

知能情報実験 iii (データマイニング班)
顔画像に基づく美男美女と一般人との比較による特徴抽出

215706D : KIM HYUNWOO, 235221E : 山脇大輝,
235701B : 松田遼平, 235732B : 長瀬一生

提出日: 2025 年 7 月 18 日

目次

1	はじめに	1
1.1	実験の目的と達成目標（アプローチの全体像を含む）	1
1.2	意図していた実験計画との違い	1
1.3	実験の全体像	2
2	実験方法	2
2.1	実験目的	2
2.2	データセット構築・前処理	2
2.3	モデル選定	4
2.4	パラメータ調整	4
3	実験	6
3.1	役割分担について	6
3.2	用意したデータセット画像	6
4	実験結果	6
4.1	Custom Model の実験結果	6
4.2	Custom Model の結果の透明性確保（仮説立て）	7
5	仮説検証 1：新規学習	8
5.1	ResNet 18	8
5.2	ResNet 34	8
5.3	ResNet 50	9
5.4	EfficientNet_b0	10
5.5	Custom Model と新規学習の比較	10
6	仮説検証 2：Grad-Cam を用いたモデルの信頼性の検証	11
6.1	検証の概要	11
6.2	検証結果	11
6.3	白背景画像を用いた Grad-Cam の改善	14
6.4	GradCam 適用結果を踏まえた全体の考察	17
7	まとめ	18
8	反省・今後の課題	18
8.1	反省点	18
8.2	時間の都合上省いた項目	19
A	付録 A GradCam を用いた分析の詳細	21
A.1	GradCam を用いた分析の実行結果	21
A.2	考察	23
B	付録 B パラメータ調整によるモデル精度向上実験	23
B.1	概要	23

B.2	実験結果	24
B.3	考察	24

概要

本研究では、FairFace データセットを一般人データとして用い、独自に構築した美男美女データセットと併せて学習を行うことで、人物の外見に基づく分類を行う機械学習システムを構築した。さらに、Grad-CAM を用いて各モデルの判断根拠を可視化し、信頼性の評価を試みた。追加学習を行った FairFace モデルにおいては、顔領域を適切に着目しながら一般人・美男美女の分類精度 100% を達成したが、新規学習した ResNet18, ResNet34, ResNet50, EfficientNet_b0 については、分類精度 95% を超える高い分類精度を示したものの顔ではなく画像の背景を見ていることが確認された。ここからデータセット構築時における内在的バイアスが含まれていることがわかった。深層学習モデルが高精度を示す場合でも、その判断根拠が実験目的と合致しているとは限らないという、モデルの信頼性における重要な課題を浮き彫りにした。また、Custom Model では GradCam の結果より目元と口元を見て一般人と美男美女を分類していることから、これらが一般人と美男美女の差異であると結論付けた。

1 はじめに

1.1 実験の目的と達成目標（アプローチの全体像を含む）

本グループでは「顔画像データを用いて、一般的に“美男美女”と称される有名人と一般人を識別・分類し、特徴差を明らかにすること」をテーマとして設定した。

人の美しさや魅力は主観的に表されることが多く、機械学習を用いて客観的に分析することを試みる。FairFace のモデル構築の際に用いられた画像群と独自に美男美女の画像から構築したデータセットを用いて、画像認識のための深層学習モデルである ResNet18, ResNet34, ResNet50 と EfficientNet_b0 を用いて、美男美女（有名人）と称される顔写真と一般人の写真を識別・分類するモデルの構築を問題として設定した。最終的には、GradCam を用いて実行結果を可視化し、美男美女と一般人の顔写真を区別する上で重要な特徴を定量的に見る。

1.2 意図していた実験計画との違い

当初の調査段階では、FairFace がデータセットのみを公開していると認識していたが、実際には高精度な学習済みモデル（ResNet-34 ベース）も提供されていることが後に判明した。FairFace と同様のデータセット構築とモデル学習を一から行うことは、実験としての新規性が少ない。そこで、方針を転換し FairFace の学習済みモデルに対して、独自に収集した美男美女・一般人のラベル付をした画像を学習させることで追加学習を行うアプローチを行うことにした。

1.3 実験の全体像

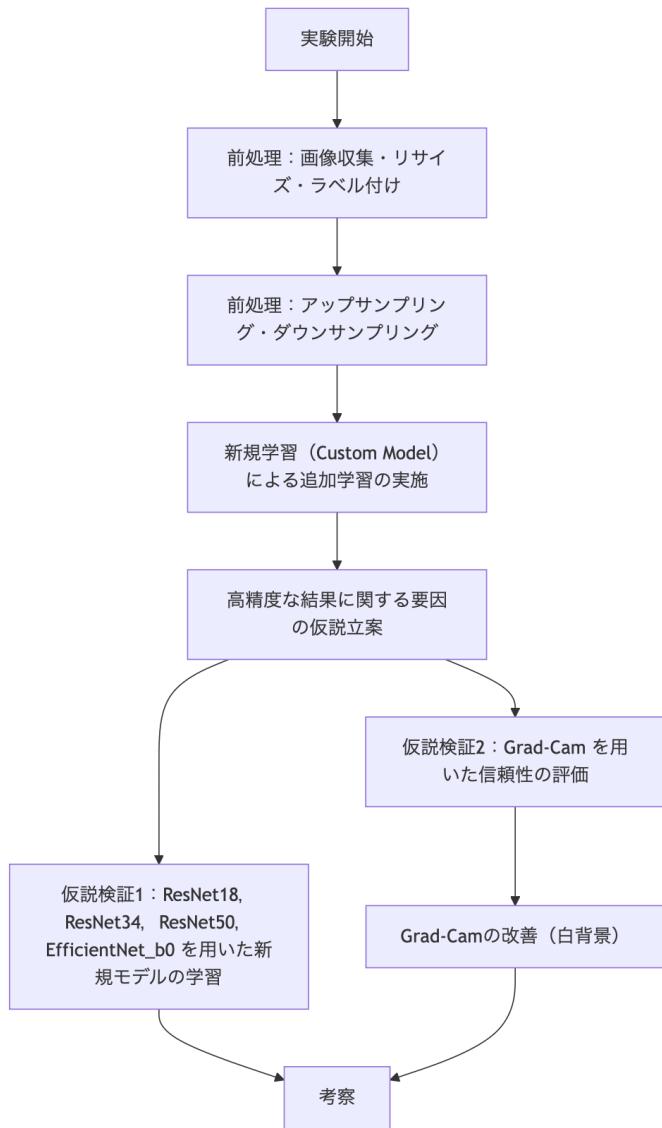


図 1: 実験のフローチャート

2 実験方法

2.1 実験目的

世界で美男美女と呼ばれる人の顔写真と一般人の顔写真を分類するモデルを作り、美男美女と一般人の分類を行う。

2.2 データセット構築・前処理

美男美女と一般人を分類するモデルを構築するために、2つのデータセットを準備し、前処理を行い、モデルの学習・評価を行う。前処理のフローチャートを図2に示す。

2.2.1 FairFace とは

本実験では FairFace というモデルを用いて、分類評価を行うと同時に、独自構築した分類モデルも作成した。ここで、FairFace というモデルについて概要を説明すると、従来の多くの顔画像データセットでは、特定の人種（白人）や性別にデータの偏りが見られる傾向にある。FairFace では、7つの人種グループ：白人、黒人、インド人、東アジア人、東南アジア人、中東人、ラテンアメリカ人でデータ数を均一して、人種差の少ないデータセット構築を行った。この論文の実験では、FairFace データセットを含む様々なデータセットでモデルを訓練し、その後に訓練されたモデルを用いて新しいデータセットでの汎化性能を評価しており、機械学習モデルの特定の人種・性別への偏り・バイアスの軽減を実現した顔画像データセット及びモデルである。構築されたデータセットについては、主に大規模なパブリックデータセット yahoo YFCC100M (Flickr 画像) から、意図的に人種バランスを考慮したサンプリング手法を用いて収集され、加えて Twitter やオンライン新聞からの画像も含まれている。YFCC100M 全体からランダムに顔画像をサンプリングし、各国の人口構成を推定し、データセットが特定（白人）の人種に偏らないように画像数を調節した。これにより他の人種を過小評価するデータセットのバイアスを軽減している。

2.2.2 定義

本実験では、以下のように一般人と美男美女の定義を行う。

- 一般人：FairFace のデータセットで作成されたデータセット。
- 美男美女：2024 年度世界の美男美女ランキング上位 top50 を美男美女として定義する。

2.2.3 データセット構築

- 一般人のデータセット：FairFace のデータセットを使用
- 美男美女のデータセット：独自構築を行った。

(1) 美男美女の基準を決定

文献 [2] や文献 [3] より 2024 年度美男美女ランキング top50 を美男美女として扱う

(2) データの収集

対象者の画像は、「2024 年度美男美女ランキング」top50 の人名から bing 検索エンジンを用いて web スクレイピングした。

- * 人名でキーワード検索を行い、人物以外のもの（例えば似顔絵やファンガッズなど）を除外した。（いわゆる宣材写真のみ収集したわけではない。）
- * ソースコードを以下の GitHub にて公開する。
- * https://github.com/e235221/info3dm_racial_classification

2.2.4 前処理

- リサイズ
 - 収集したデータの顔部分だけ切り取り、サイズを 300×300 とし、FairFace のデータセットと合わせて調整する。
- 正面・側面の判定
 - 顔の向きが学習に与える影響を排除するために、HopeNet を用いて正面を向いている画像のみ抽出した。側面を向いている画像はデータセットから除外している。
 - HopeNet とは、1枚の顔写真から Yaw(左右の向き), Pitch (上下の向き), Roll (傾き) を求め、その人がどの方向を向いているのかを推定する深層学習モデルである。感情認識などに用いられて

いる。

- アップサンプリング・ダウンサンプリング
 - アップサンプリング：データ数の少ない美男美女データの各画像を左右反転させることでデータ数を 2 倍に増やした。
 - ダウンサンプリング：データ数の多い美男美女データセットからランダムにデータを削除することで、美男美女データセットと一般人データセットのデータ件数を調節した。
 - 画像数については、train で 31000 枚ずつ、test で 7800 枚ずつの画像を用意した。
- 美男美女・一般人データセットに対して、ラベル付け
 - web スクレイピングしで収集した画像データに対して、FairFace で用意されているラベル付けと同様に美男美女データセットに対してもラベル付けを行い、それに加えて一般人か美男美女かを判別するためにそれぞれ 0 と 1 を付与した。
 - 新規作成した美男美女データセットについては、手作業で検索し、国籍・誕生日のラベルを付与した。
 - `file_name, age, gender, race, 0/1` の 5 種類を列名として csv を作成した。`file_name` に指定された path で画像を読み込み、学習を行う。
- ファイルのリネーム

2.3 モデル選定

FairFace が提供する学習済みモデルを本研究のタスクに合わせてカスタマイズした。これには ResNet34 が使用されていたため、そのまま ResNet34 を用いる。詳しくは文献 [1] を参照されたい。

(後述するが、この FairFace のカスタマイズモデルは精度が 100% になったため、別アプローチとして本実験では FairFace を用いず新規に ResNet18, ResNet34, ResNet50 も用いた。パラメータ調整はデフォルトのままである。)

2.4 パラメータ調整

本実験では FairFace のソースコードをパラメータ調整せずに使用した。GitHub に公開されている FairFace のソースコードでは、ハイパープラメータが不明であるが、文献 [1] より学習率 0.0001 の Optimizer に ADAM 最適化が用いられている。ただしこれらのハイパープラメータが選ばれた理由は GitHub や論文を読んでも不明である。

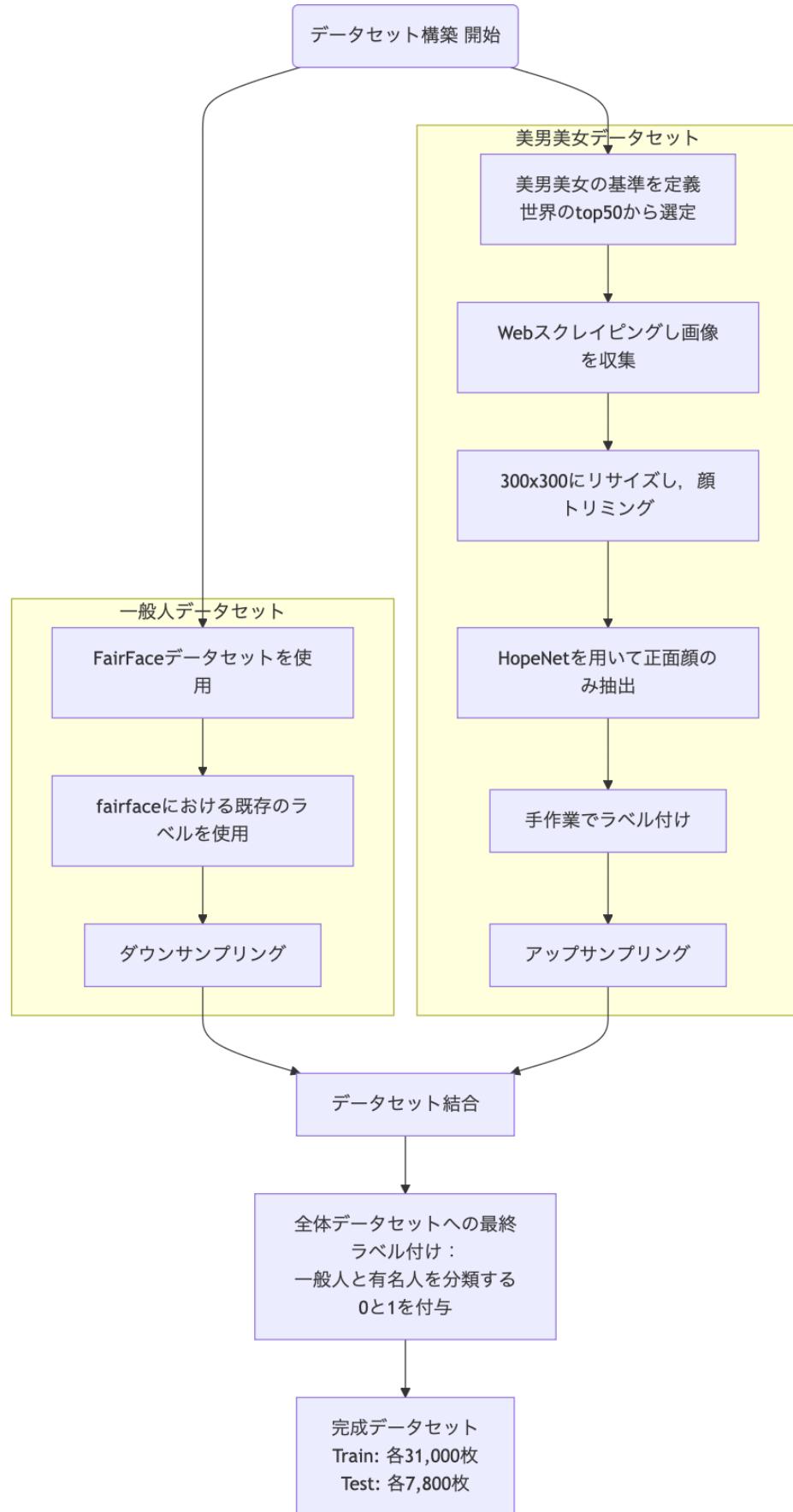


図 2: 前処理のフローチャート

3 実験

3.1 役割分担について

- 215706D : KIM HYUNWOO
 - 正面画像抽出・ラベル付・複数の ResNet と EfficientNet での実行・Grad Cam での検証・考察・プレゼンテーション作成
- 235221E : 山脇大輝
 - FairFace の調査, カスタマイズ, Web スクレイピング・リサイズ・トリミング・ラベル付・Grad Cam での検証・考察・レポート作成
- 235701B : 松田遼平
 - GitHub の操作補助
- 235732B : 長瀬一生
 - アップサンプリング・ダウンサンプリングコード作・背景のマスキング実施

3.2 用意したデータセット画像

レポートの可読性を向上させるために、図 3, 図 4 に用意したデータセットの画像の一部を示す。

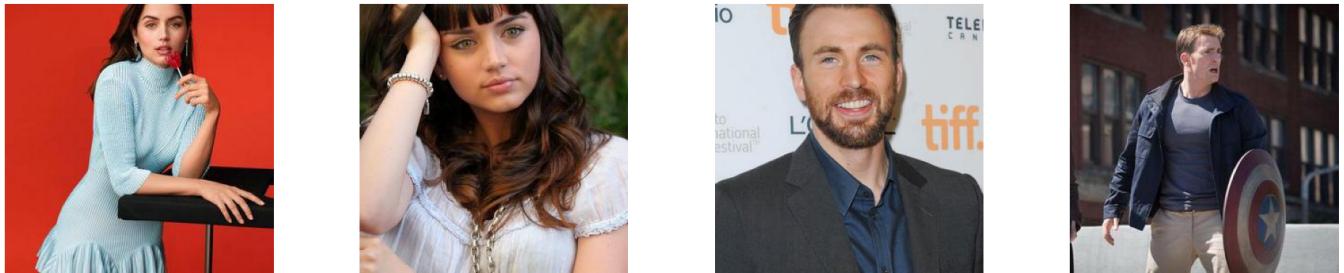


図 3: 美男美女データセットの画像例

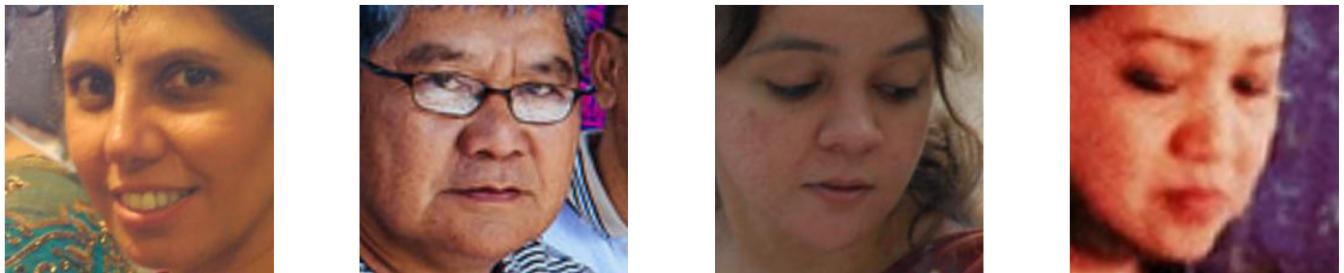


図 4: 一般人データセットの画像例

4 実験結果

4.1 Custom Model の実験結果

FairFace が提供する学習済みモデルを本研究のタスクに合わせてカスタマイズした ResNet34 を用いた実行結果を表 1 に示す。

表 1: Custom Model

Epoch	Train Accuracy	Validation Accuracy
1	99.48	100.00
2	100.00	99.99
3	100.00	99.99
4	99.97	100.00
5	100.00	100.00
6	100.00	100.00
7	100.00	100.00
8	100.00	100.00
9	100.00	100.00
10	100.00	100.00

Custom Model の考察：

学習済みモデルである FairFace に追加学習を施した Custom Model は、エポック 1 から train の精度 99.48%， val の精度 100.00% と極めて高い精度を示した。エポック 2 以降も train, val の精度はほぼ 100% に達しており、この高精度が過学習によるものなのか、あるいは学習済みモデルや追加学習が原因となったものなのか、そもそも本モデルが画像のどの部分に着目して学習を行ったのかについても明確ではなく、高精度の要因の切り分けは困難である。さらに学習済みモデルについて GitHub や論文を参照しても詳細な情報は記載されておらず、内部の動作はブラックボックス化しており、原因の特定は困難である。

4.2 Custom Model の結果の透明性確保（仮説立て）

前章で述べた過学習の可能性については、表 1 に示す結果から典型的な過学習の兆候は観察されず、過学習が生じているとは考えにくい。

以下に箇条書き形式で、なぜ表 1 のような高精度が実現できている原因について分析を行う。

- **要因 1. FairFace の学習済みモデルが持つ広い表現能力：**

FairFace の学習済みモデルが持つ広い表現能力（性別・人種・年齢分類）に対し、本タスク（美男美女/一般人分類）が比較的容易な問題であったため、追加学習によって早期に分離境界を発見でき、このような高精度になったと考えられる。

- **要因 2. パラメータ調整の違い：**

FairFace モデルが高精度を示した背景には、適切なパラメータ調整が行われていた可能性がある。（しかししながら、文献 [1] や GitHub 上に公開されているのは予測用のコードのみであり、具体的な調整手法やその根拠については明記されていない。）

追加学習を行って 100% の精度を達成したが、この精度の妥当性を検証する。上記仮説に基づき、違うアプローチとして、新規に学習を行い、それらと追加学習を行ったものを比較し、精度や追加学習を行ったため元の FairFace の性別・人種・年齢分類と比べて美男美女・一般人の二値分類タスクが簡単だったのかを確かめる。ここで、新規学習で用意したモデルは ResNet18,ResNet34,ResNet50 と EfficientNet_b0 であり、これらは Custom Model と同じパラメータ調整を行った。

5 仮説検証 1：新規学習

ResNet18,ResNet34,ResNet50 と EfficientNet のモデルで新規学習を行なった実行結果を表 2, 表 3, 表 4, 表 5 に示す。

5.1 ResNet 18

表 2: ResNet-18

Epoch	Train Accuracy	Validation Accuracy
1	97.92	99.28
2	99.23	99.22
3	99.47	99.53
4	99.61	99.48
5	99.68	99.52
6	99.76	99.38
7	99.83	99.59
8	99.83	99.51
9	99.88	99.45
10	99.88	99.52

ResNet 18 の考察：

ResNet-18 は、Train Accuracy の精度が 97.92% から 99.88% まで順調に向上した一方、Validation Accuracy の精度は 99.2% から 99.6% の範囲で安定して推移していることがわかる。Train Accuracy と Validation Accuracy の値が激しく乖離しておらず、適切に学習できていることが示唆される。またエポック 3 以降は、学習精度と検証精度がほぼ同程度の水準で推移しており、安定した汎化性能を有していると評価できる。

5.2 ResNet 34

表 3: ResNet-34

Epoch	Train Accuracy	Validation Accuracy
1	97.74	99.08
2	99.27	99.34
3	99.40	99.27
4	99.55	99.50
5	99.59	99.40
6	99.65	99.27
7	99.72	99.31
8	99.80	99.49
9	99.79	99.56
10	99.84	99.47

ResNet 34 の考察：

ResNet-34 の結果は、ResNet-18 と同様の傾向を示している。エポック 10において、Train Accuracy が 99.84, Validation Accuracy が 99.47 に達しており、ResNet18 よりも若干低い性能を示している。前述の ResNet18 よりも層が深くなっているものの、検証の精度自体に著しい向上が見られなかったことから、ResNet18 の時点では十分にモデルの表現能力が達成されており、これ以上層を増やしてもこれ以上の性能向上は見込めないだろう。

5.3 ResNet 50

表 4: ResNet-50

Epoch	Train Accuracy	Validation Accuracy
1	95.26	98.58
2	98.69	98.76
3	99.10	99.20
4	99.32	99.25
5	99.39	99.34
6	99.50	99.37
7	99.54	99.40
8	99.66	99.28
9	99.71	99.36
10	99.75	99.39

ResNet 50 の考察：

ResNet-18 や ResNet-34 の性能を上回ることはなく、ResNet では層を 18, 34, 50… と深くしても顕著な性能向上は見られなかった。さらには訓練初期の Train Accuracy が他の ResNet モデルより低い (95.26%) ことから、モデルの複雑性が増した分学習の収束に時間を要する可能性が考えられる。本データセットの分類タスクの複雑性に対して、ResNet-18 の時点ですでにモデルの表現力が十分であったことが考えられる。

5.4 EfficientNet_b0

表 5: EfficientNet

Epoch	Train Accuracy	Validation Accuracy
1	94.47	98.76
2	99.00	99.19
3	99.40	99.54
4	99.63	99.59
5	99.79	99.79
6	99.85	99.82
7	99.89	99.94
8	99.92	99.96
9	99.93	99.87
10	99.95	99.96

EfficientNet_b0 の考察 :

エポック 1 の時点で既に高い精度を示しており、以降の学習過程においても train, val の精度の両方がエポックの進行と共に安定して向上し、最終的にはエポック 10 で 99.86 という高い値を更新している。EfficientNet は本タスクにおいて極めて高い表現力を有しており、少数のエポックでも十分な収束が得られることがわかる。Train Accuracy と Validation Accuracy の推移は類似しており、両者間に大きな乖離は見られないことから、過学習の兆候も確認されない。

5.5 Custom Model と新規学習の比較

Custom Model がエポック 1 から高精度を実現できているものの、新規学習モデルについてもエポック 1 から 95% 後半と高精度で実現できており、新規学習モデルは CustomModel と同様のパラメータ調整を行ったことから Custom Model の高精度の原因がパラメータ調整によるものとは考えにくい。一方で新規学習した ResNet や EfficientNet にとって美男美女・一般人の 2 値分類タスクがそもそも簡単なタスクであった可能性がある。そこで正しく顔のみを見て分類し、ここまで高い精度を実現できているかは不透明であるため、Grad-Cam による検証も行う。

6 仮説検証 2：Grad-Cam を用いたモデルの信頼性の検証

6.1 検証の概要

Grad-CAM (Gradient-weighted Class Activation Mapping) という CNN が画像のどの部分に注目して一般人か美男美女かの判断を下したのかをヒートマップとして可視化する技術がある。アップサンプリング・ダウンサンプリングを施したデータセットに対して、前章で作成した各モデルで学習を行い、その結果を GradCam を実行することにより検証していく。図 5、図 6 に Grad-CAM を用いてモデルごとに美男美女 (good) と一般人 (normal) の各データの平均したヒートマップを示す。本研究で得られた 99% 超という検証精度の原因を探る。各画像は以下に示すモデルを適用した結果である。

- 1 行目
 - 追加学習したモデル：Custom
 - EfficientNet_b0
 - ResNet18
- 2 行目
 - ResNet34
 - ResNet50

6.2 検証結果

白背景適用後の GradCam の検証結果を図 5、図 6 に示す。

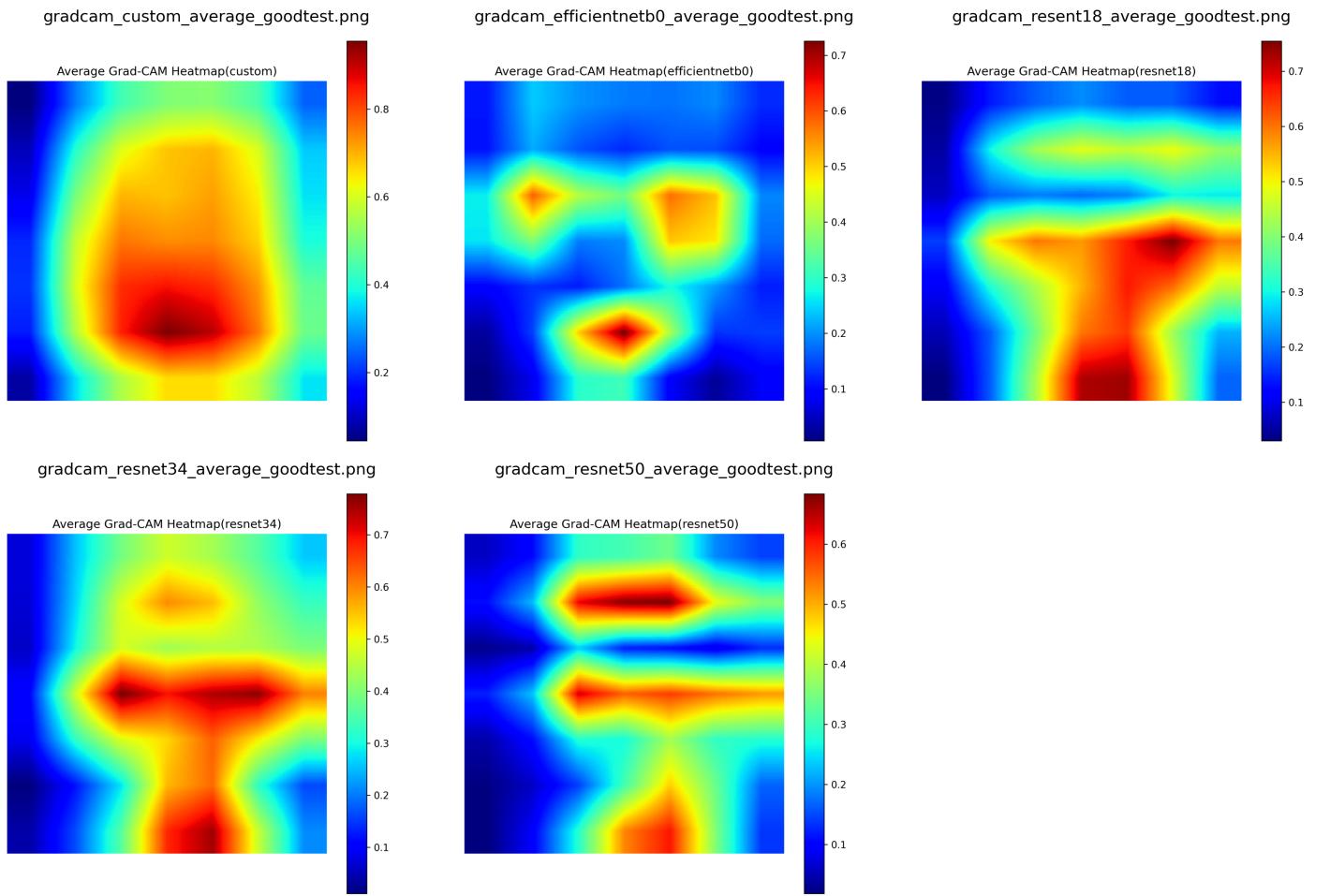


図 5: Grad-CAM を用いた美男美女の判断根拠の可視化

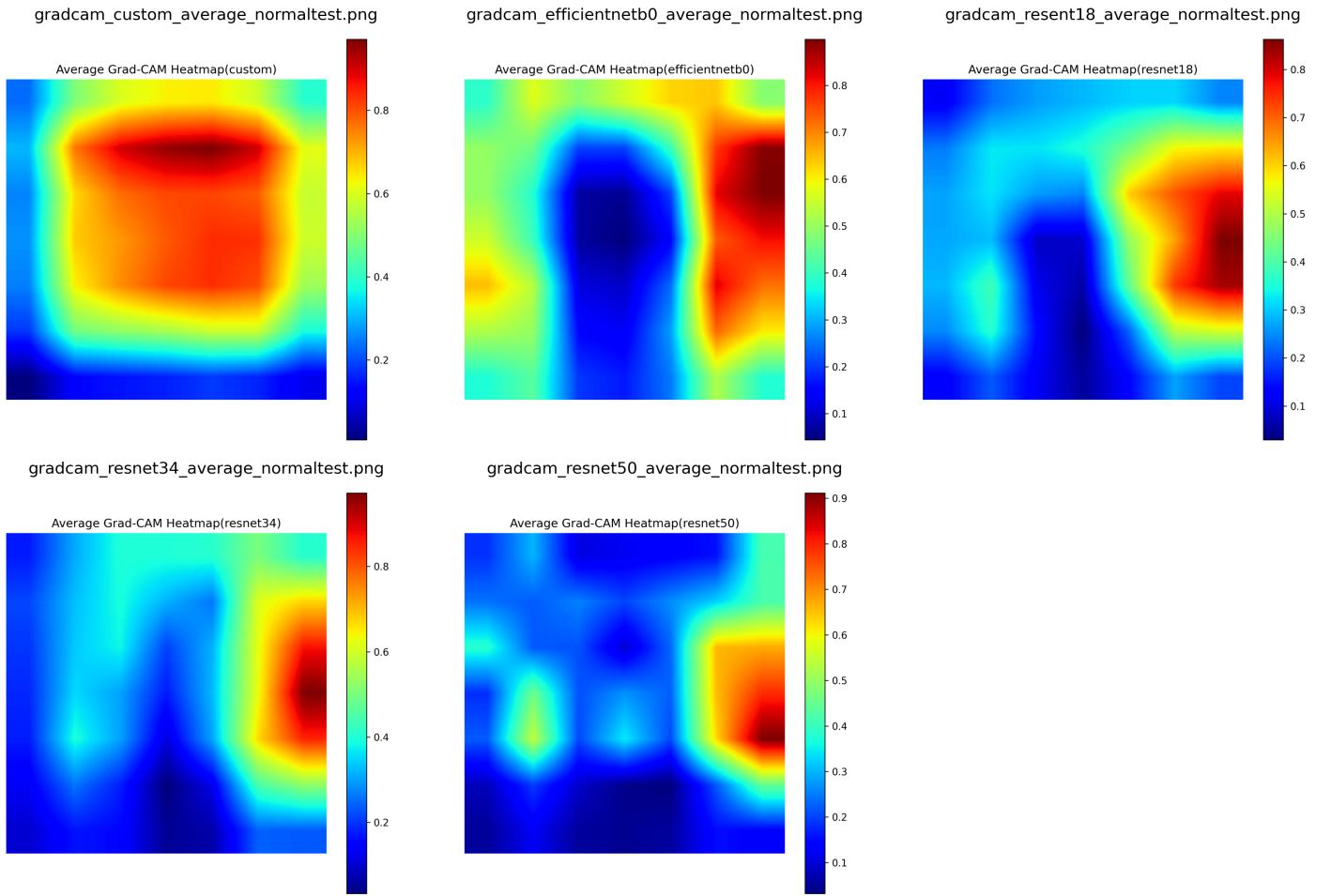


図 6: Grad-CAM を用いた一般人の判断根拠の可視化

美男美女の判断根拠について、Custom は顔全体、特に口元がよく見られているが、それ以外の ResNet・EfficientNet については、全体の傾向として、目と口が判断根拠になっていることがわかる。

一方、一般人の判断根拠について、Custom モデルは目元の上部（額と思われる）を見ており、それ以外のモデルは人物の右背景を見ていることがわかる。右背景を見ているのでは目的であった顔を見て分類する目的を果たせておらず、実験の目標は達成できていない。そこで次の章では、これらの背景の影響を抑えるために画像を白背景にして再度を学習を行った。

6.3 白背景画像を用いた Grad-Cam の改善

6.3.1 検証の概要

背景情報への依存を低減させる試みとして全画像の背景を白色に加工し、再度学習を行なった。図 7、図 8 に Grad-CAM を用いてモデルごとに美男美女 (good) と一般人 (normal) の各データの平均したヒートマップを示す。

各画像は以下に示すモデルを適用した結果である。

- 1 行目
 - 追加学習したモデル : Custom
 - EfficientNet_b0
 - ResNet18
- 2 行目
 - ResNet34
 - ResNet50

6.3.2 検証結果

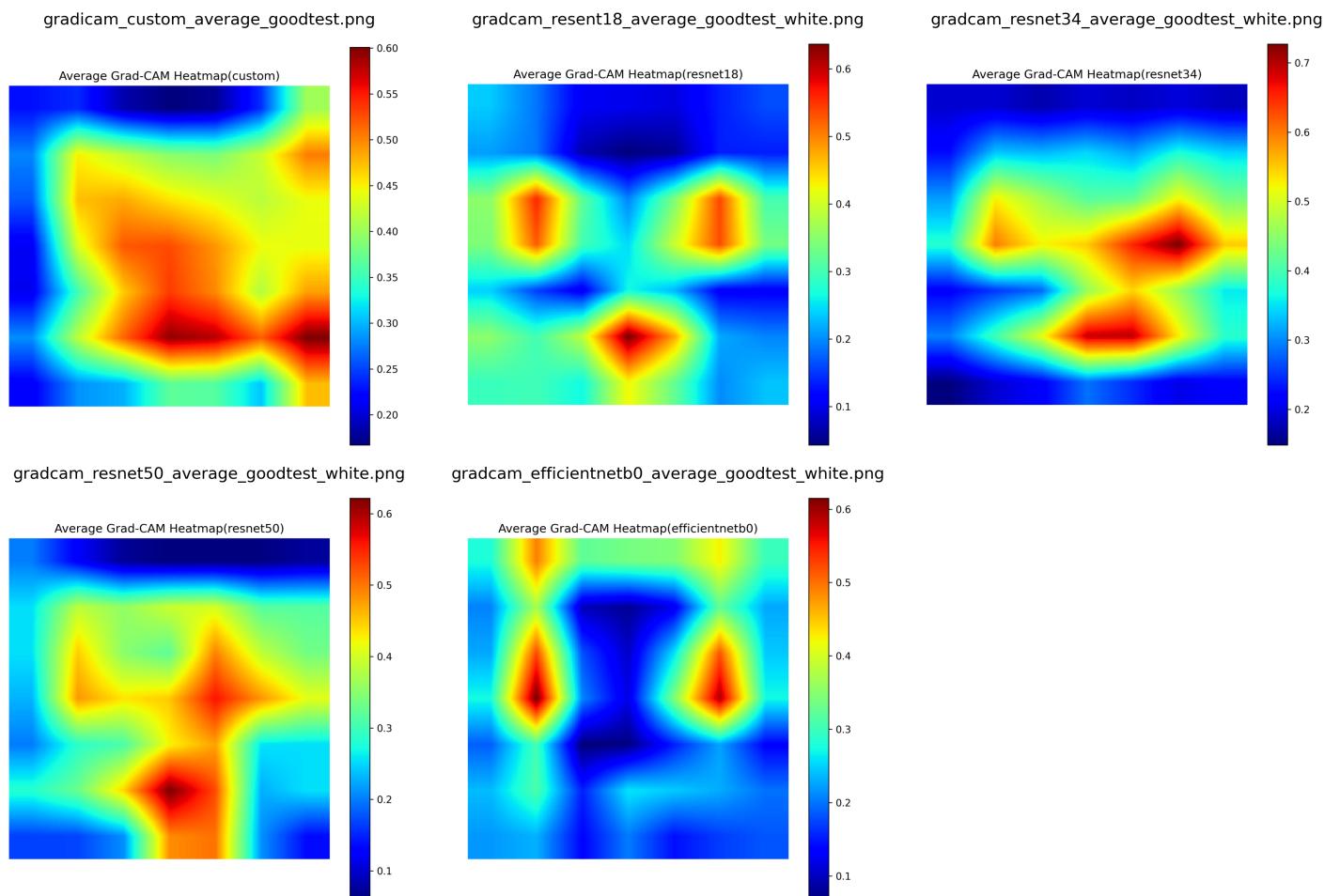


図 7: 白背景での Grad-CAM を用いた美男美女の判断根拠の可視化

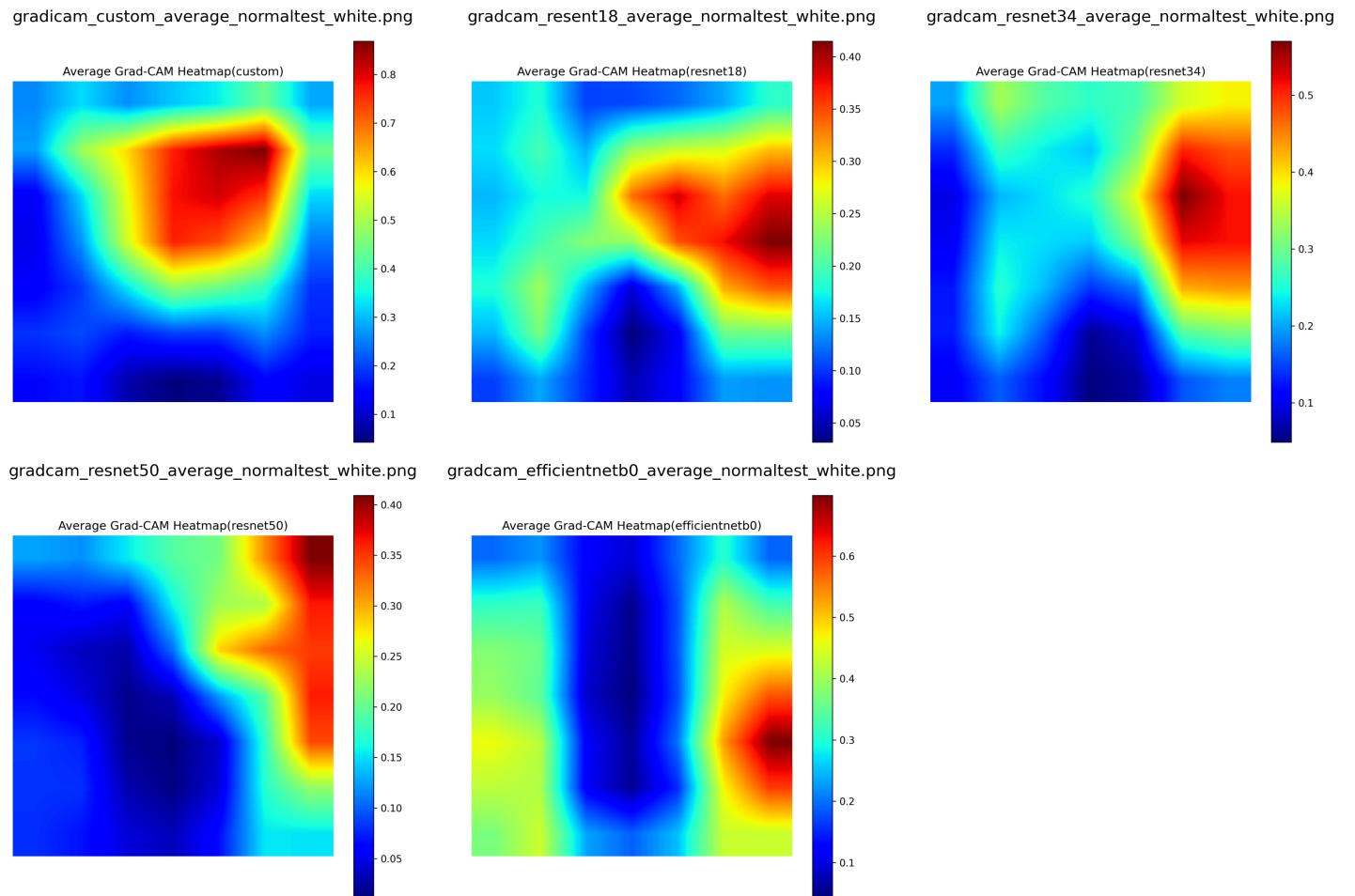


図 8: 白背景での Grad-CAM を用いた一般人の判断根拠の可視化

6.3.3 検証結果の考察

美男美女の判断根拠については、先ほどと同様に顔全体、特に目元・口元がよく見られているが、EfficientNetについては、顔から離れた背景ないしは目元を注視している。

一方、一般人の判断根拠について、Custom Model は額部分を注視している傾向にあると見受けられるが、新規学習モデルは白背景の有無に関わらず右背景を注視する傾向があり、人物には注目できていないことがわかる。

6.3.4 白背景適用後の精度

表 6 に Custom Model、表 7 に EfficientNetB0、表 8 に ResNet18、表 9 に ResNet36、表 10 に ResNet50 を示す。

表 6: Custom Model (白背景適用後)

Epoch	Train Accuracy	Validation Accuracy
1	99.06	99.35
2	99.52	99.50
3	99.58	99.47
4	99.61	99.53
5	99.63	99.58
6	99.63	99.59
7	99.65	99.57
8	99.66	99.58
9	99.65	99.57
10	99.68	99.58

表 7: EfficientNetB0 (白背景適用後)

Epoch	Train Accuracy	Validation Accuracy
1	95.42	98.99
2	98.98	99.13
3	99.33	99.34
4	99.44	99.45
5	99.55	99.45
6	99.52	99.47
7	99.57	99.21
8	99.54	99.52
9	99.59	99.52
10	99.57	99.54

表 8: ResNet18 (白背景適用後)

Epoch	Train Accuracy	Validation Accuracy
1	97.75	98.45
2	99.24	99.39
3	99.32	99.20
4	99.41	99.12
5	99.50	99.43
6	99.50	99.15
7	99.54	98.93
8	99.53	99.45
9	99.57	99.49
10	99.59	99.39

表 9: ResNet34 (白背景適用後)

Epoch	Train Accuracy	Validation Accuracy
1	97.48	98.90
2	99.13	99.25
3	99.26	99.29
4	99.33	99.04
5	99.41	99.25
6	99.45	99.31
7	99.52	99.39
8	99.54	99.12
9	99.54	99.43
10	99.56	99.45

表 10: ResNet50 (白背景適用後)

Epoch	Train Accuracy	Validation Accuracy
1	95.88	98.81
2	98.85	97.76
3	99.13	99.31
4	99.27	99.45
5	99.36	99.39
6	99.42	98.50
7	99.39	99.39
8	99.48	99.35
9	99.51	99.38
10	99.54	98.94

6.3.5 白背景適用後の精度の考察

全体的な傾向として、白背景を適用した場合には、各モデルにおいて精度の低下が見られる。特に、Epoch 10 ですべてのモデルにおいて検証精度が低下していることが確認された。白背景適用後の検証精度が、Custom Model はおよそ 0.5 ポイント、ResNet18,36,50 は 0.2 ポイント低下しており、EfficientNetにおいては、Epoch 1 の段階では白背景適用後の精度がやや高かったものの、その後のエポックにおいては一貫して精度が低下している。これらの傾向から、白背景適用後の方がモデルが正しく学習できていないことが示唆される。

6.4 GradCam 適用結果を踏まえた実験全体の考察

新規学習モデルにおいて、図 7、図 8 のようになった結果を以下に考察する。

- 背景処理の不完全性
 - 白背景化には CNN ベースの画像処理を用いたが、図 9 に示すように背景の切り抜きが不完全な画像が存在した。アップサンプリングおよびダウンサンプリングを施した後のデータに対して白背景を適用する手順を採用したため、ここからさらに図 9 のような画像を除去すると、一般人および美女のデータ数に偏りが生じる可能性があった。また、処理はサーバ上で実行されており、結果の確認に時間を要したため、時間的制約からこれらの画像を除去しなかった。
- 背景を適切に除去できた場合にも右背景を見ている傾向が強い。
 - 図 9 のように人物の境界を正確に捉え、背景を適切に除去できた場合でも、モデルが人物ではなく背景に注目する事例も確認された。図 9 のような画像が含まれた状態、かつデータ数が少なかったことが 1 要因と考えられる。
- データセットのバイアス
 - この問題の根本的な要因として、データに内在するバイアスの影響が考えられる。図 3 および図 4 に示すように、美女の画像は、記者会見におけるバックパネルやスタジオで撮影されたと考えられるものが多く、全体的に背景がシンプルである傾向にあった。また、人物の写りが良くなるように画像加工が施されていたり、いわゆる一眼レフなどの高性能なカメラが用いられていた可能性も否定できない。美女および一般人の画像は共通して 300×300 ピクセルにリサイズされており、解像度は統一されているが、色調、ボケ感、構図といった要素はデータセット構築時に考慮しておらず、明示的に制御されていなかった。
 - 一方で、一般人の画像には Flickr や Twitter、オンライン新聞などから取得した画像が含まれており、人物の横に他の人物の体の一部が写り込んでいるなど、背景が複雑なものが多く見受けられた。また、人物がカメラの方を向いていない画像も多く含まれていた。このようなデータセットの特徴により、一般人の分類においてモデルが人物ではなく背景に基づいて判断した可能性が高いと考えられる。
- 追加学習と新規学習の学習したデータセットの画像枚数の違い
 - FairFace では文献 [1] において約 108,501 枚の学習用画像が使用されている一方、本研究で用いたデータセットは、train で約 31,000 枚、test で約 7,800 枚と画像数に差がある。この画像数の違いが、学習モデルに影響を及ぼしている可能性がある。
 - 一般人のデータについては新たに収集せず、FairFace の学習済みモデルで既に使用された画像からダウンサンプリングして再利用している。そのため、モデルが顔以外の背景情報に依存して分類を行う傾向が生じた可能性がある。

図 9、図 10 に白背景適用後について、良くなかった画像の例を出す。ResNet18 で学習した GradCam の

結果も重ね合わせて示している。

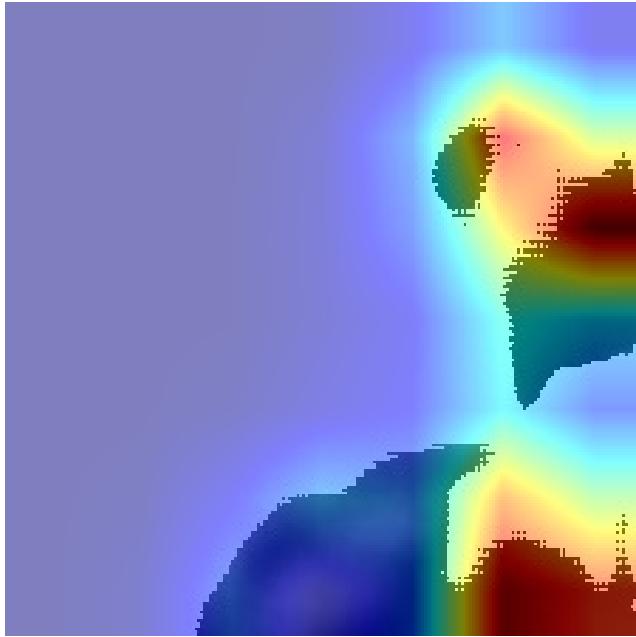


図 9: 白背景適用後の画像例 (1)



図 10: 白背景適用後の画像例 (2)

Custom Model もデータセットのバイアスは同様に受けているが、しかしながら Custom Model は正しく顔を捉えることができており、これは『新規学習モデルが背景に注目した一方、Custom Model が貫して顔領域に注目できた』という事実より、FairFace の膨大な事前学習が原因であると考察される。また、4.2 で示した仮説による影響が考えられる。つまり、本タスクに対して ResNet34 であれば十分な表現能力があり、FairFace の事前学習によって獲得された汎用的な顔特徴表現が、本タスクの学習を極めて効率的にしたと考えられる。

7 まとめ

本実験では、美男美女・一般人の二値分類タスクにおいて、追加学習を施した Custom Model と新規学習モデル（ResNet, EfficientNet）で学習を行った。実験 1 では精度を比較検討し、両モデルが高精度（95% 以上）を達成したが、実験 2 の Grad-Cam 分析により、新規学習モデルが人物ではなく背景、特に右背景に注目していることが明らかになった。これは、白背景処理の不完全性や、美男美女の画像がシンプルな背景を持つ一方、一般人の画像が複雑な背景を持つといった、データセットに内在するバイアスに起因すると考察される。対照的に、Custom Model は FairFace による大規模な事前学習の影響で、貫して顔領域に注目した特徴表現を獲得していることが示された。また、美男美女と称される人と一般人の差異は口もとと目が大きな影響を与えることが図 5、図 6 より示唆される。

このことから、タスクの精度は高いものの、モデルが正しく顔を認識しているか否かは、その学習方法とデータセットの特性に大きく依存することが判明した。

8 反省・今後の課題

8.1 反省点

データセットの誤り

新規学習モデルと追加学習モデルとの比較によって、構築したデータセットに誤りが含まれていることがあきらかになった。構築したデータセットに含まれると考えられるバイアスと、どのようにすればこのバイアス

を低減できるかの考察を以下にまとめる。

- 一般人データセットの再利用の影響
 - 想定される対処：有名人データセット同様に一般人データセットを独自構築する。
- 撮影日時の不均一性と年齢による特徴変化の未考慮（対処困難）
- 背景・撮影状況のバイアス
 - 撮影日時で画像を指定してスクレイピングを行うのは技術的に不可能であった。
 - 一般人データセットについては、人種・年齢ラベル付が行われている一方で、有名人データセットについては撮影日時が分からなかったため、2025年7月現在の年齢をラベルとして付与した。
 - 想定される対処：用意する画像の量を増やす・全て白背景を適用後、不適切な画像を除去、アップサンプリング・ダウンサンプリングを施し学習を行う。
- 画像のプロパティ（色、構図）の未考慮
 - 想定される対処：グレイスケールにする・顔のみ切り抜くなど

実験計画の見積もりミス

当初の調べ不足が原因で FairFace がデータセットのみを公開していると認識してしまっており、実験にかけられる時間が少なくなってしまった。また、実験計画よりも前処理に時間がかかり、その後の実行と結果考察のための時間が不足してしまった。

プロジェクト管理の欠如

GitHub を使おうとしたが、使える人数が少なく GitHub でのバージョン管理が実質できていなかった。さらに amane で実行する場合、さらにソースコードのバージョンが不透明になってしまった。役割分担・実験計画が適切にできておらず、実験の後半で急いで学習を行った。計画よりも前処理に時間がかかり、その後も実行しその結果を考察する時間をファイルの欠損の対応・コンテナプラットフォーム Singularity 使用のための準備といった対応に終わってしまった。

8.2 時間の都合上省いた項目

撮影日時による人物データの抽出

画像を web スクレイピングで収集したものについて、ネットに存在する美男美女の画像は撮影された日時がバラバラである。若い時の写真もあれば現在の写真もあり、年齢によって発生する特徴量の変化については考慮できていない。検索キーワードをもとにスクレイピングをする際に、人名に年齢を含むアプローチも行ったが、得られる画像が少ない・検索してもそもそも年齢で絞れないという課題が生じた。本実験で調べる限りは実験の限界と判断し、他の前処理を丁寧に行う方を優先した。

構築したデータセットの人名の偏り改善

本実験では、2024年美男美女ランキング top50 をもとに、50名ずつ抽出し、アップサンプリング（左右反転）を施した。これは一般人のデータセットと比較して、偏りが生じる。他のアップサンプリングの方法として回転・輝度変更といった手法があるが、これらでさらにデータを増やすよりかは、美男美女の人名数を50名ずつよりさらに増やす方がバイアスを削減できると考える。本実験では時間の都合・第一回の実験時の計画がずれていたため省略した。

白背景適用の改善

全て白背景を適用後、不適切な画像を除去、アップサンプリング・ダウンサンプリングを施し学習を行うことで、図7、図8のような結果を改善できる可能性がある。

参考文献

- [1] Karkkainen, Kimmo and Joo, Jungseock. FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1548–1558, 2021.
<https://arxiv.org/abs/1908.04913>
- [2] Most Handsome Man In The World 2024, shiningawards.com,
<https://shiningawards.com/most-handsome-man-in-the-world-2024/>
- [3] Most Beautiful Faces 2024, gigazine.net,
https://gigazine.net/gsc_news/en/20241229-most-beautiful-faces-2024/
- [4] Nataniel Ruiz Eunji Chong and James M. Rehg (Georgia Institute of Technology). Fine-Grained Head Pose Estimation Without Keypoints In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2074–2083, 2018.
<https://arxiv.org/abs/1710.00925>
- [5] Head Pose Estimation, <https://github.com/natanielruiz/deep-head-pose>
- [6] Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization, Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra, <https://arxiv.org/abs/1610.02391>

付録 A GradCam を用いた分析の詳細

A.1 GradCam を用いた分析の実行結果

図 11 に Custom Model (good), 図 12 に Custom Model (normal), 図 13 に EfficientNetB0 (good), 図 14 に EfficientNetB0 (normal), 図 15 に ResNet18 (good), 図 16 に ResNet18 (normal), 図 17 に ResNet34 (good), 図 18 に ResNet34 (normal), 図 19 に ResNet50 (good), 図 20 に ResNet50 (normal) の実行結果を示す。これらは、構築したデータセットの画像に対して GradCam のヒートマップを描画したものであり、客観的にモデルが顔領域のどの部分をみているのかを示す。



図 11: Custom Model (good) の実行結果



図 12: Custom Model (normal) の実行結果



図 13: EfficientNetB0 (good) の実行結果

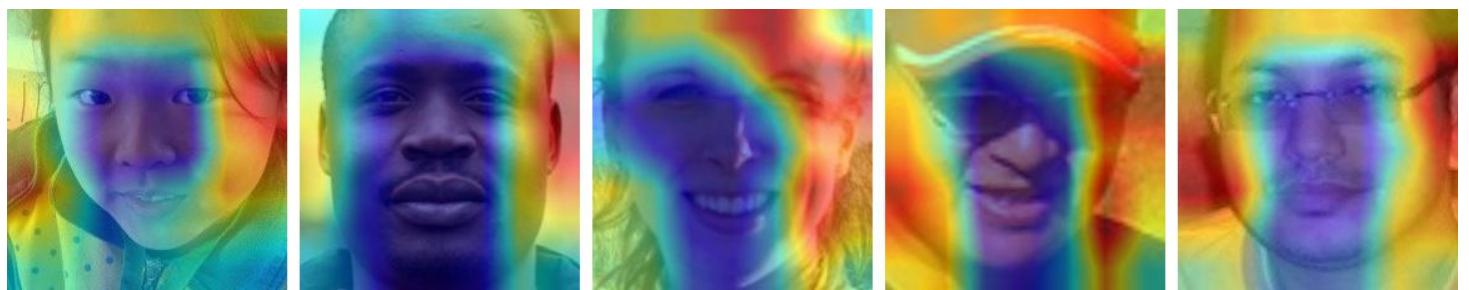


図 14: EfficientNetB0 (normal) の実行結果



図 15: ResNet18 (good) の実行結果

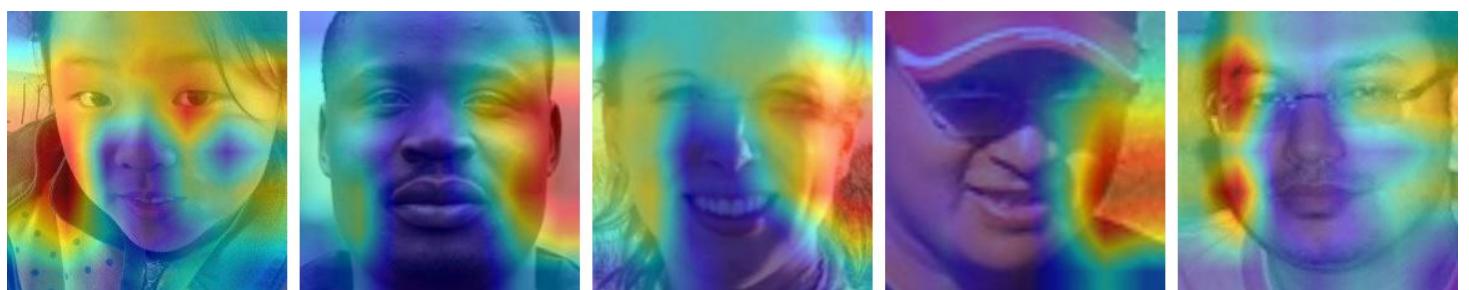


図 16: ResNet18 (normal) の実行結果

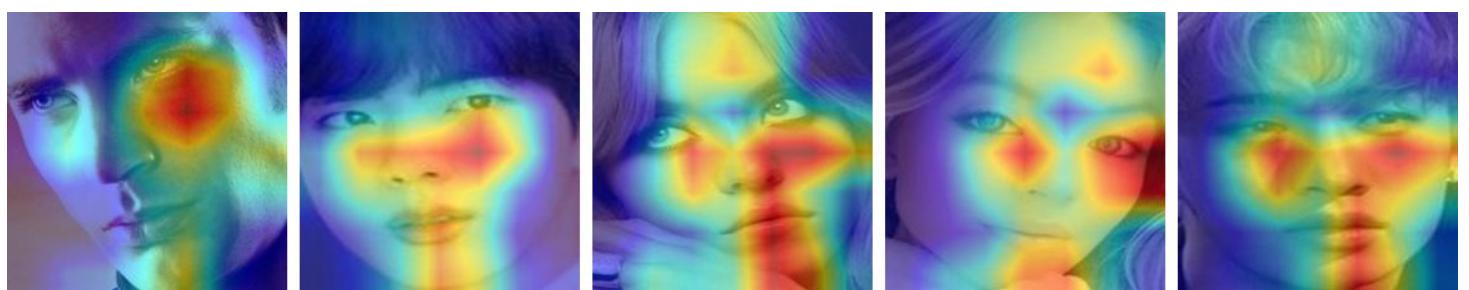


図 17: ResNet34 (good) の実行結果



図 18: ResNet34 (normal) の実行結果



図 19: ResNet50 (good) の実行結果



図 20: ResNet50 (normal) の実行結果

A.2 考察

全体的な傾向として、美男美女の顔画像では、モデルの注目は目、鼻、口といった顔の中心的なパートに強く集中している一方で、一般人の画像では注目領域が頬、額、顎の輪郭、背景など拡散している傾向が見られる。判断の根拠となる明確な特徴を見つけられていないことがわかる。

図 11 では、両目、鼻、口といった顔の中心領域をバランスよく見ていることがわかる。特定の偏りが見られない。図 12 では、額付近を注視している傾向にあるが、顔の輪郭全体から特徴を見ていることがわかる。

図 13 では、実行したモデルの中で最も目元を注目しているといえる。一方で、一般人（図 14）ではそのような傾向が見られず背景や衣服を見て判断していることがわかる。

図 15 では、顔の特に口元を注視する傾向にあり、図 16 は右背景や目元付近を学習している傾向にある。

図 17 では、目・鼻・口を注目しており、図 11 と同じような傾向が伺える。一方で、図 18 は背景を特に注目している。

図 19 は注目領域がシャープになっており、頬や額を見て判断している傾向にあるが、図 20 では依然として背景を見て判断てしまっている。

付録 B パラメータ調整によるモデル精度向上実験

B.1 概要

本実験においては、精度 100% という極めて高い性能を有するモデルの構築に成功した。そのため、本研究ではさらなるハイパーアーマーの調整による性能向上を目指すのではなく、既存の性能を維持しつつ、学習に要するコストの削減を目的とした検討を行うこととした。ハイパーアーマーの調整については、表 11 に示すように各種パラメータを変更しながら検討を行った。

表 11: ハイパーパラメータ調整表

ハイパーパラメータ	調整前	調整後
Epoch	10	5
Batch	32	64
Workers	2	4

B.2 実験結果

表 12 に、各種ハイパーパラメータの調整による実行結果を示す。本実験では、調整の有無および調整内容の違いが、学習に要する時間、学習 Loss、および学習・予測精度に与える影響について検証した。

表 12: 実行結果

	実行時間 (分)	学習 Loss	学習精度 (%)	予測精度 (%)
調整なし	32.34	0.0141	100.00	100.00
Epoch 調整	11.28	0.0022	100.00	100.00
Batch 調整	21.03	0.0003	100.00	100.00
Workers 調整	21.35	0.0026	100.00	100.00
全調整	9.84	0.0061	100.00	100.00

すべての設定において学習精度および予測精度は 100% を達成しており、性能の劣化は見られなかった。一方で、学習 Loss に注目すると、Batch サイズを調整したケースで最も小さい値 (0.0003) を記録しており、より安定した学習が行われたことが示唆される。また、Epoch 数を削減した場合も学習 Loss が低下しており、短時間で効果的な学習が行われたことが確認できる。最も実行時間が短かったのは全てのパラメータを調整したケースであり (9.84 分)、学習 Loss も 0.0061 と比較的低く、精度を保ちつつ効率的に学習を完了させることができ可能であることが分かった。

B.3 考察

本実験では、すべての設定において学習精度および予測精度が 100% を記録しており、モデルの分類性能が非常に高いことが確認された。このことから、ハイパーパラメータの調整によって精度向上を図る余地はほとんどなく、むしろ学習コストの削減が主な検討課題となった。

実行時間に注目すると、調整なしの場合は 32.34 分を要したが、全てのハイパーパラメータを調整した場合には 9.84 分と、約 70% の削減に成功した。これは、Epoch 数や Batch サイズ、DataLoader の workers 数などを適切に設定することで、計算資源の使用効率を大幅に改善できることを示している。

また、Loss の値にも一定の違いが見られた。特に Batch サイズを調整したケースでは学習 Loss が最も低く (0.0003)、学習がより安定して進んだことが示唆される。一方で、全体的に Loss の値はいずれも低く、モデルがデータに対して過剰適合している兆候は見られなかった。

以上の結果から、精度を維持したまま、学習時間や計算コストを削減するためには、ハイパーパラメータの適切な調整が非常に有効であることが示された。特に、本実験のように精度が頭打ちとなった状況においては、効率性の最適化が今後のモデル運用における重要な視点となると考えられる。