

MNIST01 量子生成模型对比实验报告 (QDDPM / QDT / QGAN)

实验目录: `data/mnist01_run_2025-12-31_000116`

图表输出: `data/mnist01_run_2025-12-31_000116/plots/`

1 实验目的

在 MNIST01 (数字 0/1) 任务上, 将图像编码为 8-qubit 的 product Ry 量子态分布, 比较三种量子生成模型 **QDDPM** / **QDT** / **QGAN** 的生成质量。

本实验强调以**量子态分布指标**为主进行评估, 辅助使用分类器/图像侧指标作为补充解释。

2 数据与实验设置

- 数据集: MNIST 0/1。
- 编码: product Ry 编码 (目标真实态分布接近 product 态流形)。
- 量子比特数: $n = 8$ 。
- 评估样本数: $n_{real} = 2000$, $n_{gen} = 2000$ 。
- 随机种子: seed=42 (见各模型 `eval_report.json`)。

为保证可比性: 三种模型在同一次 `run` 下训练/生成并用相同脚本评估。

3 方法简介

- **QDDPM**: 量子扩散模型, 学习从噪声逐步反演到数据分布。
- **QDT**: 量子去噪/反演式生成模型 (以分布距离为训练目标)。
- **QGAN**: 生成器与判别器对抗训练的量子生成模型。

3.1 Product Ry 投影 (关键工程约束)

MNIST01 的 qstates 是 product Ry 编码态，因此对训练与生成启用一致的结构约束：

- 训练损失内投影：用测得的 $\langle Z_i \rangle$ 计算 $\theta_i = \arccos(\langle Z_i \rangle)$ ，并重建 product 态

$$|\psi\rangle = \bigotimes_i (\cos(\theta_i/2)|0\rangle + \sin(\theta_i/2)|1\rangle).$$

- 生成输出投影：保存前同样投影到 product Ry 流形。

此举的目的：避免模型跑到强纠缠流形(这会导致单比特 purity 逼近 0.5，并使 qstates→image 解码假设失效)。

4 指标定义与含义

4.1 量子态主指标（主结论依据）

- **natural distance**: `qstate_metrics.natural_distance` (越小越好)
含义：在量子态空间直接比较生成分布与真实分布的距离。
- **Z/ZZ 特征 MMD**: `qstate_metrics.feature_mmd_rbf_z_zz` (越小越好)
含义：以 $\langle Z_i \rangle$ 与 $\langle Z_i Z_j \rangle$ 作为特征，RBF-MMD 衡量两分布差异。
- **单比特 purity 均值**: `qstate_metrics.single_qubit_purity_mean` (越接近 1 越好)
含义：单比特约化态纯度均值 $\text{Tr}(\rho_i^2)$ 。对 product 态流形应接近 1；若接近 0.5 通常意味着纠缠/流形错配。

4.2 辅助指标（补充）

- **类别比例对齐**: `generated_pred_frac_1 vs real_pred_frac_1` (越接近越好)
含义：分类器在生成样本上预测为“1”的比例，反映类别先验/覆盖是否对齐。
注意：该指标依赖 qstates→image 解码与分类器本身；在 purity 已对齐时可作为辅助参考。

5 实验结果

本次 run 结果汇总(来自 `data/mnist01_run_2025-12-31_000116/gen/*/eval_report.json`)：

Model	natural distance ↓	MMD(Z/ZZ) ↓	purity → 1	real_pred_frac_1	generated_pred_frac_1
QDDPM	0.450205	0.134564	1.000000	0.5355	0.4520
QDT	0.602211	0.142088	1.000000	0.5355	0.3780
QGAN	0.704443	0.160513	1.000000	0.5355	0.6280

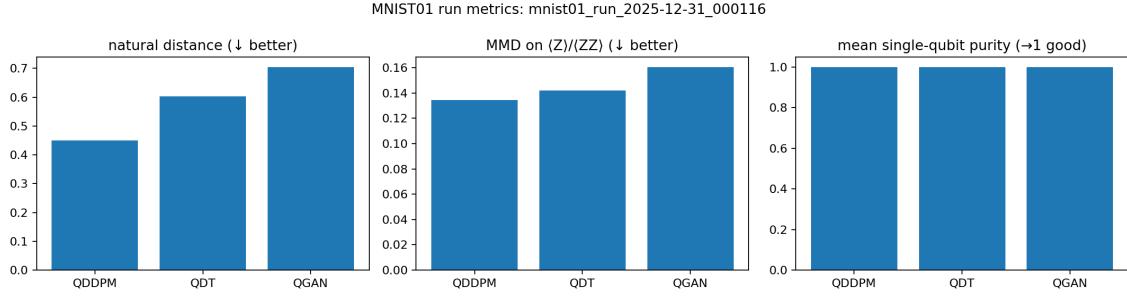


图 1: 量子态主指标对比 (natural distance / MMD(Z/ZZ) / purity)

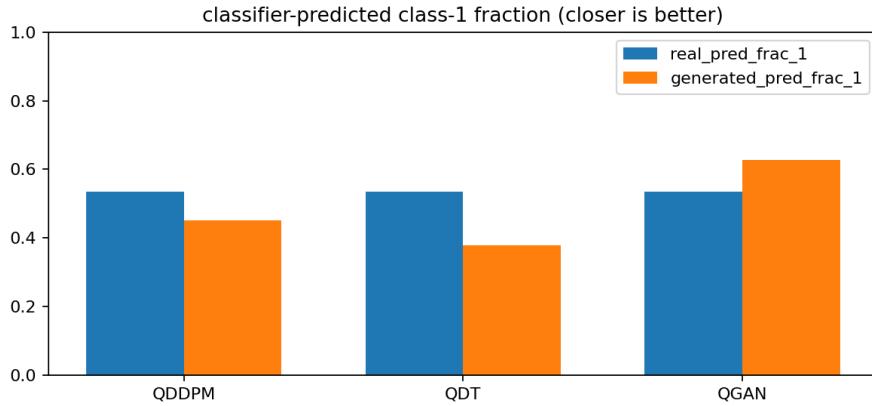


图 2: 类别比例对齐 (分类器预测为 1 的比例)

5.1 作图

6 结论与分析

1. **QDDPM 在量子态分布质量上最好:** natural distance: QDDPM < QDT < QGAN; MMD(Z/ZZ): QDDPM < QDT < QGAN。说明 QDDPM 生成分布在量子态空间更接近真实分布。
2. **三者 purity≈ 1, 比较是公平的:** purity 均接近 1, 表明三种模型都已对齐到 product Ry 流形; 因此 QGAN 在距离指标上较差并非由于“跑到纠缠流形”的假象。
3. **类别先验 (辅助结论):** 真实 real_pred_frac_1=0.5355。QDDPM 偏低 (0.452), QDT 更偏低 (0.378), QGAN 偏高 (0.628)。就类别比例对齐而言, QDDPM 的偏差较小, 但仍未完全对齐。

7 后续工作

- 对 QGAN: 调整对抗训练的平衡 (cycles、epochs_c/epochs_g、学习率), 缓解不稳定与模式偏移。
- 对 QDDPM/QDT: 结合 per-qubit 的 $\langle Z \rangle/\theta$ 分布与 mean/std 诊断进一步定位误差来源。

- 评估方面：多 seed 重复实验，报告均值 \pm 方差提升可信度。

复现方式

- 训练/生成/评估(同一次 run 跑完三种模型并出图):
`nohup bash train.sh > train.log
2>&1 &`
- 单独重画本次 run 的汇总图:
`conda run -n qml_gpu python scripts/mnist01_plot_run_reports.py --exp data/mnist01_run_--out data/mnist01_run_2025-12-31_000116/plots`