**Signate日本舞踊の画像解析チャレンジ**

**コンペの概要**

開催期間 : 2024/9/28 ~2024/11/1

URL : <https://signate.jp/competitions/1506>

**目的**

日本舞踊を踊っている人物の画像データセットを使い、未知のデータに対しても「扇子をもっている / 扇子を持っていない」を機械学習を使い正しく分類することを目指す。

**使用データ**

訓練データ：扇子を所持している人 / していない人の画像それぞれ 34472/32974件

ポーズをとる女性たち

中程度の精度で自動的に生成された説明検証データ：画像34299件

図1. 使用画像の例

左が扇子を所持しており、右が所持していない画像。

大きさは360×640となっている。

**作業環境**

Jupyter Lab, Google Colab Pro

**評価指標**

コンテストではF1Scoreによって予測精度の良しあしが決定される。ただし自分が使用した学習モデルは、クロスエントロピーが最小になるように学習している。

**テーブル が含まれている画像

自動的に生成された説明**

図２. 評価指標の説明

F1ScoreはPrecisionとRecallの調和平均となっている。

**分析の概要と結果**

**分析の概要**

主な方針として、2種類のモデルを使用し分析を進めた。

1. EfficientNet

Kaggleに公開されたコードである

<https://www.kaggle.com/code/motono0223/isic-pytorch-training-baseline-image-only>

を参考に画像認識を進めた。

モデル採用理由は、計算が高速なためである。最高のF1Scoreは0.9968701であった。

②EVA-02

本コンテストの掲示板内の記事、<https://signate.jp/competitions/1506/discussions/submitcolab>

をそのまま使用した。学習設定を下表のように変更した(たくさん学習した)ところ、

最高のF1Scoreである0.9981033を得ることができたため、これを最終結果として提出した。

**テーブル

自動的に生成された説明**

表1. EVA-02の他人のコードと自分の分析の学習設定の比較。

変化させた学習設定だけ載せている。一番左の列の値を変化させた。

次に、各モデルで行った分析の詳細について説明する。

**EfficientNet**

ダイアグラム, QR コード

自動的に生成された説明

図3. EfficientNetのモデルの構造の図

まず画像データを訓練データと検証データに分割する。その後、訓練データでモデルを学習し、検証データは訓練したモデルで予測ラベルを算出する。

上の図は、今回使用したEfficientNetのモデルの構造の図である。モデル訓練時には事前に大規模な画像データセットで学習したパラメーターのモデルを使い、全結合層や活性化関数を新たに付け加えたあと、今回の日本舞踊の画像解析タスクに最適化するためにファインチューニングした。

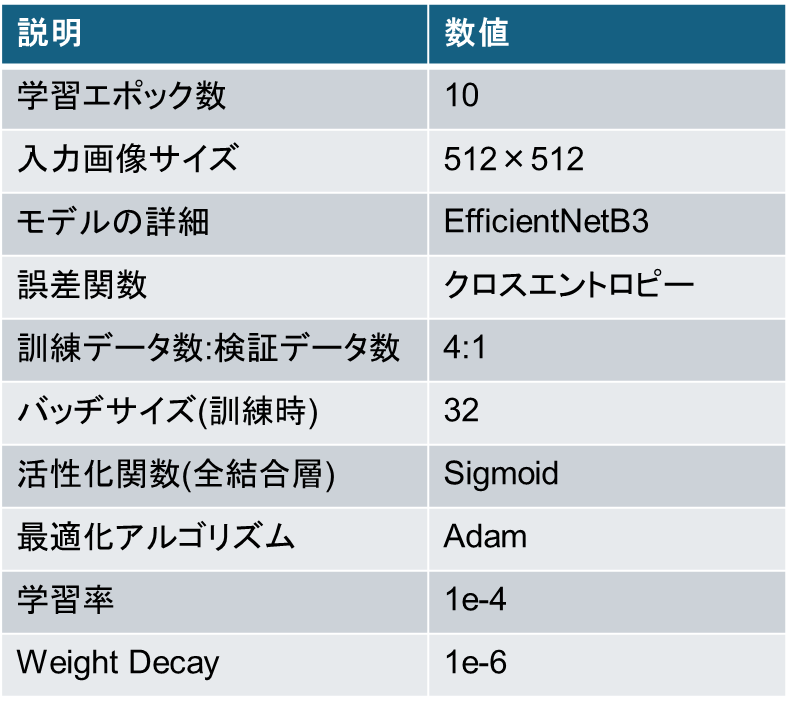
****

表2. EfficientNetで一番良い精度が出た時の学習設定

グラフ, 折れ線グラフ

自動的に生成された説明

図4. EfficientNetの訓練データと検証データのモデルの学習による誤差の推移

左の軸は誤差で、下の軸は学習エポック数を表す。青の曲線が訓練データの誤差で、オレンジの曲線が検証データの誤差となっている。

グラフ, 折れ線グラフ

自動的に生成された説明

図5. EfficientNetの訓練データと検証データのモデルの学習によるAUROCの推移

左の軸はAUROCで、下の軸は学習エポック数を表す。青の曲線が訓練データのAUROCで、オレンジの曲線が検証データのAUROCとなっている。

また、学習の設定は表2のようにし、得られた結果が図4と図5となる。

図4では学習epoch数が進むにつれ、訓練データの曲線と検証データの曲線を見ると誤差が下がっており、正しく学習できていることがわかる。

図5では、学習epoch数が進むにつれ、訓練データの曲線と検証データの曲線のAUROCが1に近づいており、モデルの分類の精度が高いことがわかる。

**EVA-02**

QR コード

自動的に生成された説明

図6. EVA-02ののモデルの構造の図

まず画像データを訓練データは、図のように層化K分割交差検証でK種類のデータ分割を行う。その後、K種類のFoldごとに出力確率値を平均し、その値から予測ラベルを算出する。

上の図は、今回使用したEVA-02のモデルの構造の図である。EfficientNetの場合と同様に、モデル訓練時には事前に大規模な画像データセットで学習したパラメーターのモデルを使い、全結合層や活性化関数を新たに付け加えたあと、今回の日本舞踊の画像解析タスクに最適化するためにファインチューニングした。

**テーブル

自動的に生成された説明**

表3. EVA-02の学習設定

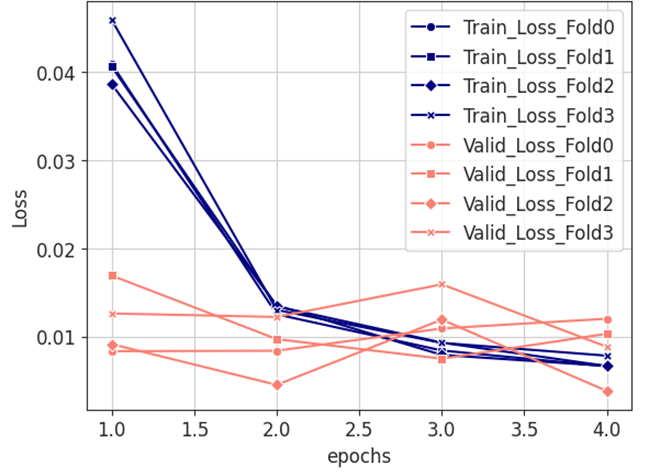


図7. EVA-02の各Fold毎の訓練データと検証データのモデルの学習による誤差の推移

左の軸は誤差で、下の軸は学習エポック数を表す。青の曲線が訓練データの誤差で、オレンジの曲線が検証データの誤差となっている。

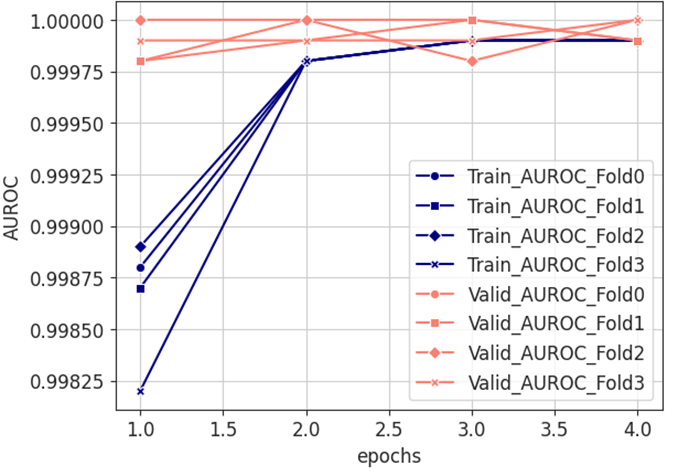
****

図8. EVA-02の各Fold毎の訓練データと検証データのモデルの学習によるAUROCの推移

左の軸はAUROCで、下の軸は学習エポック数を表す。青の曲線が訓練データのAUROCで、オレンジの曲線が検証データのAUROCとなっている。

また、学習の設定は表３のようにし、得られた結果が図7と図８となる。

図7ではどのFoldでも学習epoch数が進むにつれ、訓練データの曲線と検証データの曲線を見ると誤差が下がっており、正しく学習できていることがわかる。

図8では、どのFoldでも学習epoch数が進むにつれ、訓練データの曲線と検証データの曲線のAUROCが1に近づいており、モデルの分類の精度が高いことがわかる。

**まとめ**

・大規模データで学習済の画像認識モデルを使用し、再学習することで扇子を所持しているか否かの予測に取り組んだ。

・使用したEfficientNet, EVA-02双方で高い予測精度が得られた。