arXiv:2304.01919v1 [cs.HC] 4 Apr 2023

viz2viz：使用扩散模型进行提示词驱动的风格化可视化生成

吴佳琪、John Joon Young Chung 和 Eytan Adar

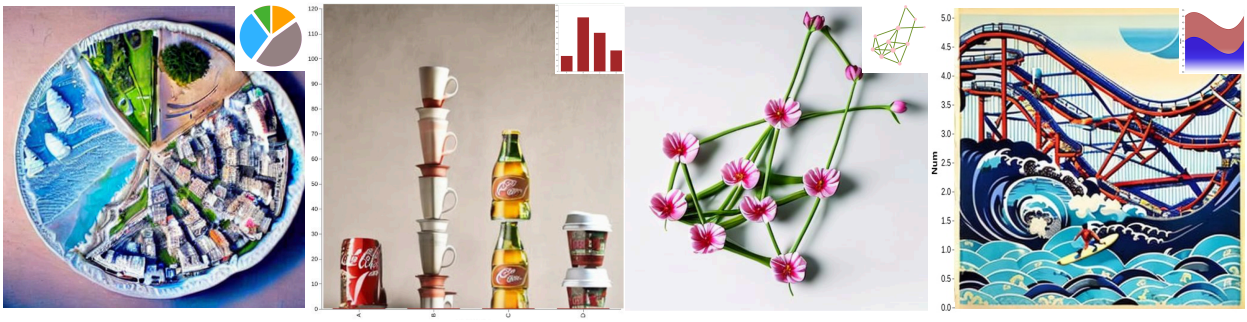


图 1：viz2viz 生成的样例（嵌入原始数据）。从左至右：绿意盎然、树木繁茂的森林鸟瞰图，蓝色海洋与船只，建筑物林立的灰色城市，橙色沙漠；写实风格的一摞摞红色可口可乐罐、棕色茶杯、玻璃葡萄酒瓶、星巴克纸杯；写实风格的粉色郁金香；浮世绘风格的红色木制过山车侧视图，浮世绘风格的蓝色海浪与冲浪者。请注意，本文中的某些示例所选提示词旨在展示 viz2viz 的能力，而非体现其美学或设计品质。

摘要——创建风格化可视化需要超越大多数工具生成的有限、抽象的几何标记。设计师会构建风格化惯例，在其中，标记不仅经过转换（例如，用蜡烛的照片代替条形图），还会被合成为一个“场景”，从而拓展传统可视化的边界。为此，我们介绍 viz2viz：一个通过文本提示词将可视化转换为风格化形式的系统。该系统遵循一套高层次配方，运用多种生成方法生成新的可视化，同时保留原始数据集的属性。尽管这套基础配方适用于多种可视化类型，我们还展示了如何针对不同类型的可视化（条形图、面积图、饼图和网络可视化）进行专门调整。我们的方法引入了针对不同标记使用不同提示词的技术（即每个条形都可以完全不同），同时仍能保持图像的“连贯性”。最后，我们评估了这一方法，并讨论其扩展方向与局限性。

索引词—风格化可视化、生成算法、Stable Diffusion

1 引言

大多数可视化软件旨在通过简单、抽象的二维图形标记准确地再现数据：条形图中的矩形（▭）；饼图中的弧线（◅）；以及散点图中圆形（○）、矩形（▭）和偶尔出现的星形（☆）的组合 [28]。在许多情况下，这些表示方式无法满足创意设计构想。图形上的“装饰”需要额外的样式处理和设计工作。设计师可能会通过用其他表示方式替换简单的抽象标记，来对可视化进行风格化：用蜡烛代替条形图中的条形，用水果切片代替饼图中的弧线，用花卉代替散点图中的圆形。除了简单地替换标记外，可视化样式设计还可以采用其他艺术形式和媒介（如草图、绘画、照片、黏土等）。我们特别关注那些在标记呈现方式上实现显著风格变化（例如，层叠的咖啡杯），但仍能以原有惯常形式（例如，条形图）辨认出来的风格化形式[1]。由于许多此类表示方式独具特色，设计师要实现自己的构想，就需要额外的工具、技能和流程。例如，要手动制作图 1 中的图表，设计师需要创建实体模型并拍摄照片，或者使用复杂的绘画工具。制作单个实例已相当耗时，更不用说围绕设计反复迭代了。为此，我们提出 viz2viz：一套利用生成式方法创作这些图像的工作流程。我们的方法让设计师能够以“初始”可视化（基础条形图、面积图、饼图和网络图）为起点，通过简单的文本提示生成风格化可视化。

在这项工作中，我们基于生成图像的系统（如 DALL-E 2 [22]、Stable Diffusion [25]、和 Midjourney）开展研究。这些模型功能强大，但单独使用时难以“约束”。如果不提供任何指导，生成结果便与底层数据毫无关联。例如，可以要求生成“一张由点燃的蜡烛组成、带有背光效果的条形图照片”，但蜡烛的高度不会对应任何实际的数据分布。img2img [19] 或 ControlNet [38]等新型架构允许我们输入一张初始图像（如标准条形图）。然而，这些方法存在相反的问题：源图像的限制过多——以高大的二维矩形为引导时，系统无法生成一摞咖啡杯。我们的目标是生成风格化可视化，使其既能灵活呈现广泛的创意构想，又能遵循数据的底层分布。例如，云朵形状的饼图扇区仍应准确体现原始数据中的比例。如果数据要求如此，代表网络中两个节点的花朵就应彼此相连。借助 viz2viz，我们提供一套适用于多种输入（包括数据和提示词）的“配方”，以创作照片级真实和非照片级真实图像。为确定风格化可视化可能涵盖的范围，我们利用人工创作的实例构建了这类形式的分类体系。

我们方法的一个优势是，它不依赖于明确的

• 吴佳琪、*John John Young Chung* 和 *Eytan Adar* 任职于密歇根大学。电子邮箱：*wujiaq*、*jjyc*、*eadar@umich.edu*。

[1]这类可视化有许多实例，但似乎没有一个统一的名称（有些称为装饰性可视化、信息图或照片可视化）。我们将这一广义类别称为风格化可视化。

使用可视化（普通或风格化）进行训练。这一点至关重要，因为很难找到足够多的示例，同时包含数据、描述和风格化形式。相反，我们专注于寻找拆解可视化普通形式中所用标记的方法。通过按需“解构”可视化，viz2viz 可以有针对性地应用不同的生成管线（例如，将条形图变成建筑物的粗略形状）。完成后，viz2viz 可以重建连贯的可视化（例如，将多栋建筑物并排排列），并应用额外的生成步骤，生成构图更完整的图像（例如，更连贯的天际线）。在这项工作中，我们描述了如何通过组合图像处理程序与经过修改的生成管线（例如，depth2Img、图生图、ControlNet 等），实现由草图、合成和精修步骤构成的基本高层配方。viz2viz 方法展示了如何将传统图像处理（例如，掩模处理、变换/平移等）与生成元素结合使用，从而实现一种受控的创造性。除了四种特定的可视化类型（条形图、区域图、饼图和网络图），我们还提供了一些指导，说明如何为其他类型构建特定的工作流程。虽然本研究重点关注将文本 提示词用作驱动生成系统的方式，但我们也介绍了向其他提示机制扩展的可能性。

我们在本研究中的贡献如下：

- 识别风格化可视化的设计空间和分类体系。
- viz2viz 的设计与实现：一种包含特定工作流程的通用配方，用于支持风格化可视化的创作。
- 将 viz2viz“配方”修改为不同类型的图表和提示的指南。

2 相关工作

两条研究路线为 viz2viz 提供了启发：（1）可视化文献，重点关注由人类和机器驱动的风格化可视化生成方法；（2）基于扩散模型的图像生成。

2.1 可视化构建

风格化可视化主要借助特定媒介的工具创作，而这些工具通常在使用可视化软件生成内容后才发挥作用（例如，Matplotlib、R、Excel、Altair 等生成的图像文件）。可视化生成工具很少支持大规模风格化。例如，在一项涵盖 40 多种工具的调查中，只有三种支持“信息图定制”[18]。因此，设计师会使用矢量图形软件（如 Adobe Illustrator）或“栅格”编辑器（如 Adobe Photoshop）来创建风格化形式 [2]。为此，人们开发了新工具，以衔接可视化软件与矢量工具（如 D3 和 Illus- trator [3]）。在编辑器中，矢量表示可以替换为图标、矢量艺术，并通过各种方式加以修饰。Data Illustrator 等工具 [17], 、Charticulator [24] 以及（在一定程度上）Tableau，旨在同时提供可视化和设计功能。这些工具可用于创建高度风格化的矢量艺术风格可视化。不过，它们都不支持位图编辑，而处理照片或其他媒介时可能需要这一功能。对于栅格图像（如照片），风格化处理更为复杂，但可以借助图像编辑器实现。图像操作（如复制、粘贴、平移和变换）既可用于实现所需外观、确保图像符合编码要求（例如，使用适当的尺寸编码底层数据），也可用于叠加惯用结构（例如，添加坐标轴或标签）。

为可视化设计样式的另一种方法，是将模板或样式应用于基本形态。在某些情况下，风格化形态是通过创建与数据绑定的手工制作示例来驱动的（例如，[14, 35]。例如，DataInk [35] 允许最终用户手绘图形符号（例如一棵草图树）。这类工具具有很高的灵活性，并支持不同的可视化惯用形式。不过，它们仍然需要一定的设计技能，而且可能需要进行后处理，以确保设计具有多样性。InfoNice [33] 则采用了另一种方法，即使用定制的（由人设计的）图形符号，例如一个部分填充的酒瓶。其他方法（例如 Diatoms [4]）可以提出新颖但可能较为抽象的图形符号类型。我们对 viz2viz 的设想是，设计师对他们希望生成的标记已有概念，这些标记不一定是图形符号（无论抽象与否）。

使用机器学习创建可视化的想法已通过不同方式得到应用 [16, 34, 39]。然而，大多数方法旨在根据数据推荐可视化类型，或修改现有可视化（例如添加注释或标题）。与之最相关的示例提供了进行某种可视化风格迁移的方法（例如，[5, 7, 20, 29]）。尽管这些方法功能强大，但它们大多局限于已有示例的情形，或只能将一种简单的标记类型或表达方式转换为另一种 [10]。值得注意的是，迄今所述系统的一个共同局限在于，它们无法生成照片级真实感可视化。

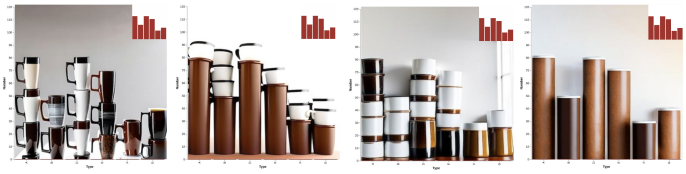
在生成可视化时使用照片或摄影图像是一个棘手的问题。DataQuilt 系统为设计师提供了一种使用图像来设计可视化风格的方法 [37]。用户可以从现有图像来源中借用组件和风格化特征，并将它们转化为可映射数据的元素（例如，根据某个数据属性调整提取出的咖啡杯图像的大小）。大多数“现成”的图像可能与底层数据并不相符。例如，设计师很难甚至无法找到或创建与底层数据比例相符的物体照片（例如，拉伸的瓶子、阴影恰当的一摞杯子等）。Infomages 系统提出了另一种方法，即让主题照片作为可视化的“背景”（例如，用于展示化妆品销售情况的嘴唇照片）[6]。Infomages 允许用户以半自动方式将传统图表数据嵌入现有图像中（例如，将嘴唇的不同部位着色，形成伪饼图）。尽管该系统能够找到可能适用于数据的图像，但它不会生成这些图像，而是将可视化“填充”或“叠加”在图像之上。前景可视化标记与背景结构之间的交互不在 viz2viz 的研究范围之内。

我们的 viz2viz 旨在从标准图表类型出发，用自然语言描述所需的样式。因此，这与使用自然语言进行可视化构建有着天然联系（参见 [30] 中的详尽综述）。尽管这类工具大多利用语言（和数据）生成标准可视化形式，但有些工具提供了更丰富的模板集（例如 [8]）。Lumina [13] 等系统采用了一种有趣的方法：它们试图“理解”数据所代表的内容（例如某个国家或特定人物），并据此提供样式（例如国旗或该人物的图像）。类似地，MetaGlyph [36] 提出根据数据语义构建矢量化字形（例如，当数据涉及快餐食材时，使用分层汉堡图形）。对于 viz2viz，我们假设基准可视化已预先构建完成（使用最合适的工具），并且可以通过自然语言提示驱动风格化过程，而无需示例或模板形式。

最后，我们注意到，可视化设计师已开始利用生成式方法来构建可视化。这些方法主要用于生成背景或剪裁图像，以便将其放置在现有标记中（例如，为地图中的每个州生成一张代表该州州花的图片）。大多数实验都指出，为确保输出符合输入分布，需要进行大量额外的后处理步骤。这些实验尚未成功映射到特定的图表形式，也未能持续一致地编码数据。然而，这些实验催生了极具创意的作品（例如，[9]），并表明我们需要更好的工具。我们提出 viz2viz 来应对这一挑战。我们定义了一种通用“配方”，支持风格化可视化生成，使其既能广泛适用于各种样式和形式，又能确保与原始数据保持一致。

2.2 扩散模型

如今有多种生成式图像模型正在使用。然而，我们特别关注扩散概率模型：这类模型逐步对加噪变量进行去噪，从而得到感兴趣的样本。这种方法由 Sohl-Dickstein 等人提出 [31]，但在效率和可控性方面已取得显著进展。例如，**潜在扩散模型（LDM）** [26] 在图像的低维压缩表示上执行扩散过程



(a) viz2viz (b) 图生图 (c) depth2img (d) ControlNet

图 2：使用不同方法生成的白色桌子上装满咖啡的逼真咖啡杯堆叠效果。我们将 (a) 我们的系统与 (b) 图生图（强度=0.8）、(c) depth2img（强度=0.95）和 (d) ControlNet（使用 Canny 边缘检测，强度=0.95）进行了比较。

（潜变量）（而非实际像素空间），以提高训练效率。

借助新的扩散模型，还可以通过文本输入或提示 [22,27]执行图像生成任务。这是通过使用 CLIP 等预训练的对比式视觉语言模型，将文本输入编码为引导向量来实现的 [21]。将这种方法与潜在扩散模型相结合，**Stable Diffusion（SD）**便成为一种潜变量文本到图像扩散模型，能够根据任意文本输入生成照片级真实感图像。由于 SD 是开源的，社区迅速发展了这一方法，以生成更好的结果。例如，SD 用户经常调整的一个参数是引导尺度 [11], ，它决定文本输入对扩散过程的引导强度。为了实现更强的可控性，开发者设计了许多工作流，即使用 SD 模型端到端的图像生成流程。在构建 viz2viz 时，我们采用了一种通用配方来生成风格化方案。不过，根据数据、提示词和所需图表形式的特性，每种配方都需要对扩散模型进行一些调整。在许多情况下，我们会借鉴并扩展现有方法，以满足特定约束。

图像到图像（**图生图**）[19] 是一种由额外输入图像进行条件控制的可调控方法。在图生图中，它不再从随机潜图像开始扩散过程，而是将添加了噪声的输入图像用作初始潜变量。因此，生成图像在视觉上与初始图像相似，同时受到引导提示词的影响。在图生图中，通常使用一个强度 值（0-1）来调整初始图像的影响程度。基于图生图的**图像修复** [1]会将图像掩模作为额外输入，使生成图像仅在特定区域生成。其他方法则会对生成图像的视觉结构施加更严格的条件 [12, 15] **2** [26] 。例如，深度图到图像生成（**深度图图像**）允许使用初始输入图像的深度图（通过 MiDaS [23]推断）作为扩散过程的条件。与图生图类似，depth2img 允许用户通过强度指定转换程度，但它能更好地保留原图的结构信息。此外，Zhang 和 Agrawala [38]提供了一种添加条件控制模块的方法，可加速模型调优。这种方法，即 **ControlNet**，允许使用多种类型的条件图像（例如深度图、边缘或分割信息），为生成结果提供结构信息。在生成风格化可视化时，当 depth2img 和图生图无法生成连贯且风格化的可视化结果时，可将搭配深度图或 Canny 边缘检测的 ControlNet 作为替代方案。

正如我们将在下文详细介绍的，直接将现有模型应用于可视化问题存在诸多困难：既要满足风格化提示词，又要保留图表中的精确信息，并不容易。单独使用这些模型无法稳定地产生可靠结果（图 2）。例如，采用较低强度时，图生图（图2b）会保留初始图像的结构信息（这是优点），但无法根据提示词转换条形图，只是添加了“噪声”。采用较高强度时，图像可能会发生变化，却无法遵循底层数据的分布。结构条件方法能够保留可视化数据，但往往过于严格地遵循结构信息，无法将标记的形状风格化（例如，图 2d）。此外，如果用户希望可视化中的多个对象标记具有特定的大小、位置和形状（例如，第一个条形由咖啡杯构成，第二个由瓶子构成，等等），现有的可控方法就很难满足这些要求。借助 viz2viz，我们可以通过组合并改造这些现有工作流，更可靠地生成风格化可视化。我们的输出既符合数据，也符合提示词。

3 风格化可视化的分类体系

为了了解风格化可视化的领域，我们收集了 250 多个示例。我们纳入了使用标准图表形式、但标记呈现为“具象”形式的示例。也就是说，这些示例使用抽象几何结构以外的对象来表示数据。为收集这些示例，我们搜索了“信息图”和“照片可视化”等词语，并纳入符合我们定义的示例。我们还利用 Pinterest 的推荐引擎寻找更多示例。

作者将这些图像分组，并归纳出三个主要类别。图 3 展示了基于真实案例制作的若干合成变体。这些图像使用 DALL-E 2 生成，之后在 Illustrator 中进行了少量修改，以添加标签并突出某些标记。这些图像展现了生成式方法的潜力。然而，由于我们没有为这些示例指定数据，其分布并无意义。我们在补充材料中将这些示例与 viz2viz 生成的示例进行了比较。

值得一提的是我们排除了哪些内容。有些可视化可能经过了风格化处理，但采用的是比本文所用定义更宽泛的标准。Federica Fragapane、Giorgia Lupi 和 Stefanie Posavec 的作品可以说具有高度风格化的特点。然而，她们的设计常运用新颖的抽象形式或图表形式。我们也不将抽象标记转换为另一种媒介中的相同标记视为本文所讨论的风格化（例如，用油画颜料绘制的条形图）。这些属于“风格迁移”，而非我们关注的风格化（对于这类情况，简单使用 ControlNet 或许就足够了）。下文将介绍我们用于归纳这些示例的三个主要标签。

3.1 风格与媒介

风格与媒介指所采用的艺术风格和媒介（颜料、摄影、雕塑等）。由于这一类别的变化极为多样，我们不会在其中进一步划分子类（艺术分类远超本研究的范围）。不过，我们在收集到的图像中发现了一些反复出现的主题。

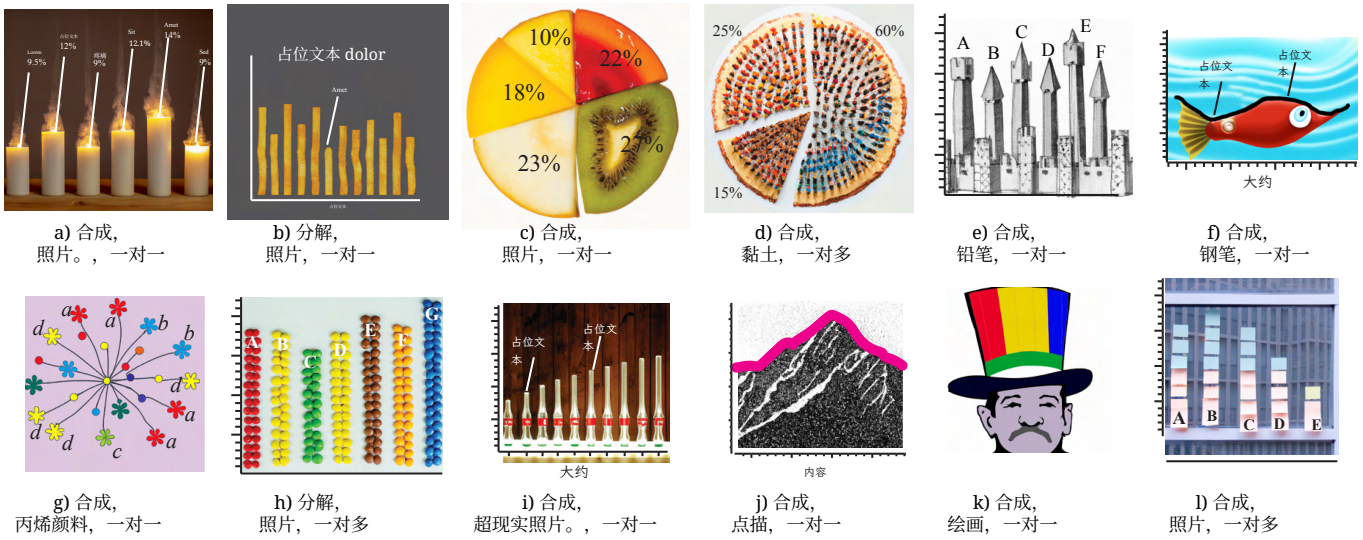
我们发现，许多设计采用实体对象的照片来制作可视化。Sarah Illenberger、Peter Ørntof 和OliverUberti 等艺术家，以及 TheDataMadeofThings项目（https://datamadeofthings.tumblr.com/）都代表了这种设计风格。图 3 中的示例包括图像 *a-c, h, i*，以及 *l*。我们将这类作品与使用其他媒介（如黏土）制作后再拍摄的图像区分开来。

非摄影风格化通常会用到某种颜料或绘画媒介 [32]。在现实案例中，这些可视化通常由手工绘制（以数字方式或在实体媒介上完成）。数据新闻记者 Mona Chalabi 创作了这类作品，Nigel Holmes 的部分作品也可归入此类。图中的示例包括 *e, f, g, j* 和 *k*。这类作品的吸引力在于，它们可以呈现非照片级真实的表现形式（例如，将时间序列表现为一条绘制出来的鱼的背部）。虽然超现实物体也可以通过照片呈现（例如*i*中的拉长瓶子），但设计师和观众可能更偏好非摄影媒介。

虽然我们注意到手工制作的风格化可视化中存在一些共同主题，但我们强调，这些主题的流行或许受到某些限制因素的影响。例如，某些文化审美偏好和规范可能导致某些图像风格出现得过于频繁。更重要的是，有些图像可能仅凭生成式工具之外的方法难以创作。例如，一种看起来像照片却带有超现实主义风格的可视化，若没有大量图像编辑工作，可能无法实现。由于创作难度很高，我们可能很少见到这类作品。生成式方法或许会改变风格化可视化的格局。

3.2 合成与分解

我们的第二个类别基于场景连贯性。具体而言，我们关注可视化是合成 的还是分解的。对于合成图像，表示标记的物体和



a) 合成，照片。，一对一　b) 分解，照片，一对一　c) 合成，照片，一对一　d) 合成，黏土，一对多　e) 合成，铅笔，一对一　f) 合成，钢笔，一对一

g) 合成，丙烯颜料，一对一　h) 分解，照片，一对多　i) 合成，超现实照片。，一对一　j) 合成，点描，一对一　k) 合成，绘画，一对一　l) 合成，照片，一对多

图3：基于广泛的“现成”类别创作的各种风格化可视化示例。这些图像由我们使用 DALL-E 2 生成（例如，一张由点燃的蜡烛组成、置于桌子上并由背光照明的条形图照片），但未使用特定的数据提示词。我们在 Illustrator 和 Photoshop 中修改了生成图像，以添加注释（例如标签、坐标轴）。

背景会以某种方式发生交互（例如，物体彼此叠放、某个条形图元素的阴影落在另一个元素上等）。此类别的示例包括图 3 *a*、*c-g*、*i* 和 *j-l*。在这些示例中，物体要么彼此“接触”，要么属于同一场景（同一张桌子上的蜡烛、同一扇窗户中的便利贴，或同一顶帽子中的条形）。分解图像可能具有相似的标记，但物体之间以及物体与其背景之间并不发生交互。例如，如果我们分别拍摄条形图中每个元素的照片，再将它们裁剪并粘贴到一张图像中，我们会将其视为分解可视化（例如，图 3*b* 和 *h*）。

3.3 一对一与一对多

最后，我们根据标记的构成方式对可视化进行分类。具体而言，我们关注它由“什么东西”构成。在许多情况下，原始表现形式中的一个标记，在风格化可视化中也由一个标记表示。一个条形对应一根蜡烛或一根薯条，或一个山峰对应一个时间序列，都是这样的例子（见图3 *a-c*、*e-g*和*i*、*j*、*k*）。另一类是一对多的例子，即一个标记由若干较小的“部分”合成，例如将一个饼图切片表示为一群人，或将一个条形图表示为许多糖果（见*d*、*h*和*l*）。一对多的例子常见于象形图，其中使用图标来构成一个“标记”。

在某些情况下，这一类别可能存在歧义。例如，我们可能用山脉来表示时间序列。山脉（整体）由多座山峰（部分）合成。在这种情况下，我们选择根据我们认为视觉解释最“简单”（从格式塔意义上说）的对象（此处为整片山脉）对可视化进行分类。

viz2viz 的目标是尽可能广泛地支持生成这类风格化可视化。除了确保标记正确编码数据外，我们还希望生成合成和解构的图像、照片级真实和非照片级真实的图像，以及由单个或多个物体构成的标记。

4 VIZ2VIZ：生成风格化可视化

我们提出了 viz2viz，这是一种利用基于扩散的图像生成模型来生成风格化可视化的工作流程。由于可视化输入和风格化输出存在广泛差异，viz2viz 的核心是一个可适配不同场景的中央“配方”。viz2viz 基于相同的三步框架，该框架结合了现有及新颖的架构与工作流（图 4）。虽然有一个中央核心配方，但我们发现，没有一种工作流程适用于所有可视化类型和所有提示。为此，我们提供了多种替代路径，设计师可以借此实现期望的结果。我们将完全自动化的工作流程子程序选择留待未来工作。一旦定义完成，viz2viz 工作流程便可在无需干预的情况下端到端运行。不过，设计师可以在不同环节介入，以实现不同效果（例如，在运行工作流程的其余部分之前修改可视化草图）。

作为输入，viz2viz 接收普通可视化、提示词以及其他控制变量。普通输入可视化（条形图、饼图等）可以通过各种工具生成（在我们的实现中，我们使用了多种基于 Python 的工具包）。提示词是描述我们希望应用的风格化效果的文本。在 viz2viz 中，提示词由不同部分组成，可在扩散步骤中有选择地应用。具体而言，一个提示词包括：(a) 应用于整个过程的上下文提示词（例如，一张条形图的照片）；(b) 一个或多个描述标记应转换成何种对象的子提示词（例如，所有条形都应变成点燃的蜡烛，或者第一根条形应变成摩天大楼，第二根条形则应变成宝塔）；(c) 可选的背景提示词（例如，一座图书馆 或简单的渐变）。

4.1 viz2viz流程

viz2viz 的第一步是**草图**。在这一步中，viz2viz 会创建粗略的风格化可视化或草图。如下文所述，viz2viz 会将每个标记（以及背景）拆分为单独的组件，并为每个组件生成风格化表示（例如，条形图中的每个条形都会根据提示词进行转换）。这样可以确保标记不会彼此融合；当它们彼此接近或接触时，就可能发生这种情况。重新组合后，草图可视化在视觉上可能不够协调，因为各个部分的光照或风格可能有所不同。随后，草图可视化会进入第二步：**合成**。这一步使用扩散流水线处理整幅图像，使可视化整体协调。图 4（左）展示了整体工作流程，并附有可视化在各个步骤中的效果截图。

虽然经过合成步骤后的可视化通常比草图可视化在视觉上更加协调，但我们发现，有些情况下还需要额外进行一次**精修**步骤，以清理图像（例如，减少模糊）。这条最终流水线可以生成分辨率更高、细节更丰富的图像。与合成不同，这些流水线仅对图像进行少量转换，以提升质量，同时保留可视化信息。下文将详细介绍各个步骤。

4.1.1 草图步骤

草图步骤包括：**1)标记预生成**，将普通可视化中的每个标记转换为风格化对象；**2) 背- 景预生成**，生成风格化背景；以及

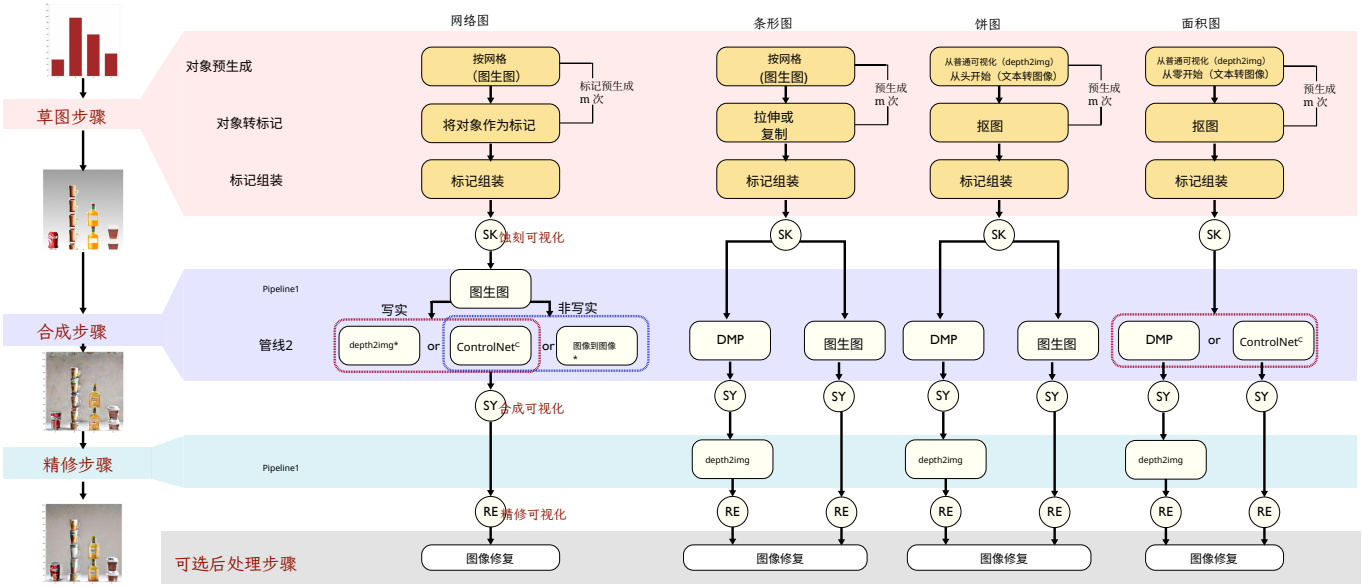


图 4：viz2viz 的工作流程。最左侧路径描述了通用配方，右侧则展示了具体变体。请注意，ControlNet^{C}表示带有 Canny 边缘条件的 ControlNet 流程，而 depth2img*/img2img*表示边缘增强的 depth2img/img2img 流程，它们会在运行 depth2img/img2img 流程之前，将一张网络边图像叠加到初始图像上。

3) 标记组装，将预生成的标记和背- 景组装成草图可视化。

标记预生成在 viz2viz 中，我们采用多种方法来生成每个标记。如第3.3节所述，“单个”标记实际上可能指不同的事物：单个物体（一个瓶子代表整个标记）、多个物体（多个堆叠的杯子），或较大物体中一个连贯的部分（例如饼图）。下面介绍针对这一问题的不同方法。

预生成物体：在许多情况下，当“绘制”替换标记时，viz2viz 会先利用原始普通可视化中创建的标记：将条形和饼图切片分开，并拆分节点和边。根据标记的数量（例如，许多条形或许多节点），viz2viz 会将标记排列在$N \times N$网格中。通过将标记清晰分隔开，我们可以确保为每个原始标记生成独立的物体。例如，要创建由水果构成的社交网络图，需要许多水果，每个水果的颜色和大小都应大致对应原始节点。普通可视化中的每个原始节点标记（例如圆形）都会放置在一个$N \times N$网格中。利用这幅经过修改的图像，我们可以执行图生图扩散（图5a）。通过实验我们发现，在许多情况下，不必将分离出的标记缩放到正确的相对尺寸。例如，对于条形图，我们只需保证条形数量正确，但条形可以大小相同（图5b）。后续步骤可以恢复正确的比例。尽管网格方法在许多情况下都有效，但我们发现，在某些情况下并非必不可少。例如，当标记数量有限（如单个数据集面积图）、物体之间有足够间隔（如彼此不接触的饼图切片），或希望生成图像严格遵循原始标记的轮廓（如面积图）时，就不需要构建网格。在这些情况下，可以将保持原始布局的原始标记作为预生成扩散阶段的输入（见图6）。值得注意的是，由于 viz2viz 以标记作为生成条件，因此图像条件扩散方法（例如 img2img 或 depth2img）最适合这一步。对于提示词包含多个形状子提示词的情况（例如希望每个条形对应不同物体），草图预生成步骤会针对每个子提示词分别运行一次。

在某些情况下，使用原始标记进行预生成可能无法生成理想的草图。例如，当提示词要求采用非常规的摄像机角度，或要求标记由许多元素合成时，就会出现这种情况

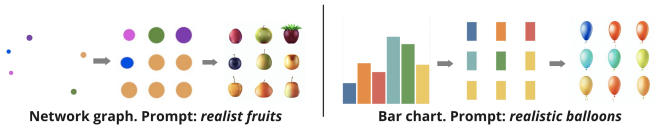


图 5：条形图和网络图按网格进行预生成。



图 6：饼图（提示词：写实的饼干切片）和面积图（提示词：火焰）的直接预生成。

（例如，由人群中许多个体的鸟瞰视图构成的饼图切片）。在这些情况下，viz2viz 不受原始标记的限制，因此可以生成更好的草图。然而，由于生成图像不受原始标记形状的约束，我们需要对其进行转换。

要转换为标记的对象： viz2viz 通过多种方式将生成的对象转换为一个或多个风格化标记：a) 直接将对象转换为标记（例如，将生成的花 → 转换为网络图节点）；b) 调整多个生成的对象（例如，一摞可口可乐罐），使其构成一个标记（一摞可口可乐罐 → 转换为条形图中的单个条形）；或 c) 修改生成的对象，使其符合原始标记的约束（例如，将苹果派 → 转换为一片饼图）。所有这些操作都通过图像变换的各种组合来完成（见图 7）。

在最简单的情况下，除了平移和基本的重新缩放之外，无需进行其他变换。例如，如果我们希望网络图中的每个节点都呈现为苹果的风格，只需生成一组苹果（按原始标记网格生成），然后用这些图像替换节点即可。

当生成对象中只有部分内容可用作标记时，viz2viz 会直接在生成的图像上使用掩码抠出对象的一部分（见图7中的抠图）。我们发现，这种方法对于更复杂的形状（例如面积图中的饼图切片或时间序列）最为有用。举例来说，这种方法使我们能够生成一个完整的



图7：将对象转换为标记的图像操作。

苹果派，或一张人群的大幅图像，然后抠出与饼图标记相对应的部分。

在某些情况下，生成的标记尺寸不正确，且无法轻松重新缩放（即宽度和高度无法一致地重新缩放）。当标记的尺寸要求生成图像具有另一种尺寸时，viz2viz 会进行拉伸（图7）。请注意，在许多情况下，我们并不希望拉伸整个对象。例如，我们可能只想拉伸对象的特定部分（例如铅笔的主体部分，而不是笔尖或橡皮）。在当前实现中，viz2viz 无法理解生成对象的语义。相反，我们采用了一种在实践中效果良好的启发式方法：将生成的对象分为三个相等的部分（“头部”“主体”和“尾部”）。应用此工作流程时，viz2viz 只会将生成对象的主体（中间三分之一）拉伸到所需尺寸，同时保持头部和尾部的尺寸不变。

当设计需要将物体“堆叠”起来时（例如，一摞咖啡杯），viz2viz 可以复制生成的对象，并将它们彼此叠放（参见图7中的复制示例）。饼图风格化也可以采用这种方法，用特定对象的副本填充饼图区域。

背景预生成 如果提供了背景提示词，viz2viz 会使用文本生成图像为草图可视化生成背景。由于背景可能会与可视化标记（可以说更为重要）产生视觉冲突，viz2viz 会对背景进行模糊和提亮。如果未提供背景提示词，viz2viz 可以在此步骤中生成简单的渐变背景。

标记组装 为每个标记创建草图后，viz2viz 会将这些标记重新组装起来，把相应的生成图像放置到原始可视化指定的位置。例如，viz2viz 会将生成对象的网格“拆解”，再将它们重新组装到网络图中的适当位置。

4.1.2 合成步骤

viz2viz 前一步生成的草图通常包含形状、大小和位置大致正确的标记。然而，这些图像缺乏整体协调性：图像被裁切的部分可能显得不完整，杯子可能悬浮在空中，标记的光照和风格差异也可能影响设计效果。合成步骤用于解决其中许多问题。与草图步骤一样，

根据输入可视化类型和提示词，可以采用多种替代工作流程。由于网络中相连元素之间难以保持连贯性（这些元素并不总能在草图步骤中得到体现），因此这类可视化需要分两步处理。对于网络图，合成步骤首先使用低强度的图生图流程。这会平滑草图可视化中可见的不自然对象呈现和构图。

此时，所有可视化都会经过相同（概念上）的合成步骤，以确保各种标记看起来更加协调一致。实际上，这里可以使用不同的流水线来实现不同的效果。讨论具体的图表类型时，我们会再说明为什么有人可能更倾向于使用某种流水线。除了可用于草图可视化的图生图、depth2img 和 ControlNet 流程外，我们还设计了另外三种流水线：**多提示词扩散（DMP）**、**边缘增强深度图生图** 和 **边缘增强图生图**。第一种 DMP 用于支持希望为每个标记指定不同物体的设计师（即使用多个子提示词）。后两种则用于确保网络可视化中的边得以保留，因为在扩散过程中，细边很容易消失。

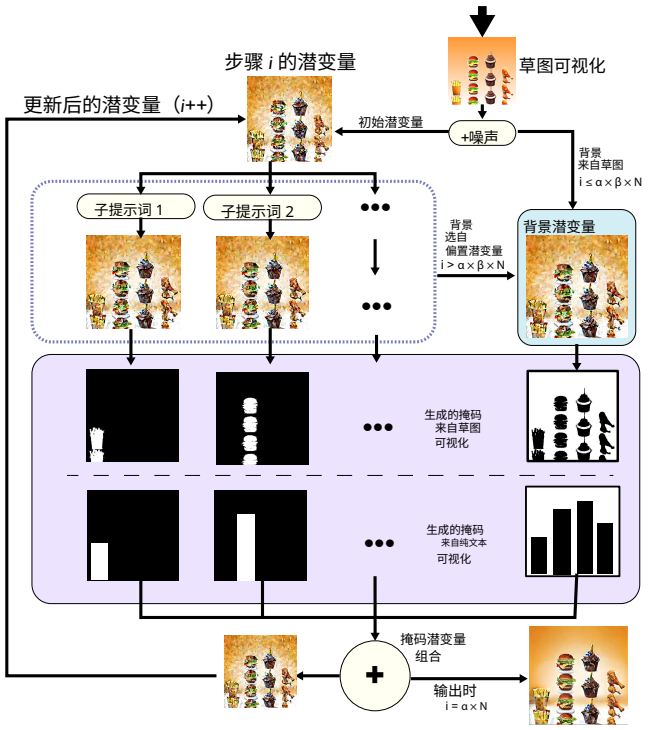
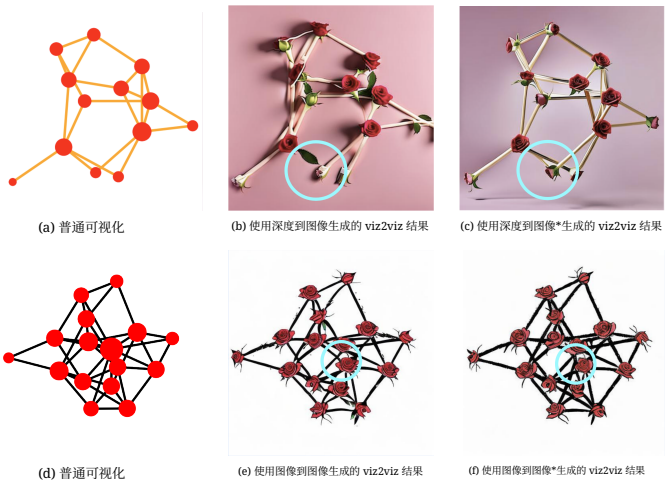


图8：DMP 流程示例。在此示例中，有四个子提示词：薯条、汉堡、纸杯蛋糕、炸鸡 。

多提示词扩散（DMP） DMP 流程用于需要为不同标记分别应用不同子提示词的情况。使用标准流水线时，多个提示词可能会产生不理想的混合效果。例如，如果使用图生图或 depth2img 流程，以提示词 巧克力和苹果派切片 合成饼图，我们可能会发现有些切片把苹果和巧克力混在了一起。图像修复以及对图像区域应用扩散等方法在这里行不通，因为每个区域基本上都会忽略其邻近区域和整幅图像。这与合成步骤的目标大相径庭；我们希望得到一幅协调一致的图像，只不过在这种情况下，它由不同的物体组成。

我们提出的解决方案是一种**多提示词扩散（DMP）**流程。DMP 可以在单次扩散过程中，将可视化中的不同标记关联到不同的子提示词，同时尽量减少不同提示词在单个标记中的混合（图8）。DMP 的基本思路是针对每个提示词，分别对整个潜变量执行扩散，然后再将生成的多个潜变量合并为一个潜变量。由于这一过程会在每一步迭代执行，由不同子提示词引导的各个区域也能兼顾其他区域，从而生成连贯的可视化结果。

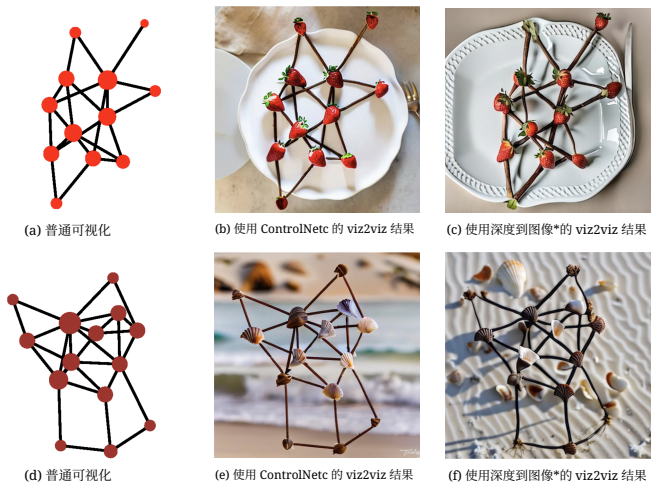
在 DMP 流程开始时，该过程类似于图生图。DMP 会根据强度 α，向草图可视化潜变量添加特定程度的噪声。这就成为初始潜变量。对于每个步骤（$i = 1 \cdots \alpha \times N$，其中 N 是从纯噪声生成图像时的完整扩散步骤数），DMP 使用 depth2img 和不同的子提示词（例如“汉堡”“薯条”等）分别对每个标记进行扩散。结果是一组潜变量，每个标记对应一个。在我们的示例中，这些潜变量中的每一个都会偏向于某个子提示词。这确保了各标记开始“合成”在一起（即，汉堡堆开始影响薯条堆），但也意味着它们会发生融合（即，汉堡堆变得更像薯条）。为避免后一个问题，DMP 会对每个条形图应用掩码。当β介于 0 和 1 之间（即一个超参数）时，在 $i \leq \alpha \times \beta \times N$ 的情况下，使用从草图可视化生成的掩码对图像进行过滤（确保其紧密遵循草图的形状）。当 $i > \alpha \times \beta \times N$ 时，我们切换为使用由普通可视化定义的掩码。



(a) 普通可视化 (b) 使用深度到图像生成的 viz2viz 结果 (c) 使用深度到图像*生成的 viz2viz 结果

(d) 普通可视化 (e) 使用图像到图像生成的 viz2viz 结果 (f) 使用图像到图像*生成的 viz2viz 结果

图 9：原始 depth2img/img2img 与 depth2img*/img2img* 的结果对比。(b) 和 (c) 的提示词：粉色天鹅绒背景上的写实红玫瑰。(e) 和 (f) 的提示词：手绘风格的玫瑰。青色圆圈标出了边缘信息得到增强的位置。



(a) 普通可视化 (b) 使用 ControlNetc 的 viz2viz 结果 (c) 使用深度到图像*的 viz2viz 结果

(d) 普通可视化 (e) 使用 ControlNetc 的 viz2viz 结果 (f) 使用深度到图像*的 viz2viz 结果

图 10：ControlNet^c 与 depth2img* 的结果对比。(b) 和 (c) 的提示词：白色盘子上的写实草莓和树枝。(e) 和 (f) 的提示词：白色沙滩上的写实贝壳和树枝

由于扩散过程已经稳定，这个更大的掩码提供了略多的灵活性。与每个标记类似，我们对背景的潜变量应用相同的思路。当 $i \leq \alpha \times \beta \times N$ 时，背景潜变量是添加了噪声的草图可视化潜变量（掩码则源自草图可视化）。当 $i > \alpha \times \beta \times N$ 时，背景潜变量从某个受提示词影响的标记潜变量中选取。我们在每个步骤中迭代选择背景潜变量（在不同子提示词的输出之间切换），以尽量减少任一子提示词对背景的影响。直观地说，这使背景能够在早期阶段根据草图独立发展，随后再受到新扩散标记的更直接影响。来自标记和背景的掩码潜变量会被合并，并成为步骤 $i+1$ 的潜变量。该策略有两个好处：（1）降低计算复杂度（即无需对背景进行额外扩散）；（2）相比使用单独的提示词对背景进行扩散，能够获得更高的视觉连贯性。请注意，如果只有一个子提示词（例如所有条形图使用相同的提示词），DMP 就会简化为 depth2img。

边缘增强的 depth2img 和 img2img depth2img 的优势在于，它在融合风格化标记的同时，能够稳健地保留结构信息。我们发现，img2img 在非写实样式上的表现很好。遗憾的是，对于网络图，这两种工作流都容易破坏「脆弱」的边。为此，我们对这两种工作流（分别为 depth2img* 和 img2img*）做了小幅改进：将原始网络可视化中的边叠加到初步平滑后的草图可视化上。这样一来，草图中消失的边就会被「强制」加回图像中。这种策略能够在合成可视化中保留边缘信息（图 9）。

4.1.3 精炼步骤

虽然合成步骤生成的可视化效果可能优于草图（因为标记之间会“交互”），但图像可能质量较低，并带有噪点。在这种情况下，我们发现，额外运行一次低强度的 depth2img 流水线，并加入提升质量的提示词（例如，高分辨率逼真的清晰照片，*4K*），可以得到更好的最终图像。请注意，这些额外提示词会与完整提示词（即子提示词和背景提示词）拼接在一起。我们专门使用 depth2img，因为精修步骤通常用于让可视化效果更加写实。在其他情况下，精修步骤是可选的，也可以替换为任何分辨率增强流程（例如，超分辨率[2]）。

4.2 viz2viz 变化

根据可视化类型和提示词，可以使用 viz2viz 的不同变体（见图 4）。我们提供一些直观说明，解释为何可能会选择某种特定的工作流程配方（例如，标记的结构和形状、图表形式，以及期望输出的写实程度）。我们强调每种工作流程的独特属性，以及这些属性对 viz2viz流程所需的调整。我们还强调，尽管每种工作流程各不相同，但其中一些子元素会反复出现，可能有助于应用于新的图表形式。

4.2.1 网络图

网络图的 **草图步骤** 会对节点的 $N \times N$ 网格使用图生图扩散流程。使用多个子提示词时，viz2viz 会多次执行此步骤。对于**合成步骤**，我们发现网络图需要先经过一个初始图生图流程，以平滑草图图像。

此步骤之后，viz2viz 会根据提示词要求生成写实或非写实输出，提供不同的选项。对于写实类提示词，扩散流水线可以采用 ControlNet^c（带有 Canny 边缘条件的 ControlNet）或深度到图像*。使用 ControlNet^c 可以生成形状更精确的标记，而深度到图像*则能生成风格化程度更高、视觉上更连贯的结果（对比见图。10）。对于非写实样式，第二个扩散流水线可以采用 ControlNet^c 或图生图*。与写实图像一样，使用 ControlNet^c 可以获得更好的视觉连贯性和更平滑的边缘结构。不过，图生图可以生成更符合提示词、且更具「草图感」的结果。我们发现，网络图不需要额外的**精修步骤**。

4.2.2 条形图

条形图通常使用彼此分离的单个标记，适合采用以下配方。与网络图一样，在**草图步骤**中，会对由等尺寸条形组成的网格应用图生图流水线，作为起始图像。不过，与网络中的节点（通常是尺寸相同的标记）不同，条形图需要先将生成的标记拉伸或复制，再重新组合。

我们在**合成步骤** 中省略了平滑处理，因为条形图的草图可视化通常很干净。对于写实提示词，我们发现 DMP 流程最适用于条形图。对于非写实图像，图生图能产生更好的效果。由于 DMP 往往会产生

[2]https://huggingface.co/docs/diffusers/api/pipelines/stable_diffusion/upscale



(a) 精细化可视化　(b) 重新生成背景　(c) 重新生成背景和一个风格化切片

图 11：通过重新生成背景和风格化标记进行后处理。使用的提示词为写实红玫瑰/郁金香/紫色雏菊/橙色向日葵田地的鸟瞰图。

一些低质量伪影，我们在**精修步骤**中使用了低强度的 depth2img 来修复这些伪影。

4.2.3 饼图

饼图需要一个独立的草图步骤，但除此之外，大致可以沿用与条形图相同的工作流程。这里假设饼图扇区可以清晰分离。与条形图和网络节点不同，饼图切片的形状更复杂。根据提示词，viz2viz 可以在**草图步骤**中使用 depth2img（基于原始可视化）或 text2img（从头开始生成）。在这两种情况下，viz2viz 都会先从生成的内容中裁剪出饼图切片，再重新组合。其余步骤**合成** 和**精修**与条形图采用相同的格式。

4.2.4 面积图

面积图与饼图类似，标记的复杂度较高。我们采用了与饼图相同的**草图**和**精修**步骤。对于**合成**步骤，我们发现，使用 DMP 或 ControlNet^{C} 流水线都可以生成写实或非写实图像。由于面积标记通常较大，因此会形成简单的深度结构，既适用于 ControlNet^{C}，也能生成形状精确的标记。DMP 则可以生成风格化程度更高、视觉上更协调的结果。支持这两种工作流，既能让用户权衡取舍，也能生成更好的风格化可视化。

4.2.5 可选后处理

在某些情况下，viz2viz 生成的可视化总体上具有我们想要的风格，但可能未能成功处理某个子提示词（即，除了风格化的汉堡条之外，其他部分我们都很满意）。为了解决这些问题，用户可以选择使用图像修复流水线，重新生成特定标记或背景。应用此功能时，viz2viz 会使用提示词以及与所选标记/背景对应的掩码进行图像修复。此步骤的示例见图 11。用户还可以选择使用自定义区域和初始生成时未使用的提示，重新生成可视化。最后，如果可视化的某些部分风格化效果良好，用户可以将这些部分复制到效果较差的区域，并以较低强度执行图像修复流水线。显然，要生成一幅“完成”的可视化，可能还需要或希望进行额外编辑（例如，更改字体、调整坐标轴以匹配新风格等）。

4.3 实现细节

我们使用 Python 和 PyTorch 实现了 viz2viz。我们使用 Google Colaboratory 笔记本来指定 viz2viz 配方。普通可视化使用 Matplotlib、NetworkX 和 Altair 生成。图像处理（例如掩码生成）通过 OpenCV 和 Pillow 实现。我们使用 Hugging Face 的 Diffusers，以及采用 SD 模型的扩散模型。具体而言，ControlNet 流程使用 SD v1.5[3]，，其他所有工作流均使用 SD v2[4]。所有图像的尺寸均为 512×512 。

通过内部实验，我们发现了一些适合作为默认值的参数。虽然 viz2viz 使用默认参数也能很好地运行，

[3]https://huggingface.co/lllyasviel/ControlNet[4] https://huggingface.co/stabilityai/stable-diffusion-2-1

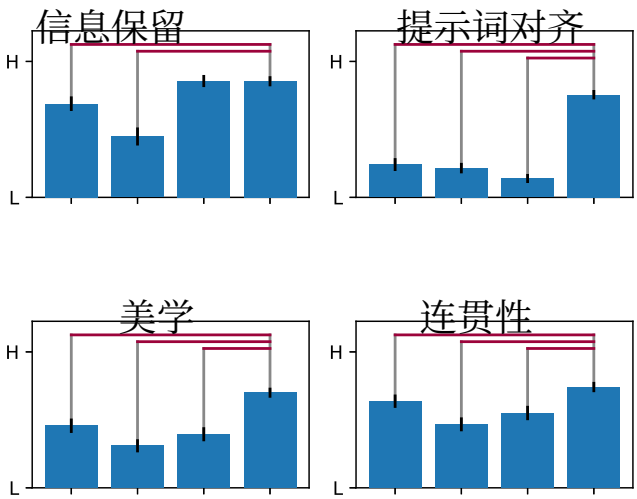


图 12：评估结果。误差线表示 95% 置信区间，红色连接线显示哪些条件与 viz2viz 存在显著差异。。

但这些值可以根据具体需求进行微调。我们将所有工作流的引导尺度设为 20。在草图步骤中，depth2img 的强度设为 0.98，图生图的强度设为 0.82。在合成步骤中，图生图的强度设为 0.4，而其他所有工作流（DMP、depth2img、ControlNet）均设为 0.8。在精修步骤中，0.25 到 0.35 之间的值最适合作为默认强度。对于 DMP 流程，默认情况下 β 设为 0.5。

5 评估

为评估 viz2viz，我们在四个维度上考察了三种基线（depth2img、图生图和 ControlNet^{c}）：1）数据保留得如何（信息保留）？ 2）该方法遵循提示词的程度如何（提示词对齐）？ 3）生成的可视化在视觉上有多美观（美学）？ 4）输出在视觉上有多连贯（连贯性）？总计而言，对于四种图表类型，我们各使用了三种不同的提示（包括单个和多个子提示词），共计 12 个“组”。我们允许每个系统使用默认参数[5]生成四种可能的输出。一位作者创建了样本并将结果随机排序（但未参与评估）。对于所有条件，我们均使用了负面提示词：“低质量、普通质量、最差质量、绘制粗劣、错误、抽象、模糊”。对于 viz2viz，我们选择了最符合提示词中指定的可视化类型和风格的工作流程（即写实或非写实）。如果有两种可能的工作流程，则选择质量更高的工作流程（这也是 viz2viz 用户很可能会采取的做法）。请注意，我们没有对 viz2viz 执行可选后处理。我们的提示词来自不同来源。一些是对现有风格化可视化的描述，另一些则由数据可视化协会或我们研究小组的设计师提出。三位评估者（两男一女）在不知道生成条件的情况下，使用五级李克特量表，依据四项指标对 192 张生成图像逐一进行评估。对于每项指标的分析，我们对所有条件进行 Kruskal-Wallis检验，随后进行 Dunn检验作为事后分析。

我们的结果汇总于图12。对于“信息保留”，我们观察到显著差异（$H = 115.7$, $p < 0.001$），viz2viz 的表现显著优于 depth2img（$p < 0.001$），并且

[5]对于基线，默认强度设为 0.8（图生图）和 0.95（depth2img、ControlNet^{c}），因为这些数值的整体表现最佳。

图生图（$p < 0.001$）。自然地，ControlNet 能很好地沿着原始可视化的轮廓生成结果（这也是我们将其用作 viz2viz 子组件的原因）。在“提示词对齐”方面（$H = 270.8$，$p < 0.001$），viz2viz 的质量显著高于所有其他条件（$p < 0.001$ 均如此）。美学指标也呈现类似趋势（$H = 117.6$，$p < 0.001$），viz2viz 的得分显著高于所有其他条件（$p < 0.001$ 均如此）。最后，在连贯性方面，各条件之间存在显著差异（$H = 60.9$，$p < 0.001$）；viz2viz 再次优于所有其他条件（与 depth2img 比较时 $p < 0.05$，与其他条件比较时 $p < 0.001$）。信息保留、提示词对齐、美学和连贯性这四项指标的 Spearman's ρ 值平均分别为 0.20、0.31、0.23 和 0.13。这表明标注者之间的一致性较低至一般。

作为一项案例研究，我们使用 DALL- E 2 生成的图像（见图 3），并尝试用 viz2viz 生成近似图像。在大多数情况下，以真实数据作为输入，我们都成功生成了近似图像。下文介绍一些效果不太理想的示例。两项研究的提示词和材料均收录于补充材料中。

6 讨论与未来工作

6.1 viz2viz 的可泛化性

我们设计 viz2viz，旨在使其能够泛化到其他可视化惯用形式。通用配方的三个步骤（草图-合成-细化）旨在兼顾可视化信息的保留与可视化风格的一致性。根据我们的经验，我们认为，要创建一个适用于所有输入和提示词的流水线，本身就存在固有挑战。尽管如此，我们认为，除了高层工作流程的可泛化性之外，具体实现中的子组件也可应用于新的可视化类型。例如，可以基于网络图的工作流程构建散点图，也可以基于饼图构建环形图。

6.2 提示词工程

viz2viz 中的可视化对提示的质量非常敏感。虽然我们将提示词工程的系统化评估留待未来工作，但我们总结了观察到的一些模式。

在 viz2viz 中，与抽象或虚构的物体相比，提示词用于更具体、常见的对象时效果更好。例如，木制过山车的侧视图 比宇宙红色过山车的侧视图，超现实主义、迷幻主题更有可能得到更好的结果。对于底层扩散模型不擅长生成的内容，通常也难以用于风格化可视化。例如，我们发现模型在生成医用注射器时会遇到困难。在标准流水线中不使用任何数据点测试提示词（例如，文本转图像、DALL-E 等），可能有助于调整提示词，提高其奏效的可能性。例如，我们注意到薯条 通常比薯片 效果更好（因为薯片这一含义相对少见）。在某些情况下，单个提示词可能会导致渲染结果含义不明确，此时添加一些文本会有所帮助。例如，在橙子 前添加水果 或颜色，可以提高获得预期风格化效果的可能性。最后，我们发现不同生成算法的能力存在差异。例如，Stable Diffusion 在生成带品牌标识的可口可乐瓶时效果不佳。

与生成式提示一样，总体而言，我们发现，加入对摄像机视角、物体数量、样式、颜色、物体的具体类型、形状、物体名称、布局和背景的详细描述会有所帮助。明确描述所需的样式或媒介（例如照片级真实或油画）似乎也有帮助。处理一对多风格可视化时，加入平铺摆拍可能会很有用。我们尚未深入实验的一个方面是负向提示词。在 viz2viz 的最后“清理”步骤中，我们会使用这类提示词，但它们也可能有助于在早期操作中引导系统生成更好的结果。

6.3 获取高质量图像

我们使用 viz2viz（以及其他生成式系统）的经验表明，通常需要多次尝试才能获得理想结果。在本文中，我们尝试梳理关键决策点，以帮助人们选择某一种工作流程或流水线，而非其他方案。我们还尝试说明需要人工干预或重复操作的各种情形。我们的实验发现，在给定优质提示词的情况下，找到一张理想图像通常最多需要生成约 10 张图像（不过很多时候所需数量远少于此）。

并非所有提示词都能轻松生成。当普通可视化标记的颜色或形状与目标对象差异过大时，生成可能会失败。因此，初始输入与期望输出之间需要保持一定的一致性（例如，对于提示词一片披萨，红色饼图切片会比黄色切片更合适）。

除了提示词之外，大多数工作流还具有其他可调参数。我们根据实验结果提供了较好的默认值。不过，某些风格化需求很可能会受益于进一步调参。例如，用户可能希望为 DMP 中的 β 参数设置一个合适的值（默认为 0.5）。根据中间结果（例如草图可视化的质量），用户可以增大（如果图像效果有潜力）或减小该值。最后，我们强调，viz2viz 可能会完全或部分失败。使用生成后图像修复及类似功能，有助于从部分失败中恢复。

6.4 局限性与未来工作

目前，viz2viz 是在笔记本环境中实现的。虽然它可以轻松用作图书馆，但未来需要解决的一个重要领域是实现图形用户界面。能够轻松查看更改的影响、查看中间结果、调整参数空间，并且通常能够使用直接操控操作（例如用于图像修复），很可能会带来更好的输出。

虽然 viz2viz 能够生成许多不同类型的标记，但有时，更好的语义理解（包括对语言和对象的理解）能够带来更好的结果。例如，很难指定我们希望用注射器中的液体量（而非注射器的长度）来表示图表中的条形。一种可行的方法是将手绘或搜集到的示例作为额外输入。由于我们无法严格控制前景对象与背景对象之间的交互，viz2viz 无法生成某些信息图像 [6]（例如图 3中的大礼帽）。未来工作若能增加预生成背景的步骤，并指定可视化的渲染区域，或许就能让 viz2viz 更广泛地支持这些风格化可视化。

与其他生成式系统一样，有些工作流可能会产生嘈杂的结果。对于 viz2viz，问题可能更加突出，因为各种生成约束可能会对可视化的外观造成不利的副作用。未来可通过改进参数调优、创建流水线、优化掩码（用于 DMP 和图像修复）以及改进界面来解决这一问题。

最后，viz2viz 目前能够支持的标记数量有限。随着标记数量增加，它们在扩散过程中可能会开始不恰当地融合。分块生成较小的可视化子集本身并无不妥。未来工作可探究这种方法的局限，并从技术上确保图像能够以美观的方式重建。

7 结论

本文介绍了 viz2viz，这是一种生成工作流程，用户只需提供文本提示和普通可视化，即可获得高质量的风格化可视化。我们提出了一个三步流程，确保风格化可视化能够正确编码数据，同时满足图像连贯性等复杂要求。为此，我们采用了先进的可控扩散模型以及自行设计的流水线（DMP）。我们还介绍了如何将 viz2viz 推广至其他风格化任务。我们认为，在可视化领域采用生成式方法既面临内在挑战，也蕴含机遇。我们很高兴为从业者和研究人员提供一个探索该领域的切入点。

致谢

我们谨向 Jason Forrest、Elsie Lee-Robbins 和 Matt Brehmer，以及为我们提供提示和反馈的早期试用用户表示感谢。本研究部分由 NSF IIS-1815760 资助。

参考文献

[1] O. Avrahami、D. Lischinski 和 O. Fried。用于自然图像文本驱动编辑的混合扩散。在 IEEE/CVF 计算机视觉与模式识别会议论文集，第 18208 18218, 2022 3 – 页。A.

[2] Bigelow、S. Drucker、D. Fisher 和 M. Meyer。关于设计师如何使用数据进行设计的反思。在 2014 年高级可视化界面国际工作会议论文集，AVI '14，第 17–24 页。美国纽约州纽约市：Association for Computing Machinery，2014 10 1145 2598153 2598175 2 。doi：。/。

[3] A. Bigelow、S. Drucker、D. Fisher 和 M. Meyer。在工具之间迭代以创建和编辑可视化。IEEE 可视化与计算机图形学汇刊，23 1 481 490, 2017 10 1109 2016 2598609 2 ( ): –。doi：。/TVCG。。

[4] M. Brehmer、R. Kosara 和 C. Hull。利用 Diatoms 为图形符号生成设计灵感。IEEE 可视化与计算机图形学汇刊，28(1)：389–399, 2022 10 1109 2021 3114792 2 。doi：。/TVCG。。Z. Chen、Y.

[5] Wang、Q. Wang、Y. Wang 和 H. Qu。迈向自动化信息图设计：基于深度学习的可扩展时间线自动提取。IEEE 可视化与计算机图形学汇刊，26(1)：917–926，2020。doi： 10 1109 2019 2934810 2 。/TVCG。。D. Coelho 和 K. Mueller。信

[6] 息图像：将数据嵌入主题图像。在《计算机图形学论坛》，第 39 卷，第 593–606 页。Wiley Online Library， 2020 2, 9 。W. Cui、J. Wang、H. Huang、Y.

[7] Wang、C.-Y. Lin、H. Zhang 和 D. Zhang。复用信息图图表的混合主动式方法。IEEE 可视化与计算机图形学汇刊，28(1)：173–183，2022。doi：10.1109/TVCG.2021. 3114856 2[8] W. Cui、X. Zhang、Y. Wang、H. Huang、B. Chen、L. Fang、H. Zhang、J.-G. Lou 和 D. Zhang。Text-to-viz：根据与比例相关的自然语言陈述自动生成信息图。IEEE 可视化与计算机图形学汇刊，26(1)：906–916，2020。doi：10.1109/TVCG.2019. 2934785 2[9] J. Forrest。我让人工智能绘制了一张图表，2022 年。访问日期： 22, 2023 2 。J. Harper 和 M.

[10] Agrawala。将基本的 d3 图表转换为可复用的样式模板。IEEE 可视化与计算机图形学汇刊，24 1286，2018 3 1274 10 1109 2017 ( ): –。doi：。/TVCG。。 2699744 2

[11] J. Ho 和 T. Salimans。无分类器扩散引导。arXiv 预印本，arXiv： *2207 12598*, 2022 3 。。L. Huang、D. Chen、Y.

[12] Liu、Y. Shen、D. Zhao 和 J. Zhou。Composer：通过可组合条件实现富有创意且可控的图像合成，2023 年。3[13] K. Kagkelidis、I. Dimitriadis 和 A. Vakali。Lumina：用于探索、丰富和可视化数据的自适应、自动化且可扩展的原型。J. Vis., 24 3 631 655, 2021 10 1007 12650 020 00718 2 ( ): – jun。

[14] doi：。/s - - -y N. W. Kim、E. Schweickart、Z. Liu、M. Dontcheva、W. Li、J. Popovic 和 H. Pfister。数据驱动指南：支持信息图形的表现性设计。IEEE 可视化与计算机图形学汇刊，23(1)：491– 500, 2017 10 1109 2016 2598620 2 。

[15] doi：。/TVCG。。Y. Li、H. Liu、Q. Wu、F. Mu、J. Yang、J. Gao、C. Li 和 Y. J. Lee。Gligen：开放集有基础的文本到图像生成，2023 年。doi：10.48550/ARXIV. 2301 07093 3 。Y. Liu、E. Jun、Q. Li 和 J. Heer。潜空间制图：向量空间嵌入的

[16] 可视化分析。在《计算机图形学论坛》，第 38 卷，第 67 78 2019 2 – 页。Wiley Online Library，。Z. Liu、J.

[17] Thompson、A. Wilson、M. Dontcheva、J. Delorey、S. Grigg、B. Kerr 和 J. Stasko。Data Illustrator：通过惰性数据绑定增强矢量设计工具，以支持表现性可视化创作。在 2018 年 CHI 人机交互系统中的人为因素会议论文集，CHI '18，第 1–13 页。美国纽约州纽约市：Association for Computing Machinery，2018 10 1145 3173574 3173697 2 。doi：。/。H. Mei、Y. Ma、Y. Wei 和

[18] W. Chen。信息可视化构建工具的设计空间：综述。《视觉语言与计算学报》，44 120 132, 2018 10 1016 2017 10 001 2 : –。doi：。/j.j

[19] vlc。。C. Meng、Y. He、Y. Song、J. Song、J. Wu、J.-Y. Zhu 和 S. Ermon。SDEdit：使用随机微分方程进行引导式图像合成与编辑。在国际学习表征会议，2022 1, 3 。C. Qian、S. Sun、

[20] W. Cui、J.-G. Lou、H. Zhang 和 D. Zhang。检索后适配：面向与比例相关信息图的基于示例的自动生成。IEEE 可视化与计算机图形学汇刊，27 452，2021 2 443 10 1109 2020 3030448 2 ( ): –。doi：。/TVCG。。

[21] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark 等。通过自然语言监督学习可迁移的视觉模型。载于《国际机器学习会议》，第 8748 8763 2021 3 – 页。

[22] PMLR，。A. Ramesh, P. Dhariwal, A. Nichol, C. Chu 和 M. Chen。利用 CLIP 潜变量进行层次化文本条件图像生成，2022。doi: 10.48550/ARXIV. 2204 06125 1, 3 。R. Ranftl, K. Lasinger, D. Hafner, K. Schindler 和 V.

[23] Koltun。迈向稳健的单目深度估计：混合数据集实现零样本跨数据集迁移。《IEEE 模式分析与机器智能汇刊》， 44 3 1623 1637, 2020 3 ( ): –。D. Ren, B. Lee 和

[24] M. Brehmer。Charticulator：交互式构建定制图表布局。《IEEE 可视化与计算机图形学汇刊》， 25 1 789 799, 2019 10 1109 2018 2865158 2 ( ): –。doi: . /

[25] TVCG. . R. Rombach, A. Blattmann, D. Lorenz, P. Esser 和 B. Ommer。利用潜变量扩散模型进行高分辨率图像合成，2021 1 。R. Rombach, A.

[26] Blattmann, D. Lorenz, P. Esser 和 B. Ommer。利用潜变量扩散模型进行高分辨率图像合成。载于《IEEE/CVF计算机视觉与模式识别会议论文集》，第10684–10695页，2022。2, 3[27] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. L. Denton, K. Ghasemipour, R. G. Lopes, B. K. Ayan, T. Salimans, J. Ho, D. J. Fleet 和 M. Norouzi。具备深度语言理解能力的照片级真实文本到图像扩散模型。《神经信息处理系统进展》， 35 36479 36494, 2022 3 : –。A. Satyanarayan,

[28] B. Lee, D. Ren, J. Heer, J. Stasko, J. Thompson, M. Brehmer 和 Z. Liu。对可视化创作系统的批判性反思。《IEEE 可视化与计算机图形学汇刊》， 26 471 2934281 1 461 10 1109 2019 ( ): –。doi: . /TVCG. .[29]

M. Savva, N. Kong, A. Chhajta, L. Fei-Fei, M. Agrawala 和 J. Heer。Re-vision：图表图像的自动分类、分析与重新设计。载于《第24届ACM用户界面软件与技术研讨会论文集》，UIST '11，第393–402页。计算机协会，纽约，纽约州，美国，2011 10 1145 2047196 2047247 2 。doi: . / . L. Shen, E.

[30] Shen, Y. Luo, X. Yang, X. Hu, X. Zhang, Z. Tai 和 J. Wang。迈向数据可视化的自然语言界面：一项综述。《IEEE 可视化与计算机图形学汇刊》，第1–1页，2022。doi: 10.1109/TVCG.2022.3148007 2[31] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan 和 S. Ganguli。利用非平衡热力学进行深度无监督学习。载于《国际机器学习会议》，第 2256 2265 2015 2 –页。PMLR，。

[32] M. Sturdee, S. Knudsen 和 S. Carpendale。数据绘画：富有表现力的自由形式可视化。《设计研究学会论文集》，2022, 2022 3 。Y.

[33] Wang, H. Zhang, H. Huang, X. Chen, Q. Yin, Z. Hou, D. Zhang, Q. Luo 和 H. Qu。InfoNice：轻松创建信息图形。载于《2018年CHI人机交互系统会议论文集》，CHI '18，第1–12页。计算机协会，纽约，纽约州，美国，2018 10 1145 3173574 3173909 2 。doi: . / . A. Wu, Y. Wang, X. Shu, D.

[34] Moritz, W. Cui, H. Zhang, D. Zhang 和 H. Qu。AI4Vis：数据可视化人工智能方法综述。《IEEE 可视化与计算机图形学汇刊》，2021 2 。H. Xia, N.

[35] Henry Riche, F. Chevalier, B. De Araujo 和 D. Wigdor。DataInk：直接且富有创意的数据导向绘画。载于《2018年CHI人机交互系统会议论文集》，CHI '18，第1–13页。计算机协会，纽约，纽约州，美国，2018。doi: 10 1145 3173574 3173797 2 . / . L. Ying, X. Shu, D. Deng, Y. Yang, T. Tang,

[36] L. Yu 和 Y. Wu。MetaGlyph：基于隐喻字形的可视化自动生成。《IEEE 可视化与计算机图形学汇刊》， 29 1 331 341, 2022 2 ( ): –。J. E. Zhang, N. Sultanum, A. Bezerianos 和 F. Chevalier。DataQuilt：从图像中提取视觉元素以创作图示化可视化。载于《2020年CHI人机交互系统会议论文集》，第 1 13, 2020 2 –页。L. Zhang 和 M. Agrawala。为文本到图像扩散

[37] [38] 模型添加条件控制。arXiv预印本 arXiv: *2302 05543*, 2023 1, 3 。. S. Zhu, G. Sun,

[39] Q. Jiang, M. Zha 和 R. Liang。自动信息图与可视化推荐综述。《可视化信息学》，4(3):24–40，2020。doi: 10.1016/j.visinf.2020.07.002 2

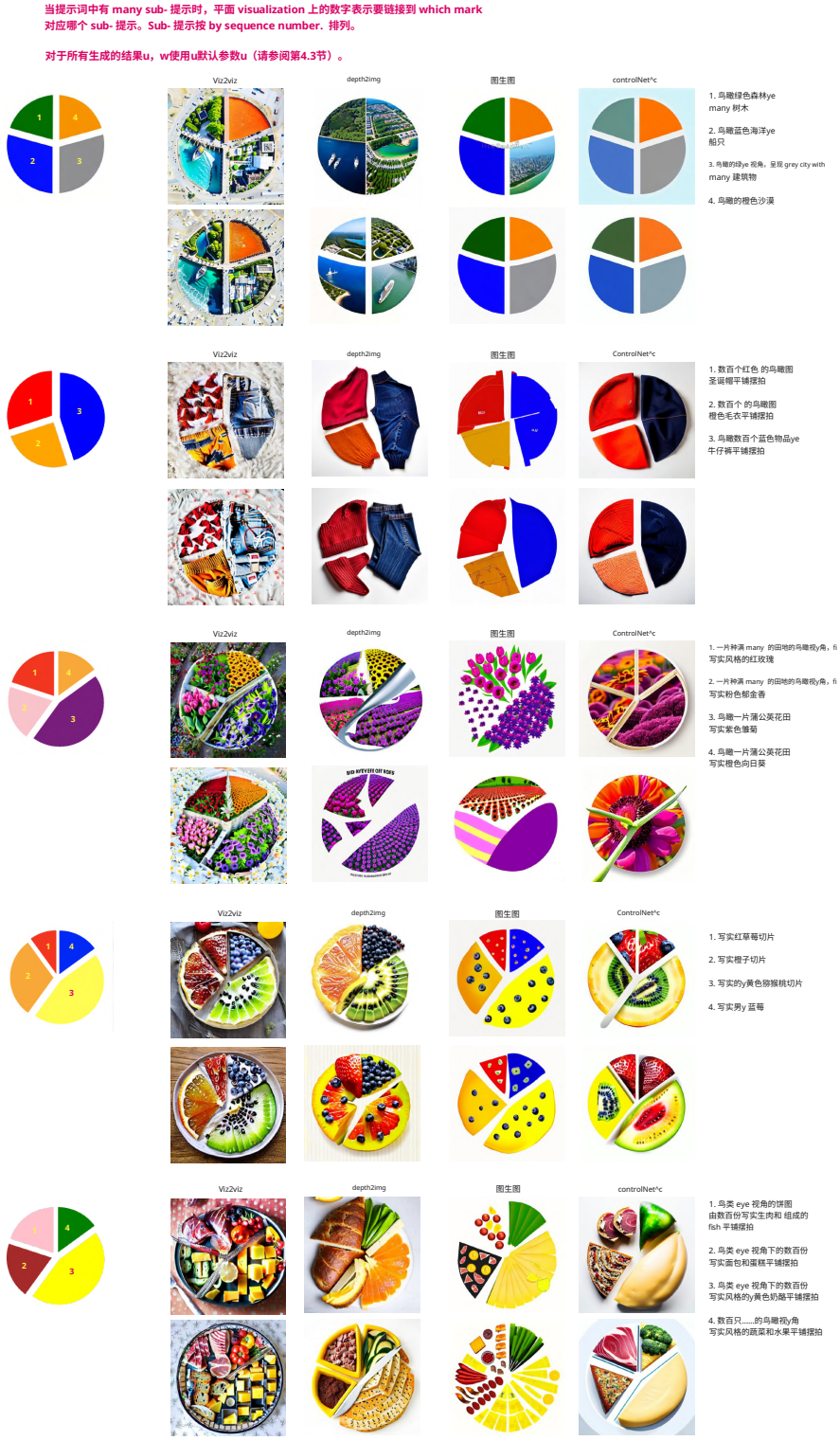

当提示词中有 many sub- 提示时，平面 visualization 上的数字表示要链接到 which mark
对应哪个 sub- 提示。Sub- 提示按 by sequence number. 排列。
对于所有生成的结果u，w使用u默认参数u（请参阅第4.3节）。
Viz2viz
depth2img
图生图
controlNet^c
1. 鸟瞰绿色森林ye many 树木
2. 鸟瞰蓝色海洋ye 船只
3. 鸟瞰的绿ye 视角，呈现 grey city with many 建筑物
4. 鸟瞰的橙色沙漠
Viz2viz
depth2img
图生图
ControlNet^c
1. 数百个红色 的鸟瞰图 圣诞帽平铺摆拍
2. 数百个 的鸟瞰图 橙色毛衣平铺摆拍
3. 鸟瞰数百个蓝色物品ye 牛仔裤平铺摆拍
Viz2viz
depth2img
图生图
ControlNet^c
1. 一片种满 many 的田地的鸟瞰视y角，fi 写实风格的红玫瑰
2. 一片种满 many 的田地的鸟瞰视y角，fi 写实粉色郁金香
3. 鸟瞰一片蒲公英花田 写实紫色雏菊
4. 鸟瞰一片蒲公英花田 写实橙色向日葵
Viz2viz
depth2img
图生图
ControlNet^c
1. 写实红草莓切片
2. 写实橙子切片
3. 写实的y黄色猕猴桃切片
4. 写实男y 蓝莓
Viz2viz
depth2img
图生图
controlNet^c
1. 鸟类 eye 视角的饼图 由数百份写实生肉和 组成的 fish 平铺摆拍
2. 鸟类 eye 视角下的数百份 写实面包和蛋糕平铺摆拍
3. 鸟类 eye 视角下的数百份 写实风格的y黄色奶酪平铺摆拍
4. 数百只.....的鸟瞰视y角 写实风格的蔬菜和水果平铺摆拍

图 13：补充结果 1。

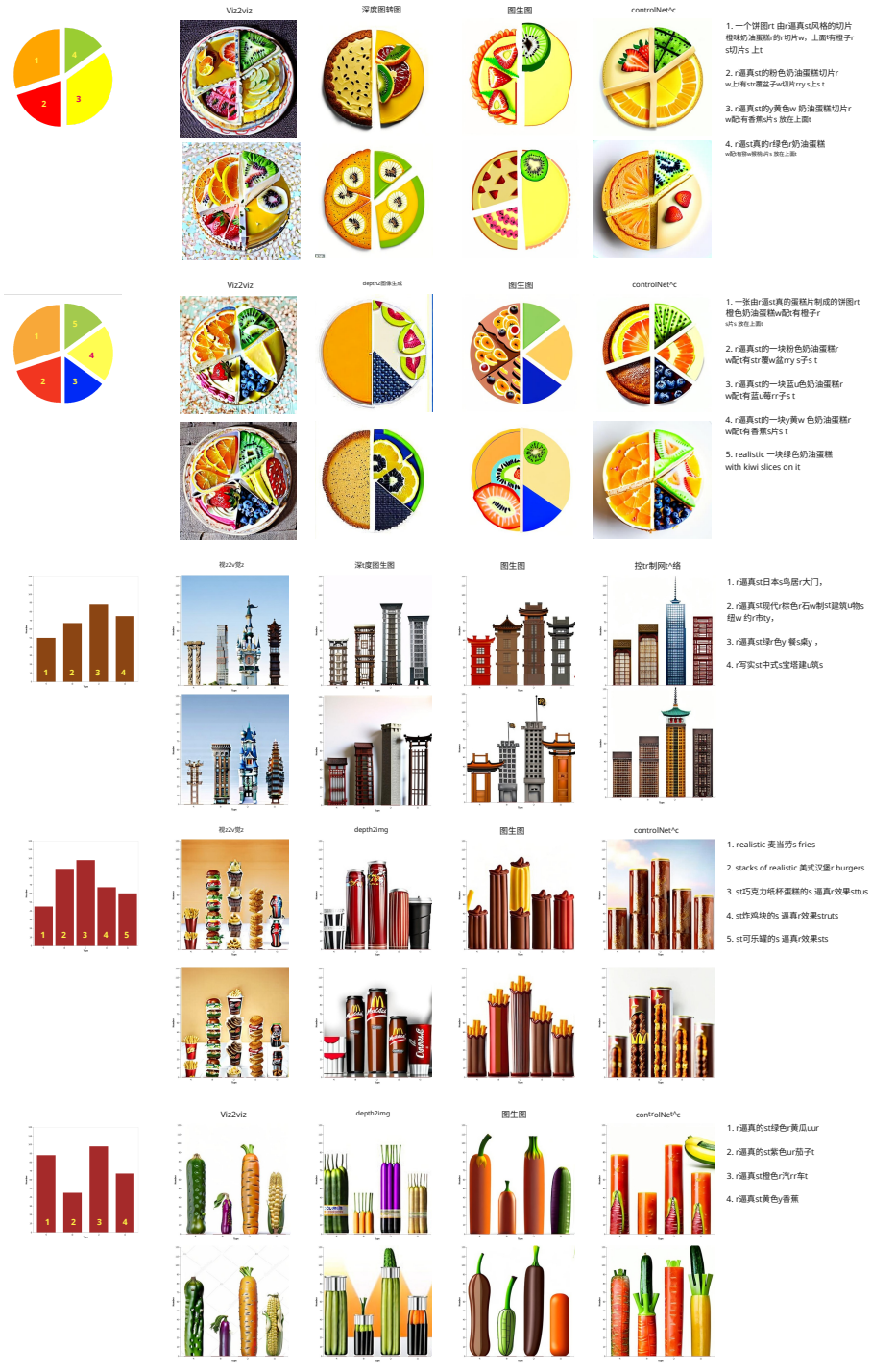

Viz2viz
深度图转图
图生图
controlNet^c
1. 一个饼图rt 由r逼真st风格的切片橙味奶油蛋糕r的r切片w，上面r有橙子r s切片s 上t
2. r逼真st的粉色奶油蛋糕切片r w上有str覆盖子w切片rry s上s t
3. r逼真st的y黄色w 奶油蛋糕切片r w配有香蕉sR s 放在上面t
4. r逼st真的r绿色r奶油蛋糕
Viz2viz
depth2图像生成
图生图
controlNet^c
1. 一张由r逼st真的蛋糕片制成的饼图rt 橙色奶油蛋糕w配r有橙子r sR s 放在上面t
2. r逼真st的一块粉色奶油蛋糕r w配r有str覆w盆rry s子s t
3. r逼真st的一块蓝u色奶油蛋糕r w配r有蓝u糖rr子s t
4. r逼真st的一块y黄w 色奶油蛋糕r w配r有香蕉sR s t
5. realistic 一块绿色奶油蛋糕 with kiwi slices on it
视2v视2
深t度图生图
图生图
控tr制网n^络
1. r逼真st日本s鸟居r大门，
2. r逼真st现代r棕色r石w制st建筑u物s 纽w 约r市ty，
3. r逼真st绿r色y 餐s桌y，
4. r写实st中式s宝塔建u筑s
视2v视2
depth2img
图生图
controlNet^c
1. realistic 麦当劳s fries
2. stacks of realistic 美式汉堡r burgers
3. st5克力纸杯蛋糕的s 逼真r效果sttus
4. st炸鸡块的s 逼真r效果struts
5. st可乐罐的s 逼真r效果sts
Viz2viz
depth2img
图生图
controlNet^c
1. r逼真的st绿色r黄瓜uur
2. r逼真的st紫色ur茄子t
3. r逼真st橙色r汽rr车t
4. r逼真st黄色y香蕉

图14：补充结果2。

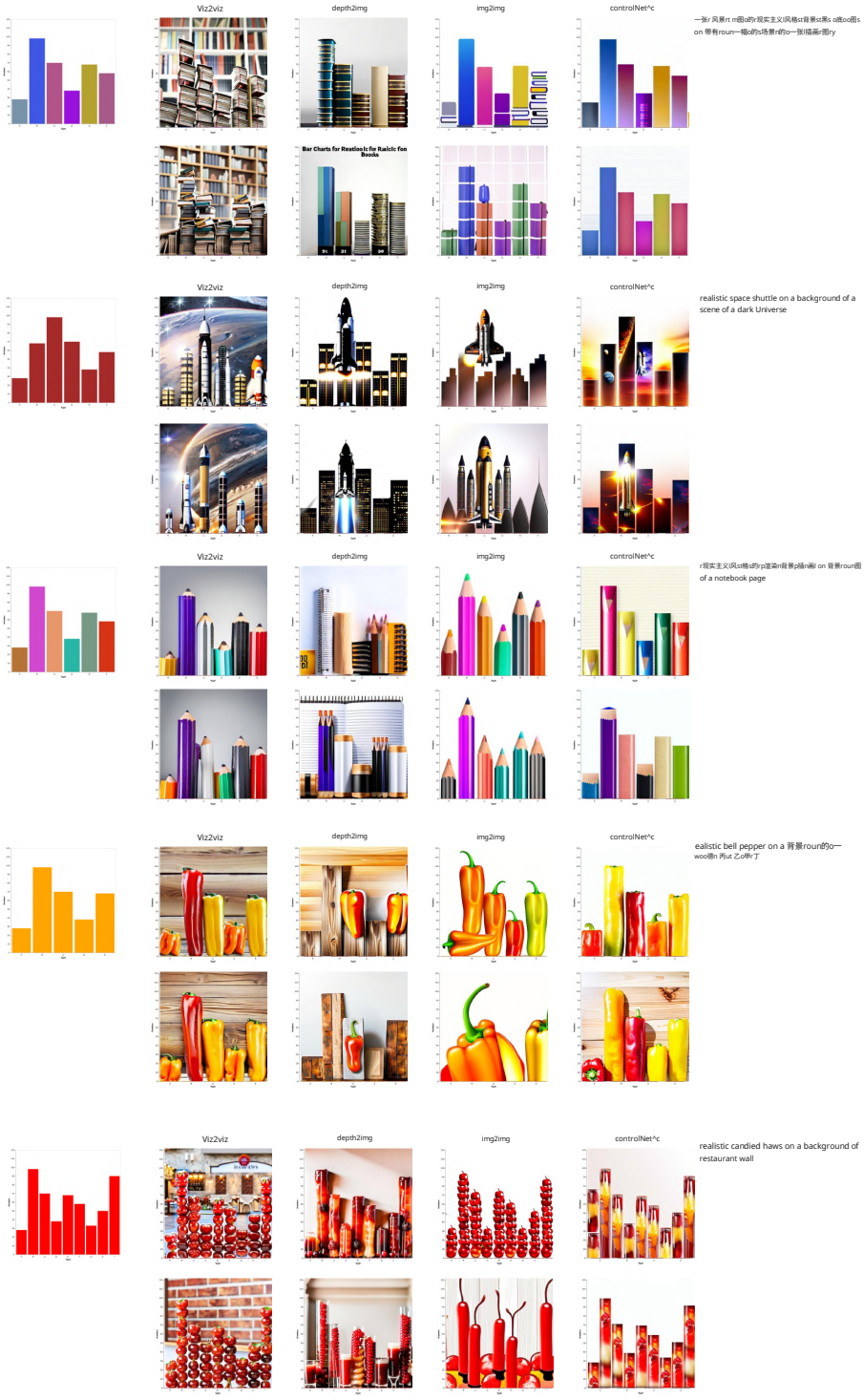

Viz2viz
depth2img
img2img
controlNet^c
一张r 风景rt rt图d的r现实主义风格ct背景ss果s oll图oo图s on 带有roun一幅d的s场景r的o一张描画r图ry
Bar Charts for Realistic for Kalcic fon Books
realistic space shuttle on a background of a scene of a dark Universe
r现实主义风ct格d的rp建筑m背景p插m画 on 背景roun图 of a notebook page
ealistic bell pepper on a 背景roun的o一
woo图n 内ut 乙o种了
realistic candied haws on a background of restaurant wall

图15：补充结果3。

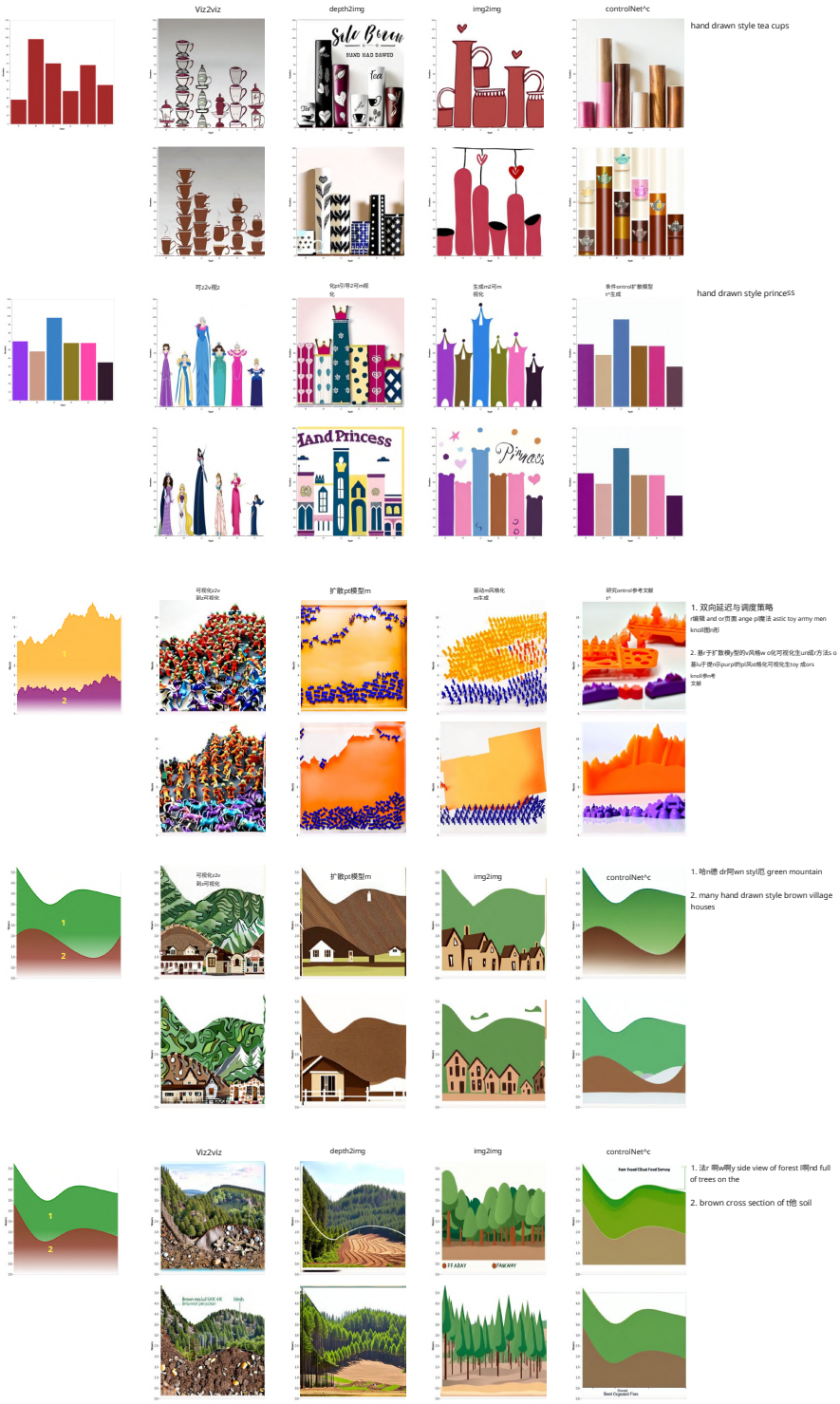

Viz2viz
depth2img
img2img
controlNet^c
hand drawn style tea cups
hand drawn style princess
1. 双向延迟与调度策略
2. many hand drawn style brown village houses
1. 哈n德 dr同wn styl尼 green mountain
2. many hand drawn style brown village houses
1. 这r 啊w啊y side view of forest 啊nd full of trees on the
2. brown cross section of t他 soil

图16：补充结果4。

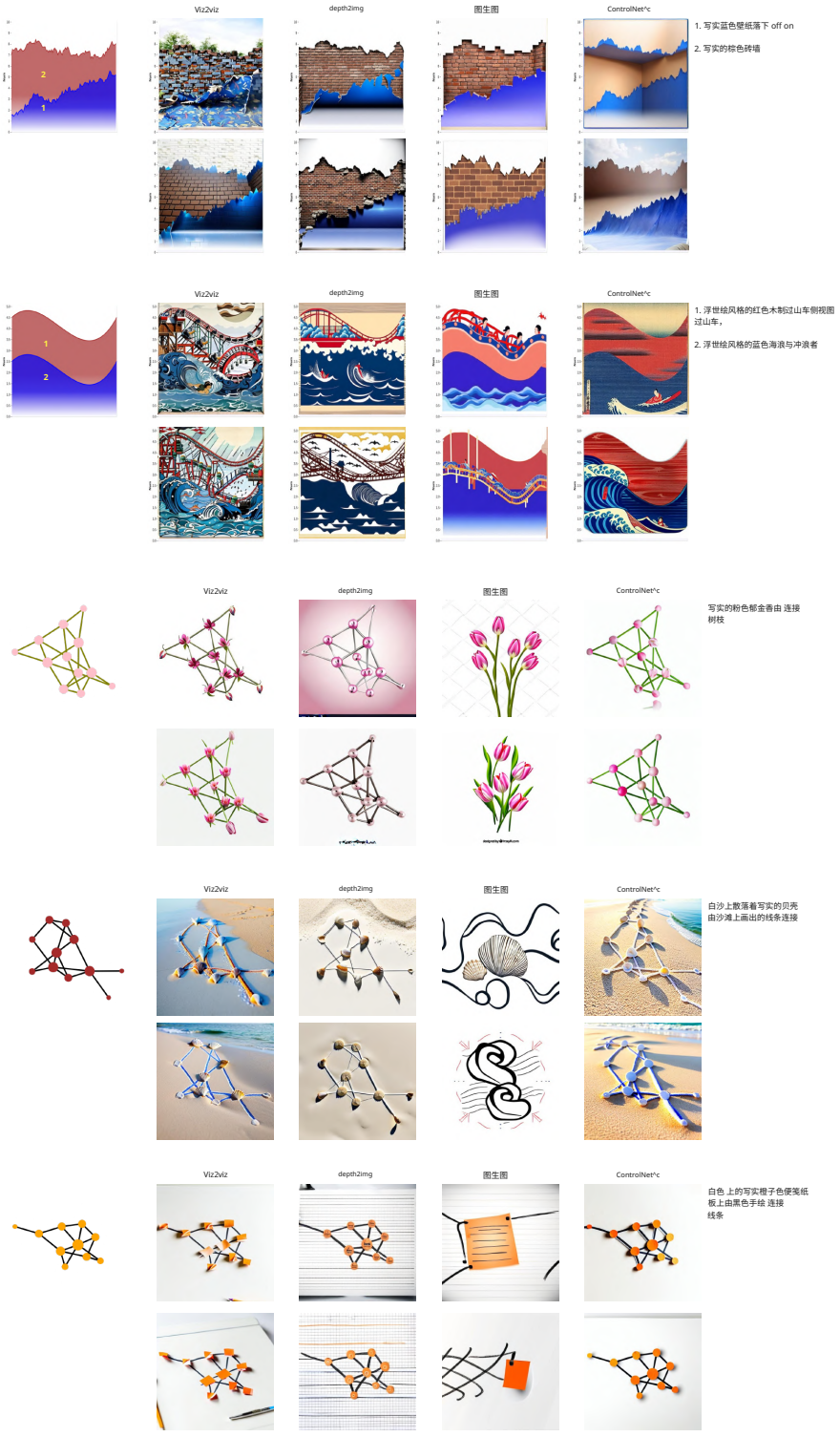

Viz2viz
depth2img
图生图
ControlNet^c
1. 写实蓝色壁纸落下 off on
2. 写实的棕色砖墙
1. 浮世绘风格的红色木制过山车侧视图过山车，
2. 浮世绘风格的蓝色海浪与冲浪者
写实的粉色郁金香由 连接树枝
白沙上散落着写实的贝壳由沙滩上画出的线条连接
白色 上的写实橙子色便笺纸板上由黑色手绘 连接线条

图 17：补充结果 5。

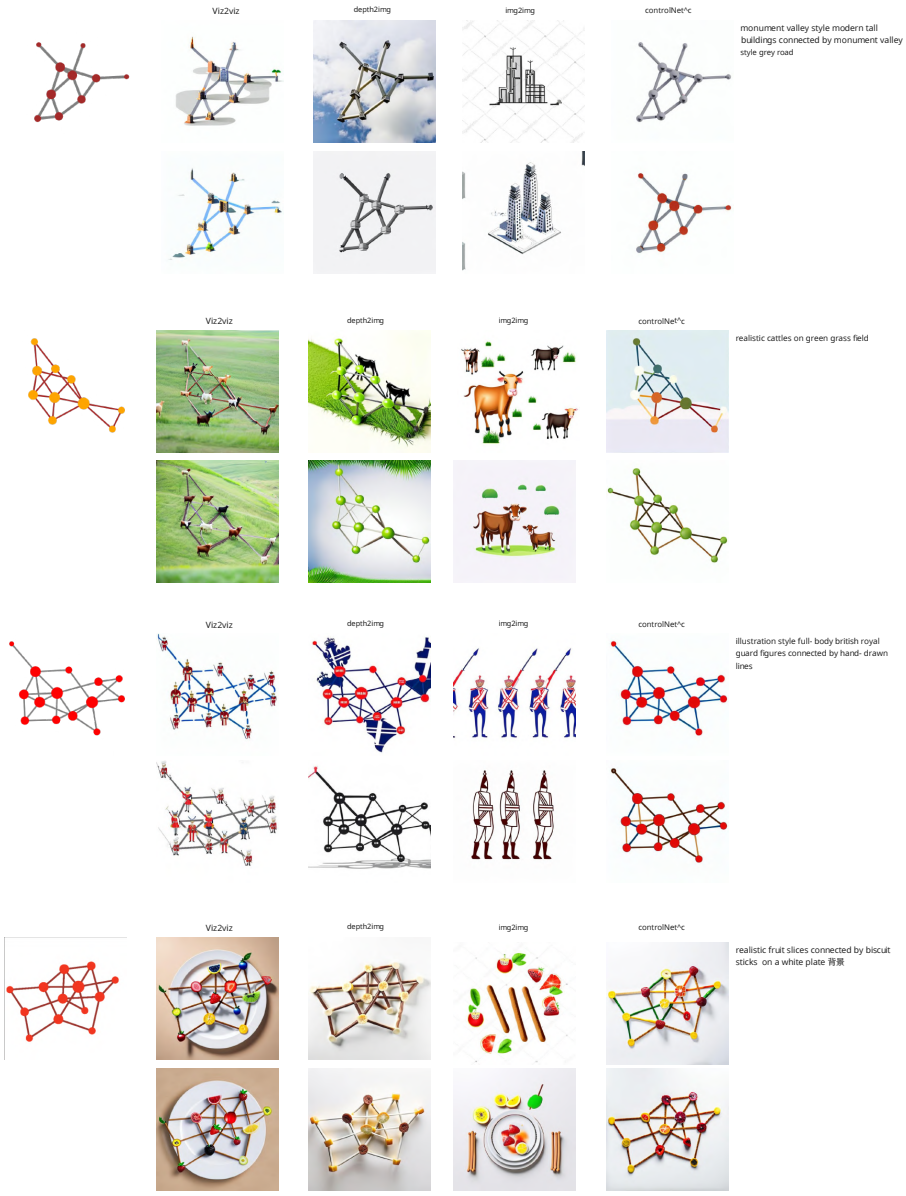

Viz2viz
depth2img
img2img
controlNet^c
monument valley style modern tall buildings connected by monument valley style grey road
Viz2viz
depth2img
img2img
controlNet^c
realistic cattles on green grass field
Viz2viz
depth2img
img2img
controlNet^c
illustration style full- body british royal guard figures connected by hand- drawn lines
Viz2viz
depth2img
img2img
controlNet^c
realistic fruit slices connected by biscuit sticks on a white plate 背景

图 18：补充结果 6。

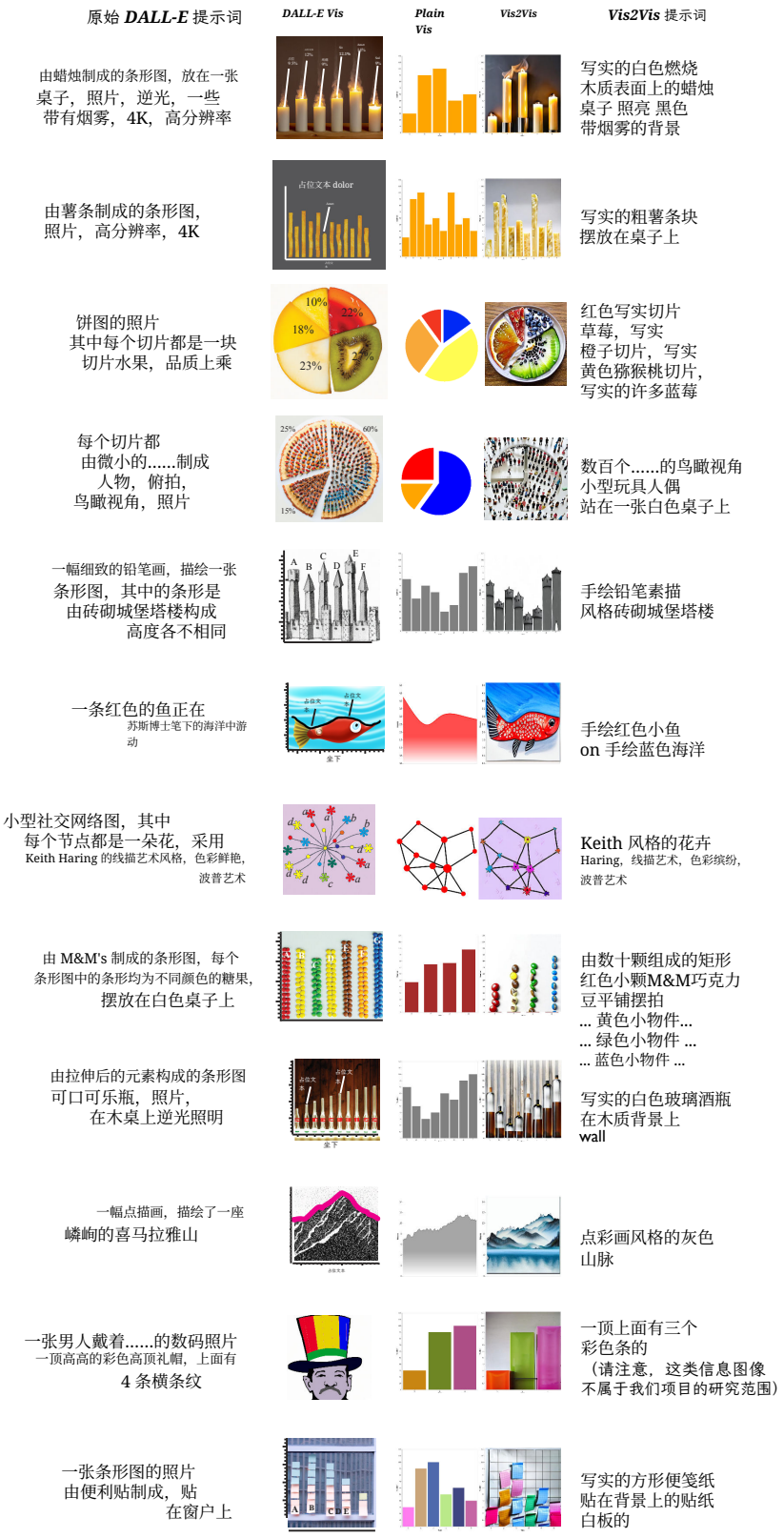


图 19：比较　　DALL-E 2 与 viz2viz 之间的比较。请注意，DALL-E 2 无法指定底层数据，而 viz2viz 可以　　n.