

## 進捗報告

### 1 テストデータについて

テストデータとして用いている鳥画像 64 枚とそのキャプションをまとめたテキストファイルを github の test フォルダにアップロードしました。

画像生成のサンプリングの際には、この 64 文を入力として生成させており、この 64 文内に未知語は含まれないように設定しています。

### 2 生成画像に対するスコアについて

#### 2.1 Inception Score

Inception Score では生成画像の「良さ」を以下 2 つの観点から測っている。

1. 画像識別器が識別しやすい
2. 物体クラスのバリエーションが豊富

$x_i$  を画像 ( $i$  は画像のインデックスで総数は  $N$  とする),  $y$  をラベルとすると, 上記 1.. は条件付き分布  $p(y|x_i)$  のエントロピーが小さくなることを意味する。一方で 2.. は  $p(y|x_i)$  を  $i$  について積分した周辺分布  $p(y)$  のエントロピーが大きくなることを意味する。そこで、両分布間の距離を測るために KL ダイバージェンスを用い、これを  $i$  について平均化したうえで exp を取ったものが Inception Score である。Inception と呼んでいるのは、画像識別器として ImageNet で学習済みの InceptionV3 モデルを用いているため。

$$\exp\left(\frac{1}{N} \sum_i \text{KL}(p(y|x_i)||p(y))\right) \quad (1)$$

2 つの分布間の距離が大きいほど良い画像生成ができていると言えるため、Inception Score が大きいほど性能が高いということになる。

#### 2.2 FID

実画像の分布が考慮されていないという Inception Score の欠点を改善するため、実画像と生成画像の分布間の距離を測っている。Inception モデルから得られる特徴ベクトル (最後の pooling 層の出力) が多変量正規分布に従うと仮定し、実画像の分布と生成画像の分布間の距離を Fréchet 距離 (Wasserstein-2 距離) により求める。特徴ベクトルの平均と共分散行列が実画像についてそれぞれ  $\mu_r$ ,  $\Sigma_r$ , 生成画像について  $\mu_g$ ,  $\Sigma_g$  と得られているとすると、FID は次式で定義される。

$$\|\mu_r - \mu_g\|_2^2 + \text{tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{\frac{1}{2}}) \quad (2)$$

## 2.3 SWD

ワッサーズタイン計量（ワッサーズタインけいりょう, 英: Wasserstein metric）とは, 与えられた距離空間  $M$  上の確率分布の間に定義される距離関数である.

直感的な説明としては, 各分布を  $M$  上に堆積した土の単位量と見なすとき, ワッサーズタイン計量とは一つの堆積を別の物へと移すときにかかる最小のコストである. そのようなコストは, 移されるべき土の量に移す距離を掛けた値であるとされる. (Wikipedia より)

SWD とはラプラシアンピラミッドの各レベルから  $7 \times 7$  のパッチを取り出し, それぞれを 1 次元に変換してワッサーズタイン距離を計算する.

Inception Score と違い, Inception モデルのドメインや識別能に依存しない.

## 2.4 参考文献

How good is my GAN? Sliced Wasserstein Distance for Learning Gaussian Mixture Models