

JacobiNeRF

JacobiNeRF: NeRF Shaping with Mutual Information Gradient(CVPR 2023)

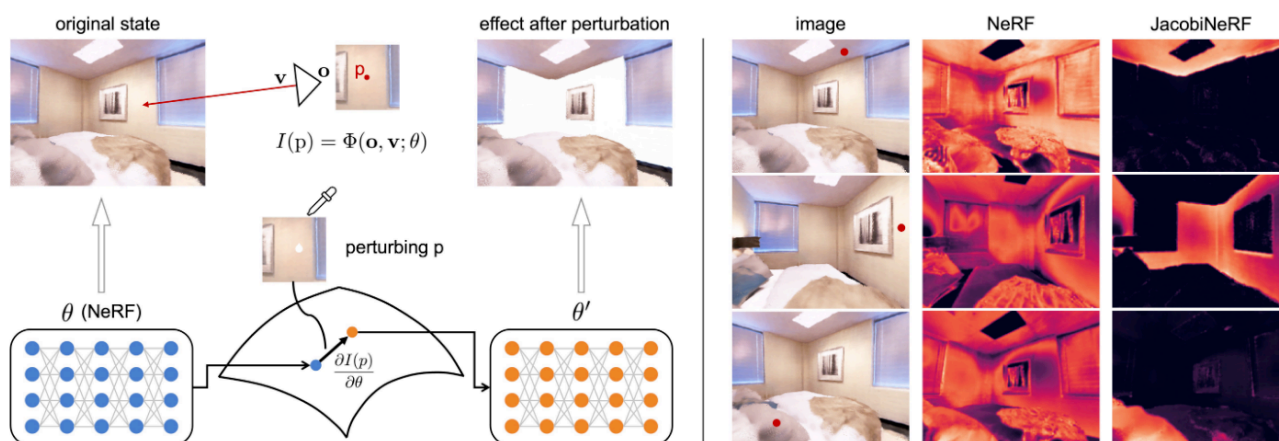


Figure 1. *Left*: a physical scene is composed of mutually-correlated entities, e.g., points belonging to the same part or object. We pursue a subspace of the parameters of an implicit scene representation so that when the scene is perturbed along the gradient of a specific point, a resonance emerges with other points having high mutual information. For example, if the brightness of a point p on the wall changes, the other wall points should also change coherently (while points on the ceiling should be unaffected). *Right*: a comparison between responses to perturbations along the gradient of the highlighted red point in the scene. The response from the original NeRF spreads all over the scene, while the localized response from the proposed JacobiNeRF demonstrates coherence between points with similar semantics.

Abstract

这篇文章主要介绍了一种称为 **JacobiNeRF** 的方法，目的是增强神经辐射场（NeRF）模型在捕捉和传播场景中语义关系的能力。传统的 NeRF 主要关注图像重建，即通过多视角图像生成一个场景的 3D 表示，但在处理不同场景实体之间的语义关联性上有所局限。

JacobiNeRF 则通过互信息和二阶信息建模，使模型能够更好地理解和利用场景中不同实体之间的关联性，从而实现更精确的标签传播和语义分割。以下是该方法的主要思想和技术要点：

1. 互信息调整（MI-shaping）

- **基本思想：**JacobiNeRF 通过对 NeRF 模型的梯度进行调整，使模型不仅能够重现场景的外观，还能捕捉场景实体之间的语义关系。
- **互信息的计算：**在模型参数空间中引入小的随机扰动，观察扰动对不同像素或实体的影响，从而衡量它们之间的互信息。若两个实体在扰动下表现出相似的变化模式，则说明它们存在较高的关联性。
- **余弦相似度：**使用梯度的余弦相似度来表示互信息的大小。相关性强的像素对的梯度方向相似，而无关的像素对则相互正交。

2. 标签传播

- **标签传播过程：**通过互信息，JacobiNeRF 可以将一个视图中的标签信息（如语义标签或实例标签）传播到其他视图或场景的其他点。标签传播的实现依赖于找到目标像素点和源像素点之间的最大互信息。
- **应用于稀疏标注：**在稀疏标注的情况下，JacobiNeRF 能够将少量的标签信息传递给未标注的区域，这在语义分割和实例分割任务中尤为重要。

3. 损失函数与优化

- **损失函数设计：**文章设计了一个包含多个部分的损失函数：NeRF 重建损失、互信息梯度损失和梯度单位球体约束。互信息梯度损失引导模型学习实体之间的关联性，单位球体约束确保梯度向量的长度保持稳定，有助于标签传播。

- **InfoNCE 损失**：通过 InfoNCE 损失，模型最大化了正样本对的梯度相似性，最小化负样本对的梯度相似性，从而实现有效的互信息编码。

4. 实验验证

- **性能验证**：文章通过实验验证了 JacobiNeRF 的性能，尤其是在稀疏标签场景下。与传统的 NeRF 相比，JacobiNeRF 在语义和实例分割任务中表现得更好，因为它能够利用互信息在不同视图或点之间传播标签。

总结

JacobiNeRF 通过互信息的建模来调整 NeRF，使其能够更好地捕捉场景中实体之间的语义关联性。这种方法不仅可以在 NeRF 模型的重建能力上有所提升，还能通过标签传播在稀疏标注的情况下实现更精确的语义分割和实例分割。

JacobiNeRF 论文中的公式解释

公式 (1)

$$I(p) = \Phi(o, v; \theta) = \int_0^{+\infty} w(t; \theta) c(p(t), v; \theta) dt \quad (1)$$

- **解释**：这是 NeRF 的基本体渲染方程，用于计算像素 p 的颜色 $I(p)$ 。通过沿射线 v 穿过场景体积，对密度权重函数 $w(t; \theta)$ 和颜色函数 $c(p(t), v; \theta)$ 进行积分，从而获得累积颜色值。

公式 (2) 和 (3)

$$\hat{I}(p_i) = I(p_i) + \sigma n \cdot \frac{\partial \Phi(o_i, v_i; \theta)}{\partial \theta^D} \quad (2)$$

$$\hat{I}(p_j) = I(p_j) + \sigma n \cdot \frac{\partial \Phi(o_j, v_j; \theta)}{\partial \theta^D} \quad (3)$$

- **解释：**这是通过 Taylor 展开得到的扰动公式，用于计算在噪声 n 影响下像素值的变化。 σ 控制噪声的大小， $\frac{\partial \Phi(o, v; \theta)}{\partial \theta^D}$ 是关于参数 θ^D 的雅可比矩阵，表示参数扰动对像素值的影响。

公式 (4)

$$I(\hat{I}(p_i), \hat{I}(p_j)) = H(\hat{I}(p_j)) - H(\hat{I}(p_j) | \hat{I}(p_i)) \quad (4)$$

- **解释：**计算扰动后的像素值 $\hat{I}(p_i)$ 和 $\hat{I}(p_j)$ 之间的互信息。互信息 $I(\cdot, \cdot)$ 通过两个熵的差值定义， $H(\hat{I}(p_j))$ 表示 $\hat{I}(p_j)$ 的熵， $H(\hat{I}(p_j) | \hat{I}(p_i))$ 表示给定 $\hat{I}(p_i)$ 后 $\hat{I}(p_j)$ 的条件熵。

公式 (5)

$$I(\hat{I}(p_i), \hat{I}(p_j)) = \log \left(\frac{1}{\sqrt{1 - \cos^2 \alpha}} \right) + \text{const.} \quad (5)$$

- **解释：**通过将噪声和雅可比矩阵用球坐标表示，互信息进一步简化。公式中的 α 是 $\frac{\partial \Phi_i}{\partial \theta^D}$ 和 $\frac{\partial \Phi_j}{\partial \theta^D}$ 之间的夹角，表示两个像素梯度方向的相似性（余弦相似度），从而互信息越大。

公式 (6)

$$\frac{\|\partial\Phi_i^T \partial\Phi_{i+}\|}{\|\partial\Phi_i\| \|\partial\Phi_{i+}\|} > \frac{\|\partial\Phi_i^T \partial\Phi_{i-}\|}{\|\partial\Phi_i\| \|\partial\Phi_{i-}\|} \quad (6)$$

- **解释：**该不等式用于确保高互信息的正样本对 (i, i+) 的梯度相似性大于低互信息的负样本对 (i, i-) 的梯度相似性，以此来保持相关实体之间的梯度方向一致。简单来说，就是相似的应该尽可能平行，不相似的应该正交。

公式 (7)

$$\mathcal{L}_{\text{MIG}} = -\log \frac{\exp(\|\cos(\partial\Phi_i, \partial\Phi_{i+})/\tau\|)}{\sum_{i+\cup i-} \exp(\|\cos(\partial\Phi_i, \partial\Phi_{i-})/\tau\|)} \quad (7)$$

- **解释：**这是 InfoNCE 损失，用于互信息梯度优化。通过最大化正样本的余弦相似性（分子部分）和最小化负样本的相似性（分母部分），此损失函数鼓励高相关性的像素对具有更高的梯度相似性。

公式 (8)

$$\mathcal{L}_{\text{TM}} = \mathcal{L}_{\text{NeRF}} + \lambda \mathcal{L}_{\text{MIG}} + \gamma(1.0 - \|\partial\Phi_i\|^2) \quad (8)$$

- **解释：**这是 JacobiNeRF 的完整训练损失函数，包含三个部分： $\mathcal{L}_{\text{NeRF}}$ 是用于场景重建的光度损失， \mathcal{L}_{MIG} 是互信息梯度调整项，第三项将梯度压缩到单位球体上，以改善标签传播效果。

公式 (9)

$$l_i^t = \arg \max_{l_i^t} P(l_i^t | \{(\mathbf{p}_k^s, l_k^s)\}, \Phi^J) \quad (9)$$

- **解释：**用于标签传播的最大后验估计（MAP）公式，表示在给定源标签的情况下，选择最大后验概率的标签作为预测标签 l_i^t 。

公式 (10) 和 (11)

$$k^* = \arg \max_k \mathbb{I}(\Phi^J(\mathbf{o}^s, \mathbf{v}_k^s; \theta), \Phi^J(\mathbf{o}^t, \mathbf{v}_i^t; \theta)) \quad (10)$$

$$= \arg \max_k \frac{\|\partial \Phi_k^{J,s} \cdot \partial \Phi_i^{J,t}\|}{\|\partial \Phi_k^{J,s}\| \|\partial \Phi_i^{J,t}\|} \quad (11)$$

- **解释：**公式 (10) 和 (11) 表示标签传播过程中选择互信息最大的源标签。通过梯度的余弦相似度衡量互信息，确保标签传播准确反映目标像素与源像素的关系。

公式 (12) 和 (13)

$$\Phi^J(:, \theta + \sigma \partial \Phi_k^{J,s}), k = 1 \dots K \quad (12)$$

$$I_k(p_i^t) = \Phi^J(\mathbf{o}^t, \mathbf{v}_i^t; \theta + \sigma \partial \Phi_k^{J,s}) \quad (13)$$

- **解释：**生成扰动后的 JacobiNeRF，用于在目标视图中生成标签图像。通过沿梯度方向的扰动，使得目标视图的标签传播更有效。

公式 (14)

$$R_k(p_i^t) = |I_k(p_i^t) - I(p_i^t)|, k = 1 \dots K \quad (14)$$

- **解释：** 计算扰动响应 $R_k(p_i^t)$ 作为目标图像中的标签分配依据。该公式使用扰动前后像素的绝对差值来量化标签的相关性，从而支持语义分割。