相互联接构成。在神经网络中，神经元（常被称为处理单元）可表示不同的对象，如特征、字母、概念，或者一些有意义的抽象模式。神经网络中处理单元的类型可以分为三类：输入单元、输出单元和隐单元。输入单元接受外部世界的信号与数据；输出单元实现处理结果的输出；隐单元是处在输入和输出单元之间，不能由系统外部观察的单元。神经元之间的连接权值反映了单元间的连接强度，信息的表示和处理体现在处理单元的连接关系中。神经网络是一种非程序化、类大脑风格的信息处理方式，神经网络的本质是通过神经网络的变换和动力学行为得到一种并行分布式的信息处理功能，并在不同程度和层次上模仿人脑神经系统的信息处理功能。在视频处理领域，常用的神经网络可以包括但不限于：卷积神经网络 (Convolutional Neural Network, CNN)、循环神经网络 (Recurrent Neural Network, RNN)、全连接网络等。

[29]卷积神经网络 (CNN)：卷积神经网络是一种前馈神经网络，是深度学习技术中极具代表的网络结构之一，卷积神经网络的人工神经元可以响应一部分覆盖范围内的周围单元，对于大型图像处理有出色表现。卷积神经网络的基本结构包括两层，其一为特征提取层（也称卷积层），每个神经元的输入与前一层的局部接受域相连，并提取该局部的特征。一旦该局部特征被提取后，它与其它特征间的位置关系也随之确定下来。其二是特征映射层（也称激活层），神经网络的每个计算层由多个特征映射组成，每个特征映射是一个平面，平面上所有神经元的权值相等。特征映射结构可采用 Sigmoid 函数、ReLU (Rectified Linear Unit) 函数、Leaky-ReLU 函数、PReLU (Parametric Rectified Linear Unit) 函数、GDN (Generalized Divisive Normalization) 函数等作为卷积神经网络的激活函数。此外，由于一个映射面上的神经元共享权值，因而减少了网络自由参数的个数。

[30]示例性的，卷积神经网络相较于图像处理算法的优点之一在于，避免了对图像复杂的前期预处理过程（提取人工特征等），可直接输入原始图像进行端到端的学习。卷积神经网络相较于普通神经网络的优点之一在于，普通神经网络都是采用全连接的方式，即输入层到隐藏层的神经元都是全部连接的，这样做将导致参数量巨大，使得网络训练耗时甚至难以训练，而卷积神经网络则通过局部连接、权值共享等方法避免了这一困难。

[31]反卷积层 (Deconvolution)：又称为转置卷积层，反卷积层和卷积层的工作过程很相似，主要区别在于，反卷积层会通过 padding（填充），使得输出大于输入（也可以保持相同）。若 stride（步长）为 1，则表示输出尺寸等于输入尺寸；若 stride 为 N，则表示输出特征的宽为输入特征的宽的 N 倍，输出特征的高为输入特征的高的 N 倍。

[32]深度可分离卷积 (Depthwise Separable Convolution)：需要进行两次卷积操作，第一次卷积操作中，先进行 deep_wise 卷积（即收集每一层的特征）， $\text{kernel_size} = K * K * 1$ ，第一次卷积总的参数量为 $K * K * C_{in}$ 。第二次卷积操作中，为了得到 C_{out} 维度的输出， $\text{kernel_size} = 1 * 1 * C_{in}$ ，第二次卷积总的参数量为 $1 * 1 * C_{in} * C_{out}$ 。第二次卷积的输出可以为深度可分离卷积的输出，第一次卷积的输入可以为深度可分离卷积的输入。

[33]泛化能力 (Generalization Ability): 指机器学习算法对新鲜样本的适应能力, 学习的目的是学到隐含在数据背后的规律, 对具有同一规律的学习集以外的数据, 经过训练的网络也能够给出合适的输出, 该能力就可以称为泛化能力。

[34]特征 (feature): 是 $C \times W \times H$ 的三维特征矩阵或张量 (tensor)。参见图 1 所示, 是三维特征矩阵的示意图, 在三维特征矩阵中, C 表示通道 (channel) 数, H 表示特征高度, W 表示特征宽度。三维特征矩阵可以是神经网络的输入, 也可以是神经网络的输出。

[35]像素重组 (pixel shuffle): 也可以称为特征位置重组, 通过对低分辨率的特征进行多通道间的重组得到高分辨率的特征图。参见图 2 所示, 是像素重组的示意图, 在图 2 中, 给出了上采样系数为 2 的情况下的像素重组示意图。显然, 通道数为 4, 分辨率为 3×3 的特征通过特征位置重组层后变成通道数为 1, 分辨率为 6×6 的特征。

[36]率失真原则 (Rate-Distortion Optimized): 评价编码效率的有两大指标: 码率和 PSNR (Peak Signal to Noise Ratio, 峰值信噪比), 比特流越小, 则压缩率越大; PSNR 越大, 则重建图像质量越好。在模式选择时, 判别公式实质上也就是对二者的综合评价。例如, 模式对应的代价:

$$J(\text{mode}) = D + \lambda * R,$$

[37]其中, D 表示 Distortion (失真), 通常可以使用 SSE (sum-square error) 指标来进行衡量, SSE 是指重建图像块与源图像的差值的均方和, 为了实现代价考虑, 也可以使用 SAD (Sum of Absolute Difference) 指标, SAD 是指重建图像块与源图像的差值绝对值之和; λ 是拉格朗日乘子, R 就是该模式下图像块编码所需的实际比特数, 包括编码模式信息、运动信息、残差等所需的比特总和。在模式选择时, 若使用率失真原则去对编码模式做比较决策, 通常可以保证编码性能最佳。

[38]针对编码端的各个模块, 提出了非常多的编码工具, 而每个工具往往又有多种模式。对于不同视频序列, 能获得最优编码性能的编码工具往往不同。因此, 在编码过程中, 通常采用 RDO (Rate-Distortion Optimize) 比较不同工具或模式的编码性能, 以选择最佳模式。在确定最优工具或模式后, 再通过比特流中编码标记信息的方法传递工具或模式的决策信息。这种方法虽然带来了较高的编码复杂度, 但可以针对不同内容, 自适应选择最优的模式组合, 获得最优的编码性能。解码端可通过直接解析标志信息获得相关模式信息, 复杂度影响较小。

[39]在端到端的图像编码通用框架中, 主要包括特征主信息部分与超先验边信息部分, 特征主信息部分包括分析变换网络、量化、正态熵编码、正态熵解码与合成变换网络, 超先验边信息包括超先验分析网络、量化、因子化熵编码、因子化熵解码、超先验合成网络。图像分量 x 分别由特征主信息部分的分析变换网络与合成变换网络进行压缩编码与重建恢复; 超先验边信息部分主要用于建模特征主信息的概率, 指导进行特征主信息的熵编解码。在端到端的图像编码通用框架中, 存在计算复杂度高, 解码性能差等问题。

[40]针对上述发现, 本实施例中, 利用端到端的图像编码框架特点, 提出一种注意力机制的合成变换网络。该合成变换网络包括注意力模块, 且注意力模块为级联混合注意力模块。通过级联混合注意力模块实现合成变换网络时, 在保证合成图像质量的同时, 有效降低网络复杂度和计算复杂度, 提高解码性能。

[41]以下结合几个具体实施例, 对本申请实施例中的解码方法进行详细说明。

[42]实施例 1: 本申请实施例中提出一种解码方法, 参见图 3 所示, 为该解码方法的流程示意图, 该方法可以应用于解码端 (也称为视频解码器), 该方法可以包括:

[43]步骤 301、对当前图像块对应的一个码流进行解码, 得到当前图像块的系数超参特

征。

[44]步骤 302、基于该系数超参特征确定概率分布参数，并基于该概率分布参数对当前图像块对应的另一码流进行解码，得到当前图像块对应的重建特征。

[45]步骤 303、将该重建特征输入合成变换网络，得到当前图像块对应的重建图像块。其中，该合成变换网络至少包括注意力模块，且该注意力模块可为级联混合注意力模块。级联混合注意力模块可包括第一注意力子模块和第二注意力子模块，且第一注意力子模块和第二注意力子模块为串行的两个子模块，比如说，第一注意力子模块的输出特征为第二注意力子模块的输入特征。

[46]其中，当前图像块对应的一个码流可以是编码有当前图像块对应的系数超参特征的码流；当前图像块对应的另一码流可以是编码有当前图像块对应的残差特征的码流。

[47]示例性的，若注意力模块为级联混合注意力模块，可以通过第一注意力子模块对第一注意力子模块的输入特征进行第一处理，得到第一注意力子模块的输出特征，第一注意力子模块的输出特征为第二注意力子模块的输入特征；其中，第一处理可以为层归一化、卷积操作和维度变换操作的至少一种。可以通过第二注意力子模块对第二注意力子模块的输入特征进行第二处理，得到第二注意力子模块的输出特征；其中，第二处理可以为深度可分离卷积操作、层归一化、多层感知操作、线性操作、激活操作、下采样操作、残差卷积操作、上采样操作的至少一种。比如说，第二处理可以为深度可分离卷积操作、层归一化、多层感知操作的至少一种。或者，第二处理可以为层归一化、线性操作、激活操作的至少一种。或者，第二处理可以为下采样操作、残差卷积操作、上采样操作、激活操作的至少一种。

[48]在一种可能的实施方式中，通过第一注意力子模块对第一注意力子模块的输入特征进行第一处理，得到第一注意力子模块的输出特征，包括但不限于：对第一注意力子模块的输入特征进行层归一化，得到层归一化后的特征；对层归一化后的特征进行三路卷积操作，得到查询向量、键向量和值向量；对查询向量、键向量和值向量分别进行维度变换操作，得到维度变换后的查询向量、维度变换后的键向量和维度变换后的值向量；基于维度变换后的查询向量和维度变换后的键向量确定注意力权重，基于注意力权重和维度变换后的值向量确定该输入特征对应的修正特征；基于该输入特征和修正特征确定第一注意力子模块的输出特征。

[49]在一种可能的实施方式中，通过第一注意力子模块对第一注意力子模块的输入特征进行第一处理，得到第一注意力子模块的输出特征，可以包括但不限于：对第一注意力子模块的输入特征进行层归一化，得到层归一化后的特征；对层归一化后的特征进行窗口划分，得到多个小尺寸特征；针对每个小尺寸特征，对该小尺寸特征进行三路卷积操作，得到该小尺寸特征对应的小尺寸查询特征、小尺寸键特征和小尺寸值特征；可以对多个小尺寸特征对应的小尺寸查询特征进行拼接，得到查询向量；可以对多个小尺寸特征对应的小尺寸键特征进行拼接，得到键向量；可以对多个小尺寸特征对应的小尺寸值特征进行拼接，得到值向量。然后，对该查询向量、该键向量和该值向量分别进行维度变换操作，得到维度变换后的查询向量、维度变换后的键向量和维度变换后的值向量；基于维度变换后的查询向量和维度变换后的键向量确定注意力权重，并基于注意力权重和维度变换后的值向量确定该输入特征对应的修正特征；基于该输入特征和修正特征确定第一注意力子模块的输出特征。

[50]在一种可能的实施方式中，通过第一注意力子模块对第一注意力子模块的输入特征进行第一处理，得到第一注意力子模块的输出特征，可包括但不限于：对第一注意力子模块的输入特征进行层归一化，得到层归一化后的特征；对层归一化后的特征进行三路卷积操作，得到查询向量、键向量和值向量；基于查询向量和键向量确定注意力权重，

并基于注意力权重和值向量确定该输入特征对应的修正特征；基于该输入特征和修正特征确定第一注意力子模块的输出特征。

[51]在一种可能的实施方式中，通过第一注意力子模块对第一注意力子模块的输入特征进行第一处理，得到第一注意力子模块的输出特征，可以包括但不限于：对第一注意力子模块的输入特征进行层归一化，得到层归一化后的特征；对层归一化后的特征进行窗口划分，得到多个小尺寸特征；针对每个小尺寸特征，对该小尺寸特征进行三路卷积操作，得到该小尺寸特征对应的小尺寸查询特征、小尺寸键特征和小尺寸值特征；对多个小尺寸特征对应的小尺寸查询特征进行拼接，得到查询向量；对多个小尺寸特征对应的小尺寸键特征进行拼接，得到键向量；对多个小尺寸特征对应的小尺寸值特征进行拼接，得到值向量；基于查询向量和键向量确定注意力权重，并基于注意力权重和值向量确定该输入特征对应的修正特征。然后，可以基于该输入特征和修正特征确定第一注意力子模块的输出特征。

[52]在一种可能的实施方式中，通过第二注意力子模块对第二注意力子模块的输入特征进行第二处理，得到第二注意力子模块的输出特征，可以包括但不限于：对第二注意力子模块的输入特征进行层归一化，得到层归一化后的特征；对层归一化后的特征进行第一线性操作，得到第一线性操作后的特征；对第一线性操作后的特征进行激活操作，得到激活后的特征；对激活后的特征进行第二线性操作，得到第二线性操作后的特征；基于该输入特征和第二线性操作后的特征确定第二注意力子模块的输出特征。

[53]在一种可能的实施方式中，通过第二注意力子模块对第二注意力子模块的输入特征进行第二处理，得到第二注意力子模块的输出特征，可以包括但不限于：对第二注意力子模块的输入特征进行下采样操作，得到下采样后的特征；对这个下采样后的特征进行残差卷积操作，得到残差卷积后的特征；对这个残差卷积后的特征进行上采样操作，得到上采样后的特征；可以对这个上采样后的特征进行激活操作，得到激活后的特征；以及，对该输入特征进行残差卷积操作，得到残差卷积后的卷积特征。在此基础上，可以基于该输入特征、激活后的特征和卷积特征确定第二注意力子模块的输出特征。

[54]在一种可能的实施方式中，合成变换网络还可以包括至少一个反卷积层；其中：注意力模块可以位于其中一个反卷积层的后面。

[55]示例性的，合成变换网络还可以包括残差层、第一反卷积层、第一裁剪层、第一残差激活层、第二反卷积层、第二裁剪层、第二残差激活层、第三反卷积层、第三裁剪层、第三残差激活层、第四反卷积层、第四裁剪层。其中，合成变换网络可以包括至少一个注意力模块，其中一个注意力模块位于第一反卷积层的后面。

[56]示例性的，合成变换网络至少包括残差层、第一反卷积层、第一裁剪层、第一残差激活层、第二反卷积层。合成变换网络包括至少一个注意力模块，其中一个注意力模块位于第一反卷积层的后面。

[57]基于与上述解码方法同样的构思，本申请实施例中还提出一种编码方法，该方法可以应用于编码端（也称为视频编码器），该方法可以包括：对当前图像块对应的一个码流进行解码，得到当前图像块对应的系数超参特征；基于系数超参特征确定概率分布参数，基于概率分布参数对当前图像块对应的另一码流进行解码，得到当前图像块对应的重建特征；将重建特征输入合成变换网络，得到当前图像块对应的重建图像块；其中，合成变换网络至少包括注意力模块，注意力模块为级联混合注意力模块；其中，级联混合注意力模块包括第一注意力子模块和第二注意力子模块，且第一注意力子模块和第二注意力子模块为串行的两个子模块。

[58]其中，当前图像块对应的一个码流可以是编码有当前图像块对应的系数超参特征的

码流；当前图像块对应的另一码流可以是编码有当前图像块对应的残差特征的码流。

[59]示例性的，编码方法与解码方法的实现过程类似，重复内容不再赘述。

[60]示例性的，上述执行顺序只是为了方便描述给出的示例，在实际应用中，还可以改变步骤之间的执行顺序，对此执行顺序不做限制。而且，在其它实施例中，并不一定按照本说明书示出和描述的顺序来执行相应方法的步骤，其方法所包括的步骤可以比本说明书所描述的更多或更少。此外，本说明书中所描述的单个步骤，在其它实施例中可能被分解为多个步骤进行描述；本说明书中所描述的多个步骤，在其它实施例中也可能被合并为单个步骤进行描述。

[61]由以上技术方案可见，本申请实施例中，针对基于神经网络的编解码技术，提出一种注意力机制的合成变换网络，该合成变换网络包括注意力模块，且注意力模块为级联混合注意力模块，通过级联混合注意力模块实现合成变换网络时，在保证合成图像质量的同时，有效降低网络复杂度和计算复杂度，提高解码性能，在降低复杂度的同时，有效保证重建图像块的质量，提升编码性能和解码性能。

[62]实施例 2：编码端的处理过程可参见图 4 所示，当然，图 4 只是编码端的处理过程的一个示例，对此编码端的处理过程不作限制，采用神经网络实现编码处理即可。

[63]编码端在得到当前图像块 x （当前图像块 x 可以为原始图像块 x ，即输入的图像块）之后，可以通过分析变换网络（即神经网络）对当前图像块 x 进行分析变换，得到当前图像块 x 对应的图像特征 y 。其中，通过分析变换网络对当前图像块 x 进行特征变换是指：将当前图像块 x 变换到 latent 域的图像特征 y ，从而方便后续所有过程在 latent 域上进行操作。

[64]图像可以划分成 1 个图像块，也可以划分成多个图像块，若图像划分成 1 个图像块，则当前图像块 x 也可以为图像，即对于图像块的编解码过程也可以直接用于图像。

[65]编码端在得到图像特征 y 之后，对图像特征 y 进行系数超参特征变换，得到系数超参特征 z ，比如说，可以将图像特征 y 输入超参编码网络（即神经网络），由超参编码网络对图像特征 y 进行系数超参特征变换，得到系数超参特征 z 。其中，超参编码网络可以是已训练的神经网络，对此超参编码网络的训练过程不作限制，能够对图像特征 y 进行系数超参特征变换即可。其中，latent 域的图像特征 y 经过超参编码网络之后得到超先验 latent 信息 z 。

[66]编码端在得到系数超参特征 z 之后，可以对系数超参特征 z 进行量化，得到系数超参特征 z 对应的超参量化特征，即图 4 中的 Q 操作为量化过程。在得到系数超参特征 z 对应的超参量化特征之后，对超参量化特征进行编码，得到当前图像块对应的 Bitstream#1（可称为第一码流），即图 4 中的 AE 操作表示编码过程，如熵编码过程。或者，编码端也可以直接对系数超参特征 z 进行编码，得到当前图像块对应的 Bitstream#1。其中，Bitstream#1 中携带的超参量化特征或者系数超参特征 z 主要用于获取均值和概率分布模型的参数。

[67]在得到当前图像块对应的 Bitstream#1 之后，编码端可以将当前图像块对应的 Bitstream#1 发送给解码端，关于解码端针对当前图像块对应的 Bitstream#1 的处理过程，参见后续实施例。

[68]编码端在得到当前图像块对应的 Bitstream#1 之后，还可以对 Bitstream#1 进行解码，得到超参量化特征，即图 4 中的 AD 表示解码过程，然后，对超参量化特征进行反量化，得到系数超参特征 $\hat{z}$ ，系数超参特征 $\hat{z}$ 与系数超参特征 z 可以相同或者不同，图 4 中的 IQ 操作为反量化过程。或者，编码端在得到当前图像块对应的 Bitstream#1 之后，

还可以对 Bitstream#1 进行解码,得到系数超参特征 z_{hat} ,而不涉及超参量化特征的反量化过程。

[69]针对 Bitstream#1 的编码过程,可以采用固定概率密度模型的编码方法,针对 Bitstream#1 的解码过程,可以采用固定概率密度模型的解码方法,对此编码和解码过程不做限制。

[70]编码端在得到系数超参特征 z_{hat} 之后,可以基于当前图像块的系数超参特征 z_{hat} 和前面图像块的重建特征 y_{hat} (重建特征 y_{hat} 的确定过程参见后续实施例),进行基于上下文的预测,得到当前图像块对应的预测值 μ (即均值 μ),比如说,将系数超参特征 z_{hat} 和重建特征 y_{hat} 输入均值预测网络,由均值预测网络基于系数超参特征 z_{hat} 和重建特征 y_{hat} 确定预测值 μ ,对此预测过程不作限制。其中,针对基于上下文的预测过程,输入包括系数超参特征 z_{hat} 和已解码的重建特征 y_{hat} ,两者联合输入获取更精准的预测值 μ ,预测值 μ 用于与原始特征作差获得残差和与解码的残差相加得到重建特征 y_{hat} 。

[71]在另一个实施方案中,均值预测网络也可以仅基于当前图像块的系数超参特征 z_{hat} 得到当前图像块对应的预测值,即不利用前面图像块的重建特征 y_{hat} 获得当前图像块对应的预测值,这样可以大大降低获得预测值的复杂度,加速编解码过程。下面编码端的预测值获取方法,都可以采用上述两种方案中的一种方案,不再一一赘述。

[72]需要注意的是,均值预测网络为可选的神经网络,即也可以没有均值预测网络,即不需要通过均值预测网络确定预测值 μ ,图 4 中的虚线框表示均值预测网络为可选。

[73]编码端在得到图像特征 y 之后,可以基于图像特征 y 和预测值 μ 确定残差特征 r ,如将图像特征 y 与预测值 μ 的差值作为残差特征 r 。然后,对残差特征 r 进行特征处理,得到图像特征 s ,对此特征处理过程不做限制,可以是任意特征处理方式。在该情况下,需要部署均值预测网络,由均值预测网络提供预测值 μ 。或者,编码端在得到图像特征 y 之后,可以对图像特征 y 进行特征处理,得到图像特征 s ,对此特征处理过程不做限制,可以是任意特征处理方式。在该情况下,不需要部署均值预测网络,通过虚线框表示残差过程为可选过程。

[74]编码端在得到图像特征 s 之后,可以对图像特征 s 进行量化,得到图像特征 s 对应的图像量化特征,即图 4 中的 Q 操作为量化过程。在得到图像特征 s 对应的图像量化特征之后,编码端可以对该图像量化特征进行编码,得到当前图像块对应的 Bitstream#2 (可称为第二码流),即图 4 中的 AE 操作表示编码过程,如熵编码过程。或者,编码端也可以直接对图像特征 s 进行编码,得到当前图像块对应的 Bitstream#2,而不涉及图像特征 s 的量化过程。

[75]在得到当前图像块对应的 Bitstream#2 之后,编码端可以将当前图像块对应的 Bitstream#2 发送给解码端,关于解码端针对当前图像块对应的 Bitstream#2 的处理过程,参见后续实施例。

[76]编码端在得到当前图像块对应的 Bitstream#2 之后,还可以对 Bitstream#2 进行解码,得到图像量化特征,即图 4 中的 AD 表示解码过程,然后,编码端可以对图像量化特征进行反量化,得到图像特征 s' ,图像特征 s' 与图像特征 s 可以相同或者不同,图 4 中的 IQ 操作为反量化过程。或者,编码端在得到当前图像块对应的 Bitstream#2 之后,还可以对 Bitstream#2 进行解码,得到图像特征 s' ,而不涉及图像量化特征的反量化过程。

[77]编码端在得到图像特征 s' 之后,可以对图像特征 s' 进行特征恢复 (即特征处理的逆过程),对此特征恢复过程不做限制,可以是任意特征恢复方式,得到残差特征 r_{hat} ,残差特征 r_{hat} 与残差特征 r 可以相同或者不同。编码端得到残差特征 r_{hat} 后,基于残

差特征 r_{hat} 和预测值 μ 确定重建特征 y_{hat} , 重建特征 y_{hat} 与图像特征 y 可以相同或不同, 比如说, 将残差特征 r_{hat} 与预测值 μ 的和作为重建特征 y_{hat} 。在该情况下, 需要部署均值预测网络, 由均值预测网络提供预测值 μ 。或者, 编码端在得到图像特征 s' 之后, 可以对图像特征 s' 进行特征恢复(即特征处理的逆过程), 得到重建特征 y_{hat} , 重建特征 y_{hat} 与图像特征 y 可以相同或不同。在该情况下, 不需要部署均值预测网络, 通过虚线框表示残差过程为可选过程。

[78]编码端在得到重建特征 y_{hat} 之后, 可以对重建特征 y_{hat} 进行合成变换, 得到当前图像块 x 对应的重建图像块 x_{hat} , 比如说, 将重建特征 y_{hat} 输入合成变换网络, 由合成变换网络对重建特征 y_{hat} 进行合成变换, 得到重建图像块 x_{hat} , 至此, 完成图像重建过程。

[79]在一种可能的实施方式中, 编码端在对图像量化特征或者图像特征 s 进行编码, 得到当前图像块对应的 Bitstream#2 时, 需要先确定概率分布模型, 然后基于该概率分布模型对图像量化特征或者图像特征 s 进行编码。此外, 编码端在对 Bitstream#2 进行解码时, 也需要先确定概率分布模型, 然后基于该概率分布模型对 Bitstream#2 进行解码。

[80]为了得到概率分布模型, 继续参见图 4 所示, 编码端在得到系数超参特征 z_{hat} 之后, 可以对系数超参特征 z_{hat} 进行系数超参特征反变换, 得到概率分布参数 p , 比如说, 将系数超参特征 z_{hat} 输入概率超参解码网络, 由概率超参解码网络对系数超参特征 z_{hat} 进行系数超参特征反变换, 得到概率分布参数 p , 在得到概率分布参数 p 之后, 可以基于概率分布参数 p 生成概率分布模型。其中, 概率超参解码网络可以是已训练的神经网络, 对此概率超参解码网络的训练过程不作限制, 能够对系数超参特征 z_{hat} 进行系数超参特征反变换即可。

[81]在一种可能的实施方式中, 上述编码端的处理过程, 可以由深度学习模型或者神经网络模型执行, 从而实现端到端的图像压缩和编码过程, 对此编码过程不作限制。

[82]实施例 3: 解码端的处理过程可以参见图 5 所示, 当然, 图 5 只是解码端的处理过程的一个示例, 对此解码端的处理过程不作限制, 采用神经网络实现解码处理即可。

[83]解码端在得到当前图像块对应的 Bitstream#1 之后, 可以对 Bitstream#1 进行解码, 得到超参量化特征, 即图 5 中的 AD 表示解码过程, 然后, 对超参量化特征进行反量化, 得到系数超参特征 z_{hat} , 系数超参特征 z_{hat} 与系数超参特征 z 可以相同或者不同, 图 5 中的 IQ 操作为反量化过程。或者, 解码端在得到当前图像块对应的 Bitstream#1 之后, 可以对 Bitstream#1 进行解码, 得到系数超参特征 z_{hat} , 而不涉及超参量化特征的反量化过程。其中, 针对 Bitstream#1 的解码过程, 可以采用固定概率密度模型的解码方法, 对此不做限制。

[84]图像可以划分成 1 个图像块, 也可以划分成多个图像块, 若图像划分成 1 个图像块, 则当前图像块 x 也可以为图像, 即对于图像块的解码过程也可以直接用于图像。

[85]解码端在得到系数超参特征 z_{hat} 之后, 可以基于当前图像块的系数超参特征 z_{hat} 和前面图像块的重建特征 y_{hat} (重建特征 y_{hat} 的确定过程参见后续实施例), 进行基于上下文的预测, 得到当前图像块对应的预测值 μ (即均值 μ), 比如说, 将系数超参特征 z_{hat} 和重建特征 y_{hat} 输入均值预测网络, 由均值预测网络基于系数超参特征 z_{hat} 和重建特征 y_{hat} 确定预测值 μ , 对此预测过程不作限制。其中, 针对基于上下文的预测过程, 输入包括系数超参特征 z_{hat} 和已解码的重建特征 y_{hat} , 两者联合输入获取更精准的预测值 μ 。

[86]在另一个实施方案中, 解码端在得到当前图像块的系数超参特征 z_{hat} 之后, 可以基于系数超参特征 z_{hat} 直接获取当前图像块对应的预测值 μ , 即不依赖前面图像块的

重建特征 y_{hat} 获取预测值, 这样可以大大降低获取预测值的复杂度, 加速编解码过程。下面解码端的预测值获取方法, 都可以采用上述两种方案中的一种方案, 不再一一赘述。

[87]需要注意的是, 均值预测网络为可选的神经网络, 即也可以没有均值预测网络, 即不需要通过均值预测网络确定预测值 μ , 图 5 中的虚线框表示均值预测网络为可选。

[88]解码端在得到当前图像块对应的 Bitstream#2 之后, 可以对 Bitstream#2 进行解码, 得到图像量化特征, 即图 5 中的 AD 表示解码过程, 然后, 解码端可以对图像量化特征进行反量化, 得到图像特征 s' , 图像特征 s' 与图像特征 s 可以相同或者不同, 图 5 中的 IQ 操作可以为反量化过程。或者, 解码端在得到当前图像块对应的 Bitstream#2 之后, 可以对 Bitstream#2 进行解码, 得到图像特征 s' , 而不涉及图像量化特征的反量化过程。

[89]解码端在得到图像特征 s' 之后, 可以对图像特征 s' 进行特征恢复 (即特征处理的逆过程), 得到残差特征 r_{hat} , 残差特征 r_{hat} 与残差特征 r 相同或者不同。解码端在得到残差特征 r_{hat} 之后, 基于残差特征 r_{hat} 和预测值 μ 确定重建特征 y_{hat} , 重建特征 y_{hat} 与图像特征 y 相同或不同, 如将残差特征 r_{hat} 与预测值 μ 的和作为重建特征 y_{hat} 。在该情况下, 需要部署均值预测网络, 由均值预测网络提供预测值 μ 。或者, 解码端在得到图像特征 s' 之后, 可以对图像特征 s' 进行特征恢复, 得到重建特征 y_{hat} , 重建特征 y_{hat} 与图像特征 y 可以相同或不同。在该情况下, 不需要部署均值预测网络, 通过虚线框表示残差过程为可选过程。

[90]解码端在得到重建特征 y_{hat} 之后, 可以对重建特征 y_{hat} 进行合成变换, 得到当前图像块 x 对应的重建图像块 x_{hat} , 比如说, 将重建特征 y_{hat} 输入合成变换网络, 由合成变换网络对重建特征 y_{hat} 进行合成变换, 得到重建图像块 x_{hat} , 至此, 完成图像重建过程。

[91]在一种可能的实施方式中, 解码端在对 Bitstream#2 进行解码时, 需要先确定概率分布模型, 然后基于该概率分布模型对 Bitstream#2 进行解码。为了得到概率分布模型, 继续参见图 5 所示, 解码端在得到系数超参特征 z_{hat} 之后, 可以对系数超参特征 z_{hat} 进行系数超参特征反变换, 得到概率分布参数 p , 比如说, 将系数超参特征 z_{hat} 输入概率超参解码网络, 由概率超参解码网络对系数超参特征 z_{hat} 进行系数超参特征反变换, 得到概率分布参数 p , 在得到概率分布参数 p 之后, 可以基于概率分布参数 p 生成概率分布模型。其中, 概率超参解码网络可以是已训练的神经网络, 对此概率超参解码网络的训练过程不作限制, 能够对系数超参特征 z_{hat} 进行系数超参特征反变换, 得到概率分布参数 p 即可。

[92]在一种可能的实施方式中, 上述解码端的处理过程, 可以由深度学习模型或者神经网络模型执行, 从而实现端到端的图像压缩和编码过程, 对此解码过程不作限制。

[93]实施例 4: 针对实施例 1、实施例 2 和实施例 3, 可以部署均值预测网络, 也可以不部署均值预测网络, 为了提升特征编码性能, 以部署均值预测网络为例。在部署均值预测网络时, 为了获得图像特征的准确预测值, 可基于当前图像块的系数超参特征 z_{hat} 和前面图像块的重建特征 y_{hat} , 进行基于上下文的预测, 得到当前图像块对应的预测值 μ ; 或者, 基于当前图像块的系数超参特征 z_{hat} 得到当前图像块对应的预测值 μ 。

[94]关于编码端的处理过程, 可以包括以下步骤:

[95]步骤 S11、在得到原始图像之后, 基于图像分辨率确定是否对原始图像进行分块, 比如, 若图像分辨率大于阈值, 则对原始图像进行分块, 否则不对原始图像进行分块。若是, 则将原始图像分成若干个图像块 x , 相邻图像块之间存在交叠部分, 否则图像块 x 为原始图像, 即原始图像只有 1 个图像块。为了方便描述, 后续以一个图像块为例, 将这个图像块记为当前图像块 x , 后续以当前图像块 x 的处理过程为例进行说明。

[96]步骤 S12、在得到当前图像块 x 之后，通过分析变换网络对当前图像块 x 进行分析变换，得到当前图像块 x 对应的图像特征 y （即特征块）。比如，通过分析变换网络将当前图像块 x 变换到 latent 域的图像特征 y ，以方便后续所有过程在 latent 域上进行操作。

[97]步骤 S13、对图像特征 y 进行系数超参特征变换，得到当前图像块 x 对应的系数超参特征 z 。比如说，可以将图像特征 y 输入超参编码网络（即神经网络），由超参编码网络对图像特征 y 进行系数超参特征变换，得到当前图像块 x 对应的系数超参特征 z 。

[98]步骤 S14、将系数超参特征 z 编码到当前图像块对应的第一码流（Bitstream#1）。比如说，对系数超参特征 z 进行量化，得到超参量化特征，并对该超参量化特征进行编码，得到当前图像块对应的第一码流。或者，直接对系数超参特征 z 进行编码，得到当前图像块对应的第一码流。即，当前图像块对应的第一码流是编码有当前图像块 x 对应的系数超参特征 z 的码流。在得到当前图像块对应的第一码流后，将当前图像块对应的第一码流发送给解码端。

[99]步骤 S15、在得到当前图像块对应的第一码流之后，对第一码流进行解码，得到当前图像块 x 对应的系数超参特征 z_{hat} 。比如说，可对第一码流进行解码，得到系数超参特征 z_{hat} 对应的超参量化特征，对超参量化特征进行反量化，得到系数超参特征 z_{hat} 。或者，对第一码流进行解码，得到系数超参特征 z_{hat} ，而不涉及反量化过程。

[100]步骤 S16、基于系数超参特征 z_{hat} 确定概率分布参数。比如说，对系数超参特征 z_{hat} 进行系数超参特征反变换，得到概率分布参数 p ，比如说，将系数超参特征 z_{hat} 输入概率超参解码网络，由概率超参解码网络对系数超参特征 z_{hat} 进行系数超参特征反变换，得到概率分布参数 p ，在得到概率分布参数 p 之后，基于概率分布参数 p 生成概率分布模型。

[101]步骤 S17、将系数超参特征 z_{hat} 输入均值预测网络，由均值预测网络基于系数超参特征 z_{hat} 和前面图像块的重建特征 y_{hat} ，进行基于上下文的预测，得到预测值 μ 。或者，将系数超参特征 z_{hat} 输入均值预测网络，由均值预测网络基于系数超参特征 z_{hat} 得到预测值 μ 。

[102]步骤 S18、基于图像特征 y 和均值特征 μ （即预测值 μ ）确定残差特征 r ，如将图像特征 y 与均值特征 μ 的差值作为残差特征 r 。对残差特征 r 进行特征处理，得到特征处理后的残差特征 r ，对此特征处理过程不做限制。特征处理是可选步骤，也可以不对残差特征 r 进行特征处理。

[103]步骤 S19、对残差特征 r （或特征处理后的残差特征 r ）进行量化，得到图像量化特征，并对该图像量化特征进行编码，得到当前图像块对应的第二码流（Bitstream#2）。或者，也可以直接对残差特征 r （或特征处理后的残差特征 r ）进行编码，得到当前图像块对应的第二码流。即，当前图像块对应的第二码流是编码有当前图像块 x 对应的残差特征 r 的码流。在得到当前图像块对应的第二码流之后，将当前图像块对应的第二码流发送给解码端。

[104]示例性的，在对图像量化特征或者残差特征 r 进行编码时，可以采用概率分布参数 p 对应的概率分布模型对图像量化特征或者残差特征 r 进行编码，得到第二码流。

[105]步骤 S20、在得到当前图像块对应的第二码流之后，采用概率分布参数 p 对应的概率分布模型对第二码流进行解码得到图像量化特征，对图像量化特征进行反量化，对反量化后的特征进行特征恢复，得到残差特征 r_{hat} ，或，对图像量化特征进行反量化后直接得到残差特征 r_{hat} ，而不涉及特征恢复过程。或者，对第二码流进行解码之后，对解码后的特征进行特征恢复，得到残差特征 r_{hat} ，或者，对第二码流进行解码后直接得到残差特征 r_{hat} 。

[106]步骤 S21、在得到残差特征 r_{hat} 之后，基于残差特征 r_{hat} 和均值特征 μ 确定重建特征 y_{hat} ，比如说，将残差特征 r_{hat} 与均值特征 μ 之和作为重建特征 y_{hat} 。

[107]步骤 S22、将重建特征 y_{hat} 输入合成变换网络，由合成变换网络基于重建特征 y_{hat} 确定当前图像块对应的重建图像块 x_{hat} ，并输出当前图像块对应的重建图像块 x_{hat} 。

[108]示例性的，以部署均值预测网络为例，关于解码端的处理过程，可以包括以下步骤：

[109]步骤 S31、在得到当前图像块对应的第一码流之后，对第一码流进行解码，得到当前图像块 x 对应的系数超参特征 z_{hat} 。比如说，可对第一码流进行解码，得到系数超参特征 z_{hat} 对应的超参量化特征，对超参量化特征进行反量化，得到系数超参特征 z_{hat} 。或者，对第一码流进行解码，得到系数超参特征 z_{hat} ，而不涉及反量化过程。

[110]步骤 S32、基于系数超参特征 z_{hat} 确定概率分布参数。比如说，对系数超参特征 z_{hat} 进行系数超参特征反变换，得到概率分布参数 p ，比如说，将系数超参特征 z_{hat} 输入概率超参解码网络，由概率超参解码网络对系数超参特征 z_{hat} 进行系数超参特征反变换，得到概率分布参数 p ，在得到概率分布参数 p 之后，基于概率分布参数 p 生成概率分布模型。

[111]步骤 S33、在得到当前图像块对应的第二码流之后，基于概率分布参数 p 对当前图像块对应的第二码流进行解码，得到当前图像块对应的重建特征。比如说，可以确定当前图像块对应的残差特征 r_{hat} ，并确定当前图像块对应的预测值 μ ，基于残差特征 r_{hat} 和均值特征 μ 确定重建特征 y_{hat} ，比如说，将残差特征 r_{hat} 与均值特征 μ （即预测值 μ ）之和作为重建特征 y_{hat} 。

[112]其中，为了确定当前图像块对应的预测值 μ ，在得到系数超参特征 z_{hat} 之后，可以将系数超参特征 z_{hat} 输入均值预测网络，由均值预测网络基于系数超参特征 z_{hat} 和前面图像块的重建特征 y_{hat} ，进行基于上下文的预测，得到当前图像块对应的预测值 μ 。或者，可以将系数超参特征 z_{hat} 输入均值预测网络，由均值预测网络基于系数超参特征 z_{hat} 得到当前图像块对应的预测值 μ 。

[113]为了确定当前图像块对应的残差特征 r_{hat} ，可以采用概率分布参数 p 对应的概率分布模型对第二码流进行解码得到图像量化特征，对图像量化特征进行反量化，对反量化后的特征进行特征恢复，得到残差特征 r_{hat} ，或，对图像量化特征进行反量化后直接得到残差特征 r_{hat} ，而不涉及特征恢复过程。或者，对第二码流进行解码之后，对解码后的特征进行特征恢复，得到残差特征 r_{hat} ，或者，对第二码流进行解码后直接得到残差特征 r_{hat} 。

[114]步骤 S34、将重建特征 y_{hat} 输入合成变换网络，由合成变换网络基于重建特征 y_{hat} 确定当前图像块对应的重建图像块 x_{hat} ，并输出当前图像块对应的重建图像块 x_{hat} 。

[115]实施例 5：在实施例 1、实施例 2、实施例 3 和实施例 4 中，均涉及合成变换网络。该实施例的实施方式与实施例 1、实施例 2、实施例 3 或实施例 4 的实施方式相同，重复内容不再赘述。下面对其中所涉及的合成变换网络进行详细描述。

[116]该合成变换网络的一个示例可以参见图 6A 所示，即合成变换网络可以依次包括残差层、反卷积层、裁剪层、残差激活层、反卷积层、裁剪层、残差激活层、反卷积层、注意力模型、裁剪层、残差激活层、反卷积层、裁剪层等。其中，该合成变换网络的输入特征可以为重建特征 y_{hat} ，该合成变换网络的输出特征可以为重建图像块 x_{hat} ，也就是说，将重建特征 y_{hat} 输入合成变换网络之后，依次经过上述网络层的处理，得到重建图像块 x_{hat} 。

[117]示例性的，合成变换网络中的残差层的一个示例可以参见图 6B 所示，即残差层可

以依次包括卷积层、整流线性单元（Rectified Linear Unit, ReLU）（如泄露整流线性单元 LeakyReLU）、卷积层和叠加层（用于实现特征的相加）。第 1 个卷积层可以为 1x1 的卷积层、或 3x3 的卷积层、或 5x5 的卷积层，对此不作限制，如可以选取 3x3 的卷积层。第 2 个卷积层可以为 1x1 的卷积层、或 3x3 的卷积层、或 5x5 的卷积层，对此不作限制，如可以选取 3x3 的卷积层。

[118]示例性的，合成变换网络中的残差激活层可参见图 6C 所示，即残差激活层可以依次包括激活层（如 LeakyRelu 激活层）、卷积层、激活层（如 tanh 激活层）、相乘层（用于实现特征的相乘）和叠加层（用于实现特征的相加）。卷积层可以为 1x1 的卷积层、或 3x3 的卷积层、或 5x5 的卷积层，对此不作限制，如可以选取 1x1 的卷积层。

[119]示例性的，合成变换网络中的注意力模型可为残差非局部注意力模块（Residual Non-local Attention Block, RNAB），即 RNAB 被引入作为注意力模型，RNAB 用于获取图像的非局部信息，并将非局部信息作为注意力权重以提升解码性能。RNAB 的一个示例可以参见图 6D 所示，即 RNAB 可以依次包括残差块、残差块、下采样卷积层（如 2 倍下采样卷积层，如 3x3 的卷积层）、残差块、残差块、上采样卷积层（如 2 倍上采样卷积层，如 3x3 的卷积层）、残差块、残差块、卷积层（如 3x3 的卷积层）、S 形激活函数（Sigmoid）、残差块、残差块、残差块、相乘层（用于实现特征的相乘）和叠加层（用于实现特征的相加）。

[120]针对残差块，该残差块可以参见图 6B 所示，即残差块与图 6B 中的残差层类似，也可以依次包括卷积层、整流线性单元（Rectified Linear Unit, ReLU）（如泄露整流线性单元 LeakyReLU）、卷积层和叠加层（用于实现特征的相加）。这样，残差块的输入特征经过卷积层、整流线性单元和卷积层提取特征后，与该残差块的输入特征相加得到该残差块的输出特征。

[121]由残差块、残差块、下采样卷积层、残差块、残差块、上采样卷积层、残差块、残差块、卷积层和 S 形激活函数组成的子模块可以为非局部注意力提取子模块，由 3 个残差块组成的子模块可以为特征提取子模块，RNAB 的输入特征经过非局部注意力提取子模块产生注意力权重后与特征提取子模块的输出相乘，并与 RNAB 的输入特征相加后得到 RNAB 的输出特征。

[122]实施例 6：在实施例 1、实施例 2、实施例 3 和实施例 4 中，均涉及合成变换网络。该实施例的实施方式与实施例 1、实施例 2、实施例 3 或实施例 4 的实施方式相同，重复内容不再赘述。下面对其中所涉及的合成变换网络进行详细描述。

[123]该合成变换网络至少可以包括注意力模块，且该注意力模块可以为级联混合注意力模块。除了注意力模块，合成变换网络还可以包括其它网络层，本实施例中对此合成变换网络的结构不做限制，只要合成变换网络包括注意力模块即可。

[124]在一种可能的实施方式中，除了注意力模块，该合成变换网络还可以包括至少一个反卷积层，此外，除了包括注意力模块和至少一个反卷积层，该合成变换网络还可以包括其它网络层，对此不做限制。在合成变换网络包括至少一个反卷积层时，注意力模块可以位于任意一个反卷积层的后面。或者，注意力模块可以位于每个反卷积层的后面；其中，位于不同反卷积层后面的注意力模块可以完全相同，或者，位于不同反卷积层后面的注意力模块可以不完全相同。或者，注意力模块可以位于部分反卷积层（如所有反卷积层中的 K 个反卷积层，K 大于 1，且 K 小于反卷积层总数量）的后面；其中，位于不同反卷积层后面的注意力模块可以完全相同，或者，位于不同反卷积层后面的注意力模块可以不完全相同。

[125]在一种可能的实施方式中，除了注意力模块，合成变换网络可以包括以下至少一种

网络层：残差层、第一反卷积层、第一裁剪层、第一残差激活层、第二反卷积层、第二裁剪层、第二残差激活层、第三反卷积层、第三裁剪层、第三残差激活层、第四反卷积层、第四裁剪层，即合成变换网络可以包括上述的部分网络层或者全部网络层。当然，除了上述网络层，合成变换网络还可以包括其它网络层，本申请对此不做限制，只是以上述网络层为例进行说明。

[126]参见图 7A 所示，为合成变换网络的一个示例，即合成变换网络可以依次包括残差层、反卷积层 1（即第一反卷积层）、裁剪层 1（即第一裁剪层）、残差激活层 1（即第一残差激活层）、反卷积层 2（即第二反卷积层）、裁剪层 2（即第二裁剪层）、残差激活层 2（即第二残差激活层）、反卷积层 3（即第三反卷积层）、裁剪层 3（即第三裁剪层）、残差激活层 3（即第三残差激活层）、反卷积层 4（即第四反卷积层）、裁剪层 4（即第四裁剪层）。示例性的，该合成变换网络的输入特征可以为重建特征 y_{hat} ，该合成变换网络的输出特征可以为重建图像块 x_{hat} ，也就是说，在将重建特征 y_{hat} 输入合成变换网络之后，依次经过上述网络层的处理，合成变换网络可以得到重建图像块 x_{hat} ，并输出重建图像块 x_{hat} 。

[127]在图 7A 中，示出了三个注意力模型位置，即注意力模型位置 1、注意力模型位置 2 和注意力模型位置 3。示例性的，若合成变换网络只包括一个注意力模块（即级联混合注意力模块），则该注意力模块可以位于反卷积层 1 的后面，即该注意力模块位于注意力模型位置 1。或者，该注意力模块可以位于反卷积层 2 的后面，即该注意力模块位于注意力模型位置 2。或者，该注意力模块可以位于反卷积层 3 的后面，即该注意力模块位于注意力模型位置 3，比如说，在反卷积层 3 的后面放置该注意力模块。当然，上述只是注意力模块的位置的几个示例，对此不做限制，可以位于该合成变换网络的任意位置。

[128]示例性的，合成变换网络可包括位于反卷积层 1 的后面的注意力模块 1、位于反卷积层 2 的后面的注意力模块 2、位于反卷积层 3 的后面的注意力模块 3，即注意力模块 1 位于注意力模型位置 1，注意力模块 2 位于注意力模型位置 2，注意力模块 3 位于注意力模型位置 3。当然，注意力模块也可位于合成变换网络的其它位置，对此不做限制。

[129]其中，注意力模块 1、注意力模块 2 和注意力模块 3 可以完全相同，即注意力模块 1、注意力模块 2 和注意力模块 3 可以采用相同的网络结构。或者，注意力模块 1、注意力模块 2 和注意力模块 3 可以不完全相同，比如说，注意力模块 1 的网络结构与注意力模块 2 的网络结构可以相同，注意力模块 1 的网络结构与注意力模块 3 的网络结构可以不同；或，注意力模块 1 的网络结构与注意力模块 3 的网络结构可以相同，注意力模块 1 的网络结构与注意力模块 2 的网络结构可以不同；或，注意力模块 2 的网络结构与注意力模块 3 的网络结构可以相同，注意力模块 2 的网络结构与注意力模块 1 的网络结构可以不同；或，注意力模块 1 的网络结构与注意力模块 2 的网络结构可以不同，注意力模块 1 的网络结构与注意力模块 3 的网络结构可以不同，注意力模块 2 的网络结构与注意力模块 3 的网络结构可以不同。

[130]示例性的，合成变换网络可以包括位于反卷积层 1 的后面的注意力模块 1、位于反卷积层 2 的后面的注意力模块 2，即注意力模块 1 位于注意力模型位置 1，注意力模块 2 位于注意力模型位置 2。当然，注意力模块也可以位于合成变换网络的其它位置，对此不做限制。

[131]其中，注意力模块 1 和注意力模块 2 可以完全相同，比如说，注意力模块 1 和注意力模块 2 可以采用相同的网络结构。或者，注意力模块 1 和注意力模块 2 可以不完全相同，比如说，注意力模块 1 的网络结构与注意力模块 2 的网络结构可以不同。

[132]示例性的，合成变换网络可以包括位于反卷积层 1 的后面的注意力模块 1、位于反

卷积层 3 的后面的注意力模块 2, 即注意力模块 1 位于注意力模型位置 1, 注意力模块 2 位于注意力模型位置 3。当然, 注意力模块也可以位于合成变换网络的其它位置, 对此不做限制。

[133]其中, 注意力模块 1 和注意力模块 2 可以完全相同, 比如说, 注意力模块 1 和注意力模块 2 可以采用相同的网络结构。或者, 注意力模块 1 和注意力模块 2 可以不完全相同, 比如说, 注意力模块 1 的网络结构与注意力模块 2 的网络结构可以不同。

[134]示例性的, 合成变换网络可以包括位于反卷积层 2 的后面的注意力模块 1、位于反卷积层 3 的后面的注意力模块 2, 即注意力模块 1 位于注意力模型位置 2, 注意力模块 2 位于注意力模型位置 3。当然, 注意力模块也可以位于合成变换网络的其它位置, 对此不做限制。

[135]其中, 注意力模块 1 和注意力模块 2 可以完全相同, 比如说, 注意力模块 1 和注意力模块 2 可以采用相同的网络结构。或者, 注意力模块 1 和注意力模块 2 可以不完全相同, 比如说, 注意力模块 1 的网络结构与注意力模块 2 的网络结构可以不同。

[136]合成变换网络中的残差层的示例可以参见图 6B 所示, 即残差层可以依次包括卷积层、整流线性单元 (如泄露整流线性单元 LeakyReLU)、卷积层和叠加层, 即输入特征经过卷积层、整流线性单元和卷积层提取特征后, 与输入特征相加得到残差层的输出特征。其中, 第 1 个卷积层可以为 1x1 的卷积层、或 3x3 的卷积层、或 5x5 的卷积层, 对此不作限制, 如可以选取 3x3 的卷积层。此外, 第 2 个卷积层可以为 1x1 的卷积层、或 3x3 的卷积层、或 5x5 的卷积层, 对此不作限制, 如可以选取 3x3 的卷积层。

[137]合成变换网络中的残差激活层的示例可以参见图 6C 所示, 即残差激活层可以依次包括激活层 (如 LeakyRelu 激活层)、卷积层、激活层 (如 tanh 激活层)、相乘层和叠加层, 对此残差激活层的结构不做限制。其中, 卷积层可以为 1x1 的卷积层、或 3x3 的卷积层、或 5x5 的卷积层, 对此不作限制, 如可以选取 1x1 的卷积层。

[138]实施例 7: 此实施例是在实施例 6 的基础上进一步对其中涉及的注意力模块进行描述。在实施例 6 中, 涉及合成变换网络中的注意力模块, 该注意力模块可以为级联混合注意力模块, 通过级联混合注意力模块替换 RNAB, 以在尽可能保证解码性能的同时大量降低解码的计算复杂度。示例性的, 级联混合注意力模块能够提取输入特征的注意力权重, 并基于注意力权重和输入特征确定级联混合注意力模块的输出特征。

[139]示例性的, 级联混合注意力模块可包括第一注意力子模块 (也可以称为变换器注意力子模块, 变换器注意力子模块记为 Transformer-based Attention Module) 和第二注意力子模块 (也可以称为升级卷积子模块, 记为 ConvNext Block)。其中, 第一注意力子模块和第二注意力子模块可以为串行的两个子模块, 即第一注意力子模块的输出特征为第二注意力子模块的输入特征。比如说, 参见图 7B 所示, 为级联混合注意力模块的示意图, 级联混合注意力模块的输入特征 (或输入图像) 作为第一注意力子模块 (图 7B 中以变换器注意力子模块为例) 的输入特征, 第一注意力子模块的输入特征经过第一注意力子模块, 由第一注意力子模块基于该输入特征进行处理, 得到第一注意力子模块的输出特征。第一注意力子模块的输出特征作为第二注意力子模块 (图 7B 中以升级卷积子模块为例) 的输入特征, 第二注意力子模块的输入特征经过第二注意力子模块, 由第二注意力子模块基于该输入特征进行处理, 得到第二注意力子模块的输出特征, 第二注意力子模块的输出特征作为级联混合注意力模块的输出特征。

[140]示例性的, 在将第一注意力子模块的输入特征输入第一注意力子模块之后, 通过第一注意力子模块对该输入特征进行第一处理, 得到第一注意力子模块的输出特征。其中, 第一处理可以包括但不限于: 层归一化、卷积操作和维度变换操作的至少一种, 对此第

一处理不做限制，第一处理与第一注意力子模块的网络结构有关。可以任意配置第一注意力子模块的网络结构。

[141]示例性的，在将第二注意力子模块的输入特征输入第二注意力子模块之后，通过第二注意力子模块对该输入特征进行第二处理，得到第二注意力子模块的输出特征。其中，第二处理可以包括但不限于：深度可分离卷积操作、层归一化、多层感知操作、线性操作、激活操作、下采样操作、残差卷积操作、上采样操作的至少一种，对此第二处理不做限制，第二处理与第二注意力子模块的网络结构有关。可以任意配置第二注意力子模块的网络结构。

[142]比如说，第二处理可以包括深度可分离卷积操作、层归一化、多层感知操作的至少一种。或者，第二处理可以包括层归一化、线性操作、激活操作的至少一种。或者，第二处理可以包括下采样操作、残差卷积操作、上采样操作、激活操作的至少一种。

[143]实施例 8：此实施例是在实施例 6/实施例 7 的基础上进一步对级联混合注意力模块中涉及的第一注意力子模块进行描述。在实施例 6 和实施例 7 中，级联混合注意力模块包括第一注意力子模块，第一注意力子模块用于对第一注意力子模块的输入特征进行第一处理，得到第一注意力子模块的输出特征。比如说，对第一注意力子模块的输入特征进行层归一化，得到层归一化后的特征；对层归一化后的特征进行三路卷积操作，得到查询向量、键向量和值向量；对该查询向量、该键向量和该值向量分别进行维度变换操作，得到维度变换后的查询向量、维度变换后的键向量和维度变换后的值向量；基于维度变换后的查询向量和维度变换后的键向量确定注意力权重，并基于该注意力权重和维度变换后的值向量确定该输入特征对应的修正特征；基于该输入特征和修正特征确定第一注意力子模块的输出特征。

[144]参见图 8A 所示，为第一注意力子模块的结构示意图，这里只是第一注意力子模块的示例，对此第一注意力子模块的结构不做限制。

[145]现以图 8A 的第一注意力子模块为例进行说明。第一注意力子模块可以包括层归一化层，可以通过层归一化层对第一注意力子模块的输入特征（或输入图像）进行层归一化，得到层归一化后的特征。其中，层归一化（layer norm）是神经网络正则化技术，用于规范化每层神经网络的输入特征，以提高训练和泛化性能。其中，层归一化的操作可以包括计算均值和方差、标准化、缩放和平移，使得每个隐藏单元的输出在一个相对较小的范围内，并且每个隐藏单元在整个数据集中都有相似的分布。

[146]在得到层归一化后的特征之后，可以对层归一化后的特征进行三路卷积操作，得到查询向量 Q、键向量 K 和值向量 V。比如说，可以通过两个卷积层对层归一化后的特征进行卷积操作，得到查询向量 Q，第 1 个卷积层可以为 1x1 的卷积层、或 3x3 的卷积层、或 5x5 的卷积层，对此不作限制，如可以选取 1x1 的卷积层。第 2 个卷积层可以为 3x3 的卷积层、或 5x5 的卷积层、或 7x7 的卷积层，对此不作限制，如可以选取 3x3 的卷积层。可以通过两个卷积层对层归一化后的特征进行卷积操作，得到键向量 K，第 1 个卷积层可以为 1x1 的卷积层、或 3x3 的卷积层、或 5x5 的卷积层，如可以选取 1x1 的卷积层。第 2 个卷积层可以为 3x3 的卷积层、或 5x5 的卷积层、或 7x7 的卷积层，如可以选取 3x3 的卷积层。可以通过两个卷积层对层归一化后的特征进行卷积操作，得到值向量 V，第 1 个卷积层可以为 1x1 的卷积层、或 3x3 的卷积层、或 5x5 的卷积层，如可以选取 1x1 的卷积层。第 2 个卷积层可以为 3x3 的卷积层、或 5x5 的卷积层、或 7x7 的卷积层，如可以选取 3x3 的卷积层。

[147]对查询向量 Q、键向量 K 和值向量 V 分别进行维度变换操作（即 R 操作），得到维度变换后的查询向量 Q、维度变换后的键向量 K 和维度变换后的值向量 V。其中，维度变换操作的目的是：将变换后的张量变换成多头自注意力机制所需要满足的维度形式。

多头自注意力机制是一种在自然语言处理 (Natural Language Processing, NLP) 任务中广泛使用的技术。该技术通过自注意力机制来构建上下文词之间的联系, 从而根据上下文来重新构建目标词的表示。多头注意力机制是多组自注意力构件的组合, 能够学习到多种不同类型的上下文影响情况。通过将多个自注意力的输出再经过参数矩阵得到一个新输出, 多头注意力机制能够捕获更大更多范围的相关性特征, 增加模型的表达能力。

[148]基于维度变换后的查询向量 Q 和维度变换后的键向量 K 确定注意力权重。比如说, 对维度变换后的查询向量 Q 与维度变换后的键向量 K 进行矩阵乘法, 将矩阵乘法的结果经过软最大值函数 (Softmax), 得到注意力权重, 对此 Softmax 的处理过程不做限制。

[149]基于注意力权重和维度变换后的值向量 V 确定输入特征对应的修正特征。比如说, 对维度变换后的值向量 V 与注意力权重进行矩阵乘法, 对矩阵乘法的结果进行维度变换操作 (即 R 操作), 并对维度变换操作后的特征进行卷积操作, 得到输入特征对应的修正特征, 卷积层可以为 1×1 的卷积层、或 3×3 的卷积层、或 5×5 的卷积层, 如可以选取 1×1 的卷积层。

[150]基于第一注意力子模块的输入特征和修正特征确定第一注意力子模块的输出特征, 比如说, 对该输入特征和该修正特征进行矩阵加法, 得到第一注意力子模块的输出特征。

[151]至此, 完成第一注意力子模块的处理过程, 得到第一注意力子模块的输出特征。

[152]实施例 9: 此实施例是在实施例 6/实施例 7 的基础上进一步对级联混合注意力模块中涉及的第一注意力子模块进行描述。在实施例 6 和实施例 7 中, 级联混合注意力模块均包括第一注意力子模块, 第一注意力子模块用于对第一注意力子模块的输入特征进行第一处理, 得到第一注意力子模块的输出特征。比如说, 对第一注意力子模块的输入特征进行层归一化, 得到层归一化后的特征; 对层归一化后的特征进行窗口划分, 得到多个小尺寸特征; 针对每个小尺寸特征, 对该小尺寸特征进行三路卷积操作, 得到该小尺寸特征对应的小尺寸查询特征、小尺寸键特征和小尺寸值特征; 对多个小尺寸特征对应的小尺寸查询特征进行拼接得到查询向量; 对多个小尺寸特征对应的小尺寸键特征进行拼接得到键向量; 对多个小尺寸特征对应的小尺寸值特征进行拼接得到值向量。对查询向量、键向量和值向量分别进行维度变换操作, 得到维度变换后的查询向量、维度变换后的键向量和维度变换后的值向量; 基于维度变换后的查询向量和维度变换后的键向量确定注意力权重, 基于注意力权重和维度变换后的值向量确定修正特征; 基于该输入特征和修正特征确定第一注意力子模块的输出特征。

[153]比如说, 在图 8A 的第一注意力子模块的基础上, 可以在层归一化层的后面增加窗口划分层。对此第一注意力子模块的结构不做限制, 例如, 第一注意力子模块的结构示例可以参见图 8A 所示。

[154]示例性的, 第一注意力子模块可以包括层归一化层, 可以通过层归一化层对第一注意力子模块的输入特征 (或输入图像) 进行层归一化, 得到层归一化后的特征。

[155]可以通过窗口划分层对层归一化后的特征进行窗口划分, 得到多个小尺寸特征, 比如说, 将层归一化后的特征划分成多个小尺寸特征, 每个小尺寸特征的宽度为 w_i , 每个小尺寸特征的高度为 h_i , 不同小尺寸特征的宽度可以相同或不同, 不同小尺寸特征的高度可以相同或不同, 通过划分多个小尺寸特征进行后续运算, 能够减低复杂度。

[156]针对每个小尺寸特征, 对该小尺寸特征进行三路卷积操作, 得到该小尺寸特征对应的小尺寸查询特征、小尺寸键特征和小尺寸值特征。比如说, 通过两个卷积层对该小尺寸特征进行卷积操作, 得到小尺寸查询特征, 第 1 个卷积层可以为 1×1 的卷积层、或 3×3 的卷积层、或 5×5 的卷积层, 如选取 1×1 的卷积层。第 2 个卷积层可以为 3×3 的卷积层、或 5×5 的卷积层、或 7×7 的卷积层, 如选取 3×3 的卷积层。通过两个卷积层对该小尺寸

特征进行卷积操作，得到小尺寸键特征，第1个卷积层可以为1x1的卷积层、或3x3的卷积层、或5x5的卷积层，如选取1x1的卷积层。第2个卷积层可以为3x3的卷积层、或5x5的卷积层、或7x7的卷积层，如选取3x3的卷积层。通过两个卷积层对该小尺寸特征进行卷积操作，得到小尺寸值特征，第1个卷积层可以为1x1的卷积层、或3x3的卷积层、或5x5的卷积层，如选取1x1的卷积层。第2个卷积层可以为3x3的卷积层、或5x5的卷积层、或7x7的卷积层，如选取3x3的卷积层。

[157]可以对所有小尺寸特征对应的小尺寸查询特征进行拼接得到查询向量Q，对此特征拼接过程不做限制，以及，可以对所有小尺寸特征对应的小尺寸键特征进行拼接得到键向量K，以及，可以对所有小尺寸特征对应的小尺寸值特征进行拼接得到值向量V。

[158]可以对查询向量Q、键向量K和值向量V分别进行维度变换操作（即R操作），得到维度变换后的查询向量Q、维度变换后的键向量K和维度变换后的值向量V。

[159]可基于维度变换后的查询向量Q和维度变换后的键向量K确定注意力权重。比如，对维度变换后的查询向量Q与维度变换后的键向量K进行矩阵乘法，将矩阵乘法的结果经过软最大值函数(Softmax)，得到注意力权重，对此Softmax的处理过程不做限制。

[160]可以基于注意力权重和维度变换后的值向量V确定输入特征对应的修正特征。比如说，对维度变换后的值向量V与注意力权重进行矩阵乘法，对矩阵乘法的结果进行维度变换操作（即R操作），并对维度变换操作后的特征进行卷积操作，得到输入特征对应的修正特征，卷积层可以为1x1的卷积层、或3x3的卷积层、或5x5的卷积层，如可以选取1x1的卷积层。

[161]可以基于第一注意力子模块的输入特征和修正特征确定第一注意力子模块的输出特征，比如说，对该输入特征和该修正特征进行矩阵加法，得到第一注意力子模块的输出特征。

[162]至此，完成第一注意力子模块的处理过程，得到第一注意力子模块的输出特征。

[163]实施例10：此实施例是在实施例6/实施例7的基础上进一步对级联混合注意力模块中涉及的第一注意力子模块进行描述。在实施例6和实施例7中，级联混合注意力模块包括第一注意力子模块，第一注意力子模块用于对第一注意力子模块的输入特征进行第一处理，得到第一注意力子模块的输出特征。比如说，对第一注意力子模块的输入特征进行层归一化，得到层归一化后的特征；对层归一化后的特征进行窗口划分，得到多个小尺寸特征；针对每个小尺寸特征，对该小尺寸特征进行三路卷积操作，得到该小尺寸特征对应的小尺寸查询特征、小尺寸键特征和小尺寸值特征。对多个小尺寸特征对应的小尺寸查询特征进行拼接得到查询向量；对多个小尺寸特征对应的小尺寸键特征进行拼接得到键向量；对多个小尺寸特征对应的小尺寸值特征进行拼接得到值向量。基于该查询向量和该键向量确定注意力权重，基于该注意力权重和该值向量确定该输入特征对应的修正特征。基于该输入特征和修正特征确定第一注意力子模块的输出特征。

[164]参见图8B所示，为第一注意力子模块的结构示意图，这里只是第一注意力子模块的示例。对此第一注意力子模块的结构不做限制，下面以图8B的第一注意力子模块为例进行说明。

[165]示例性的，第一注意力子模块可以包括层归一化层，可以通过层归一化层对第一注意力子模块的输入特征（或输入图像）进行层归一化，得到层归一化后的特征。

[166]第一注意力子模块可以包括窗口划分层，通过窗口划分层对层归一化后的特征进行窗口划分，得到多个小尺寸特征，比如说，通过窗口划分层将层归一化后的特征划分成多个小尺寸特征，每个小尺寸特征的宽度为 w_i ，每个小尺寸特征的高度为 h_i ，不同小尺

寸特征的宽度可以相同，也可以不同，不同小尺寸特征的高度可以相同，也可以不同。

[167]针对每个小尺寸特征，对该小尺寸特征进行三路卷积操作，得到该小尺寸特征对应的小尺寸查询特征、小尺寸键特征和小尺寸值特征。比如说，通过卷积层（如 1×1 的卷积层和 3×3 的卷积层）对该小尺寸特征进行卷积操作得到小尺寸查询特征。通过卷积层（如 1×1 的卷积层和 3×3 的卷积层）对该小尺寸特征进行卷积操作得到小尺寸键特征。通过卷积层（如 1×1 的卷积层和 3×3 的卷积层）对该小尺寸特征进行卷积操作得到小尺寸值特征。对所有小尺寸特征对应的小尺寸查询特征进行拼接得到查询向量 Q ，对所有小尺寸特征对应的小尺寸键特征进行拼接得到键向量 K ，对所有小尺寸特征对应的小尺寸值特征进行拼接得到值向量 V 。

[168]基于查询向量 Q 和键向量 K 确定注意力权重。比如说，对查询向量 Q 与键向量 K 进行矩阵乘法，将矩阵乘法的结果经过软最大值函数（Softmax），得到注意力权重。

[169]对于所得到的注意力权重和值向量 V ，在一个示例中，可直接基于该注意力权重和该值向量 V 得到输入特征对应的修正特征。例如，对该注意力权重和值向量 V 进行矩阵乘法，将矩阵乘法后的结果作为输入特征对应的修正特征，也就是说，可以去除图 8B 中的第二个层归一化层。

[170]进一步地，可以基于第一注意力子模块的输入特征和该修正特征确定第一注意力子模块的输出特征，比如说，对该输入特征和该修正特征进行矩阵加法，得到第一注意力子模块的输出特征。

[171]至此，完成第一注意力子模块的处理过程，得到第一注意力子模块的输出特征。

[172]实施例 11：此实施例是在实施例 6/实施例 7 的基础上进一步对级联混合注意力模块中涉及的第一注意力子模块进行描述。在实施例 6 和实施例 7 中，级联混合注意力模块包括第一注意力子模块，第一注意力子模块用于对第一注意力子模块的输入特征进行第一处理，得到第一注意力子模块的输出特征。比如说，对第一注意力子模块的输入特征进行层归一化，得到层归一化后的特征；对层归一化后的特征进行三路卷积操作，得到查询向量、键向量和值向量；基于该查询向量和该键向量确定注意力权重，并基于注意力权重和该值向量确定该输入特征对应的修正特征；基于第一注意力子模块的输入特征和修正特征确定第一注意力子模块的输出特征。

[173]比如说，在图 8B 的第一注意力子模块的基础上，可以去除层归一化层后面的窗口划分层。对此第一注意力子模块的结构不做限制，例如，第一注意力子模块的结构示例可以参见图 8B 所示。

[174]示例性的，第一注意力子模块可以包括层归一化层，可以通过层归一化层对第一注意力子模块的输入特征（或输入图像）进行层归一化，得到层归一化后的特征。

[175]第一注意力子模块可包括卷积层，对层归一化后的特征进行三路卷积操作，得到查询向量 Q 、键向量 K 和值向量 V 。比如，可通过卷积层（如 1×1 的卷积层和 3×3 的卷积层）对层归一化后的特征进行卷积操作得到查询向量 Q 。可通过卷积层（如 1×1 的卷积层和 3×3 的卷积层）对层归一化后的特征进行卷积操作得到键向量 K 。可通过卷积层（如 1×1 的卷积层和 3×3 的卷积层）对层归一化后的特征进行卷积操作得到值向量 V 。

[176]基于查询向量 Q 和键向量 K 确定注意力权重。比如说，对查询向量 Q 与键向量 K 进行矩阵乘法，将矩阵乘法的结果经过软最大值函数（Softmax），得到注意力权重。

[177]对于所得到的注意力权重和值向量 V ，在一个示例中，可直接基于该注意力权重和该值向量 V 得到输入特征对应的修正特征。例如，对该注意力权重和值向量 V 进行矩阵乘法，将矩阵乘法后的结果作为输入特征对应的修正特征，也就是说，可以去除图 8B

中的第二个层归一化层。

[178]进一步地,可以基于第一注意力子模块的输入特征和该修正特征确定第一注意力子模块的输出特征,比如说,对该输入特征和该修正特征进行矩阵加法,得到第一注意力子模块的输出特征。

[179]至此,完成第一注意力子模块的处理过程,得到第一注意力子模块的输出特征。

[180]在实施例 8-实施例 11 中,给出了第一注意力子模块的几个示例,本本申请对第一注意力子模块的结构不做限制,可以对输入特征进行第一处理得到输出特征即可。

[181]实施例 12: 此实施例是在实施例 6/实施例 7 的基础上进一步对级联混合注意力模块中涉及的第二注意力子模块进行描述。在实施例 6 和实施例 7 中,级联混合注意力模块和包括第二注意力子模块,第二注意力子模块用于对第二注意力子模块的输入特征进行第二处理,得到第二注意力子模块的输出特征。比如说,对第二注意力子模块的输入特征进行深度可分离卷积操作,得到卷积后的特征(即深度可分离卷积操作后的特征);对卷积后的特征进行层归一化,得到层归一化后的特征;对层归一化后的特征进行多层感知操作,得到多层感知后的特征;基于该输入特征和多层感知后的特征确定第二注意力子模块的输出特征。

[182]参见图 8C 所示,为第二注意力子模块的结构示意图,这里只是第二注意力子模块的示例。对此第二注意力子模块的结构不做限制,下面以图 8C 的第二注意力子模块为例进行说明。

[183]示例性的,第二注意力子模块可以包括深度可分离卷积层,深度可分离卷积层可以为 1x1 的卷积层、或 3x3 的卷积层、或 5x5 的卷积层、或 7x7 的卷积层,对此不作限制,如可以选取 7x7 的卷积层,可以通过深度可分离卷积层对第二注意力子模块的输入特征进行深度可分离卷积操作,得到卷积后的特征(即深度可分离卷积操作后的特征),对此卷积操作不做限制。

[184]第二注意力子模块可以包括层归一化层,可以通过层归一化层对卷积后的特征进行层归一化,得到层归一化后的特征。层归一化的操作包括计算均值和方差、标准化、缩放和平移,使得每个隐藏单元的输在相对较小范围内,且每个隐藏单元在数据集中都有相似分布。

[185]第二注意力子模块可以包括多层感知器,可以通过多层感知器对层归一化后的特征进行多层感知操作,得到多层感知后的特征。参见图 8D 所示,为多层感知器的结构示意图,多层感知器可以包括线性层、激活层、随机失活层、线性层和随机失活层。基于此,层归一化后的特征依次经过线性层、激活层、随机失活层、线性层和随机失活层,得到多层感知后的特征。其中,激活层可以为 relu 层、或 leaky relu 层、或 sigmoid 层、或 tanh 层、或 Gelu 层,对此激活层的类型不作限制。随机失活层可以为 Dropout 层,对此随机失活层的类型不作限制。需要说明的是,图 8D 所示只是多层感知器的示例,对此多层感知器的结构不做限制。

[186]基于第二注意力子模块的输入特征和多层感知后的特征确定第二注意力子模块的输出特征,比如说,对该输入特征和多层感知后的特征进行矩阵加法,得到该输出特征。

[187]至此,完成第二注意力子模块的处理过程,得到第二注意力子模块的输出特征。

[188]实施例 13: 此实施例是在实施例 6/实施例 7 的基础上进一步对级联混合注意力模块中涉及的第二注意力子模块进行描述。在实施例 6 和实施例 7 中,级联混合注意力模块均包括第二注意力子模块,第二注意力子模块用于对第二注意力子模块的输入特征进行第二处理,得到第二注意力子模块的输出特征。比如说,可以对第二注意力子模

块的输入特征进行层归一化,得到层归一化后的特征;可以对层归一化后的特征进行第一线性操作,得到第一线性操作后的特征;可以对第一线性操作后的特征进行激活操作,得到激活后的特征;可以对激活后的特征进行第二线性操作,得到第二线性操作后的特征;可以基于第二注意力子模块的输入特征和第二线性操作后的特征确定第二注意力子模块的输出特征。

[189]参见图 8E 所示,为第二注意力子模块的结构示意图,这里只是第二注意力子模块的示例。对此第二注意力子模块的结构不做限制,下面以图 8E 的第二注意力子模块为例进行说明。

[190]示例性的,第二注意力子模块可以包括层归一化层,可以通过层归一化层对第二注意力子模块的输入特征进行层归一化,得到层归一化后的特征。第二注意力子模块可以包括线性层 1,可以通过线性层 1 对层归一化后的特征进行第一线性操作,得到第一线性操作后的特征。

[191]第二注意力子模块可以包括激活层,可以通过激活层对第一线性操作后的特征进行激活操作,得到激活后的特征。其中,激活层可以为 relu 层、或 leaky relu 层、或 sigmoid 层、或 tanh 层、或 Gelu 层,对此激活层的类型不作限制,如选取 relu 层作为激活层。第二注意力子模块可以包括线性层 2,可以通过线性层 2 对激活后的特征进行第二线性操作,得到第二线性操作后的特征。

[192]基于第二注意力子模块的输入特征和第二线性操作后的特征确定第二注意力子模块的输出特征,比如说,对该输入特征和第二线性操作后的特征进行矩阵加法,得到该输出特征。

[193]至此,完成第二注意力子模块的处理过程,得到第二注意力子模块的输出特征。

[194]实施例 14: 此实施例是在实施例 6/实施例 7 的基础上进一步对级联混合注意力模块中涉及的第二注意力子模块进行描述。在实施例 6 和实施例 7 中,级联混合注意力模块包括第二注意力子模块,第二注意力子模块用于对第二注意力子模块的输入特征进行第二处理,得到第二注意力子模块的输出特征。比如说,对第二注意力子模块的输入特征进行下采样操作,得到下采样后的特征;对下采样后的特征进行残差卷积操作,得到残差卷积后的特征;对这个残差卷积后的特征进行上采样操作,得到上采样后的特征;对上采样后的特征进行激活操作,得到激活后的特征;对该输入特征进行残差卷积操作,得到残差卷积后的卷积特征。基于该输入特征、激活后的特征和卷积特征确定第二注意力子模块的输出特征。

[195]参见图 8F 所示,为第二注意力子模块的结构示意图,这里只是第二注意力子模块的示例。对此第二注意力子模块的结构不做限制,下面以图 8F 的第二注意力子模块为例进行说明。

[196]示例性的,第二注意力子模块可以包括下采样层,该下采样层可以为核大小为 3x3、stride 为 2 的卷积层,或者,核大小为 3x3、stride 为 4 的卷积层,或者,核大小为 4x4、stride 为 2 的卷积层,或者,核大小为 4x4、stride 为 4 的卷积层,或者,核大小为 5x5、stride 为 2 的卷积层,或者,核大小为 5x5、stride 为 4 的卷积层,或者,该下采样层还可以为反像素重组层(inverse PixelShuffle)。对此下采样层的结构不做限制,如可以选取核大小为 3x3、stride 为 2 的卷积层作为下采样层。在此基础上,可以通过下采样层对第二注意力子模块的输入特征进行下采样操作,得到下采样后的特征,对此下采样操作的实现不做限制。

[197]第二注意力子模块可以包括 M 个残差卷积层,M 可以为正整数,如 1、2、3、4 等,可以通过 M 个残差卷积层对下采样后的特征进行残差卷积操作,得到残差卷积后

的特征。比如说,通过第1个残差卷积层对下采样后的特征进行残差卷积操作,得到的结果经过第2个残差卷积层进行残差卷积操作,以此类推,一直到最后一个残差卷积层输出残差卷积后的特征。

[198]参见图 8G 所示,为残差卷积层的一个结构示意图,该残差卷积层可以包括卷积层 1、激活层和卷积层 2,可以通过卷积层 1 对残差卷积层的输入特征进行第一卷积处理,得到第一卷积处理后的特征,可以通过激活层对第一卷积处理后的特征进行激活处理,得到激活处理后的特征,可以通过卷积层 2 对激活处理后的特征进行第二卷积处理,得到第二卷积处理后的特征。然后,对残差卷积层的输入特征与第二卷积处理后的特征进行矩阵相加,得到残差卷积层的输出特征。

[199]卷积层 1 可以为 1x1 的卷积层、或 3x3 的卷积层、或 5x5 的卷积层、或 7x7 的卷积层,对此不做限制,如选取 3x3 的卷积层。卷积层 2 可以为 1x1 的卷积层、或 3x3 的卷积层、或 5x5 的卷积层、或 7x7 的卷积层,对此不做限制,如选取 3x3 的卷积层。激活层为 relu 层、或 leaky relu 层、或 sigmoid 层、或 tanh 层、或 Gelu 层,对此不作限制,如选取 relu 层或 sigmoid 层。

[200]参见图 8H 所示,为残差卷积层的另一个结构示意图,该残差卷积层可以包括卷积层和激活层,可以通过卷积层对残差卷积层的输入特征进行卷积处理,得到卷积处理后的特征,可以通过激活层对卷积处理后的特征进行激活处理,得到激活处理后的特征。然后,对残差卷积层的输入特征与激活处理后的特征进行矩阵相加,得到残差卷积层的输出特征。

[201]卷积层可以为 1x1 的卷积层、或 3x3 的卷积层、或 5x5 的卷积层、或 7x7 的卷积层,对此不做限制,如可以选取 3x3 的卷积层。激活层可以为 relu 层、或 leaky relu 层、或 sigmoid 层、或 tanh 层、或 Gelu 层,对此不作限制,如可以选取 relu 层或 sigmoid 层。

[202]第二注意力子模块可以包括上采样层,该上采样层可以为核大小为 3x3、stride 为 2 的反卷积层,或者,核大小为 3x3、stride 为 4 的反卷积层,或者,核大小为 4x4、stride 为 2 的反卷积层,或者,核大小为 4x4、stride 为 4 的反卷积层,或者,核大小为 5x5、stride 为 2 的反卷积层,或者,核大小为 5x5、stride 为 4 的反卷积层,或者,该上采样层还可以为像素重组层。对此上采样层的结构不做限制,如可以选取核大小为 4x4、stride 为 2 的反卷积层作为上采样层。在此基础上,可以通过上采样层对残差卷积后的特征(即经过 M 个残差卷积层进行残差卷积操作后得到的特征)进行上采样操作,得到上采样后的特征,对此上采样操作的实现不做限制。

[203]第二注意力子模块可以包括激活层,激活层可以为 relu 层、或 leaky relu 层、或 sigmoid 层、或 tanh 层、或 Gelu 层,对此激活层的结构不作限制,如可以选取 relu 层。在此基础上,可以通过激活层对上采样后的特征进行激活操作,得到激活后的特征。

[204]第二注意力子模块可以包括 N 个残差卷积层,N 可以为正整数,如 1、2、3、4 等,N 与 M 可以相同,也可以不同。可以通过 N 个残差卷积层对第二注意力子模块的输入特征进行残差卷积操作,得到残差卷积后的卷积特征。比如说,通过第 1 个残差卷积层对第二注意力子模块的输入特征进行残差卷积操作,得到的结果经过第 2 个残差卷积层进行残差卷积操作,以此类推,一直到最后一个残差卷积层输出残差卷积后的卷积特征。

[205]参见图 8G 所示,为残差卷积层的一个结构示意图,该残差卷积层可以包括卷积层 1、激活层和卷积层 2。参见图 8H 所示,为残差卷积层的另一个结构示意图,该残差卷积层可以包括卷积层和激活层,可以采用上述任一残差卷积层执行残差卷积操作,在此不再赘述。

[206]参见图 8F 所示,在得到残差卷积后的卷积特征(即经过 N 个残差卷积层进行残差

卷积操作后输出的特征)、激活后的特征(即经过激活层进行激活操作后输出的特征)之后,就可以基于第二注意力子模块的输入特征、激活后的特征和残差卷积后的卷积特征确定第二注意力子模块的输出特征。比如说,可以对激活后的特征与残差卷积后的卷积特征进行矩阵相乘,对相乘后结果与第二注意力子模块的输入特征进行矩阵相加,从而得到第二注意力子模块的输出特征。

[207]至此,完成第二注意力子模块的处理过程,得到第二注意力子模块的输出特征。

[208]在实施例 12-实施例 14 中,给出了第二注意力子模块的几个示例,本申请对第二注意力子模块的结构不做限制,可以对输入特征进行第二处理得到输出特征即可。

[209]实施例 15: 该实施例的实施方式与实施例 1、实施例 2、实施例 3 或实施例 4 的实施方式相同,重复内容不再赘述。下面对其中所涉及的合成变换网络进行详细描述。

[210]在实施例 1、实施例 2、实施例 3 和实施例 4 中,均涉及合成变换网络,该合成变换网络至少可以包括注意力模块,且该注意力模块可以为第一注意力子模块。当然,除了第一注意力子模块,合成变换网络还可以包括其它网络层,本实施例中对此合成变换网络的结构不做限制,只要合成变换网络包括第一注意力子模块即可。比如说,该合成变换网络还可以包括至少一个反卷积层,第一注意力子模块可以位于任意一个反卷积层的后面。或,第一注意力子模块可以位于每个反卷积层的后面。或,第一注意力子模块可以位于部分反卷积层(如所有反卷积层中的 K 个反卷积层, K 大于 1, 且 K 小于反卷积层总数量)的后面。

[211]示例性的,第一注意力子模块用于对第一注意力子模块的输入特征进行第一处理,得到第一注意力子模块的输出特征。在一种可能的实施方式中,第一注意力子模块的结构可以参见实施例 8 所示,在另一种可能的实施方式中,第一注意力子模块的结构可以参见实施例 9 所示,在另一种可能的实施方式中,第一注意力子模块的结构可以参见实施例 10 所示,在另一种可能的实施方式中,第一注意力子模块的结构可以参见实施例 11 所示。当然,上述只是第一注意力子模块的几个示例,对此第一注意力子模块的结构不做限制。

[212]实施例 16: 该实施例的实施方式与实施例 1、实施例 2、实施例 3 或实施例 4 的实施方式相同,重复内容不再赘述。下面对其中所涉及的合成变换网络进行详细描述。

[213]在实施例 1、实施例 2、实施例 3 和实施例 4 中,均涉及合成变换网络,该合成变换网络至少可包括注意力模块,且该注意力模块可为第二注意力子模块。当然,除了第二注意力子模块,合成变换网络还可包括其它网络层,本实施例中对此合成变换网络的结构不做限制,只要合成变换网络包括第二注意力子模块即可。比如说,该合成变换网络还可包括至少一个反卷积层,第二注意力子模块可位于任意一个反卷积层的后面。或,第二注意力子模块位于每个反卷积层的后面。或,第二注意力子模块位于部分反卷积层(如所有反卷积层中的 K 个反卷积层, K 大于 1, 且 K 小于反卷积层总数量)的后面。

[214]示例性的,第二注意力子模块用于对第二注意力子模块的输入特征进行第二处理,得到第二注意力子模块的输出特征。在一种可能的实施方式中,第二注意力子模块的结构可以参见实施例 12 所示,在另一种可能的实施方式中,第二注意力子模块的结构可以参见实施例 13 所示,在另一种可能的实施方式中,第二注意力子模块的结构可以参见实施例 14 所示。当然,上述只是第二注意力子模块的几个示例,对此第二注意力子模块的结构不做限制。

[215]由以上技术方案可见,本申请实施例中,针对基于神经网络的编解码技术,提出一种注意力机制的合成变换网络,该合成变换网络包括注意力模块,且注意力模块为级联混合注意力模块,通过级联混合注意力模块实现合成变换网络时,在保证合成图像质量

的同时,有效降低网络复杂度和计算复杂度,提高解码性能,在降低复杂度的同时,有效保证重建图像块的质量,提升编码性能和解码性能。通过将残差非局部注意力模块替换为混合注意力模块(级联混合注意力模块),大幅降低了解码器的计算复杂度。比如说,混合注意力模块中的第二注意力子模块仅占用 21K 次浮点乘加操作/像素的计算复杂度,而第一注意力子模块的复杂度也远低于残差非局部注意力模块。此外,第二注意力子模块起到了残差非局部注意力模块中注意力机制的作用,并且由于引入性能较优的 transformer 结构,从而带来优异的性能,第二注意力子模块中的 7*7 大小的卷积操作起到了残差非局部注意力模块中引入非局部信息的作用,第二注意力子模块中的随机失活层起到了增加网络鲁棒性的作用。基于上述功能,使得级联混合注意力模块在降低计算复杂度的同时,并不会有过多的性能损失。

[216]示例性的,上述各实施例可以单独实现,也可以组合实现,比如说,实施例 1-实施例 17 中的每个实施例,可以单独实现,实施例 1-实施例 16 中的至少两个实施例,可以组合实现。

[217]示例性的,上述各实施例中,编码端的内容也可以应用到解码端,即解码端可以采用相同方式处理,解码端的内容也可以应用到编码端,即编码端可以采用相同方式处理。

[218]基于与上述方法同样的构思,本申请实施例中还提出一种解码装置,所述解码装置应用于解码端,所述解码装置包括:存储器,其经配置以存储视频数据;解码器,其经配置以实现上述实施例 1-实施例 16 中的解码方法,即解码端的处理流程。

[219]比如说,在一种可能的实施方式中,解码器,其经配置以实现:对当前图像块对应的一个码流进行解码,得到所述当前图像块对应的系数超参特征;基于所述系数超参特征确定概率分布参数,基于所述概率分布参数对所述当前图像块对应的另一码流进行解码,得到所述当前图像块对应的重建特征;将所述重建特征输入合成变换网络,得到所述当前图像块对应的重建图像块;其中,所述合成变换网络至少包括注意力模块,所述注意力模块为级联混合注意力模块;其中,所述级联混合注意力模块包括第一注意力子模块和第二注意力子模块,且所述第一注意力子模块和所述第二注意力子模块为串行的两个子模块。

[220]基于与上述方法同样的构思,本申请实施例中还提出一种编码装置,所述编码装置应用于编码端,所述编码装置包括:存储器,其经配置以存储视频数据;编码器,其经配置以实现上述实施例 1-实施例 16 中的编码方法,即编码端的处理流程。

[221]比如说,在一种可能的实施方式中,编码器,其经配置以实现:对当前图像块对应的一个码流进行解码,得到所述当前图像块对应的系数超参特征;基于所述系数超参特征确定概率分布参数,基于所述概率分布参数对所述当前图像块对应的另一码流进行解码,得到所述当前图像块对应的重建特征;将所述重建特征输入合成变换网络,得到所述当前图像块对应的重建图像块;其中,所述合成变换网络至少包括注意力模块,所述注意力模块为级联混合注意力模块;其中,所述级联混合注意力模块包括第一注意力子模块和第二注意力子模块,且所述第一注意力子模块和所述第二注意力子模块为串行的两个子模块。

[222]基于与上述方法同样的构思,本申请实施例提供的解码端设备(也可以称为视频解码器),从硬件层面而言,其硬件架构示意图具体可以参见图 9A 所示。包括:处理器 901 和机器可读存储介质 902,机器可读存储介质 902 存储有能够被处理器 901 执行的机器可执行指令;处理器 901 用于执行机器可执行指令,以实现本申请上述实施例 1-17 的解码方法。

[223]机器可读存储介质 902 可以是任何电子、磁性、光学或其它物理存储装置,可以包

含或存储信息，如可执行指令、数据，等等。例如，机器可读存储介质可以是随机存取存储器（Random Access Memory, RAM）、易失存储器、非易失性存储器、闪存、存储驱动器（如硬盘驱动器）、固态硬盘、任何类型的存储盘（如光盘、dvd 等），或者类似的存储介质，或者它们的组合。

[224]基于与上述方法同样的构思，本申请实施例提供的编码端设备（也可以称为视频编码器），从硬件层面而言，其硬件架构示意图具体可以参见图 9B 所示。包括：处理器 911 和机器可读存储介质 912，机器可读存储介质 912 存储有能够被处理器 911 执行的机器可执行指令；处理器 911 用于执行机器可执行指令，以实现本申请上述实施例 1-17 的编码方法。

[225]机器可读存储介质 912 可以是任何电子、磁性、光学或其它物理存储装置，可以包含或存储信息，如可执行指令、数据，等等。例如，机器可读存储介质可以是 RAM、易失存储器、非易失性存储器、闪存、存储驱动器（如硬盘驱动器）、固态硬盘、任何类型的存储盘（如光盘、dvd 等），或者类似的存储介质，或者它们的组合。

[226]基于与上述方法同样的构思，本申请实施例提供一种电子设备。包括：处理器和机器可读存储介质，机器可读存储介质存储有能够被处理器执行的机器可执行指令；处理器用于执行机器可执行指令，以实现本申请上述实施例 1-16 的解码方法或编码方法。

[227]基于与上述方法同样的构思，本申请实施例还提供一种机器可读存储介质，所述机器可读存储介质上存储有若干计算机指令，所述计算机指令被处理器执行时，以实现本申请上述实施例 1-16 的解码方法或编码方法，如上述各实施例中的解码方法或者编码方法。例如，所述机器可读存储介质可以是 ROM、RAM、CD-ROM、磁带、软盘和光数据存储设备等。

[228]基于与上述方法同样的构思，本申请实施例还提供一种计算机应用程序，所述计算机应用程序令被处理器执行时，能够实现本申请上述示例公开的解码方法或者编码方法。

[229]基于与上述方法同样的构思，本申请实施例中还提出一种解码装置，可以应用于解码端，所述解码装置包括：解码模块，用于对当前图像块对应的一个码流进行解码，得到所述当前图像块对应的系数超参特征；基于所述系数超参特征确定概率分布参数，基于所述概率分布参数对所述当前图像块对应的另一码流进行解码，得到所述当前图像块对应的重建特征；处理模块，用于将所述重建特征输入合成变换网络，得到所述当前图像块对应的重建图像块；其中，所述合成变换网络至少包括注意力模块，所述注意力模块为级联混合注意力模块或；其中，所述级联混合注意力模块包括第一注意力子模块和第二注意力子模块，且所述第一注意力子模块和所述第二注意力子模块为串行的两个子模块。

[230]示例性的，若所述注意力模块为级联混合注意力模块，所述处理模块，还用于通过所述第一注意力子模块对所述第一注意力子模块的输入特征进行第一处理，得到所述第一注意力子模块的输出特征，所述第一注意力子模块的输出特征为所述第二注意力子模块的输入特征；其中，所述第一处理为层归一化、卷积操作和维度变换操作的至少一种；通过所述第二注意力子模块对所述第二注意力子模块的输入特征进行第二处理，得到所述第二注意力子模块的输出特征；其中，所述第二处理为深度可分离卷积操作、层归一化、多层感知操作、线性操作、激活操作、下采样操作、残差卷积操作、上采样操作的至少一种。

[231]示例性的，所述处理模块通过所述第一注意力子模块对所述第一注意力子模块的输入特征进行第一处理，得到所述第一注意力子模块的输出特征时用于：对所述第一注意

力子模块的输入特征进行层归一化,得到层归一化后的特征;对所述层归一化后的特征进行三路卷积操作,得到查询向量、键向量和值向量;对所述查询向量、所述键向量和所述值向量分别进行维度变换操作,得到维度变换后的查询向量、维度变换后的键向量和维度变换后的值向量;基于所述维度变换后的查询向量和所述维度变换后的键向量确定注意力权重,并基于所述注意力权重和所述维度变换后的值向量确定该输入特征对应的修正特征;基于该输入特征和所述修正特征确定所述第一注意力子模块的输出特征。

[232]示例性的,所述处理模块通过所述第一注意力子模块对所述第一注意力子模块的输入特征进行第一处理,得到所述第一注意力子模块的输出特征时具体用于:对所述第一注意力子模块的输入特征进行层归一化,得到层归一化后的特征;对所述层归一化后的特征进行窗口划分,得到多个小尺寸特征;针对每个小尺寸特征,对该小尺寸特征进行三路卷积操作,得到该小尺寸特征对应的小尺寸查询特征、小尺寸键特征和小尺寸值特征;对所述多个小尺寸特征对应的小尺寸查询特征进行拼接,得到查询向量;对所述多个小尺寸特征对应的小尺寸键特征进行拼接,得到键向量;对所述多个小尺寸特征对应的小尺寸值特征进行拼接,得到值向量;对所述查询向量、所述键向量和所述值向量分别进行维度变换操作,得到维度变换后的查询向量、维度变换后的键向量和维度变换后的值向量;基于维度变换后的查询向量和维度变换后的键向量确定注意力权重,并基于所述注意力权重和维度变换后的值向量确定该输入特征对应的修正特征;基于该输入特征和所述修正特征确定所述第一注意力子模块的输出特征。

[233]示例性的,所述处理模块通过所述第一注意力子模块对所述第一注意力子模块的输入特征进行第一处理,得到所述第一注意力子模块的输出特征时具体用于:对所述第一注意力子模块的输入特征进行层归一化,得到层归一化后的特征;对所述层归一化后的特征进行三路卷积操作,得到查询向量、键向量和值向量;基于所述查询向量和所述键向量确定注意力权重,并基于所述注意力权重和所述值向量确定该输入特征对应的修正特征;基于该输入特征和所述修正特征确定所述第一注意力子模块的输出特征。

[234]示例性的,所述处理模块通过所述第一注意力子模块对所述第一注意力子模块的输入特征进行第一处理,得到所述第一注意力子模块的输出特征时具体用于:对所述第一注意力子模块的输入特征进行层归一化,得到层归一化后的特征;对所述层归一化后的特征进行窗口划分,得到多个小尺寸特征;针对每个小尺寸特征,对该小尺寸特征进行三路卷积操作,得到该小尺寸特征对应的小尺寸查询特征、小尺寸键特征和小尺寸值特征;对所述多个小尺寸特征对应的小尺寸查询特征进行拼接,得到查询向量;对所述多个小尺寸特征对应的小尺寸键特征进行拼接,得到键向量;对所述多个小尺寸特征对应的小尺寸值特征进行拼接,得到值向量;基于所述查询向量和所述键向量确定注意力权重,并基于所述注意力权重和所述值向量确定该输入特征对应的修正特征;基于该输入特征和所述修正特征确定所述第一注意力子模块的输出特征。

[235]示例性的,所述处理模块通过所述第二注意力子模块对所述第二注意力子模块的输入特征进行第二处理,得到所述第二注意力子模块的输出特征时具体用于:对所述第二注意力子模块的输入特征进行层归一化,得到层归一化后的特征;对所述层归一化后的特征进行第一线性操作,得到第一线性操作后的特征;对所述第一线性操作后的特征进行激活操作,得到激活后的特征;对所述激活后的特征进行第二线性操作,得到第二线性操作后的特征;基于该输入特征和所述第二线性操作后的特征确定所述第二注意力子模块的输出特征。

[236]示例性的,所述处理模块通过所述第二注意力子模块对所述第二注意力子模块的输入特征进行第二处理,得到所述第二注意力子模块的输出特征时具体用于:对所述第二注意力子模块的输入特征进行下采样操作,得到下采样后的特征;对所述下采样后的特

征进行残差卷积操作，得到残差卷积后的特征；对所述残差卷积后的特征进行上采样操作，得到上采样后的特征；对所述上采样后的特征进行激活操作，得到激活后的特征；对该输入特征进行残差卷积操作，得到残差卷积后的卷积特征，并基于所述第二注意力子模块输入特征、所述激活后的特征和所述卷积特征确定所述第二注意力子模块的输出特征。

[237]示例性的，所述合成变换网络还包括至少一个反卷积层；其中：所述注意力模块位于其中一个反卷积层的后面。

[238]示例性的，所述合成变换网络还包括残差层、第一反卷积层、第一裁剪层、第一残差激活层、第二反卷积层、第二裁剪层、第二残差激活层、第三反卷积层、第三裁剪层、第三残差激活层、第四反卷积层、第四裁剪层；所述合成变换网络包括至少一个注意力模块，其中一个注意力模块位于所述第一反卷积层的后面。

[239]基于与上述方法同样的构思，本申请实施例中还提出一种编码装置，所述装置应用于编码端，所述装置包括：解码模块，用于对当前图像块对应的一个码流进行解码，得到所述当前图像块对应的系数超参特征；基于所述系数超参特征确定概率分布参数，基于所述概率分布参数对所述当前图像块对应的另一码流进行解码，得到所述当前图像块对应的重建特征；处理模块，用于将所述重建特征输入合成变换网络，得到所述当前图像块对应的重建图像块；其中，所述合成变换网络至少包括注意力模块，所述注意力模块为级联混合注意力模块；其中，所述级联混合注意力模块包括第一注意力子模块和第二注意力子模块，且所述第一注意力子模块和所述第二注意力子模块为串行的两个子模块。

[240]本领域内的技术人员应明白，本申请实施例可提供为方法、系统、或计算机程序产品。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。本申请实施例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。以上所述仅为本申请的实施例而已，并不用于限制本申请。

[241]对于本领域技术人员来说，本申请可以有各种更改和变化。凡在本申请的精神和原理之内所作的任何修改、等同替换、改进等，均应包含在本申请的权利要求范围之内。

权利要求书

1、一种解码方法，其特征在于，应用于解码端，所述方法包括：

对当前图像块对应的一个码流进行解码，得到所述当前图像块对应的系数超参特征；
基于所述系数超参特征确定概率分布参数，基于所述概率分布参数对所述当前图像块对应的另一码流进行解码，得到所述当前图像块对应的重建特征；

将所述重建特征输入合成变换网络，得到所述当前图像块对应的重建图像块；

其中，所述合成变换网络至少包括注意力模块，所述注意力模块为级联混合注意力模块；

其中，所述级联混合注意力模块包括第一注意力子模块和第二注意力子模块，且所述第一注意力子模块和所述第二注意力子模块为串行的两个子模块。

2、根据权利要求1所述的方法，其特征在于，若所述注意力模块为级联混合注意力模块，所述方法还包括：

通过所述第一注意力子模块对所述第一注意力子模块的输入特征进行第一处理，得到所述第一注意力子模块的输出特征，所述第一注意力子模块的输出特征为所述第二注意力子模块的输入特征；其中，所述第一处理为层归一化、卷积操作和维度变换操作中的至少一种；

通过所述第二注意力子模块对所述第二注意力子模块的输入特征进行第二处理，得到所述第二注意力子模块的输出特征；其中，所述第二处理为深度可分离卷积操作、层归一化、多层感知操作、线性操作、激活操作、下采样操作、残差卷积操作、上采样操作中的至少一种。

3、根据权利要求2所述的方法，其特征在于，通过所述第一注意力子模块对所述第一注意力子模块的输入特征进行第一处理，得到所述第一注意力子模块的输出特征，包括：

对所述第一注意力子模块的输入特征进行层归一化，得到层归一化后的特征；

对所述层归一化后的特征进行三路卷积操作，得到查询向量、键向量和值向量；

对所述查询向量、所述键向量和所述值向量分别进行维度变换操作，得到维度变换后的查询向量、维度变换后的键向量和维度变换后的值向量；

基于维度变换后的查询向量和维度变换后的键向量确定注意力权重，并基于所述注意力权重和维度变换后的值向量确定该输入特征对应的修正特征；

基于该输入特征和所述修正特征确定所述第一注意力子模块的输出特征。

4、根据权利要求2所述的方法，其特征在于，通过所述第一注意力子模块对所述第一注意力子模块的输入特征进行第一处理，得到所述第一注意力子模块的输出特征，包括：

对所述第一注意力子模块的输入特征进行层归一化，得到层归一化后的特征；

对所述层归一化后的特征进行窗口划分，得到多个小尺寸特征；

针对每个小尺寸特征，对该小尺寸特征进行三路卷积操作，得到该小尺寸特征对应的小尺寸查询特征、小尺寸键特征和小尺寸值特征；对所述多个小尺寸特征对应的小尺寸查询特征进行拼接，得到查询向量；对所述多个小尺寸特征对应的小尺寸键特征进行拼接，得到键向量；对所述多个小尺寸特征对应的小尺寸值特征进行拼接，得到值向量；

对所述查询向量、所述键向量和所述值向量分别进行维度变换操作，得到维度变换后的查询向量、维度变换后的键向量和维度变换后的值向量；

基于维度变换后的查询向量和维度变换后的键向量确定注意力权重，并基于所述注意力权重和维度变换后的值向量确定该输入特征对应的修正特征；

基于该输入特征和所述修正特征确定所述第一注意力子模块的输出特征。

5、根据权利要求2所述的方法，其特征在于，通过所述第一注意力子模块对所述第一注意力子模块的输入特征进行第一处理，得到所述第一注意力子模块的输出特征，

包括:

对所述第一注意力子模块的输入特征进行层归一化,得到层归一化后的特征;

对所述层归一化后的特征进行三路卷积操作,得到查询向量、键向量和值向量;

基于所述查询向量和所述键向量确定注意力权重,并基于所述注意力权重和所述值向量确定该输入特征对应的修正特征;

基于该输入特征和所述修正特征确定所述第一注意力子模块的输出特征。

6、根据权利要求2所述的方法,其特征在于,通过所述第一注意力子模块对所述第一注意力子模块的输入特征进行第一处理,得到所述第一注意力子模块的输出特征,包括:

对所述第一注意力子模块的输入特征进行层归一化,得到层归一化后的特征;

对所述层归一化后的特征进行窗口划分,得到多个小尺寸特征;

针对每个小尺寸特征,对该小尺寸特征进行三路卷积操作,得到该小尺寸特征对应的小尺寸查询特征、小尺寸键特征和小尺寸值特征;对所述多个小尺寸特征对应的小尺寸查询特征进行拼接,得到查询向量;对所述多个小尺寸特征对应的小尺寸键特征进行拼接,得到键向量;对所述多个小尺寸特征对应的小尺寸值特征进行拼接,得到值向量;

基于所述查询向量和所述键向量确定注意力权重,并基于所述注意力权重和所述值向量确定该输入特征对应的修正特征;

基于该输入特征和所述修正特征确定所述第一注意力子模块的输出特征。

7、根据权利要求2所述的方法,其特征在于,通过所述第二注意力子模块对所述第二注意力子模块的输入特征进行第二处理,得到所述第二注意力子模块的输出特征,包括:

对所述第二注意力子模块的输入特征进行层归一化,得到层归一化后的特征;

对所述层归一化后的特征进行第一线性操作,得到第一线性操作后的特征;

对所述第一线性操作后的特征进行激活操作,得到激活后的特征;

对所述激活后的特征进行第二线性操作,得到第二线性操作后的特征;

基于该输入特征和所述第二线性操作后的特征确定所述第二注意力子模块的输出特征。

8、根据权利要求2所述的方法,其特征在于,通过所述第二注意力子模块对所述第二注意力子模块的输入特征进行第二处理,得到所述第二注意力子模块的输出特征,包括:

对所述第二注意力子模块的输入特征进行下采样操作,得到下采样后的特征;对所述下采样后的特征进行残差卷积操作,得到残差卷积后的特征;对所述残差卷积后的特征进行上采样操作,得到上采样后的特征;对所述上采样后的特征进行激活操作,得到激活后的特征;

对该输入特征进行残差卷积操作,得到残差卷积后的卷积特征;

基于该输入特征、所述激活后的特征和所述卷积特征确定所述第二注意力子模块的输出特征。

9、根据权利要求1-8任一所述的方法,其特征在于,

所述合成变换网络还包括至少一个反卷积层;其中:

所述注意力模块位于其中一个反卷积层的后面。

10、根据权利要求9所述的方法,其特征在于,所述合成变换网络还包括残差层、第一反卷积层、第一裁剪层、第一残差激活层、第二反卷积层、第二裁剪层、第二残差激活层、第三反卷积层、第三裁剪层、第三残差激活层、第四反卷积层、第四裁剪层;

所述合成变换网络包括至少一个注意力模块,其中一个注意力模块位于所述第一反卷积层的后面。

11、一种编码方法,其特征在于,应用于编码端,所述方法包括:

对当前图像块对应的一个码流进行解码,得到所述当前图像块对应的系数超参特征;
基于所述系数超参特征确定概率分布参数,基于所述概率分布参数对所述当前图像块对应的另一码流进行解码,得到所述当前图像块对应的重建特征;

将所述重建特征输入合成变换网络,得到所述当前图像块对应的重建图像块;

其中,所述合成变换网络至少包括注意力模块,所述注意力模块为级联混合注意力模块;其中,所述级联混合注意力模块包括第一注意力子模块和第二注意力子模块,且所述第一注意力子模块和所述第二注意力子模块为串行的两个子模块。

12、一种解码装置,其特征在于,应用于解码端,所述装置包括:

解码模块,用于对当前图像块对应的一个码流进行解码,得到所述当前图像块对应的系数超参特征;基于所述系数超参特征确定概率分布参数,基于所述概率分布参数对所述当前图像块对应的另一码流进行解码,得到所述当前图像块对应的重建特征;

处理模块,用于将所述重建特征输入合成变换网络,得到所述当前图像块对应的重建图像块;其中,所述合成变换网络至少包括注意力模块,所述注意力模块为级联混合注意力模块;其中,所述级联混合注意力模块包括第一注意力子模块和第二注意力子模块,且所述第一注意力子模块和所述第二注意力子模块为串行的两个子模块。

13、一种编码装置,其特征在于,应用于编码端,所述装置包括:

解码模块,用于对当前图像块对应的一个码流进行解码,得到所述当前图像块对应的系数超参特征;基于所述系数超参特征确定概率分布参数,基于所述概率分布参数对所述当前图像块对应的另一码流进行解码,得到所述当前图像块对应的重建特征;

处理模块,用于将所述重建特征输入合成变换网络,得到所述当前图像块对应的重建图像块;其中,所述合成变换网络至少包括注意力模块,所述注意力模块为级联混合注意力模块;其中,所述级联混合注意力模块包括第一注意力子模块和第二注意力子模块,且所述第一注意力子模块和所述第二注意力子模块为串行的两个子模块。

14、一种解码端设备,其特征在于,包括:处理器和机器可读存储介质,所述机器可读存储介质存储有能够被所述处理器执行的机器可执行指令;

所述处理器用于执行机器可执行指令,以实现权利要求 1-10 任一项所述的方法。

15、一种编码端设备,其特征在于,包括:处理器和机器可读存储介质,所述机器可读存储介质存储有能够被所述处理器执行的机器可执行指令;

所述处理器用于执行机器可执行指令,以实现权利要求 11 所述的方法。

16、一种电子设备,其特征在于,包括:处理器和机器可读存储介质,所述机器可读存储介质存储有能够被所述处理器执行的机器可执行指令;

所述处理器用于执行机器可执行指令,以实现权利要求 1-11 任一项所述的方法。

17、一种机器可读存储介质,其特征在于,所述机器可读存储介质上存储有若干计算机指令,所述计算机指令被处理器执行时,实现权利要求 1-11 任一项所述的方法。

18、一种计算机应用程序,所述计算机应用程序令被处理器执行时,实现权利要求 1-11 任一项所述的方法。