

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2022年9月1日 (01.09.2022)



(10) 国际公布号
WO 2022/179509 A1

- (51) 国际专利分类号:
H04N 21/2343 (2011.01)
- (21) 国际申请号: PCT/CN2022/077341
- (22) 国际申请日: 2022年2月23日 (23.02.2022)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202110212042.1 2021年2月25日 (25.02.2021) CN
- (71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
- (72) 发明人: 葛运英 (GE, Yunying); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong

518129 (CN)。王晶 (WANG, Jing); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。师一博 (SHI, Yibo); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。高尚银 (GAO, Shangyin); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK,

(54) Title: AUDIO/VIDEO OR IMAGE LAYERED COMPRESSION METHOD AND APPARATUS

(54) 发明名称: 音视频或图像分层压缩方法和装置

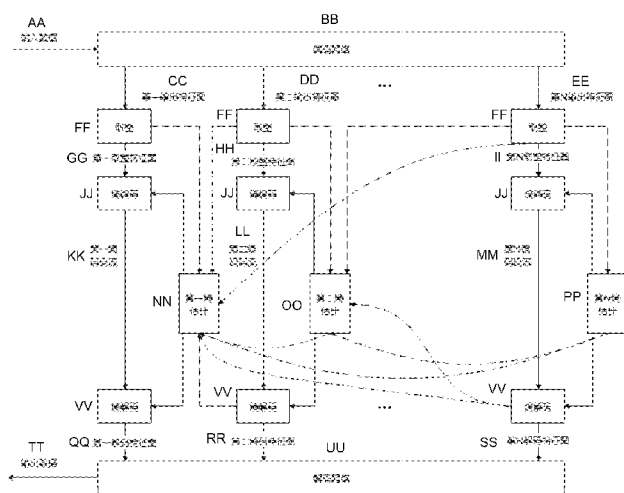


图 14A

- AA Input data
- BB Encoding network
- CC First output feature map
- DD Second output feature map
- EE Nth output feature map
- FF Rounding
- GG First rounding feature map
- HH Second rounding feature map
- II Nth rounding feature map
- JJ Entropy encoding
- KK First coding code stream
- LL Second coding code stream
- MM Nth coding code stream
- NN First entropy estimation
- OO Second entropy estimation
- PP Nth entropy estimation
- QQ First decoding feature map
- RR Second decoding feature map
- SS Nth decoding feature map
- TT Output data
- UU Decoding network
- VV Entropy decoding

(57) Abstract: Provided are an audio/video or image layered compression method and apparatus. The present application relates to the technical field of audio/video or image compression based on artificial intelligence (AI), and in particular to the technical field of audio/video or image compression based on neural networks. The method comprises: transforming an original audio/video or image into a feature space by means of a multi-layer convolution operation, and extracting features of different layers in the feature space; and outputting rounded feature signals of different layers, predicting a probability distribution of shallow-layer feature signals by using deep-layer feature signals or entropy estimation results, and performing entropy encoding on the rounded feature signals. In the present application, signal correlation between different layers is utilized. In this way, audio/video or image compression performance can be improved.

(57) 摘要: 本申请提供了音视频或图像压缩方法和装置。涉及基于人工智能(AI)的音视频或图像压缩技术领域, 具体涉及基于神经网络的音视频或图像压缩技术领域。其中方法包括: 将原始音视频或图像经过多层卷积操作变换到特征空间并提取特征空间不同层的特征, 输出不同层取整后特征信号利用深层特征信号或熵估计结果预测浅层特征信号的概率分布, 并对取整后的特征信号进行熵编码。本申请不同层间的信号相关性得以利用。这样, 可以提升音视频或图像压缩性能。

WO 2022/179509 A1

SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, WS, ZA, ZM, ZW。

- (84) 指定国 (除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告 (条约第21条(3))。

音视频或图像分层压缩方法和装置

本申请要求于2021年02月25日提交中国专利局、申请号为202110212042.1、申请名称为“音视频或图像分层压缩方法和装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本发明实施例涉及基于人工智能(AI)的音视频或图像压缩技术领域，尤其涉及一种音视频或图像分层压缩方法及装置。

背景技术

音视频或图像编解码(视频编码和解码)广泛用于数字音视频或图像应用，例如广播数字电视、互联网和移动网络上的音视频或图像传输、音视频聊天和音视频会议等实时会话应用、DVD和蓝光光盘、音视频或图像内容采集和编辑系统以及可携式摄像机的安全应用。

即使在影片较短的情况下也需要对大量的音视频数据进行描述，当数据要在带宽容量受限的网络中发送或以其它方式传输时，这样可能会造成困难。因此，音视频或图像数据通常要先压缩然后在现代电信网络中传输。由于内存资源可能有限，当在存储设备上存储音视频时，音视频或图像的大小也可能成为问题。音视频或图像压缩设备通常在信源侧使用软件和/或硬件，以在传输或存储之前对视频数据进行编码，从而减少用来表示数字音视频或图像所需的数据量。然后，压缩的数据在目的地侧由视频解压缩设备接收。在有限的网络资源以及对更高视频质量的需求不断增长的情况下，需要改进压缩和解压缩技术，这些改进的技术能够提高压缩率而几乎不影响音图像质量。

近年来，将深度学习应用于在音视频或图像编解码领域逐渐成为一种趋势。如谷歌已连续几年在CVPR会议上组织CLIC(Challenge on Learned Image Compression)专题竞赛，CLIC专注使用深度神经网络来提升图像的压缩效率，在2020年CLIC中还加入了视频或图像挑战类别。基于竞赛方案的性能评估，当前基于深度学习技术的图像编解码方案的综合压缩效率已经与最新一代视频图像编解码标准VVC(Versatile Video Coding)相当，而且在提升用户感知质量方面有独特优势。

VVC的视频标准制定工作已于2020年6月完成，标准收纳几乎所有能够带来显著压缩效率提升的技术算法。因此沿传统信号处理路径继续研究新型的压缩编码算法在短时间内难以获得大的技术突破。区别于传统图像算法通过人工设计来对图像压缩的各模块进行优化，端到端AI的图像压缩是作为一个整体共同进行优化，因此AI图像压缩方案的压缩效果更好。变分自编码器(Variational Autoencoder, AE)方法是当前AI图像有损压缩技术的主流技术方

案。在目前的主流技术方案中，存在着在固定模型的高码率下模型压缩质量趋于平稳的 AE 极限问题，虽然通过增加算力可以缓解 AE 极限问题，但是会导致高算力问题。

发明内容

本申请提供一种音视频或图像分层压缩方法和装置，能够提高保证低码率压缩质量的情况下，高码率突破 AE 极限以及减少模型算力和运行时间。

第一方面，提供了一种音视频或图像的编码方法，包括：对待编码音频或图像经过编码网络获得所述待编码音频或图像的第一输出特征图和所述待编码音频或图像的第二输出特征图；对所述第一输出特征图经过取整以获得所述待编码音频或图像的第一取整特征图；对所述第二输出特征图经过取整以获得所述待编码音频或图像的第二取整特征图；根据第二熵估计的概率估计结果，对所述第二取整特征图进行熵编码得到第二编码码流；根据所述第二取整特征图或/和所述第二熵估计的概率估计结果，进行第一熵估计的概率估计；根据所述第一熵估计的概率估计结果，对所述第一取整特征图进行熵编码得到第一编码码流。

上述技术方案中，通过根据第二取整特征图或/和第二熵估计的概率估计结果进行第一熵估计的概率估计，使得不同网络间的信号相关性得以利用。这样，可以提升音频或图像压缩性能。

上述技术方案中，根据奈奎斯特采样定律高频信号需要更小感受野，低频信号需要更大感受野。编码网络内的浅层网络对应所述第一编码网络的感受野小，有利于捕获高频信息，编码网络内的深层网络对应待编码信号经过了第一编码网络和第二编码网络，感受野大，有利于捕获低频信号。高低频信息分层传输，高频信息只进行浅层网络计算，能够减小算力。

上述技术方案中，取整操作可以包括向下取整操作，或者采用四舍五入取整操作或者采用向上取整操作。

上述技术方案中，熵估计是用于对取整特征图中的各点进行概率估计，实现模型压缩的码率估计。

上述技术方案中，熵编码方案包括且不限于变长编码和算术编码方式。其中变长编码方案可以包括指数哥伦布编码(Exponential Golomb Codes)、上下文自适应可变长编码(CAVLC)；算术编码方式可以包括基于上下文的自适应二进制算术编码(CABAC)。

在一种可能的实现方式中，所述方法还包括：对所述待编码音频或图像经过编码网络获得所述待编码音频或图像的第三输出特征图，以及对所述第三输出特征图取整获得所述待编码音频或图像的第三取整特征图；根据第三熵估计的概率估计结果，对所述第三取整特征图进行熵编码得到第三编码码流。

在一种可能的实现方式中，所述待编码音频或图像的第三熵估计结果和/或所述第三取整

特征图作为所述第二熵估计的输入信息。

在一种可能的实现方式中，所述方法还包括：对所述待编码音频或图像经过编码网络获得所述待编码音频或图像的第 k 输出特征图，以及对所述第 k 输出特征图取整获得所述待编码音频或图像的第 k 取整特征图，其中 $k > 3$ 整数。

上述技术方案中，所述第 k 取整特征图和/或第 k 熵估计结果作为第 n 熵估计的输入信息，其中 $n \leq k$ 整数。

在一种可能的实现方式中，所述方法还包括：对所述至少一个取整特征图作为待编码音频或图像的解码网络输入，经过所述解码网络后获得所述待编码音频或图像的重构音频或图像；对待编码音频或图像与所述重构音频或图像的对应像素点求差以获得所述待编码音频或图像的残差图像；根据以下 3 种信息中的至少一种信息得到待解码音频或图像的无损熵估计的概率估计：

- 1) 所述任意一个或多个熵估计结果；
- 2) 所述任意一个或多个取整特征图；或
- 3) 所述重构音频或图像；

根据所述无损熵估计结果，对所述残差图像进行熵编码以获得所述待编码音频或图像的无损编码码流。

在一种可能的实现方式中，所述任意一个编码网络为单个或多个的卷积网络。

在一种可能的实现方式中，所述第 r 熵估计的概率估计结果输出为第 r 取整特征图的均值和方差信息。

在一种可能的实现方式中，所述方法还包括：对所述第一输出特征图经过取整以获得所述待编码音频或图像的第一取整特征图之前，所述方法还包括：对所述第一输出特征图中的像素点的像素值与第一熵估计输出的均值相减得到第一偏移特征图。对所述第一输出特征图经过取整以获得所述待编码音频或图像的第一取整特征图包括：对所述第一偏移特征图经过取整以获得所述待编码音频或图像的所述第一取整特征图。

应理解，上一种可能的实现方式也可能应用于任意一个取整特征图的获得方式。

应理解，对上述图像压缩方法可以应用在神经网络的训练阶段，或者也可以应用在神经网络的推理阶段，本申请对此不做具体限定。

应理解，将待编码音频或图像经过编码网络操作变换到了特征空间，通过多个编码网络所对应的多个输出特征图从而提取到了特征空间中不同层的特征用来压缩，例如 1, 2, ..., n 层或者其中若干层，1, 2, ..., n 层特征频率依次降低，另外用来压缩的特征层数可变，每层的特征通道数可变，以达到动态调整压缩能力。

第二方面，提供了一种音频或图像的解码方法，包括：获取待解码音频或图像的第一待

解码码流；获取待解码音频或图像的第二待解码码流；根据所述待解码音频或图像的第二熵估计的概率估计结果，对所述解码图像的第二解码码流进行熵解码获得所述待解码音频或图像的第二解码特征图；根据所述第二特征图或/和所述第二熵估计的概率估计结果，对第一熵估计进行概率估计以获得第一熵估计的概率估计结果；根据所述第一熵估计的概率估计结果，对所述第一解码码流进行熵解码获得所述待解码音频或图像的第一解码特征图；所述第一解码特征图或/和所述第二解码特征图作为解码网络的输入，输出所述待解码音频或图像的重构音频或图像。

上述技术方案中，通过根据第二解码特征图或/和第二熵估计的概率估计结果进行第一熵估计的概率估计，使得特征空间中不同层间的信号相关性得以利用。这样，可以提升图像压缩性能。

上述技术方案中，解码过程的熵估计是用于对解码特征图中的各点进行概率估计，实现模型压缩的码率估计。

上述技术方案中，熵解码方案包括且不限于变长解码和算术解码方式。其中变长解码方案可以包括指数哥伦布解码(Exponential Golomb Codes)、上下文自适应可变长解码(CAVLC)；算术解码方式可以包括基于上下文的自适应二进制算术解码(CABAC)。

在一种可能的实现方式中，所述方法还包括：获取待解码音频或图像的第三待解码码流；根据所述待解码音频或图像的第三熵估计的概率估计结果，对所述第三待解码码流进行熵解码获得所述待解码音频或图像的第三解码特征图，根据所述第三解码特征图或/和所述第三熵估计的概率估计结果，进行待解码音频或图像的第二熵估计的概率估计；所述输出所述待解码音频或图像的重构音频或图像具体包括：根据所述第一解码特征图、第二解码特征图以及所述第三解码特征图中的至少一种解码特征图作为解码网络的输入，所述解码网络输出所述待解码音频或图像的重构音频或图像。

在一种可能的实现方式中，根据所述第三解码特征图或/和所述第三熵估计的概率估计结果，进行待解码音频或图像的第二熵估计的概率估计。

在一种可能的实现方式中，所述方法还包括：获取待解码音频或图像的第 m 待解码码流；根据所述待解码音频或图像的第 m 熵估计的概率估计结果，对所述第 m 待解码码流进行熵解码获得所述待解码音频或图像的第 m 解码特征图，其中 $m > 3$ 整数。

在上述可能的实现方式中，所述第 m 解码特征图作为第 t 熵估计的输入信息，其中 $t \leq m$ 整数。

在一种可能的实现方式中，所述方法还包括：获取待解码音频或图像的无损待解码码流；根据以下 3 种信息中的至少一种信息得到待解码音频或图像的无损熵估计的概率估计结果。

1) 所述至少一个解码特征图；

2) 所述至少一个熵估计的概率估计结果; 或

3) 所述待解码音频或图像的重构音频或图像;

根据所述待解码音频或图像的无损熵估计的概率估计结果, 对所述无损待解码码流进行熵解码获得所述待解码音频或图像的残差图像; 对所述待解码音频或图像的重构音频或图像与所述待解码音频或图像的残差图像的对应像素点求和以获得所述待解码音频或图像的无损重构音频或图像。

在一种可能的实现方式中, 所述方法还包括: 当所述第一解码特征图作为解码网络的输入, 对所述第一解码特征图中的像素点的像素值与第一熵估计输出的均值相加得到第一解码特征图。

应理解, 上一种可能的实现方式也可能应用于任意一个解码特征图的获得方式。

应理解, 对上述图像压缩方法可以应用在神经网络的训练阶段, 或者也可以应用在神经网络的推理阶段, 本申请对此不做具体限定。

由于采用了分层解码方案, 一方面不同层提取不同频率分量的特征信息, 从而对不同层的熵解码更有针对性和有效, 有利于信号的压缩和编码性能的提升。另一方面, 利用了层间特征信号的相关性, 也有利于熵估计实现。这两方面因素保障了在低码率情况, 即只利用少量分层解码特征进行重构输出也能保障压缩质量。同样的, 这两方面使得在高码率情况, 即利用更多的分层解码特征进行重构输出能进一步的提升压缩质量, 突破 AE 极限。由于采用了分层结构, 能够根据实际需求采用分层解码方案。当采用解码网络浅层输出, 得到解码特征所需要的解码网络中的网络层数更少, 从而减少了模型算力和运行时间。

第三方面, 提供了一种音频或图像压缩处理装置, 包括:

编码网络模块, 将待编码音频或图像经过编码网络操作变换到了特征空间, 通过多个编码网络所对应的多个输出特征图从而提取到了特征空间中不同层的特征用来压缩, 例如 1, 2, ..., n 层或者其中若干层, 1, 2, ..., n 层特征频率依次降低, 另外用来压缩的特征层数可变, 每层的特征通道数可变, 以达到动态调整压缩能力。

取整模块, 用于将输出特征图中的各点取整为整数以便后续编码;

熵估计网络模块: 用于对取整特征图中的各点进行概率估计, 实现模型压缩的码率估计;

熵编解码模块, 利用熵估计模型得到的概率分布, 对多个取整特征图进行熵编码得到多个编码码流, 降低了输出特征的编码冗余, 进一步降低图像压缩过程中的数据传输量。对编码码流进行解码得到各解码特征图;

解码网络模块, 用于对解码特征图进行反变换, 将解码特征图解析为音频或图像。

上述编码网络模块、取整模块和熵估计网络模块、熵编解码模块、解码网络模块的进一步实现功能可以参考第一、二方面或者第一、二方面的任意一种实现方式, 此处不再赘述。

第四方面，本申请提供一种编码器，包括处理电路，用于执行根据上述第一方面及第一方面任一项所述的方法。

第五方面，本申请提供一种解码器，包括处理电路，用于执行上述第二方面及第二方面任一项所述的方法。

第六方面，本申请提供一种计算机程序产品，包括程序代码，当其在计算机或处理器上执行时，用于执行上述第一方面及第一方面任一项、上述第二方面及第二方面任一项所述的方法。

第七方面，本申请提供一种编码器，包括：一个或多个处理器；非瞬时性计算机可读存储介质，耦合到所述处理器并存储由所述处理器执行的程序，其中所述程序在由所述处理器执行时，使得所述解码器执行上述第一方面及第一方面任一项所述的方法。

第八方面，本申请提供一种解码器，包括：一个或多个处理器；非瞬时性计算机可读存储介质，耦合到所述处理器并存储由所述处理器执行的程序，其中所述程序在由所述处理器执行时，使得所述编码器执行上述第二方面及第二方面任一项所述的方法所述的方法。

第九方面，本申请提供一种非瞬时性计算机可读存储介质，包括程序代码，当其由计算机设备执行时，用于执行上述第一方面及第一方面任一项、上述第二方面及第二方面任一项所述的方法。

第十方面，本发明涉及编码装置，具有实现上述第一方面或第一方面任一项的方法实施例中的行为的功能。所述功能可以通过硬件实现，也可以通过硬件执行相应的软件实现。所述硬件或软件包括一个或多个与上述功能相对应的模块。在一个可能的设计中，所述编码装置包括：编码网络模块，用于将原始图像经过多层卷积操作变换到特征空间，提取特征空间不同层的特征，用来压缩，例如 1, 2, ..., n 层或者其中若干层，1, 2, ..., n 层特征频率依次降低，另外用来压缩的特征层数可变，每层的特征通道数可变，以达到动态调整压缩能力；取整模块，用于将输出特征取整为整数以便后续编码；熵估计网络模块，用于对取整后的取整特征图中的各点进行概率估计，实现模型压缩的码率估计；以及熵编码模块，利用熵估计模型得到的概率分布，对取整特征图进行熵编码，降低输出特征的编码冗余，进一步降低图像压缩过程中的数据传输量。这些模块可以执行上述第一方面或第一方面任一项方法示例中的相应功能，具体参见方法示例中的详细描述，此处不做赘述。

第十方面，本发明涉及解码装置，具有实现上述第二方面或第二方面任一项的方法实施例中的行为的功能。所述功能可以通过硬件实现，也可以通过硬件执行相应的软件实现。所述硬件或软件包括一个或多个与上述功能相对应的模块。在一个可能的设计中，所述解码装置包括：熵估计网络模块，用于对解码特征图中的各点进行概率估计，实现模型压缩的码率估计；熵解码模块，利用熵估计模型得到的概率分布，对待解码码流进行熵解码得到解码特征

图，降低输出特征的编码冗余，进一步降低图像压缩过程中的数据传输量；解码网络模块，用于对压缩特征进行反变换，将解码特征图解析为音频或图像。这些模块可以执行上述第二方面或第二方面任一项方法示例中的相应功能，具体参见方法示例中的详细描述，此处不做赘述。

在现有端到端音频或图像编码主流方案中，存在着在固定模型的高码率下模型压缩质量趋于平稳的AE极限问题，虽然通过增加算力可以缓解AE极限问题，但是会导致高算力问题。本申请增加了特征空间中不同层(不同的特征层)的特征信号传输，且利用了不同层间特征信号的相关性，提高编码质量，突破AE极限。对于特征频率高的特征层所对应的编码或解码网络的算力小，降低了复杂度。根据实际的编码质量要求，可调节压缩特征的层数和通道数，兼容目前的多码率模型，可实现单模型多码率。

附图及以下说明中将详细描述一个或多个实施例。其它特征、目的和优点在说明、附图以及权利要求中是显而易见的。

附图说明

下面对本申请实施例用到的附图进行介绍。

图 1A 为用于实现本发明实施例的音视频/图像译码系统示例的框图，其中该系统利用神经网络来编码或解码视频或图像；

图 1B 为用于实现本发明实施例的音视频/图像译码系统另一示例的框图，其中该视频或图像编码器和/或视频或图像解码器使用神经网络来编码或解码视频或图像；

图 1C 为用于实现本发明实施例的音视频或图像译码设备示意性框图；

图 1D 为用于实现本发明实施例的音视频或图像译码装置示意性框图

图 2 为用于实现本发明实施例的编码器系统示例框图；

图 3A 为用于实现本发明实施例的音频或图像编码器的编码网络示例框架；

图 3B 为用于实现本发明实施例的音频或图像编码器的编码网络输入输出结构；

图 4 为用于实现本发明实施例的子编码网络示例结构图；

图 5 为编码网络示例图；

图 6A-6B 为熵估计单元示例框图；

图 7A-7M 为熵估计输入示例框图；

图 8 熵估计网络示例；

图 9 为用于实现本发明实施例的解码器系统示意框图；

图 10A-10B 解码熵估计单元框架示意图；

图 11 解码网络示意框图；

- 图 12 解码网络输入结构示意图；
- 图 13 为本申请实施例中解码网络结构示例图；
- 图 14A-14C 为本申请实施例结构示例；
- 图 15 为本申请实施例系统架构示例(无损)；
- 图 16 为本申请实施例系统架构示例；
- 图 17 为本申请实施例系统架构示例(云应用)；
- 图 18A-18B 为本申请实施例系统架构示例；和
- 图 19 为本申请实施例系统架构示例(音频)。

具体实施方式

本申请实施例所涉及的术语“第一”、“第二”等仅用于区分描述的目的，而不能理解为指示或暗示相对重要性，也不能理解为指示或暗示顺序。此外，术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含，例如，包含了一系列步骤或单元。方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

应当理解，在本申请中，“至少一个(项)”是指一个或者多个，“多个”是指两个或两个以上。“和/或”，用于描述关联对象的关联关系，表示可以存在三种关系，例如，“A和/或B”可以表示：只存在A，只存在B以及同时存在A和B三种情况，其中A，B可以是单数或者复数。字符“/”一般表示前后关联对象是一种“或”的关系。“以下至少一项(个)”或其类似表达，是指这些项中的任意组合，包括单项(个)或复数项(个)的任意组合。例如，a，b或c中的至少一项(个)，可以表示：a，b，c，“a和b”，“a和c”，“b和c”，或“a和b和c”，其中a，b，c可以是单个，也可以是多个。

本申请实施例提供一种基于AI的音视频或图像压缩技术，尤其是提供一种基于神经网络的音视频或图像压缩技术，具体提供一种基于端到端的音频或图像编解码系统。

在图像编码领域，术语“图像(picture)”、或“图片(image)”，可以用作同义词。图像编码(或通常称为编码)包括图像编码和图像解码两部分，其中视频由多个图像所组成，是连续图像的表达方式。音频编码(或通常称为编码)包括音频编码和音频解码两部分。音频或图像编码在源侧执行，通常包括处理(例如，压缩)原始视频图像以减少表示该音频或视频图像所需的数据量(从而更高效存储和/或传输)。音频或图像解码在目的地侧执行，通常包括相对于编码器作逆处理，以重建音频或图像。实施例涉及的音频或图像(或通常称为音频或图像)的“编码”应理解为音频或图像的“编码”或“解码”。编码部分和解码部分也合称为编解码(编码和解码，CODEC)。

在无损音频或图像编码情况下，可以重建原始音频或图像，即重建的音频或图像与原始音频或图像具有相同的质量（假设存储或传输期间没有传输损耗或其它数据丢失）。在传统有损图像编码情况下，通过量化等执行进一步压缩，来减少表示音频或视频图像所需的数据量，而解码器侧无法完全重建音频或视频图像，即重建的音频或视频图像的质量比原始音频或视频图像的质量较低或较差。

由于本申请实施例涉及大量神经网络的应用，为了便于理解，下面先对本申请实施例可能涉及的神经网络的相关术语和概念进行介绍。

(1) 神经网络

神经网络可以由神经元组成的，神经元可以是指以 x_s 和截距 1 为输入的运算单元，该运算单元的输出可以为：

$$h_{w,b}(x) = f(W^T x) = f(\sum_{s=1}^n W_s x_s + b) \quad (1-1)$$

其中， $s=1, 2, \dots, n$ ， n 为大于 1 的自然数， W_s 为 x_s 的权重， b 为神经单元的偏置。 f 为神经单元的激活函数（activation functions），用于将非线性特性引入神经网络中，来将神经元中的输入信号转换为输出信号。该激活函数的输出信号可以作为下一层卷积层的输入，激活函数可以是 sigmoid 函数。神经网络是将多个上述单一的神经元联结在一起形成的网络，即一个神经元的输出可以是另一个神经元的输入。每个神经元的输入可以与前一层的局部接受域相连，来提取局部接受域的特征，局部接受域可以由若干个神经元组成的区域。

(2) 深度神经网络

深度神经网络（deep neural network, DNN），也称多层神经网络，可以理解为具有多层隐含层的神经网络。按照不同层的位置对 DNN 进行划分，DNN 内部的神经网络可以分为三类：输入层，隐含层，输出层。一般来说第一层是输入层，最后一层是输出层，中间的层数都是隐含层。层与层之间是全连接的，也就是说，第 i 层的任意一个神经元一定与第 $i+1$ 层的任意一个神经元相连。

虽然 DNN 看起来很复杂，但是就每一层的工作来说，其实并不复杂，简单来说就是如下线性关系表达式： $\vec{y} = \alpha(W \vec{x} + \vec{b})$ ，其中， \vec{x} 是输入向量， \vec{y} 是输出向量， \vec{b} 是偏移向量， W 是权重矩阵（也称系数）， $\alpha(\cdot)$ 是激活函数。每一层仅仅是对输入向量 \vec{x} 经过如此简单的操作得到输出向量 \vec{y} 。由于 DNN 层数多，系数 W 和偏移向量 \vec{b} 的数量也比较多。这些参数在 DNN 中的定义如下所述：以系数 W 为例：假设在一个三层的 DNN 中，第二层的第 4 个神经元到第三层的第 2 个神经元的线性系数定义为 W_{24}^3 。上标 3 代表系数 W 所在的层数，而下标对应的是

输出的第三层索引 2 和输入的第二层索引 4。

综上，第 $L-1$ 层的第 k 个神经元到第 L 层的第 j 个神经元的系数定义为 W_{jk}^L 。

需要注意的是，输入层是没有 W 参数的。在深度神经网络中，更多的隐含层让网络更能够刻画现实世界中的复杂情形。理论上而言，参数越多的模型复杂度越高，“容量”也就越大，也就意味着它能完成更复杂的学习任务。训练深度神经网络的也就是学习权重矩阵的过程，其最终目的是得到训练好的深度神经网络的所有层的权重矩阵（由很多层的向量 W 形成的权重矩阵）。

(3) 卷积神经网络

卷积神经网络 (convolutional neuron network, CNN) 是一种带有卷积结构的深度神经网络。卷积神经网络包含了一个由卷积层和子采样层构成的特征抽取器，该特征抽取器可以看作是滤波器。卷积层是指卷积神经网络中对输入信号进行卷积处理的神经元层。在卷积神经网络的卷积层中，一个神经元可以只与部分邻层神经元连接。一个卷积层中，通常包含若干个特征平面，每个特征平面可以由一些矩形排列的神经单元组成。同一特征平面的神经单元共享权重，这里共享的权重就是卷积核。共享权重可以理解为提取图像信息的方式与位置无关。卷积核可以以随机大小的矩阵的形式初始化，在卷积神经网络的训练过程中卷积核可以通过学习得到合理的权重。另外，共享权重带来的直接好处是减少卷积神经网络各层之间的连接，同时又降低了过拟合的风险。

(4) 循环神经网络 (recurrent neural networks, RNN) 是用来处理序列数据的。在传统的神经网络模型中，是从输入层到隐含层再到输出层，层与层之间是全连接的，而对于每一层层内之间的各个节点是无连接的。这种普通的神经网络虽然解决了很多难题，但是却仍然对很多问题无能为力。例如，你要预测句子的下一个单词是什么，一般需要用到前面的单词，因为一个句子中前后单词并不是独立的。RNN 之所以称为循环神经网络，即一个序列当前的输出与前面的输出也有关。具体的表现形式为网络会对前面的信息进行记忆并应用于当前输出的计算中，即隐含层本层之间的节点不再无连接而是有连接的，并且隐含层的输入不仅包括输入层的输出还包括上一时刻隐含层的输出。理论上，RNN 能够对任何长度的序列数据进行处理。对于 RNN 的训练和对传统的 CNN 或 DNN 的训练一样。RNN 旨在让机器像人一样拥有记忆的能力。因此，RNN 的输出就需要依赖当前的输入信息和历史的记忆信息。

(5) 损失函数

在训练深度神经网络的过程中，因为希望深度神经网络的输出尽可能的接近真正想要预测的值，所以可以通过比较当前网络的预测值和真正想要的目标值，再根据两者之间的差异情况来更新每一层神经网络的权重向量（当然，在第一次更新之前通常会有初始化的过程，

即为深度神经网络中的各层预先配置参数），比如，如果网络的预测值高了，就调整权重向量让它预测低一些，不断地调整，直到深度神经网络能够预测出真正想要的目标值或与真正想要的目标值非常接近的值。因此，就需要预先定义“如何比较预测值和目标值之间的差异”，这便是损失函数（loss function）或目标函数（objective function），它们是用于衡量预测值和目标值的差异的重要方程。其中，以损失函数举例，损失函数的输出值（loss）越高表示差异越大，那么深度神经网络的训练就变成了尽可能缩小这个 loss 的过程。

(6) 反向传播算法

神经网络可以采用误差反向传播（back propagation, BP）算法在训练过程中修正初始的神经网络模型中参数的大小，使得神经网络模型的重建误差损失越来越小。具体地，前向传递输入信号直至输出会产生误差损失，通过反向传播误差损失信息来更新初始的神经网络模型中参数，从而使误差损失收敛。反向传播算法是以误差损失为主导的反向传播运动，旨在得到最优的神经网络模型的参数，例如权重矩阵。

(7) 生成式对抗网络

生成式对抗网络（generative adversarial networks, GAN）是一种深度学习模型。该模型中至少包括两个模块：一个模块是生成模型（Generative Model），另一个模块是判别模型（Discriminative Model），通过这两个模块互相博弈学习，从而产生更好的输出。生成模型和判别模型都可以是神经网络，具体可以是深度神经网络，或者卷积神经网络。GAN 的基本原理如下：以生成图片的 GAN 为例，假设有两个网络，G（Generator）和 D（Discriminator），其中 G 是一个生成图片的网络，它接收一个随机的噪声 z ，通过这个噪声生成图片，记做 $G(z)$ ；D 是一个判别网络，用于判别一张图片是不是“真实的”。它的输入参数是 x ， x 代表一张图片，输出 $D(x)$ 代表 x 为真实图片的概率，如果为 1，就代表 100% 是真实的图片，如果为 0，就代表不可能是真实的图片。在对该生成式对抗网络进行训练的过程中，生成网络 G 的目标就是尽可能生成真实的图片去欺骗判别网络 D，而判别网络 D 的目标就是尽量把 G 生成的图片和真实的图片区分开来。这样，G 和 D 就构成了一个动态的“博弈”过程，也即“生成式对抗网络”中的“对抗”。最后博弈的结果，在理想的状态下，G 可以生成足以“以假乱真”的图片 $G(z)$ ，而 D 难以判定 G 生成的图片究竟是不是真实的，即 $D(G(z)) = 0.5$ 。这样就得到了一个优异的生成模型 G，它可以用来生成图片。

在以下译码系统 10 的实施例 1 中，编码器 20 和解码器 30 根据图 1A 至图 13 进行描述。

图 1A 为示例性译码系统 10 的示意性框图，例如可以利用本申请技术的音视频或图像译码系统 10（或简称为译码系统 10）。音视频或图像译码系统 10 中的视频编码器 20（或简称为编码器 20）和音视频或图像解码器 30（或简称为解码器 30）代表可用于根据本申请中描述的各种示例执行各技术的设备等。

如图 1A 所示，译码系统 10 包括源设备 12，源设备 12 用于将编码音频或图像等编码音频或图像数据 21 提供给用于对编码音频或图像数据 21 进行解码的目的设备 14。

源设备 12 包括编码器 20，另外即可选地，可包括音频或图像源 16、音频或图像预处理器等预处理器（或预处理单元）18、通信接口（或通信单元）22。

音频或图像源 16 可包括或可以为任意类型的用于捕获现实世界图像等的音频或图像捕获设备，和/或任意类型的图像生成设备，例如用于生成计算机动画图像的计算机图形处理器或任意类型的用于获取和/或提供现实世界图像、计算机生成图像（例如，屏幕内容、虚拟现实（virtual reality, VR）图像和/或其任意组合（例如增强现实（augmented reality, AR）图像）的设备。所述音频或图像源可以为存储上述图像中的任意音频或图像的任意类型的内存或存储器。

为了区分预处理器（或预处理单元）18 执行的处理，音频或图像（或图像数据）17 也可称为原始音频或图像（或原始图像数据）17。

预处理器 18 用于接收（原始）音频或图像数据 17，并对音频或图像数据 17 进行预处理，得到预处理图像（或预处理图像数据）19。例如，预处理器 18 执行的预处理可包括修剪、颜色格式转换（例如从 RGB 转换为 YCbCr）、调色或去噪。可以理解的是，预处理单元 18 可以为可选组件。

音视频或图像编码器（或编码器）20 用于接收预处理音频或图像数据 19 并提供编码音频或图像数据 21（下面将根据图 2 等进一步描述）。

源设备 12 中的通信接口 22 可用于：接收编码音频或图像数据 21 并通过通信信道 13 向目的设备 14 等另一设备或任何其它设备发送编码音频或图像数据 21（或其它任意处理后的版本），以便存储或直接重建。

目的设备 14 包括解码器 30，另外即可选地，可包括通信接口（或通信单元）28、后处理器（或后处理单元）32 和显示设备 34。

目的设备 14 中的通信接口 28 用于直接从源设备 12 或从存储设备等任意其它源设备接收编码音频或图像数据 21（或其它任意处理后的版本），例如，存储设备为编码音频或图像数据存储设备，并将编码音频或图像数据 21 提供给解码器 30。

通信接口 22 和通信接口 28 可用于通过源设备 12 与目的设备 14 之间的直连通信链路，例如直接有线或无线连接等，或者通过任意类型的网络，例如有线网络、无线网络或其任意组合、任意类型的私网和公网或其任意类型的组合，发送或接收编码音频或图像数据（或编码数据）21。

例如，通信接口 22 可用于将编码音频或图像数据 21 封装为报文等合适的格式，和/或使用任意类型的传输编码或处理来处理所述编码后的音频或图像数据，以便在通信链路或通信

网络上进行传输。

通信接口 28 与通信接口 22 对应,例如,可用于接收传输数据,并使用任意类型的对应传输解码或处理和/或解封装对传输数据进行处理,得到编码音频或图像数据 21。

通信接口 22 和通信接口 28 均可配置为如图 1A 中从源设备 12 指向目的设备 14 的对应通信信道 13 的箭头所指示的单向通信接口,或双向通信接口,并且可用于发送和接收消息等,以建立连接,确认并交换与通信链路和/或例如编码后的图像数据传输等数据传输相关的任何其它信息,等等。

音视频或图像解码器(或解码器)30 用于接收编码音频或图像数据 21 并提供解码音频或图像数据(或解码音频或图像数据)31(下面将根据图 4 等进一步描述)。

后处理器 32 用于对解码后的音频或图像等解码音频或图像数据 31(也称为重建后的音频或图像数据)进行后处理,得到后处理后的音频或图像等后处理音频或图像数据 33。后处理单元 32 执行的后处理可以包括例如颜色格式转换(例如从 YCbCr 转换为 RGB)、调色、修剪或重采样,或者用于产生供显示设备 34 等显示的解码音频或图像数据 31 等任何其它处理。

显示设备 34 用于接收后处理音频或图像数据 33,以向用户或观看者等显示图像。显示设备 34 可以为或包括任意类型的用于表示重建后音频或图像的播放器或显示器,例如,集成或外部显示屏或显示器。例如,显示屏可包括液晶显示器(liquid crystal display, LCD)、有机发光二极管(organic light emitting diode, OLED)显示器、等离子显示器、投影仪、微型 LED 显示器、硅基液晶显示器(liquid crystal on silicon, LCoS)、数字光处理器(digital light processor, DLP)或任意类型的其它显示屏。

译码系统 10 还包括训练引擎 25,训练引擎 25 用于训练编码器 20 或解码器 30 以处理(例如处理输入音频或图像或图像区域或图像块以生成输入音频或图像或图像区域或图像块的预测值)。训练数据可以存入数据库(未示意)中,训练引擎 25 基于训练数据训练得到目标模型(例如:可以是用于音频或图像编码或解码的网络等)。需要说明的是,本申请实施例对于训练数据的来源不做限定,例如可以是云端或其他地方获取训练数据进行模型训练。

尽管图 1A 示出了源设备 12 和目的设备 14 作为独立的设备,但设备实施例也可以同时包括源设备 12 和目的设备 14 或同时包括源设备 12 和目的设备 14 的功能,即同时包括源设备 12 或对应功能和目的设备 14 或对应功能。在这些实施例中,源设备 12 或对应功能和目的设备 14 或对应功能可以使用相同硬件和/或软件或通过单独的硬件和/或软件或其任意组合来实现。

根据描述,图 1A 所示的源设备 12 和/或目的设备 14 中的不同单元或功能的存在和(准确)划分可能根据实际设备和应用而有所不同,这对技术人员来说是显而易见的。

编码器 20(例如音视频或图像编码器 20)或解码器 30(例如音视频或图像解码器 30)

或两者都可通过如图 1B 所示的处理电路实现，例如一个或多个微处理器、数字信号处理器（digital signal processor, DSP）、专用集成电路（application-specific integrated circuit, ASIC）、现场可编程门阵列（field-programmable gate array, FPGA）、离散逻辑、硬件、音视频或图像编码专用处理器或其任意组合。编码器 20 可以通过处理电路 46 实现，以包含参照图 2 编码器 20 论述的各种模块和/或本文描述的任何其它编码器系统或子系统。解码器 30 可以通过处理电路 46 实现，以包含参照图 4 解码器 30 论述的各种模块和/或本文描述的任何其它解码器系统或子系统。所述处理电路 46 可用于执行下文论述的各种操作。如图 7 所示，如果部分技术在软件中实施，则设备可以将软件的指令存储在合适的非瞬时性计算机可读存储介质中，并且使用一个或多个处理器在硬件中执行指令，从而执行本发明技术。音视频或图像编码器 20 和视频或图像解码器 30 中的其中一个可作为组合编解码器（encoder/decoder, CODEC）的一部分集成在单个设备中，如图 1B 所示。

源设备 12 和目的设备 14 可包括各种设备中的任一种，包括任意类型的手持设备或固定设备，例如，笔记本电脑或膝上型电脑、手机、智能手机、平板或平板电脑、相机、台式计算机、机顶盒、电视机、显示设备、数字媒体播放器、视频游戏控制台、视频流设备（例如，内容业务服务器或内容分发服务器）、广播接收设备、广播发射设备，等等，并可以不使用或使用任意类型的操作系统。在一些情况下，源设备 12 和目的设备 14 可配备用于无线通信的组件。因此，源设备 12 和目的设备 14 可以是无线通信设备。

在一些情况下，图 1A 所示的音视频或图像译码系统 10 仅仅是示例性的，本申请提供的技术可适用于音视频或图像编码设置（例如，音视频或图像编码或视频或图像解码），这些设置不一定包括编码设备与解码设备之间的任何数据通信。在其它示例中，数据从本地存储器中检索，通过网络发送，等等。音视频或图像编码设备可以对数据进行编码并将数据存储到存储器中，和/或音视频或图像解码设备可以从存储器中检索数据并对数据进行解码。在一些示例中，编码和解码由相互不通信而只是编码数据到存储器和/或从存储器中检索并解码数据的设备来执行。

图 1B 是根据一示例性实施例包含图 1A 的音视频或图像编码器 20 和/或图 1B 的音视频或图像解码器 30 的视频或图像译码系统 40 的实例的说明图。音视频或图像译码系统 40 可以包含成像设备 41、音视频或图像编码器 20、音视频或图像解码器 30（和/或藉由处理电路 46 实施的音视频或图像编/解码器）、天线 42、一个或多个处理器 43、一个或多个内存存储器 44 和/或显示设备 45。

如图 1B 所示，成像设备 41、天线 42、处理电路 46、音视频或图像编码器 20、音视频或图像解码器 30、处理器 43、内存存储器 44 和/或显示设备 45 能够互相通信。在不同实例中，音视频或图像译码系统 40 可以只包含音视频或图像编码器 20 或只包含音视频或图像解码器

30。

在一些实例中，天线 42 可以用于传输或接收音视频或图像数据的经编码比特流。另外，在一些实例中，显示设备 45 可以用于呈现音视频或图像数据。处理电路 46 可以包含专用集成电路 (application-specific integrated circuit, ASIC) 逻辑、图形处理器、通用处理器等。音视频或图像译码系统 40 也可以包含可选的处理器 43，该可选处理器 43 类似地可以包含专用集成电路 (application-specific integrated circuit, ASIC) 逻辑、图形处理器、通用处理器等。另外，内存存储器 44 可以是任何类型的存储器，例如易失性存储器（例如，静态随机存取存储器 (static random access memory, SRAM)、动态随机存储器 (dynamic random access memory, DRAM) 等）或非易失性存储器（例如，闪存等）等。在非限制性实例中，内存存储器 44 可以由超速缓存内存实施。在其它实例中，处理电路 46 可以包含存储器（例如，缓存等）用于实施图像缓冲器等。

在一些实例中，通过逻辑电路实施的音视频或图像编码器 20 可以包含（例如，通过处理电路 46 或内存存储器 44 实施的）音频或图像缓冲器和（例如，通过处理电路 46 实施的）音频或图形处理单元。音频或图形处理单元可以通信耦合至图像缓冲器。音频或图形处理单元可以包含通过处理电路 46 实施的音视频或图像编码器 20，以实施参照图 2 和/或本文中所描述的任何其它编码器系统或子系统所论述的各种模块。逻辑电路可以用于执行本文所论述的各种操作。

在一些实例中，音视频或图像解码器 30 可以以类似方式通过处理电路 46 实施，以实施参照图 1B 的音视频或图像解码器 30 和/或本文中所描述的任何其它解码器系统或子系统所论述的各种模块。在一些实例中，逻辑电路实施的音视频或图像解码器 30 可以包含（通过处理电路 46 或内存存储器 44 实施的）音频或图像缓冲器和（例如，通过处理电路 46 实施的）音频或图形处理单元。音频或图形处理单元可以通信耦合至图像缓冲器。音频或图形处理单元可以包含通过处理电路 46 实施的音视频或图像解码器 30，以实施参照图 4 和/或本文中所描述的任何其它解码器系统或子系统所论述的各种模块。

在一些实例中，天线 42 可以用于接收音视频或图像数据的经编码比特流。如所论述，经编码比特流可以包含本文所论述的与编码音频或视频帧相关的数据、指示符、索引值、模式选择数据等，例如与编码分割相关的数据（例如，变换系数或经量化变换系数，（如所论述的）可选指示符，和/或定义编码分割的数据）。音视频或图像译码系统 40 还可包含耦合至天线 42 并用于解码经编码比特流的视频或图像解码器 30。显示设备 45 用于呈现视频帧。

应理解，本申请实施例中对于参考音视频或图像编码器 20 所描述的实例，音视频或图像解码器 30 可以用于执行相反过程。关于信令语法元素，音视频或图像解码器 30 可以用于接收并解析这种语法元素，相应地解码相关视频数据。在一些例子中，音视频或图像编码器 20

可以将语法元素熵编码成经编码比特流。在此类实例中，音视频或图像解码器 30 可以解析这种语法元素，并相应地解码相关音视频或图像数据。

图 1C 为本发明实施例提供的音视频或图像译码设备 400 的示意图。音视频或图像译码设备 400 适用于实现本文描述的公开实施例。在一个实施例中，音视频或图像译码设备 400 可以是解码器，例如图 1A 中的音视频或图像解码器 30，也可以是编码器，例如图 1A 中的音视频或图像编码器 20。

音视频或图像译码设备 400 包括：用于接收数据的入端口 410（或输入端口 410）和接收单元（receiver unit, Rx）420；用于处理数据的处理器、逻辑单元或中央处理器（central processing unit, CPU）430；例如，这里的处理器 430 可以是神经网络处理器 430；用于传输数据的发送单元（transmitter unit, Tx）440 和出端口 450（或输出端口 450）；用于存储数据的存储器 460。音视频或图像译码设备 400 还可包括耦合到入端口 410、接收单元 420、发送单元 440 和出端口 450 的光电（optical-to-electrical, OE）组件和电光（electrical-to-optical, EO）组件，用于光信号或电信号的出口或入口。

处理器 430 通过硬件和软件实现。处理器 430 可实现为一个或多个处理器芯片、核（例如，多核处理器）、FPGA、ASIC 和 DSP。处理器 430 与入端口 410、接收单元 420、发送单元 440、出端口 450 和存储器 460 通信。处理器 430 包括译码模块 470（例如，基于神经网络 NN 的译码模块 470）。译码模块 470 实施上文所公开的实施例。例如，译码模块 470 执行、处理、准备或提供各种编码操作。因此，通过译码模块 470 为音视频或图像译码设备 400 的功能提供了实质性的改进，并且影响了音视频或图像译码设备 400 到不同状态的切换。或者，以存储在存储器 460 中并由处理器 430 执行的指令来实现译码模块 470。

存储器 460 包括一个或多个磁盘、磁带机和固态硬盘，可以用作溢出数据存储设备，用于在选择执行程序时存储此类程序，并且存储在程序执行过程中读取的指令和数据。存储器 460 可以是易失性和/或非易失性的，可以是只读存储器（read-only memory, ROM）、随机存取存储器（random access memory, RAM）、三态内容寻址存储器（ternary content-addressable memory, TCAM）和/或静态随机存取存储器（static random-access memory, SRAM）。

图 1D 为示例性实施例提供的装置 500 的简化框图，装置 500 可用作图 1A 中的源设备 12 和目的设备 14 中的任一个或两个。

装置 500 中的处理器 502 可以是中央处理器。或者，处理器 502 可以是现有的或今后将研发出的能够操控或处理信息的任何其它类型设备或多个设备。虽然可以使用如图所示的处理器 502 等单个处理器来实施已公开的实现方式，但使用一个以上的处理器速度更快和效率更高。

在一种实现方式中，装置 500 中的存储器 504 可以是只读存储器 (ROM) 设备或随机存取存储器 (RAM) 设备。任何其它合适类型的存储设备都可以用作存储器 504。存储器 504 可以包括处理器 502 通过总线 512 访问的代码和数据 506。存储器 504 还可包括操作系统 508 和应用程序 510，应用程序 510 包括允许处理器 502 执行本文所述方法的至少一个程序。例如，应用程序 510 可以包括应用 1 至 N，还包括执行本文所述方法的音视频或图像译码应用。

装置 500 还可以包括一个或多个输出设备，例如显示器 518。在一个示例中，显示器 518 可以是将显示器与可用于感测触摸输入的触敏元件组合的触敏显示器。显示器 518 可以通过总线 512 耦合到处理器 502。

虽然装置 500 中的总线 512 在本文中描述为单个总线，但是总线 512 可以包括多个总线。此外，辅助存储器可以直接耦合到装置 500 的其它组件或通过网络访问，并且可以包括存储卡等单个集成单元或多个存储卡等多个单元。因此，装置 500 可以具有各种各样的配置。

编码器和编码方法

图 2 为用于实现本申请技术的一种音视频或图像编码器 20 的示例的示意性框图。在图 2 的示例中，音视频或图像编码器 20 包括输入端 (或输入接口) 202、编码网络 204、取整 206、熵编码 208、熵估计 210 和输出端 (或输出接口) 212。图 2 所示的音视频编码器 20 也可称端到端音频或图像编码器或基于端到端音频或图像编解码器的音频或图像编码器。

编码器 20 可用于通过输入端 202 等接收音频或图像 (或音频或图像数据) 17，例如，形成音视频或视频序列的音频或图像。接收的音频或图像、音频或图像数据也可以是预处理后的音频或图像 (或预处理后的音频或图像数据) 19。为简单起见，以下描述使用音频或图像 17。音频或图像 17 也可称为当前音频或图像或待编码的音频或图像 (尤其是在音视频编码中将当前音频或图像与其它音频区分开时，其它音频或图像例如同一音频或视频序列，即也包括当前音频或图像的音频或视频序列中的之前编码后音频或图像和/或解码后音频或图像)。

(数字) 图像为或可以视为具有强度值的像素点组成的二维阵列或矩阵。阵列中的像素点也可以称为像素 (pixel 或 pel) (图像元素的简称)。阵列或图像在水平方向和垂直方向 (或轴线) 上的像素点数量决定了图像的大小和/或分辨率。为了表示颜色，通常采用三个颜色分量，即图像可以表示为或包括三个像素点阵列。在 RGB 格式或颜色空间中，图像包括对应的红色、绿色和蓝色像素点阵列。但是，在视频或图像编码中，每个像素通常以亮度/色度格式或颜色空间表示，例如 YCbCr，包括 Y 指示的亮度分量 (有时也用 L 表示) 以及 Cb、Cr 表示的两个色度分量。亮度 (luma) 分量 Y 表示亮度或灰度水平强度 (例如，在灰度等级图像中两者相同)，而两个色度 (chrominance，简称为 chroma) 分量 Cb 和 Cr 表示色度或颜色信息分量。相应地，YCbCr 格式的图像包括亮度像素点值 (Y) 的亮度像素点阵列和色度值 (Cb 和 Cr) 的两个色度像素点阵列。RGB 格式的图像可以转换或变换为 YCbCr 格式，反之亦

然，该过程也称为颜色变换或转换。如果图像是黑白的，则该图像可以只包括亮度像素点阵列。相应地，图像可以为例如单色格式的亮度像素点阵列或 4:2:0、4:2:2 和 4:4:4 彩色格式的亮度像素点阵列和两个相应的色度像素点阵列。

在一个可能性中，视频或图像编码器 20 的实施例可包括图像分割单元（图 1A 或图 1B 中未示出），用于将图像 17 分割成多个（通常不重叠）图像块 203。这些块在 H.265/HEVC 和 VVC 标准中也可以称为根块、宏块（H.264/AVC）或编码树块（Coding Tree Block, CTB），或编码树单元（Coding Tree Unit, CTU）。分割单元可用于对视频序列中的所有图像使用相同的块大小和使用限定块大小的对应网格，或在图像或图像子集或图像组之间改变块大小，并将每个图像分割成对应块。

在另一个可能性中，视频或图像编码器可用于直接接收图像 17 的块 203，例如，组成所述图像 17 的一个、几个或所有块。图像块 203 也可以称为当前图像块或待编码图像块。

与图像 17 一样，图像块 203 同样是或可认为是具有强度值（像素点值）的像素点组成的二维阵列或矩阵，但是图像块 203 的比图像 17 的小。换句话说，块 203 可包括一个像素点阵列（例如，单色图像 17 情况下的亮度阵列或彩色图像情况下的亮度阵列或色度阵列）或三个像素点阵列（例如，彩色图像 17 情况下的一个亮度阵列和两个色度阵列）或根据所采用的颜色格式的任何其它数量和/或类型的阵列。块 203 的水平方向和垂直方向（或轴线）上的像素点数量限定了块 203 的大小。相应地，块可以为 $M \times N$ （ M 列 \times N 行）个像素点阵列，或 $M \times N$ 个变换系数阵列等。

在另一个可能性中，图 1A 或图 1B 或图 2 所示的音视频或图像编码器 20 用于逐块对图像 17 进行编码，例如，对每个块 203 执行编码、取整和熵编码。

在另一个可能性中，图 1A 或图 1B 或图 2 所示的音视频或图像编码器 20 用于对图像 17 进行编码，例如，对图像 17 执行编码、量化和熵编码。

在另一个可能性中，图 1A 或图 1B 或图 2 所示的音视频或图像编码器 20 用于对音频数据 17 进行编码，例如，对音频数据 17 执行编码、量化和熵编码。

在另一个可能性中，图 1A 或图 1B 或图 2 所示的音视频或图像编码器 20 还可以用于使用片（也称为视频片）分割编码图像，其中图像可以使用一个或多个片（通常为不重叠的）进行分割或编码。每个片可包括一个或多个块（例如，编码树单元 CTU）或一个或多个块组（例如 H.265/HEVC/VVC 标准中的编码区块（tile）和 VVC 标准中的子图像（subpicture））。

在另一个可能性中，图 1A 或图 1B 或图 2 所示的音视频或图像编码器 20 还可以用于使用片段分割音频，其中音频可以使用一个或多个片段（通常为不重叠的）进行分割或编码。

在另一个可能性中，图 1A 或图 1B 或图 2 所示的音视频或图像编码器 20 还可以用于使用片/编码区块组（也称为视频编码区块组）和/或编码区块（也称为视频编码区块）对图像进

行分割和/或编码，其中图像可以使用一个或多个片/编码区块组（通常为不重叠的）进行分割或编码，每个片/编码区块组可包括一个或多个块（例如 CTU）或一个或多个编码区块等，其中每个编码区块可以为矩形等形状，可包括一个或多个完整或部分块（例如 CTU）。

编码网络单元 204

如图 2，编码网络单元 204 用于通过编码网络根据输入数据 203 来得到各特征层的输出特征图 205。其中编码网络单元输出大于等于两个特征层所对应的至少两个输出特征图 205。

在一个可能性中，编码网络单元如图 3A 所示，编码网络单元包含了 K 个子编码网络，每个子编码网络对应了所在的特征层以及对应的输出特征图。此时有 K 个输出特征图输出，其中 $K \geq 2$ 。

在一个可能性中，编码网络输入为待编码音频或图像或待编码音频片段或图像块。

在一个可能性中，编码网络 204 中的编码网络的输入输出结构如图 3B 所示，编码网络单元由 T 个网络层所构成，其中 M、L、T、K 为正整数。第 M 输出特征图与第 K 输出特征图均为编码网络的输出，且编码网络输出的第 K 输出特征图为编码网络单元的网络层 L 后输出，第 M 输出特征图在网络层 T 后输出。可以理解，可以有多个输出特征图在编码网络单元中的不同网络层位置后输出，此处不做限定。

其中图 3B 中的任意一网络层可以为卷积层、归一化层、非线性激活层等。

在一个可能性中，编码网络 204 网络中图 3A 所对应的各子编码网络结构可以互为相同或者不同，第 M 和第 N 子编码网络的网络结构如图 4 所示。其中第 M 子编码网络的网络层 1 和第 N 子编码网络中的任意一网络层结构可以互不相同，且 L 和 P 可以相同或不同，L 和 P 均为大于 0 的正整数。

在一个可能性中，编码网络 204 网络中的任意一子编码网络的网络结构示例如图 5 所示，可见示例中第 N 子编码网络包含了 5 个网络层，具体包括了一个归一化层、两个卷积层以及两个非线性激活层。

取整单元 206

取整单元 206 用于通过例如标量量化或矢量量化对输出特征图 205 进行取整，得到取整特征图 207。

在一个可能性中，音视频或图像编码器 20（对应地，取整单元 206）可用于输出取整参数（quantization parameter, QP），例如，直接输出或由熵编码单元 270 进行编码或压缩后输出，例如使得音视频解码器 30 可接收并使用量化参数进行解码。

如图 6A 所示，对 N 个特征层的输出特征图均进行取整操作，输出 N 个取整特征图，其中 N 为大于 1 的正整数。

在一个可能性中，输出特征图在进行取整前进行预处理，包括输出特征图中的像素点的

像素值与熵估计输出的均值相减得到偏移特征图,对获得的偏移特征图取整得到取整特征图。

熵编码单元 208

熵编码单元 208 用于将熵编码算法或方案(例如,可变长度编码(variable length coding, VLC)方案、上下文自适应 VLC 方案(context adaptive VLC, CALVC)、算术编码方案、二值化算法、上下文自适应二进制算术编码(context adaptive binary arithmetic coding, CABAC)、基于语法的上下文自适应二进制算术编码(syntax-based context-adaptive binary arithmetic coding, SBAC)、概率区间分割熵(probability interval partitioning entropy, PIPE)编码或其它熵编码方法或技术)应用于至少两个特征层中的取整特征图,得到可以通过输出端 212 以编码比特流 21 等形式输出的编码音频或图像数据 21,使得视频解码器 30 等可以接收并使用用于解码的参数。可将编码比特流 21 传输到视频解码器 30,或将其保存在存储器中稍后由视频解码器 30 传输或检索。

在一个可能性中,熵编码由于不知道取整特征图的真实字符概率,所以需要可以统计这些或相关的信息添加至熵编码模块中,可以把这些信息传到解码端。

熵估计单元 210

如图 6A 所示,熵估计单元 210 用于对取整特征图进行概率估计,以实现模型压缩的码率估计。其中虚线输入为可选输入。熵估计单元 210 对应了 N 个特征层所对应的 N 个熵估计模块。

如图 6B 所示,熵估计单元 210 用于对取整特征图进行概率估计,以实现模型压缩的码率估计。其中虚线输入为可选输入。熵估计单元 210 对应了 N 个特征层所对应的 N 个熵估计模块以及残差层所对应的无损熵估计模块。

在一种可能性中,如图 7A 所示,编码器中的第 P 熵估计结果作为第 L 熵估计的输入,其中 L、P 为正整数且 $L < P$ 。

在另一种可能性中,编码器中的多个熵估计结果作为第 L 熵估计的输入,且多个熵估计中每个熵估计模块所对应的特征层序号均大于 L,其中 L 为正整数。如图 7B 所示为例, L/I/T/K 均为正整数且 $L < I, L < T, L < K$ 。

在一种可能性中,如图 7C 所示,编码器中的第 P 取整特征图作为第 L 熵估计的输入,其中 L、P 为正整数且 $L < P$ 。

在另一种可能性中,编码器中的多个取整特征图作为第 L 熵估计的输入,且多个取整模块中每个取整模块所对应的特征层序号均大于 L,其中 L 为正整数。如图 7D 所示为例, L/I/T/K 均为正整数且 $L < I, L < T, L < K$ 。

在一种可能性中,编码器中的至少一个取整特征图和至少一个熵估计结果作为第 L 熵估计的输入,其中 L 为正整数。其中作为第 L 熵估计的输入的取整特征图所对应的特征层序号

大于等于 L ，作为第 L 熵估计的输入的熵估计结果所对应的特征层序号均大于 L 。如图 7E 所示为例， $L/I/T/K$ 均为正整数且 $L \leq I, L \leq T, L \leq K$ 。

在一种可能性中，如图 7F 所示，编码器中的第 P 熵估计结果作为无损熵估计的输入，其 P 为正整数。

在另一种可能性中，编码器中的多个熵估计结果作为无损熵估计的输入。如图 7G 所示为例， $I/T/K$ 均为正整数。

在一种可能性中，如图 7H 所示，编码器中的第 P 取整特征图作为无损熵估计的输入，其 P 为正整数。

在另一种可能性中，编码器中的多个取整特征图作为无损熵估计的输入，如图 7I 所示为例， $I/T/K$ 均为正整数。

在一种可能性中，编码器中的至少一个取整特征图和至少一个熵估计结果作为无损熵估计的输入。如图 7J 所示为例， $I/T/K$ 均为正整数。

在一种可能性中，编码器中的至少一个取整特征图和重构音频或图像作为无损熵估计的输入。如图 7K 所示为例， I/T 均为正整数。

在一种可能性中，编码器中的至少一个熵估计结果和重构音频或图像作为无损熵估计的输入。如图 7L 所示为例， I/T 均为正整数。

在一种可能性中，编码器中的至少一个熵估计结果和重构音频或图像以及至少一个取整特征图作为无损熵估计的输入。如图 7M 所示为例， $I/T/P$ 均为正整数。

在一种可能性中，熵估计网络为卷积网络，卷积网络中包括了卷积层和非线性激活层。，可见图 8 示例中熵估计网络中包含了 5 个网络层，具体包括了三个卷积层以及两个非线性激活层。

解码器和解码方法

图 1A-1B 示出了用于实现本申请技术的示例性音视频或图像解码器 30。如图 9 所示，音视频或图像解码器 30 用于接收例如由编码器 20 编码的编码音频或图像数据 21（例如编码比特流 21），得到解码音频或图像 308。编码图像数据或比特流包括用于解码所述编码图像数据的信息，例如表示编码视频片（和/或编码区块组或编码区块）的图像块的数据和相关的语法元素。在图 9 的示例中，解码器 30 包括熵估计单元 304、熵解码单元 306 和解码网络单元 308。

熵估计单元 304

如图 10A 所示，熵估计单元 304 用于对解码特征图进行概率估计。其中虚线输入为可选输入。熵估计单元 304 对应了 N 个特征层所对应的 N 个熵估计模块。

如图 10B 所示，熵估计单元 304 用于对解码特征图进行概率估计，以实现模型压缩的码

率估计。其中虚线输入为可选输入。熵估计单元 304 对应了 N 个特征层所对应的 N 个熵估计模块以及残差层所对应的无损熵估计模块。

在一种可能性中，如图 7A 所示，解码器中的第 P 熵估计结果作为第 L 熵估计的输入，其中 L 、 P 为正整数且 $L < P$ 。

在另一种可能性中，解码器中的多个熵估计结果作为第 L 熵估计的输入，且多个熵估计中每个熵估计模块所对应的特征层序号均大于 L ，其中 L 为正整数。如图 7B 所示为例， $L/I/T/K$ 均为正整数且 $L < I, L < T, L < K$ 。

在一种可能性中，解码器中的第 P 解码特征图作为第 L 熵估计的输入，其中 L 、 P 为正整数且 $L \leq P$ 。

在另一种可能性中，解码器中的多个解码特征图作为第 L 熵估计的输入，且多个熵解码模块所对应的特征层序号均大于 L ，其中 L 为正整数。

在一种可能性中，解码器中的至少一个解码特征图和至少一个熵估计结果作为第 L 熵估计的输入，其中 L 为正整数。其中作为第 L 熵估计的输入的解码特征图所对应的特征层序号大于等于 L ，作为第 L 熵估计的输入的熵估计结果所对应的特征层序号均大于 L 。

在一种可能性中，如图 7F 所示，解码器中的第 P 熵估计结果作为无损熵估计的输入，其 P 为正整数。

在另一种可能性中，解码器中的多个熵估计结果作为无损熵估计的输入。如图 7G 所示为例， $I/T/K$ 均为正整数。

在一种可能性中，解码器中的第 P 解码特征图作为无损熵估计的输入，其中 P 为正整数。

在另一种可能性中，解码器中的多个解码特征图作为无损熵估计的输入。

在一种可能性中，解码器中的至少一个解码特征图和至少一个熵估计结果作为无损熵估计的输入。

在一种可能性中，解码器中的至少一个解码特征图和重构音频或图像作为无损熵估计的输入。

在一种可能性中，解码器中的至少一个熵估计结果和重构音频或图像作为无损熵估计的输入。如图 7K 所示为例， I/T 均为正整数。

在一种可能性中，解码器中的至少一个熵估计结果和重构音频或图像以及至少一个解码特征图作为无损熵估计的输入。

在一种可能性中，熵估计为卷积网络，卷积网络中包括了卷积层和非线性激活层。，可见图 8 示例中熵估计网络中包含了 5 个网络层，具体包括了三个卷积层以及两个非线性激活层。

熵解码单元 306

熵解码单元 306 用于解析比特流 21 (或一般为编码图像数据 21) 并对编码音频或图像数据 21 执行熵解码, 得到解码特征图 307 和/或解码后的编码参数 (图 11 中未示出) 等。熵解码单元 304 可用于应用编码器 20 的熵编码单元 208 的编码方案对应的解码算法或方案。

在一个可能性中, 如图 10A 所示, 熵解码单元应用于至少两个特征层中所对应的待解码码流。

在一个可能性中, 如图 10B 所示, 熵解码单元应用于残差层中所对应的无损待解码码流。

在一个可能性中, 熵编码由于不知道取整特征图真实的字符概率, 所以需要可以统计这些或相关的信息添加至熵编码模块中, 可以把这些信息传到解码端。

在一个可能性中, 解码器 30 还可以用于对图像 21 或图像 21 中的图像块执行熵解码。

在一个可能性中, 解码器 30 还可以用于对音频 21 或音频 21 中的音频片段执行熵解码。

解码网络单元

解码网络 306 用于将解码特征图 307 通过解码网络 308 以在像素域中得到重构音频或图像 310。

在一个可能性中, 如图 11 所示为 N 个解码特征图输入解码网络的示意图, 其中 N 大于 1。

在一个可能性中, 解码特征图在输入解码网络前进行预处理, 包括输出特征图中的像素点的像素值与熵估计输出的均值相减得到偏移特征图, 对解码特征图中的像素点的像素值与熵估计输出的均值相加得到更新的解码特征图, 再将更新的解码特征图输入解码网络。

在一个可能性中, 解码网络 306 中的输入输出结构如图 12 所示, 解码网络 306 由 P 个网络层所构成, 其中 $P > L$ 且 M、L、T、P 均为正整数。其中第 T 解码特征图和解码网络 306 的网络层 L 后输入。第 T 解码特征图与第 M 解码特征图均为解码网络的输入, 可以理解, 可以有多个解码特征图在解码网络单元中的不同网络层位置后作为输入, 此处不做限定。

图 12 中的任意一网络层可以为卷积层、归一化层、非线性激活层等。

在一个可能性中, 解码网络 306 中可以存在包括叠加(concat)、相加、相减等操作。

在一个可能性中, 解码网络 306 中各网络层结构可以互为相同或者不同。

在一个可能性中, 解码网络 306 的结构示例如图 13 所示, 可见示例中解码网络包含了 12 个网络层, 具体包括了一个反归一化层、六个卷积层以及四个非线性激活层。其中 m, n, t 均为正整数, 解码网络输出为解码图像或重构音频或图像。第 t 解码特征图、第 n 解码特征图、第 m 解码特征图均为解码网络的输入, 其中 t、n、m 均为整数。

应理解, 在编码器 20 和解码器 30 中, 可以对当前步骤的处理结果进一步处理, 然后输出到下一步骤。例如, 在编码器单元或解码器单元之后, 可以对编码器单元或解码器单元的处理结果进行进一步的运算或处理, 例如裁剪(clip)或移位(shift)运算或滤波处理。

尽管上述实施例主要描述了音视频编解码, 但应注意的是, 译码系统 10、编码器 20 和

解码器 30 的实施例以及本文描述的其它实施例也可以用于静止图像处理或编解码,即音视频编解码中独立于任何先前或连续图像的单个图像的处理或编解码。音视频编码器 20 和音视频解码器 30 的所有其它功能(也称为工具或技术)同样可用于静态图像处理。

图 14A 示出了本申请的一种可能的系统架构,模型运行步骤如下:

编码端:

步骤 1: 输入数据通过编码网络输出了 N 个输出特征图;

步骤 2: N 个输出特征图经过取整操作得到 N 个取整特征图;

步骤 3: 结合熵估计依次(第 N 至第一顺序)估计各取整特征图的概率并对 N 个取整特征图进行熵编码,得到 N 个编码码流。

其中第 M 取整特征图可以作为第 L 熵估计的输入,所述 L, M, N 均为正整数且 $L \geq 1, L < M \leq N$;

其中第 P 熵估计可以作为第 T 熵估计的输入,所述 T, P, N 均为正整数且 $T \geq 1, T < P \leq N$;

解码端:

步骤 1: 获得 N 个待解码码流,结合熵估计依次(第 N 至第一顺序)进行熵解码获得 N 个解码特征图;

其中第 M 取整特征图可以作为第 L 熵估计的输入,所述 L, M, N 均为正整数且 $L \geq 1, L < M \leq N$;

其中第 P 熵估计可以作为第 T 熵估计的输入,所述 T, P, N 均为正整数且 $T \geq 1, T < P \leq N$;

步骤 2: N 个输出特征图经过解码网络得到输出数据。

图 14B 和图 14C 示出了本申请实施例提供的一种可能的场景下的系统架构 700,模型运行步骤如下:

编码端:

步骤 1: 编码网络模块包括了四个特征层,通过四个子编码网络提取输入图像的四个输出特征图,其中每个编码网络的网络深度、特征层数、每层特征的通道数可变,以达到动态改变模型压缩能力的作用;

步骤 2: 如图 14B 所示,取整模块将第一至第四输出特征图取整为第一至第四取整特征图,对应了图 14C 中的 y_1, y_2, y_3, y_4 ;

步骤 3: 熵估计模块依次估计压缩特征的概率,如图 14B 所示,第四取整特征图作为第三熵估计的输入,第三取整特征图作为第二熵估计的输入,第二取整特征图作为第一熵估计的输入。对应了图 14C 所示,用已知概率估计取整特征图 y_4 的概率,用取整特征图 y_4 估计取整特征图 y_3 的概率,用取整特征图 y_3 估计 y_2 的概率,用取整特征图 y_2 估计取整特征图 y_1 的概率;

步骤 4: 如图 14B 所示,熵编码模块按照从深层到浅层的顺序,根据第四熵估计结果对第四取整特征图进行熵编码,根据第三熵估计结果对第三取整特征图进行熵编码,根据第二

熵估计结果对第二取整特征图进行熵编码，根据第一熵估计结果对第一取整特征图进行熵编码。对应了图 14C 依次根据取整特征图 y_4, y_3, y_2, y_1 的概率对取整特征图 y_4, y_3, y_2, y_1 进行熵编码生成，用来网络传输或者保存成本地文件。

解码端：

步骤 1：如图 14B 所示，熵估计网络模块根据第四熵估计结果对第四编码码流或第四待解码码流进行解码得到第四解码特征图，对应了图 14B 中根据已知概率估计 y_4 概率，熵解码模块解码出 y_4 ；

步骤 2：如图 14C 所示，第三熵估计网络结合第四解码特征图，并对第三编码码流或第三待解码码流进行解码得到第三解码特征图，对应了图 14B 中用 y_4 估计 y_3 的概率，熵解码模块解码出 y_3 ；

步骤 3：如图 14B 所示，第二熵估计网络结合第三解码特征图，并对第二编码码流或第二待解码码流进行解码得到第二解码特征图，对应了图 14B 中用 y_3 估计 y_2 的概率，熵解码模块解码出 y_2 ；

步骤 4：如图 14C 所示，第一熵估计网络根据第二解码特征图，并对第一编码码流或第一待解码码流进行解码得到第一解码特征图，对应了图 14B 中用 y_2 估计 y_1 的概率，熵解码模块解码出 y_1 ；

步骤 5：解码网络模块根据解码特征图 y_4, y_3, y_2, y_1 得到重构音频或图像。

该实施例取 $N = 128$ ，不同 bpp 情况下压缩不同层的特征，具体如下

$$bpp \leq 0.5, \text{压缩 } y_4, y_3$$

$$0.5 < bpp \leq 1, \text{压缩 } y_4, y_3, y_2$$

$$bpp > 1, \text{压缩 } y_4, y_3, y_2, y_1$$

图 15 示出了本申请实施例提供的另一种可能的场景下的系统架构 800，该实施例是一个无损压缩框架。

模型运行步骤如下：

编码端：

步骤 1：编码网络模块分特征层提取输入数据 x_0 每一特征层的输出特征图；

步骤 2：取整模块将图像各个特征层的输出特征图取整为取整特征图；

步骤 3：解码网络根据各特征层的解码特征图得到重构音频或图像 x_1 ；

步骤 4：编码网络获得输入数据的残差 $x_0 - x_1$ ；

步骤 5：熵估计网络模块按照从底层到浅层(特征层 4 至特征层 1，最后残差层)的顺序依次进行熵估计获得各层的概率估计结果；

步骤 6: 熵编码模块依次(从底层到浅层)进行熵编码, 各层生成的码流用来网络传输或者保存成本地文件。

解码端:

步骤 1: 熵估计网络和熵解码模块从底向上(特征层 4 至特征层 1, 最后残差层)依次解码出各层的解码特征图和残差;

步骤 2: 解码网络模块根据各解码特征图得到重构音频或图像 x_1 ;

步骤 3: 残差和重构音频或图像 x_1 叠加得到输出数据或无损重构音频或图像。

在再一种实施例中, 图 16 示出了一种可能的场景下的系统架构 900。压缩解码运行步骤如下:

步骤 1: AI 编码单元模块用于将图像数据变换成冗余度更低的输出特征, 且产生输出特征中各点的概率估计, 其中 AI 编码单元模块包括了编码网络单元, 取整单元以及熵估计单元;

通过摄像头等图像采集设备采集图像后可以以多种的图像表达方式, 如 RAW 域, YUV 或 RGB 表征方式作为 AI 编码单元的输入。AI 编码单元模块中的数据处理过程可以在如神经网络处理器(NPU)或者图形处理器(GPU)设备上实现。

步骤 2: 熵编码/解码模块(以算术编码为例)用于降低输出特征的编码冗余, 进一步降低图像压缩过程中的数据传输量;

熵编码/解码模块的数据处理过程可以在中央处理器(CPU)上进行实现。

步骤 3: AI 解码单元模块用于对输出特征进行反变换, 将输出特征解析为图像。其中 AI 解码单元模块包括了解码网络单元以及熵估计单元;

AI 编码单元模块中的数据处理过程可以在如神经网络处理器(NPU)或者图形处理器(GPU)设备上实现。

步骤 4: 保存文件模块用于将熵编码模块生成的数据文件保存至终端的对应存储位置;

步骤 5: 加载文件模块用于从终端的对应存储位置加载数据文件, 并输入到熵解码模块。

在再一种实施例中, 图 17 示出了一种可能的场景下的系统架构 1000, 该实施例是云应用设备结构图, 包括:

步骤 1: 本地与云侧的 JPEG 编/解码模块用于在端侧或者云侧对数字图像进行压缩/解压, 以降低本地与云侧间数据传输的数据量;

步骤 2: AI 编码单元模块用于将图像变换成冗余度更低的输出特征, 且产生输出特征中各点的概率估计。其中 AI 解码单元模块包括了解码网络单元以及熵估计单元;

步骤 3: 熵编码/解码模块(以算术编码为例)用于降低输出特征的编码冗余, 进一步降低图像压缩过程中的数据传输量;

步骤 4: AI 解码单元模块用于对输出特征进行反变换, 将输出特征解析为图像。其中 AI

解码单元模块包括了解码网络单元以及熵估计单元；

步骤 5: 保存文件模块用于将熵编码（以算术编码为例）模块生成的数据文件如编码码流或待解码码流保存至云侧的对应存储位置；

步骤 6: 加载文件模块用于从云侧的对应存储位置加载数据文件，并输入到熵解码单元。

图 18A 和图 18B 示出了本申请提供的一种实施例，模型运行步骤如下：

编码端：

步骤 1: 编码网络模块包括了四个特征层，通过四个子编码网络提取输入数据的四个输出特征图，对应了图 18B 中的 y_1, y_2, y_3, y_4 。其中每个编码网络的网络深度、特征层数、每层特征的通道数可变，以达到动态改变模型压缩能力的作用；

步骤 2: 如图 18A 所示，将第一至第三输出特征图进行预处理得到第一至第三偏移特征图，预处理操作对应了图 17B 中减法操作，将输出特征图的各像素点的像素值与熵估计输出的均值 μ 相减得到偏移特征图；

步骤 3: 如图 18A 所示，取整模块将第一至第三偏移特征图与第四输出特征图取整为第一至第四取整特征图；

步骤 4: 熵估计模块依次估计压缩特征的概率，如图 18A 和 18B 所示，用已知概率估计第四取整特征图的概率，第四取整特征图估计第三取整特征图的概率，第三取整特征图进行反处理后作为第二熵估计的输入来估计第二取整特征图的概率，第二取整特征图进行反处理后作为第一熵估计的输入来估计第一取整特征图的概率。

步骤 5: 如图 18A 所示，熵编码模块按照从深层到浅层的顺序，根据第四熵估计结果对第四取整特征图进行熵编码。根据第三熵估计结果如方差信息 σ_3 ，对第三取整特征图进行熵编码。根据第二熵估计结果如方差信息 σ_2 对第二取整特征图进行熵编码。根据第一熵估计结果如方差信息 σ_1 对第一取整特征图进行熵编码。对应了图 18B 依次根据各取整特征图的概率对第四至第一取整特征图进行熵编码生成，用来网络传输或者保存成本地文件。

编码端的反处理操作是将取整特征图的各像素点的像素值与熵估计输出的均值相加

解码端：

步骤 1: 如图 18A 所示，熵估计网络模块根据第四熵估计结果对第四编码码流或第四待解码码流进行熵解码得到第四解码特征图；

步骤 2: 如图 18A 所示，第三熵估计网络根据第四解码特征图得到输出概率估计结果如方差信息 σ_3 ，结合第三熵估计输出结果对第三编码码流或第三待解码码流进行熵解码得到第三解码特征图，第三解码特征图进行反处理后作为更新的第三解码特征图，并将更新的第三解码特征图作为第二熵估计的输入来估计第二解码特征图的概率；

步骤 3: 如图 18A 所示，根据第二熵估计输出概率估计结果如方差信息 σ_2 ，对第二编码

码流或第二待解码码流进行熵解码得到第二解码特征图，第二解码特征图进行反处理后作为更新的第二解码特征图，并将更新的第二解码特征图作为第一熵估计的输入来估计第一解码特征图的概率；

步骤4：如图18A所示，根据第一熵估计输出概率估计结果如方差信息 σ_1 ，对第一编码码流或第一待解码码流进行熵解码得到第一解码特征图，第一解码特征图进行反处理后作为更新的第一解码特征图；

步骤5：解码网络模块根据更新的第一、第二、第三解码特征图得到重构音频或图像。

解码端的反处理操作是将解码特征图的各像素点的像素值与熵估计输出的均值相加。

图19示出了本申请提供的一种实施例，该实施例是一个音频压缩框架。模型运行步骤如下：

编码端：

步骤1：编码网络模块分为了N个特征层，分层提取音频x每一层的输出特征；

步骤2：取整模块将音频x各个层的输出特征取整为各个层的取整特征；

步骤3：熵估计网络模块按照从底层(第n特征层)到浅层(第1层)的顺序依次估计各层取整特征的概率；

步骤4：熵编码模块从底层(第n特征层)到浅层(第1层)的顺序依次根据熵估计结果进行熵编码，各层的编码码流用来网络传输或者保存成本地文件。

解码端：

步骤1：熵估计网络和熵解码模块从底层(第n特征层)到浅层(第1层)的顺序依次解码出各层的解码特征；

步骤2：解码网络模块根据各层的解码特征重建出音频x'。

本领域技术人员能够领会，结合本文公开描述的各种说明性逻辑框、模块和算法步骤所描述的功能可以硬件、软件、固件或其任何组合来实施。如果以软件来实施，那么各种说明性逻辑框、模块、和步骤描述的功能可作为一或多个指令或代码在计算机可读媒体上存储或传输，且由基于硬件的处理单元执行。计算机可读媒体可包含计算机可读存储媒体，其对应于有形媒体，例如数据存储媒体，或包括任何促进将计算机程序从一处传送到另一处的媒体(例如，根据通信协议)的通信媒体。以此方式，计算机可读媒体大体上可对应于(1)非暂时性的有形计算机可读存储媒体，或(2)通信媒体，例如信号或载波。数据存储媒体可为可由一或多个计算机或一或多个处理器存取以检索用于实施本申请中描述的技术的指令、代码和/或数据结构的任何可用媒体。计算机程序产品可包含计算机可读媒体。

作为实例而非限制，此类计算机可读存储媒体可包括RAM、ROM、EEPROM、CD-ROM或其它光盘存储装置、磁盘存储装置或其它磁性存储装置、快闪存储器或可用来存储指令或数据结

构的形式的所要程序代码并且可由计算机存取的任何其它媒体。并且，任何连接被恰当地称作计算机可读媒体。举例来说，如果使用同轴缆线、光纤缆线、双绞线、数字订户线(DSL)或例如红外线、无线电和微波等无线技术从网站、服务器或其它远程源传输指令，那么同轴缆线、光纤缆线、双绞线、DSL 或例如红外线、无线电和微波等无线技术包含在媒体的定义中。但是，应理解，所述计算机可读存储媒体和数据存储媒体并不包括连接、载波、信号或其它暂时媒体，而是实际上针对于非暂时性有形存储媒体。如本文中所使用，磁盘和光盘包含压缩光盘(CD)、激光光盘、光学光盘、数字多功能光盘(DVD)和蓝光光盘，其中磁盘通常以磁性方式再现数据，而光盘利用激光以光学方式再现数据。以上各项的组合也应包含在计算机可读媒体的范围内。

可通过例如一或多个数字信号处理器(DSP)、通用微处理器、专用集成电路(ASIC)、现场可编程逻辑阵列(FPGA)或其它等效集成或离散逻辑电路等一或多个处理器来执行指令。因此，如本文中所使用的术语“处理器”可指前述结构或适合于实施本文中所描述的技术的任一其它结构中的任一者。另外，在一些方面中，本文中所描述的各种说明性逻辑框、模块、和步骤所描述的功能可以提供于经配置以用于编码和解码的专用硬件和/或软件模块内，或者并入在组合编解码器中。而且，所述技术可完全实施于一或多个电路或逻辑元件中。

本申请的技术可在各种各样的装置或设备中实施，包含无线手持机、集成电路(IC)或一组 IC(例如，芯片组)。本申请中描述各种组件、模块或单元是为了强调用于执行所揭示的技术的装置的功能方面，但未必需要由不同硬件单元实现。实际上，如上文所描述，各种单元可结合合适的软件和/或固件组合在编码解码器硬件单元中，或者通过互操作硬件单元(包含如上文所描述的一或多个处理器)来提供。

以上所述，仅为本申请示例性的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到的变化或替换，都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应该以权利要求的保护范围为准。

权 利 要 求 书

1、一种音视频或图像的编码方法，其特征在于，包括：

待编码音频或图像经过编码网络获得所述待编码音频或图像的第一输出特征图；

所述待编码音频或图像经过编码网络获得所述待编码音频或图像的第二输出特征图；

对所述第一输出特征图经过取整以获得所述待编码音频或图像的第一取整特征图；

对所述第二输出特征图经过取整以获得所述待编码音频或图像的第二取整特征图；

根据第二熵估计的概率估计结果，对所述第二取整特征图进行熵编码得到第二编码码流；

根据所述第二取整特征图或/和所述第二熵估计的概率估计结果，对第一熵估计进行概率估计以获得第一熵估计的概率估计结果；以及

根据所述第一熵估计的概率估计结果，对所述第一取整特征图进行熵编码得到第一编码码流。

2、根据权利要求1的方法，其特征在于，所述方法还包括：

对所述待编码音频或图像经过编码网络获得所述待编码音频或图像的第三输出特征图，以及对所述第三输出特征图取整获得所述待编码音频或图像的第三取整特征图；

根据第三熵估计的概率估计结果，对所述第三取整特征图进行熵编码得到第三编码码流。

3、根据权利要求2所述的方法，其特征在于，所述方法还包括：

根据所述待编码音频或图像的第三熵估计结果和/或所述第三取整特征图，对所述第二熵估计进行概率估计以获得第二熵估计的概率估计结果。

4、根据权利要求1-3中任一项所述的方法，其特征在于，所述方法还包括：

所述待编码音频或图像经过编码网络获得所述待编码音频或图像的第k输出特征图，以及对所述第k输出特征图取整获得所述待编码音频或图像的第k取整特征图，其中 $k > 3$ 整数。

5、根据权利要求4所述的方法，其特征在于，所述第k取整特征图作为第n熵估计的输入信息，其中n为整数且 $n \leq k$ 。

6、根据权利要求1-5中任一项所述的方法，其特征在于，所述方法还包括：

所述至少一个取整特征图作为待编码音频或图像的解码网络的输入，输出所述待编码音频或图像的重构音频或图像；

对所述待编码音频或图像与所述重构音频或图像的对应像素点求差以获得所述待编码音频或图像的残差图像；

根据以下至少一种信息得到待解码音频或图像的无损熵估计的概率估计：

所述任意一个或多个熵估计结果；

所述任意一个或多个取整特征图；或

所述重构音频或图像；

根据所述无损熵估计的概率估计结果，对所述残差图像进行熵编码以获得所述待编码音频或图像的无损编码码流。

7、根据权利要求1所述的方法，其特征在于，所述方法还包括：

对所述第一输出特征图经过取整以获得所述待编码音频或图像的第一取整特征图之前，所述方法还包括：

对所述第一输出特征图中的像素点的像素值与第一熵估计输出的均值相减得到第一偏移特征图；

对所述第一输出特征图经过取整以获得所述待编码音频或图像的第一取整特征图包括：

对所述第一偏移特征图经过取整以获得所述待编码音频或图像的所述第一取整特征图。

8、根据权利要求2-5中任一项所述的方法，其特征在于，所述方法还包括：

对第 m 输出特征图经过取整以获得所述待编码音频或图像的第 m 取整特征图之前，所述方法还包括：

对所述第 m 输出特征图中的像素点的像素值与第 m 熵估计输出的均值相减得到第 m 偏移特征图；

对所述第 m 输出特征图经过取整以获得所述待编码音频或图像的第 m 取整特征图包括：

对所述第 m 偏移特征图经过取整以获得所述待编码音频或图像的所述第 m 取整特征图；

对第 m 取整特征图中的像素点的像素值与第 m 熵估计输出的均值相加后作为第 t 熵估计的输入，其中 m 为大于1的整数， t 为小于 m 的整数。

9、一种图像的解码方法，其特征在于，包括：

获取待解码音频或图像的第一待解码码流；

获取所述待解码音频或图像的第二待解码码流；

根据所述待解码音频或图像的第二熵估计的概率估计结果，对所述待解码音频或图像的第二解码码流进行熵解码获得所述待解码音频或图像的第二解码特征图；

根据所述第二特征图或/和所述第二熵估计的概率估计结果，对第一熵估计进行概率估计以获得第一熵估计的概率估计结果；

根据所述第一熵估计的概率估计结果，对所述第一解码码流进行熵解码获得所述待解码音频或图像的第一解码特征图；

所述第一解码特征图或/和所述第二解码特征图作为解码网络的输入，输出所述待解码音频或图像的重构音频或图像。

10、根据权利要求9中所述的方法，其特征在于，所述方法还包括：

获取所述待解码音频或图像的第三待解码码流；

根据所述待解码音频或图像的第三熵估计的概率估计结果，对所述第三待解码码流进行

熵解码获得所述待解码音频或图像的第三解码特征图；

根据所述第三解码特征图或/和所述第三熵估计的概率估计结果，进行待解码音频或图像的第二熵估计的概率估计；

所述第一解码特征图、所述第二解码特征图以及所述第三解码特征图中的至少一种解码特征图作为解码网络的输入，输出所述待解码音频或图像的重构音频或图像。

11、根据权利要求 10 所述的方法，其特征在于，根据所述第三解码特征图或/和所述第三熵估计的概率估计结果，进行待解码音频或图像的第二熵估计的概率估计；

12、根据权利要求 9-11 中任一项所述的方法，其特征在于，所述方法还包括：

获取待解码音频或图像的第 m 待解码码流；

根据所述待解码音频或图像的第 m 熵估计的概率估计结果，对所述第 m 待解码码流进行熵解码获得所述待解码音频或图像的第 m 解码特征图，其中 $m > 3$ 整数。

13、根据权利要求 12 所述的方法，其特征在于，所述第 m 解码特征图和/或第 m 熵估计的概率估计结果作为第 t 熵估计的输入信息，其中 $t \leq m$ 整数。

14、根据权利要求 9-13 中任一项所述的方法，其特征在于，所述方法还包括：

获取待解码音频或图像的无损待解码码流；

根据以下至少一种信息得到待解码音频或图像的无损熵估计的概率估计结果：

所述至少一个解码特征图；

所述至少一个熵估计的概率估计结果；或

所述待解码音频或图像的重构音频或图像；

根据所述待解码音频或图像的无损熵估计的概率估计结果，对所述无损待解码码流进行熵解码获得所述待解码音频或图像的残差图像；

对所述待解码音频或图像的重构音频或图像与所述待解码音频或图像的残差图像的对应像素点求和以获得所述待解码音频或图像的无损重构音频或图像。

15、根据权利要求 9-14 中任一项所述的方法，其特征在于，当所述第一解码特征图作为解码网络的输入，对所述第一解码特征图中的像素点的像素值与第一熵估计输出的均值相加得到第一解码特征图。

16、一种编码器，其特征在于，包括处理电路，用于执行权利要求 1-9 任一项所述的方法。

17、一种解码器，其特征在于，包括处理电路，用于执行权利要求 9-15 任一项所述的方法。

18、一种计算机程序产品，其特征在于，包括程序代码，当其在计算机或处理器上执行时，用于执行权利要求 1-15 任一项所述的方法。

19、一种编码器，其特征在于，包括：

一个或多个处理器；

非瞬时性计算机可读存储介质，耦合到所述处理器并存储由所述处理器执行的程序，其中所述程序在由所述处理器执行时，使得所述编码器执行权利要求 1-8 任一项所述的方法。

20、一种解码器，其特征在于，包括：

一个或多个处理器；

非瞬时性计算机可读存储介质，耦合到所述处理器并存储由所述处理器执行的程序，其中所述程序在由所述处理器执行时，使得所述解码器执行根据权利要求任一项所述的方法。

21、一种编码器，其特征在于，包括：

一个或多个处理器；

非瞬时性计算机可读存储介质，耦合到所述处理器并存储由所述处理器执行的程序，其中所述程序在由所述处理器执行时，使得所述编码器执行根据权利要求任一项所述的方法。

22、一种图像处理器，其特征在于，包括处理电路，用于执行根据权利要求 1 至 15 任一项所述的方法。

23、一种非瞬时性计算机可读存储介质，其特征在于，包括程序代码，当其由计算机设备执行时，用于执行根据权利要求任一项所述的方法。

10

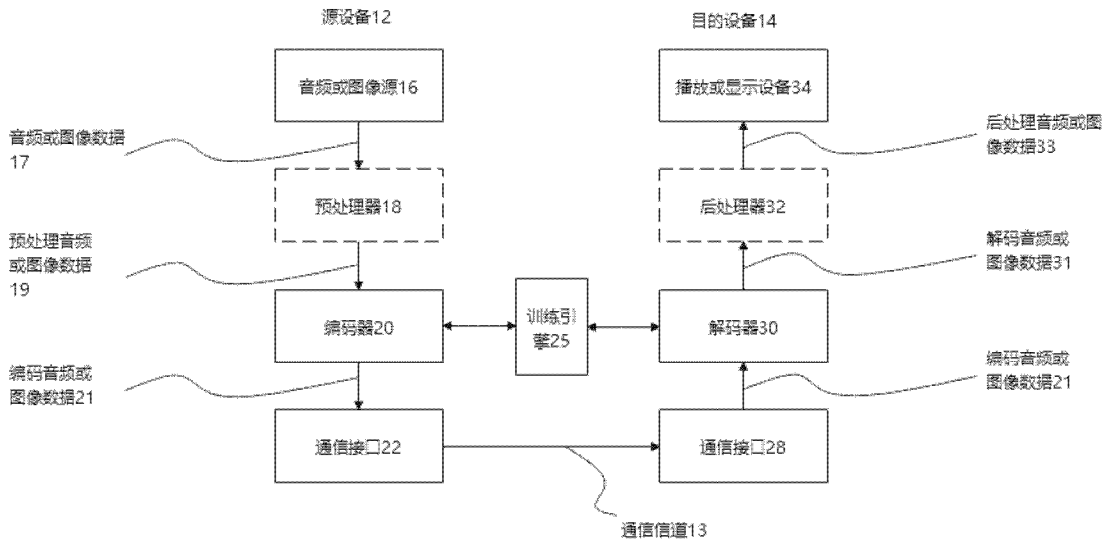


图 1A

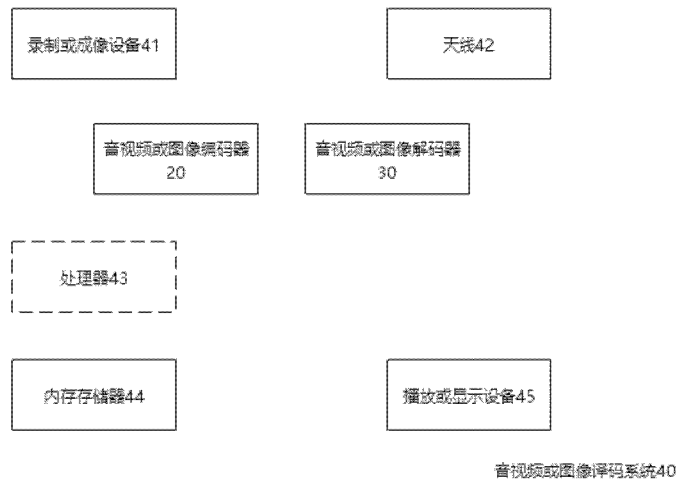


图 1B

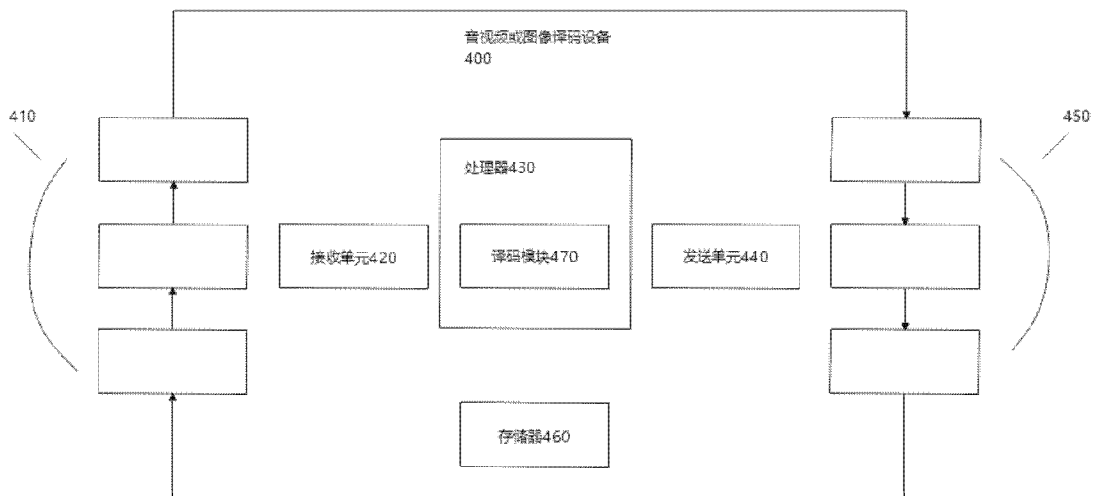


图 1C

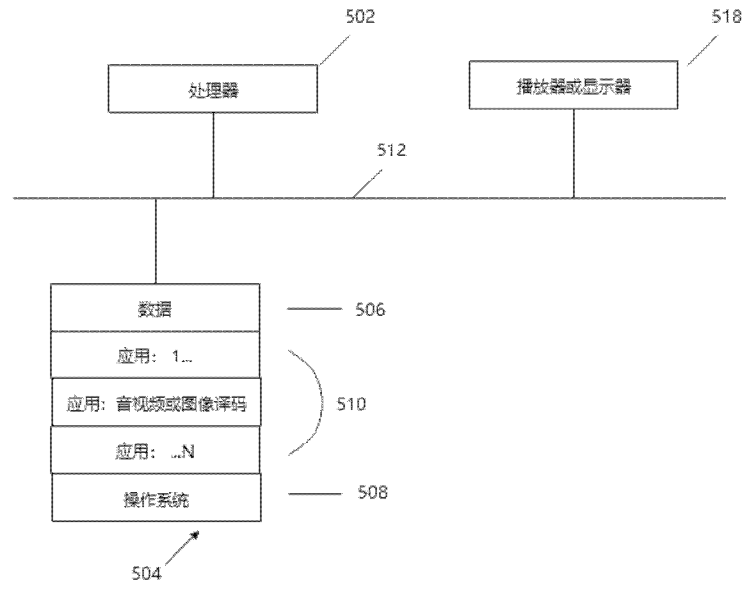


图 1D

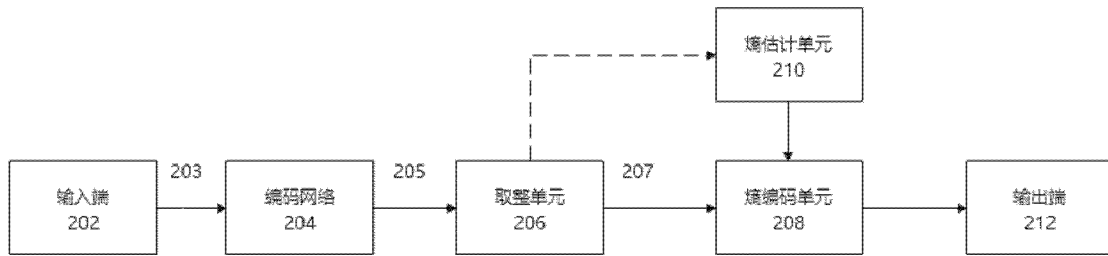


图 2

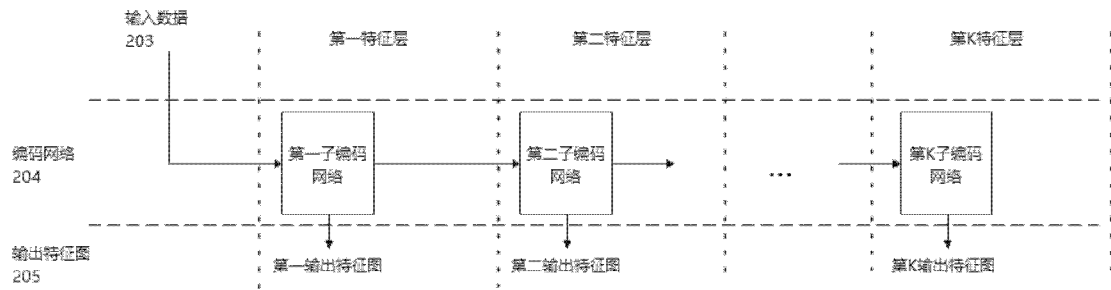


图 3A

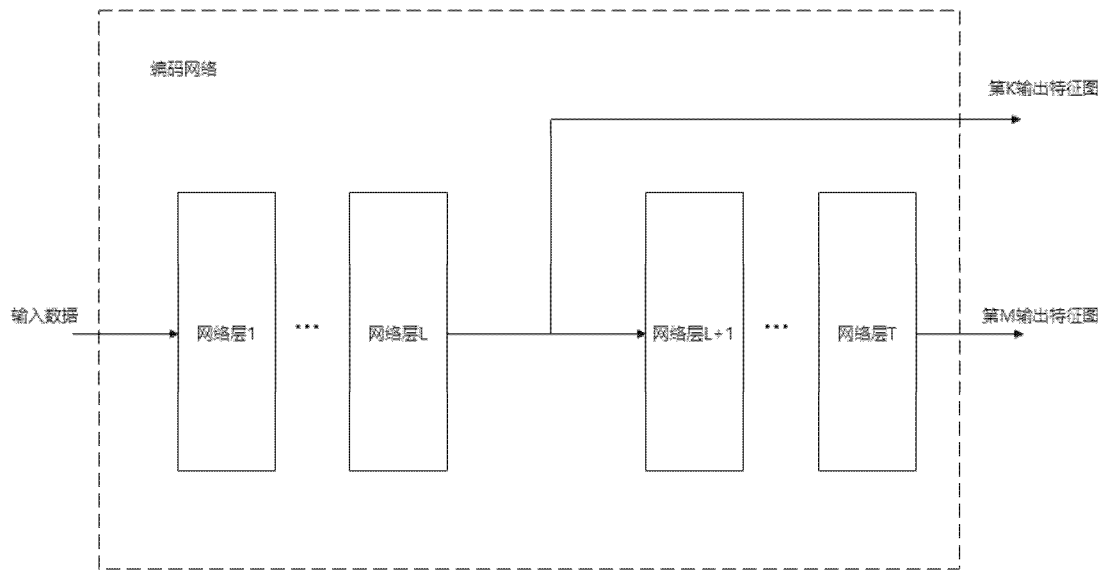


图 3B

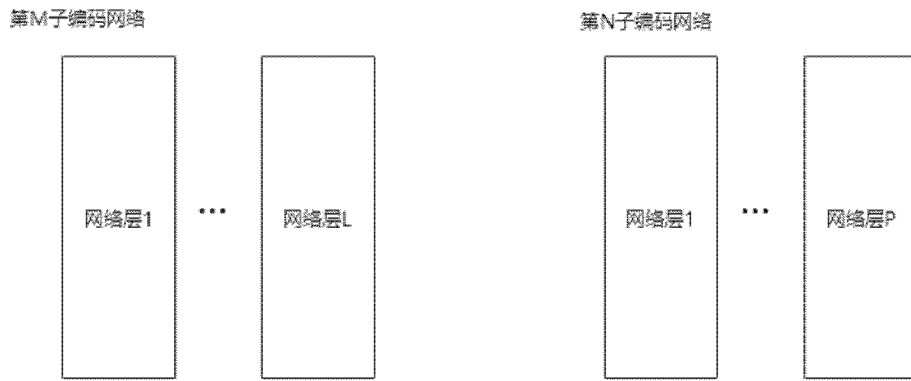


图 4

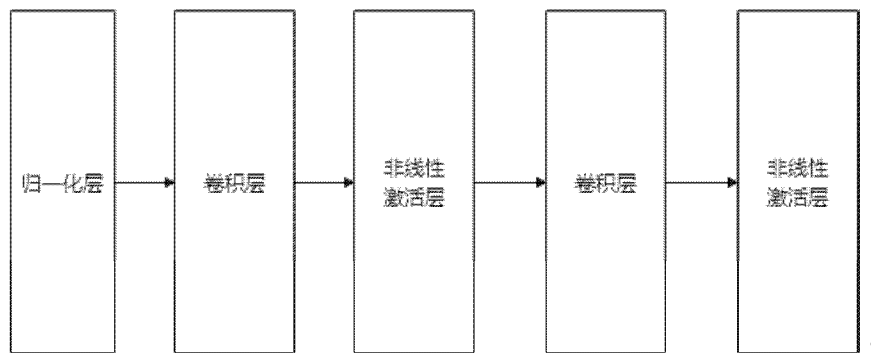


图 5

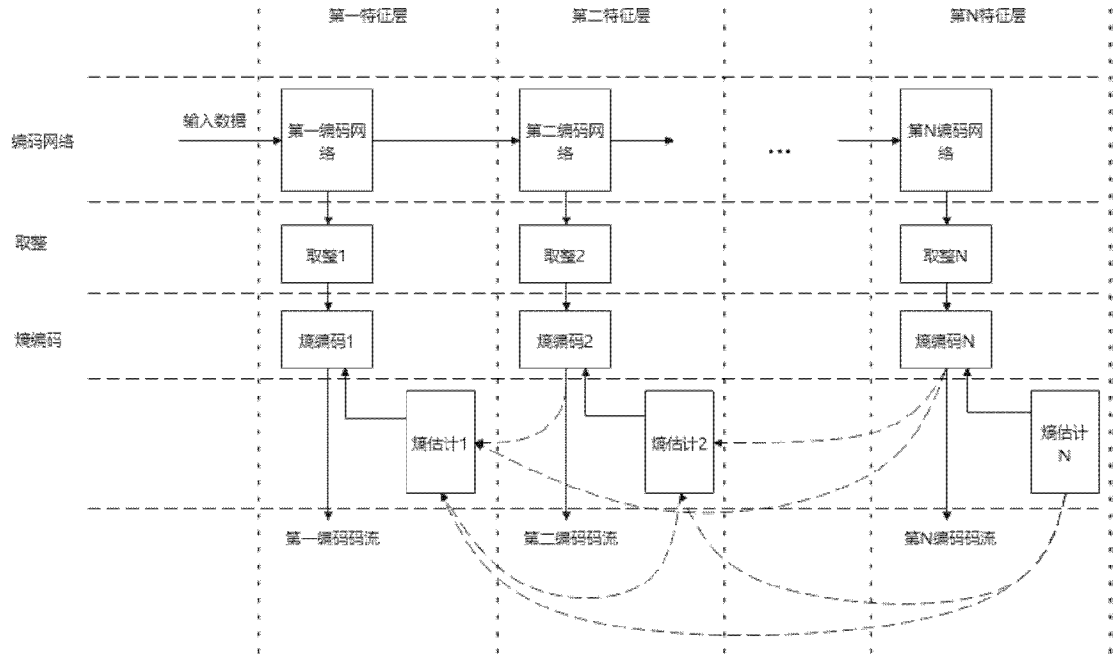


图 6A

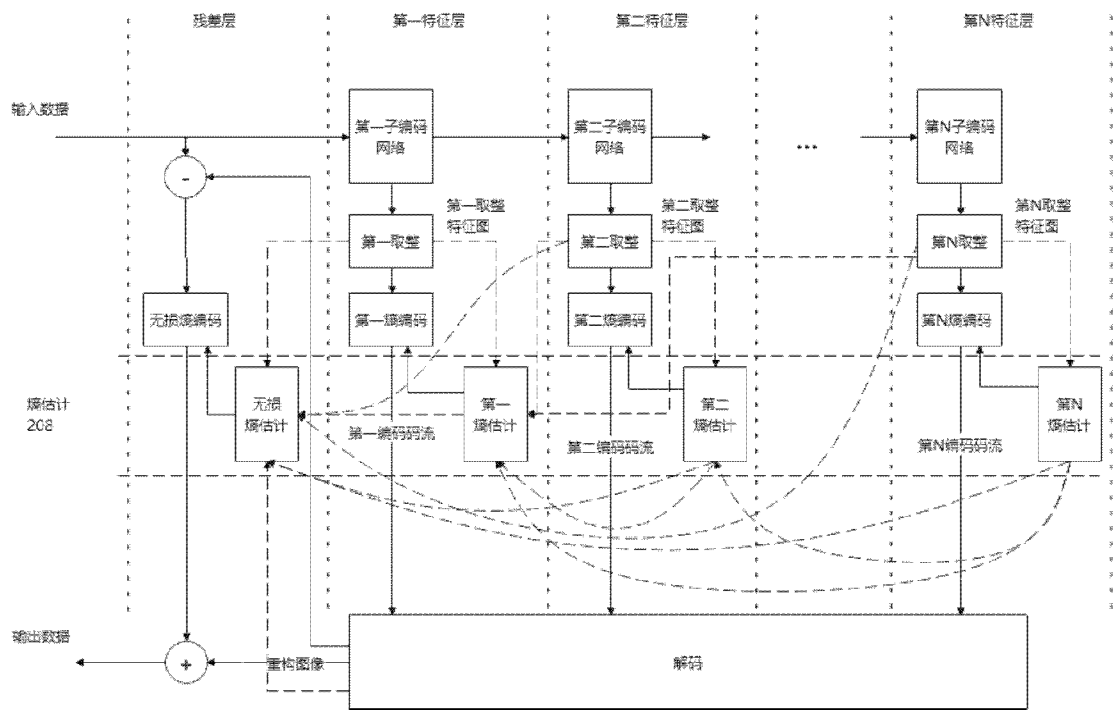


图 6B

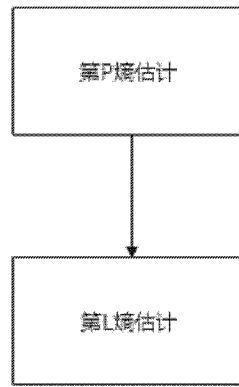


图 7A

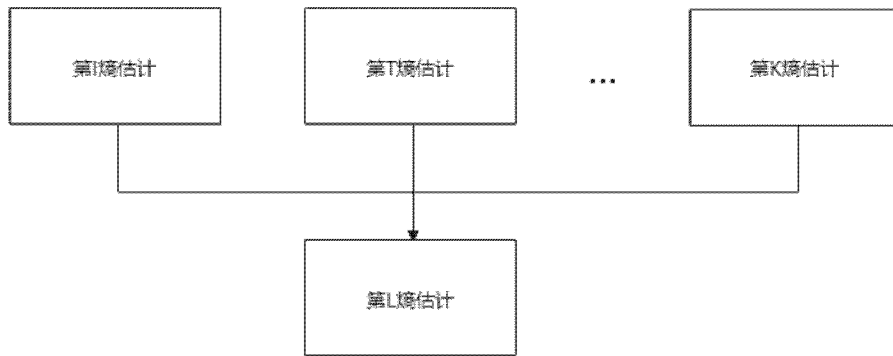


图 7B

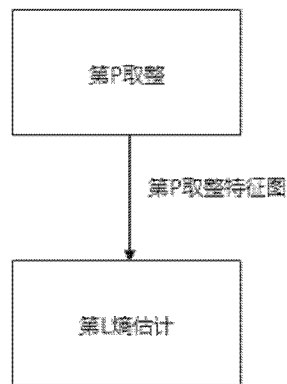


图 7C

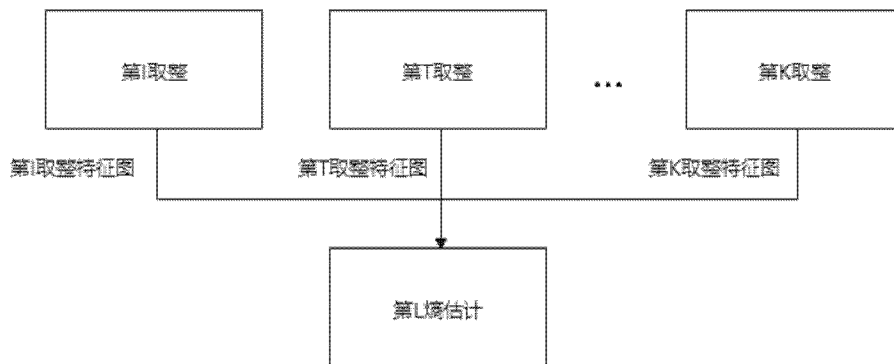


图 7D

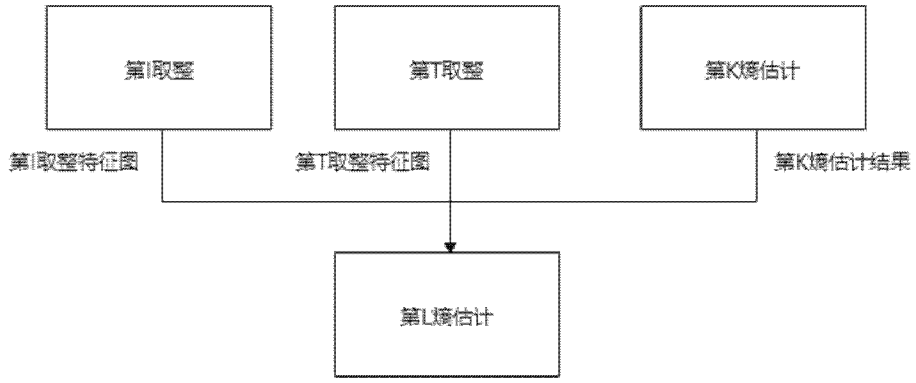


图 7E

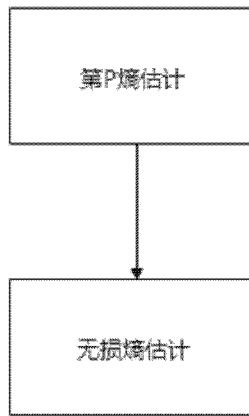


图 7F

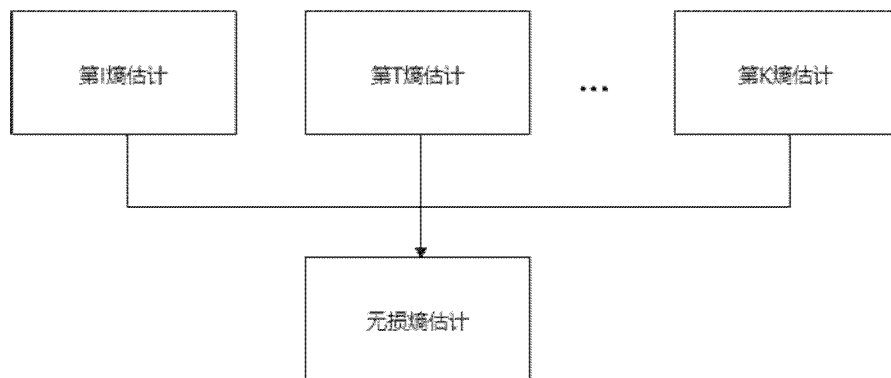


图 7G

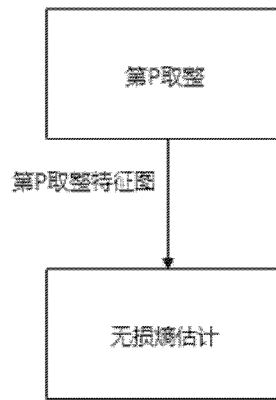


图 7H

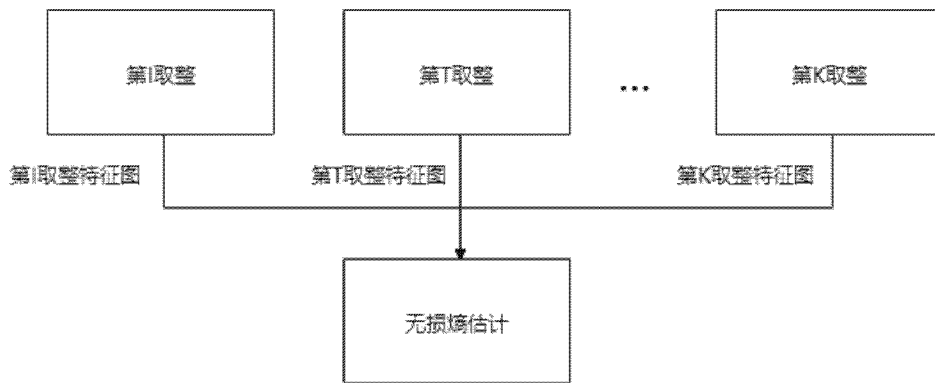


图 7I

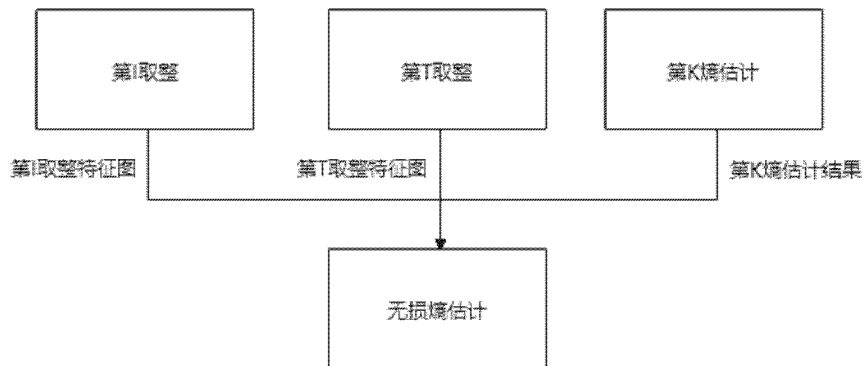


图 7J

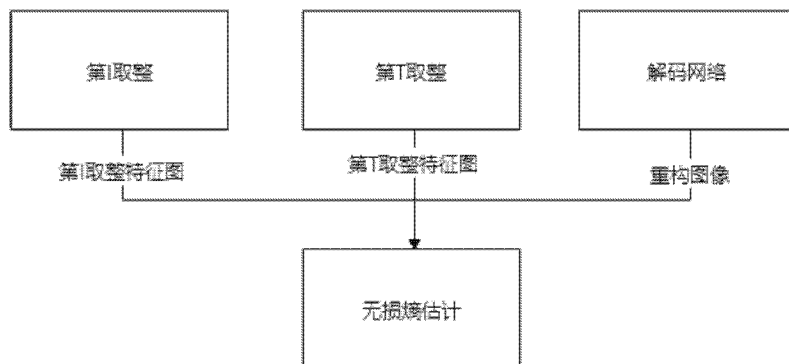


图 7K

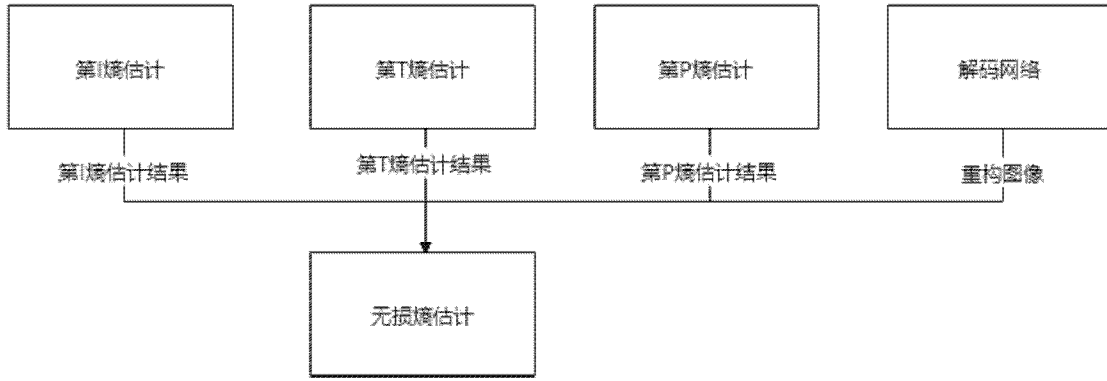


图 7L

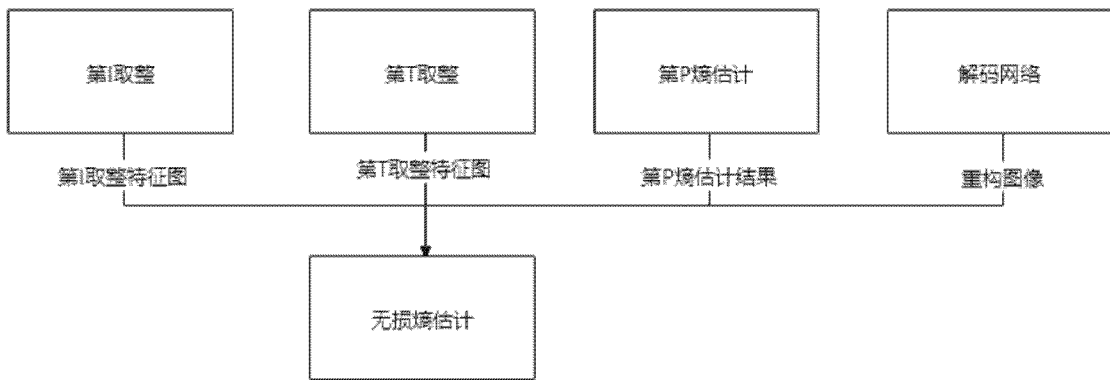


图 7M

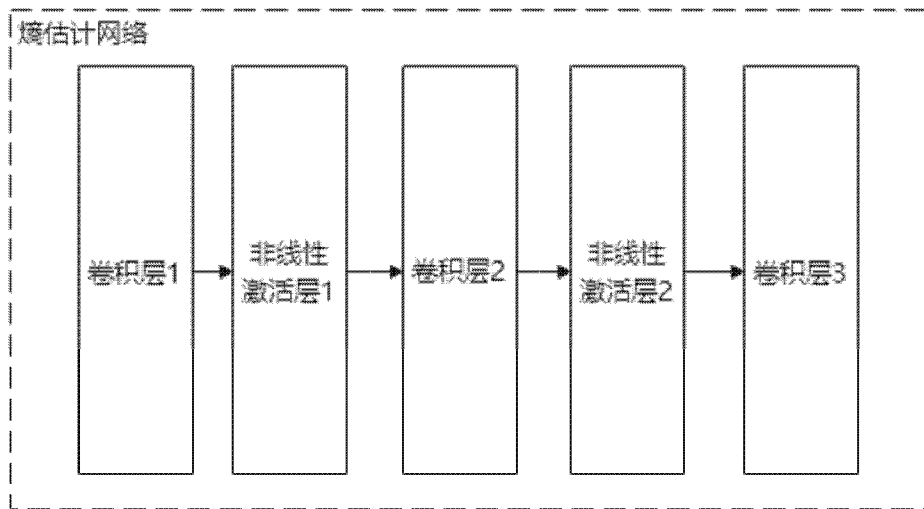


图 8

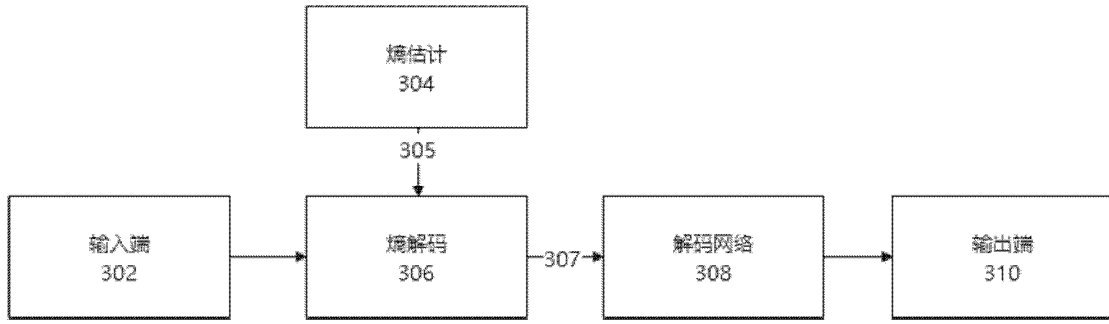


图 9

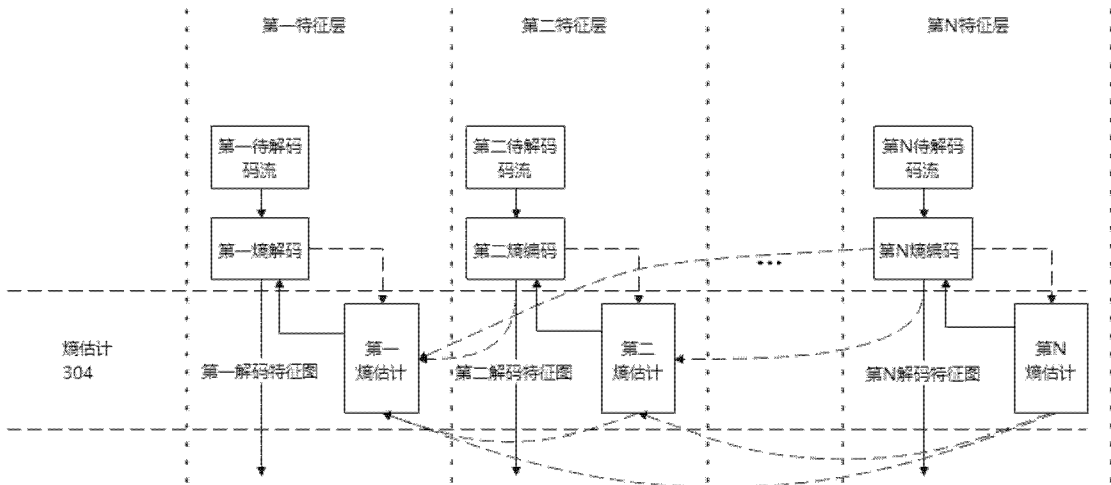


图 10A

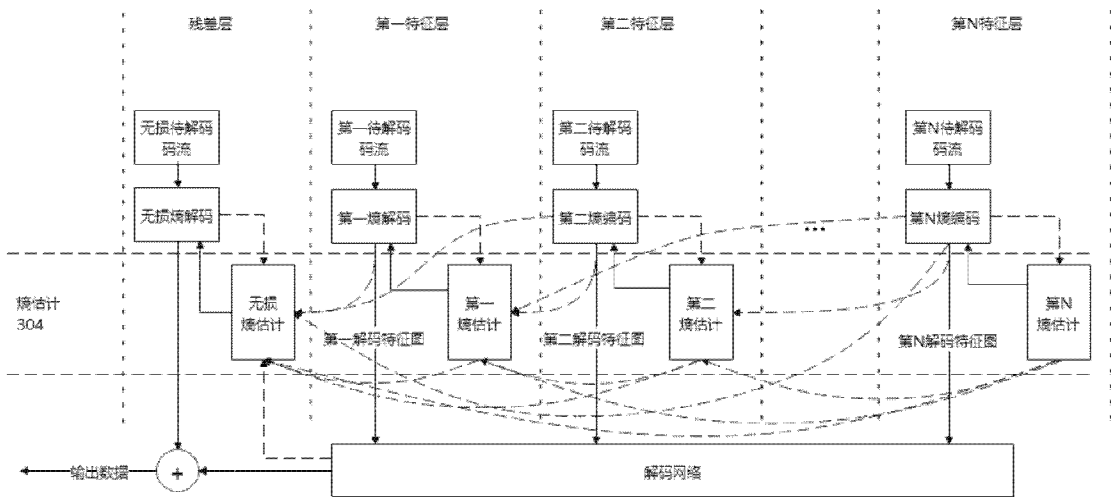


图 10B

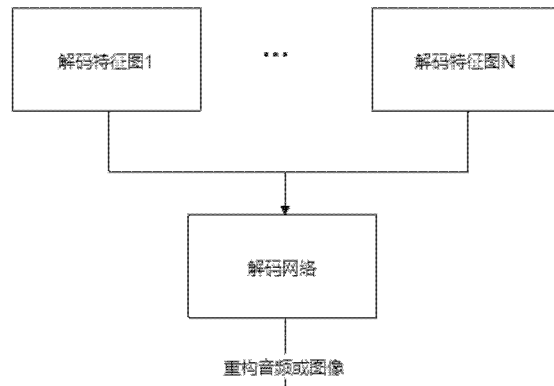


图 11

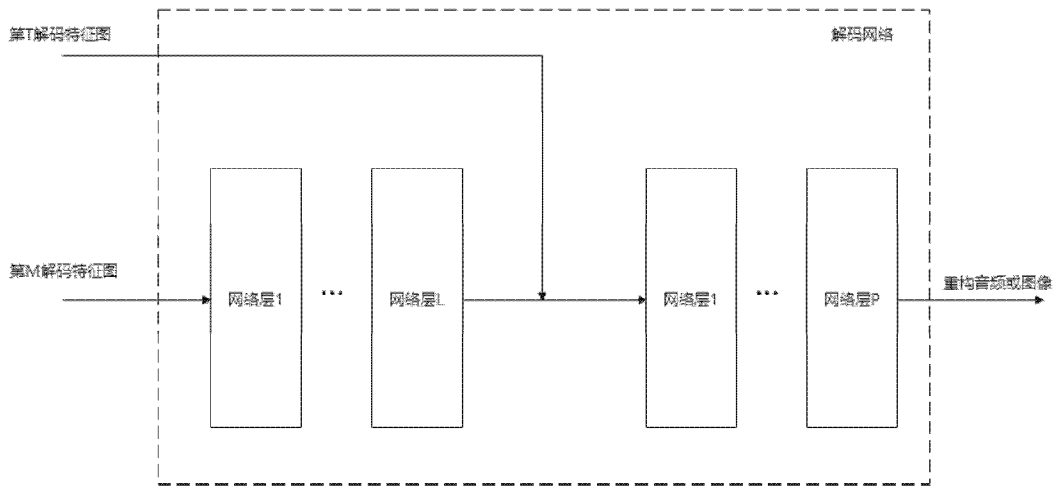


图 12

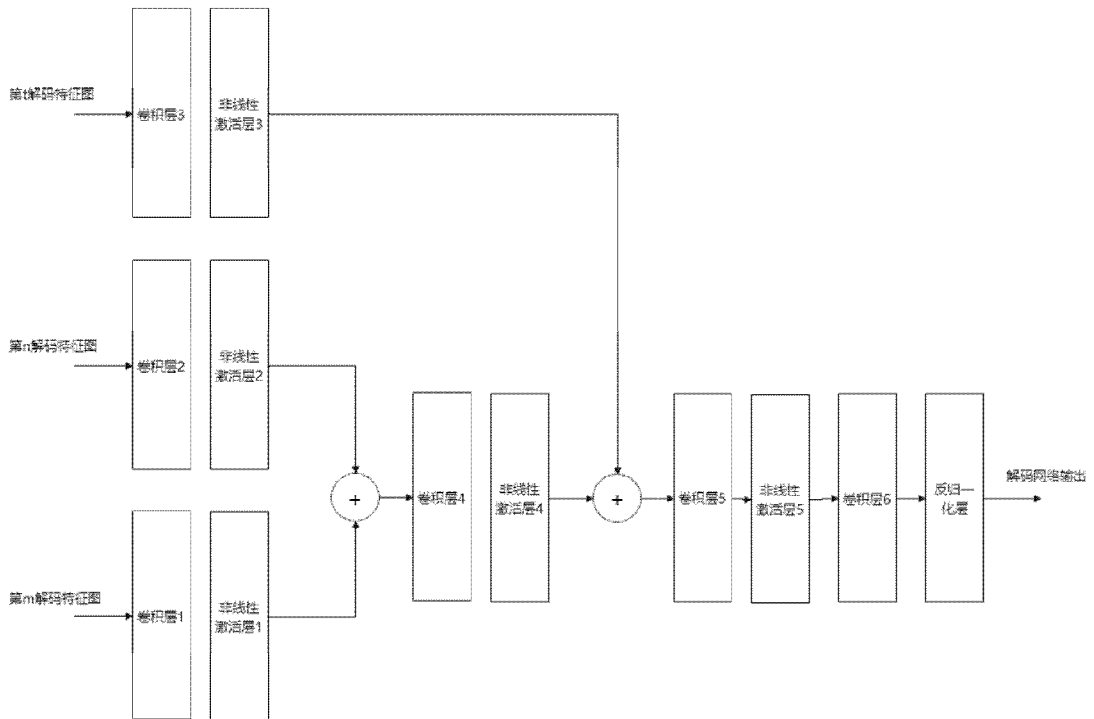


图 13

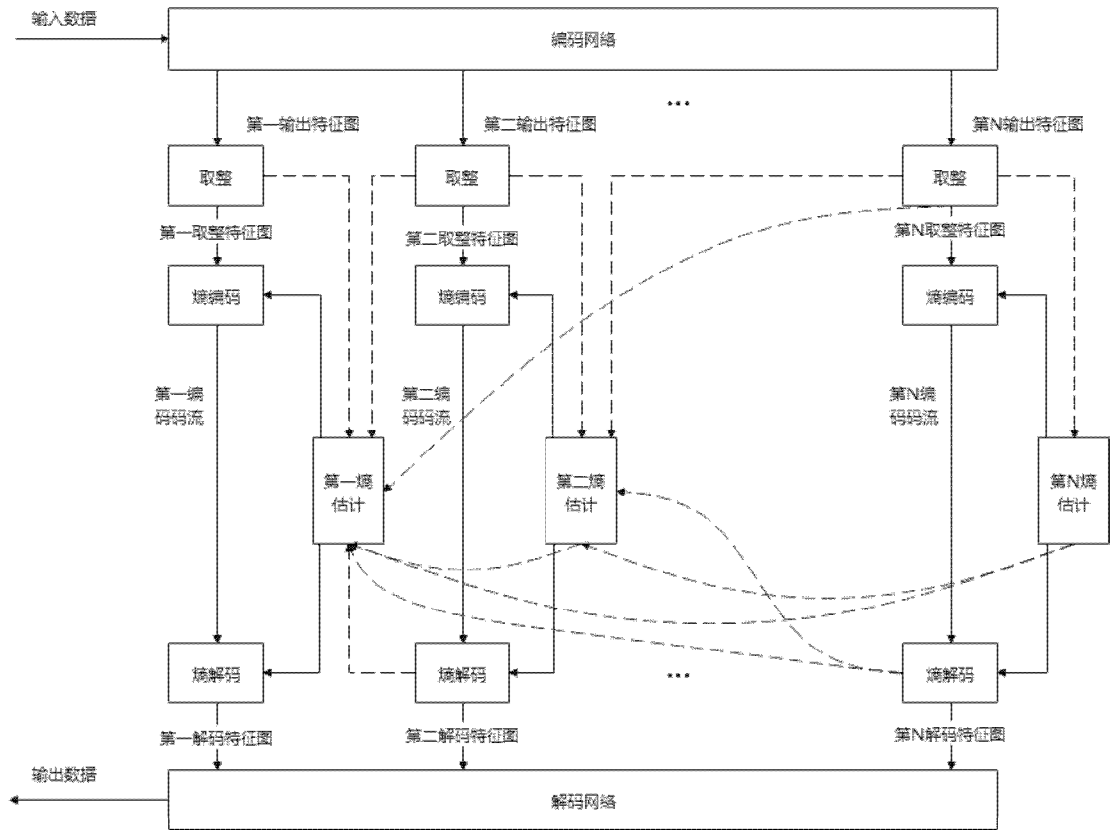


图 14A

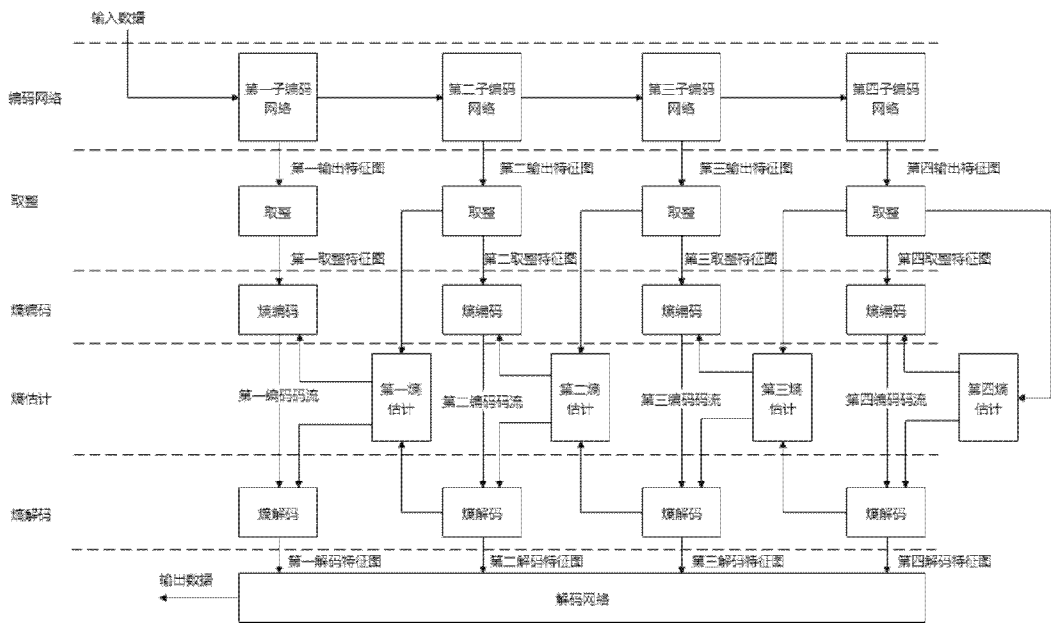


图 14B

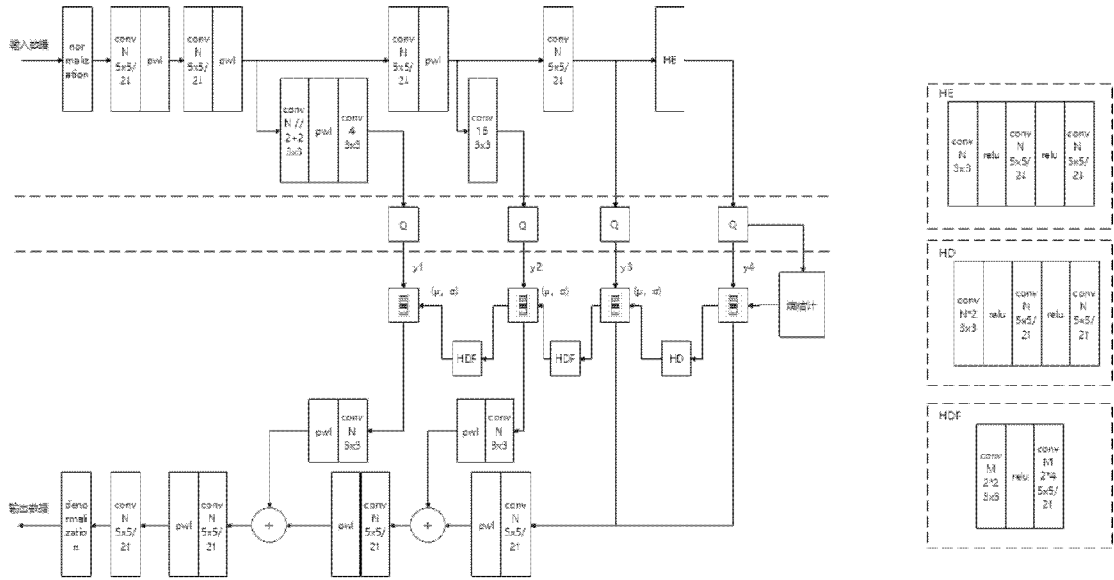


图 14C

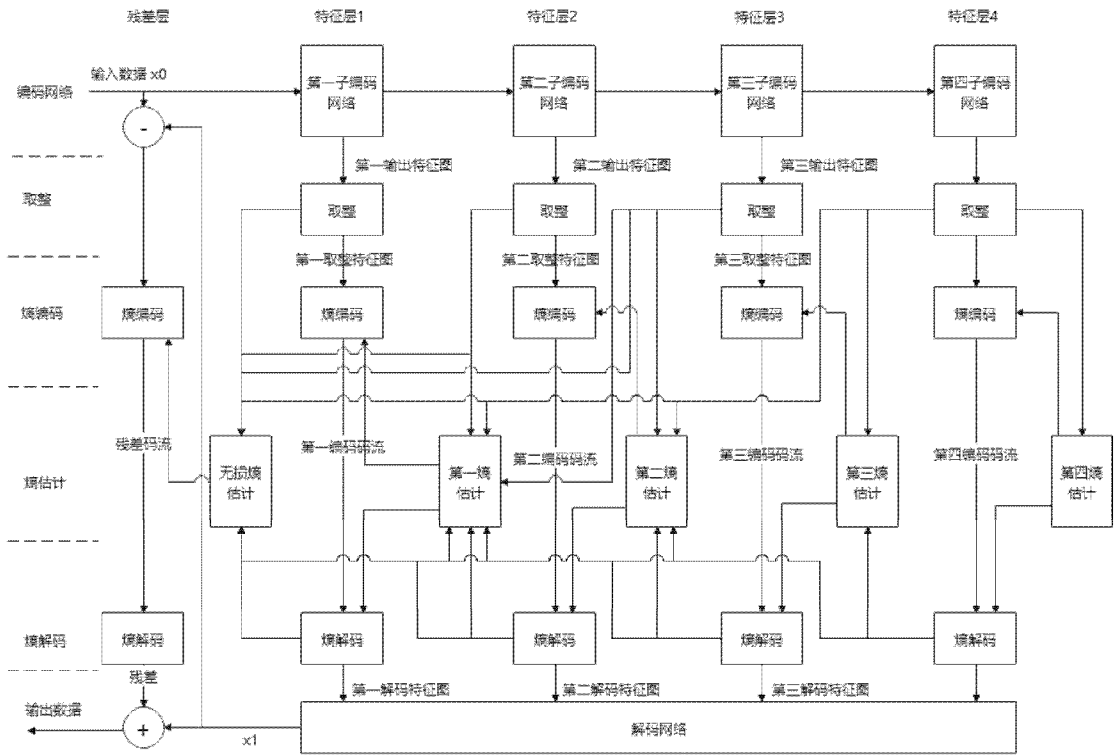


图 15

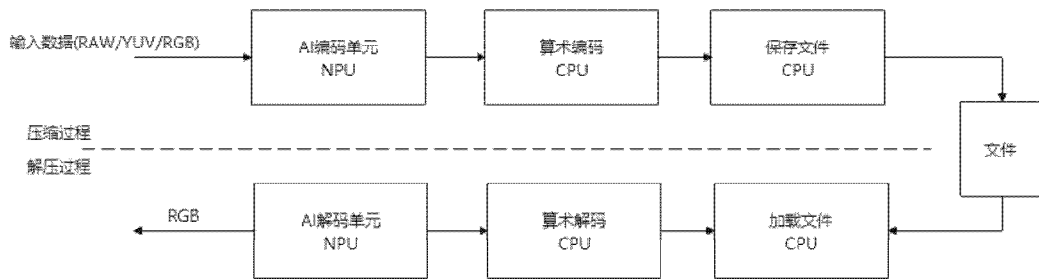


图 16

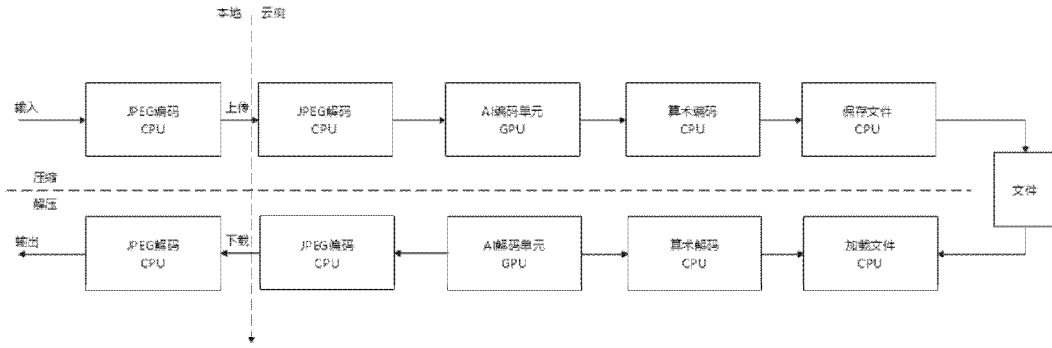


图 17

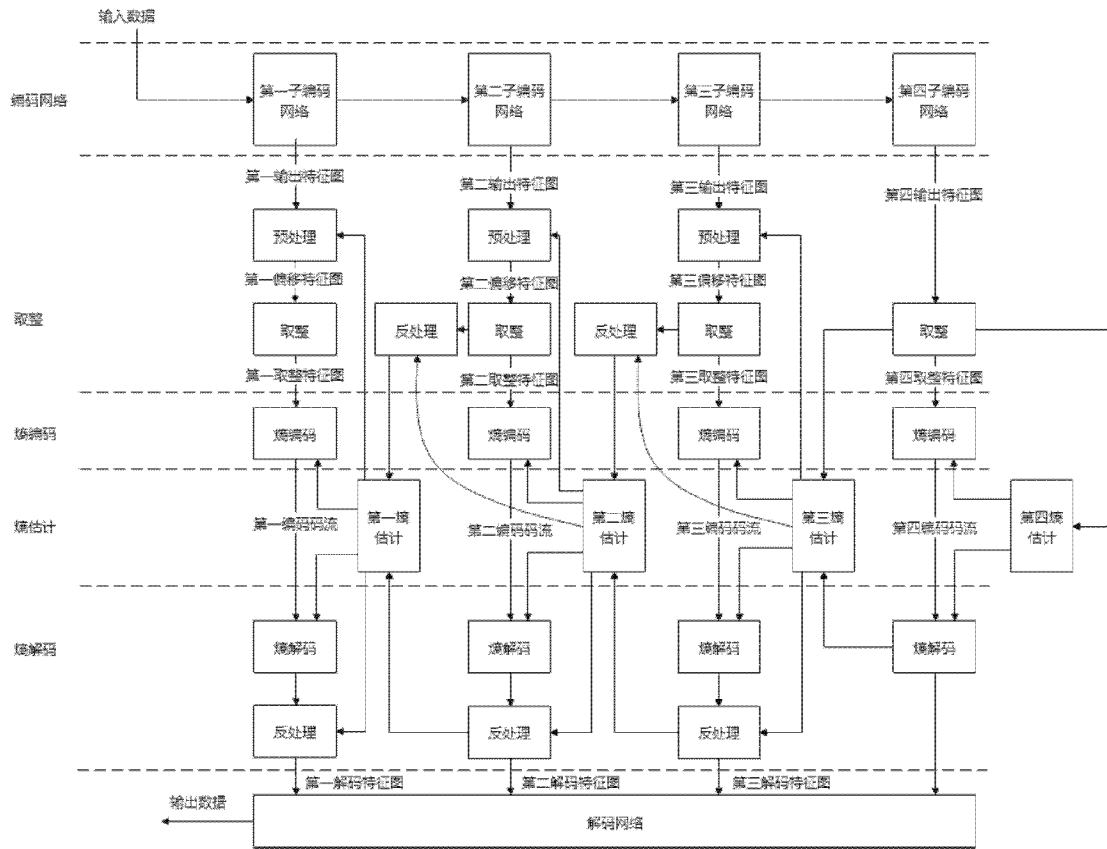


图 18A

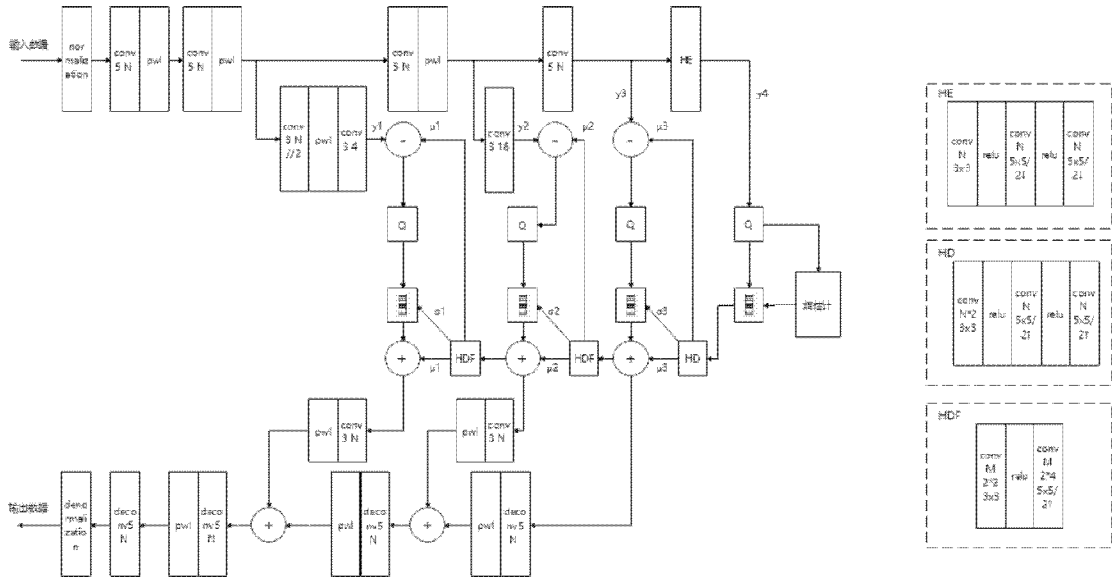


图 18B

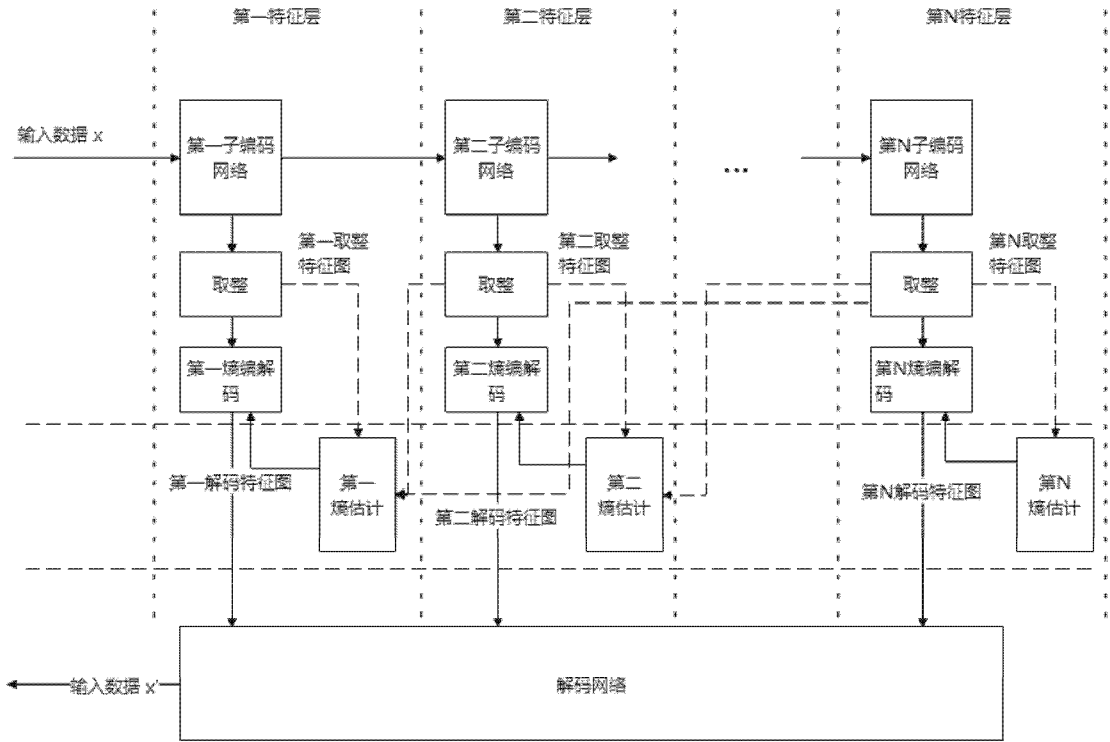


图 19

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/077341

A. CLASSIFICATION OF SUBJECT MATTER		
H04N 21/2343(2011.01)i		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols)		
H04N		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
CNTXT; CNABS; ENTXTC; VEN; CNKI: 编码, 取整, 熵, 特征图, 概率, 估计, encode, code, integer, entropy, feature map, probability, estimation		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 111818346 A (FUJITSU LIMITED) 23 October 2020 (2020-10-23) entire document	1-23
A	CN 111787323 A (TSINGHUA UNIVERSITY) 16 October 2020 (2020-10-16) entire document	1-23
A	CN 111988629 A (FUJITSU LIMITED) 24 November 2020 (2020-11-24) entire document	1-23
A	CN 109996071 A (SHANGHAI JIAO TONG UNIVERSITY) 09 July 2019 (2019-07-09) entire document	1-23
A	US 2020327701 A1 (FUJITSU LTD.) 15 October 2020 (2020-10-15) entire document	1-23
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
15 April 2022		22 April 2022
Name and mailing address of the ISA/CN		Authorized officer
China National Intellectual Property Administration (ISA/CN) No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088, China		
Facsimile No. (86-10)62019451		Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No. PCT/CN2022/077341

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	111818346	A	23 October 2020	None	
CN	111787323	A	16 October 2020	None	
CN	111988629	A	24 November 2020	None	
CN	109996071	A	09 July 2019	None	
US	2020327701	A1	15 October 2020	JP 2020173782 A	22 October 2020
				EP 3723048 A1	14 October 2020

国际检索报告

国际申请号

PCT/CN2022/077341

<p>A. 主题的分类</p> <p>H04N 21/2343(2011.01) i</p> <p>按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类</p>																				
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>H04N</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))</p> <p>CNXTX; CNABS; ENTXTC; VEN; CNKI:编码, 取整, 熵, 特征图, 概率, 估计, encode, code, integer, entropy, feature map, probability, estimation</p>																				
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>CN 111818346 A (富士通株式会社) 2020年10月23日 (2020 - 10 - 23) 全文</td> <td>1-23</td> </tr> <tr> <td>A</td> <td>CN 111787323 A (清华大学) 2020年10月16日 (2020 - 10 - 16) 全文</td> <td>1-23</td> </tr> <tr> <td>A</td> <td>CN 111988629 A (富士通株式会社) 2020年11月24日 (2020 - 11 - 24) 全文</td> <td>1-23</td> </tr> <tr> <td>A</td> <td>CN 109996071 A (上海交通大学) 2019年7月9日 (2019 - 07 - 09) 全文</td> <td>1-23</td> </tr> <tr> <td>A</td> <td>US 2020327701 A1 (FUJITSU LTD) 2020年10月15日 (2020 - 10 - 15) 全文</td> <td>1-23</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	A	CN 111818346 A (富士通株式会社) 2020年10月23日 (2020 - 10 - 23) 全文	1-23	A	CN 111787323 A (清华大学) 2020年10月16日 (2020 - 10 - 16) 全文	1-23	A	CN 111988629 A (富士通株式会社) 2020年11月24日 (2020 - 11 - 24) 全文	1-23	A	CN 109996071 A (上海交通大学) 2019年7月9日 (2019 - 07 - 09) 全文	1-23	A	US 2020327701 A1 (FUJITSU LTD) 2020年10月15日 (2020 - 10 - 15) 全文	1-23
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
A	CN 111818346 A (富士通株式会社) 2020年10月23日 (2020 - 10 - 23) 全文	1-23																		
A	CN 111787323 A (清华大学) 2020年10月16日 (2020 - 10 - 16) 全文	1-23																		
A	CN 111988629 A (富士通株式会社) 2020年11月24日 (2020 - 11 - 24) 全文	1-23																		
A	CN 109996071 A (上海交通大学) 2019年7月9日 (2019 - 07 - 09) 全文	1-23																		
A	US 2020327701 A1 (FUJITSU LTD) 2020年10月15日 (2020 - 10 - 15) 全文	1-23																		
<p><input type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><input checked="" type="checkbox"/> 见同族专利附件。</p>																				
<p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“P” 公布日先于国际申请日但迟于所要求的优先权日的文件</p> <p>“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件</p> <p>“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性</p> <p>“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性</p> <p>“&” 同族专利的文件</p>																				
<p>国际检索实际完成的日期</p> <p>2022年4月15日</p>		<p>国际检索报告邮寄日期</p> <p>2022年4月22日</p>																		
<p>ISA/CN的名称和邮寄地址</p> <p>中国国家知识产权局(ISA/CN)</p> <p>中国北京市海淀区蓟门桥西土城路6号 100088</p> <p>传真号 (86-10)62019451</p>		<p>授权官员</p> <p>王倩</p> <p>电话号码 86-(010)-62412164</p>																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/077341

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	111818346	A	2020年10月23日	无	
CN	111787323	A	2020年10月16日	无	
CN	111988629	A	2020年11月24日	无	
CN	109996071	A	2019年7月9日	无	
US	2020327701	A1	2020年10月15日	JP	2020173782 A 2020年10月22日
				EP	3723048 A1 2020年10月14日