团体标准《信息技术 视觉特征编码 第4部分：深度特征图》（征求意见稿）

编制说明

**一、工作简况**

新一代人工智能产业技术创新战略联盟（简称AITISA联盟）是在科技部的指导下于2017年成立，由近200家来自行业内的顶尖企业、高校和科研院所、资本机构、服务机构、创新创业企业组成。在AVS工作组近20年标准制定经验及标准制定团队的基础上，联盟从成立开始，便以人工智能产业发展需求为引领，围绕产业发展中存在的共性问题，建立了AI标准工作组，着手人工智能标准体系的搭建和技术标准的制定。AI标准工作组内部成立了若干专题组，分别承担相关细分领域的标准制定工作，视觉特征编码专题组是其中之一，致力于视觉特征编码相关技术标准的制订。

《信息技术 视觉特征编码 第4部分：深度特征图》规定了图像分析任务中，深度网络提取的特征图数据的编码格式和解码工具。本标准适用于图像数据中目标或场景的分类、检索、识别等应用。

根据中关村视听产业技术创新联盟2023年标准制修订计划的安排，下达了由鹏城实验室牵头制定的团体标准《信息技术 视觉特征编码 第4部分：深度特征图》（标准计划号为2021032504）的制定任务。

工作组自2017年12月第一次会议开始，讨论启动视觉特征编码标准制定工作，为视觉特征建立新一代编解码技术体系架构，以解决视觉特征编码的难题，推进领域技术发展。在标准制定的整体路线方面，前三次会议(2017年12月-2018年6月)对视觉特征编码的技术需求和应用场景进行讨论与梳理，首先形成了《视觉特征编码提案征集书(草案)》(AI M1016)技术文档与《视觉特征编码评价框架(草案)》(AI M1017)，前者明确了提案需求，并规提案阶段设置、提案内容、时间节点计划等，后者列出了典型视觉任务的验证数据集，设置核心实验中的测试任务、基本思路、测试条件、参考技术文档、性能度量等。经过后续会议(2018年8月-2019年12月)修订完善，最终形成了《视觉特征编码核心实验数据集6.0》(AI N1160)、《视觉特征编码核心实验设计6.0》(AI N1159)、《视觉特征编码测试模型V1.0》(AI N1093)技术文档。关于深度特征编码，工作组在2018年8月第4次会议首次收到《基于视频编码标准的Deep feature有损压缩框架》(AI M1030)技术提案，提出采纳预量化与视频编码相结合的方案作为特征图压缩的基础编码框架。在2018年12月第5次会议收到《基于视频编码标准的特征图有损压缩框架及测试条件》(AI M1061)技术提案，提出基于视频编码标准的深度特征图有损压缩框架，并最终审议通过。在2019年3月第6次会议收到《多任务下特征图有损压缩测试及实验平台升级》(AI M1088)技术提案，该提案报告了在目标检测任务下的特征图有损压缩性能，并建议更新实验平台。在2019年8月第8次会议收到《特征图有损压缩中预量化和重打包模块的模式拓展》(AI M1122)技术提案，并最终审议通过。在2019年12月第9次会议收到《特征图有损压缩试验补充》(AI M1154)技术提案，该提案对特征图有损压缩进行了试验补充，并对视频序列特征进行试验，探索含有时空冗余的特征序列编码压缩。在2020年3月第10次会议收到《深度特征图量化技术更新》(AI M1220)技术提案，该提案针对特征图有损压缩任务，提出了一种基于特征图统计特性的自适应量化方法，并最终审议通过。在上述技术提案的基础上，形成了《信息技术 视觉特征编码 第4部分：深度特征图》标准的第一份草案，之后，该草案在工作组内部进行公示与审阅，广泛采纳各单位专家的意见，对收到的反馈意见进行反复认真讨论和修订相应内容，并最终于2023年5月形成了标准的征求意见稿。

本标准起草单位：鹏城实验室，北京大学，中新国际联合研究院，青岛海信网络科技股份有限公司，青岛新一代人工智能技术研究院，青岛图灵科技有限公司，浙江邦盛科技股份有限公司

本标准起草人：陈卓，段凌宇，Alex C. Kot，Weisi Lin，杨文瀚，汪维，高峰，冯栋，王雯雯，王新宇，陈伟

**二、标准编制原则和确定主要内容的论据及解决的主要问题**

随着深度学习技术的发展和普及，诸多视觉分析任务开始以深度模型作为算法内核。深度模型的计算量往往十分巨大，云端更难以承受数以万计的视频分析需求，因此将特征提取转移至边缘端十分必要。深度模型的多层次结构产生不同语义层次的特征图，又使特征对视觉分析任务的泛化性得到保证。本部分即是面向深度特征图编解码而制定的，目标是提高深度特征图的编码效率以及降低编解码过程中的信息损失。

《信息技术 视觉特征编码 第4部分：深度特征图》是以我国科研机构和重点科研创新企业为首、引领信息技术领域有影响力的企业共同创新而成。为了形成优化的技术方案，专题组根据标准的通用性和实用性对编码方法的架构进行了多轮讨论和评估。技术评估的基本依据是综合考虑提案对解决深度特征图编解码的有效性、可实现性、可扩展性、对已有设备的兼容性和现有标准的复用性以及知识产权情况等，这些原则与手段为深度特征图编码的技术先进性和妥善解决知识产权问题奠定了坚实基础。

本部分规范了深度特征图编码解码的过程和编码格式，适用于图像数据中目标或场景的分类、检索、识别等应用。本部分根据深度特征图的特点，提出了深度特征图量化方法以及重打包/反重打包方法，使深度特征图能够符合传统视频编解码器输入的编码格式，并且提高了深度特征图之间的时间相关性和空间相关性。本标准则是针对深度特征图的特点，专门设计了针对性的编码方式和编码格式，采用了传统的视频编解码技术，在保证任务性能的同时减少了数据体积。

《信息技术 视觉特征编码 第4部分：深度特征图》的特色技术包括：

**1. 预量化技术**

深度特征图数据范围很广，并且数据精度很高，同时深度特征图的数据大多数集中在值较小的数据范围，少量值在较大数据范围。可通过预量化技术可以在不影响特征应用的前提下，舍弃部分数据精度，实现降低数据精度的目标以及降低数据压缩难度。

**1) 自定义标量量化**

该技术根据对特征图统计特性的分析，通过手动设计、机器学习等方法设计符合深度特征图特征值分布的量化区间划分，用于对特征数据的量化操作，生成符合传统视频编码器输入要求的数据形式。该量化方式充分利用了特征图的统计性质，对量化区间进行精确划分，有效地降低了由于量化过程产生的量化误差。

**2. 深度特征图重打包/反重打包**

深度特征图重打包/反重打包将原始特征图的三维数组变化为符合传统视频编码器输入要求的YUV400颜色编码格式，并通过改变特征图的组合方式，提高传统视频编码器对待编码特征图数据的编码效率。该技术可以根据组合方式的不同进一步划分为以下三种重打包/反重打包方式：

**1) 特征图默认顺序叠加**

特征图默认顺序叠加方式中，特征图的每个通道对应传统视频编解码器输入数据中的一个视频帧，特征图通道顺序为原始顺序，保持不变。特征图的高、宽被填充至符合传统视频编解码器输入要求的高度和宽度。

**2) 特征图指定顺序叠加**

特征图指定顺序叠加方式中，特征图的每个通道对应传统视频编解码器输入数据中的一个视频帧，同时，固定第一通道，余下通道按照与前一通道二范数最小的规则排列。特征图的高、宽被填充至符合传统视频编解码器输入要求的高度和宽度。

**3) 特征图默认顺序平铺**

特征图默认顺序平铺方式中，特征图的多个通道平铺拼接成一个二维数组，并作为传统视频编解码输入数据中的一个视频帧。拼接后数组的高、款被填充至符合传统视频编解码器输入要求的高度与宽度。拼接顺序为原始特征图通道的顺序，由数组宽方向优先，高方向依次排列，当前帧铺满后再创造下一帧继续平铺，直到特征图所有通道平铺完毕。

**4) 特征图指定顺序平铺**

特征图指定顺序平铺方式中，特征图的多个通道平铺拼接成一个二维数组，并作为传统视频编解码输入数据中的一个视频帧。拼接后数组的高、款被填充至符合传统视频编解码器输入要求的高度与宽度。按照指定的特征图通道顺序进行视频帧的拼接，当前帧铺满后再创造下一帧继续平铺，直到特征图所有通道平铺完毕。

**三、主要试验[或验证]情况分析**

高效编码同时保持较好任务性能是本标准的核心特征之一，下面给出使用本标准特色技术进行深度特征图编码在压缩率和信息损失上的表现。测试数据如表1所示，为使用VGGNet-16模型基于ImageNet数据作为输入数据抽取的深度网络特征。

**表 1 测试数据说明**

|  |  |  |
| --- | --- | --- |
| **深度特征图类型** | **深度特征图尺寸** | **深度特征图数据体积** |
| **conv1** | 224×224×64 | 12.25M |
| **pool1** | 112×112×64 | 3.0625M |
| **conv2** | 112×112×128 | 6.125M |
| **pool2** | 56×56×128 | 1568K |
| **conv3** | 56×56×256 | 3.0625M |
| **pool3** | 28×28×256 | 784K |
| **conv4** | 28×28×512 | 1568K |
| **pool4** | 14×14×512 | 392K |
| **conv5** | 14×14×512 | 98K |
| **pool5** | 7×7×512 | 392K |

测试结果数据如表2到表4所示。

**表 2 测试结果—预量化+特征图默认顺序叠加**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **QP0** | | **QP12** | | **QP22** | | **QP32** | | **QP42** | |
|  | **Comp. Rate** | **Fidelity** | **Comp. Rate** | **Fidelity** | **Comp. Rate** | **Fidelity** | **Comp. Rate** | **Fidelity** | **Comp. Rate** | **Fidelity** |
| **conv1** | 0.148 | 0.997 | 0.116 | 0.996 | 0.080 | 0.985 | 0.048 | 0.955 | 0.020 | 0.839 |
| **pool1** | 0.180 | 0.997 | 0.145 | 0.994 | 0.099 | 0.984 | 0.057 | 0.914 | 0.023 | 0.693 |
| **conv2** | 0.150 | 0.997 | 0.130 | 0.992 | 0.098 | 0.972 | 0.066 | 0.952 | 0.035 | 0.790 |
| **pool2** | 0.214 | 0.997 | 0.185 | 0.995 | 0.138 | 0.982 | 0.090 | 0.947 | 0.047 | 0.745 |
| **conv3** | 0.114 | 0.997 | 0.102 | 0.995 | 0.080 | 0.986 | 0.057 | 0.960 | 0.034 | 0.840 |
| **pool3** | 0.196 | 0.997 | 0.179 | 0.989 | 0.140 | 0.981 | 0.102 | 0.955 | 0.063 | 0.819 |
| **conv4** | 0.070 | 0.998 | 0.065 | 0.992 | 0.053 | 0.984 | 0.041 | 0.967 | 0.028 | 0.865 |
| **pool4** | 0.164 | 0.998 | 0.160 | 0.992 | 0.127 | 0.974 | 0.097 | 0.969 | 0.065 | 0.864 |
| **conv5** | 0.060 | 0.998 | 0.059 | 0.997 | 0.046 | 0.989 | 0.037 | 0.969 | 0.023 | 0.920 |
| **pool5** | 0.162 | 0.998 | 0.162 | 0.995 | 0.129 | 0.986 | 0.106 | 0.967 | 0.075 | 0.908 |

**表 3 测试结果—预量化+特征图指定顺序叠加**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **QP0** | | **QP12** | | **QP22** | | **QP32** | | **QP42** | |
|  | **Comp. Rate** | **Fidelity** | **Comp. Rate** | **Fidelity** | **Comp. Rate** | **Fidelity** | **Comp. Rate** | **Fidelity** | **Comp. Rate** | **Fidelity** |
| **conv1** | 0.148 | 0.998 | 0.100 | 0.992 | 0.064 | 0.969 | 0.028 | 0.883 | 0.006 | 0.694 |
| **pool1** | 0.182 | 0.997 | 0.124 | 0.986 | 0.078 | 0.944 | 0.032 | 0.746 | 0.005 | 0.412 |
| **conv2** | 0.155 | 0.995 | 0.116 | 0.988 | 0.082 | 0.970 | 0.046 | 0.879 | 0.012 | 0.455 |
| **pool2** | 0.220 | 0.998 | 0.165 | 0.991 | 0.116 | 0.971 | 0.062 | 0.812 | 0.014 | 0.369 |
| **conv3** | 0.118 | 0.997 | 0.093 | 0.990 | 0.069 | 0.967 | 0.043 | 0.887 | 0.013 | 0.617 |
| **pool3** | 0.208 | 0.998 | 0.164 | 0.994 | 0.124 | 0.970 | 0.079 | 0.874 | 0.023 | 0.541 |
| **conv4** | 0.073 | 0.998 | 0.059 | 0.988 | 0.046 | 0.973 | 0.032 | 0.915 | 0.012 | 0.706 |
| **pool4** | 0.178 | 0.996 | 0.142 | 0.991 | 0.108 | 0.982 | 0.076 | 0.917 | 0.028 | 0.611 |
| **conv5** | 0.057 | 0.997 | 0.046 | 0.990 | 0.036 | 0.985 | 0.026 | 0.938 | 0.015 | 0.812 |
| **pool5** | 0.152 | 0.998 | 0.127 | 0.994 | 0.099 | 0.984 | 0.078 | 0.947 | 0.053 | 0.746 |

**表 4 测试结果—预量化+特征图默认顺序平铺**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **QP0** | | **QP12** | | **QP22** | | **QP32** | | **QP42** | |
|  | **Comp. Rate** | **Fidelity** | **Comp. Rate** | **Fidelity** | **Comp. Rate** | **Fidelity** | **Comp. Rate** | **Fidelity** | **Comp. Rate** | **Fidelity** |
| **conv1** | 0.162 | 0.997 | 0.116 | 0.998 | 0.080 | 0.987 | 0.048 | 0.951 | 0.020 | 0.839 |
| **pool1** | 0.198 | 0.997 | 0.145 | 0.993 | 0.099 | 0.975 | 0.057 | 0.919 | 0.023 | 0.686 |
| **conv2** | 0.165 | 0.997 | 0.130 | 0.992 | 0.098 | 0.976 | 0.066 | 0.937 | 0.035 | 0.781 |
| **pool2** | 0.233 | 0.997 | 0.184 | 0.996 | 0.137 | 0.982 | 0.089 | 0.953 | 0.045 | 0.756 |
| **conv3** | 0.122 | 0.997 | 0.100 | 0.995 | 0.079 | 0.988 | 0.056 | 0.951 | 0.033 | 0.835 |
| **pool3** | 0.200 | 0.997 | 0.163 | 0.992 | 0.129 | 0.992 | 0.091 | 0.953 | 0.053 | 0.791 |
| **conv4** | 0.060 | 0.998 | 0.051 | 0.993 | 0.042 | 0.983 | 0.033 | 0.955 | 0.021 | 0.857 |
| **pool4** | 0.118 | 0.998 | 0.100 | 0.993 | 0.083 | 0.983 | 0.064 | 0.957 | 0.042 | 0.855 |
| **conv5** | 0.032 | 0.997 | 0.028 | 0.996 | 0.023 | 0.987 | 0.018 | 0.974 | 0.010 | 0.920 |
| **pool5** | 0.063 | 0.997 | 0.054 | 0.996 | 0.046 | 0.990 | 0.036 | 0.974 | 0.024 | 0.903 |

从表2-4中结果可以看出，本标准提出的深度特征图编码具有很好的通道间冗余去除性能，压缩效果优越，同时保持着较高的保真度。在任务性能损失不超过1%的情况下，深度神经网络模型中常用的第四、第五卷积层特征可被压缩至原始体积的5.1%和2.8%；在任务性能损失不超过2%的情况下，第四、第五卷积层特征可被压缩至原始体积的4.2%和2.3%。

**四、知识产权情况说明**

工作组中提案单位或工作组成员正在申请或拥有的自主专利构成了标准的核心技术。根据目前提案单位及工作组成员单位披露情况，北京大学与新加坡南洋理工大学共同持有专利1项(专利申请号：WO2020SG50526，专利名称：NETWORK-BASED VISUAL ANALYSIS)，工作组内联合提案单位均承诺标准所涉及的所有必要专利都提供许可。专利或潜在专利的权利人均同意将所有必要专利都纳入“视觉特征编码技术专利池”统一进行实施许可，因此这些专利将是未来视觉特征编码专利池的主要组成部分。本部分标准使用了部分已过专利保护期限的专利技术，成为公开技术。

**五、采用国际标准和国外先进标准情况**

由于目前该领域的标准均是面向单流（视频流）系统的，而本标准中的视觉特征编码，具有视频流、特征流、模型流等多流并行架构，具有端、边、云协同的视觉大数据分析处理框架，国外尚无同类标准。

**六、与现行相关法律、法规、规章及相关标准的协调性**

符合我国有关的现行法律、法规。

**七、重大分歧意见的处理经过和依据**

无。

**八、标准性质的建议**

建议发布为推荐性标准。

**九、贯彻标准的要求和措施建议**

应通过各类国家级科技计划和产业化项目资助，大力开展基于视觉特征编码结构和通信协议标准的芯片、终端、服务器以及系统的研发、试验示范系统建设、直至大规模商用，扶持视觉特征编码产业链的形成。

为了促进深度特征图编码的广泛应用和产品化，建议通过项目资助或技术手段，推广国内相关科技企业采用标准定义的视觉特征编码方式，加强国内相关科技企业、科研院所以及高校之间的协同合作。标准发布后，可向有视觉特征编码需求的各单位宣传、推荐执行本标准。促进深度特征图编码的良性发展和资源的整合利用。

**十、替代或废止现行相关标准的建议**

无。

**十一、其它应予说明的事项**

无。

团体标准《信息技术 视觉特征编码 第4部分：深度特征图》编制工作组

2023-05-10