1 Diffusion Model

基于 LDMLR (CVPRWP 2024 [Han et al.(2024)]) 的改进 1: 算法（方案）改进。

不足：依赖编码器质量，伪特征的多样性和可控性有限。

1. 端到端可训练编码-生成框架：不再是先独立训练编码器，而是将编码器与潜空间扩散生成器联合优化，使伪特征更贴合分类目标。

   原流程的问题是：Stage 2 的扩散模型是在“固定”的潜特征 $z = \mathcal{E}(x)$ 上学习分布。但当我们在 Stage 3 再微调分类器 $\mathcal{G}$ 甚至回头改动 $\mathcal{E}$ 时，这样一来，Stage 2 生成的伪特征分布，可能与最新版本的编码器提取的特征分布不一致。

   训练时，分类器 $\mathcal{G}$ 被迫在旧的伪特征和新的真实特征之间学习；但测试时只有新的真实特征。训练和测试分布不一致，泛化性能应该会下降。同时，对分类器来说，同一个类变成了两团相隔的点云。类内不紧致，决策边界被拉扯，模型难以学习清晰的 margin。分类器会被大量过时的伪特征牵着走。在训练时可能表现不错（因为它学会区分“旧猫 vs 狗”），但在真实测试分布上（只有“新猫 vs 狗”）性能掉下去。

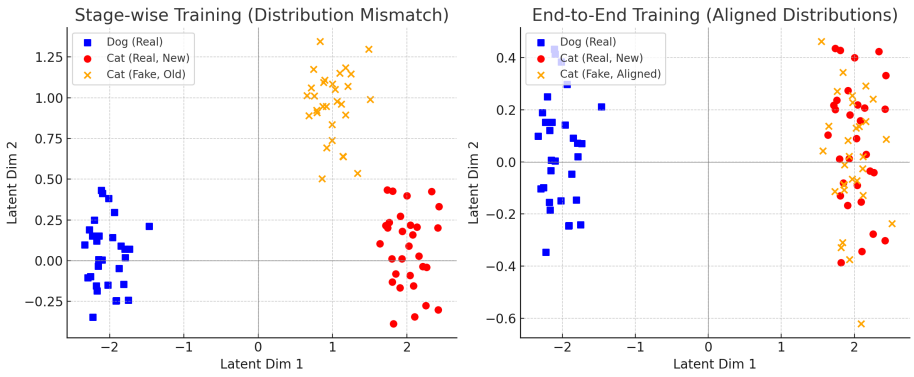

图 1: Comparison

举例：

- 假设有“猫 vs 狗”两类数据：原流程：Stage 1 学到的猫特征是某个分布，但后来微调时 $\mathcal{E}$ 又稍微移动了；此时扩散模型还在生成“老版本”的猫特征 → 伪特征和真实特征对不上。端到端：当

$\mathcal{E}$ 更新时，扩散模型同时更新；它生成的猫伪特征会自动迁移到新的分布位置，还会被引导到“猫”的分类边界附近，既真实又有判别性。

拟定方案：单次训练迭代中的四个并行阶段

Stage 1:

输入：从数据集中取出一批图像和标签 (x, y)

编码：图像 x 通过编码器 $\mathcal{E}$ 转化到潜空间中的特征向量 z：$z = \mathcal{E}(x; \theta_{\mathcal{E}})$。

$f(x) = \mathcal{G}(z)$, 计算真实数据分类损失 $\mathcal{L}_{real} = -\mathbb{E}_{x,y} \log f(x)_y$。

Stage 2 （扩散模型）：

前向加噪：和原文（9）和（10）相同，在任意时间步 t，带噪特征 z_t 的生成为：$q(z_t|z_{t-1}) := \mathcal{N}(z_t; \sqrt{\alpha_t} z_{t-1}, (1-\alpha_t)\mathbf{I})$ ($\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$), $\mathcal{L}_{LDM} = \mathbb{E}[||\epsilon - \epsilon_\theta(z_t, t, y)||_2^2]$。

语义一致性：首先，$z_t = \sqrt{\bar{\alpha}_t} z + \sqrt{1-\bar{\alpha}_t}\epsilon$, 因此知道了带噪特征 z_t 和噪声 ϵ, 我们可得 $z = \frac{z_t - \sqrt{1-\bar{\alpha}_t}\epsilon}{\sqrt{\bar{\alpha}_t}}$。在模型的训练和生成过程中，我们并不知道当时加入的真实噪声 ϵ 是什么，但是我们有一个预测它的模型 $\epsilon_\theta(z_t, t, y)$。因此，我们可以得到一个“估计出的原始特征”：$\tilde{z}_0(z_t, t, y) = \frac{z_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(z_t, t, y)}{\sqrt{\bar{\alpha}_t}}$。

- 原型拉拢：确保生成的特征在“位置”上是正确的，即与所属类别的中心保持一致。

  若我们关心的类 y（比如“猫”）有 N_y 个样本，则 $\mu_y = \frac{1}{N_y} \sum_{i:y_i=y} \mathcal{E}(x_i)$ (计算所有属于类别 y 的样本特征的平均向量)。μ_y 是预先维护的、每个类别 y 的真实特征的平均（即类原型或中心）。EMA 以真样本为主，伪样本可选低权重（如 0.3）。它通过指数移动平均（EMA）的方式不断更新，代表了该类最核心的语义位置。

  在每一个 mini-batch 上，$\mu_y^t = (1-\lambda)\mu_y^{t-1} + \lambda \hat{\mu}_y^{batch}$ ($\hat{\mu}_y^{batch}$ 是该 batch 下的均值，λ 是更新率)。

  计算 $\mathcal{L}_{\text{proto}} = \mathbb{E}_{z,y,\epsilon,t}\left[\|\tilde{z}_0(z_t, t, y) - \mu_y^t\|_2^2\right]$。整个损失函数的目标，就是通过最小化这个距离，将 $\tilde{z}_0$ “拉拢”到其所属类别的中心。

- 形状匹配: 防止“模式坍缩” (Mode Collapse)。仅有“原型拉拢”可能会导致模型只生成一模一样的、位于类中心的特征，丧失了

多样性。形状匹配旨在让生成特征的分布形状也与真实特征保持一致。(仅有“原型拉拢”模型可能会偷懒)。

(a) 协方差匹配: $\mathcal{L}_{cov} = \sum_y \|\hat{\Sigma}_y^{(\tilde{z}_0)} - \Sigma_y\|_F$。要求一批估算特征 $\tilde{z}_0$ 的协方差矩阵 $\hat{\Sigma}_y^{(\tilde{z}_0)}$，与该类真实特征的协方差矩阵 Σ_y 尽可能相似。

(b) 等半径约束: $\mathcal{L}_{rad} = \mathbb{E}\left[(\|\tilde{z}_0 - \mu_y\|_2 - r_y)^2\right]$ ($r_y \propto \sqrt{\mathrm{tr}(\sum_y)}$)。简化的替代方案，它不要求形状完全匹配，而是要求估算的特征 $\tilde{z}_0$ 到类中心的距离，平均来说应该等于一个预设的半径 r_y（这个半径 r_y 与真实特征的分布半径相关)。
“迹” (Trace) 的计算就是把矩阵主对角线上的所有元素相加。如果 $\mathrm{tr}(\sum_y)$ 很大,说明这个类的特征分布很广泛。如果 $\mathrm{tr}(\sum_y)$ 很小，说明这个类的特征非常集中和紧凑。设置合适的半径。

语义增强的扩散损失:$\mathcal{L}_{LDM}^{sem} = \mathcal{L}_{LDM} + \eta_p \mathcal{L}_{proto} + \eta_c \mathcal{L}_{cov}(\text{or} \quad (\eta_r \mathcal{L}_{rad}))$。

伪特征生成: 此阶段在每次迭代中都会执行，以生成用于计算 $\mathcal{L}_{ge}$ 的伪特征。和原文（11）和（12）相同, 反向采样生成 ($\hat{z}$)。当模型 ϵ_θ 在当前迭代中被定义后，我们采用与原文公式 (11) 相同的反向采样方法，从一个纯高斯噪声 $\hat{z}_T \sim \mathcal{N}(0, I)$ 开始，逐步生成一个干净的伪特征 $\hat{z}_0$。

每一步反向去噪的更新公式为: $\hat{z}_{t-1} = \sqrt{\bar{\alpha}_{t-1}} \left(\frac{\hat{z}_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\hat{z}_t, t, y)}{\sqrt{\bar{\alpha}_t}} \right) + \sigma_t \epsilon_t$。生成的伪特征 $\hat{z}_0$ 是即时 (on-the-fly) 使用的，并不会被收集成一个静态的数据集。它被立刻送入分类器 $\mathcal{G}$ 用于计算 $\mathcal{L}_{ge} = -\mathbb{E}_{\hat{z},y}[\log \mathcal{G}(\hat{z})_y]$，其梯度会立即反向传播以更新模型。因此，虽然概念上存在一个伪特征集合，但它在实现上是动态和瞬时的，而非像原文中那样是一个固定的中间产物。

端到端地总目标: $\mathcal{L}_{total} = \mathcal{L}_{real} + \lambda_{sem}\mathcal{L}_{LDM}^{sem} + \gamma_{ge}\mathcal{L}_{ge}$。最小化 $\mathcal{L}_{total}$，同时更新 $\{\mathcal{E}, \mathcal{G}, \theta\}$。

端到端训练时的一次迭代（运算顺序与梯度去向）

输入: 采样一批 $((x, y))$

1. 真实分支: $(z = \mathcal{E}(x))$，累积 $(\mathcal{L}_{real})$。
2. 扩散前向:可得得 z_t,累积 $\mathcal{L}_{LDM}$;计算 $\tilde{z}_0$ 并累积语义项 ($\mathcal{L}_proto$,$\mathcal{L}_cov/\mathcal{L}_{cov}$, 得到 $\mathcal{L}_{LDM}^{\text{sem}}$。

3. 反向采样：生成 $\hat{z}$，累积 $\mathcal{L}_{ge}$。

4. 合成并更新：最小化 $\mathcal{L}_{total}$，同时更新 $\{\mathcal{E}, \mathcal{G}, \theta\}$。

三种训练策略:

A. 同步联合

每个小批次同时计算 (1)(2)，并周期性（例如每 k 步）触发一次快速反采样得到 $\hat{z}$ 以训练 (3)(4)。损失：直接用 L_{total}。优点：耦合最紧，伪分布与最新 (z) 同步；缺点：采样频繁 → 计算重。实现要点：采样步数用 DDIM 减步（与原文一致，DDIM 用于加速）；语义项权重 warm-up，避免早期把 $\tilde{z}_0$ 拉到原型。

B. 交替更新（稳定实用）

步骤 A（训扩散）：冻结 $\mathcal{E}, \mathcal{G}$，用当前 $z = \mathcal{E}(x)$ 训练 ϵ_θ：最小化 $\mathcal{L}_{LDM}^{\text{sem}}$。
步骤 B（训分类）：冻结 θ，用真实 z 和即时/缓存的 $\hat{z}$ 训练 $\mathcal{E}, \mathcal{G}$。
循环往复：例如 A:B=1:1 或 2:1。性质：仍保留 "端到端" 耦合（A 步用最新的 z 训练扩散；B 步用最新 $\hat{z}$ 训练分类）；抑制目标漂移，训练更稳。

C. 渐进式联合（warm-up → 联合）warm-up：先短跑只开 (1) 训练 $\mathcal{E}, \mathcal{G}$。逐渐开：(2) $\mathcal{L}_{LDM}^{\text{sem}}$ → (3) ($\mathcal{L}_{ge}$) → (4)；比例调度：伪样本比例由 10% → 30%。

基于 LDMLR (CVPRWP 2024 [Han et al.(2024)]) 的改进 2：结合差分隐私和联邦学习的应用。

扩散模型结合差分隐私的论文参考 [Wang et al.(2024)], 扩散模型结合扩散模型的论文参考 [Peng et al.(2025)]。

dp-promise 的核心技术：利用扩散模型前向过程中的噪声本身来承担部分差分隐私保证，从而减少额外注入的 DP 噪声。

具体设计为两阶段训练:

- Phase I (Non-private Training): 在扩散时间步 [S, T] 内训练。此时图像已被注入较强的高斯噪声，作者理论上证明这些噪声等价于高斯机制，能够直接提供 GDP（Gaussian DP）保证。因此 Phase I 无需再注入额外 DP 噪声，只需进行普通训练。

- Phase II (Private Training): 在较早的扩散步 [1, S−1]，噪声较少，隐私保障不足。此时采用 DP-SGD（梯度裁剪 + 高斯噪声），确保 DP。引入子采样（Poisson Sub-sampling）进一步增强隐私放大。

FL for DM 的核心技术:

- 提出 FEDDDPM：在标准联邦流程（随机抽取客户端、本地多步 SGD、服务器聚合）后，服务器利用合成的辅助数据集对全局模型做一次 "方向校正"，用以抵消非 IID 导致的偏置更新；并给出非凸情形的收敛分析。
- 提出 FEDDDPM+：当检测到全局学习"边际收益"变小（通过 QUICK-TEST）时，一次性用辅助数据集在服务器侧对全局模型做多步微调然后提前终止训练，以显著降低总训练开销，同时保持与 FEDDDPM 接近的性能。

沿用 LDMLR 的思想：始终在特征/潜空间工作，只生成伪特征（带标签），不生成图像。

DP 就是在扩散模型中沿用 dp-promise。联邦扩散沿用 FL for DM。

需要讨论的点：

1. Warm-up 阶段是上传 DP 的扩散模型给服务器，还是直接本地先生成一部分伪特征上传给服务器。这个问题的点在于通信开销上上传模型和上传伪特征哪个更优（比如 UNet 和 200 个 feature 哪个开销大）.
2. FL for DM 方案是生成辅助数据就不再更新了，校准一直用这个数据集。我们考虑的是，客户端训练完上传模型，服务器聚合后进行校准，我们补充一个对这个校准模型的评分，如果评分在我们设定的范围内，认为是比较好的更新，那么用这个校准模型再生成一定比例的数据补充或者替换原始的数据。

   直觉上我们这是在指标显著改善时触发生成；不像固定库那样 "一次大生成" 后可能长期失配。长期看，生成样本/收益比更高。保留高质量样本、淘汰低质量或过时样本，库的性价比随轮次提高。
3. 生成数据的阶段都可以引入 AID/OOD [Shao et al.(2024)], 用来判断和保留更好用的伪特征。

4. 这个 idea 主要是在 [Peng et al.(2025)] 的基础上补了 DP, 校准所使用的的辅助数据修改为渐进式生成方法。实验上我们可以做得充分一些。FL-DM 考虑长尾数据。DP 考虑 MIA 和 Auditing 的工作。

参考文献

[Han et al.(2024)] Pengxiao Han, Changkun Ye, Jieming Zhou, Jing Zhang, Jie Hong, and Xuesong Li. 2024. Latent-based diffusion model for long-tailed recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2639–2648.

[Peng et al.(2025)] Zihao Peng, Xijun Wang, Shengbo Chen, Hong Rao, Cong Shen, and Jinpeng Jiang. 2025. Federated Learning for Diffusion Models. *IEEE Transactions on Cognitive Communications and Networking* (2025).

[Shao et al.(2024)] Jie Shao, Ke Zhu, Hanxiao Zhang, and Jianxin Wu. 2024. DiffuLT: Diffusion for Long-tail Recognition Without External Knowledge. *Advances in Neural Information Processing Systems* 37 (2024), 123007–123031.

[Wang et al.(2024)] Haichen Wang, Shuchao Pang, Zhigang Lu, Yihang Rao, Yongbin Zhou, and Minhui Xue. 2024. dp-promise: Differentially private diffusion probabilistic models for image synthesis. In *33rd USENIX Security Symposium (USENIX Security 24)*. 1063–1080.