

Machine Learning Assignment 11

Unanswered questions

- When performing linear interpolation between two latent noise samples $x_T^{(1)}$ and $x_T^{(2)}$, and then running the backward ODE (or probability flow ODE) to generate an image, should we always expect the resulting image to visually resemble a "mixture" of the two source images? For instance, if $x_T^{(1)}$ and $x_T^{(2)}$ correspond to a dog image and a cat image respectively, is it possible that the generated interpolated image does not resemble either a dog or a cat?
 - 可以使用 guidance 讓插值仍然保持語義：[Ho & Salimans, 2022 - Classifier-Free Diffusion Guidance](https://arxiv.org/abs/2207.12598) (<https://arxiv.org/abs/2207.12598>).
 - 或者使用 latent interpolation：[Song et al., 2021 - Denoising Diffusion Implicit Models](https://arxiv.org/abs/2010.02502) (<https://arxiv.org/abs/2010.02502>).

Toy model in final project

參考在上一個作業中，對於考古領域關於這個 Toy model 的描述：[連結](https://hackmd.io/dlfiWSANSd2-9zmGIzIRQ#Question-2)
(<https://hackmd.io/dlfiWSANSd2-9zmGIzIRQ#Question-2>).

我想關注在其中「少樣本學習」這個方法上，因為在古物上通常相同類別的樣本數可能不多，因此我認為這是最核心的一項技能。

1. 問題設計

給定一個類別集合 \mathcal{C} ，其中包含至少 10 種的幾何的紋飾類別（例如：條紋、點陣、格狀、同心圓，波浪線等等）

每次測試隨機抽取 N 個類別，每個類別提供 K 張支持樣本與若干查詢樣本，我們希望達到「用支持樣本集合建立分類器，並對查詢樣本集合做出分類」

其中，這些樣本影像與查詢影像皆由我們自行合成得到。

2. 模型與方法

在每一次測試內做以下的事：

1. 先從 \mathcal{C} 內抽 N 個類別

2. 從這些類別內抽 K 張支援樣本，形成 support set S
3. 再從這些類別內各抽 M 張查詢樣本，形成 query set Q
4. 當模型看到 S ，根據 S 構造一個分類規則，並且用這個規則分類 Q 裡的每一張圖

模型的訓練方法：

1. 先用一個 encoder (f_θ) 把每張影像轉換成特徵向量： $z = f_\theta(x)$
2. 對每個類別，所有該類的支持特徵向量取平均得到中心：

$$c_k = \frac{1}{|S_k|} \sum_{x_i: y_i=k} f_\theta(x_i)$$

3. 對要分類的查詢樣本 $x^{(q)}$ ，算它的特徵向量 $z^{(q)} = f_\theta(x^{(q)})$ ，看它與哪個中心的距離最近。

4. 詳細上我們以 $P(y = k|x^{(q)}) = \frac{\exp(-d(z^{(q)}, c_k))}{\sum_j \exp(-d(z^{(q)}, c_j))}$ 來衡量距離，其中我們

$$\text{採餘弦距離：} d(a, b) = 1 - \frac{a^T b}{\|a\| \|b\|}$$

每次測試的 Loss function 定義如下：

$$\mathcal{L}_{episode} = - \sum_{(x^{(q)}, y^{(q)}) \in Q} \log P(y^{(q)} | x^{(q)}, S)$$