DT-UFC: Universal Large Model Feature Coding via Peaky-to-Balanced Distribution Transformation

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

Zijie Liu
jeffrey101@stu.xmu.edu.cn
Xiamen University
Xiamen, China

Li Li
lil1@ustc.edu.cn
University of Science and Technology of China
Hefei, China

Dong Liu
dongeliu@ustc.edu.cn
University of Science and Technology of China
Hefei, China

Xiaoyan Sun
sunxiaoyan@ustc.edu.cn
University of Science and Technology of China
Hefei, China

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

Abstract

Like image coding in visual data transmission, feature coding is essential for the distributed deployment of large models by significantly reducing transmission and storage burden. However, prior studies have mostly targeted task- or model-specific scenarios, leaving the challenge of universal feature coding across diverse large models largely unexplored. In this paper, we present the first systematic study on universal feature coding for large models. The key challenge lies in the inherently diverse and distributionally incompatible nature of features extracted from different models. For example, features from DINOv2 exhibit highly peaky, concentrated distributions, while those from Stable Diffusion 3 (SD3) are more dispersed and uniform. This distributional heterogeneity severely hampers both compression efficiency and cross-model generalization. To address this, we propose a learned peaky-to-balanced distribution transformation, which reshapes highly skewed feature distributions into a common, balanced target space. This transformation is non-uniform, data-driven, and plug-and-play, enabling effective alignment of heterogeneous distributions without modifying downstream codecs. With this alignment, a universal codec trained on the balanced target distribution can effectively generalize to features from different models and tasks. We validate our approach on three representative large models (LLaMA3, DINOv2, and SD3) across multiple tasks and modalities. Extensive experiments show that our method achieves notable improvements in both compression efficiency and cross-model generalization over task-specific baselines. All source code has been made available at https://github.com/chansongoal/DT-UFC.

CCS Concepts

• **Computing methodologies** → **Image compression**; *Computer vision*; *Natural language processing*.

Keywords

Coding for Machines, Feature Coding, Large Models

ACM Reference Format:
Changsheng Gao, Zijie Liu, Li Li, Dong Liu, Xiaoyan Sun, and Weisi Lin†. 2025. DT-UFC: Universal Large Model Feature Coding via Peaky-to-Balanced Distribution Transformation. In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25), October 27–31, 2025, Dublin, Ireland.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3746027.3755814

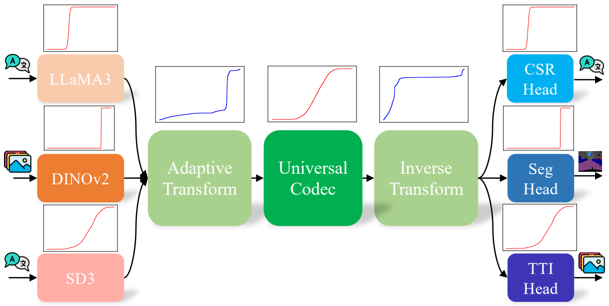

Figure 1: The overall framework of DT-UFC. Large models generate heterogeneous features (red CDF curves). The proposed adaptive transforms (blue curves) reshape these features into a unified target distribution space. The decoded features are inverse-transformed and passed to task-specific heads to complete predictions. (Sec. 4.1)

1 Introduction

In recent years, large foundation models such as GTP-4 [1], DeepSeek [21], and Gemini [49] have demonstrated remarkable capabilities across a wide range of reasoning, discriminative, and generative tasks. As these models are increasingly integrated into distributed and resource-constrained systems [43, 53, 63], the transmission of intermediate representations has become a bottleneck for scalable deployment. Similar to image and video coding [16, 30–34, 39, 48, 69], feature coding aims to encode intermediate features compactly while preserving semantic fidelity. However, existing feature coding approaches are predominantly model- or task-specific

MM '25, Dublin, Ireland

ACM ISBN 979-8-4007-2035-2/2025/10
https://doi.org/10.1145/3746027.3755814

[2, 8, 23, 38, 59, 62], where the encoder is tailored to a particular feature distribution. Such specialization severely limits their universality in realistic scenarios where multiple large models coexist and are applied to various downstream tasks.

These limitations highlight the need for **universal feature coding** — a single codec capable of compressing and reconstructing intermediate representations from diverse models and tasks. Despite its practical importance, universal feature coding remains largely unexplored. The fundamental challenge lies in the distributional heterogeneity of large model features. As illustrated in Fig. 2, features extracted from different models exhibit diverse statistical properties and value distributions. For example, features from DINOv2 [41] exhibit highly peaky distributions with dense value concentrations, while Stable Diffusion 3 (SD3) [12] features are more evenly spread over a narrower range. Such heterogeneity stems from fundamental differences in model architecture, training objectives, and source modalities. This poses a severe obstacle to universal feature coding: a codec trained on a specific feature distribution often generalizes poorly to others. Moreover, these peaky distributions introduce additional challenges for feature coding: they often lead to sparse latent activations, poor entropy modeling efficiency, and unstable codec training. These issues jointly degrade compression performance and further hinder generalization to unseen models or tasks.

To tackle these challenges, we propose a simple yet reasonable insight: rather than forcing a universal encoder to directly accommodate heterogeneous inputs, we introduce a learnable transformation module that reshapes peaky, model-specific feature distributions into a common, balanced target space. By aligning feature distributions through a preprocessing step, we decouple the codec from model-specific statistics, enabling robust generalization across feature distributions. This idea draws inspiration from classical signal processing, where histogram equalization improves encoding robustness by balancing value distributions.

Building on this insight, we propose **DT-UFC**, a novel framework for **U**niversal **F**eature **C**oding via Peaky-to-Balanced **D**istribution **T**ransformation. The core of DT-UFC is the learnable non-uniform peaky-to-balanced transform that can adaptively transforms diverse input feature distributions into a common, balanced target distribution. This transformation is designed to be plug-and-play and does not require any modification to the downstream codec. Once transformed, the features can be compressed by a universal codec and then inversely transformed for downstream tasks. This not only bridges the distribution gap between features but also enhances the expressiveness of the latent representation and improves entropy coding efficiency. To evaluate the generalization ability of DT-UFC, we conduct experiments across three representative tasks: common sense reasoning (CSR), semantic segmentation (Seg), and text-to-image generation (TTI). These tasks cover both discriminative and generative paradigms and span vision and language modalities, providing a comprehensive evaluation for universal feature coding. Our main contributions are summarized as follows:

- We formally introduce the problem of universal feature coding and identify distributional heterogeneity as the key bottleneck preventing generalization across various models and tasks.
- We propose DT-UFC, a plug-and-play framework that learns a peaky-to-balanced transformation to align heterogeneous feature distributions into a balanced target space, enabling a single codec to handle diverse inputs effectively.
- We demonstrate the effectiveness and generalization ability of DT-UFC across multiple large models, tasks, and modalities, achieving superior rate-distortion performance compared to task-specific baselines.

2 Related Work

2.1 Feature Coding

Feature coding, as a key branch of coding for machines [9, 18, 24, 35, 45, 46, 50–52, 65, 68], was originally introduced to compress intermediate features in collaborative intelligence scenarios, where deep models are split across edge and cloud. Early work in [10] evaluated coding performance using standard image codecs but suffered from low efficiency due to distribution mismatch. Subsequent research focused on task-specific feature coding [7, 11, 17, 23, 38, 59], where compression pipelines are tailored to single-task scenarios like image classification or detection. While effective, these methods lack generalizability across diverse tasks or models.

To broaden applicability, recent efforts explored multi-task feature coding [2, 3, 13–15, 22, 55, 57, 62, 67]. For instance, [62] proposed a scalable coding scheme for coarse-to-fine tasks. However, these methods typically rely on task-specific decoders or side information and remain constrained to fixed sets of tasks or models. Recognizing the growing importance of feature coding, standardization efforts have also emerged, notably through MPEG's ongoing work on feature coding standards [60].

Both academic research and standardization have predominantly focused on features extracted from CNN-based models [25, 36, 40, 61, 67]. In contrast, the current landscape of artificial intelligence has been dominated by large models [1, 5, 12, 21, 41, 54]. To bridge this gap, [19] introduced the first large model feature coding benchmark, providing a dataset, unified test conditions, and baseline methods. This work highlights the critical role of feature coding in large model deployment. However, the proposed baselines require model-specific training and do not address the universal feature coding problem.

Therefore, a universal feature codec that can handle diverse models and tasks is increasingly needed. Our work takes a step toward this goal by proposing a learnable transformation that aligns feature distributions, enabling efficient, model-agnostic compression with a single codec.

2.2 Feature Alignment

Feature alignment is a common technique in representation learning to reduce domain shift and improve generalization. It has been widely applied in domain adaptation [26, 37, 47, 56, 64], multi-modal learning [20, 27–29, 42], and multi-task learning [44, 58], enabling models to share or transfer representations across distributions, modalities, or tasks.

Despite its effectiveness, feature alignment has never been applied to feature coding. Existing codecs typically assume fixed input statistics and fail to generalize across heterogeneous features. In this work, we introduce alignment as a pre-coding transformation,

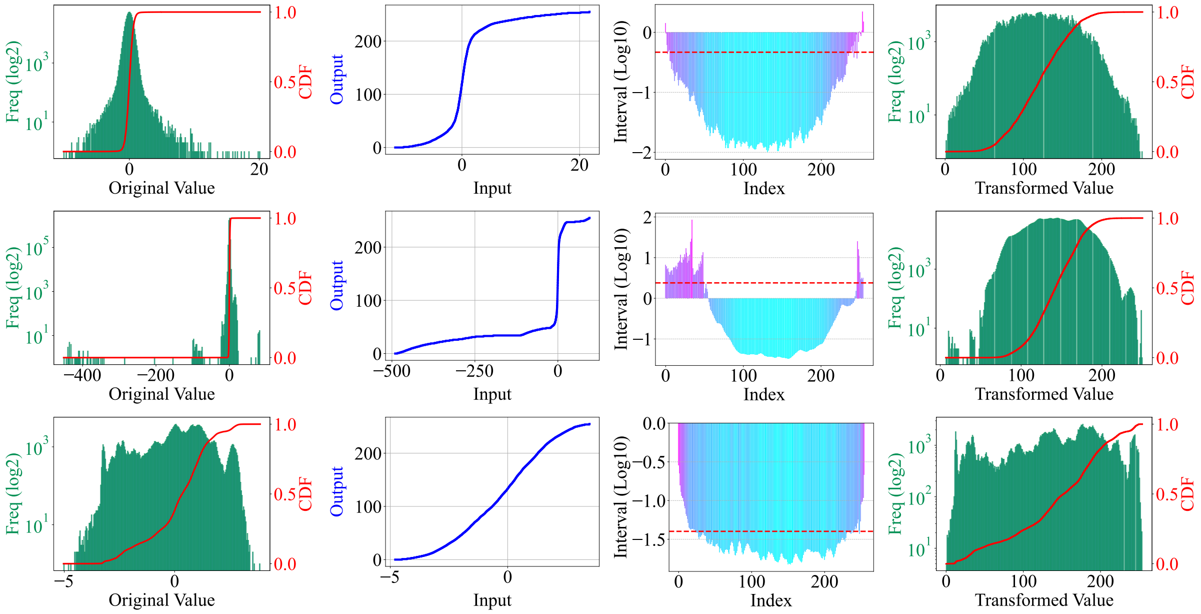

Figure 2: Visualizations of the original feature distribution (1st column), the nonlinear transformation function (2nd column), the width of the split intervals in the original feature space (3rd column), and the transformed feature distribution (4th column). The first, second, and third rows correspond to the CSR, Seg, and TTI tasks, respectively. (Sec. 4.2)

enabling a universal codec to handle diverse large model features for the first time.

3 Challenges in Universal Feature Coding

This section outlines the key challenges hindering universal feature coding across large models.

3.1 Feature Distribution Analysis

A core challenge of universal feature coding lies in the heterogeneous nature of feature distributions across models, stemming from differences in architecture, task objectives, and training data. We visualize the feature distribution in Fig. 2. The green histogram shows the log-scaled frequency of the original feature values, while the red curve represents their empirical cumulative distribution function (CDF).

The feature distributions vary significantly in both statistical range and shape. DINOv2 features exhibit a wide value range with a majority of values densely concentrated in a narrow interval, resulting in a highly peaky distribution. LLaMA3 features follow a similar trend, though with a slightly narrower range and less extreme concentration. In contrast, SD3 produces features that are more balanced and flatly distributed over a limited range. These differences highlight the incompatibility of feature distributions across models, posing a fundamental challenge to universal coding.

3.2 Challenges in Universality

Encoders trained on features from a specific large model tend to internalize distribution-specific priors and inductive biases. When applied to features from another model with a substantially different statistical distribution, these learned priors are violated, resulting in a significant domain shift that degrades both reconstruction quality and generalization performance [43].

Table 1: KL Divergence Comparison between the Original and Transformed Feature Distributions (Sec. 3.2)

Original	**CSR** (q)	**Seg** (q)	**TTI** (q)
CSR (p)	0	19.87	2.95
Seg (p)	18.06	0	4.92
TTI (p)	2.95	15.18	0
Transformed	**CSR** (q)	**Seg** (q)	**TTI** (q)
CSR (p)	0	0.50	0.57
Seg (p)	0.25	0	0.50
TTI (p)	1.15	2.23	0

Moreover, peaky feature distributions further exacerbate the generalization issue. Such inputs tend to encourage neural networks to memorize dominant patterns while neglecting rare but semantically meaningful variations [4, 66]. This tendency reduces the network’s ability to learn robust representations, especially when exposed to data outside the training domain. As a result, the learned representations become brittle and narrowly focused, unable to generalize across diverse input variations or support cross-distribution deployment.

To validate this, we characterize the similarity between different feature distributions using Kullback-Leibler (KL) divergence. Each feature distribution is approximated using a discretized histogram. The KL divergence is computed as $D_{\mathrm{KL}}(p\|q) = \sum p \log(p/q)$. As

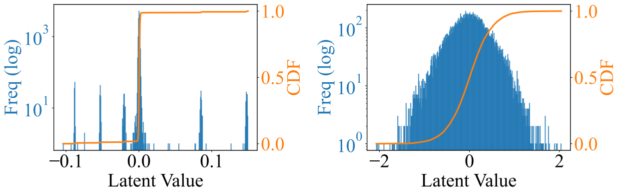

Figure 3: Visualization of the latent space. Left: latent representations obtained from an encoder trained on the original Seg features. Right: latent representations obtained from an encoder trained on the transformed Seg features. (Sec. 3.3)

shown in Table 1, original features from different models exhibit large KL divergence values, confirming their inherent distributional misalignment. This implies that an encoder trained on one feature distribution must allocate additional bitrates to accurately compress features from another distribution. In contrast, the KL divergence among transformed feature distributions is significantly lower, demonstrating that our proposed distribution transformation (see Section 4) effectively aligns statistical properties across different models. This alignment serves as a critical enabler for universal feature coding.

3.3 Challenges in Coding Performance

Beyond generalization, peaky input distributions also negatively impact the efficiency and expressiveness of learned encoders. Specifically, they restrict the effective exploitation of the latent space, destabilize quantization, and reduce entropy coding efficiency.

When most of the input mass resides in a narrow range, the encoder's analysis transform is biased toward over-representing these dominant intervals. Consequently, the latent space is sparsely and unevenly activated, as shown in Fig. 3, where only a few regions are heavily populated while large portions remain underutilized. Such non-uniform activation patterns limit the encoder's representational capacity and are suboptimal for downstream entropy modeling. In addition, this indicates that the encoder transform capacity is not fully exploited, leading to limited transform capacity and poor adaptability to out-of-distribution inputs.

This sparse activation pattern violates the Lloyd-Max principle, which dictates that optimal quantizers should allocate bins proportional to the input density to minimize quantization distortion. However, under peaky distributions, large parts of the quantization space are wasted, resulting in inefficient bit allocation and increased distortion.

In addition, this non-uniform latent usage undermines the assumptions underlying many entropy models, which typically rely on smooth and continuous latent distributions for accurate likelihood estimation. Sharp discontinuities and empty regions lead to unstable training, inaccurate bitrate estimation, and ultimately deteriorate rate–distortion performance.

4 The Proposed DT-UFC Method

4.1 Overall Framework

Based on the challenges discussed in Section 3, we identify a critical requirement for universal feature coding:

Feature distributions should be aligned across models and avoid excessive concentration.

This insight motivates us to seek a feature transformation that reshapes diverse feature distributions into a common and balanced target distribution. Such a transformation reduces distributional discrepancies and enables a universal encoder to compress heterogeneous features efficiently.

We propose the DT-UFC framework, illustrated in Fig. 1. Given that input features are extracted from various large models, we first apply an adaptive transformation that normalizes and redistributes the original feature values into a common, balanced distribution space. A universal encoder is then trained to compress features in this space. After decoding, the features are mapped back to their original domain via an inverse transformation, and then fed into task-specific heads for inference.

The core challenge in the DT-UFC framework lies in designing a transformation that aligns heterogeneous input distributions to a common target, while minimizing transformation-induced distortion.

4.2 Non-uniform Feature Transformation

To address the problem outlined above, we first define two essential criteria for a desirable transformed feature distribution:

- **Consistent Distribution Range:** Feature values from different models should be mapped to a common range to facilitate model-agnostic feature coding.
- **Balanced Distribution:** Feature values should be evenly distributed within the range to fully exploit the latent capacity of the encoder and enable efficient entropy coding.

To satisfy the first criterion, we decided to map the original features into a fixed space. Inspired by image coding, we define a fixed target space of $p_t \in \{0, 1, \ldots, 255\}$ for transformed features. This standardization facilitates compatibility with existing image codecs and allows for direct reuse of quantization and entropy modeling techniques. To satisfy the second criterion, we design a learnable non-uniform transformation function that redistributes the original feature values to form a balanced histogram in the target space. This is especially important for features from models like LLaMA3 and DINOv2, whose raw distributions are highly concentrated.

Instead of deriving a closed-form function, we treat the transformation as a data-driven optimization problem. We aim to partition the input space into 256 non-overlapping regions and map each region to a unique integer in p_t. Now the problem turns to how to split the original distribution range into 256 regions with minimal transformation distortion.

Formally, given an input feature $\mathbf{f} \in \mathbb{R}^d$ sampled from distribution p_f, we define a transformation $T : \mathbb{R} \to \{0, 1, \ldots, 255\}$ such that:

- The range of $\mathbf{f}$ is partitioned into 256 non-overlapping intervals $\mathcal{R} = \{R_0, R_1, \ldots, R_{255}\}$, with each $R_k \subset \mathbb{R}$.
- All feature values in R_k are mapped to the discrete value k via T.
- For each k, we define an inverse transformed value $c_k \in \mathbb{R}$.

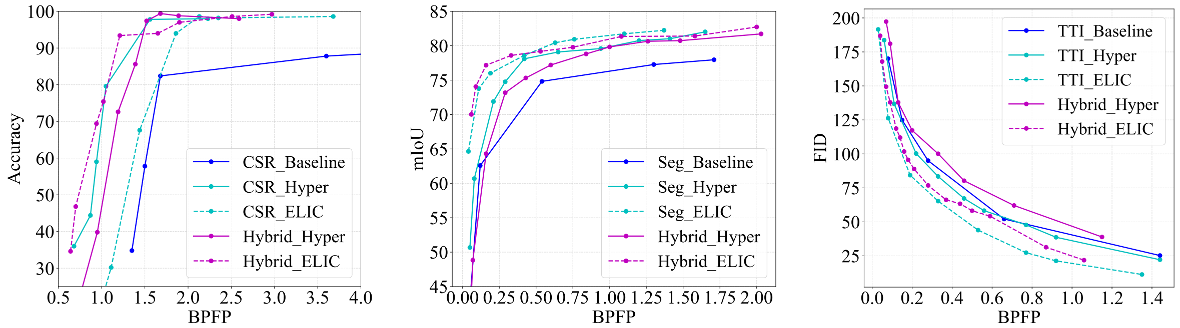

Figure 4: Rate-accuracy performance comparisons between the baseline and the proposed method. The left, middle, and right figures correspond to the CSR, Seg, and TTI tasks, respectively. (Sec. 5.3)

The forward transformation is defined as:

$$T(f_i) = k \quad \text{if} \quad f_i \in R_k, \tag{1}$$

and the inverse transformation is:

$$T^{-1}(k) = c_k. \tag{2}$$

The objective is to minimize the transformation distortion:

$$\mathcal{D}_{\text{transform}} = \mathbb{E}_{\mathbf{f}\sim p_f}\left[\sum_{i=1}^{d}\left(f_i - T^{-1}(T(f_i))\right)^2\right]. \tag{3}$$

This optimization is carried out iteratively through two steps:

Assignment step: Given current inverse transformed values $\{c_k\}$, update each region R_k as:

$$R_k = \left\{ f_i \in \mathbb{R} \mid k = \arg\min_{j\in\{0,\ldots,255\}} \left|f_i - c_j\right| \right\}. \tag{4}$$

Update step: Given updated regions $\{R_k\}$, recompute the inverse transformed values:

$$c_k = \frac{1}{|R_k|}\sum_{f_i\in R_k} f_i. \tag{5}$$

We alternate between these steps until convergence of $\mathcal{D}_{\text{transform}}$. The resulting region-to-integer mapping forms the non-uniform transformation $T(\cdot)$, and $\{c_k\}$ are stored for inverse transform $T^{-1}(\cdot)$. In our experiment, we randomly select 10 features to learn the non-uniform transform.

Figure 2 illustrates the full transformation pipeline of the proposed DT-UFC method. Each row corresponds to a different task, and each column shows a specific stage:

Original Distribution: The original feature distribution is often highly skewed or concentrated, as seen in LLaMA3 and DINOv2.

Learned Transform: The blue curve plots the learned nonlinear transform function $T(f)$. For peaky distributions, the mapping function becomes steep around highly concentrated regions, assigning more indices to concentrated values. For more uniform inputs (e.g., TTI), the function appears smoother, resembling a near-linear transform.

Region Interval Width: We visualize the logarithmic width of each region R_k. We observe shorter intervals (denser coverage) near peaks of the original distribution, and longer intervals in sparse regions. The dashed red lines denote the average region interval.

Transformed Distribution: The transformed distribution over integers in p_t becomes much more balanced. The CDF curve is now smooth and monotonic, indicating successful distribution alignment. This transformation not only reduces distributional divergence across models (Table 1) but also improves latent expressiveness (See Fig. 3).

In summary, the proposed non-uniform transformation bridges the gap between heterogeneous input distributions and a common encoding space. It enables a universal codec to achieve both high generalization and rate-distortion efficiency across diverse models and tasks.

5 Experiments

5.1 Dataset

For each task, we first extract 10,000 original features from pre-trained large models to construct the training set. These features are then transformed using the proposed non-uniform transformation to generate training inputs for the codec.

For evaluation, we follow the unified test condition proposed in [19], using 100, 500, and 500 features for the Seg, CSR, and TTI tasks, respectively. All features are packed using the same method in [19].

5.2 Codec and Training Details

We adopt two representative feature codecs: Hyperprior [6] and ELIC [23]. To investigate the impact of training data on performance, we consider two strategies: hybrid training and task-specific training. In the hybrid training setting, all 30000 features across tasks are used. In the task-specific training setting, only the 10000 features corresponding to each task are used for that task.

The feature codecs are optimized using the following loss function:

$$L = BPFP + \lambda \times ||(F_o - \hat{F_r})||^2 \tag{6}$$

where BPFP measures the bitrate. λ is a scaling factor used to adjust the bitrate. F_o and F_r denote the original input feature and

Table 2: Rate-Accuracy Performance on CSR, Seg, and TTI with *Hybrid-Trained Hyperprior* Codecs (Sec. 5.3, Sec. 5.4)

Task	Common Sense Reasoning			Semantic Segmentation			Text-to-Image Synthesis		
Lambda	BPFP	Accuracy ↑	MSE	BPFP	mIoU ↑	MSE	BPFP	FID ↓	MSE
0.0008	0.07	0.00	0.3006	0.04	34.33	3.5556	0.05	226.74	0.1388
0.001	0.11	0.00	0.2836	0.07	48.83	3.3804	0.06	234.27	0.1413
0.0013	0.20	0.00	0.2400	0.13	60.66	4.5479	0.07	197.31	0.1220
0.0018	0.49	3.80	0.1943	0.29	73.17	4.6615	0.12	140.52	0.0902
0.0019	0.52	4.00	0.1892	0.31	73.85	4.2202	0.13	137.88	0.0884
0.0023	0.67	15.60	0.1672	0.43	75.32	2.5743	0.16	137.83	0.0824
0.0025	0.95	39.80	0.1494	0.47	75.77	3.1560	0.17	127.98	0.0788
0.0028	1.19	72.60	0.1342	0.54	76.95	2.8723	0.19	119.43	0.0750
0.003	1.23	75.40	0.1177	0.60	77.20	2.4534	0.20	117.31	0.0707
0.004	1.39	85.60	0.1142	0.84	78.81	1.8782	0.27	112.95	0.0625
0.005	1.52	97.40	0.0904	1.00	79.83	1.8823	0.33	100.11	0.0551
0.007	1.68	99.40	0.0782	1.26	80.65	1.8499	0.46	80.26	0.0441
0.01	1.89	98.80	0.0715	1.48	80.74	1.6486	0.71	62.06	0.0337
0.02	2.59	98.00	0.0454	2.03	81.72	1.0889	1.15	38.91	0.0217

Table 3: Rate-Accuracy Performance on CSR, Seg, and TTI with *Hybrid-Trained ELIC* Codecs (Sec. 5.3, Sec. 5.4)

Task	Common Sense Reasoning			Semantic Segmentation			Text-to-Image Synthesis		
Lambda	BPFP	Accuracy ↑	MSE	BPFP	mIoU ↑	MSE	BPFP	FID ↓	MSE
0.0001	0.04	0.00	0.3316	0.005	15.23	7.0575	0.02	226.86	0.1751
0.0003	0.21	0.00	0.2986	0.06	70.02	3.0153	0.04	186.87	0.1267
0.001	0.36	3.80	0.2199	0.16	77.17	3.0283	0.09	137.92	0.0890
0.0015	0.64	34.60	0.1702	0.28	78.28	2.6021	0.12	118.64	0.0765
0.0019	0.70	46.80	0.1579	0.33	78.59	2.5045	0.14	112.00	0.0716
0.0021	0.94	69.40	0.1433	0.47	78.92	2.3994	0.16	101.74	0.0663
0.0025	1.02	75.40	0.1332	0.53	79.18	2.4558	0.18	95.59	0.0625
0.003	1.21	93.40	0.1125	0.61	79.24	2.2981	0.21	88.92	0.0586
0.004	1.65	94.00	0.0875	0.75	79.78	2.0942	0.28	76.84	0.0507
0.005	1.90	97.00	0.0641	0.94	80.13	1.9586	0.37	66.24	0.0440
0.007	2.21	94.20	0.0577	1.08	81.35	1.3101	0.44	63.32	0.0399
0.01	2.51	98.60	0.0464	1.58	81.41	1.2771	0.59	54.21	0.0327
0.015	2.97	99.20	0.0317	1.79	81.64	1.2243	0.87	31.37	0.0225
0.02	3.16	98.80	0.0192	2.00	82.73	0.6090	1.06	21.85	0.0174

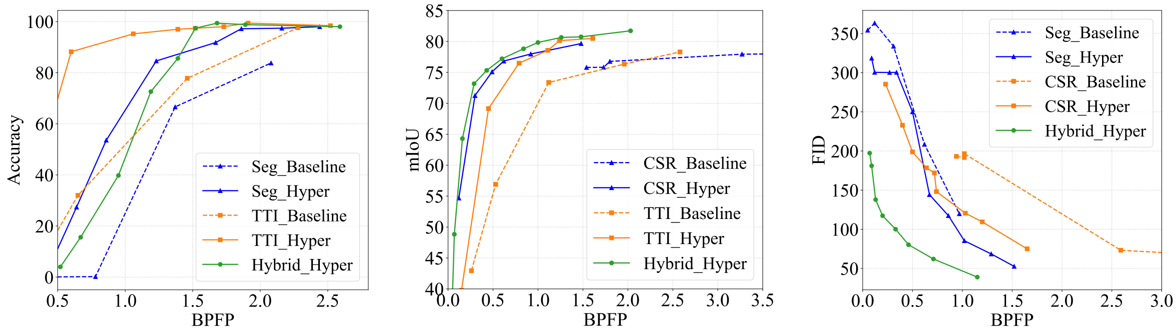

Figure 5: Universality comparison between the baseline and the proposed method. The left, middle, and right figures correspond to CSR, Seg, and TTI, respectively. (Sec. 5.4)

Table 4: Rate-Accuracy Performance on CSR, Seg, and TTI with *Task-Specifically Trained Hyperprior* Codecs

	Task	CSR		Seg		TTI	
Metric	**Lambda**	**BPFP**	**Acc.**	**BPFP**	**mIoU**	**BPFP**	**FID**
CSR Models	0.0005	0.11	0.00	0.05	10.61	0.13	327.38
	0.0008	0.30	0.00	0.12	54.69	0.23	285.36
	0.0017	0.68	36.00	0.39	73.10	0.50	198.86
	0.0019	0.87	44.40	0.49	75.09	0.64	178.49
	0.002	0.94	59.00	0.52	75.62	0.72	171.93
	0.0025	1.05	79.60	0.62	76.83	0.74	148.21
	0.003	1.56	97.80	0.92	77.97	1.03	120.52
	0.0035	1.74	97.40	1.03	77.84	1.20	109.46
	0.006	2.23	98.00	1.48	79.66	1.65	75.02
Seg Models	0.0007	0.17	0.00	0.05	50.67	0.09	318.53
	0.0008	0.21	0.00	0.08	60.69	0.12	300.42
	0.002	0.46	6.40	0.21	71.89	0.27	300.25
	0.0025	0.64	27.40	0.29	74.76	0.34	300.40
	0.003	0.86	53.60	0.42	78.10	0.50	250.03
	0.004	1.23	84.60	0.65	79.09	0.67	144.28
	0.005	1.67	91.80	0.94	79.59	0.86	117.41
	0.006	1.86	97.20	1.10	80.05	1.01	111.02
	0.01	2.16	97.40	1.41	81.02	1.29	68.50
	0.015	2.44	98.00	1.65	82.02	1.52	52.64
TTI Models	0.001	0.09	0.00	0.06	28.64	0.06	183.68
	0.002	0.21	16.20	0.15	39.78	0.11	136.78
	0.004	0.60	88.20	0.45	69.15	0.22	100.26
	0.006	1.06	95.20	0.79	76.49	0.33	83.42
	0.008	1.39	97.00	1.11	78.56	0.46	67.20
	0.01	1.52	97.40	1.24	80.13	0.56	58.27
	0.015	1.73	98.00	1.43	78.62	0.77	47.79
	0.02	1.91	99.40	1.61	80.49	0.92	38.69
	0.05	2.52	98.40	2.16	80.62	1.44	22.16

reconstructed feature, respectively. Both training strategies use the same training configuration. We employ the Adam optimizer with an initial learning rate of 1e-4, which is adjusted by the ReduceLROnPlateau scheduler. Training continues until the learning rate drops to 1e-8, followed by 20 additional epochs. The model with the lowest validation loss is selected. To control the bitrate, we adjust the Lagrange multiplier λ. Detailed settings are provided in Tables 2, 3, and 4. All experiments are conducted on Nvidia RTX 3090 GPUs.

5.3 Rate-Accuracy Performance

This subsection validates the performance of DT-UFC in terms of rate-accuracy across three representative tasks: CSR, Seg, and TTI. We compare DT-UFC against baseline methods in both the task-specifically trained and hybrid-trained strategies. Fig. 4 shows the rate-accuracy curves, with numerical results in Tables 2, 3, and 4. For clarity, some points are omitted from the figures.

In both CSR and Seg tasks, DT-UFC consistently outperforms the baselines under both training strategies, especially at higher bitrates. This demonstrates that our proposed feature transformation effectively addresses the issue of peaky and concentrated feature distributions. Generally, the CSR task exhibits a higher correlation between the feature MSE and task accuracy than the Seg task. We attribute this to the broader original feature distribution range observed in the Seg task, which introduces greater nonlinearity in the rate–accuracy relationship.

In the TTI task, while the hybrid-trained ELIC codecs outperform both baselines consistently, the hybrid-trained Hyperprior codecs slightly underperform their task-specifically trained counterparts at low bitrates. This indicates that TTI features, which are most similar to image distributions, benefit more from high-capacity codecs like ELIC.

Additionally, in the CSR task, the task-specifically trained Hyperprior codec achieves better performance than the more sophisticated ELIC codec. This phenomenon reveals an inherent distinction between textual and visual feature distributions: codecs that excel on visual features may not perform well on textual features and may even hurt performance.

In summary, the proposed distribution transformation provides a more encoder-friendly feature representation space that enables both lightweight and high-capacity codecs to operate effectively. This is the key to realizing high-performance, task-agnostic feature coding.

5.4 Cross-Task Universality

To evaluate the universality of DT-UFC, we test its ability to compress features from unseen tasks using models trained on a different task. Fig. 5 shows the rate-accuracy performance across tasks.

In the CSR task, baseline models trained on Seg or TTI perform poorly as they fail to generalize to the highly distinct textual CSR features. In contrast, DT-UFC achieves strong performance even when trained on other tasks, demonstrating the effectiveness of our transformation in aligning heterogeneous distributions. By transforming task-specific features into a unified space, DT-UFC reduces distributional divergence and enables successful cross-task compression.

In the Seg task, the baseline CSR models fail to achieve low bitrates, while all the DT-UFC models achieve smooth rate-accuracy curves. For the TTI models, both hybrid-trained and task-specifically trained DT-UFC models outperform their baseline counterparts by a large margin, indicating higher robustness of DT-UFC in the TTI task.

In the TTI task, all the baseline models and task-specifically trained models show unstable rate-accuracy behaviors. Nevertheless, DT-UFC still outperforms the baselines, especially when trained on CSR. The hybrid-trained variant demonstrates notable robustness and achieves the best overall performance.

Furthermore, comparing task-specific and hybrid training reveals that hybrid training generally improves generalization, especially in Seg and TTI tasks, by exposing the codec to more diverse feature distributions. However, in the CSR task, hybrid training slightly degrades performance at low bitrates, possibly due to the model’s adaptation to visual features. This highlights the importance of balancing feature diversity in multi-task scenarios.

Overall, DT-UFC’s distribution alignment strategy ensures reliable encoder performance across tasks, even in zero-shot scenarios. This confirms its potential as a universal feature coding solution across tasks, models, and modalities.

Table 5: Rate-Accuracy Performance Comparison between the Baseline, Baseline with Transformed Features, and DT-UFC with Baseline Features (Sec. 5.5 and Sec. 5.6)

Task	CSR		Seg		TTI	
Metric	**BPFP**	**Acc.**	**BPFP**	**mIoU**	**BPFP**	**FID**
Original	**32**	**100**	**32**	**83.39**	**32**	**0**
Baseline	1.35	34.80	0.03	37.11	0.08	170.01
	1.50	57.80	0.12	62.58	0.15	124.84
	1.68	82.40	0.54	74.82	0.28	95.03
	3.60	87.80	1.30	77.27	0.66	52.12
	6.34	91.40	1.71	77.96	1.44	25.25
Baseline with Transformed Features (Sec. 5.5)	2.08	94.00	0.04	31.60	0.12	161.72
	2.34	91.40	0.12	64.60	0.26	115.86
	2.49	92.00	0.51	77.72	0.56	80.91
	3.91	96.00	1.23	80.14	1.18	35.26
	7.02	96.80	1.65	81.34	1.93	22.60
DT-UFC with Baseline Features (Sec. 5.6)	0.30	6.60	0.05	37.32	0.08	166.57
	0.35	14.80	0.17	62.41	0.14	144.11
	0.50	20.80	0.45	73.50	0.18	128.57
	0.97	71.00	1.09	76.55	0.33	96.84
	1.41	97.20	1.55	77.94	0.54	72.97

Table 6: Rate-Accuracy Performance of DT-UFC on CNN Features (Sec. 5.7)

Lambda	BPFP	Accuracy	MSE
Original	**32**	**86.40**	**0**
0.0005	0.06	84.40	0.5006
0.0008	0.09	85.00	0.4627
0.0015	0.19	85.80	0.3350
0.0025	0.43	86.00	0.1749
0.004	0.88	86.40	0.0950

5.5 Universality of Transformed Features

To isolate the contribution of the distribution transformation, we directly feed transformed features into the baseline codecs without any retraining and compare their rate-accuracy performance with the baseline.

As shown in Table 5, this simple substitution leads to significant improvements across all tasks. In the CSR task, the baseline codecs consistently encode the transformed feature with high bitrates yet retain high accuracy. In the Seg and TTI tasks, the trends in bitrate and accuracy remain similar to the baseline, indicating no degradation. These results demonstrate that the transformation alone enhances feature compressibility by reshaping feature distributions. Even without retraining the codec, the benefits are clearly observed. In essence, the transformation serves not only as an enabler for universal encoder training but also as a standalone mechanism for improving structural expressiveness and compressibility of features.

5.6 Codec's Robustness to Imperfect Inputs

We further examine the robustness of DT-UFC codecs when given imperfect input. Specifically, we feed features truncated by baseline pre-processing to the hybrid-trained feature codecs.

As reported in Table 5, DT-UFC codecs maintain a competitive performance under these degraded conditions. For the CSR task, DT-UFC substantially outperforms the baseline, achieving over 40% accuracy improvement at 1.5 BPFP. In Seg and TTI tasks, DT-UFC maintains comparable performance with the baseline. This demonstrates that even when inputs deviate from the ideal transformed space, codecs trained using the transformed features still maintains their effectiveness.

Moreover, the decline of metrics such as MSE and FID across bitrate indicates that codecs trained with transformed features have not merely memorized specific input patterns but have learned to model underlying feature structures more generally. Therefore, DT-UFC is not only effective under ideal conditions but also robust in real-world deployment scenarios where features may be distorted or misaligned. This resilience is critical for practical adoption.

5.7 Architecture-Agnostic Generalization

To assess cross-architecture generalization, we apply DT-UFC to CNN-derived features. Specifically, we extract features from ResNet-50 on an image classification task. A transformation is learned on 10 features and applied before encoding. Instead of training feature codecs on CNN features, we feed them to the hybrid-trained Hyperprior codecs.

Table 6 shows that DT-UFC achieves nearly lossless classification accuracy and decreasing MSE across bitrates. This demonstrates the codec's strong generalization from Transformer features to CNN features. Importantly, this result highlights our method's ability to align feature distributions across architectural paradigms, enabling universal feature compression across diverse model backbones. In addition, the generalization ability also indicates that the proposed transformation can align feature distributions across model architectures, enabling universal feature coding across Transformer-based and CNN-based backbones. The strong performance without retraining further validates the scalability and practicality of DT-UFC as a truly architecture-agnostic feature codec.

6 Conclusion

In this paper, we present DT-UFC, a novel framework for universal feature coding across large models by introducing a peaky-to-balanced distribution transformation. By aligning diverse feature distributions into a common, balanced space, DT-UFC enables a single codec to generalize across models and tasks without retraining. Extensive experiments demonstrate that DT-UFC not only improves rate-distortion performance over task-specific baselines, but also supports architecture-agnostic and cross-task generalization, confirming its practical value in large model deployment scenarios.

In future work, we aim to integrate DT-UFC into real-world distributed AI systems to evaluate its impact on latency, energy efficiency, and scalability. We hope this work inspires further research on universal and practical feature coding.

Acknowledgments

This work was supported by the Ministry of Education of Singapore under Grant T2EP20123-0006. We acknowledge the support of GPU cluster built by MCC Lab of Information Science and Technology Institution, USTC.

References

[1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).

[2] 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

[3] 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

[4] Devansh Arpit, Stanisław Jastrzundefinedbski, Nicolas Ballas, David Krueger, Emmanuel Bengio, Maxinder S. Kanwal, Tegan Maharaj, Asja Fischer, Aaron Courville, Yoshua Bengio, and Simon Lacoste-Julien. 2017. A closer look at memorization in deep networks *(ICML’17)*. JMLR, 233–242.

[5] Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, et al. 2023. Qwen technical report. *arXiv preprint arXiv:2309.16609* (2023).

[6] 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).

[7] 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

[8] Qiaoxi Chen, Changsheng Gao, and Dong Liu. 2024. End-to-End Learned Scalable Multilayer Feature Compression For Machine Vision Tasks. In *ICIP*. 1781–1787. doi:10.1109/ICIP51287.2024.10647798

[9] Yi-Hsin Chen, Ying-Chieh Weng, Chia-Hao Kao, Cheng Chien, Wei-Chen Chiu, and Wen-Hsiao Peng. 2023. TransTIC: Transferring transformer-based image compression from human perception to machine perception. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 23297–23307.

[10] 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

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

[12] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. 2024. Scaling rectified flow transformers for high-resolution image synthesis. In *ICML*.

[13] 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.

[14] 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

[15] 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

[16] Changsheng Gao, Li Li, Dong Liu, Zhibo Chen, Weiping Li, and Feng Wu. 2022. Two-Step Fast Mode Decision for Intra Coding of Screen Content. *IEEE Transactions on Circuits and Systems for Video Technology* 32, 8 (2022), 5608–5622.

[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, Dong Liu, Li Li, and Feng Wu. 2023. Towards Task-Generic Image Compression: A Study of Semantics-Oriented Metrics. *IEEE Transactions on Multimedia* 25 (2023), 721–735. doi:10.1109/TMM.2021.3130754

[19] 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).

[20] Xuri Ge, Fuhai Chen, Joemon M Jose, Zhilong Ji, Zhongqin Wu, and Xiao Liu. 2021. Structured multi-modal feature embedding and alignment for image-sentence retrieval. In *ACM MM*. 5185–5193.

[21] 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).

[22] 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

[23] 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

[24] Chia-Hao Kao, Cheng Chien, Yu-Jen Tseng, Yi-Hsin Chen, Alessandro Gnutti, Shao-Yuan Lo, Wen-Hsiao Peng, and Riccardo Leonardi. 2024. Bridging compressed image latents and multimodal large language models. *arXiv preprint arXiv:2407.19651* (2024).

[25] 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

[26] JoonHo Lee and Gyemin Lee. 2023. Feature Alignment by Uncertainty and Self-Training for Source-Free Unsupervised Domain Adaptation. *Neural Networks* 161 (2023), 682–692. doi:10.1016/j.neunet.2023.02.009

[27] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International conference on machine learning*. PMLR, 19730–19742.

[28] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. 2022. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International conference on machine learning*. PMLR, 12888–12900.

[29] Junnan Li, Ramprasaath Selvaraju, Akhilesh Gotmare, Shafiq Joty, Caiming Xiong, and Steven Chu Hong Hoi. 2021. Align before fuse: Vision and language representation learning with momentum distillation. *Advances in neural information processing systems* 34 (2021), 9694–9705.

[30] Yao Li, Zhuoyuan Li, Li Li, Dong Liu, and Houqiang Li. 2022. Global homography motion compensation for versatile video coding. In *2022 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. IEEE, 1–5.

[31] Zhuoyuan Li, Jiacheng Li, Yao Li, Li Li, Dong Liu, and Feng Wu. 2024. In-Loop Filtering via Trained Look-Up Tables. In *2024 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. 1–5.

[32] Zhuoyuan Li, Yao Li, Chuanbo Tang, Li Li, Dong Liu, and Feng Wu. 2024. Uniformly Accelerated Motion Model for Inter Prediction. In *2024 IEEE International Conference on Visual Communications and Image Processing (VCIP)*. 1–5.

[33] Zhuoyuan Li, Junqi Liao, Chuanbo Tang, Haotian Zhang, Yuqi Li, Yifan Bian, Xihua Sheng, Xinmin Feng, Yao Li, Changsheng Gao, et al. 2024. USTC-TD: A test dataset and benchmark for image and video coding in 2020s. *arXiv preprint arXiv:2409.08481* (2024).

[34] Zhuoyuan Li, Zikun Yuan, Li Li, Dong Liu, Xiaohu Tang, and Feng Wu. 2024. Object Segmentation-Assisted Inter Prediction for Versatile Video Coding. *IEEE Transactions on Broadcasting* 70, 4 (2024), 1236–1253.

[35] Jinming Liu, Ruoyu Feng, Yunpeng Qi, Qiuyu Chen, Zhibo Chen, Wenjun Zeng, and Xin Jin. 2024. Rate-distortion-cognition controllable versatile neural image compression. In *European Conference on Computer Vision*. Springer, 329–348.

[36] Kang Liu, Dong Liu, Li Li, Ning Yan, and Houqiang Li. 2021. Semantics-to-Signal Scalable Image Compression with Learned Revertible Representations. *Int. J. Comput. Vision* 129, 9 (Sept. 2021), 2605–2621. doi:10.1007/s11263-021-01491-7

[37] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. 2015. Learning transferable features with deep adaptation networks. In *International conference on machine learning*. PMLR, 97–105.

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

[39] Rui Mao, Xinmin Feng, Changsheng Gao, Li Li, Dong Liu, and Xiaoyan Sun. 2024. Perceptual Image Compression With Conditional Diffusion Transformers. In *VCIP*. 1–5.

[40] 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

[41] 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).

[42] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *ICML*. PMLR, 8748–8763.

[43] Shiori Sagawa*, Pang Wei Koh*, Tatsunori B. Hashimoto, and Percy Liang. 2020. Distributionally Robust Neural Networks. In *ICLR*.

[44] Dmitry Senushkin, Nikolay Patakin, Arseny Kuznetsov, and Anton Konushin. 2023. Independent component alignment for multi-task learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 20083–20093.

[45] Xuelin Shen, Haoqiao Ou, and Wenhan Yang. 2024. Image Coding For Machine Via Analytics-Driven Appearance Redundancy Reduction. In *2024 IEEE International Conference on Image Processing (ICIP)*. IEEE, 1883–1889.

[46] Xuelin Shen, Kangsheng Yin, Xu Wang, Yulin He, Shiqi Wang, and Wenhan Yang. 2024. Image coding for analytics via adversarially augmented adaptation. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 3605–3609.

[47] Baochen Sun and Kate Saenko. 2016. Deep coral: Correlation alignment for deep domain adaptation. In *Computer vision–ECCV 2016 workshops: Amsterdam,*

the Netherlands, October 8-10 and 15-16, 2016, proceedings, part III 14. Springer, 443–450.

[48] Chuanbo Tang, Xihua Sheng, Zhuoyuan Li, Haotian Zhang, Li Li, and Dong Liu. 2024. Offline and online optical flow enhancement for deep video compression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 5118–5126.

[49] Gemini Team, Petko Georgiev, Ving Ian Lei, Ryan Burnell, Libin Bai, Anmol Gulati, Garrett Tanzer, Damien Vincent, Zhufeng Pan, Shibo Wang, et al. 2024. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530* (2024).

[50] Yuan Tian, Guo Lu, Yichao Yan, Guangtao Zhai, Li Chen, and Zhiyong Gao. 2024. A coding framework and benchmark towards low-bitrate video understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 8 (2024), 5852–5872.

[51] Yuan Tian, Guo Lu, and Guangtao Zhai. 2024. Free-VSC: Free semantics from visual foundation models for unsupervised video semantic compression. In *European Conference on Computer Vision*. Springer, 163–183.

[52] Yuan Tian, Guo Lu, Guangtao Zhai, and Zhiyong Gao. 2023. Non-semantics suppressed mask learning for unsupervised video semantic compression. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 13610–13622.

[53] Yuanyishu Tian, Yao Wan, Lingjuan Lyu, Dezhong Yao, Hai Jin, and Lichao Sun. 2022. FedBERT: When federated learning meets pre-training. *ACM Transactions on Intelligent Systems and Technology* 13, 4 (2022), 1–26.

[54] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).

[55] 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

[56] Shuai Wang, Daoan Zhang, Zipei Yan, Jianguo Zhang, and Rui Li. 2023. Feature alignment and uniformity for test time adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 20050–20060.

[57] Weiqian Wang, Ping An, Xinpeng Huang, Kunqiang Huang, and Chao Yang. 2023. Intermediate deep feature coding for human–machine vision collaboration. *Journal of Visual Communication and Image Representation* 95 (2023), 103859.

[58] Xiaoling Wang, Qi Kang, MengChu Zhou, Siya Yao, and Abdullah Abusorrah. 2022. Domain adaptation multitask optimization. *IEEE Transactions on Cybernetics* 53, 7 (2022), 4567–4578.

[59] Zixi Wang, Fan Li, Yunfei Zhang, and Yuan Zhang. 2024. Low-Rate Feature Compression for Collaborative Intelligence: Reducing Redundancy in Spatial and Statistical Levels. *IEEE Transactions on Multimedia* 26 (2024), 2756–2771. doi:10.1109/TMM.2023.3303716

[60] 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).

[61] Xu Wu, Xianxu Hou, Zhihui Lai, Jie Zhou, Ya-nan Zhang, Witold Pedrycz, and Linlin Shen. 2025. A codebook-driven approach for low-light image enhancement. *Engineering Applications of Artificial Intelligence* 156 (2025), 111115.

[62] 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

[63] Rui Ye, Wenhao Wang, Jingyi Chai, Dihan Li, Zexi Li, Yinda Xu, Yaxin Du, Yanfeng Wang, and Siheng Chen. 2024. OpenFedLLM: Training large language models on decentralized private data via federated learning. In *ACM SIGKDD*. 6137–6147.

[64] Hao-Wei Yeh, Baoyao Yang, Pong C Yuen, and Tatsuya Harada. 2021. SOFA: Source-data-free feature alignment for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 474–483.

[65] Kangsheng Yin, Quan Liu, Xuelin Shen, Yulin He, Wenhan Yang, and Shiqi Wang. 2025. Unified Coding for Both Human Perception and Generalized Machine Analytics with CLIP Supervision. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 9517–9525.

[66] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. 2021. Understanding deep learning (still) requires rethinking generalization. *Commun. ACM* 64, 3 (Feb. 2021), 107–115. doi:10.1145/3446776

[67] 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

[68] 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 *2024 IEEE International Conference on Image Processing (ICIP)*. 1788–1794.

[69] Yanchen Zuo, Changsheng Gao, Dong Liu, Li Li, Yueyi Zhang, and Xiaoyan Sun. 2024. Learned Rate-Distortion Cost Prediction for Ultrafast Screen Content Intra Coding. *IEEE Transactions on Circuits and Systems for Video Technology* 34, 3 (2024), 1976–1980.