

Computational Photograph

Generative Novel View Synthesis with 3D-Aware Diffusion Models

主要看一下他的 `related work` 部分

The article discusses novel view synthesis (NVS), particularly focusing on methods that leverage both geometry-based priors and generative models, like diffusion models. Here's a summary of the related work discussed: 本文讨论了新颖的视图合成 (NVS), 特别关注利用基于几何的先验和生成模型 (如扩散模型) 的方法。以下是讨论的相关工作的摘要:

- Geometry-based NVS:** Many earlier methods for NVS used explicit 3D reconstructions from images, such as multi-view stereo (MVS) or depth maps. These approaches recover the 3D structure of a scene by estimating camera parameters or using multi-view stereo. However, these methods often struggle with photorealistic rendering and fail to generate complete views, especially in occluded areas. Some newer methods combine MVS with deep learning to improve NVS, but the overall performance can suffer if the MVS pipeline is inaccurate .
 - 基于几何的 NVS :** 许多早期的 NVS 方法使用来自图像的显式 3D 重建, 例如多视图立体 (MVS) 或深度图。这些方法通过估计相机参数或使用多视图立体来恢复场景的 3D 结构。然而, 这些方法通常难以实现照片级逼真的渲染, 并且无法生成完整的视图, 尤其是在遮挡区域。一些较新的方法将 MVS 与深度学习相结合以改进 NVS, 但如果 MVS 管道不准确, 整体性能可能会受到影响。
- Regression-based NVS:** Several deep learning-based NVS approaches focus on regression models that predict novel views based on input images. These methods often rely on 3D representations and differentiable neural rendering. While these methods are often designed for a specific scene, few-shot NVS approaches have emerged that generalize across different 3D scenes, allowing predictions from just one or a few input images. Some use meta-learning or test-time optimization to handle the challenge of extrapolation from sparse data
 - 基于回归的 NVS :** 几种基于深度学习的 NVS 方法专注于基于输入图像预测新视图的回归模型。这些方法通常依赖于 3D 表示和可微分神经渲染。虽然这些方法通常是针对特定场景设计的, 但出现了适用于不同 3D 场景的少样本 NVS 方法, 允许仅从一张或几张输入图像进行预测。有些方法使用元学习或测试时优化来应对从稀疏数据推断的挑战。
- Generative Models for NVS:** Generative models, especially for extrapolating views far from the original ones, have become an important focus. These methods work by generating plausible scenes based on weak or no geometry priors, which allows them to fill in missing parts of a scene. While they perform well for generating diverse outputs, their consistency in long-range view extrapolation can suffer. Recent works aim to combine generative models with pose-conditioned image diffusion models to improve consistency in view synthesis .
 - NVS 的生成模型 :** 生成模型, 尤其是用于推断远离原始视图的视图的生成模型, 已成为一个重要的焦点。这些方法通过基于弱几何先验或无几何先验生成可信场景来工作, 这使得它们能够填补场景中缺失的部分。虽然它们在生成多样化输出方面表现良好, 但它们在长距离视图推断中的一致性可能会受到影响。最近的研究旨在将生成模型与姿势条件图像扩散模型相结合, 以提高视图合成的一致性。

4. **3D GANs:** 3D GANs integrate adversarial training with implicit neural scene representations to generate 3D objects. These models, while effective for object-scale generation, require a canonical camera pose and are challenging to train for complex scenes. A 3D GAN can be inverted for NVS, offering a good balance between generative power and scene consistency, but it still suffers from challenges like camera pose inaccuracies and brittleness without robust 3D priors .
 1. **3D GAN :** 3D GAN 将对抗性训练与隐式神经场景表示相结合，以生成 3D 对象。这些模型虽然对于对象级生成有效，但需要规范的相机姿势，并且难以训练复杂场景。3D GAN 可以反转用于 NVS，在生成能力和场景一致性之间提供良好的平衡，但它仍然面临相机姿势不准确和没有稳健的 3D 先验的脆弱性等挑战。
5. **Recent Advances in Diffusion Models:** The article introduces the use of 2D diffusion models for NVS, extending them to be 3D-aware by conditioning them on 3D neural features. This allows the model to generate realistic and consistent novel views even with minimal input. The key innovation is the introduction of a latent 3D feature field, which helps maintain geometrical consistency across generated views. This method is shown to outperform earlier diffusion-based and generative approaches
 1. **扩散模型的最新进展：** 本文介绍了 2D 扩散模型在 NVS 中的应用，通过对 3D 神经特征进行调节，将其扩展为 3D 感知。这使得模型即使在输入最少的情况下也能生成逼真且一致的新视图。关键创新是引入了潜在的 3D 特征场，这有助于在生成的视图中保持几何一致性。事实证明，该方法优于早期的基于扩散和生成的方法。

In summary, while traditional NVS methods have struggled with geometric consistency, especially for long-range extrapolation, the combination of geometry priors and generative models, particularly diffusion models, shows promise in generating realistic, consistent views from limited input data .

总之，虽然传统的 NVS 方法在几何一致性方面存在困难，特别是对于长距离外推，但几何先验和生成模型（特别是扩散模型）的结合，有望从有限的输入数据中生成真实、一致的视图。

对应的论文代表作

以下是从不同类型的新视角合成（NVS）方法中选出的代表作，以及每个论文方法的简单介绍：

1. **几何基础的NVS:**
 - **论文名称:** "Photo Tourism: Exploring Photo Collections in 3D" (Snavely et al., 2006)
 - **方法简介:** 该方法通过从图像集合中恢复三维结构并估计相机参数，利用多视角立体重建（MVS）技术来生成新视角的图像。通过对不同角度的照片进行匹配，重建出物体或场景的三维模型，从而实现视角合成。
2. **回归基础的NVS:**
 - **论文名称:** "Neural 3D Mesh Renderer" (Kato et al., 2018)
 - **方法简介:** 该方法通过神经网络结合3D物体的几何信息进行视图合成。它使用深度学习来学习物体表面的几何特征，并利用回归方法在训练数据上进行预测。通过这种方法，可以从少量输入图像预测不同视角下的物体图像。
 - **论文名称:** "PixelNeRF: Neural Radiance Fields from a Single Image" (Chen et al., 2020)

- **方法简介:** PixelNeRF通过神经辐射场（NeRF）生成模型，基于单张输入图像预测新视角下的图像。该方法结合了3D场景表示和回归损失函数，成功实现了单张图像的视图合成，适用于相对简单的场景。

3. 生成模型的NVS:

- **论文名称:** "Look Outside The Room: Synthesizing a Consistent Long-term 3D Scene Video from a Single Image" (Ren et al., 2022)
 - **方法简介:** 该方法通过生成模型来进行长距离视角外推，适用于更复杂的场景。它通过一个基于变换器（Transformer）的生成模型，无需显式的3D几何先验，可以从单张图像中生成与输入图像一致的远距离视图。
- **论文名称:** "DreamFusion: Text-to-3D using 2D Diffusion Models" (Poole et al., 2022)
 - **方法简介:** DreamFusion使用2D扩散模型进行文本引导的3D生成。通过优化一个神经辐射场（NeRF），结合文本条件和图像条件，能够从文本描述中生成三维场景。该方法使用扩散模型的优势进行高质量的3D生成，特别是在多视图一致性和几何一致性方面。

4. 3D GANs:

- **论文名称:** "3D-GANs: Learning a Generative Model of 3D Objects" (Wu et al., 2016)
 - **方法简介:** 该论文提出了3D GAN，它结合了生成对抗网络（GAN）和3D对象的隐式场景表示。通过训练GAN模型，它能够从无结构数据中生成3D对象的三维视图，尤其适用于单一物体的生成任务。

5. 基于2D扩散模型的NVS:

- **论文名称:** "Diffusion Models Beat GANs on Image Synthesis" (Dhariwal & Nichol, 2021)
 - **方法简介:** 该论文提出了扩散模型在图像生成方面优于GAN的效果。扩散模型通过逐步去噪的方式生成图像，能够在多个生成任务中展示出优越的性能。虽然此方法主要是用于2D图像生成，但其原理也为3D视角合成任务提供了启示。

这些论文代表了不同类型的NVS方法，从几何基础方法、回归方法、生成模型到GAN和扩散模型，每种方法都有其独特的优势和局限。

大致timeline

为了帮助你撰写**基于NVS的综述**，我将梳理一条**时间线**，包括代表性的研究成果以及每个阶段的关键方法。这些方法包括几何基础的NVS、回归基础的NVS、生成模型的NVS、以及结合了深度学习和扩散模型的现代方法。

1. 早期几何基础的NVS（2000年-2010年）

- **2006年 - Snavely et al., "Photo Tourism: Exploring Photo Collections in 3D"**
 - **方法:** 利用图像集合恢复3D结构并估计相机参数，通过多视角立体重建（MVS）来实现视图合成。
 - **贡献:** 该方法为基于图像的三维重建和视角合成奠定了基础，推动了多视角重建技术的发展。
- **2007年 - Goesele et al., "Multi-view Stereo for Community Photo Collections"**
 - **方法:** 提出了用于社区照片集合的多视角立体重建（MVS），从多张图片中提取深度信息来合成新视角。

- 贡献: 提供了一种处理不完全照片集（例如社交媒体照片）的方案，使得从多个视角重建场景成为可能。

2. 回归基础的NVS方法（2010年-2017年）

- **2016年 - Flynn et al., "Deep Stereo: Learning to Predict New Views from the World's Imagery"**
 - 方法: 通过卷积神经网络（CNN）预测新视角图像，使用深度学习从多张图像中提取特征，生成新的视图。
 - 贡献: 该方法是深度学习与传统立体视觉技术的结合，标志着回归方法在NVS中的首次应用。
- **2018年 - Kato et al., "Neural 3D Mesh Renderer"**
 - 方法: 引入神经网络渲染器，通过3D物体的几何信息与深度学习模型结合来进行视角合成。
 - 贡献: 该方法有效地利用了3D几何信息并结合深度学习进行优化，推动了基于神经网络的回归方法发展。
- **2020年 - Chen et al., "PixelNeRF: Neural Radiance Fields from a Single Image"**
 - 方法: 利用神经辐射场（NeRF）生成模型，通过回归方法根据单张输入图像生成新视角图像。
 - 贡献: 提出了基于神经辐射场的NVS方法，为少量输入图像的视角合成提供了新的思路。

3. 生成模型基础的NVS方法（2018年-2022年）

- **2018年 - Niemeyer et al., "GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields"**
 - 方法: 提出了GIRAFFE模型，将场景表示为组合生成神经特征场，可以有效地生成新视角图像。
 - 贡献: 通过生成特征场，方法能够生成不同场景的一致性视图，特别适用于复杂场景的生成。
- **2019年 - Wiles et al., "PixelSynth: Generating a 3D-consistent Experience from a Single Image"**
 - 方法: 提出了PixelSynth，通过生成模型从单张图像中生成3D一致性的视图，填补了深度信息缺失的空白。
 - 贡献: 在生成模型和几何一致性之间建立了桥梁，推动了生成模型在NVS中的应用。
- **2021年 - Rombach et al., "Geometry-free View Synthesis: Transformers and No 3D Priors"**
 - 方法: 提出了无需3D先验的生成视图合成方法，使用变换器（Transformer）网络来进行生成，适应了没有几何先验的场景。
 - 贡献: 在没有几何信息的条件下，通过生成模型实现了长时间跨度的视角外推，展示了生成模型的强大能力。
- **2022年 - Ren et al., "Look Outside The Room: Synthesizing a Consistent Long-term 3D Scene Video from a Single Image"**
 - 方法: 该方法使用变换器网络结合生成模型，进行长时间跨度的3D场景视频生成。即使在只有单张图像的条件下，也能保持场景的一致性。
 - 贡献: 提出了从单张图像生成长时间序列的可行方法，是生成视角合成领域的一次突破。

4. 基于扩散模型的NVS方法（2021年-现在）

- **2021年 - Dhariwal & Nichol, "Diffusion Models Beat GANs on Image Synthesis"**

- 方法: 提出了扩散模型在图像生成上的优势, 通过逐步去噪的方式生成图像, 并展示了在生成任务中优于GAN。
- 贡献: 确立了扩散模型作为图像生成技术的领先地位, 奠定了后续扩散模型应用于NVS的基础。
- **2022年 - Poole et al., "DreamFusion: Text-to-3D using 2D Diffusion Models"**
 - 方法: 使用2D扩散模型来进行文本到3D的生成, 通过优化NeRF模型来生成三维场景。
 - 贡献: 扩散模型首次被成功应用于3D生成, 并通过文本引导实现高质量的3D场景合成。
- **2022年 - Watson et al., "Novel View Synthesis with Diffusion Models"**
 - 方法: 使用扩散模型进行NVS, 通过将生成模型与几何先验结合, 生成更加一致和真实的新视角图像。
 - 贡献: 将扩散模型与3D几何先验结合, 解决了传统生成模型在视角一致性和几何一致性方面的不足。
- **2023年 - Kulháněk et al., "Consistent View Synthesis with Pose-Guided Diffusion Models"**
 - 方法: 提出了基于姿势引导的扩散模型, 通过结合姿势信息生成一致的视角, 适用于复杂场景的长时间序列生成。
 - 贡献: 提供了一种更加灵活的生成方式, 可以根据不同的视角需求生成一致的图像, 解决了多视图一致性问题。

时间线总结:

- **2000-2010年:** 几何基础方法占主导地位, 主要集中在通过多视角重建和MVS进行NVS。
- **2010-2017年:** 回归基础方法兴起, 利用深度学习进行视角合成, 推动了基于神经网络的NVS发展。
- **2018-2022年:** 生成模型开始应用于NVS, 尤其是在没有几何先验的条件下, 通过GAN和变换器网络解决了长时间跨度的视角合成问题。
- **2021-至今:** 扩散模型成为NVS领域的重要技术, 尤其是与几何先验相结合, 进一步提高了生成视图的一致性和真实感。

这些代表作涵盖了从几何基础到生成模型的进展, 展示了NVS技术的演变和发展趋势。希望这些信息能为你写综述提供清晰的框架和参考。

Neural Radiance Fields (NeRF) 在新视角合成 (NVS) *的发展历史中具有重要的里程碑地位, 可以插入到*生成模型基础的NVS方法阶段, 并且它的提出和发展进一步推动了基于神经网络的生成模型应用。

插入位置: 2019年-2021年 (生成模型基础的NVS方法)

具体时间节点:

- **2019年: NeRF (Neural Radiance Fields) 被提出并首次应用于新视角合成。** NeRF利用深度学习进行三维场景建模, 通过卷积神经网络学习三维空间中的体积信息, 从而合成出从新视角生成的图像。

论文: "**NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis**" (Mildenhall et al., 2020)

方法简介:

- NeRF通过一个深度神经网络表示三维场景，每个空间点都由一个颜色和透明度组成，利用体积渲染技术计算从不同视角看场景的图像。
- 它能够生成非常高质量、具有高度几何一致性的视图，特别适用于细节丰富的场景和复杂的光照环境。

NeRF的贡献:

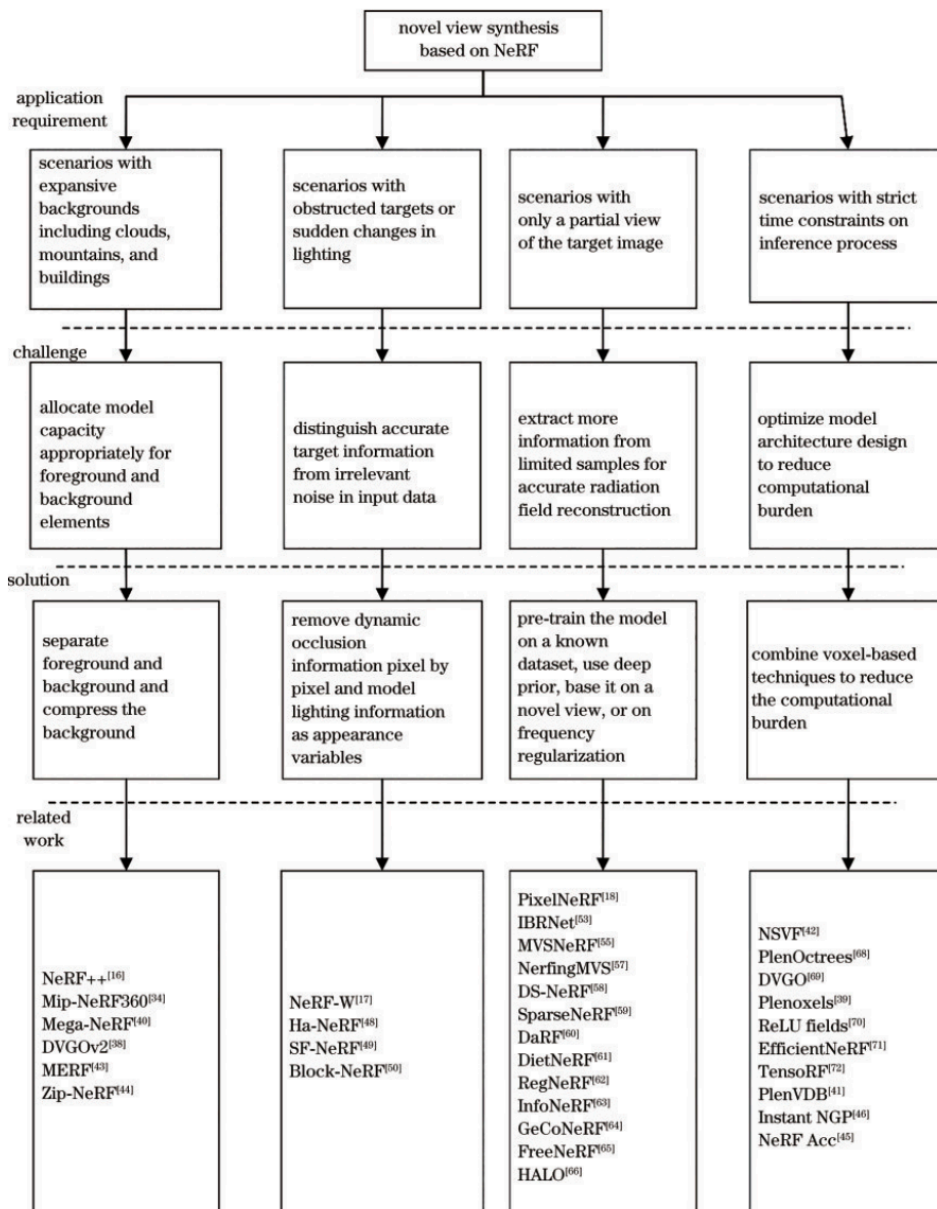
- **高质量新视角生成**：NeRF能够生成极为逼真的新视角图像，尤其是在需要细致的光照和几何一致性的场景中表现出色。
- **精确的三维表示**：NeRF方法引入了体积渲染的概念，成功实现了对场景的三维建模，避免了传统方法中的几何模糊和视图一致性问题。
- **开启了基于神经网络的NVS新时代**：NeRF的提出推动了后续许多基于深度学习的生成方法，如基于生成对抗网络（GANs）、扩散模型的NVS应用。

在时间线中的位置：

- **之前（2010-2017年）**：回归基础的NVS方法仍然占主导地位，基于回归网络的图像预测方法逐渐演化。
- **2019年**：NeRF的提出标志着**生成模型**在NVS中的应用取得了突破。NeRF的方法结合了生成模型的优点（能够处理视图的多样性和复杂性），并通过对三维场景的深度表示进行建模，解决了传统方法的许多限制。
- **之后（2021年及以后）**：随着NeRF的成功，研究者开始扩展NeRF的应用，提出了如Mip-NeRF、FastNeRF等改进版本，进一步提高了渲染效率和大规模场景的处理能力。同时，也有很多基于NeRF的变种在NVS领域中找到了广泛应用。

综述建议：

在撰写综述时，你可以将**NeRF及其变种**作为生成模型基础的NVS方法的代表之一，特别是在讨论基于神经网络的**三维一致性**和高质量生成时，NeRF是非常关键的技术进展。



这个图表总结了与神经辐射场（NeRF）相关的应用需求、挑战、解决方案和相关工作。以下是梳理后的内容：

1. 应用需求（Application Requirement）

- 场景需求：
 - 包含广阔背景（如云、山、建筑）的场景。
 - 目标被遮挡或光照突然变化的场景。
 - 仅能获取目标部分视图的场景。
 - 对推理过程有严格时间限制的场景。

2. 挑战（Challenge）

- **模型容量分配**：如何合理分配模型容量给前景和背景元素。
 - **噪声处理**：如何从输入数据中区分准确的目标信息与无关噪声。
 - **有限样本信息提取**：如何从有限样本中提取更多信息以重建准确的辐射场。
 - **计算优化**：如何优化模型架构设计以减少计算负担。
-

3. 解决方案 (Solution)

- **前景与背景分离**：分离前景和背景，并对背景进行压缩。
 - **动态遮挡与光照建模**：
 - 逐像素移除动态遮挡信息。
 - 将光照信息建模为外观变量。
 - **有限样本信息增强**：
 - 在已知数据集上预训练模型。
 - 使用深度先验 (deep prior)、基于新视角 (novel view) 或频率正则化 (frequency regularization) 。
 - **计算优化**：
 - 结合基于体素 (voxel-based) 的技术以减少计算负担。
-

4. 相关工作 (Related Work)

相关工作分为四类，主要围绕NeRF的扩展与优化：

4.1 大规模场景处理

- **代表性工作**：
 - NeRF++[16], Mip-NeRF360[94], Mega-NeRF[40], DVGOv2[38], MERF[43], Zip-NeRF[44]
 - **特点**：专注于处理广阔背景或大规模场景的NeRF变体。

4.2 动态与复杂光照场景

- **代表性工作**：
 - NeRF-W[17], Ha-NeRF[48], SF-NeRF[49], Block-NeRF[50]
 - **特点**：解决动态遮挡、光照变化或复杂外观建模问题。

4.3 有限视角或样本重建

- **代表性工作**：
 - PixelNeRF[18], IBRNet[53], MVNeRF[55], NerfingMVS[57], DS-NeRF[58], SparseNeRF[59], DaRF[60], DietNeRF[61], RegNeRF[62], InfoNeRF[63], GeCoNeRF[64], FreeNeRF[65], HALO[66]
 - **特点**：从稀疏或部分输入中重建完整场景，或通过正则化提升泛化能力。

4.4 高效计算与加速

- 代表性工作：
 - NSVF[42], PlenOctrees[68], DVGO[69], Plenoxels[39], ReLU fields[70], EfficientNeRF[71], TensorRF[72], PlenVDB[41], Instant NGP[46], NeRF Acc[45]
 - 特点：通过体素、张量分解、稀疏化等技术加速NeRF的训练或推理。
-

5. 总结

图表系统性地归纳了NeRF在不同应用场景下面临的挑战及解决方案，并分类列举了相关研究工作。核心方向包括：

1. **场景扩展**：处理大规模或复杂背景。
2. **鲁棒性提升**：解决遮挡、光照变化等问题。
3. **数据效率**：从有限输入中提取更多信息。
4. **计算优化**：降低计算成本，满足实时性需求。

如果需要进一步了解某类方法或具体论文，可以深入查阅对应文献。