arXiv:2305.01857v1 [cs.IT] 3 May 2023

迈向文本变换编码*

Tsachy Weissman 电气工程系
斯坦福大学 tsachy@stanford.edu

2023年5月4日

摘要

受到近期关于面向年轻人类及由年轻人类参与的压缩、信息处理中的基于变换的方法取得的成功，以及强大的基于语言的人工智能兴起的启发，我们提出 *textualtransform coding*。它与传统的基于变换的编码共享一些关键特性，而后者构成了当今许多多媒体压缩技术的基础。它可作为压缩的基础，使比特率低至直到最近仍被认为低得毫无用处的水平，并在更传统的比特率下提升人类对重建结果的满意度。

1 上下文与动机

1.1 我们的压缩器现状

我们的音频、图像和视频压缩与流传输技术，是否正逐渐逼近某种率失真复杂度帕累托前沿，抑或仍有很大的改进空间？鉴于人们对采集、存储和传输此类数据的需求日益增长，以及点云和来自自动驾驶车辆的多传感器数据等更加庞大的数据类型的出现，过去十年间，这个问题一直在我的研究组及其他地方酝酿。

1.2 香农会怎么做?

在 [Sha48], 香农发表相关论文后，他将注意力转向了英语的可压缩性。除其他见解外， [Sha51] 他提出，一旦我们的算法预测模型赶上 Mary E. Shannon夫人大脑中的模型，英语文本应当能够被压缩到约 1.3 比特/字符。后续研究 [CK78; RTT19] 表明，这一数字在很大程度上适用于英语使用者。我们的技术花了七十多年才赶上，并实现了 [Bel21; Kno21]。

受这一发展历程的启发，我们希望了解：当多媒体数据压缩领域实现以下条件后，人类认为能达到怎样的效果：

- 我们的算法模型赶上人脑中的模型
- 我们的码本／编码／解码充分利用人类公开可用的“边信息”
- 我们根据人类真正关心的内容定制重建结果

2018 年夏季，我们很幸运地接待了 3 名高中生实习生，他们专门尝试解决这些问题。

*改编自一场题为“向人类学习：如何改进有损数据压缩”的讲座。该讲座于2022年夏季在芬兰埃斯波举行的国际信息论研讨会上发表。

1.3 图 像压缩：人工编码器、解码器和评分者

其中一名实习生（人类编码器）拿到了一张此前从未见过且网上没有的照片。他的任务是通过内置的 Skype 文字聊天功能（关闭自己的音频/视频输出），用文字向另一名实习生（人类解码器）描述这张照片。解码器可以利用 Skype 的屏幕共享功能，将部分且尚在进行中的重建结果分享给人类编码器[1]。我们请人类评分者比较了人类压缩系统下的重建结果质量与使用 WebP 时的质量。在大多数图像类型上，在相近比特率的限制下，人类压缩器取得了比 WebP 更高的质量 [Bho+19a；Bho+19b；Fis+21]。

1.4 SHTEM

受该项目启发，我们于2019年夏季启动了面向高中和社区学院学生的 SHTEM 暑期实习项目 [SHT23]。H 既代表人文学科，也代表人的因素；项目强调跨学科研究，将人文学科与科学、技术、工程和数学相结合，并运用人的判断来评估、指导和衡量我们的技术应当努力实现的目标。

SHTEM 2019年的几个项目延续了前一个夏季的项目：其中一个项目致力于从警方画像师绘制图像的方式中获取启示，探索提升人脸图像压缩效果的潜力 [Cha+19]。另一个项目则在 [Bho+19a] 的方案中移除了作为解码器的人，同时将比特率进一步降低了一个数量级，而且没有降低人类满意度。该项目让人工编码器用 Python 代码描述图像，使计算机能够运行这些代码并生成重建图像 [Mah+19]。

1.5 视频

2020年夏季的 SHTEM 项目之一，对后来成为 StageCast [Ro21],的项目起到了关键作用。该项目诞生于疫情初期，探索并开发了相关技术，使来自不同地理位置的演员阵容成员和观众能够进行实时表演。其理念是检测并流传输人脸和人体关键点的位置，然后由解码器 [Pra+21]重建其数字人偶。所需比特率的大幅降低，使此前借助现有（完整视频流传输）技术无法实现的实时互动成为可能。该项目还提出了一种类似的低比特率方案，用于流传输真实视频，在重建端用原人物的深度伪造 [PS21] 取代数字人偶。显然，英伟达在 [Max]中采纳了这一建议。

胃口被吊起后，在 [Tan+23] 中，我们继续探索一种框架，旨在将视频会议压缩至更低的比特率范围。其核心在于认识到：在典型的远程会议视频中，一旦背景和发言者都已被学习，包括该发言者如何依据其话语内容发声和移动，从信息论角度看，剩下的主要内容就是其话语的内容。我们设计了一种利用现有技术的视频流系统（例如从音频中提取文本以及根据文本合成语音的技术），在所需比特率低三个数量级的情况下，仍实现了与现有标准相近的用户满意度。我们的 SHTEMers 实现了一个可在标准网页浏览器上运行的版本 [TW23]。

在 [Per+22] 中，我们利用人工输入训练了一个小型深度神经网络，使其能够预测视频中的重要性区域，并据此引导比特率分配，从而提升现有视频编解码器（x264）的性能。自然而然，学到的重要性区域对应于可用一两个词语标记的物体。

1.6 文本的涌现

这项工作主要由 SHTEMers 及其导师完成，揭示了改进现有多媒体数据压缩技术的巨大潜力，而人类语言也已成为评估并实现这一潜力的关键。与此同时，自然语言处理和机器学习取得了长足进展，强大的语言模型不断涌现（例如 ChatGPT、

[1]中间重建结果基于从编码器流出的比特，因此这类“反馈”方案是合理的有损压缩器。

LaMDA、Bard 和 Copilot）、基于文本描述生成多媒体数据的生成器（例如 DALL-E2）、生成多媒体数据文本描述的生成器（例如 inGPT4）等。

2 寻求一种新变换

我们最广泛使用的信息处理技术都基于变换：变换、在变换域中处理、逆变换。最常用的是线性变换（快速傅里叶变换、小波等），而非线性变换近年来也日益发挥重要作用 [Bal+21; Vas+17]。

2.1 有效变换

在特定应用中，有效变换具备以下全部或大部分特征：

- 变换域中的系数对应于有意义的元素（基函数）（在生物学、物理学、感知或概念层面上）
- 变换系数具有稀疏性，且彼此之间关系简单（不相关、独立、弱相关等）
- 正向变换和逆变换相对于相关度量的平滑性（即，两个“相似”的输入在变换域中仍然“相似”，反之亦然）
- 低复杂度（正向变换和逆变换均如此）

An 此类变换的另一项优势在于，它们能够导出易于处理且符合现实的数据模型。

2.2 将快速傅里叶变换作为寓言

最著名的有效变换示例，是以傅里叶命名的变换。一位经验丰富的音频工程师只需快速浏览一段音频的傅里叶变换，就能判断它是否是一段音乐；如果是，还能辨认出哪些乐器演奏了哪些音符、背景噪声的强度等。变换域中的峰值对应于主导频率，而主导频率又对应于“谐波”，即与我们的听觉系统产生共振的信号分量。

在傅里叶上下文中，谐波对我们而言具有明确意义：我们知道它在示波器上是什么样子，知道它对我们的听觉系统而言是什么感觉（听起来如何），并且有一种（数值上的）方式来标记它（频率）。那么，我们概念上的“谐波”是什么？什么“事物”会激活我们神经元中非平凡的子集？

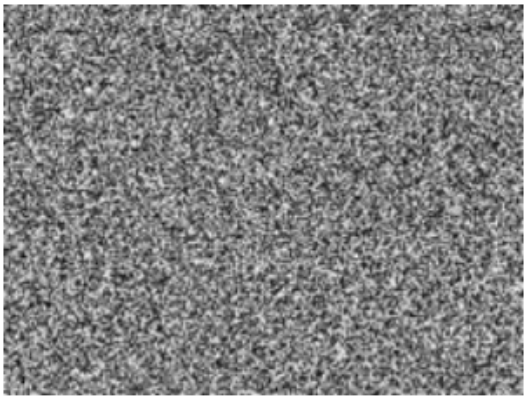
答案或许是我们的词语，它们指代并描述着与我们产生共鸣的概念或事物。在图像数据的上下文中，以下类比或许能助我们一臂之力：

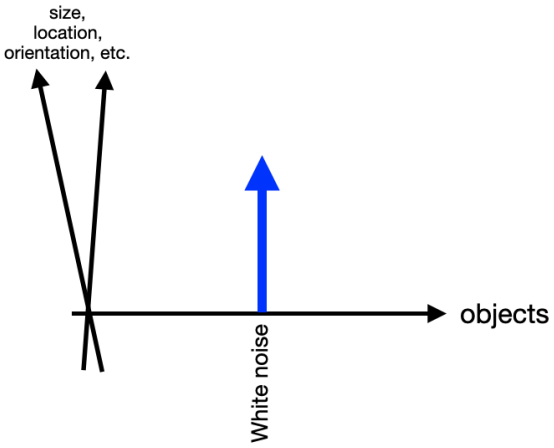
傅里叶变换	新变换
谐波	我们有词语指代的物体
频率	词语
谐波的振幅和相位	物体的大小、位置和方向
振幅和相位的精度	大小和位置的精度
频率分辨率	每个对象的词语数量
逐分量保真度	对所讲述故事的保真度
FFT	计算高效的图像描述

词语对于我们寻求的变换至关重要，正如频率对于傅里叶变换至关重要。

2.3 定性示例

不妨想想，我们要寻找的变换在几个熟悉的极端情形下可能是什么样子。先来看白噪声。经过傅里叶变换或小波等其他正交变换后，它仍然是白噪声。但在新变换域中，它会完全集中在一个元素上（即“狄拉克δ函数”）：白噪声。接下来，我们来看看图2中的《南方公园》图像，以及一幅卡通



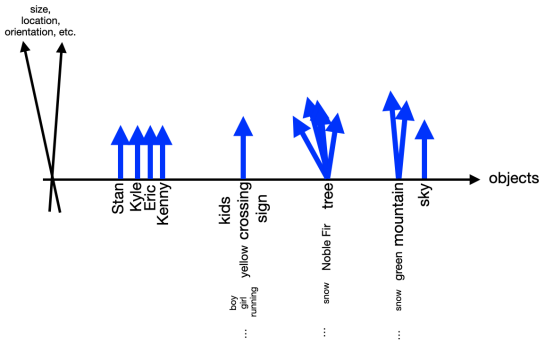


（a）在传统变换域中（仍为白噪声）

（b）在文本变换域中

图1：白噪声

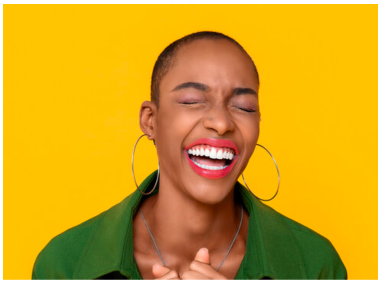




（a）在原始域中

（b）在文本变换域中

图2：南方公园图像



(a) 源数据

Woman with shaved head
brown skin
and hoop earrings
laughing merrily
with her eyes closed
in a green shirt
against a yellow background

(b) 文本表示



（c）DALL-E2 重建结果

图3：通过文本变换编码实现极限压缩

关于它在这种变换域中可能呈现的某种形式。关键在于，某些类型的图像——其中所讲述的故事比逐像素保真度更为重要——在这种变换下能够获得稀疏近似表示。这些图像可能包括节日贺卡、婚礼照片、卡通图像、计算机生成图像、宣传海报和广告。

2.4 文本变换

从定性层面转向具体层面，一种可能的（有损）图像表示方式，可采用图1和图2所暗示的那类变换域，包含以下要素：

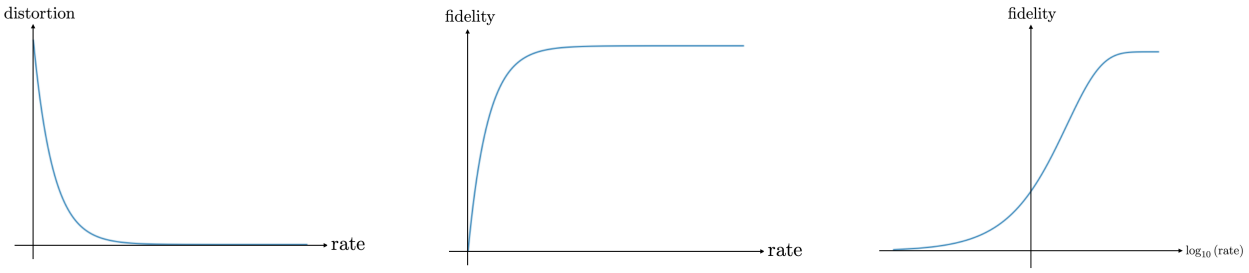
- 对物体数量的描述（不超过规定数量 L，以控制认知负荷）
- 列出这些物体是什么（每个用一个词语表示）
- 它们的大小、位置和方向（按规定的物理分辨率 R）
- 对每个物体的描述词语，以及它们之间的关系（按规定的“文本分辨率” W）

对 (L、 R、 W) 的不同选择，对应不同的比特率和描述质量。这些要素都可以完全用文本描述，归根结底就是对图像进行文字描述，并将每个物体可用的词语数、句子数、总长度等限制在给定范围内。

这种变换似乎满足第2.1小节所列的要求，包括相对于文本相似性度量的“平滑性”（参见 [WD20] 及其中的参考文献）和“低复杂度”。用于实现此类变换（图像描述）及其逆变换（从文本生成图像）的技术正在迅速发展 [Osm+23; GCV21]。

这种变换对于有损压缩尤其有用。它构成压缩表示，如第1.3小节所述，但无需人工参与。除了满足良好变换的各项要求外，这种变换还赋予压缩表示另一层意义：它可供人类阅读和查询。在这里，编码和解码各自都具有独立价值：编码是对数据所作的、人类可读的文本摘要；解码器则根据文本描述，在原始数据域中生成一个点。

图3展示了这一思想的一种极端形式，利用现有技术即可实现。它表明，按传统意义上的每符号（像素）比特数衡量的压缩率，可以低得近乎荒谬，同时重建结果在某些应用中仍可能令人满意。或许更有意义的是，压缩表示所需的总比特数以及实现的复杂度，取决于所讲述故事的细节丰富程度，而非图像的物理（像素）分辨率。



(a) 失真率函数　(b) 保真率函数　(c) 保真度与对数 $_{10}$（速率）

图 4：在经典符号级或像素级标准下的失真率函数与保真率函数

2.5 经典机制与文本机制

香农率失真理论关注的是逐符号失真标准。图4a展示了无记忆高斯源在平方误差失真下的失真率函数，它具有代表性：对于任何非退化信源和逐符号失真标准，定性上都会呈现相似的形态。速率以每个源符号的比特数衡量，与总比特数随所表示的源符号数量线性增长的情形相对应。速率足够大时，失真可以任意小（事实上，对于离散信源，当速率达到或超过熵率时，失真为零）。等价地，也可以将保真度定义为失真的相反数，此时零失真对应最大保真度。图4b展示了图4a中的失真率函数所对应的保真率函数；在本例中，最大保真度设为信源的方差。图4c则展示了另一种等价表示，其中速率采用对数尺度。这清楚地表明，在这种逐符号失真的框架下，极低的信息率至多只能使保真度从其最低值略微提升。

这种行为与我们在多媒体数据压缩中观察到的、基于文本的重建所带来的人类满意度表现截然不同，如图5所示。我们在第1.3小节讨论的初步实验表明，即使比特率极低——在横轴采用对数比特率的图中远处于左侧——也足以显著提升人类满意度水平，使其超越最低值。第1.4小节提及的后续工作表明，将比特率再降低一个数量级，仍可实现相似的重建质量。我们目前的实验（尚未公开）表明，通过文本变换编码，在段落、少数句子（如图3所示）、少数词语等粒度下，进一步将比特率降低若干数量级，仍能达到切实为正的人类满意度水平。

3 后续步骤

3.1 与人类满意度一致的保真度度量

我们猜想，在多媒体数据的上下文中，给定一个源及其可能的重建结果，人类满意度主要由两个因素决定。第一，是对所讲述故事的保真程度，这可以通过文本域中的相似性来衡量。第二，是传统意义上的保真度。换言之，人类满意度可以通过如下形式的函数得到较好的近似或预测：

$$s(x, \hat{x}) = g\left(T_f, P_f\right), \tag{1}$$

其中，x 是源文本，$\hat{x}$ 是其重建结果，$T_f(x, \hat{x})$表示在相应文本变换之间测得的相似度（T_f 代表“文本保真度”），而 $P_f(x, \hat{x})$则表示一种传统类型的相似度度量，例如峰值信噪比（PSNR）或结构相似性指数（SSIM）（P_f 代表“逐像素保真度”）。图5表明，高满意度主要可以通过较高的传统保真度来实现，并且在这种情况下，满意度在很大程度上由传统保真度决定。与此相反，当传统保真度较低时，满意度水平将在很大程度上取决于文本域中的保真度水平。换言之，假设不失一般性，T_f 和 P_f 是

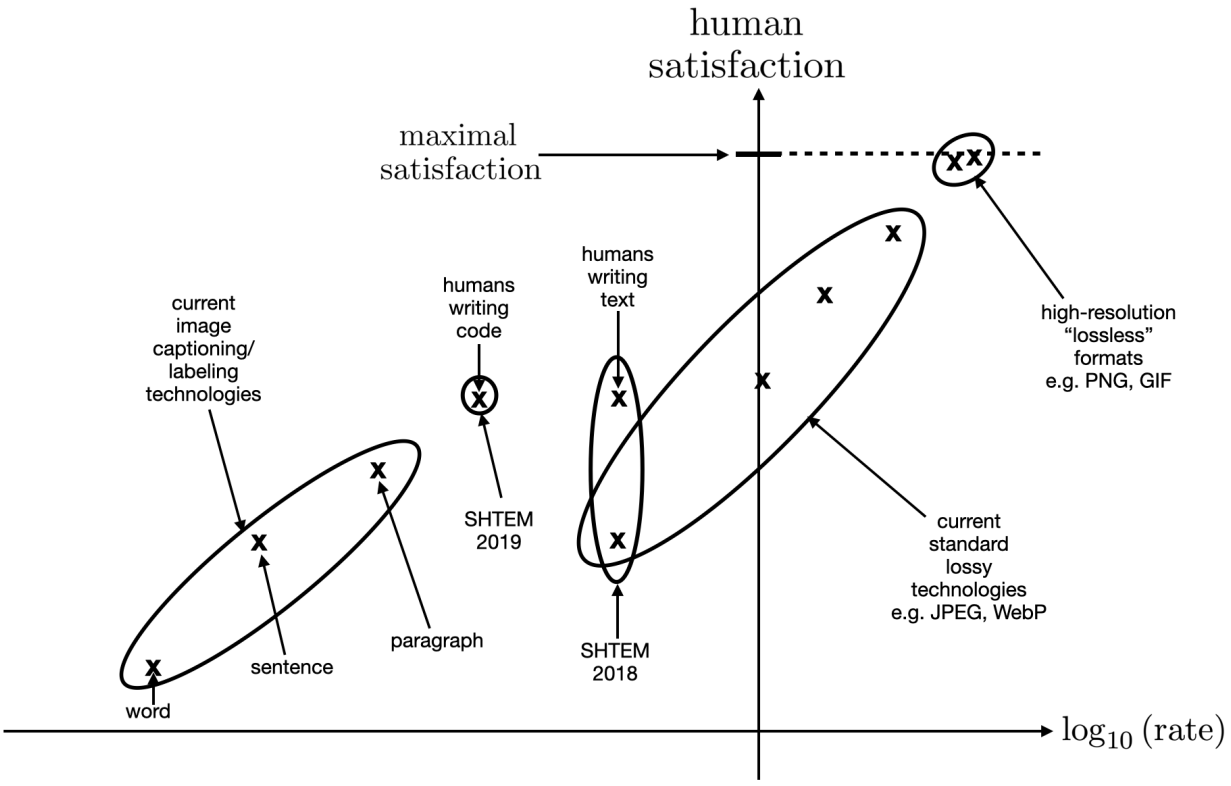


图5：可实现的人类满意度水平

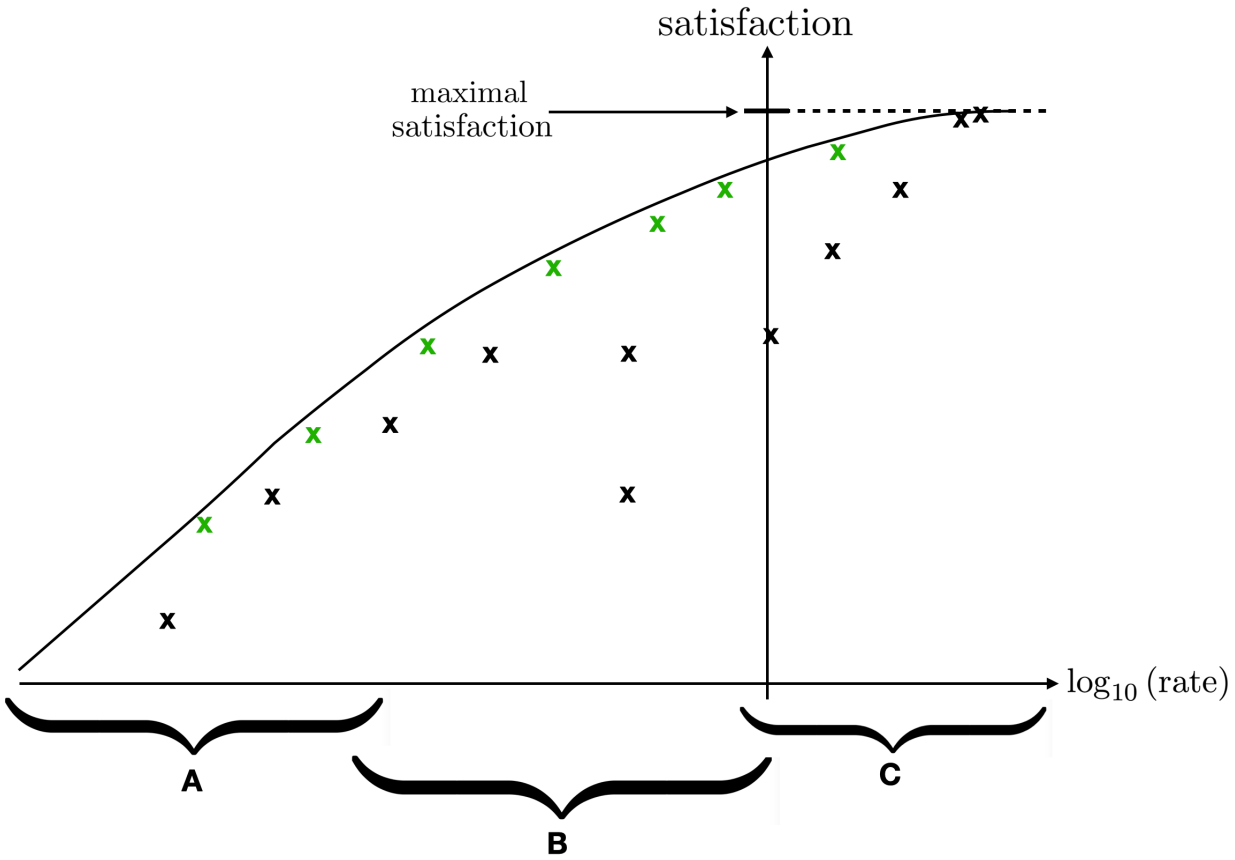


图6：由实线曲线表示的 $S(R)$；黑点来自图5，绿点表示有待开发的方案。A 是超低码率区域，文本变换编码在其中至关重要；C 是目前大多数压缩技术所处的区域；B 是一个尚未充分探索的中间区域，需要采用新方法来衔接文本与传统机制。

归一化至 $[0,1]$，$g(\alpha,\ \beta)$ 后，很大程度上将由 α 决定：当 β 较低时，以及当 β 较高时的 β。我们计划在 SHTEM 2023 项目中安排几个项目，用于衡量和量化人类满意度，并通过实验评估我们的猜想以及函数 g 的形式是否成立。

3.2 表征并逼近 S（R）

如前所述，文本变换可作为一种有用的数据建模框架。与符号逐一处理的方法不同，后者往往会产生过于简单、无法反映实际数据的模型，或过于复杂、难以发挥作用的模型；我们可以从文本域入手，因为该域适合进行相当简单且符合实际的建模，然后再建立一个从文本到原始域的生成模型。采用这种方法，并确定前一小节讨论的满意度函数后，我们便可以运用香农理论工具在文本域和生成域中发展成熟的方法，刻画满意度与码率（的对数）之间可实现的最佳权衡。

在构造性方面，对 $S(R)$进行刻画将有助于并指导新型实用方案的开发。第1.6小节提及的那类文本技术取得进展，将直接转化为超低码率区域中方案的改进。另一个同样值得关注的领域，是连接超低码率区域与传统的逐符号（逐像素）保真度区域的广阔地带，而目前大多数技术都处于后者。这一区域很可能受益于新方法：有效利用并结合来自文本域的信息，以及来自更传统技术的低分辨率版本，如图6所示。

3.3 超越压缩

3.3.1 文本变换与隐私

文本变换编码所实现的超低码率机制，不仅能大幅节省空间／带宽，也有利于保护隐私。通过文本表示源数据，并将其压缩至仅需几个 k 位，便可轻松获得隐私保证：图像与其表示之间的互信息上界为 k。另一种做法是让这些 k 位表示一组关于源数据的查询答案；这些查询经约定不属于隐私信息，并且可以自然地从文本域中提取。近期已有工作沿着这一方向展开，见 [Guo+22a]。

3.3.2 去噪

可以设想某些应用和数据类型，在这些场景中，能够在文本域内以尤为显著且有效的方式进行去噪。例如，只需从文本描述中删除“嘈杂”一词，或将其替换为“清晰”，就可能大幅提升质量；这种提升可以根据第 3.1 小节中的满意度函数进行测量和优化。

4 结论

文本变换是一个在概念上有用且日益实用的框架，可用于多媒体信息处理；它基于一种经过人类进化多年优化的编码。它生成可由人阅读和搜索的表示形式，其实现复杂度有界，不会随像素分辨率等传统物理特征而增长。它或许是实现前所未有的低比特率压缩与流传输的关键，也有助于刻画比特率与人类满意度之间的基本权衡。

本文以图像为例进行说明，但这些思想同样适用于其他传统和新型多媒体数据类型。我们的工作与近期其他尝试将感知和语义概念融入传统压缩与通信范式的研究一脉相承，参见 [BM19；Zha+21；Wan+22；Guo+22b；Gun+23；Agu+23] 及其中的参考文献。该领域的进一步发展很可能来自信息科学家、工程师、神经科学家和心理学家之间的跨学科合作。

5 致谢

我感谢许多为我们的 SHTEM项目不断取得更大成功而付出努力的人。特别感谢 Cindy Nguyen、Sylvia Chin 和 Suzanne Sims，自2019年夏季以来，她们一直不知疲倦、无私地负责各项事务；感谢 Shubham Chandak博士、Kedar Tatwawadi博士、Pulkit Tandon博士、Irena Hwang博士、Yanjun Han博士、Devon Baur、Noah Huffman、Dmitri Pavlichin博士、Jay Mardia 以及许多其他斯坦福教师和学生，为各个项目提供了卓越的指导；也感谢 SHTEMers 所做的鼓舞人心的工作。感谢 Lara Arikan 在图3中提供示例并参与富有启发性的讨论。

参考文献

[Sha48] C. E. Shannon. “通信的数学理论”。载于：贝尔系统技术期刊（1948）。[Sha51] C. E. Shannon. “印刷英语的预测与熵”。载于：贝尔系统技术期刊30.1（1951），第 50–64 页。 DOI: 10.1002/j.1538-7305.1951.tb01366.x。[CK78] T. Cover 和 R. King. “英语熵的收敛赌博估计”。载于：*IEEE*信息论汇刊24.4（1978），第 413–421 页。 DOI: 10 . 1109 / TIT. 1978 . 1055912。

[RTT19] Geng Ren、Shuntaro Takahashi 和 Kumiko Tanaka-Ishii。《通过使用 Mechanical Turk 进行的大型认知实验估计英语的熵率》。载于：Entropy 21.12（2019年12月），第 1201 页。 ISSN：1099-4300。 DOI：10.3390/e21121201。 URL：http://dx.doi.org/10.3390/e21121201.[Bel21] F. Bellard。NNCP。https://bellard.org/nncp/。2021年。[Kno21] B. Knoll。CMIX。http://www.byronknoll.com/cmix.html。2021年。[Bho+19a] Ashutosh Bhown、Soham Mukherjee、Sean Yang、Shubham Chandak、Irena Fischer-Hwang、Kedar Tatwawadi 和 Tsachy Weissman。《人类仍是最佳有损图像压缩器》。载于：2019年数据压缩会议（DCC）。2019年。 DOI：10.1109/DCC.2019.00070。[Bho+19b] Ashutosh Bhown、Soham Mukherjee、Sean Yang、Shubham Chandak、Irena Fischer-Hwang、Kedar Tatwawadi、Judith Fan 和 Tsachy Weissman。《改进有损图像压缩：利用公有领域图像进行人类图像重建》。2019年。arXiv：1810.11137[eess.IV]。[Fis+21] Irena Fischer-Hwang、Shubham Chandak、Kedar Tatwawadi 和 Tsachy Weissman。《别管 JPEG：人会如何压缩图片？》。载于：IEEE Spectrum（2021年）。[SHT23] SHTEM。暑期实习项目。https://compression.stanford.edu/outreach/shtem-summer-internships-high-schoolers-and-community-college-students。2019 年 - 2023年。[Cha+19] H. Chau、T. Pauly、D. Phan、R. Prabhakar、L. Quach 和 C. Vo。《构建以人为中心的人脸图像有损压缩器》。载于：The Informaticists Journal for High Schoolers（2019年）。URL：https://theinformaticists.com/2019/08/28/building-a-human-centric-lossy-compressor-for-facial-images/.[Mah+19] V. Mahtab、G. Pimpale、J. Aldama 和 P. Truong。《基于人类的图像压缩：使用确定性计算机算法重建预分割图像》。载于：The Informaticists Journal for High Schoolers（2019年）。 URL：https://theinformaticists.com/2019/08/29/human-based-image-compression-using-a-deterministic-computer-algorithm-to-reconstruct-pre-segmented-images/.[Ro21] M. Rau 等人。StageCast：表演与技术实验。https://taps.stanford.edu/stagecast/。2021年。[Pra+21] Roshan Prabhakar、Shubham Chandak、Carina Chiu、Renee Liang、Huong Nguyen、Kedar Tatwawadi 和 Tsachy Weissman。《利用关键点提取和数字木偶操控降低视频流媒体的延迟与带宽》。载于：2021年数据压缩会议（DCC）。2021年。 DOI：10.1109/DCC50243.2021.00057。[PS21] Swathi P 和 Saritha Sk。《深度伪造的创建与检测：综述》。载于：2021年第三届国际创新计算应用研究会议（ICIRCA）。2021年，第 584–588 页。 DOI：10.1109/ICIRCA51532.2021.9544522。[Max] NVIDIA Maxine。https://developer.nvidia.com/maxine。[Tan+23] Pulkit Tandon、Shubham Chandak、Pat Pataranutaporn、Yimeng Liu、Anesu M. Mapuranga、Pattie Maes、Tsachy Weissman 和 Misha Sra。《Txt2Vid：通过文本实现人头说话视频的超低比特率压缩》。载于：IEEE Journal on Selected Areas in Communications 41.1（2023年），第 107–118 页。 DOI：10.1109/JSAC.2022.3221953。[TW23] Arjun Barrett；Laura Gomezjurado；Shuvam Mukherjee；Arz Bshara；Sahasrajit Sarmasarkar；Pulkit Tandon 和 Tsachy Weissman。《Txt2Vid-Web：基于网页的文本转视频会议流程》。2023年。 URL：https://sigport.org/documents/txt2vid-web-web-based-text-video-video-conferencing-pipeline。[Per+22] Evgenya Pergament、Pulkit Tandon、Oren Rippel、Lubomir Bourdev、Alexander Anderson、Bruno Olshausen、Tsachy Weissman、Sachin Katti 和 Kedar Tatwawadi。PIM：使用感知重要性图进行视频编码。2022年12月。 DOI：10.48550/arXiv.2212.10674。

[Bal+21] Johannes Ball´e、Philip A. Chou、David Minnen、Saurabh Singh、Nick Johnston、Eirikur Agusts- son、Sung Jin Hwang 和 George Toderici。“非线性变换编码”。载于：IEEE《信号处理精选主题期刊》15.2（2021年），第 339–353 页。 DOI：10.1109/JSTSP.2020. 3034501。[Vas+17] Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan Gomez、Lukasz Kaiser 和 Illia Polosukhin。“Attention Is All You Need”。载于：神经信息处理系统进展。2017年。[WD20] J. Wang 和 Y. Dong。“文本相似度测量：综述”。载于：Information 11.9（2020年）。 DOI：10.3390/info11090421。[Osm+23] Asmaa A. E. Osman、A. Wahby Shalaby Mohamed、Mona M. Soliman 和 Khaled M. Elsayed。“图像描述中基于注意力的模型综述”。载于：《国际高级计算机科学与应用期刊》14.2（2023年）。 URL：https://www.proquest.com/scholarly-journals/survey-on-attention-based-models-image-captioning/docview/ 2791786135/se-2。[GCV21] Federico Galatolo、Mario Cimino 和 Gigliola Vaglini。“通过 CLIP 引导的生成式潜在空间搜索从描述生成图像及反向生成”。载于：国际图像处理与视觉工程会议论文集。SCITEPRESS - Science and Technology Publications，2021年。 DOI：10.5220/0010503701660174。 URL：https://doi.org/10.5220%2F0010503701660174。[Guo+22a] Tao Guo、Jie Han、Huihui Wu、Yizhu Wang、Bo Bai 和 Wei Han。“ 利用高效密钥保护语义信息 ”。载于：2022 年 IEEE 国际信息论研讨会（ ISIT ）。2022 年，第 2660–2665 页。 DOI：10.1109/ISIT50566.2022.9834462。[BM19] Yochai Blau 和 Tomer Michaeli。《重新思考有损压缩：率失真感知权衡》。2019年。arXiv：1901。07821 [cs.LG]。[Zha+21] George Zhang、Jingjing Qian、Jun Chen 和 Ashish J Khisti。“面向有损压缩的通用率失真感知表征”。载于：神经信息处理系统进展。A. Beygelzimer、Y. Dauphin、P. Liang 和 J. Wortman Vaughan 编。2021年。URL：https://openreview.net/forum?id=_wdgJCH-Jf。[Wan+22] Yizhu Wang、Tao Guo、Bo Bai 和 Wei Han。“语义通信系统中的估计-压缩分离”。载于：2022年IEEE信息论研讨会（ITW）。2022年，第 315–320 页。 DOI：10.1109/ITW54588.2022.9965794。[Guo+22b] Tao Guo、Yizhu Wang、Jie Han、Huihui Wu、Bo Bai 和 Wei Han。《带边信息的语义压缩：率失真视角》。2022年。arXiv：2208。06094 [cs.IT]。[Gun+23] D Gunduz、Z Qin、IE Aguerri、HS Dhillon、Z Yang、A Yener、KK Wong 和 C-B Chae。“超越比特传输：上下文、语义与面向任务的通信”。载于：IEEE Journal on Selected Areas in Communications 41（2023年），第 5–41 页。 DOI：10. 1109/JSAC.2022.3223408。 URL：http://dx.doi.org/10.1109/JSAC.2022.3223408。[Agu+23] Eirikur Agustsson、David Minnen、George Toderici 和 Fabian Mentzer。《借助条件生成器实现多真实感图像压缩》。2023年。arXiv：2212。13824 [cs.CV]。