面向视频压缩后向兼容的连续学习

陆 明 丛吾洋 黄建凯 石峻奇 马 展
（南京大学电子科学与工程学院 南京 210023）

摘 要 本文面向基于深度学习的视频压缩，探索在保证视频压缩模型后向兼容能力的前提下，通过参数微调提升预训练视频编码器在新数据或新的目标码率范围的压缩性能。后向兼容是指能够使用微调后的解码模型解码原始模型编码的码流。本文将此问题定义为视频压缩的连续学习。初步结果表明，现有解决方案（如端到端微调）无法保持编码器所需的后向兼容能力。为解决这一问题，本文提出了一种基于知识回放的训练策略，有效解决了无法正确解码旧码流的问题。此外，本文立足于多种代表性的视频编码框架，尝试优化改进不适用于连续学习的熵模型设计，使得连续学习在各类编码结构上都可以行之有效。本文针对两种实验场景进行了测试：数据连续学习和码率范围连续学习，并在不同视频编码框架上进行验证。在数据连续学习中，本文选取的模型框架可以在基准预训练模型基础上实现压缩率提升最高达4%。在码率范围连续学习中，本文选取的模型框架可以在基准预训练模型基础上实现压缩率提升最高达5%。大幅度提升了视频压缩连续学习方法的后向兼容能力。此外，消融实验分别验证了本文提出的知识回放策略的有效性。本文所提出的视频压缩连续学习方法，可以根据新数据和新码率范围微调压缩模型，相较于预训练版本可以取得更好的压缩性能，同时不会损害后向兼容的能力。

关键词 视频压缩；连续学习；后向兼容；知识回放；深度学习
中图法分类号 TP391 **DOI号** 10.11897/SP.J.1016.2026.00828

Towards Backward-Compatible Continual Learning of Video Compression

LU Ming CONG Wu-Yang HUANG Jian-Kai SHI Jun-Qi MA Zhan
（School of Electronic Science and Engineering, Nanjing Universitty, Nanjing 210023）

Abstract This paper focuses on deep learning-based video compression, aiming to enhance the compression performance of pre-trained video codecs on new data or new target bitrate ranges through fine-tuning of parameters, while ensuring the backward compatibility of the video coding models. Backward compatibility refers to the ability to decode bitstreams encoded by the original model using the fine-tuned decoder. This paper defines this challenge as continual learning for video compression. Preliminary results indicate that existing solutions, such as end-to-end fine-tuning, fail to maintain the required backward compatibility of the codec. To address this issue, we propose a knowledge replay-based training strategy that effectively resolves the problem of incorrect decoding of old bitstreams. Additionally, this study explores various representative video coding frameworks to optimize entropy models that are not suited for continual learning, ensuring that continual learning can be applied effectively across different coding structures. This work evaluates two experimental scenarios: continual learning on data and continual learning across bitrate ranges, with validation conducted on different video coding frameworks. In the data

收稿日期：2025-07-16；在线发布日期：2025-11-13。本课题得到国家重点研发计划（2022YFF0902402）、国家自然科学基金重点项目（62431011）和国家自然科学基金青年科学基金项目（C项）（62401251）资助。**陆 明**，博士，副研究员，中国计算机学会（CCF）会员，主要研究领域为图像视频处理，数据压缩和深度学习。E-mail：minglu@nju.edu.cn。**丛吾洋**，博士研究生，主要研究领域为智能视频压缩和强化学习。**黄建凯**，硕士研究生，主要研究领域为智能音视频压缩和连续学习。**石峻奇**，博士研究生，主要研究领域为模型量化，神经表示和智能视频压缩。**马 展**（通信作者），博士，教授，中国计算机学会（CCF）会员，主要研究领域为类脑视频通信和计算摄像。E-mail：mazhan@nju.edu.cn。

continual learning scenario, the selected model framework achieves a compression performance improvement of up to 4% compared to the baseline pre-trained model. In the bitrate range continual learning scenario, the selected model framework achieves a compression performance improvement of up to 5% compared to the baseline pre-trained model. Our method significantly enhances the backward compatibility of the continual learning for video compression. Additionally, ablation studies are conducted to verify the effectiveness of the proposed knowledge-replay strategy. The continual learning method for video compression proposed in this paper allows for fine-tuning the compression model based on new data and new bitrate ranges, achieving better compression performance compared to the pre-trained version, without compromising backward compatibility.

Keywords　video compression; continual learning; backward compatibility; knowledge replay; deep learning

1　引　言

近年来，基于深度学习的视频压缩技术取得了令人瞩目的发展。通过使用神经网络替换传统人为定义的算法模块，并在数据驱动下端到端联合训练，这类方法可以获得更加紧致的数据表示，从而提升整体的率失真(Rate-Distortion，R-D)性能。绝大多数现有工作都立足于离线学习模式，即神经网络模型一旦训练完成，其参数在部署后就保持固定不变。然而，真实世界中的视频应用需求往往是复杂且动态多变的，一个理想的视频编码器应该可以通过连续增量学习来适应各种应用场景。例如，在一个视频存储应用中，编码器通常是在自然视频场景上针对特定的目标码率范围进行预训练的。当遇到新的视频输入源(如视频会议、屏幕内容或卡通视频等)时，人们可能希望更新编码器参数以提升其在新数据上的表现，并支持不同的目标码率范围。这引发了一个有趣的问题：基于神经网络的视频编码器能否实现连续学习？如果可以，是否能在预训练模型的基础上带来进一步的性能提升？

要实现视频编码的连续学习，有人可能认为简单地微调预训练模型就足够了。然而，这样做会破坏编码器模型的后向兼容能力，如图1(b)所示。由于预训练的编码模块和微调后的解码模块之间的不匹配，原始编码模块生成的码流无法被新的解码模块正确解码。保持视频编码器的后向兼容能力至关重要，如果做不到这一点，保存在存储中的既有码流(或从其他设备发送来的码流)将无法被正确访问。值得注意的是，这种后向兼容问题不同于神经网络中著名的“灾难性遗忘”问题[1]。视频编码的独特属性，包括发送端编码与接收端解码的关系，以及熵编码的存在，使其区别于其他视频处理或视觉任务。因此，现有的用于视觉任务的连续学习方法[2-3]，不能直接应用于压缩任务，必须开发新的策略，使得编码器模型在适应新数据和码率范围时保持解码模块的后向兼容能力。

为了实现后向兼容，最直接的方法是在适应新数据或新码率范围时，仅修改预训练模型的编码模块[4-6]。这一策略在传统编码器中较为常见，即通常有一个标准化的解码模块，但可以设计各种复杂度或专门的编码模块以满足不同的应用需求。尽管这种方法看起来简单有效，但保持解码模块不变限制了模型对新数据和新的码率范围的适应能力，导致整体的压缩性能不佳。

本文的研究表明，在连续优化编码器的编码模块和解码模块的同时，也可以保持编码器模型具备后向兼容的能力。本文首先注意到，只要保持编码器的熵模型不变，它就能够解码旧的码流并获得潜在特征表示。基于这一观察，本文提出了一种知识回放(Knowledge Replay)的训练方案，用于在不破坏后向兼容的情况下调整编码和解码模块网络，使得模型的表达与泛化能力能够兼容已有知识和新知识。此外，本文设计了一种新的模型架构，使熵模型只消耗少量参数，从而使大部分模型参数在微调过程中是可学习的。本文还制定了两个实验场景：数据连续学习和码率范围连续学习。实验结果表明，本文提出的方法能够在不破坏模型后向兼容的情况下，提升视频编码器在新数据和新码率范围的压缩性能。

本文的主要贡献如下：

(1) 本文提出了一种基于知识回放的训练策

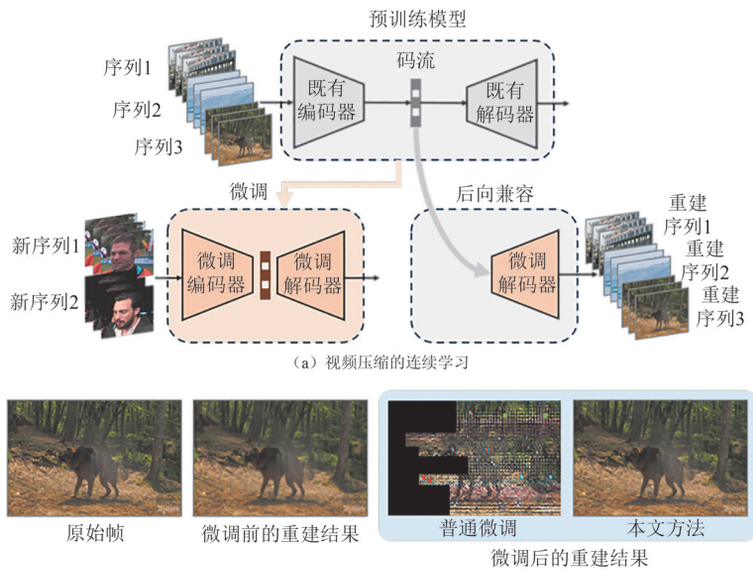


图1 视频压缩连续学习下的后向兼容问题

略，可用于连续微调视频编码器模型，并保持后向兼容的能力；

（2）本文专门针对不同的视频压缩框架重新调整了熵模型，支持视频编码器有效地连续学习；

（3）本文制定了视频压缩的两种连续学习场景：数据连续学习和码率范围连续学习。实验结果表明，在这两种场景中，本文的方法均优于基线对比方法。

2 背景和相关工作

2.1 基于深度学习的视频压缩

现有的基于深度学习的视频压缩（Deep Learning-based Video Compression）方法[7-10]基本都沿用传统的混合编码架构，即先通过运动估计得到参考帧与当前帧的运动向量，用来对齐参考帧到当前帧，再编码对齐后的参考帧与当前帧的残差并与对齐帧融合，可以得到最后的重建结果。这类方法需要前后编码两次，即先编码估计的运动向量，再编码残差。两种编码方式思路相近，一般运动向量的编码会采用较为简单的网络模型。

相比于直接相减得到的显式待编码残差[8-9]，条件残差编码[10-11]因其更低的理论熵，取得了更好的压缩性能。虽然条件残差编码并没有显式计算残差，但是这类方法同样也遵循混合编码框架，即同时编码了运动向量和隐式的残差信息，因此本文方法对这两类工作不作具体区分。

类似于图像压缩，基于深度学习的有损视频压缩方法的两阶段编码一般都通过熵约束的非线性变换编码框架[12]来进行编码。以DCVC[10]为例（如图2所示），假设$x_t \sim p_{data}$表示服从潜在数据分布的样本。在该框架中，基于神经网络的运动编码模块将运动向量映射为潜在变量v_t，而运动解码模块则将其映射回运动向量，用来指导运动对齐。上下文（残差）编码模块则将输入x_t映射为潜在隐变量y_t，而上下文（残差）解码模块将映射回重建帧$\hat{x}_t$。可学习的熵模型用来建模y_t的边际分布，运动编码部分通常也有熵模型（图中省略）。优化目标是最小化率失真（R-D）损失：

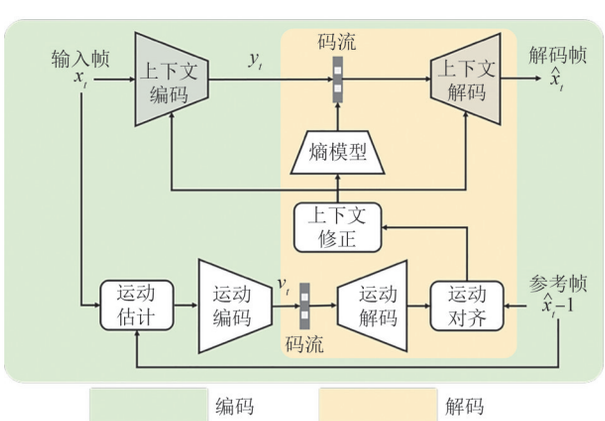


图2 视频编码框架示意图

$$\min \mathbb{E}_{x_t \sim p_{data}}[-\log_2 p_{v_t}(v_t) - \log_2 p_{y_t}(y_t) + \lambda \cdot d(x_t, \hat{x}_t)] \quad (1)$$

其中，$d(\cdot)$用来度量失真（如均方误差），λ是在码率

和失真之间进行权衡的拉格朗日因子，优化过程涉及编码模块、解码模块以及熵模型的网络参数。该框架也可以扩展到可变码率压缩[13-14]，其中编码模块、解码模块和熵模型条件依赖于λ。在可变码率训练中，整个模型参数根据不同的$\Lambda\sim p_\Lambda$进行优化：

$$\min \mathbb{E}_{x_t\sim p_{data},\ \Lambda\sim p_\Lambda}[-\log_2 p_{v_t}(v_t|\Lambda)-\log_2 p_{y_t}(y_t|\Lambda)+\Lambda\cdot d(x_t,\hat{x}_t)] \tag{2}$$

无论传统视频编码，还是近十年基于深度学习的视频编码，对于未见内容高效编码的兼容支持一直都是一个挑战。标准组织通常依赖定义一个新的标准或同一标准下的不同档次来实现对新数据的高效编码，费时费力。如屏幕内容编码的制定耗费了3年以上的时间。而基于深度学习的视频压缩还没有对新数据持续学习的相关解决方案，而这正是本文研究的创新点。与本文工作最相关的现有研究方向主要是内容自适应的视频压缩，其目标是根据每个不同的视频序列，自适应地调整编码网络及解码网络以应对新的视频场景或新的目标码率范围采取新的特征变换处理。对应的解决方案包括编码侧优化和解码侧适应。编码侧优化方法[4-6]是直接在编码过程中针对新的隐空间变量重新进行率失真优化。解码侧适应方法[15-17]则额外编码一个高效的神经网络模块，在解码侧执行以改进解码过程。其中许多方法需要在编码过程中进行复杂的迭代优化计算，在实际部署时会带来额外的计算负载。

本文的研究范围在如下几个方面不同于内容自适应视频压缩：(1)本文的目标是在不引入额外参数的情况下，逐步地微调优化编码器模型的参数；(2)相较于测试推理时对每个视频序列进行优化，本文的方法采用一次性训练的策略，且在测试推理时不引入额外的计算开销。本文的研究与内容自适应视频压缩相辅相成，两者结合可以进一步提高编码器的性能。

2.2　连续学习与知识回放

连续学习(Continual Learning)，又称增量学习(Incremental Learning)或终身学习(Lifelong Learning)，是机器学习的一个重要分支，其核心目标是使模型能够像人类一样，在一系列任务中持续积累知识，同时避免在学习新任务时对已学任务性能的显著下降，即克服“灾难性遗忘”(Catastrophic Forgetting)问题。根据任务序列中任务边界、数据分布及类别划分方式的不同，连续学习场景通常被划分为任务增量学习、领域增量学习和类别增量学习等典型设定。近年来，连续学习在图像分类[1]、图像识别[18]和语义分割[19]等视觉任务上得到了广泛研究。这些任务旨在连续学习新任务的同时，不遗忘已经学习到的知识。在现有的连续学习方法中，基于知识回放的方法[20-22]尤为有效，它通过开辟额外的内存缓冲区，存储从已学任务中选取的样本数据，并与新任务数据结合，在连续学习过程中执行知识回放。本文提出的知识回放训练方法遵循了类似的原则。然而，视频压缩与典型的计算机视觉任务在原理上有着很大的区别，现有的连续学习策略并不能直接应用于视频压缩任务。因此，在本文的问题场景中需要根据视频压缩任务量身定制一种新的知识回放策略。

3　问题阐述

3.1　视频压缩的连续学习

给定一个预训练的可变码率视频压缩模型，其码率范围由拉格朗日因子确定，即$\lambda\in[\lambda_{low}^{(0)},\lambda_{high}^{(0)}]$。为了简化描述，下文中将不再具体区分运动编码和残差编码，而是通过引入符号$f_{enc}^{(0)}$，$f_{dec}^{(0)}$和$f_{em}^{(0)}$分别表示预训练模型的编码模块、解码模块和熵模型。如图2所示，$f_{enc}^{(0)}$包括了运动估计、运动向量编解码模块、运动对齐和残差编码模块，$f_{dec}^{(0)}$包括了运动向量解码模块、运动对齐和残差解码模块，$f_{em}^{(0)}$则包括了运动向量编码和残差编码的熵模型。如图3(a)上半部分所示，$X_{test}^{(0)}$表示待压缩的测试视频序列，其编码后生成相应的码流$b_{test}^{(0)}$。这种基于离线训练的视频编码系统无法根据现实场景的变化来动态优化模型参数，其编码性能会大大受限。因此，一个理想的编码器需要具备连续学习的能力以累积知识，提升在新数据场景和新码率范围的性能。视频压缩的连续学习问题可以定义为：给定原始训练数据$X_{train}^{(0)}$下预训练的视频压缩模型，通过微调模型参数，使其在兼容既有码流的条件下，在新数据和目标码率范围上获得率失真性能更优的视频压缩模型，其中，目标训练样本由$X_{train}^{(1)}$给出，目标码率范围由拉格朗日因子$\lambda\in[\lambda_{low}^{(1)},\lambda_{high}^{(1)}]$确定。同样地，本文使用$f_{enc}^{(1)}$，$f_{dec}^{(1)}$和$f_{em}^{(1)}$来表示新的模型的各个部分。如图3(a)下半部分，连续学习可以进一步解耦为两个子问题：1)后向兼容：新模型能够完全解码旧模型生成的既有码流(如图3(b)上半部分)；2)增量学习：新模型在新的数据和码率范围上获得相比旧模型更高的压缩效率(如图3(b)下半部分)。

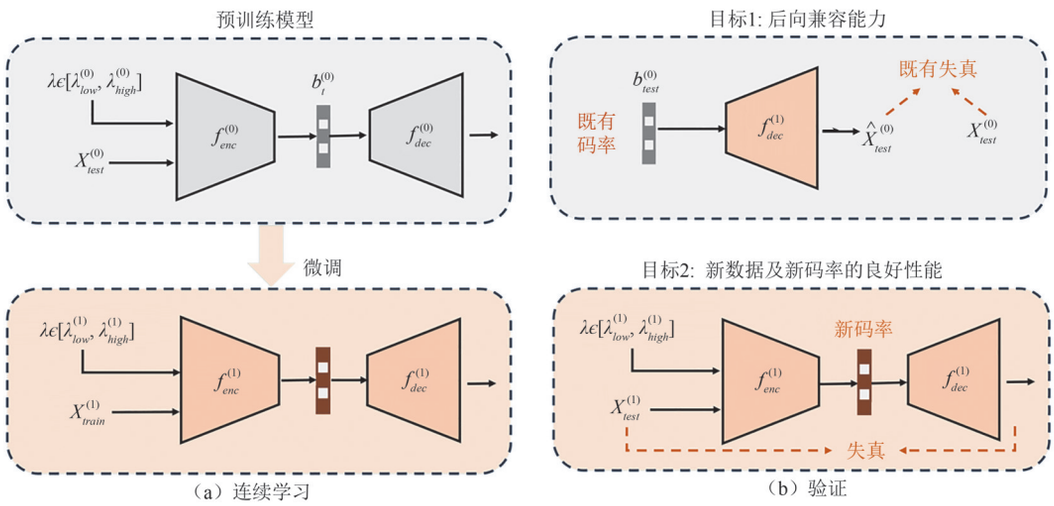


图3 问题阐述

3.2 熵解码的后向兼容

典型的连续学习方案致力于端到端地微调模型参数，以实现对目标的增量感知。然而，在视频编码系统中，完全端到端地微调预训练模型会破坏模型的后向兼容能力，如图1所示。具体地，不同于分类、检测等单一目标导向的任务，视频编码通常面向率-失真双目标联合优化，致力于在给定的保真度下最小化其编码码率。当前基于变分自编码器的智能视频编码系统是对视频信号进行概率分布建模，因此正确解码的必要条件是获得码流中每个符号（Symbol）正确的概率分布。其概率分布在旧模型编码时由熵模型$f_{em}^{(0)}$推理得到，而微调后的熵模型$f_{em}^{(1)}$改变了概率分布，从而破坏了后向兼容性。根据Hong等人[23]在之前的工作中指出的那样，即使千分之一的概率估计误差都可能造成熵解码的失败。

3.3 微调中冻结熵模型

基于上述分析，在视频编码的连续学习中，熵模型f_{em}需要固定不变，从而使得解码模块能够从既有码流中无损地恢复原有视频的隐特征表示。当满足这一必要条件时，连续学习问题可以表述为有约束的最优化求解：保证既有知识稳定的前提下，持续优化编码器f_{enc}和解码器f_{dec}以学习新知识（目标数据域和目标码率范围）。

本文通过对现有视频编码框架的调研，目前的熵模型通常会在多帧优化中与参考帧建立联系，编解码模块的更新对参考帧的改变会导致熵模型的预测分布发生变化。因此，为了有效避免冻结熵模型参数以外造成的后向不兼容问题，本文尝试调整现有熵模型结构，在对原有视频压缩方法性能基本没有影响的情况下，很好地保证了编码器的后向兼容能力。

4 本文方法

4.1 知识回放引导的连续学习

遵循第三章的符号定义，本文将预训练模型表示为$f_{enc}^{(0)}$，$f_{dec}^{(0)}$和f_{em}，其原始训练数据和原始码率范围分别由$X_{train}^{(0)}$和$\lambda\in[\lambda_{low}^{(0)},\lambda_{high}^{(0)}]$确定。其中，熵模型$f_{em}$在连续学习中被冻结，故省略其上标。知识回放引导的连续学习可以定义为受既有知识约束的率失真最优求解过程，本文通过构建拉格朗日方程求解该优化问题，该方程由约束项和目标项联合给出。

具体地，目标项$\mathcal{L}_{new}$定义为目标新数据$X_{train}^{(1)}$和目标码率范围$\lambda\in[\lambda_{low}^{(1)},\lambda_{high}^{(1)}]$下的标准率失真联合损失，即

$$\mathcal{L}_{new}=\mathbb{E}[R^{(1)}+\Lambda^{(1)}\cdot D^{(1)}] \tag{4}$$

$$R^{(1)}=-\log_2 p_Y\left(Y^{(1)}|\Lambda^{(1)}\right)-\log_2 p_V\left(V^{(1)}|\Lambda^{(1)}\right) \tag{5}$$

$$D^{(1)}=d(X_{train}^{(1)},\hat{X}_{train}^{(1)}) \tag{6}$$

其中，$\Lambda^{(1)}$是区间$[\lambda_{low}^{(1)},\lambda_{high}^{(1)}]$中的随机变量。其概率密度$p_\Lambda^{(1)}$决定$\lambda$在训练中是如何进行采样的。$R^{(1)}$包含了新的运动向量隐特征表示$V^{(1)}$的编码码率和新的残差隐特征表示$Y^{(1)}$的编码码率。直观地说，最小化$\mathcal{L}_{new}$等效于调整模型参数去适应新的数据分布和码率范围，但是并不足以保证后向兼容。

因此，本文引入约束项$\mathcal{L}_{KR}$，通过既有数据$X_{train}^{(0)}$和编解码器$f_{enc}^{(0)}$，$f_{dec}^{(0)}$和f_{em}进行知识回放，来约束目标项的求解空间，避免模型过拟合造成对先前知识的灾难性遗忘，从而保证后向兼容。具体地，本文首先使用$f_{enc}^{(0)}$编码$X_{train}^{(0)}$，得到相应的视频隐特征表示$V^{(0)}$和$Y^{(0)}$，以及对应的码流$b_{train}^{(0)}$。在解码时，相应的重

建视频 $\hat{X}_{train}^{(0)}$ 由当前新的解码模块 $f_{dec}^{(1)}$ 解码 $V^{(0)}$ 和 $Y^{(0)}$ 得到。值得注意的是，本文固定了熵模型，但是在对每一个当前帧解码的过程中，参考帧也均由新解码模型得到，这样才能保证在实际部署时彻底摆脱对旧编码模型的依赖。因此，本文中知识回放约束项可以定义为

$$\mathcal{L}_{KR}=\mathbb{E}[\Lambda^{(0)}\cdot D^{(0)}] \tag{7}$$

$$D^{(0)}=d(X_{train}^{(0)}, f_{dec}^{(1)}(f_{enc}^{(0)}(X_{train}^{(0)}))) \tag{8}$$

这里，期望由 $\hat{X}_{train}^{(0)}$ 和 $\Lambda^{(0)}$ 给出，其中 $\Lambda^{(0)}\sim p_{\Lambda}^{(0)}$ 控制 λ 在知识回放中是如何进行采样的。特别地，由于既有码流是固定的，因此约束项 $\mathcal{L}_{KR}$，完全由原始数据下的失真损失给出，与码率无关。通过知识回放，新模型在对新的目标视频高效编码的同时，能够后向兼容既有视频码流。

综上所述，该拉格朗日方程由目标项和约束项共同决定，如图4所示。

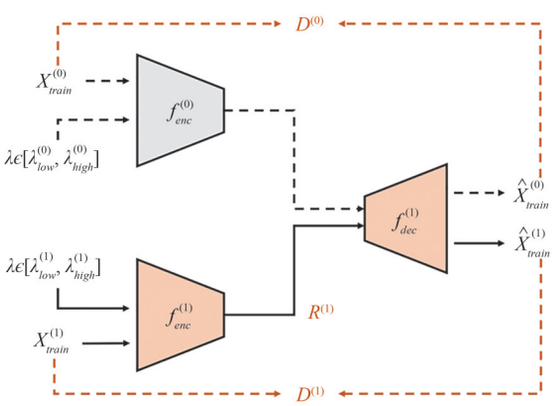


图4　基于知识回放的训练策略

$$\min(1-\alpha)\cdot\mathcal{L}_{new}+\alpha\cdot\mathcal{L}_{KR}, \tag{9}$$

其中，超参数 $\alpha\in[0, 1]$ 固定为常数，加权知识回放约束项。调节 α 可以实现目标新数据/新码率范围与原始数据/码率范围的不同权衡，即可以实现增量感知与后向兼容的权衡。值得注意的是，约束项 $\mathcal{L}_{KR}$ 依赖于既有知识 $f_{enc}^{(0)}$ 和 $X_{train}^{(0)}$，故本文默认既有码流和既有训练数据在微调过程中仍然可以获得。该假设的立足点在于，本文所提出的基于知识回放的连续学习方案，对于用来回放的旧数据没有过于严苛的要求。在进行持续学习优化前，我们可以首先从既有码流的解码结果中获取之前被编码的数据，从而得到编码器适应的旧数据类型，按照该数据类型我们可以寻找相似的开源数据集或搜集近似的数据用于回放训练，而不严格要求跟原始训练集保持一致。本文不对训练资源进行限制，并假设连续学习的过程中能够访问全部数据。本文所提的知识回放引导的连续学习策略是泛化的，可以集成到现有的各种智能视频编码模型中，实现后向兼容的连续学习。

4.2　模型结构调整

如前所述，冻结熵模型对于实现连续学习下视频编码器的后向兼容是必需的。然而，要完全冻结视频编码器的熵模型，不仅要保证熵模型部分的参数不更新，还需要考虑没有时序信息进一步传入到熵模型中。由于时序特征在编解码模块微调后会发生变化，一旦引入到熵模型中，这同样会造成熵模型预测的分布发生变化。

因此，本文尝试修改代表性视频编码工作DCVC系列来解决该问题，其他方法按照本文所述的方式，同样可以解决这个问题。如图2所示，通常情况下，经过上下文修正后的对齐特征可以作为时序先验带来概率估计准确度的提升，从而被引入到熵模型中。然而，由于编解码器在对新数据学习的过程会发生微调，从而会改变生成的对齐特征。而这也会导致即使冻结熵模型后，概率分布的估计仍然会出现偏差。因此，本文设计了双支路熵估计模块对上下文编码中的均值和方差进行分别预测来消除因参考特征变化导致对概率估计的影响。由于运动编码模块中并没有引入参考特征，因此我们仅修改原始上下文编码的熵编码部分（如图5(a)所示）。修改后的熵编码部分如图5(b)所示，本文尝试做如下修改：1）保留了对齐特征和当前帧的先验特征联合对均值的预测；2）去掉了对齐特征对方差的预测，只通过当前帧的先验特征来估计方差。实验证明，由于对齐特征在上下文编码/解码过程中都会引入以降低需要编码的残差信息量，同时还对当前帧

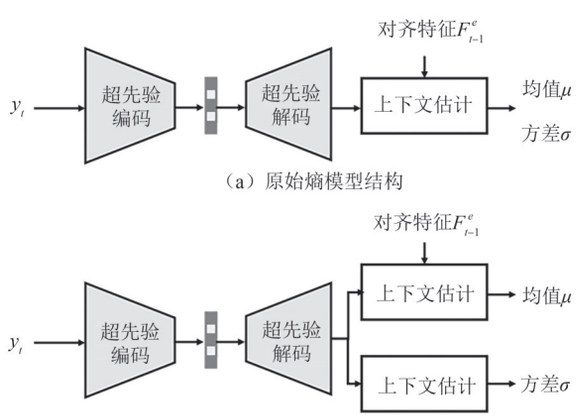


图5　熵模型结构

隐特征的均值做了较好的预测，因此即使在方差估计中去掉了对齐特征先验，仍然可以取得与原始模型相近的压缩性能。

5 实 验

5.1 实验设置

本文考虑了视频压缩中的两种连续学习场景：数据连续学习和码率连续学习。在数据连续学习场景中，本文基于预训练模型在新数据集上进行微调；而在码率连续学习场景中，模型将在新的码率范围内进行微调（可以是更高或更低）。以下是本文所使用的数据集和评估指标，详细配置如表1所示。

表1 本文实验所采用的详细配置

	码率范围 $[\lambda_{low}^{(0)}, \lambda_{high}^{(0)}]$	训练集 $X_{train}^{(0)}$	测试集 $X_{test}^{(0)}$
预训练模型	[256, 2048]	Vimeo90k	UVG
	码率范围 $[\lambda_{low}^{(1)}, \lambda_{high}^{(1)}]$	训练集 $X_{train}^{(1)}$	测试集 $X_{test}^{(1)}$
数据连续学习	[256, 2048]	Vimeo90k VFHQ	VFHQ
码率连续学习（低→高）	[256, 4096]	Vimeo90k	UVG
码率连续学习（高→低）	[64, 2048]	Vimeo90k	UVG

（1） 训练数据集

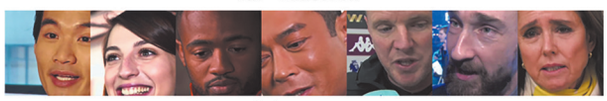
• Vimeo90k[24]：这是用于训练基线模型的通用视频数据集。该数据集包含64,612个分辨率为480×270的视频序列。在训练过程中，本文将每帧图像随机裁剪成256×256的图像块输入到网络。

• VFHQ[25]：这是数据连续学习使用到的高质量人脸视频数据集，常用于视频超分辨率等视觉任务的训练。数据集共有16,000段高保真视频序列，主要包含各种采访场景。本文使用公开数据集缩放到512×512后的版本，并在训练时将每帧图像随机裁剪成256×256的图像块输入到网络。

• SCVQA[26]：这是数据连续学习使用到的屏幕内容数据集。该数据集包括1600段远程办公、演示场景、游戏视频等内容。本文选取数据集的前90%的视频作为训练数据集，并且在训练时将每帧图像随机裁剪成256×256的图像块输入到网络。

三种训练数据集在内容、运动模式、分辨率帧率等数据特性分布上表现出了显著的差异，其各自的场景示例如图6所示。没有特别的说明，在本文中主要实验部分的新数据集（5.3节）均采用VFHQ数据集，SCVQA数据集上的实验结果详见消融实验部分（5.4节）。

（a）Vimeo90k



（b）VFHQ

（c）SCVQA

图6 训练数据集样本

（2） 测试数据集

• UVG[27]：这是常用的视频编码器测试数据集。该数据集包含7个分辨率为1920×1080的，长度为600帧（个别序列为300帧）的高分辨率视频序列，覆盖不同内容、运动特性的视频，在测试时将会被填充到1920×1088分辨率输入网络。

• VFHQ[25]：我们使用了VFHQ官方公开的测试集进行人脸新数据的测试。

• SCVQA[26]：我们选取了数据集后10%的视频作为屏幕内容新数据的测试集，并且经过验证训练和测试数据集之间并没有数据重叠。

三种测试数据集的场景示例如图7所示。

（a）UVG

（b）VFHQ

（c）SCVQA

图7 测试数据集样本

（3） 性能指标

本文使用以下标准指标来衡量压缩性能：

• Bits per pixel（bpp）：每像素比特数。

• Peak Signal-to-Noise Ratio（PSNR）：峰值信噪比（本文中所有测试在RGB色彩空间下进行）。

• BD-Rate：率-失真压缩性能的通用衡量指标。

（4） 评估方法

正如第2、3节所述，本文通过以下两个目标来

评估每个方法：

① 后向兼容能力：本文使用微调后的模型对预训练模型编码后既有码流进行解码，以获得重建结果。bpp用于表示既有码流的码率大小（这是一个独立于微调策略的常数），PSNR则用于计算重建视频和原始视频之间每一帧平均的像素差异。

②新数据与新码率范围下的性能：本文基于新的数据或新的码率范围计算bpp和PSNR指标，并计算与x265编码器*very slow*模式的BD-Rate，命令行指令如下：

```
# 编码
ffmpeg -i input.yuv \
-c:v libx265 \
-preset veryslow \
-crf 23 \
-x265-params "profile=main:level-idc=4.1:high-tier=1" \
-pix_fmt yuv420p \
-f mp4 output.hevc
# 解码
ffmpeg -i output.hevc \
-c:v libx265 \
-pix_fmt yuv420p decoded.yuv
```

(5)　实验环境及配置

本文涉及的所有训练和测试都是基于PyTorch 2.6.0框架和Python 3.11环境，在2张NVIDIA RTX 3090 GPU上进行的。本文所有的训练均采用Adam优化器，设置初始学习率为1e-4，经过10个epoch之后减少到1e-5，批次大小固定为4，最终经过15到20个epoch后达到模型收敛。

5.2　对比方法

在实验中，本文选择了三种不同范式的编码器模型：DVC[8]、DCVC[10]和最新的DCVC-RT[28]。这三种模型是基于深度学习的视频压缩中的代表性模型，基于这三种模型的实验结论可以推广到其他现有模型。由于需要进行码率连续学习，本文按照Duan等人[14]的工作为定码率的DVC和DCVC构建了可变码率版本，并按照各自原文中所介绍的方法进行训练。得到的模型分别称为DVC-VR和DCVC-VR。而DCVC-RT自身支持可变码率功能，我们仅对于其熵模型做了简单修改以方便在调整适应新数据/码率范围的同时兼容旧数据/码率范围，具体地，DCVC-RT原本的熵模型集成了时空上下文融合模块，同时利用时间、空间上下文预测需要编码的特征的均值和方差。在本文中，为了兼容旧码流准确的概率估计，本文修改其均值预测使用时间和空间上下文信息，方差预测则仅使用空间上下文信息，以适应编码器、解码器变化后时间上下文发生的偏移。

本文对每个模型设计以下微调策略，并比较其性能，以验证本文提出的连续学习方法的有效性：

• 预训练模型（Pre-trained）：不使用新数据或码率进行微调，仅最简单的可变码率基线模型。

• 仅微调编码器（FT Enc）：仅微调编码器，并冻结其他参数。由于熵模型和解码器保持不变，保证了后向兼容能力。

• 微调编码器和解码器（FT Enc & Dec）：微调除熵模型参数外的编码器和解码器等所有模型参数。由于解码器的变化，可能会破坏编码器模型在旧数据上的表现。

• 本文的方法（KR）：使用知识回放策略微调模型的编码器和解码器，冻结熵模型参数。

通过这些微调策略的对比，本文能够评估不同方法在数据连续学习和码率连续学习上的表现。

5.3　实验结果

数据连续学习：表2显示了模型在Vimeo90k上预训练并在VFHQ上进行微调后的结果，其中既有码流代表微调前模型编码生成的码流。本文首先比较DVC-VR模型的微调策略。仅微调编码器（FT Enc）对新数据的BD-Rate改善不大（从11.54%降至10.64%），这是因为只微调编码器只能减少摊销差距（Amortization Gap）[5]，而不能显著提高模型容量。当同时微调编码器和解码器（FT Enc & Dec）时，新数据的BD-Rate有了更大的提升（从11.54%降至7.62%），这表明更新解码器对新数据性能的重要性。然而，这带来了后向兼容时重建性能的问题，旧比特流的BD-Rate大幅增加（从15.04%骤增到204.91%）。这表明解码器虽然对新数据的拟合效果更好，但同时并没有保留对旧比特流的重建能力。使用本文的知识回放策略（DVC-VR w/ KR），编码器能够在不牺牲后向兼容能力（15.08%）的前提下实现新数据性能上的提升的BD-Rate（从13.29%降至12.49%）。基于DCVC-RT的各类训练方法的主观效果及对应的编码结果（编码码率和重建PSNR值）对比如图8所示。值得注意的是，使用本文的方法微调基本不影响旧比特流的性能，这是其他策略无法实现的。总体来看，知识回放策略明显优于其他策略。这些观察结果在DCVC-VR和DCVC-RT中也保持一致，证明了知识回放在数据增量学习中的有效性。

表2 编码器基于不同训练方式在新数据上的性能(评价指标:BD-Rate(%),基线:x265(very slow),下同)

	既有码流(Vimeo90k)	新数据(VFHQ)	平均值
DVC-VR, Pre-trained	15.04	11.54	13.29
DVC-VR, w/ FT. Enc	15.04	10.64	12.84
DVC-VR, w/ FT. Enc & Dec	204.91	7.62	106.27
DVC-VR, w/ KR(本文方法)	15.08	9.91	**12.49**
DCVC-VR, Pre-trained	−17.88	−25.23	−21.56
DCVC-VR, w/ FT. Enc	−17.88	−25.98	−21.93
DCVC-VR, w/ FT. Enc & Dec	64.92	−32.28	16.37
DCVC-VR, w/ KR(本文方法)	−19.26	−29.15	**−24.21**
DCVC-RT, Pre-trained	−74.29	−80.15	**−77.22**
DCVC-RT, w/ FT. Enc	−74.29	−80.91	-77.60
DCVC-RT, w/ FT. Enc & Dec	−10.14	−85.83	-47.99
DCVC-RT, w/ KR(本文方法)	−72.12	−84.14	**−78.13**

码率范围连续学习:表3展示了码率范围连续学习的实验结果。在码率连续学习实验中,本文省略了仅微调编码器(FT Enc)的基线,因为单独微调编码器无法有效扩展任何用来评估的模型的码率范围。从DVC-VR模型开始,本文观察到同时微调编码器和解码器(FT. Enc & Dec)能够扩展模型的码率范围,并在新码率下表现出良好的性能。然而,旧比特流的性能显著下降,类似于数据连续学习实验中的观察结论。相比之下,使用本文的知识回放策略(DVC-VR w/ KR),模型在保持后向兼容能力的同时,能够在新码率下获得具有竞争力的BD-Rate(低→高为23.82%,高→低为18.08%)。DCVC-VR和DCVC-RT的结果也显示出类似的模式。总体而言,本文测试的模型结合KR策略表现均优于基线,验证了其在码率连续学习上的有效性。无论是从低码率到高码率还是从高码率到低码率的码率连续学习,上述观察结果都保持一致。

(a) Pretrained (b) w/FT.Enc (b) w/FT.Enc & Dec (b) w/KR(本文方法)

图8 不同训练策略的主观效果对比(上方:旧数据;下方:新数据)

熵模型结构的调整:对于在熵模型概率建模过程中引入了时序上下文参考特征的编码器,如DCVC-RT,本文通过仅使用时序上下文预测分布均值,解耦了其均值和方差的预测过程。表4展现了这种结构调整带来的性能和复杂度变化,可以看出,在经过本文提出的持续学习方案后,原始结构无法实现对旧数据的正常编解码,这是因为更新后的时序信息影响了原始熵模型的概率预测,出现了编解码不一致。而本文为了实现后向兼容对熵模型修改后,不仅可以实现对旧数据的正常解码,还能提升在新数据上的压缩性能。这得益于熵模型修改后可以正常实现只是回放训练,从而实现模型更好地收敛。与此同时,修改熵模型后并不会造成明显的复杂度、编解码速度的增加。此外,对于DVC这类简单的编码结构,其时序参考信息仅包含在参考帧,并没有引入到对当前帧的概率建模过程。因此,我们只需要固定熵模型,仅调整编解码网络就可以实现后向兼容。

SCVQA数据集上的实验结果:为了进一步验证本文所提出的连续学习方法对于数据内容差异明显的新旧数据的兼容性,本文进一步在原始旧数据为通用视频Vimeo90k数据集,新数据为屏幕内容数据集SCVQA上进行了持续学习验证。我们在DCVC编码器上的验证结果如表5所示。可以发

表3　两种编码器基于不同训练方式实现新码率兼容的性能

	既有码流($\lambda\in[256,2048]$)	新码率($\lambda\in[256,4096]$)	平均值
DVC-VR, Pre-trained	15.04	/	/
DVC-VR, w/ FT. Enc & Dec	153.84	10.22	82.03
DVC-VR, w/ KR(本文方法)	29.85	17.79	**23.82**
DCVC-VR, Pre-trained	−17.88	/	/
DCVC-VR, w/ FT. Enc & Dec	46.88	-22.26	12.31
DCVC-VR, w/ KR(本文方法)	−18.52	-27.95	**−23.24**
	既有码流($\lambda\in[256,2048]$)	新码率($\lambda\in[64,2048]$)	平均值
DVC-VR, Pre-trained	15.04	/	/
DVC-VR, w/ FT. Enc & Dec	189.22	11.03	100.13
DVC-VR, w/ KR(本文方法)	20.01	16.15	**18.08**
DCVC-VR, Pre-trained	−17.88	/	/
DCVC-VR, w/ FT. Enc & Dec	30.37	-24.33	3.02
DCVC-VR, w/ KR(本文方法)	−20.04	-29.77	**−24.91**

表4　消融实验结果：探索熵模型结构调整的影响

	旧数据性能(BD-Rate %)	新数据性能(BD-Rate %)	计算复杂度(KMACs/pixel)	编码时间(ms)	解码时间(ms)
原始结构	编解码不匹配	−81.67	421.31	9.18	10.15
修改熵模型	−72.12	−84.14	423.56	9.18	10.88

现，在与原始训练数据差异更加明显的屏幕内容数据上，预训练模型的压缩性能出现了明显的劣化(BD-Rate退化到19.97%)；同样地，采用微调编码器的方法也出现了明显的性能降低(12.15%)；同时微调编解码器的方案虽然在新数据上有不错的性能，但是无法保证旧数据集的性能。而采用了本文提出的后向兼容的连续学习方案后，有效实现了保障新数据性能的同时兼容旧知识，最终获得了新旧数据上平均−12.47%的码率节约，远超预训练模型和其他微调方案。

表5　消融实验结果：SCVQA数据集上的实验结果

	既有码流(Vimeo90k)	新数据(SCVQA)	平均值
DCVC-VR, Pre-trained	−17.88	19.97	1.05
DCVC-VR, w/ FT. Enc	−17.88	12.15	-2.87
DCVC-VR, w/ FT. Enc & Dec	89.17	−11.86	38.66
DCVC-VR, w/ KR(本文方法)	−16.34	-8.60	**−12.47**

5.4　实验分析：知识回放

本文提出的知识回放策略的有效性已在上述的实验中得到了验证。接下来本文将通过在DCVC-VR上的消融实验分析以下两个问题，进一步探究知识回放策略的有效性来源。

哪些因素有助于保持后向兼容能力？ 在数据连续学习中，本文的知识回放策略中有两个组成部分可能有助于后向兼容能力：知识回放的训练数据和损失函数。为了分析每个组成部分的贡献，本文冻结模型的熵模型参数，并分别对其进行微调。在分析时，本文从“微调编码器和解码器的模型(FT Enc & Dec)”开始，逐一评估这两个组成部分。表6展示了结果。配置0是在训练集Vimeo90k上的预训练模型，配置1是在VFHQ上“微调编码器和解码器”的基线对比模型，该模型虽然在新数据上取得了更好的率失真性能，但是在旧数据上的测试结果大幅度下降。通过两种训练数据一起微调(配置2)，后向兼容性能显著提高(BD-Rate从64.92%降至46.05%)，但仍大幅度低于预训练模型。当应用知识回放损失(配置3)在新数据上微调时，新数据的测试效果得到了提升的同时，也一定程度上保留了后向兼容的性能。而当联合两种训练数据集一起进行知识回放(配置4)时，旧码流上的性能甚至超过了预训练模型，且新数据上也表现出更好的结果。因此，本文得出结论，回放的训练数据和损失函数都对后向兼容性能有贡献，但知识回放损失更为重要。

参数对知识回放的影响是什么？ 本文在数据连续学习中训练模型时，调整了参数α的值，控制了每次训练迭代中回放数据的比例。表7展示了结果。首先，随着α从0增加到1，旧码流上的性能逐步改善(即BD-Rate降低)。如果当α为0时，不进行任何重放，模型仅使用新数据进行训练；当α为1时，

表6 消融实验结果:探索后向兼容能力的来源

配置	微调		BD-Rate w.r.t. x265.(*very slow*)		
	数据	KR损失	旧码流	新数据	平均
0	/	/	−17.88	−25.23	−21.56
1	VFHQ		64.92	−32.28	16.37
2	VFHQ + Vimeo90k		46.05	−28.74	8.66
3	VFHQ	√	−12.32	−26.83	−19.58
4	VFHQ + Vimeo90k	√	−20.04	−29.77	**−24.91**

模型仅使用回放数据进行训练。结果与直觉一致:更多的回放数据导致更好的后向兼容性能。在新数据上的表现则相反:随着α的增加,新数据的性能逐步下降,这也符合直觉。总体来看,α在[0.25, 0.75]范围内的表现是相当的。本文可以得出结论,本文的方法对α的选择并不敏感,并且可以在后向兼容和新数据性能之间实现良好的权衡。本文的实验默认选择α为0.5,但在实际应用中,α的选择可以作为超参数,根据应用需求进行调整。

表7 消融实验结果:探索不同α值对性能的影响

α	0.0	0.25	0.5	0.75	1.0
旧码流	64.92	12.71	−19.26	−21.90	**−24.67**
新数据	**−32.28**	−30.44	−29.15	−25.11	−18.38
平均值	16.37	−8.87	**−24.21**	−23.51	−21.53

5.5 讨 论

本文的实验表明,通过使用提出的训练策略,基于深度学习的视频编码器可以以后向兼容的方式适应新的数据和码率。除了在连续学习应用中的意义外,这一发现还为相关研究提供了启示,例如关于基于深度学习的视频编码的标准化问题,应当对神经网络的哪些部分进行标准化。本文的研究结果提出了一个可能的方向:只需标准化熵模型而不是整个模型架构及其参数,其他组件(如编码、解码模块网络)可以通过后向兼容的训练策略(如本文的知识回放)进行微调,而不会导致编解码不匹配。

6 结 论

本文提出了一种基于知识回放的训练策略来实现视频压缩的连续学习。本文的知识回放策略使现有编码器能够适应新数据和目标码率范围,同时确保先前压缩的码流仍然可解码。通过在不同编码框架下广泛的实验验证,本文得出了明确的结论:基于深度学习的视频编码器可以在保证后向兼容的情况下连续学习,并在新数据、新码率范围和旧码率范围上均能实现压缩性能的提升。

参 考 文 献

[1] Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu A A, Milan K, Quan J, Ramalho T, Grabska-Barwinska A, Hassabis D, Clopath C, Kumaran D, Hadsell R. Overcoming catastrophic forgetting in neural networks. Proceedings of the National Academy of Sciences, 2017, 114(13): 3521-3526

[2] De Lange M, Aljundi R, Masana M, Parisot S, Jia X, Leonardis A, Slabaugh G, Tuytelaars T. A continual learning survey: defying forgetting in classification tasks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021,44(7): 3366-3385

[3] Wang L, Zhang X, Su H, Zhu J. A comprehensive survey of continual learning: theory, method and application. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024,46(8): 5362-5383

[4] Campos J, Meierhans S, Djelouah A, Schroers C. Content adaptive optimization for neural image compression//Proceedings of the CVPR 2019 Workshops (CLIC workshop). Long Beach, USA, 2019, 2: 1-5

[5] Yang Y, Bamler R, Mandt S. Improving inference for neural image compression//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada, 2020: 573-584

[6] Gao C, Xu T, He D, Wang Y, Qin H. Flexible neural image compression via code editing//Proceedings of the 36th International Conference on Neural Information Processing Systems. Red Hook, USA, 2022: 12184-12196

[7] Jia C, Ma H, Yang W, Ren W, Pan J, Liu D, Liu J, Ma S. Video processing and compression technologies. Journal of Image and Graphics, 2021, 26(6): 1179-1200
(贾川民, 马海川, 杨文瀚, 任文琦, 潘金山, 刘东, 刘家瑛, 马思伟. 视频处理与压缩技术. 中国图象图形学报, 2021,26(6): 1179-1200)

[8] Lu G, Ouyang W, Xu D, Zhang X, Cai C, Gao Z. Dvc: an end-to-end deep video compression framework//2019 IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 10998-11007

[9] Liu H, Shen H, Huang L, Lu M, Chen T, Ma Z. Learned video compression via joint spatial-temporal correlation exploration//Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA, 2020: 11580-11587

[10] Li J, Li B, Lu Y. Deep contextual video compression//Proceedings of the 35th International Conference on Neural Information Processing Systems. Red Hook, USA, 2021: 18114-18125

[11] Sheng X, Li J, Li B, Li L, Liu D, Lu Y. Temporal context

mining for learned video compression. IEEE Transactions on Multimedia, 2022, 25: 7311-7322

[12] Ballé J, Chou P A, Minnen D, Singh S, Johnston N, Agustsson E, Hwang S, Toderici G. Nonlinear transform coding. IEEE Journal of Selected Topics in Signal Processing, 2020, 15(2): 339-353

[13] Choi Y, El-Khamy M, Lee J. Variable rate deep image compression with a conditional autoencoder//2019 IEEE/CVF International Conference on Computer Vision. Seoul, Republic of Korea, 2019: 3146-3154

[14] Duan Z, Lu M, Ma J, Huang Y, Ma Z, Zhu F. Qarv: quantization-aware resnet vae for lossy image compression. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(1): 436-450

[15] Pan G, Lu G, Hu Z, Xu D. Content adaptive latents and decoder for neural image compression//17th European Conference on Computer Vision. Tel Aviv, Israel, 2022: 556-573

[16] Shen S, Yue H, Yang J. Dec-adapter: exploring efficient decoder-side adapter for bridging screen content and natural image compression//2023 IEEE/CVF International Conference on Computer Vision. Paris, France, 2023: 12841-12850

[17] Tsubota K, Akutsu H, Aizawa K. Universal deep image compression via content-adaptive optimization with adapters//2023 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, USA, 2023: 2528-2537

[18] Jiang J, Deng W. Facial expression recognition improved by continual learning. Journal of Image and Graphics, 2020, 25(11): 2361-2369
(江静, 邓伟洪. 持续学习改进的人脸表情识别. 中国图象图形学报, 2020, 25(11): 2361-2369)

[19] Cermelli F, Mancini M, Bulo S R, Ricci E, Caputo B. Modeling the background for incremental learning in semantic segmentation//2020 IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 9230-9239

[20] Liu Y, Su Y, Liu A A, Schiele B, Sun Q. Mnemonics training: multi-class incremental learning without forgetting//2020 IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 12242-12251

[21] Lopez-Paz D, Ranzato M A. Gradient episodic memory for continual learning//Proceedings of the 36th International Conference on Neural Information Processing Systems. Red Hook, USA, 2017: Curran Associates Inc.: 6470-6479

[22] Rebuffi S A, Kolesnikov A, Sperl G, Lampert C H. Icarl: incremental classifier and representation learning//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 5533-5542

[23] Hong W, Chen T, Lu M, Pu S, Ma Z. 2020. Efficient neural image decoding via fixed-point inference. IEEE Transactions on Circuits and Systems for Video Technology, 31(9): 3618-3630

[24] Xue T, Chen B, Wu J, Wei D, Freeman W T. 2019. Video enhancement with task-oriented flow. International Journal of Computer Vision, 127(8): 1106-1125

[25] Xie L, Wang X, Zhang H, Dong C, Shan Y. Vfhq: a high-quality dataset and benchmark for video face super-resolution//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans, USA, 2022: 656-665

[26] Safonov N, Rakhmanov M, Vatolin D. 2025. Screen Content Video Dataset and Benchmark//Proceedings of the 33rd ACM International Conference on Multimedia. New York, USA, 2025: 1-9

[27] Mercat A, Viitanen M, Vanne J. UVG dataset: 50/120fps 4K sequences for video codec analysis and development//Proceedings of the 11th ACM multimedia systems conference. Istanbul, Türkiye, 2020: 297-302

[28] Jia Z, Li B, Li J, et al. Towards practical real-time neural video compression//Proceedings of the Computer Vision and Pattern Recognition Conference. Nashville, USA, 2025: 12543-12552

LU Ming, Ph. D., associate researcher. His main research interests include image and video processing, data compression, and deep learning.

CONG Wu-Yang, Ph. D. candidate. His main research interests include learned video compression and reinforcement learning.

HUANG Jian-Kai, M. S. candidate. His main research interests include learned audio and video compression and continual learning.

SHI Jun-Qi, Ph. D. candidate. His main research interests include model quantization, neural representation, and learned video compression.

MA Zhan, Ph. D., professor. His main research interests include brain-inspired video communication and computing photography.

Background

This paper focuses on deep learning-based video compression and addresses a problem that has received increasing attention: enabling a pretrained video codec to continually improve its compression performance on new data or new bitrate ranges without losing backward compatibility.

Recent advances in neural video compression replace hand-crafted modules with data-driven neural networks and jointly optimize motion estimation, latent feature compression, and reconstruction. These approaches consistently achieve superior rate-distortion performance compared with traditional codecs. However, current learned codecs follow a fixed, offline training paradigm—the model parameters remain unchanged after deployment.

In real-world applications, video content and bitrate requirements are dynamic and diverse. For example, a codec pretrained on natural scenes may later need to handle screen content, video conferencing, or extremely low-bitrate storage scenarios. Ideally, the codec should be able to learn from new data incrementally, instead of being retrained from scratch.

A straightforward solution is to fine-tune the pretrained codec on new data or new bitrate ranges. However, this naïve strategy destroys backward compatibility: after parameter updates, the decoder can no longer correctly decode bitstreams generated by the original encoder. In contrast to typical continual learning tasks where forgetting only reduces accuracy, losing backward compatibility in video compression leads to irreversible decoding failure—previously stored or transmitted bitstreams become unusable. Existing adaptive video compression methods either optimize encoder-side parameters during inference or add auxiliary modules on the decoder side, but none ensure continual learning while preserving the ability to decode old bitstreams. Furthermore, the entropy model in a learned codec tightly couples the encoder and decoder; modifying it during fine-tuning fundamentally changes the latent representation and breaks compatibility.

To address this limitation, this paper formulates continual learning for video compression: the codec should be incrementally fine-tuned on new data or new bitrate ranges, while preserving the decodability of previously encoded bitstreams. This paper proposes a knowledge replay-based training strategy and redesigns entropy modeling to ensure stable latent representation during fine-tuning. As a result, the codec can evolve after deployment, improving compression performance on new data, yet maintaining full backward compatibility with historical bitstreams.