适用于机器图像编码的通用可变码率自适应

Yui Tatsumi*FSE*研究生院，早稻田大学日本东京
yui.t@fuji.waseda.jp

Hiroshi Watanabe*FSE*研究生院，早稻田大学日本东京
hiroshi.watanabe@waseda.jp

摘要—机器图像编码（ICM）旨在在有限比特率下保留机器视觉所需的任务相关信息。然而，大多数ICM模型都是针对固定工作点训练的，由于不同任务和模型设计的训练目标各不相同，将其扩展到可变码率压缩颇具挑战。本文提出一种适用于可变码率ICM的与损失无关的蒸馏框架。该框架使用预训练的固定码率ICM模型作为冻结的教师模型，同时通过重要性自适应量化和基于蒸馏的优化，使学生模型学会可变码率行为。这样，只需一个模型即可实现连续的比特率控制，同时保留任务相关潜在分量，且无需原始的特定任务损失或下游任务标签。基于现有ICM模型开展的实验表明，与固定码率模型相比，所提方法能够实现连续的比特率控制，并取得有竞争力的识别性能。

arXiv:2611.29914v1 [eess.IV] 23 Nov 2026

索引词—**机器图像编码，可变码率**

I. 引言

学习式图像压缩（LIC）凭借优于 JPEG [1]和 VVC-Intra [2]等传统编解码器的率失真性能，受到了越来越多的关注。近年来，在交通监控和智慧农业等应用中，机器视觉模型越来越多地用于分析图像。在这些场景中，感知质量并非始终至关重要；相反，压缩表征应保留与下游识别任务相关的信息。这推动了机器图像编码（ICM）的发展。

在实际的 ICM 场景中，所需比特率可能因网络状况和应用需求而异。然而，许多学习式压缩模型只能在固定工作点运行，并且需要针对不同的比特率目标分别训练模型。这会增加存储和训练成本。为解决这一问题，研究人员提出了多种面向人类的压缩可变码率学习图像压缩方法。相比之下，ICM 的可变码率控制仍未得到充分研究。由于面向人类和面向机器的编码具有不同的优化目标，ICM 需要专门的可变码率方法。

开发可变码率ICM的一大挑战在于，现有ICM模型采用多种不同的目标进行训练，例如任务损失、语义监督以及任务相关潜在表征保留。尽管这种多样性是面向机器的图像压缩的一大优势，却也阻碍了依赖特定训练目标的可变码率机制在不同ICM模型中的应用。

本文提出了一种面向机器图像编码的、基于重要性自适应蒸馏的可变码率框架。该方法采用一个预训练的固定码率ICM模型作为教师，并训练学生模型，使其在实现比特率控制的同时保留教师模型学到的压缩特性。我们将估计得到的熵模型尺度参数用作潜在表征重要性，并据此调整量化步长和蒸馏损失。这样，学生模型便能在低码率下更有效地保留任务相关潜在分量。由于该框架仅依赖预训练ICM模型的输出，因此无需额外的任务标签、显著性图或对下游识别模型进行修改，即可应用于不同的ICM架构。

II. 相关工作

A. 机器图像编码

早期的学习式图像压缩方法主要面向人类视觉压缩 [3]- [7]。大多数方法采用 J. Ball´ 于等人 [3]提出的超先验架构，对量化潜变量的分布进行建模。自回归熵模型通过利用空间依赖关系 [4] 和通道依赖关系 [5]等多种依赖关系，进一步提升了性能。这些方法通常采用率失真损失进行训练，其中 λ 用于控制比特率与重建失真之间的权衡。由于训练期间 λ 固定不变，因此必须训练并存储多个独立模型，以支持不同的比特率目标。

与面向人类的压缩不同，面向机器的图像压缩采用多种针对机器视觉任务的优化目标。基于任务损失的方法直接优化下游任务的性能，例如目标检测或分割 [8]- [10], ；而基于 ROI 的方法则利用分析驱动的重要性图，为任务相关区域分配更多比特 [11]- [13]。近年来，研究者开发了任务泛化的面向机器图像压缩方法，以减少对特定下游任务的依赖 [14]- [18]。例如，SA-ICM [14] 能够保留目标边界信息用于任务无关的机器分析，而 UG-ICM [15]采用基于 CLIP [19]的监督和偏好条件

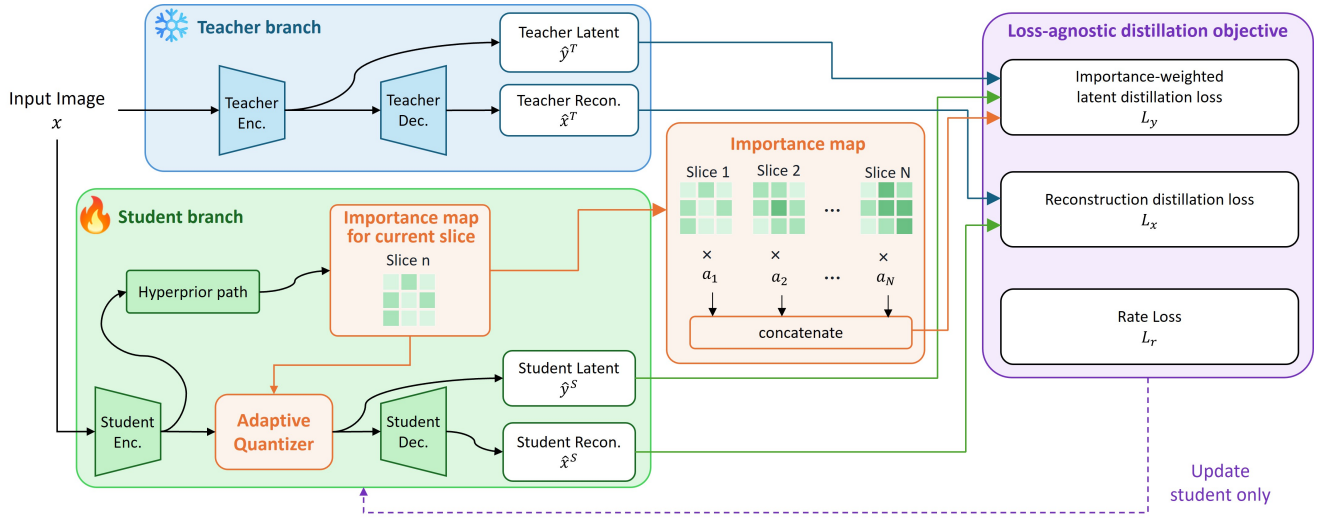


图 1. 所提出训练框架的概览。

解码，通过统一码流同时支持人类感知和通用机器分析。

B. 可变码率压缩

可变码率LIC已广泛应用于面向人类的压缩 [20]- [29]，但在面向机器的图像压缩方面仍未得到充分研究。M. Song等 [30]使用质量图控制空间比特分配。尽管该编解码器可适用于任务感知压缩，但其训练目标是图像重建，而非面向机器的目标。J. Liu等 [31] 提出了一种率失真认知可控编解码器，但该方法依赖专用架构和面向认知的目标。因此，这些方法无法为具有不同优化目标的多种面向机器的图像压缩模型提供通用解决方案。AQVR-ICM [32]利用熵模型尺度参数，无需重新训练即可对预训练ICM模型进行免训练的可变码率控制。然而，由于模型并未针对可变码率运行进行优化，其性能受到限制，尤其是在低码率下。这些局限促使我们提出一种可训练、与损失函数无关的可变码率框架，适用于具有不同目标的预训练ICM模型。

C. 知识蒸馏

知识蒸馏已广泛用于将模型能力从规模较大、性能较强的教师模型迁移至精简的学生模型，主要用于模型压缩和高效推理 [33]。近期研究还将蒸馏应用于 LIC，以降低计算复杂度，同时保持率失真性能 [34]。然而，现有的基于蒸馏的压缩方法大多侧重于模型加速或率失真保持，而很少关注将蒸馏用于其他目标，例如 ICM 中的通用可变码率自适应。

III. 提出的方法

A. 总体框架

本文提出了一种适用于可变码率ICM的重要性自适应、与损失函数无关的框架。如图1所示，该框架将预训练的固定码率ICM模型调整为支持可变码率压缩，同时保留其从原始优化目标中学到的行为。

所提出的框架由两个部分组成。首先，自适应量化器根据由熵模型尺度参数导出的重要性图调整量化步长，从而在控制比特率的同时保留任务相关潜在分量。其次，与损失函数无关的蒸馏目标使框架能够适配不同的预训练ICM模型，而无需特定任务损失或识别标签。此外，潜在蒸馏损失还通过重要性图进行加权，以优先保留任务相关潜在分量，并减轻识别性能下降。

B. 尺度感知重要性估计

所提出的方法根据熵模型尺度参数估计潜在表征重要性。在仅超先验模型中，尺度参数直接从解码后的超潜变量中获得；而自回归模型则按照解码顺序依次预测这些参数。因此，重要性估计必须遵循相应的自回归结构。这里，我们介绍通道级自回归模型的公式，并可按照空间自回归模型的解码顺序将其扩展到空间自回归模型。

在第 III-B 节和第 III-C 节中，所有潜变量和熵模型参数均由学生模型获得。在按通道自回归熵模型中，潜在表示沿通道维度拆分为 N 个切片，并按顺序处理。对于第 n 个切片（$n = 1, 2, \ldots\ldots, N$），熵模型会预测均值和尺度参数，我们使用尺度参数 $\sigma^{(n)}$ 估计每个潜在分量的重要性。由于编码器和解码器都会根据解码后的超潜变量 $\hat{z}$ 以及之前重建的切片，按照相同的自回归顺序预测 $\sigma^{(n)}$，因此无需传输额外的重要性图，就能获得相同的重要性图和量化步长。我们首先计算稳定化尺度表示：

$$s^{(n)} = \log\left(1 + \left|\sigma^{(n)}\right|\right). \tag{1}$$



(a) ground truth (b) compressed image (SA-ICM) (c) detection result (SA-ICM) (d) compressed image (Ours) (e) detection result (Ours)

(f) ground truth (g) compressed image (UG-ICM) (h) detection result (UG-ICM) (i) compressed image (Ours) (j) detection result (Ours)

图 2。 解码图像示例及 YOLOv5 目标检测 结果。

为同时捕获通道和空间趋势，我们计算空间平均的通道尺度分数 $\bar{s}_c^{(n)}$、通道平均的空间尺度分数 $\bar{s}_{h,w}^{(n)}$，以及 $s^{(n)}$的全局平均值 $\bar{s}^{(n)}$。切片级重要性图随后定义为

$$I_{c,h,w}^{(n)} = \frac{\bar{s}_c^{(n)}}{\bar{s}^{(n)} + \epsilon} \cdot \frac{\bar{s}_{h,w}^{(n)}}{\bar{s}^{(n)} + \epsilon}. \tag{2}$$

式（2）中定义的切片级重要性图 $I^{(n)}$ 用于对当前切片进行自适应量化。在自回归熵模型中，此阶段无法确定各切片之间的相对重要性。然而，对于潜在蒸馏，在处理完完整的潜在表示后，即可获得所有切片级尺度分数。因此，我们进一步为蒸馏损失引入切片级权重：

$$a_n = \frac{\bar{s}^{(n)}}{\frac{1}{N}\sum_{n'=1}^{N} \bar{s}^{(n')} + \epsilon}. \tag{3}$$

然后，通过拼接加权后的切片级重要性图，得到损失函数的重要性图：

$$I = conc\left(a_1 I^{(1)}, a_2 I^{(2)}, \ldots, a_N I^{(N)}\right). \tag{4}$$

C. 重要性自适应可变码率量化

我们使用重要性图 $I^{(n)}$，为当前切片中的每个潜变量元素分配一个自适应量化步长。令 $\tilde{I}^{(n)} = I^{(n)}/(\bar{I}^{(n)} + \epsilon)$ 表示归一化切片级重要性图。首先计算依赖重要性的相对量化因子：

$$R^{(n)} = \left(\tilde{I}^{(n)} + \epsilon\right)^{-\gamma}, \tag{5}$$

其中，γ 控制基于重要性的比特分配强度。重要性值越大，相对因子越小，因此重要的潜变量分量会进行更精细的量化。

为了控制总体比特率，我们引入一个标量码率控制参数 α。全局码率控制因子和保护因子定义如下：

$$G(\alpha) = \exp(k\alpha), \qquad P(\alpha) = 1 - \exp(-c|\alpha|), \tag{6}$$

其中，k 控制比特率对 α 的敏感度，而 c 控制保护因子 $P(\alpha)$ 随着 α 偏离零而增加的速度。$P(\alpha)$ 会抑制原始教师工作点附近的重要性自适应量化。最终的自适应量化步长为：

$$\Delta^{(n)} = G(\alpha)\left[1 + P(\alpha)\left(R^{(n)} - 1\right)\right]. \tag{7}$$

改变 α 会改变全局码率，同时根据重要性图重新分配比特。

给定第 n 个切片的潜在表示 $y^{(n)}$、预测均值 $\mu^{(n)}$ 和自适应步长 $\Delta^{(n)}$，量化过程如下：

$$\hat{y}^{(n)} = \text{round}\left(\frac{y^{(n)} - \mu^{(n)}}{\Delta^{(n)}}\right)\Delta^{(n)} + \mu^{(n)}. \tag{8}$$

由于量化区间取决于 Δ，因此对每个潜变量元素，似然是在宽度为 Δ_i 的区间上计算的：

$$p_{\hat{y}|\hat{z}}(\hat{y}_i|\hat{z}, \Delta_i) = \int_{\hat{y}_i - \Delta_i/2}^{\hat{y}_i + \Delta_i/2} \mathcal{N}(t|\mu_i, (\sigma_i)^2)dt. \tag{9}$$

我们不对超先验路径应用自适应量化，因为对超潜变量 z 进行粗量化会降低尺度预测的精度，而 $\hat{z}$ 对整体码率控制影响较小，正如先前研究所述 [28], [32]。

D. 与损失无关的蒸馏目标

所提出的方法采用基于蒸馏的目标来训练学生模型，而不依赖原始的 ICM损失函数。令 T 和 S 分别表示教师模型和学生模型。训练损失函数由潜变量蒸馏、重建蒸馏和码率损失组成：

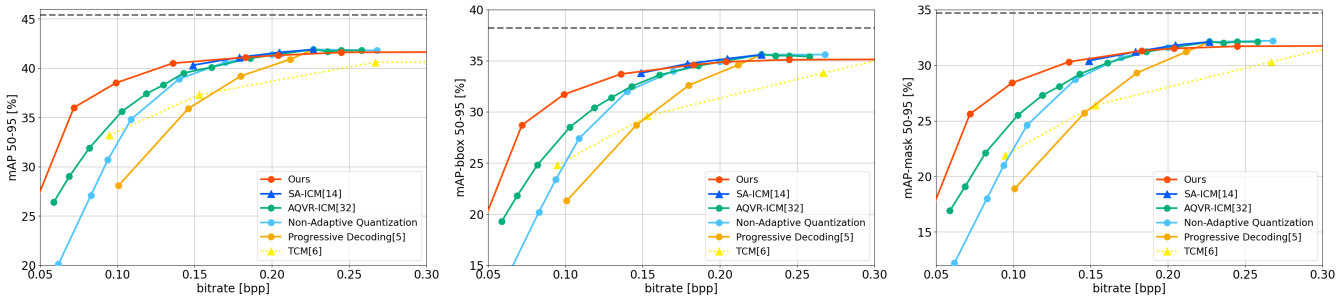
$$\mathcal{L} = \lambda_y \mathcal{L}_y + \lambda_x \mathcal{L}_x + \lambda_r \mathcal{L}_r. \tag{10}$$

~ ¯使用归一化重要性图 $I = I/(I + \epsilon)$，潜变量蒸馏损失和重建蒸馏损失定义为

$$\mathcal{L}_y = \mathbb{E}\left[\tilde{I}\left(\hat{y}^S - \hat{y}^T\right)^2\right], \tag{11}$$

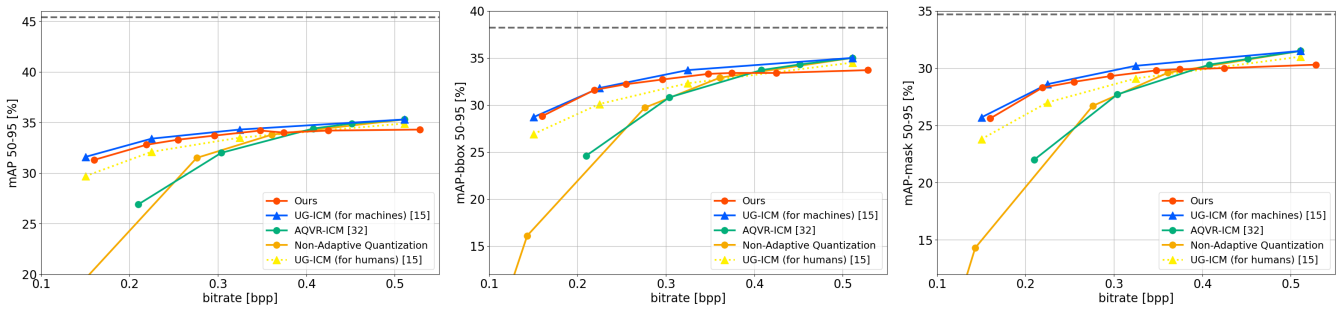
$$\mathcal{L}_x = \mathbb{E}\left[\left(\hat{x}^S - \hat{x}^T\right)^2\right]. \tag{12}$$

其中，$\mathcal{L}_r$ 表示根据学生模型的 $\hat{y}^S$ 和 $\hat{z}^S$的熵模型似然计算得出的估计每像素比特数。



(a) 使用 YOLOv5 进行目标检测　(b) 使用 Mask R-CNN 进行目标检测　(c) 使用 Mask R-CNN 进行实例分割

图 3. 基于 SA-ICM 的图像压缩在不同识别任务上的性能。实线/点线分别表示面向机器/面向人类的编解码器，圆形/三角形标记分别表示可变码率/固定码率结果。灰色点线表示未压缩图像上的结果。



(a) 使用 YOLOv5 进行目标检测　(b) 使用 Mask R-CNN 进行目标检测　(c) 使用 Mask R-CNN 进行实例分割

图 4. 基于 UG-ICM [15] 在不同识别任务上的图像压缩性能。实线/点线分别表示面向机器/面向人的编解码器，圆形/三角形标记分别表示可变码率/固定码率结果。灰色点线表示未压缩图像上的结果。

IV. 实验

我们使用 COCO训练集进行训练，并使用 COCO验证集进行评估 [35]。学生模型使用 Adam优化器训练 10 个 epoch，学习率为 1×10^{-4}，批大小为 8。对于每个基础模型，使用最高码率预训练检查点作为固定教师模型，并使用相同的权重初始化学生模型。式 (10) 中的损失权重为 $\lambda_y = 1.0$、$\lambda_x = 0.03$ 和 $\lambda_r = 0.05$。训练期间，码率控制参数从 $\alpha \in \{0, 0.5, 1.0, 2.0\}$ 中采样，并在 $\alpha \in \{-0.5, 0, 0.3, 0.5, 1.0, 1.5, 2.0, 3.0\}$处进行评估。报告的 bpp 值根据 $\hat{}y$ 和 $\hat{}z$的熵模型似然估算，不包括算术编码开销和 α 信令。由于 α 是每张图像仅发送一次的单个标量，其信令开销很小。

所提出的方法适用于任何具有超先验结构的固定码率 ICM 模型。我们在两个训练目标不同的模型上对其进行验证：SA-ICM [14]和基于 MLIC++ [7]的 UG-ICM [15]。对于 SA-ICM，我们使用 (k，c，γ) =(1。0，1。0，0。5)，对于 UG-ICM，使用 (0。35，0。25，0。25)，两者均采用 $\epsilon = 10^{-6}$。在基于 UG-ICM 的实验中，教师模型和学生模型的解码器偏好参数均固定为面向机器设置（$\beta = 1$）。使用 YOLOv5 [36] 和 Mask R-CNN [37] 评估目标检测的识别性能，并使用 Mask R-CNN 评估实例分割的识别性能。

我们将所提出的方法与相应的固定码率ICM模型、AQVR-ICM [32]，以及非自适应量化进行比较；后者采用均匀步长缩放，且不使用自适应控制。对于基于 SA-ICM 的实验，我们还纳入了渐进式解码 [5] 和 TCM [6]，即 SA-ICM 的基础编解码器。对于基于 UG-ICM 的实验，我们还比较了 UG-ICM 面向人的输出。

图2展示了解码图像和YOLOv5检测结果示例。与固定码率模型的输出相比，所提出的方法在低得多的比特率下保留了任务相关信息。图3和图4分别展示了以SA-ICM和UG-ICM为基础模型时的Rate-mAP曲线。所提出的方法通过单个模型实现了连续比特率控制，并在不同任务和基础架构下保持了具有竞争力的识别性能。这些结果表明，重要性自适应蒸馏能够有效保留任务相关潜在分量，同时以与损失函数无关的方式实现灵活的码率控制。

V. 结论

本文提出了一种重要性自适应蒸馏框架，用于将固定码率ICM模型转换为可变码率模型。通过使用熵模型尺度参数作为重要性线索，所提出的方法能够在无需原始损失函数或下游标签的情况下，自适应地进行量化和蒸馏。实验表明，所提出的框架适用于具有不同目标的现有ICM模型。结果还表明，单个经过自适应调整的模型能够在较宽的比特率范围内保持接近固定码率ICM模型的识别性能。未来的工作包括自动码率分配。

致谢

本文是早稻田大学特别研究项目资助下开展的研究成果之一（项目编号：2026R-021）。

参考文献

[1] G. K. Wallace, 《JPEG静态图像压缩标准》, IEEE消费电子汇刊, 第38卷, 第1期, 第18–34页, 1992年。[2] 通用视频编码, 标准ISO/IEC 23090-3, 2020年。[3] J. Ball´e、D. Minnen、S. Singh、S. J. Hwang和N. Johnston, 《采用尺度超先验的变分图像压缩》, 国际学习表征会议（ICLR）, 2018年, 第1–10页。[4] D. Minnen、J. Ball´e和G. Toderici, 《学习式图像压缩的联合自回归与分层先验》, 第32届神经信息处理系统会议（NeurIPS）, 2018年, 第10794–10803页。[5] D. Minnen和S. Singh, 《学习式图像压缩的通道级自回归熵模型》, 2020年IEEE国际图像处理会议（ICIP）, 2020年, 第3339–3343页。[6] J. Liu、H. Sun和J. Katto, 《采用混合Transformer-CNN架构的学习式图像压缩》, 2023年IEEE/CVF计算机视觉与模式识别会议（CVPR）, 2023年, 第14388–14397页。[7] W. Jiang、J. Yang、Y. Zhai、F. Gao和R. Wang, 《MLIC++: 用于学习式图像压缩的线性复杂度多参考熵建模》, ACM多媒体计算、通信与应用汇刊, 第21卷, 第5期, 第1–25页, 2025年。[8] N. Le、H. Zhang、F. Cricri、R. Ghaznavi-Youvalari和E. Rahtu, 《面向机器的图像编码: 端到端学习方法》, 2021年IEEE声学、语音与信号处理国际会议（ICASSP）, 2021年, 第1590–1594页。[9] N. Le、H. Zhang、F. Cricri、R. Ghaznavi-Youvalari、H. R. Tavakoli和E. Rahtu, 《面向机器的学习式图像编码: 内容自适应方法》, 2021年IEEE多媒体与博览会国际会议（ICME）, 2021年, 第1–6页。[10] X. Shen、H. Ou和W. Yang, 《通过分析驱动的外观冗余消减实现面向机器的图像编码》, 2024年IEEE国际图像处理会议（ICIP）, 2024年, 第1883–1889页。[11] H. Choi和I. V. Bajic, 《面向目标检测的高效压缩》, 2018年IEEE声学、语音与信号处理国际会议（ICASSP）, 2018年, 第1792–1796页。[12] Z. Huang、C. Jia、S. Wang和S. Ma, 《视觉分析驱动的图像编码率失真模型》, 2021年IEEE多媒体与博览会国际会议（ICME）, 2021年, 第1–6页。[13] J. I. Ahonen、N. Le、H. Zhang、F. Cricri和E. Rahtu, 《支持感兴趣区域的面向机器学习式图像编码》, 2023年IEEE第25届多媒体信号处理国际研讨会（MMSP）, 2023年, 第1–6页。[14] T. Shindo、K. Yamada、T. Watanabe和H. Watanabe, 《利用Segment Anything进行边缘信息学习的面向机器图像编码》, 2024年IEEE国际图像处理会议（ICIP）, 2024年, 第3702–3708页。[15] K. Yin 等, 《通过CLIP监督统一编码, 以兼顾人类感知与通用机器分析》, AAAI人工智能大会论文集, 第39卷, 9525, 2025年, 第9517–9529页。[16] T. Shindo、T. Watanabe、Y. Tatsumi和H. Watanabe, 《基于轮廓特征学习与灵活目标选择的目标识别任务图像编码》, 载于IEEE Access, 第13卷, 106606–106617页, 2025年。[17] T. Shindo、T. Watanabe、K. Yamada和H. Watanabe, 《采用目标区域学习的面向机器图像编码》, 2024年IEEE第21届消费者通信与网络会议（CCNC）, 2024年, 第1040–1041页。[18] T. Shindo、T. Watanabe、Y. Tatsumi和H. Watanabe, 《Delta-ICM: 采用狄拉克δ函数熵建模的学习式图像压缩》, 2025年IEEE国际消费电子会议（ICCE）, 2025年, 第1–6页。[19] A. Radford 等, 《通过自然语言监督学习可迁移视觉模型》, 第38届国际机器学习会议（ICML）论文集, 第139卷, 第8748–8763页, 2021年。[20] G. Toderici 等, 《采用循环神经网络的可变码率图像压缩》, arXiv预印本, arXiv: 1511.06085, 2015年。[21] G. Toderici 等, 《采用循环神经网络的全分辨率图像压缩》, 2017年IEEE计算机视觉与模式识别会议（CVPR）, 2017年, 第5435–5443页。[22] N. Johnston 等, 《通过预热和空间自适应比特率改进循环网络的有损图像压缩》, 2018年IEEE/CVF计算机视觉与模式识别会议, 2018年, 第4385–4393页。[23] Y. Choi、M. El-Khamy 和 J. Lee, 《基于条件自编码器的可变码率深度图像压缩》, 2019 IEEE/CVF国际计算机视觉会议（ICCV）, 2019年, 第3146-3154页。[24] F. Yang、L. Herranz、J. v. d. Weijer、J. A. I. Guiti´an、A. M. L´opez 和 M. G. Mozerov, 《基于调制自编码器的可变码率深度图像压缩》, 发表于《IEEE信号处理快报》, 第27卷, 335, 2020年, 第331页。[25] Z. Cui、J. Wang、S. Gao、T. Guo、Y. Feng 和 B. Bai, 《具有连续码率自适应的非对称增益深度图像压缩》, 2021年IEEE/CVF计算机视觉与模式识别会议（CVPR）, 2021年, 第10527-10536页。[26] F. Mei、L. Li 和 D. Liu, 《采用自适应量化步长的可变码率学习式图像压缩》, 2023年IEEE国际视觉通信与图像处理会议（VCIP）, 2023年, 第1-5页。[27] K. Tong、Y. Wu、Y. Li、K. Zhang、L. Zhang 和 X. Jin, 《QVRF: 面向学习式图像压缩的量化误差感知可变码率框架》, 2023年IEEE国际图像处理会议（ICIP）, 2023年, 第1310-1314页。[28] F. Kamisli、F. Racap´e 和 H. Choi, 《采用多目标优化和量化重建偏移的可变码率学习式图像压缩》, 2024年数据压缩会议（DCC）, 2024年, 第193-202页。[29] J. Lee、S. Jeong 和 M. Kim, 《面向可变码率图像压缩的潜在表示选择性压缩学习》, 发表于《神经信息处理系统大会论文集》第36卷, 2022年, 第13146–13157页。[30] M. Song、J. Choi 和 B. Han, 《通过空间自适应特征变换实现可变码率深度图像压缩》, 2021年IEEE/CVF国际计算机视觉会议（ICCV）, 2021年, 第2360-2369页。[31] J. Liu 等, 《率失真认知可控的通用神经图像压缩》, 欧洲计算机视觉会议（ECCV）, 施普林格, 2024年, 第329–348页。[32] Y. Tatsumi、Z. Zeng 和 H. Watanabe, 《用于机器可变码率图像编码的免训练自适应量化》, 2026年IEEE国际消费电子会议（ICCE）, 2026年, 第1-6页。[33] G. Hinton、O. Vinyals 和 J. Dean, 《神经网络中的知识蒸馏》, arXiv预印本 arXiv: 1503.02531, 2015年。[34] Y. Chen 等, 《学习式图像压缩的知识蒸馏》, 2025年IEEE/CVF国际计算机视觉会议（ICCV）, 2025年, 第4996-5006页。[35] T. Y. Lin 等, 《Microsoft COCO: 上下文中的常见物体》, 《计算机视觉—ECCV 2014》。ECCV 2014。《计算机科学讲义》, 第8693卷, 第740-755页, 2014年。[36] G. Jocher 等, 《Ultralytics/YOLOv5: V7.0-YOLOv5 最先进的实时实例分割》, Zenodo, 2022年。[37] K. He、G. Gkioxari、P. Dollar 和 R. Girshick, 《Mask R-CNN》, 2017年IEEE国际计算机视觉会议（ICCV）, 2017年, 第2980-2988页。