

毕业设计开题报告

郑鑫裕

2024 年 12 月 31 日

目录

1	项目分析	2
1.1	项目目标分析	2
1.2	解决方案整体架构	3
1.2.1	目标检测	3
1.2.2	边缘检测	3
1.2.3	图像压缩	4
1.2.4	实时性	4
2	边缘检测	5
3	图像压缩	5
3.1	图像压缩算法结构	5
3.1.1	预处理	6
3.1.2	变换	6
3.1.3	量化	6
3.1.4	熵编码	6
3.1.5	解压缩	7
3.1.6	后处理	7
3.2	传统图像压缩算法/编码格式	7
3.2.1	无损图像编码	7

3.2.2	有损图像编码	8
3.3	AI 算法	10
3.3.1	训练	11
3.3.2	主干网络	12
3.3.3	INN	13
3.3.4	GAN	15
3.3.5	Stable Diffusion ATTACH	16
3.4	评价指标	17
3.5	总结	18
4	实现计划	19
4.1	基础框架搭建与初步实验	19
4.2	模型优化与架构探索	19
4.3	实验记录与总结	20
4.4	最终成果与应用验证	20

1 项目分析

1.1 项目目标分析

本项目的目标是在窄带或极低带宽下实现针对图像轮廓信息的高压缩比智能图像编解码算法。具体而言，困难主要体现在以下几个方面：

1. 低带宽与图像大小之间的矛盾：在某些情况下，数据传输的速率可能不高于 1kbit/s, 而为满足人眼的视觉暂留效应条件，每秒在接收端至少需要 10 帧以上的图像播放，这对视频的编解码提出了巨大的挑战。而现实情形中，一张图像哪怕是采用了最先进的 AVC, HEVC, BPG 等图像压缩编码技术，其尺寸也往往在 MB 级，难以满足需求。就本项目而言，初步目标是要达到 500 倍以上。
2. 计算复杂度与硬件条件之间的矛盾：窄带应用场景往往出现在资源受限的情况下，设备的计算能力极其有限，因此必需要考虑使用轻量化的模型与算法作为解决方案。

3. 信息保真与极端压缩比之间的矛盾：为了实现极端压缩比，不得不采用极高比例的有损压缩编码。进而，我们需要考虑哪些信息是需要保留的，哪些信息是可以丢弃的。
4. 抗噪性：在窄带传输中，信息极易受到噪声影响而丢失。如何在保证极高压缩比的前提下，确保解码后的信息准确性是一大难点。

1.2 解决方案整体架构

1.2.1 目标检测

在应用场景中，可能不是图像中的所有区域信息都是我们需要的。例如，在行人检测中，背景道路是我们不关心的信息。因此，我们可以通过目标检测来将图像中我们不关心的部分隐去，只传输我们关心的区域的信息。

1.2.2 边缘检测

为了保证图片重要信息的前提下实现图像的高压缩比例，我们首先采用边缘检测技术提取图像的边缘信息。

在计算机视觉和图像处理中，边缘检测涉及定位灰度图像的显著变化并识别引起这些变化的物理现象。这些信息对 3D 重建、运动、识别、图像增强和恢复、图像配准、图像压缩等应用非常有用。^[1]

边缘检测 (Edge-detection)：根据维基百科上的解释：图像边缘检测大幅度地减少了数据量，并且剔除了可以认为不相关的信息，保留了图像重要的结构属性。因此，使用图像边缘检测可以有效地提高图像压缩效率。

github 上的 MarkMoHR/Awesome-Edge-Detection-Papers 项目中搜集了众多 edge detection 相关的论文，其中^[2]中提出的基于随机决策森林的方法能够实现实时帧率，同时达到最先进的准确性。这能够实现需要高质量边缘检测和高效率的应用情况。例如，此方法可能非常适用于视频分割或对时间敏感的对象识别任务，如行人检测。

1.2.3 图像压缩

在完成了边缘检测以后，我们需要将图像传入编码器，通过编码器对图像进行压缩。

传统的图像压缩算法有 JPEG, JPEG2000, BPG 等有损压缩算法，也有 PNG, GIF, WEBP 等无损压缩算法。但这些算法在低比特率情形下可能会有严重的块效应。

学术界也有很多基于机器学习的算法，其中很多在压缩效率上超过了传统算法。

基于机器学习的算法中被使用到的主要 backbone 模型有 RNN^[3,4], VAE, INN^[5], GAN^[6], Stable Diffusion^[7], transformer^[8] 等。有的模型是多种方法的结合。

1.2.4 实时性

由于本文的目的是在低比特率下实现更好的信息传输，能够实现满足实时帧率要求的算法会更受青睐。

在边缘检测方面，传统方法，如 canny 算法，sobel 算法等都能简单地实现实时帧率边缘检测。

^[2] 则提出了一种使用随机决策森林算法进行边缘检测的办法，同样可以实现实时帧率，并且实现了极高的准确率。相比于传统方法，其优势是可以考虑到更多的视觉特征，而不是仅考虑颜色梯度。

在图像压缩方面，^[6] 提出的基于 GAN 的算法可以从标签映射中完全合成解码图像中的不重要区域，如街道和树木，从而相应降低存储成本。最终，该方案可以实现在低比特率（低于 0.1 bpp）下效果显著的图像压缩效果，并且相当程度地保证了图片质量。

^[7] 综合使用了 VAE 与 stable-diffusion 技术。先通过 VAE 对图像进行压缩与初步的解压缩，再通过预训练的 stable-diffusion 模型来重建图像。

2 边缘检测

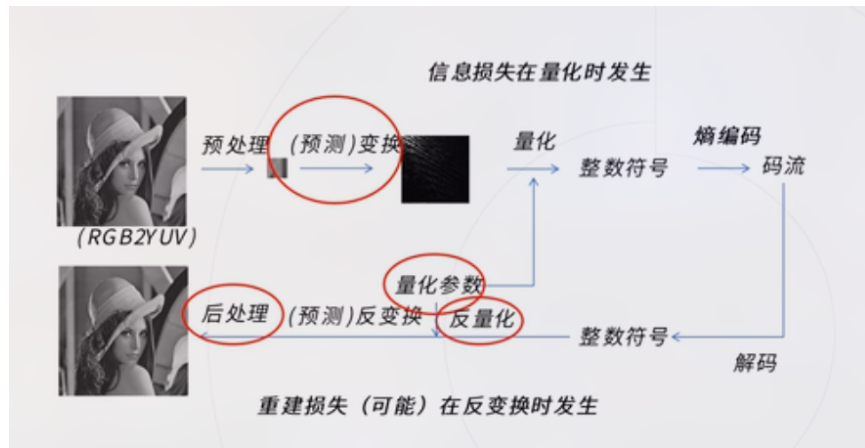
由于传统的 Canny 边缘检测算法在边缘提取任务中已表现出较为满意的效果，且边缘检测并非本次毕业设计的研究重点，为了简化问题并使研究主题更聚焦于图像压缩处理，理应在项目中优先采用 Canny 算法进行边缘检测。这一选择能够有效降低技术复杂度，并将更多精力投入到图像压缩这一核心议题上。

如若 Canny 算法在本项目中的实际应用效果未达到预期，在后续计划中则可以再尝试复现^[2]所提出的快速边缘检测算法。这一算法在计算效率和检测精度上有较大提升，可能更适用于本项目的场景需求。

如果上述方案依然无法获得令人满意的结果，则进一步对边缘检测算法进行深入调研，以探索更符合项目需求的替代方法。分阶段的策略选择旨在高效实现边缘检测的目标，同时将研究重点集中于图像压缩处理的优化上。

3 图像压缩

3.1 图像压缩算法结构



本项目的核心内容在于图像压缩算法，故而本节大致介绍图像压缩算法的结构、主要编码格式。

图像压缩算法主要由预处理、变换、量化、熵编码，以及解码构成。

图像压缩算法的本质是在去除图像中的冗余信息，因此图像压缩算法压缩过程中的每一步都是在去除图像中的一部分冗余。

3.1.1 预处理

假设我们有一张图像，记为 x ，我们先对其进行需要的预处理，以 jpeg 编码为例，这一步要做的是色域转换。之所以进行色域转换，是因为人眼的视杆细胞（对亮度敏感）的数量远多于视锥细胞（对颜色敏感）。因此通过压缩颜色信息，可以有效减少数据量。这个过程是在去除图像的视觉冗余。

3.1.2 变换

而后要做的是变换：如 jpeg 中做的是 DCT(Discrete Cosine Transform) 变换。而 jpeg2000 中做的是 DWT(Discrete Wavelet Transform, 小波变换)。一般化的表达是：图像 x 经过了一个 g_a 函数，从而得到了 y 。DCT 变换的过程实现了图像从空间域到频域的转换，有效地去除了空间冗余。依笔者理解，这一步是整个压缩算法的核心之一。大多数基于人工智能的图像压缩算法就是将这一步换成了神经网络，从而获得了超过传统算法的性能。

3.1.3 量化

再之后的量化过程主要目的在于去除人眼不敏感的频段。经过量化，我们从 y 得到了 q 。这个过程去除了视觉冗余。量化过程是图像失真的根本来源。

3.1.4 熵编码

压缩过程的最后一步是熵编码。熵编码根据数据中不同符号出现的概率分布，对高概率的符号分配较短的编码，低概率符号分配较长的编码，从而减少平均编码长度。这种方式可以更高效地表示图像数据，降低码率 (bit-rate)。

常见的熵编码算法有 Huffman 编码与算术编码。

经过熵编码，我们得到了编码后的码流 H 。

3.1.5 解压缩

解压缩过程整体而言是将这个过程反过来。重新得到 \hat{y} 。在解压缩过程中，熵编码过程得到的码流先被解码，得到整数符号，而后再对整数符号进行反量化，最后进行反变换，从而得到解码图像。

3.1.6 后处理

在有一些算法（如^[7]）中，会在反量化之后进行一些后处理，如：对图像进行重缩放，利用 diffusion-model 或 GAN 重建图像，或应用各种超分辨率 (SR) 提高图像的分辨率。

3.2 传统图像压缩算法/编码格式

图像压缩编码算法分为有损编码与无损编码。常见的编码格式中，JPEG 与 JPEG2000, WEBP, AVC(H.264), HEVC(H.265), 与 BPG(基于 HEVC) 为有损编码（但也支持无损编码），PNG 与 GIF 为无损编码（但 GIF 只支持 256 色，是伪无损编码。）

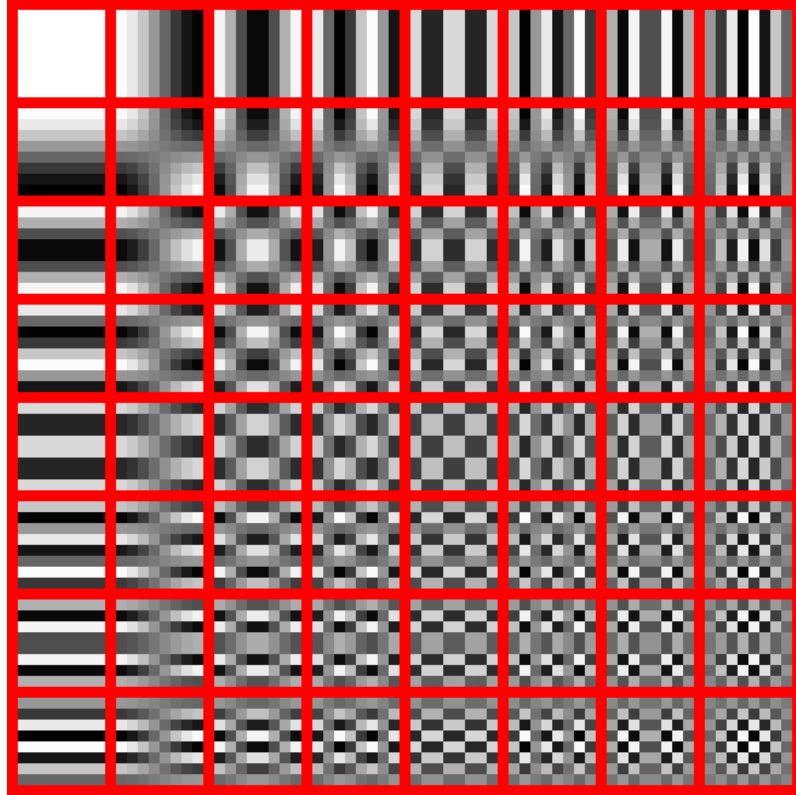
3.2.1 无损图像编码

根据^[9]，以下格式支持无损图像编码：

名称	发布年份	压缩方法与备注
PNG	1992	DEFLATE 字典压缩方法。
JPEG-LS	1998	后续上下文熵编码。
JPEG 2000	2000	离散小波变换 (DWT) 结合上下文算术编码。
JPEG XR	2009	分层离散余弦变换 (hierarchical DCT) 和哈夫曼编码。
WebP	2010	预测器结合 LZ77 和哈夫曼编码。
H.265 (HEVC)	2013	基于空间预测的混合编码，BPG 基于 HEVC。
FLIF	2015	预测器和复杂算术编码器 MANIAC。
AVIF	2018	基于帧内参考像素。
JPEG XL	2020	变换预测、上下文建模和熵编码。

3.2.2 有损图像编码

1. JPEG JPEG 是第一个成标准的有损压缩编码格式，其假设任何图像都可以由 64 幅基本图像的加权余弦之和近似。



其过程是：

- (a) 将图片切割成 8×8 的 block;
- (b) 对每一个 block 做 DCT，得到 transform 之后的系数;
- (c) 对每个系数量化;
- (d) 量化后对每个 block zigzag 扫描，得到最终的 encoded bit stream。

JPEG 的主要缺点是会有比较严重的块效应。

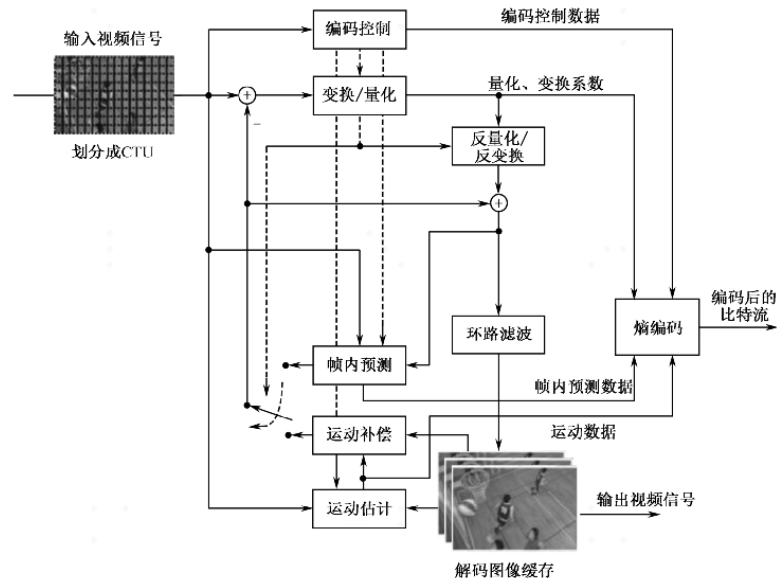
2. 常见有损编码比较 由于 JPEG 在图像压缩上并不高效, 后续人们引起了许多新技术来改进图像压缩技术。

简略而言:

- (a) JPEG->JPEG2000: 使用了小波系数, 利用多尺度分析降低块效应。
- (b) jpeg2000->AVC 帧内预测: 使用相邻的块来预测当前块的内容, 降低空间结构冗余
- (c) AVC 帧内预测->BPG: 使用不同大小在块进行编码, 变化平缓的地方用 32x32, 变化剧烈的地方则用 4x4. 有效减轻块效应。

压缩能力: JPEG < JPEG2000 < AVC 帧内预测 < HEVC 帧内预测 (BPG)

3. H.266/VVC(Versatile Video Coding) H.265 之后最新推出的视频编码标准 VVC, 是目前非基于机器学习的图像压缩技术中的 SOTA 技术。
[10]



其使用混合编码框架，综合了帧内预测，帧间预测，主变换与二次变换结合、标量量化、CABAC 熵编码、环路滤波、率失真优化等多维度的技术。

在其编码结构中，一个视频编码流包含了一个或多个视频编码序列 (Coded Video Sequence, CVS), 一个 CVS 又有多个 AU(Access-unit). AU 又包含多个 PU(Picture-unit)

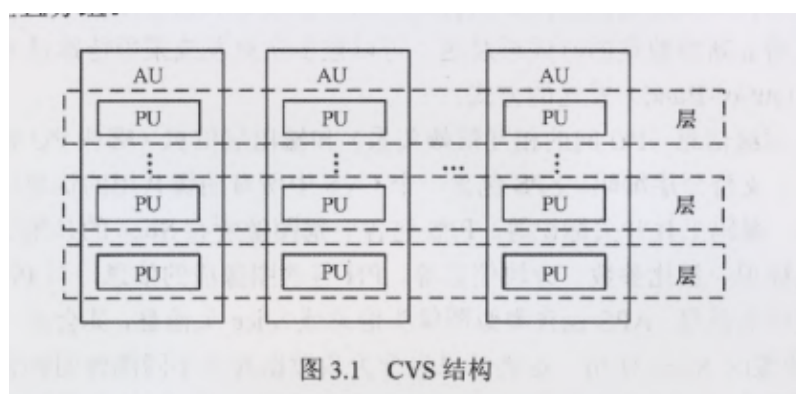


图 3.1 CVS 结构

每个 PU 为一幅编码图像，PU 又分成 Slice, Slice 之间进行独立的编解码。Slice 再由 CTU 组成。

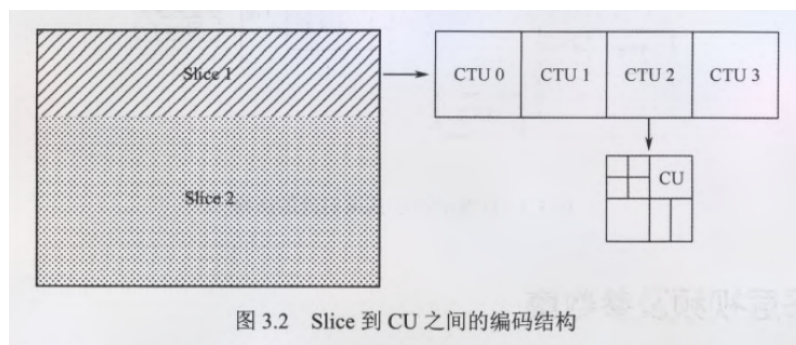


图 3.2 Slice 到 CU 之间的编码结构

3.3 AI 算法

根据^[11]，传统的图像压缩算法主要有以下几个问题：

1. 分块效应 (Blocking Effects): 传统算法通常基于分块的图像处理方法, 将图像分割为小块进行压缩。这种方式容易在解码时引入块状伪影, 尤其是在高压缩比的情况下, 影响图像的视觉质量。
2. 模块之间复杂的依赖性: 压缩算法的各个模块 (如变换、量化、熵编码等) 具有高度的相互依赖性。这种复杂的依赖性使得很难独立优化某个模块, 同时确保整体性能的提升。
3. 整体优化难度大: 由于模型无法整体优化, 即便某个模块有所改进, 可能无法转化为整体性能的显著提升。这种局限性阻碍了复杂框架的进一步改进。

由此, 我们希望引入基于学习的图像压缩编码。来使得潜层系数之间相关性更小, 同时熵编码模型系数概率分布预测更准确, 从而使得压缩编码更高效。

3.3.1 训练

想要使用 AI 算法进行图像压缩, 面临的第一个问题是训练, 大多数基于反向传播算法的 AI 模型要求计算过程整体可导。然而, 在标准图像压缩过程中, 量化这一步要求将连续的像素值或频率转化为离散值, 从而减少信息量, 这一步从定义上就决定了其不可导。

面对量化过程不可导的问题, 有以下解决方案被采用:

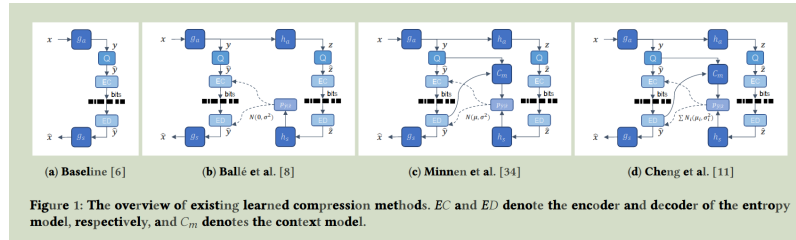
1. 在训练过程中, 使用均匀噪声代替量化操作。这种方法通过在量化前后添加噪声来模拟量化的效果, 从而保持可导性。
2. 在前向传播中直接进行舍入操作, 而在反向传播中使用近似的梯度。这种方法通过在反向传播中忽略舍入的不可导性来简化训练。
3. 采用软量化方法进行训练, 然后在推理阶段使用硬量化。这种方法通过在训练中使用可导的软量化函数来近似硬量化。^[12]

3.3.2 主干网络

在当前基于学习的图像压缩中，最主流的主干网络是变分自编码器（VAE）。除此以外，也有基于 RNN 的算法，如^[3]；基于窗口注意力机制的算法，如^[8]；基于 GAN 的算法，如^[13] 等。

1. 基于变分自编码器的图像压缩算法

VAE 的核心思想是通过编码器将图像映射到一个低维的潜在空间，并利用解码器将该潜在表示还原为原始图像。与传统的编码器-解码器不同，VAE 的编码器输出的是潜在空间的概率分布参数（均值和方差），而不是确定性特征向量。通过对潜在空间的分布施加正则化，VAE 能够消除数据的冗余性，从而实现高效的压缩。进一步的改进方法如 Hyperprior 模型、自动回归先验（Auto-regressive Priors）以及高斯混合可能性模型（Gaussian Mixture Likelihood Model），显著提升了潜在空间的表达能力和对数据复杂分布的适应性。



(a) 熵建模

熵编码的核心目标是通过准确预测潜在表示的概率分布来最大程度地压缩数据。基于 VAE 的图像压缩方法在这一方面提出了多种创新。

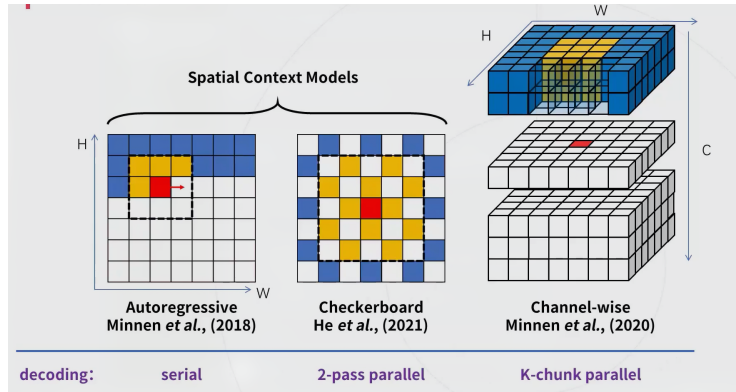
首先，通过引入 3D 上下文熵模型，空间和通道的特征冗余能够被联合建模，大幅提升了概率估计的准确性。同时，基于通道的熵模型（Channel-wise Model）对特征通道分别进行建模，降低了整体的计算复杂度。此外，分层熵模型（Hierarchical Entropy Model）通过逐层建模潜在分布，进一步提升了预测精度。而高斯混合分布（Gaussian Mixture Likelihood）的引入，为捕捉复杂

的潜在表示提供了更大的灵活性，有效改善了模型的重建质量和压缩效率。

基于上下文的自适应二元算术编码（CABAC）以及 Range Coder 等熵编码技术在基于 VAE 的图像压缩中得到了广泛应用。这些方法通过高效利用上下文信息，实现了接近理论最优的压缩性能。CABAC 尤其适用于捕获复杂的上下文相关性，而 Range Coder 提供了更加灵活和高效的实现方案。

(b) 具体实例

ELIC（Efficient Learned Image Compression）^[14] 是基于 VAE 的一种高效图像压缩算法，提出了一系列针对性优化策略。首先，ELIC 引入了改进的上下文模型，通过自回归方法对当前解码的符号进行条件建模。并且，由于串行解码效率较低，ELIC 进一步结合空间通道信息和棋盘格分组技术，显著提升了解码速度。此外，ELIC 利用了非线性分组的策略，通过动态调整通道分组的粒度优化了计算效率。实验表明，仅前 40% 的通道即可包含大部分的语义信息，因此在初期通道使用细粒度分组，而后期则采用粗粒度分组，从而大幅减少了计算开销。



3.3.3 INN

根据^[5]，可逆神经网络（Invertible Neural Networks, INNs）在图像压缩领域的应用呈现出显著的优势，主要体现在其严格的可逆性和高效的信息

保留能力上。INNs 通过其双射映射 (bijective mapping) 特点, 使得输入和输出之间的映射可以精确逆转, 从而在编码-解码过程中有效减少信息丢失问题。此外, INNs 提供可解的雅可比行列式, 这一特性使得可以明确计算后验概率, 为生成式任务中的精确似然估计奠定了基础。

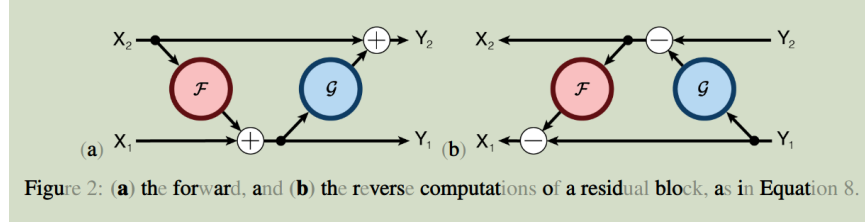
下图是一个 INN 的例子。图源于^[15]。可以发现,

$$y_1 = x_1 + F(x_2)y_2 = x_2 + g(y_1)$$

对应的解码公式为

$$x_2 = y_2 - G(y_1)x_1 = y_1 - F(y_2)$$

这个过程完全可逆。缺点在于, 由于需要同时计算 \mathcal{F} 与 \mathcal{G} , 计算量更大。



INN 在图像领域也有很多的应用。例如, SRFlow 使用条件 INN 架构, 相较于基于生成对抗网络 (GAN) 的方法, 在解决超分辨率问题上表现出更优越的能力。同时, 为了进一步提升 INN 在图像压缩中的表现, 增强型可逆编码网络 (Enhanced Invertible Encoding Network) 通过引入专门设计的特征增强模块和注意力通道压缩层, 不仅提升了网络的非线性表示能力, 还在稳定训练和特征维度调整方面表现出显著的灵活性。^[16]

^[17] 中则利用 INN 的可逆结构, 通过双射 (bijective) 将图像特定丢失内容的分布转换为预先指定的与图像无关的分布, 并生成退化图像, 该可逆框架能够建模丢失信息并在可逆模型中保留分布转换的知识。

尽管 INNs 在信息保留和严格可逆性上表现出独特的优势, 其有限的非线性变换能力仍是一大挑战。因此, 通过优化网络设计和引入辅助模块, 研究者们正不断提升 INN 在图像压缩领域的应用潜力。

3.3.4 GAN

根据^[6], 生成对抗网络 (Generative Adversarial Networks, GANs) 在图像压缩领域的应用展示了其独特的潜力和优势。GAN 通过生成器 G 和判别器 D 的对抗训练机制, 能够有效捕捉数据的全局语义信息和局部纹理结构, 从而在极低比特率下实现高质量的图像重建。

1. GAN 在图像压缩中的框架设计 GAN 在图像压缩中的典型框架通常由编码器 E 、生成器/解码器 G 和量化器 q 组成。编码器 E 将输入图像映射到潜在特征图 w , 随后通过量化器 q 对特征进行有限量化, 生成可以编码为比特流的表示 \hat{w} 。解码器 G 则利用 \hat{w} 重建图像 \hat{x} 。为了解决量化器 q 的不可微问题, 通常引入其可微松弛形式, 以支持反向传播。
2. GAN 在图像压缩中的模式
 - (a) 生成压缩 (**Generative Compression, GC**) 生成压缩旨在保留整体图像内容, 同时通过 GAN 生成高质量的局部细节 (如树叶或建筑物的窗户)。GC 不依赖语义标签图进行训练或部署, 非常适合带宽受限的场景。在这些场景中, 当无法存储原始像素时, GC 使用生成的内容替代, 避免出现块状或模糊伪影。
 - (b) 选择性生成压缩 (**Selective Generative Compression, SC**) SC 结合语义/实例标签图, 针对不重要的图像区域 (如街道、树木) 进行完全合成, 而对用户定义的重要区域 (如人物) 进行高精度保留。这种方法特别适用于视频通话等场景, 通过合成背景减少带宽需求的同时保持核心内容的视觉效果。
3. GAN 在压缩中的优势
 - (a) 对抗损失的引入
通过对抗损失, GAN 在生成压缩伪影较少的图像时表现优越。这种损失能够捕捉全局语义信息和局部纹理, 使得重建的图像既具有整体一致性, 又包含高质量的细节。

(b) 灵活的生成能力

无条件条件和条件 GAN 的结合使得框架既可以进行全局生成（如带宽受限场景下的内容合成），也可以结合语义标签进行选择性压缩，满足多样化需求。

(c) 改进传统失真度量的局限性

传统的失真度量（如 PSNR 和 MS-SSIM）在极低比特率下失去意义，而 GAN 通过生成视觉上更自然的图像，超越了单纯的像素级相似性。

4. 挑战与未来方向

尽管 GAN 在图像压缩中展现了诸多优势，仍然面临一些挑战，例如对抗训练的稳定性、伪影控制以及生成内容的可解释性。此外，如何有效结合上下文模型和先进的编码/解码策略进一步提升压缩性能，也是未来的研究方向之一。通过结合编码器、解码器、量化器与对抗训练机制，GAN 在图像压缩领域为实现高质量、低比特率的图像传输提供了创新路径。

3.3.5 Stable Diffusion

ATTACH

通过Stable Diffusion Based Image Compression 这篇博文中的对比可以看出，stable diffusion 在图像压缩方面的表现要远优于 JPEG 和 WEBP 算法。

此文中对比了 Stable Diffusion 与 JPEG 以及 WEBP 等算法的压缩效率，可以直观地看出，stable diffusion 进行图像压缩效果显著，远好于 JPEG 与 WEBP 压缩算法。

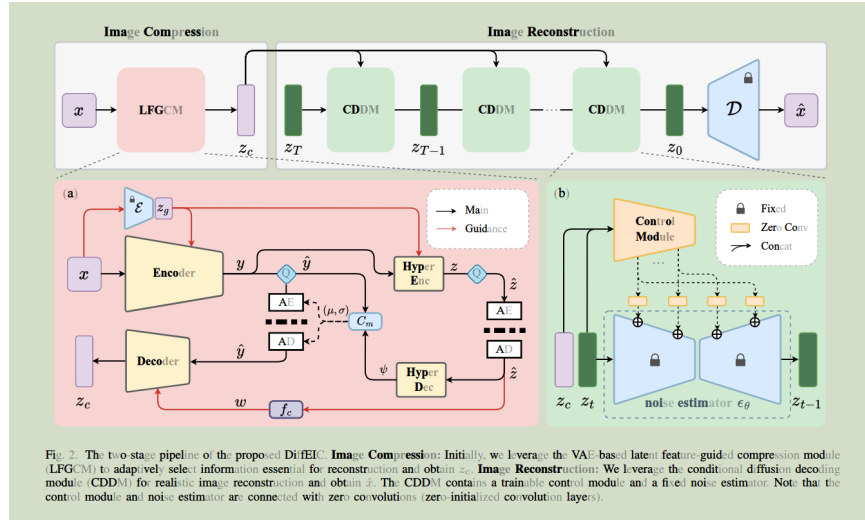
本文的 stable diffusion 主要依赖于三个关键组件：

变分自编码器 (VAE)：负责将原始图像（如 512×512 分辨率，3×8 或 4×8 位深）映射到隐层表示（latent space representation），该表示为更低分辨率（如 64×64），但更高精度（4×32 位）的形式。

U-Net 模型：用于对隐层表示进行降噪（denoise），从而增强重建图像的细节质量。

文本编码器 (**Text-Encoder**): 进一步利用多模态语义信息提升图像重建的生成能力。

不仅如此, 近年来的研究 (如^[7]) 进一步优化了 Stable Diffusion 在图像压缩中的表现。例如, 将^[18] 提出的带有 Hyper Prior 结构的 VAE 用作主干网络, 以更高效地提取图像的隐层表示。在图像重建阶段, 通过预训练的 Stable Diffusion 模型完成高质量图像的生成与还原。



3.4 评价指标

图像压缩算法的性能通常通过一系列定量指标和主观方法来评价, 以平衡压缩率与图像质量之间的权衡。这些指标主要分为以下几类:

1. 传统失真度量指标

- **PSNR(Peak Signal-to-Noise Ratio)**: 峰值信噪比, 衡量原始图像与压缩图像之间像素值的差异。PSNR 数值越高, 图像质量越好。
- **SSIM(Structural Similarity Index Measure)**: 结构相似性评价, 评估原始图像与压缩图像在结构上的相似性。其扩展版本 **MS-SSIM(Multi-Scale SSIM)** 能够在多个尺度上更全面地评估图像的结构保真度。

2. 主观评价指标

- **MOS (Mean Opinion Score)**: 意见平均分, 基于人类视觉感知的主观评分方法, 用于评估图像的视觉质量。

3. 比特率

- **BD-Rate (Bjontegaard Delta-Rate)**: 用于比较两种编码器在不同比特率下的性能, 衡量编码器在压缩率和图像质量之间的平衡能力。
- **Rate-Distortion**: 通过绘制比特率 (Rate) 与失真 (Distortion) 曲线来评估算法。理想的算法点分布在曲线的左上角, 即在更低的比特率下实现更高的图像质量。

在有损压缩 (lossy compression) 中, 算法需权衡压缩率与图像质量, 避免信息丢失过多。然而, 如^[6]指出, 在极低比特率 (低于 0.1 bpp) 下, 传统指标 (如 PSNR 和 MS-SSIM) 可能失效。这是因为这些指标倾向于逐像素保留局部高熵结构 (local structure), 而忽略纹理和全局语义信息的保真度。在这种情况下, 对抗损失 (**Adversarial Loss**) 被认为是一种更有效的评价方法, 能够捕捉图像的全局语义和局部纹理信息, 从而生成视觉上更吸引人的高质量图像。

3.5 总结

在图像压缩方面, 为满足本项目低带宽情形下高效图像传输要求, 一个合理的做法是使用 GAN 或 Stable Diffusion 这类生成算法, 先将图像进行压缩, 而后使用生成模型对图像进行重建。

如此一来, 在保证关键信息可以得到传输的前提下, 接收端可以自行生成不重要的信息, 这有效地在低比特率条件下保证了高质量的信息传输。

对于基于 INN 的算法, 由于需要更高的算力去去持, 可能不完成适用于宽带宽低息传输。

基于注意力机制的算法, 由于计算复杂度与输入大小呈平方关系, 其在实时性和资源受限的场景下表现受限。

4 实现计划

4.1 基础框架搭建与初步实验

- 框架设计
 - 使用 Canny 算子提取图像边缘信息，作为图像的关键结构特征。
 - 使用 Stable Diffusion 模型的降级版本进行生成式重建。此阶段的目标是验证生成模型在图像重建中的基本可行性。
- 数据集准备
 - 使用公共图像数据集（如 Flickr-2W 或 COCO 数据集），涵盖多种场景与复杂度，为模型训练提供广泛样本支持。
- 初步实验与调整
 - 通过训练基础模型，对压缩率与图像失真之间的关系进行初步评估。
 - 使用 PSNR（峰值信噪比）、SSIM（结构相似性指数）和 MOS（主观评分）等指标评估图像质量。

4.2 模型优化与架构探索

- 模型复杂度升级
 - 在初步框架基础上，引入更多生成模型的优化版本（如扩散模型的优化算法、轻量化 GAN）以提升生成效果。
 - 设计自适应压缩方案，根据图像内容自动调整保留的关键信息比例。
- 寻找平衡点
 - 在压缩率与图像失真之间找到最优点，确保关键信息的有效传输，同时实现尽可能高的压缩率。
 - 测试不同场景下的性能，例如低带宽网络和多种设备类型。

4.3 实验记录与总结

- 实验过程的系统化记录
 - 对每次实验的模型参数、训练数据、性能结果进行详细记录。
 - 使用表格和可视化工具（如折线图或散点图）分析压缩率与图像失真的关系。
- 论文写作
 - 梳理研究思路，从问题背景、算法设计、实验结果到未来展望，形成完整的学术论文。
 - 引用相关领域最新成果，与自己的方法进行对比，突出创新性与应用前景。

4.4 最终成果与应用验证

- 模型落地应用
 - 开发一个演示系统，模拟图像压缩与传输流程，并展示在接收端的重建效果。
 - 将模型部署到资源受限设备上（如树莓派或嵌入式平台），验证其实用性。
- 成果展示
 - 在学术会议或毕业设计答辩中展示成果。通过清晰的实验结果和视觉对比图，阐述算法的优势。

- [1] ZIOU D, TABBONE S. Edge detection techniques-an overview[C]. 1998.
- [2] DOLLÁR P, ZITNICK C L. Fast edge detection using structured forests[A/OL]. arXiv, 2014. DOI:10.48550/arXiv.1406.5549.

- [3] TODERICI G, O' MALLEY S M, HWANG S J, et al. Variable rate image compression with recurrent neural networks[A/OL]. arXiv, 2016. DOI:10.48550/arXiv.1511.06085.
- [4] TODERICI G, VINCENT D, JOHNSTON N, et al. Full resolution image compression with recurrent neural networks[A/OL]. arXiv, 2017. DOI:10.48550/arXiv.1608.05148.
- [5] XIE Y, CHENG K L, CHEN Q. Enhanced invertible encoding for learned image compression[A/OL]. arXiv, 2021. DOI:10.48550/arXiv.2108.03690.
- [6] AGUSTSSON E, TSCHANNEN M, MENTZER F, et al. Generative adversarial networks for extreme learned image compression[A/OL]. arXiv, 2019. DOI:10.48550/arXiv.1804.02958.
- [7] LI Z, ZHOU Y, WEI H, et al. Towards Extreme Image Compression with Latent Feature Guidance and Diffusion Prior[A/OL]. arXiv, 2024. DOI:10.48550/arXiv.2404.18820.
- [8] ZOU R, SONG C, ZHANG Z. The devil is in the details: Window-based attention for image compression[A/OL]. arXiv, 2022. DOI:10.48550/arXiv.2203.08450.
- [9] BARINA D. Comparison of lossless image formats[A/OL]. arXiv, 2021. DOI:10.48550/arXiv.2108.02557.
- [10] 新一代通用视频编码 H.266VVC: 原理、标准与实现 [A].
- [11] HU Y, YANG W, MA Z, et al. Learning end-to-end lossy image compression: A benchmark[A/OL]. arXiv, 2021. DOI:10.48550/arXiv.2002.03711.
- [12] AGUSTSSON E, MENTZER F, TSCHANNEN M, et al. Soft-to-hard vector quantization for end-to-end learning compressible representations[A/OL]. arXiv, 2017. DOI:10.48550/arXiv.1704.00648.

- [13] ZHONG Z, CHAI L, ZHOU Y, et al. Faithful extreme rescaling via generative prior reciprocated invertible representations[C/OL]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 5708-5717. DOI:10.1109/CVPR52688.2022.00562.
- [14] HE D, YANG Z, PENG W, et al. ELIC: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding[A/OL]. arXiv, 2022. DOI:10.48550/arXiv.2203.10886.
- [15] GOMEZ A N, REN M, URTASUN R, et al. The reversible residual network: Backpropagation without storing activations[A/OL]. arXiv, 2017. DOI:10.48550/arXiv.1707.04585.
- [16] LUGMAYR A, DANELLJAN M, GOOL L V, et al. SRFlow: Learning the super-resolution space with normalizing flow[A/OL]. arXiv, 2020. DOI:10.48550/arXiv.2006.14200.
- [17] XIAO M, ZHENG S, LIU C, et al. Invertible rescaling network and its extensions[A/OL]. arXiv, 2022. DOI:10.48550/arXiv.2210.04188.
- [18] BALLÉ J, MINNEN D, SINGH S, et al. Variational image compression with a scale hyperprior[A/OL]. arXiv, 2018. DOI:10.48550/arXiv.1802.01436.