

【要点まとめ】

VAEの派生技術で、以下のような違いがある。

- VAE：潜在変数 z がガウス分布に従うベクトルになるよう学習を行う
- VQ_VAE：潜在変数が離散的な数値となるように学習を行う

離散的な潜在変数を用いるメリットは以下の2点が挙げられている。

- 離散変数のほうがデータの特徴を捉えるのに適している → 人間が画像を見て、それを「猫」「車」など簡潔な表現をすることができる。言語が離散的であるように、離散的表現のほうが特徴を捉えやすい。
- posterior collapseを防ぐことができる → VAEでは強力なデコーダー（Pixel CNNなど）と組み合わせた時に、潜在変数がデータの特徴をうまく捉えることが出来なくなるposterior collapseと呼ばれる現象が確認されている。特に画像生成では輪郭がぼやけるなどの問題に発展する。

【実装演習】

ソースコードが無いため省略します

【実装演習考察】

ソースコードが無いため省略します

【自己学習】

※参考資料：<https://compvis.github.io/taming-transformers/>

VQ_VAEはVAEを更に安定化させた素晴らしい技術ですが、まだまだ拡張の可能性を秘めており
例えばそこにtransformer→CNN decoderと入力することで高解像度画像を生成するようなVQGANなど
ある完成されたように見える技術も、工夫や組み合わせ次第でまだまだ拡張が可能であることを学びました。