

采用扩散模型生成道路垃圾图像

丁莎欧 S2302W0294

摘要：本文旨在应用《Diffusion Models Beat GANs on Image Synthesis》论文中的方法，生成道路垃圾图像。采用改进的扩散模型架构及分类器指导技术，以提高图像生成的质量和精确度。最终目标是生成高质量的图像，并在生成过程中增加特定类型的道路垃圾图像。

关键词：扩散模型；神经网络；图像生成

1. 引言

近年来，图像生成技术取得了显著进展，特别是在生成高质量的合成图像方面。GANs 尽管在这一领域表现出色，但其在捕捉图像多样性和稳定训练方面存在局限性。扩散模型作为一种新兴技术，展示了其在稳定性和分布覆盖率上的优势。本文将应用扩散模型来生成包含特定类型道路垃圾的图像，通过分类器指导技术确保生成的图像高质量。

2. NN 方法在扩散模型中的应用

扩散模型通常采用神经网络架构，如 UNet。UNet 是一种流行的卷积神经网络（CNN）架构，广泛应用于图像生成和分割任务。扩散模型通过 UNet 架构来实现逐步去噪的过程，从而生成高质量的图像。扩散模型的训练过程依赖于神经网络的反向传播算法。通过最小化预测噪声的均方误差（MSE）损失，模型学习如何从噪声图像中逐步恢复原始图像。在训练过程中，神经网络不断更新其权重，以提高去噪和生成图像的能力。扩散模型可以结合分类器指导（classifier guidance）来增强图像生成的效果。分类器本质上是一个预训练的神经网络，用于提供图像分类的梯度信息。分类器指导通过调节分类器的输出，使得扩散模型在生成图像的过程中能够有选择性地增加或去除特定特征，从而提高图像的质量和精确度。

2.1 多分辨率注意力机制

扩散模型利用神经网络中的多分辨率注意力机制来捕捉不同尺度下的图像特征。这些机制可以帮助模型更好地理解 and 生成复杂的图像结构。多分辨率注意力机制通过在不同分辨率层次上引入注意力模块，提升了模型的表达能力和生成质量。

2.2 大规模数据训练

扩散模型需要在大规模数据集上进行训练，以学习丰富的图像特征。这与其他基于 NN 的方法一致，如 GANs 和 VAE（变分自编码器），都需要大量的数据和计算资源来训练高性能的生成模型。扩散模型通过大量的数据训练，可以更好地捕捉图像中的复杂模式和细节，从而生成更逼真的图像。

2.3 优化技术

扩散模型的优化过程中，使用了神经网络中的一些先进技术，如 Adam 优化器、学习率调度等，以加速训练和提高模型性能。这些优化技术在神经网络训练中广泛应用，并在扩散模型中也同样适用。

3. 道路垃圾图像生成

3.1 图像生成方法

改进的扩散模型架构：基于 UNet 架构，增加模型的深度和宽度，采用多分辨率注意力机制，以及 BigGAN 风格的上采样和下采样方法。

分类器指导：使用分类器的梯度在采样过程中进行指导，通过调节分类器权重，在生成高质量图像的同时增加特定的垃圾。

3.2 图像生成实验步骤

数据准备：收集并标注包含垃圾和不包含垃圾的图像数据集，分为训练集和测试集。

模型训练：使用改进的扩散模型架构，结合分类器指导技术，在训练集上进行模型训练。训练过程中，模型通过预测噪声来逐步还原图像，并利用分类器指导增加特定类型的垃圾。

图像生成与评估：在测试集中生成新的图像，并评估生成图像的质量和垃圾增加的效果。主要评估指标包括 Fréchet Inception Distance (FID)、Inception Score (IS) 和使用扩充后数据进行训练的目标检测模型的精度和召回率。

3.3 扩散模型的训练和采样算法

扩散模型通过逆转逐步加噪过程生成样本。该过程如下：

加噪过程：从一个真实样本 x_0 开始，逐步添加噪声，生成一系列样本 x_1, x_2, \dots, x_T 。

噪声添加的过程可以表示为： $q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t I)$ ，其中 β_t 是每一步的噪声强度。

去噪过程：模型学习一个去噪函数 $\epsilon_\theta(x_t, t)$ ，用于从加噪样本 x_t 生成去噪样本 x_{t-1} 。去噪过程可以表示为： $p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$ ，其中均值 $\mu_\theta(x_t, t)$ 由预测的噪声 $\epsilon_\theta(x_t, t)$ 计算得到，方差 $\Sigma_\theta(x_t, t)$ 可以固定或由模型学习。

训练目标：训练目标是 minimized 预测噪声与真实噪声之间的均方误差，该均方误差表示为：

$$L_{simple} = \mathbb{E}_{x_0, t, \epsilon} [||\epsilon - \epsilon_\theta(x_t, t)||^2]。$$

3.4 分类器指导算法

分类器指导算法用于在生成过程中引入类别信息，提高有条件生成样本的质量。其主要步骤如下：

训练分类器：训练一个分类器 $C(x)$ 来预测输入图像的类别。

计算分类器梯度：在生成过程中，使用分类器的梯度来引导生成样本朝向特定类别。对于每个生成的样本 x_t ，计算分类器梯度： $g_t = \nabla_{x_t} \log p(y|x_t)$ 。

梯度引导生成：在每一步去噪过程中，使用分类器梯度调整去噪方向： $\tilde{\mu}_\theta(x_t, t) = \mu_\theta(x_t, t) + s \cdot \Sigma_\theta(x_t, t) \cdot g_t$ ，其中 s 是梯度缩放因子，用于控制样本质量和多样性之间的权衡。

3.5 样本生成流程

初始化：从噪声分布 $x_T \sim \mathcal{N}(0, I)$ 开始。

迭代去噪：对于 $t = T, T-1, \dots, 1$ ，执行以下步骤：

(1) 计算去噪样本的均值和方差： $\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}} \epsilon_\theta(x_t, t) \right)$, $\Sigma_\theta(x_t, t) = \beta_t \cdot I$ 。

(2) 使用分类器指导调整均值： $\tilde{\mu}_\theta(x_t, t) = \mu_\theta(x_t, t) + s \cdot \Sigma_\theta(x_t, t) \cdot \nabla_{x_t} \log p(y|x_t)$ 。

(3) 从调整后的均值和方差中采样： $x_{t-1} \sim \mathcal{N}(\tilde{\mu}_\theta(x_t, t), \Sigma_\theta(x_t, t))$ 。

输出最终样本：当 $t = 0$ 时，输出生成的样本 x_0 。

3.6 训练图像生成样本图及结果图



图 1 训练集样本



图 2 训练集样本



图 3 生成图示例



图 4 生成图示例



图 5 生成图示例



图 6 生成图示例



图 7 生成图示例



图 8 生成图示例

采用扩散模型，使用例如图 1、图 2 的数据集训练生成道路垃圾图像的模型，具体程序编写流程为：

（1）噪声注入：扩散模型从随机噪声图像开始，这些噪声图像是通过将高斯噪声注入到原始图像中生成的。

（2）逐步去噪：模型通过一系列逐步去噪的过程来生成清晰的图像。每一步都使用一个神经网络来预测噪声，并从当前状态中减去该噪声，从而逐渐还原出清晰的图像。

（3）训练过程：扩散模型的训练过程涉及最大化似然估计，这需要对所有中间步骤的去噪过程进行建模。这种方法使得模型能够学习如何从噪声中生成逼真的图像。

（4）分类器引导：对于条件图像生成任务，使用了分类器引导（classifier guidance）技术。这种方法通过调整分类器梯度的尺度来平衡生成图像的多样性和保真度。

（5）上采样：通过结合引导和上采样技术，可以进一步提高高分辨率条件图像合成的样本质量。

图 3-图 8 为采用扩散模型生成的像素为 256×256 的道路垃圾图像效果展示。

4. 结论

通过应用《Diffusion Models Beat GANs on Image Synthesis》中的方法，有效生成了高质量的图像，并有效增加了特定的道路垃圾图像。实验结果表明，改进的扩散模型架构和分类器指导技术在提升图像质量和增加特定道路垃圾图像方面具有显著优势。