压缩特征质量评估：数据集和基准

Changsheng Gao
changsheng.gao@ntu.edu.sg
Nanyang Technological University
Singapore

Wei Zhou
ZhouW26@cardiff.ac.uk
Cardiff University
UK

Guosheng Lin
gslin@ntu.edu.sg
Nanyang Technological University
Singapore

Weisi Lin
wslin@ntu.edu.sg
Nanyang Technological University
Singapore

Abstract

在资源受限的环境中广泛部署的大型模型突显了高效传输中间特征表示的需求。在这种背景下，特征编码，即将特征压缩成紧凑的比特流，成为涉及特征传输、存储和重用的场景中的关键组件。然而，这一压缩过程引入了固有的语义退化，而传统指标很难对其进行量化。为了解决这一问题，本文引入了压缩特征质量评估（CFQA）的研究问题，该问题旨在评估压缩特征的语义保真度。为了推进 CFQA 研究，我们提出了第一个基准数据集，包含来自三项视觉任务和四种特征编解码器的 300 个原始特征和 12000 个压缩特征。提供了任务特定的性能下降作为评估 CFQA 指标的真实语义失真。我们评估了三种广泛使用的指标（MSE、余弦相似度和中心化核对齐）在捕捉语义退化方面的表现。结果强调了数据集的代表性，并突出了需要更精细的指标来解决压缩特征中的语义失真细微差别。为了促进 CFQA 研究的持续发展，我们在 https://github.com/chansongoal/Compressed-Feature-Quality-Assessment 发布了数据集和所有相关的源代码。这一贡献旨在推进该领域的发展并为社区提供基础资源以探索 CFQA。

CCS Concepts

• **Information systems** → **Multimedia databases**; • **Computing methodologies** → **Image compression**; • **Human-centered computing** → *Ubiquitous and mobile computing design and evaluation methods.*

Keywords

Compressed Feature Quality Assessment, Coding for Machines, Feature Coding

ACM Reference Format:
Changsheng Gao, Wei Zhou, Guosheng Lin, and Weisi Lin. 2025. 压缩特征质量评估：数据集和基准. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym '25)*. ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 介绍

大型基础模型（例如，DINOv2，LLaMA3）在分布式和资源受限环境中的快速部署，产生了传输中间特征而非原始信号的日益增长的需求。特征编码通过压缩这些中间表示，在实现可扩展性、保护隐私和提高系统效率方面发挥了至关重要的作用。与优先考虑感知质量的传统图像或视频编码不同，特征编码的目标是保持深层表示中嵌入的任务相关语义。

然而，特征编码不可避免地会引入语义下降：即识别信息的丢失，这可能会影响后续工作的性能。这种下降与像素级的失真根本不同，通常无法通过传统的失真指标（如 MSE 或 PSNR）捕捉到。虽然任务准确性是语义质量更可靠的指标，但在实际应用中是不切实际的：下游任务可能无法访问、运行成本高，或在压缩时不可用。这些限制突显了一个紧迫且未被充分研究的问题：压缩特征质量评估（CFQA）——我们如何在不依赖下游推理的情况下，估计压缩特征的语义质量？

解决 CFQA 存在几个挑战。首先，没有标准数据集提供跨多个任务和编码器的压缩特征及其相应的任务性能。其次，现有的相似性度量缺乏对高维特征的验证，并且常常不能在任务之间推广。第三，没有统一的协议来评估给定的度量是否忠实地反映语义退化。

为了弥合这一差距，我们迈出了系统研究 CFQA 的第一步。与依赖于任务监督或任务特定编解码器的先前工作相比，我们将 CFQA 视为一个独立的问题，旨在基准测试其核心组件：数据集、指标和评估协议。

我们做出了以下贡献。

- 多任务基准数据集：我们构建了第一个 CFQA 数据集，该数据集由来自三个视觉任务（分类、分割、深度估计）和四个代表性编解码器（包括手工制作和学习的模型）的 300 个原始特征和 12,000 个压缩特征组成。该数据集使在任务、比特率和编解码器之间的语义降级进行定量分析成为可能。
- 真值语义失真推导：对于每个压缩特征，我们提供任务特定的语义失真标签（排名变化，mIoU 下降，RMSE 差异），这些是通过比较使用原始和压缩特征的任务头输

Conference acronym '25, Dublin, Ireland

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

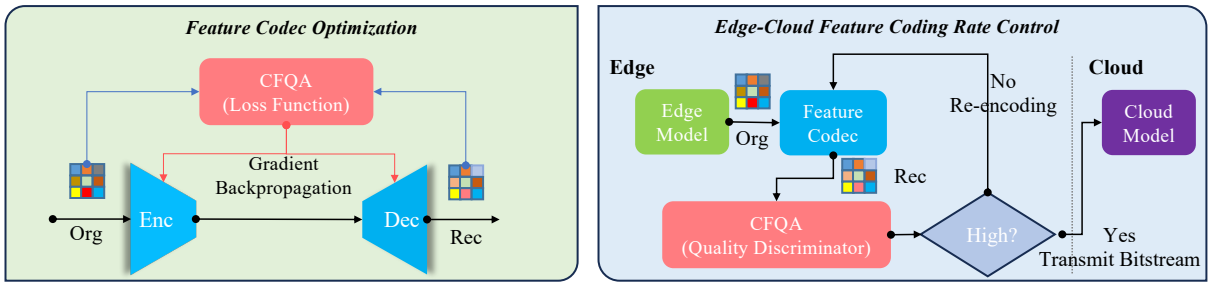

Figure 1: 压缩特征质量评估的示例应用场景。

出计算得出的。这些标签作为训练和评估质量指标的真实值。

- 现有指标评估：我们评估了三种代表性信号基础指标（MSE、余弦相似性和居中核对齐（CKA）），并分析了它们对压缩的敏感性及与任务特定退化的相关性。我们的分析揭示了它们在有效性、稳定性和失效案例方面的不同表现，为其在语义质量预测中的适用性提供了关键见解。

我们希望这项工作为未来的 CFQA 研究奠定基础，并弥合低级压缩和高级语义实用性之间的差距。通过提供一个标准化的基准和分析流程，我们的工作鼓励开发轻量化、可推广和任务无关的质量指标，以支持压缩特征在实际中的应用。

2 相关工作

2.1 特征编码

特征编码在云边协同智能场景的早期阶段受到了越来越多的关注 [1, 2, 6, 7, 9, 15, 16, 20, 24, 26, 27, 34, 35, 38, 39] 。最近，大模型特征编码引起了更多的兴趣 [19, 30] 。这些工作将特征编码扩展到了特征被传输、存储和重用的场景。

2.2 语义失真测量

语义失真测量方法分为三个主要类别: 信号保真度指标、语义保真度指标和基于任务的指标。信号保真度指标 [5, 9, 24, 29, 32] 专注于使用传统的指标如均方误差（MSE）来测量原始与重构特征之间的失真。这种方法直接借鉴自图像压缩技术，其中信号相似性被视为压缩失真。

语义保真度度量 [2, 3, 6, 8, 10, 12, 14, 22, 23] 评估语义信息的保留，其中失真是根据特定机器视觉任务中的性能下降来衡量的。这些方法提供了一种与任务更相关的失真度量，将重建质量直接与任务性能联系起来。

基于任务的指标 [1, 17, 21, 25, 27, 31, 33, 35–38] 通过任务性能直接衡量语义失真。这些方法是为特定任务而专门设计的，缺乏通用性。

给定一个预训练模型 $\mathcal{M}$ ，我们将其从输入 $x \in \mathcal{X}$ 中提取的中间特征记为 $f = \mathcal{M}(x) \in \mathbb{R}^d$ ，其中 d 是特征维度。一个特征编解码器 $\mathcal{C}$ 将 f 编码为一个紧凑的比特流并解码为压缩表示 $\hat{f} = \mathcal{C}^{-1}(\mathcal{C}(f))$ 。我们的目标是在不访问下游任务的真实标签或执行推理的情况下，评估 $\hat{f}$ 相对于 f 保留了多少语义信息。

我们将此评估任务定义为特征质量评估（FQA）：给定一对原始和压缩特征 $(f, \hat{f})$ ，估计压缩特征的语义质量 $q \in \mathbb{R}$ ，使得 q 与其在下游任务中的表现强相关。

假设一个下游任务 $\mathcal{T}$ 具有一个头网络 $h_{\mathcal{T}}$ ，产生输出 $y = h_{\mathcal{T}}(\cdot)$ 。设任务特定的性能指标记为 $\mathcal{A}_{\mathcal{T}}(\cdot)$ 。当使用压缩特征时，任务性能变为：$s = \mathcal{A}_{\mathcal{T}}(h_{\mathcal{T}}(\hat{f}))$ 。该值 $s \in \mathbb{R}$ 反映了压缩特征的实际语义效用。然而，计算 s 需要任务执行和标签，而这些可能不可用或成本高。FQA 的目标是估计一个得分 $q = Q(f, \hat{f})$ ，其中 $Q(\cdot)$ 是与任务无关的质量指标，以便在样本之间的相关性上 $q \approx s$ 。

我们在图 1 中展示了 CFQA 的两个典型应用场景。

首先，CFQA 可以直接集成到编解码器的训练过程中，作为监督信号以使压缩目标与下游任务性能保持一致。最近的研究 [15, 31] 表明，任务感知的特征编码在语义感知监督中受益显著。CFQA 提供了这种指导，而无需端到端任务标签。如图 1 左侧所示，原始和重构后的特征都被输入到 CFQA 模块中，该模块估计语义失真。然后使用失真分数生成梯度，以优化编码器和解码器。通过精心设计的 CFQA 度量，生成的编解码器即便在低比特率限制下也能学习保留语义内容。

其次，CFQA 在边缘云协作系统中至关重要，其中特征在边缘处被提取并传输到云端进行推理。由于下游模型位于云端，无法直接在边缘测量语义失真。在这种情况下，CFQA 被用作代理来估计与任务相关的降级。如 1 图右侧所示，压缩特征首先由 CFQA 模块进行评估。如果被判断为高质量，则将数据流传输到云端。否则，边缘设备将以更高的比特率重新编码特征以更好地保留语义信息。该策略确保了传输特征的可靠性，同时减少了不必要的带宽消耗。

尽管我们的工作专注于由于压缩导致的语义降级，但我们强调，CFQA 的价值不仅限于编解码器基准测试：它在各种系统中起着至关重要的作用，这些系统中特征被提取、压缩、传输、缓存或重用。

3 数据集构建

3.1 概述

我们提出的数据集旨在支持分析有损压缩引入的语义失真。它包括三个任务，300 个原始特征，以及来自四种特征编解码器的 12000 个压缩特征。特征覆盖了多种图像处理方法和多种特征提取策略。数据集的摘要信息如表 1 所示。

由于特征编码仍然关注视觉信号，我们从视觉特征开始 CFQA。我们选择 DINOv2 [30] 作为特征提取的骨干模型，因为它具有很强的泛化能力，并在通用视觉任务中得到广泛应用。为了涵盖广泛的语义复杂性范围，我们选择了三个被广泛研究的视觉任务：图像分类（Cls）、语义分割（Seg）和深度估计（Dpt）。这些任务涵盖了从粗略到精细的语义理解：Cls 关

Task	Source	Num. of Org. Feat.	Propocessing	Feature Codecs	Num of Comp. Feat.	Feature Shape	GT Distortion
Cls	ImageNet	100	Resize	HM, VTM	4000	257×1536	Rank
Seg	VOC 2012	100	Flip and Crop	Multi-task Hyperprior	4000	$2 \times 1370 \times 1536$	mIoU difference
Dpt	NYUv2	100	Flip	Task-specific Hyperprior	4000	$2 \times 4 \times 161 \times 1536$	RMSE difference

Table 1: 所提出的 CFQA 数据集的摘要信息。（详见第 3 节）

注图像级别的类别预测，Seg 在像素级引入了类对齐的空间语义，Dpt 则需要细致的几何预测。

通过包括这三个任务，我们的目标是确保语义失真在多个抽象层次上得到评估，这对于评估 CFQA 度量指标的一般适用性和敏感性至关重要。

为确保语义多样性，我们为每个任务选择 100 个具有代表性的样本。对于 Cls，我们从 100 个不同类别中抽取 100 张 ImageNet [11] 图像，每张图像都被 DINOv2 分类器正确预测。对于 Seg，我们从 Pascal VOC 2012 [13] 验证集中抽取 100 张图像，涵盖所有 20 个语义类别。对于 Dpt，我们从 NYUv2 [28] 数据集中抽取 100 张图像，覆盖所有 16 个场景。

这种采样在确保特征多样性的同时保持了适合于对照评估的数据集规模。

3.2 原始特征集合

对于每张图片，我们从 DINOv2 中指定的分割点提取任务特定的中间特征。这些点是根据其语义丰富性和与分布计算中的常用做法的对齐来选择的：对于 Cls，我们将图片调整为 224×224，并从第 40 个 $^{\text{th}}$ ViT 块提取特征，生成形状为 257×1536（256 个补丁令牌和 1 个类别令牌）的特征。对于 Seg，我们水平翻转原始图片，并从相同的第 40 个 $^{\text{th}}$ ViT 块提取特征，结果是两个 $2 \times 1370 \times 1536$ 特征。对于 Dpt，我们从第 10 个 $^{\text{th}}$，第 20 个 $^{\text{th}}$，第 30 个 $^{\text{th}}$ 和第 40 个 $^{\text{th}}$ ViT 块收集多尺度特征。原始和水平翻转后的图片生成形状为 $2 \times 4 \times 1611 \times 1536$ 的堆叠张量。

各种图像预处理和分割点模拟了真实世界的输入变化并提高了泛化能力。

3.3 压缩特征集合

为了模拟不同类型和强度的语义降级，我们通过四种编解码器压缩原始特征。所有原始特征在编码之前都被展平为二维数组，并在解码后将压缩特征重新塑造成原始格式。

手工编解码器. 我们将原始特征均匀量化到 $[0, 1023]$，然后使用两个人工设计的编解码器：HM 帧内编码（配置文件为 encoder_intra_main_rext.cfg）和 VTM 帧内编码（配置文件为 encoder_intra_vtm.cfg）。对于这两个编解码器，我们使用 YUV-400 格式进行特征编码，并将量化参数设置为 $\in 2, 4, 6, \ldots, 20$，以模拟不同的比特率水平和失真强度。

我们采用了 Hyperprior 模型 [4]，它是学习特征编码的里程碑。Hyperprior 模型在建模能力和计算效率之间提供了良好的平衡。大多数现有的压缩方法都遵循 Hyperprior 架构。因此，它能够模拟基于学习的方法所引入的语义失真。

为了分析优化效果，我们训练了两个变体：任务特定：编解码器在从一个特定任务中提取的特征上独立进行训练。多任务：编解码器在从三个任务中提取的特征上进行训练。所有编解码器均按照 [18] 中提出的训练策略进行训练，相关的 lambda 信息可以在我们发布的模型中找到。

为了评估 CFQA 指标的有效性，我们需要将它们与真正的语义失真进行比较。我们将真正的语义失真定义为压缩特征 $\hat{f}$ 与原始特征 f 之间的任务性能差异。

对于类别任务（Cls），我们计算在由 $\hat{f}$ 得到的预测逻辑值中，真实标签的排名。排名越高表示语义退化越严重。原始特征的排名值为 1。对于分割任务（Seg），我们计算从 $\hat{f}$ 和 f 预测的分割掩码之间的 mIoU 差异。对于深度任务（Dpt），我们计算从 $\hat{f}$ 和 f 预测的深度图之间的 RMSE 差异。

这些任务特定分数作为评估特征质量指标有效性的真实语义失真。

为涵盖不同类型的失真和特征之间的相似关系，我们选择了三个具有代表性的基于信号的度量：MSE、余弦相似度和居中核对齐 (CKA)。这些度量形成了从元素级到结构级失真度量的多样化基线。

我们定义一个评估协议，以评估 CFQA 指标在预测压缩特征语义失真方面的表现。在我们的实验中使用了 PLCC 和 SROCC。具体地，对于每个原始特征 f_i，我们计算这两个指标在其 10 个预测质量得分和 10 个真实语义失真上的表现。每个指标都会在所有三个任务和所有编解码器上进行评估。

3.4 速率-准确性性能分析

表 **??** 展示了手工编写的编解码器的码率-准确性表现。对于所有三个任务，这两种编解码器都表现出广泛的比特率范围及相应的准确性水平，从近乎无损的重建到显著的语义退化。这种广泛的范围强调了它们在有效模拟传统编码失真方面的实用性。然而，Dpt 任务在码率和性能上的变化较小。这可以归因于使用了多尺度特征，其中包括具有更高冗余性和减少语义抽象的较低层次，导致比特率和失真方面的波动较小。

表 **??** 显示了两种基于学习的编解码器的速率-准确性性能。与手工制作的编解码器相比，基于学习的编解码器在各种任务中表现出更加多样化的性能行为。具体来说，在 Cls 任务中，多任务编解码器优于特定任务编解码器，而在 Dpt 任务中，特定任务编解码器表现出更好的性能。这表明结合细粒度的任务特征能够增强粗粒度特征的压缩训练，而细粒度特征通过使用针对特定任务量身定制的单一分布进行训练效果更佳。

这四种编解码器全面模拟了广泛的压缩失真，使它们不仅非常适合，而且对于支持压缩特征质量评估的研究也是必不可少的。

3.5 CFQA 性能分析

3.5.1 **平均性能分析**. 表 **??** 展示了三个基准指标的平均 CFQA 性能。总体而言，三个基准指标在手工编解码器上显示出更高的 PLCC 和 SROCC 值，相较于基于学习的编解码器。这表明手工编解码器引入的语义失真更加稳定。观察到基于学习的编解码器的语义失真波动较大可能是由于训练过程中的固有变异性，因为这些模型是数据驱动的并且任务特定。此外，由于手工编解码器是基于块的，它们往往表现出更一致和可预测的失真模式。

在三个基线度量中，余弦相似度始终显示出与真实语义扭曲更高程度的线性和单调性。这是由于该度量能够捕捉特征向量之间的角度关系，使其对于诸如由 ViT 生成的高维特征集特别有效。

对于手工编写的编解码器，相较于 VTM，三个基线指标在捕捉 HM 生成的失真方面表现更好。这可能是因为 VTM 采用了更复杂的编码工具，导致失真模式更为复杂，且可能更难以预测。相反，对于基于学习的编解码器，三个基线指标中观察到了显著的性能差异。这种变异是预料之中的，因为基于学习的编码器会产生更广泛的失真类型，这使得简单的指标更难以实现线性拟合。

在这三个任务中，Seg 呈现出最复杂和最难拟合的失真。对于大多数编解码器，基线指标显示 Seg 的拟合效果比 Cls 和 Dpt 差。这反映了 Seg 中较高的语义复杂性，这对于这些指标来说更难以准确捕捉。

图 **??** 展示了 PLCC 的分布情况。为了提供更清晰的视图，所有 PLCC 值在计算频率直方图前都四舍五入到十分位。

总体而言，与基于学习的编码器相比，手工设计编码器的 PLCC 分布更为集中。这与手工设计编码器引入更一致和可预测的语义失真模式的事实相符。

在三个基线指标中，余弦相似度表现出最集中的 PLCC 分布。这个观察结果与表 **??** 和表 **??** 中报道的其较高的平均 PLCC 值一致，表明余弦相似度在不同条件下提供了更稳定的质量评估。

在所有三个指标中，Seg 的 PLCC 分布比 Cls 和 Dpt 更分散。在某些情况下，例如 CKA，在同一指标中观察到正负相关性。这突显了分割特征中失真模式的高复杂性。更广泛的分布进一步证实了现有基线指标难以准确建模这种复杂语义退化的困难。

虽然评估的指标在某些任务和编解码器中表现出中等性能，但它们也显示出显著的局限性。首先，像 MSE 这样的信号基础指标缺乏任务敏感性，往往无法反映有意义的语义退化，尤其是在分割和生成任务中。其次，余弦和 CKA 指标虽然可以捕捉结构对齐，但在学习编解码器引入的非线性失真下有时不稳定。最后，没有一个指标能在所有任务和压缩策略中一致泛化，这表明仅仅依靠手工打造的相似度可能不足以实现通用的 FQA。这些局限性突出了需要更具适应性、任务感知或学习的语义质量估计器的重要性，这是我们未来的工作考虑。

这篇论文介绍了压缩特征质量评估（CFQA）的概念，这是评估在特征被传输、存储和重用的系统中压缩特征语义退化的一个关键领域。我们提出了第一个用于 CFQA 的基准数据集，为该领域的进一步研究奠定了基础。我们评估了 CFQA 中广泛使用的指标，并为感兴趣的研究人员提供了见解。未来的研究将集中于开发能够在不同任务和特征编码策略中通用的自适应 CFQA 指标。

References

[1] Saeed Ranjbar Alvar and Ivan V. Bajić. 2019. Multi-Task Learning with Compressible Features for Collaborative Intelligence. In *ICIP*. 1705–1709. doi:10.1109/ICIP.2019.8803110

[2] Saeed Ranjbar Alvar and Ivan V. Bajić. 2020. Bit Allocation for Multi-Task Collaborative Intelligence. In *ICASSP*. 4342–4346. doi:10.1109/ICASSP40776.2020.9054770

[3] Saeed Ranjbar Alvar and Ivan V. Bajić. 2021. Pareto-Optimal Bit Allocation for Collaborative Intelligence. *IEEE Transactions on Image Processing* 30 (2021), 3348–3361. doi:10.1109/TIP.2021.3060875

[4] Johannes Ballé, David C. Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston. 2018. Variational image compression with a scale hyperprior. *ArXiv* abs/1802.01436 (2018).

[5] Yangang Cai, Peiyin Xing, and Xuesong Gao. 2022. High Efficient 3D Convolution Feature Compression. *IEEE Transactions on Circuits and Systems for Video Technology* (2022), 1–1. doi:10.1109/TCSVT.2022.3200698

[6] Zhuo Chen, Ling-Yu Duan, Shiqi Wang, Weisi Lin, and Alex C. Kot. 2020. Data Representation in Hybrid Coding Framework for Feature Maps Compression. In *ICIP*. 3094–3098. doi:10.1109/ICIP40778.2020.9190843

[7] Zhuo Chen, Kui Fan, Shiqi Wang, Lingyu Duan, Weisi Lin, and Alex Chichung Kot. 2020. Toward Intelligent Sensing: Intermediate Deep Feature Compression. *IEEE Transactions on Image Processing* 29 (2020), 2230–2243. doi:10.1109/TIP.2019.2941660

[8] Zhuo Chen, Kui Fan, Shiqi Wang, Ling-Yu Duan, Weisi Lin, and Alex Kot. 2019. Lossy Intermediate Deep Learning Feature Compression and Evaluation. In *Proceedings of the 27th ACM International Conference on Multimedia (MM '19)*. Association for Computing Machinery, New York, NY, USA, 2414–2422. doi:10.1145/3343031.3350849

[9] Hyomin Choi and Ivan V. Bajić. 2018. Deep Feature Compression for Collaborative Object Detection. In *ICIP*. 3743–3747. doi:10.1109/ICIP.2018.8451100

[10] Hyomin Choi and Ivan V. Bajić. 2021. Latent-Space Scalability for Multi-Task Collaborative Intelligence. In *ICIP*. 3562–3566. doi:10.1109/ICIP42928.2021.9506712

[11] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *CVPR*. 248–255. doi:10.1109/CVPR.2009.5206848

[12] Lingyu Duan, Jiaying Liu, Wenhan Yang, Tiejun Huang, and Wen Gao. 2020. Video Coding for Machines: A Paradigm of Collaborative Compression and Intelligent Analytics. *IEEE Transactions on Image Processing* 29 (2020), 8680–8695. doi:10.1109/TIP.2020.3016485

[13] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. 2010. The pascal visual object classes (voc) challenge. *International journal of computer vision* 88 (2010), 303–338.

[14] Ruoyu Feng, Xin Jin, Zongyu Guo, Runsen Feng, Yixin Gao, Tianyu He, Zhizheng Zhang, Simeng Sun, and Zhibo Chen. 2022. Image coding for machines with omnipotent feature learning. In *ECCV*. Springer, 510–528.

[15] Changsheng Gao, Yiheng Jiang, Li Li, Dong Liu, and Feng Wu. 2024. DMOFC: Discrimination Metric-Optimized Feature Compression. In *PCS*. 1–5. doi:10.1109/PCS60826.2024.10566361

[16] Changsheng Gao, Yiheng Jiang, Siqi Wu, Yifan Ma, Li Li, and Dong Liu. 2025. IMOFC: Identity-Level Metric Optimized Feature Compression for Identification Tasks. *IEEE Transactions on Circuits and Systems for Video Technology* 35, 2 (2025), 1855–1869. doi:10.1109/TCSVT.2024.3467124

[17] Changsheng Gao, Zhuoyuan Li, Li Li, Dong Liu, and Feng Wu. 2024. Rethinking the Joint Optimization in Video Coding for Machines: A Case Study. In *DCC*. 556–556.

[18] Changsheng Gao, Yifan Ma, Qiaoxi Chen, Yenan Xu, Dong Liu, and Weisi Lin. 2024. Feature Coding in the Era of Large Models: Dataset, Test Conditions, and Benchmark. *arXiv preprint arXiv:2412.04307* (2024).

[19] Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948* (2025).

[20] Sha Guo, Zhuo Chen, Yang Zhao, Ning Zhang, Xiaotong Li, and Lingyu Duan. 2023. Toward Scalable Image Feature Compression: A Content-Adaptive and Diffusion-Based Approach. In *Proceedings of the 31st ACM International Conference on Multimedia* (Ottawa ON, Canada) *(MM '23)*. Association for Computing Machinery, New York, NY, USA, 1431–1442. doi:10.1145/3581783.3611851

[21] Robert Henzel, Kiran Misra, and Tianying Ji. 2022. Efficient Feature Compression for the Object Tracking Task. In *ICIP*. 3505–3509. doi:10.1109/ICIP46576.2022.9897802

[22] Yuzhang Hu, Sifeng Xia, Wenhan Yang, and Jiaying Liu. 2020. Sensitivity-Aware Bit Allocation for Intermediate Deep Feature Compression. In *VCIP*. 475–478. doi:10.1109/VCIP49819.2020.9301807

[23] Ademola Ikusan and Rui Dai. 2021. Rate-Distortion Optimized Hierarchical Deep Feature Compression. In *ICME*. 1–6. doi:10.1109/ICME51207.2021.9428228

[24] Yeongwoong Kim, Hyewon Jeong, Janghyun Yu, Younhee Kim, Jooyoung Lee, Se Yoon Jeong, and Hui Yong Kim. 2023. End-to-End Learnable Multi-Scale Feature Compression for VCM. *IEEE Transactions on Circuits and Systems for Video Technology* (2023), 1–1. doi:10.1109/TCSVT.2023.3302858

[25] Shibao Li, Chenxu Ma, Yunwu Zhang, Longfei Li, Chengzhi Wang, Xuerong Cui, and Jianhang Liu. 2023. Attention-Based Variable-Size Feature Compression Module for Edge Inference. *The Journal of Supercomputing* (2023).

[26] Yifan Ma, Changsheng Gao, Qiaoxi Chen, Li Li, Dong Liu, and Xiaoyan Sun. 2024. Feature Compression With 3D Sparse Convolution. In *VCIP*. 1–5.

[27] Kiran Misra, Tianying Ji, Andrew Segall, and Frank Bossen. 2022. Video Feature Compression for Machine Tasks. In *ICME*. 1–6. doi:10.1109/ICME52920.2022.9859894

[28] Pushmeet Kohli Nathan Silberman, Derek Hoiem and Rob Fergus. 2012. Indoor Segmentation and Support Inference from RGBD Images. In *ECCV*.

[29] Benben Niu, Xiaoran Cao, Ziwei Wei, and Yun He. 2021. Entropy Optimized Deep Feature Compression. *IEEE Signal Processing Letters* 28 (2021), 324–328. doi:10.1109/LSP.2021.3052097

[30] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. 2023. DINOv2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193* (2023).

[31] Saurabh Singh, Sami Abu-El-Haija, Nick Johnston, Johannes Ballé, Abhinav Shrivastava, and George Toderici. 2020. End-to-End Learning of Compressible Features. In *ICIP*. 3349–3353. doi:10.1109/ICIP40778.2020.9190860

[32] Shurun Wang, Shiqi Wang, Wenhan Yang, Xinfeng Zhang, Shanshe Wang, Siwei Ma, and Wen Gao. 2022. Towards Analysis-Friendly Face Representation With Scalable Feature and Texture Compression. *IEEE Transactions on Multimedia* 24 (2022), 3169–3181. doi:10.1109/TMM.2021.3094300

[33] Zixi Wang, Fan Li, Yunfei Zhang, and Yuan Zhang. 2023. Low-Rate Feature Compression for Collaborative Intelligence: Reducing Redundancy in Spatial and Statistical Levels. *IEEE Transactions on Multimedia* (2023), 1–16. doi:10.1109/TMM.2023.3303716

[34] WG2. April 2023. Call for Proposals on Feature Compression for Video Coding for Machines. ISO/IEC JTC 1/SC 29/WG 2, N282 (April 2023).

[35] Ning Yan, Changsheng Gao, Dong Liu, Houqiang Li, Li Li, and Feng Wu. 2021. SSSIC: Semantics-to-Signal Scalable Image Coding With Learned Structural Representations. *IEEE Transactions on Image Processing* 30 (2021), 8939–8954. doi:10.1109/TIP.2021.3121131

[36] Wenhan Yang, Haofeng Huang, Yueyu Hu, Ling-Yu Duan, and Jiaying Liu. 2024. Video Coding for Machines: Compact Visual Representation Compression for Intelligent Collaborative Analytics. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2024), 1–18. doi:10.1109/TPAMI.2024.3367293

[37] Zhongzheng Yuan, Samyak Rawlekar, Siddharth Garg, Elza Erkip, and Yao Wang. 2022. Feature Compression for Rate Constrained Object Detection on the Edge. In *2022 IEEE 5th International Conference on Multimedia Information Processing and Retrieval*. 1–6. doi:10.1109/MIPR54900.2022.00008

[38] Zhicong Zhang, Mengyang Wang, Mengyao Ma, Jiahui Li, and Xiaopeng Fan. 2021. MSFC: Deep Feature Compression in Multi-Task Network. In *ICME*. 1–6. doi:10.1109/ICME51207.2021.9428258

[39] Lingyu Zhu, Binzhe Li, Riyu Lu, Peilin Chen, Qi Mao, Zhao Wang, Wenhan Yang, and Shiqi Wang. 2024. Learned Image Compression for Both Humans and Machines via Dynamic Adaptation. In *ICIP*. IEEE, 1788–1794.