進捗報告

1　今週やったこと

- データセットの許可メール
- 使用するモデルの検討.

2　メールの結果

サイトの運営者より使用許可を頂きました．大阪公立大学高専 (美術担当) の方とのことです.
得られた許可：画像・文章・曲・動画

3　データセットの整形

イラストに含まれる枠が学習の妨げになりそうなので，枠なしバージョンも生成した．若干線のアンチエイリアスが変化したがそこまで大きな問題ではないと思われる.

図 1: 変換前

図 2: 変換後

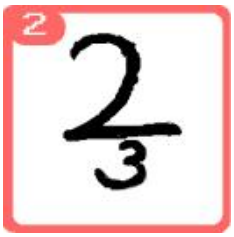

図 3: 変換前

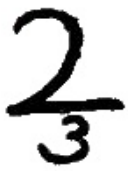

図 4: 変換後

4　問題設定

このデータセットを利用する問題をいくつか考えた.

- 順序推定
- 絵描き歌のクラス推定

- 最終状態の予測

順序推定はいちばん簡単な問題として，画像や文章を様々な順序で入力し正順かどうかを判定する問題.

クラス推定は単純に画像と文章の列を入力し描かれたもののクラスを推定する問題であるが，このデータセットは 1 クラスにつき 1 つのデータしか存在しないためこれをテストデータに使うと 0-shot での推定となりかなり困難であると思われる.

最終状態の予測は最後以外の画像や文章を入力し，最後の画像や文章を予測させる問題.

4.1 モデルの調査

どの問題を解くにしても必要な，画像の時系列データ（動画）を入力とするモデルを調べたところ，似たモデルが同時期に複数提案されていることが分かったため，まず Video Vision Transformer (ViViT) を使用できるようにした．実装と事前学習パラメータも使用できることを確認した.

次に必要なのは自前のデータセットを ViViT の入力形式に変換することである.

5 予定

- ViViT の予備実験