

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の識別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	3
2.1	実験目的	3
2.2	データセット構築・前処理	3
2.3	モデル選定	6
2.4	パラメータ調整	6
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	7
4.1	Custom Model の実験結果	7
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 18	8
5.3	ResNet 34	9
5.4	ResNet 50	9
5.5	EfficientNet	10
5.6	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	12
6.3	白背景画像を用いた Grad-Cam の改善	13
6.4	検証結果・考察	15
7	まとめ	16
8	反省・今後の課題	16
8.1	時間の都合上省いた項目	16

## 概要

本グループでは、美男美女（美男美女）と一般人を分類する機械学習システムを構築した。美男美女ランキングから対象者の顔写真をスクレイピングで収集し、一般人データとのバランスを図るために前処理を実施した。前処理として画像サイズ調整や顔位置の正規化を行い、分類器を構築した。実験では美男美女と一般人の顔画像の識別を試み、機械学習モデルから人間の美的感覚に近い判断をどの程度行えるかというデータを明らかにする。

# 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”と称される美男美女と一般人を識別・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（美男美女）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量的的に見る。

## 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

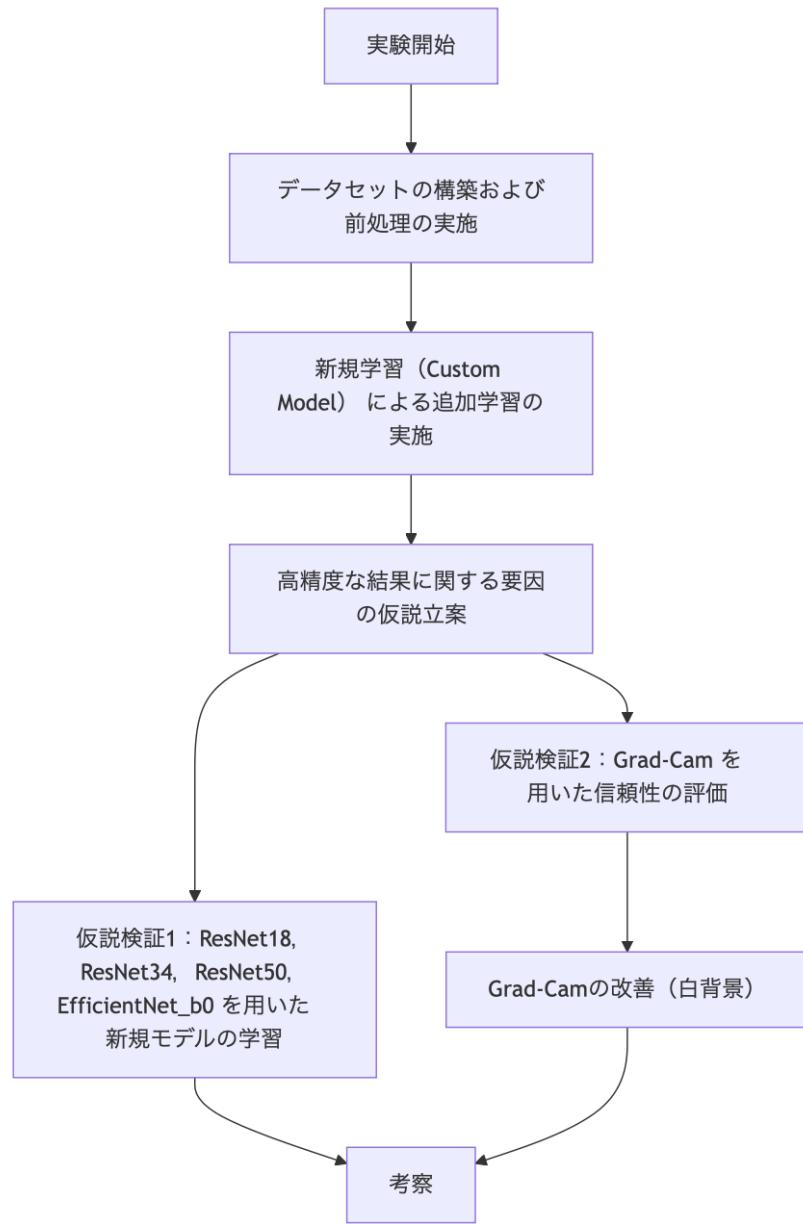


図 1: 実験のフローチャート

## 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。

#### 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

[2] や [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グッズなど）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](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  $300 \times 300$  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

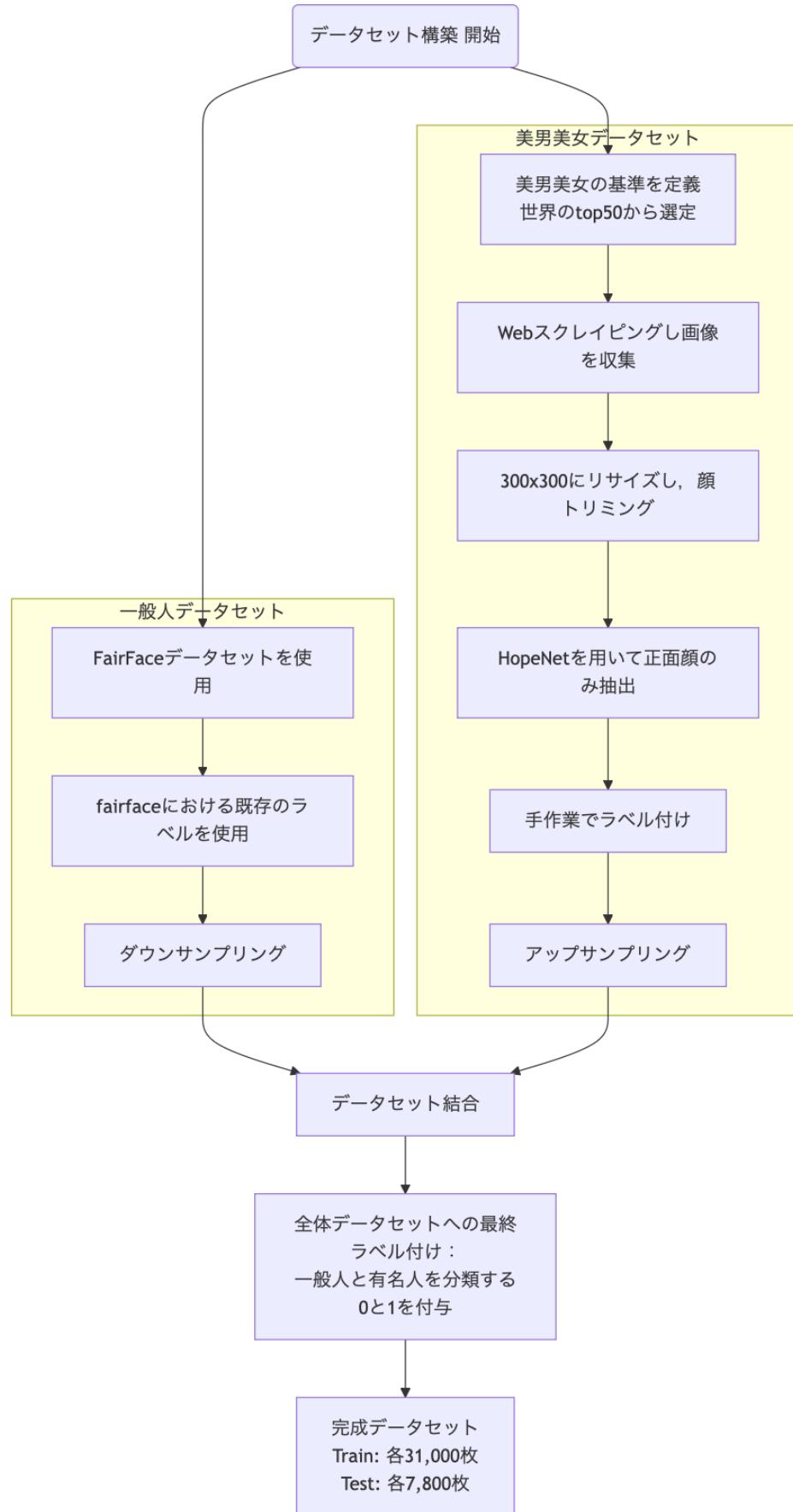


図 2: 前処理のフローチャート

## 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 [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、[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。

# 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
- 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
- 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

図 3, 図 4 に用意したデータセットの画像の一部を示す。これは筆者がランダムに選択したものであり、レポートの可読性を向上させるために用意した。

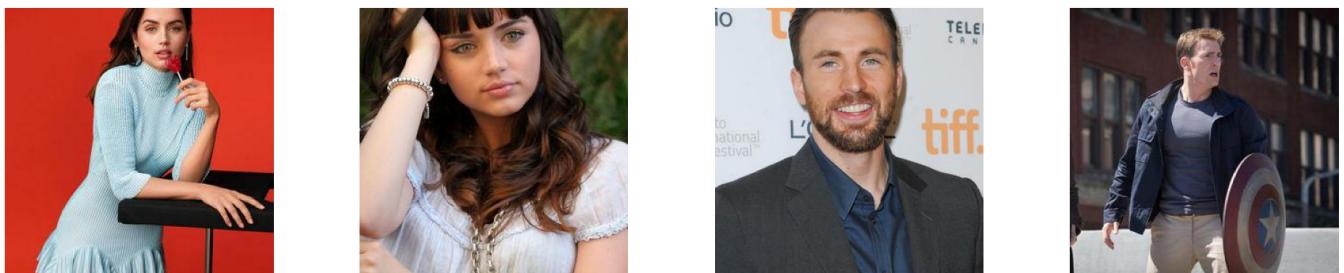


図 3: 美男美女データセットの画像例



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

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（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

- **要因 2. パラメータ調整の違い**

- FairFace のモデルが高精度を実現した原因として適切なパラメータ調整が考えられるが、[1] や GitHub 等を見ても与えられているのは学習済みモデルを使って予測するコードであり、具体的にどのようにパラメータ調整が行われたのか、またその理由は不明である。

- **要因 3. 画像の枚数の違い**

- \* fairface では [1] より、108,501 枚の学習データが使用されている。私たちが用意したデータセットの画像数は、およそ train で 31000 枚ずつ、test で 7800 枚ずつであり、そもそも学習している画像の枚数に違いがあるのではないかと考えられる。

- **要因 4. 構築したデータセットの問題**

- 構築したデータセットは一般人のデータセットは新規で用意しておらず、FairFace の学習済みモデルすでに使用された画像データをダウンサンプリングした上で再度美男美女の画像群とともに学習させている。そのため一般人のデータを学習時に背景を見る傾向にある可能性がある。

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

### 5.2 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

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。

### 5.3 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

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、ResNet-18 よりも若干低い性能を示している。前述の ResNet-18 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、ResNet-18 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。

### 5.4 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

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 モデルよりも低い

(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。

## 5.5 EfficientNet

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

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。

## 5.6 Custom Model と新規学習の比較

図??に Custom Model と新規学習の実行結果をグラフとして示す。

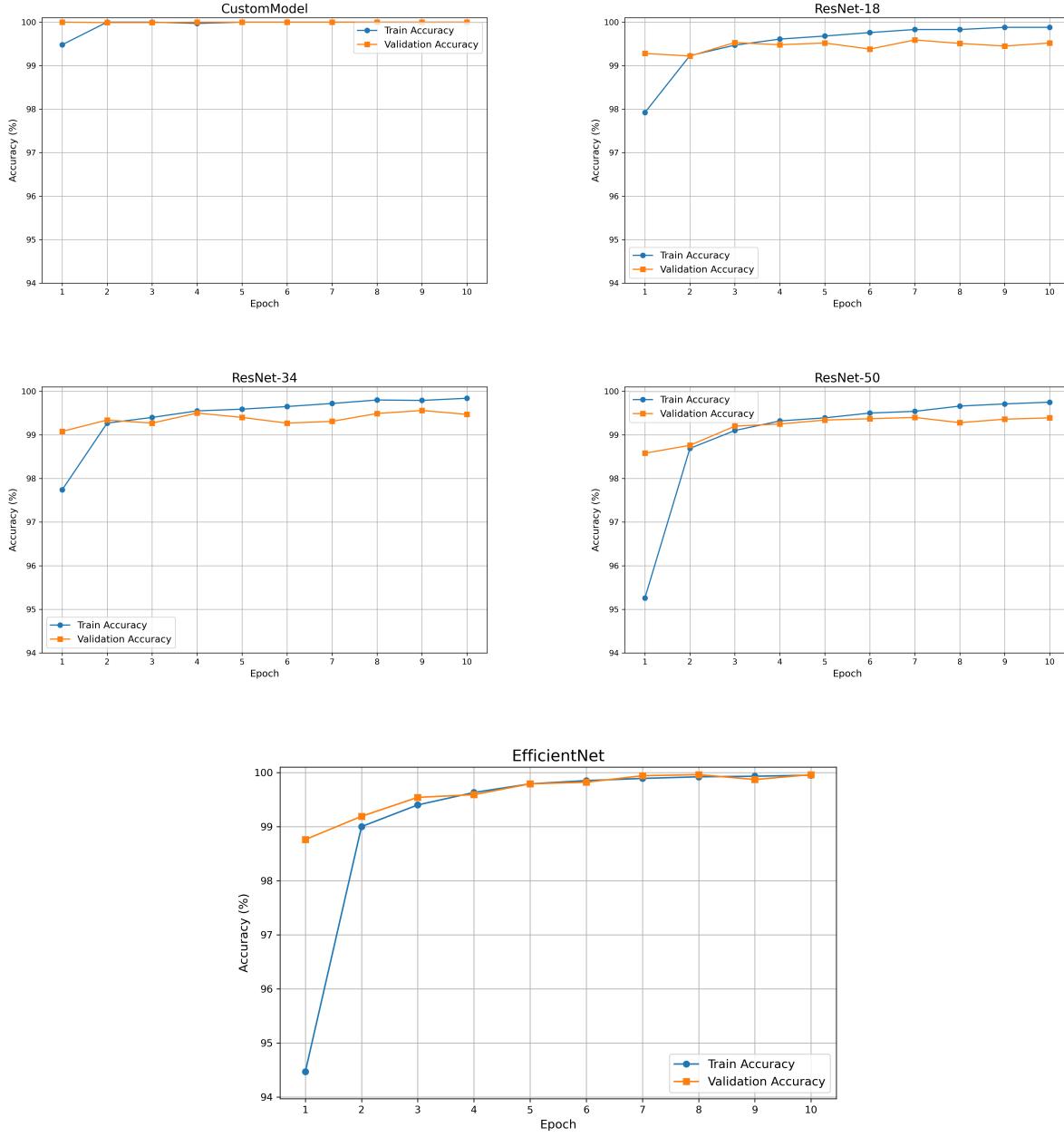


図 5: 各モデルの出力結果 (Custom Model と新規学習モデル)

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。図 6, 図 7 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

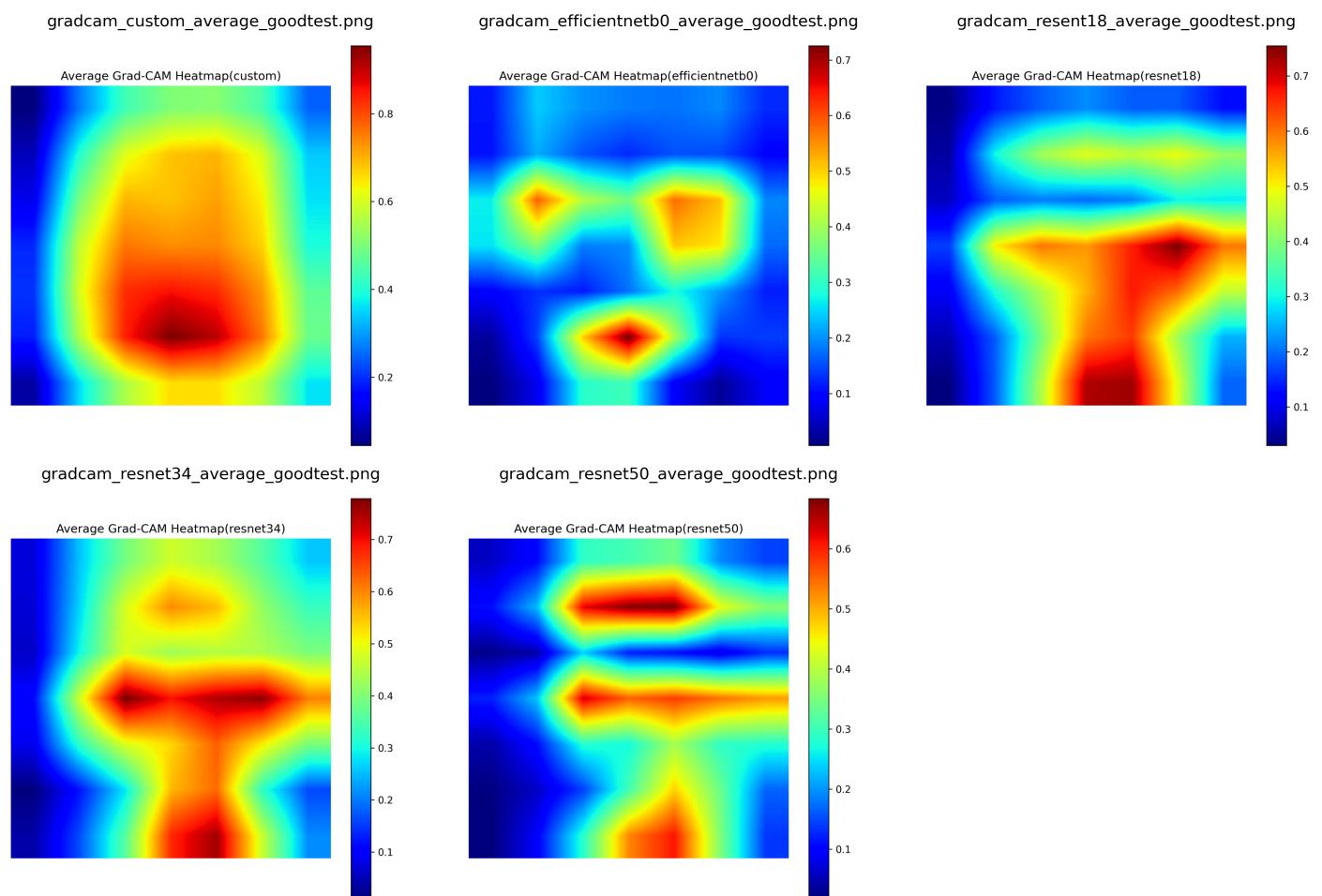


図 6: Grad-CAM を用いた美男美女の判断根拠の可視化

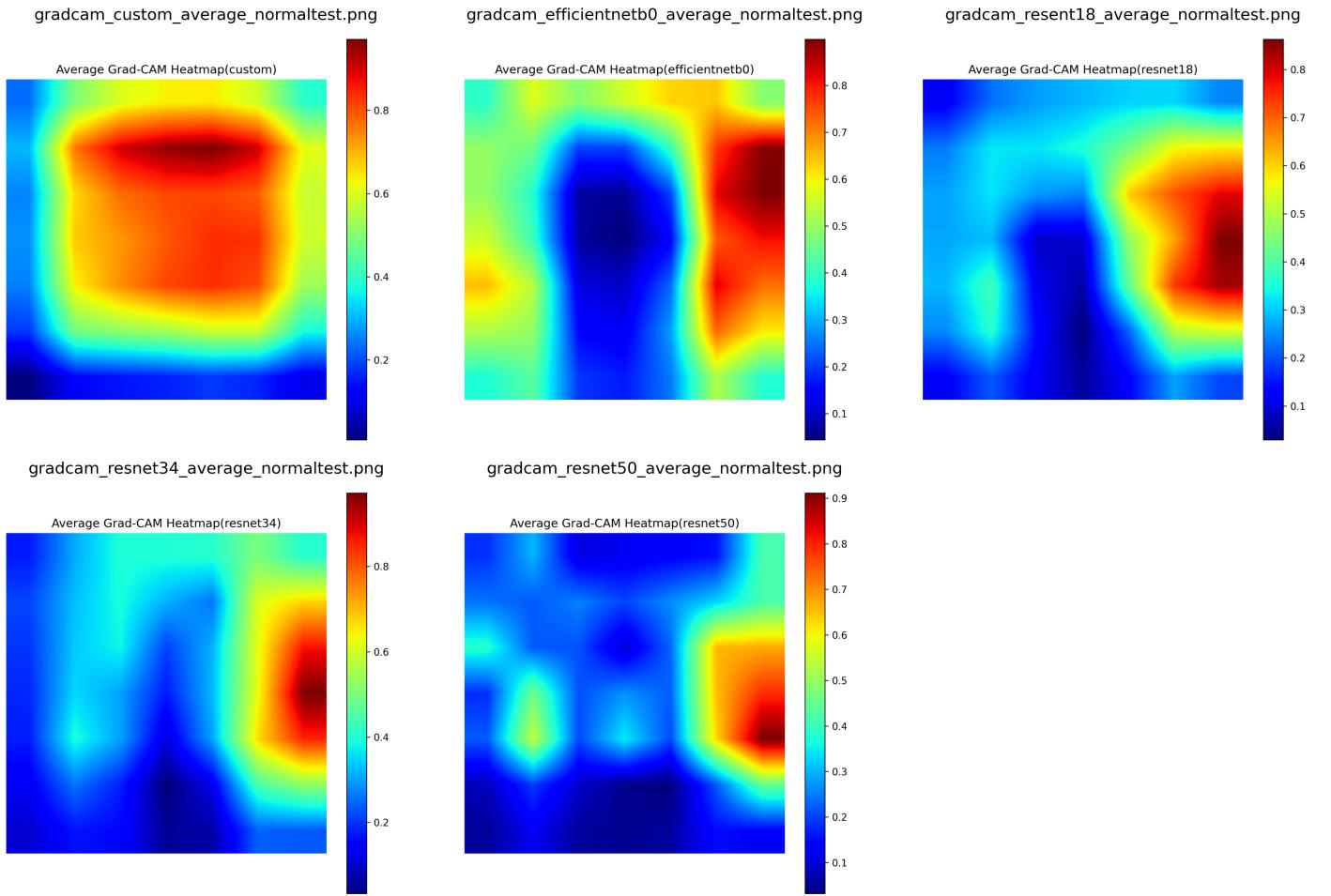


図 7: Grad-CAM を用いた一般人の判断根拠の可視化

### 6.2.1 検証結果の考察

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 8、図 9 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

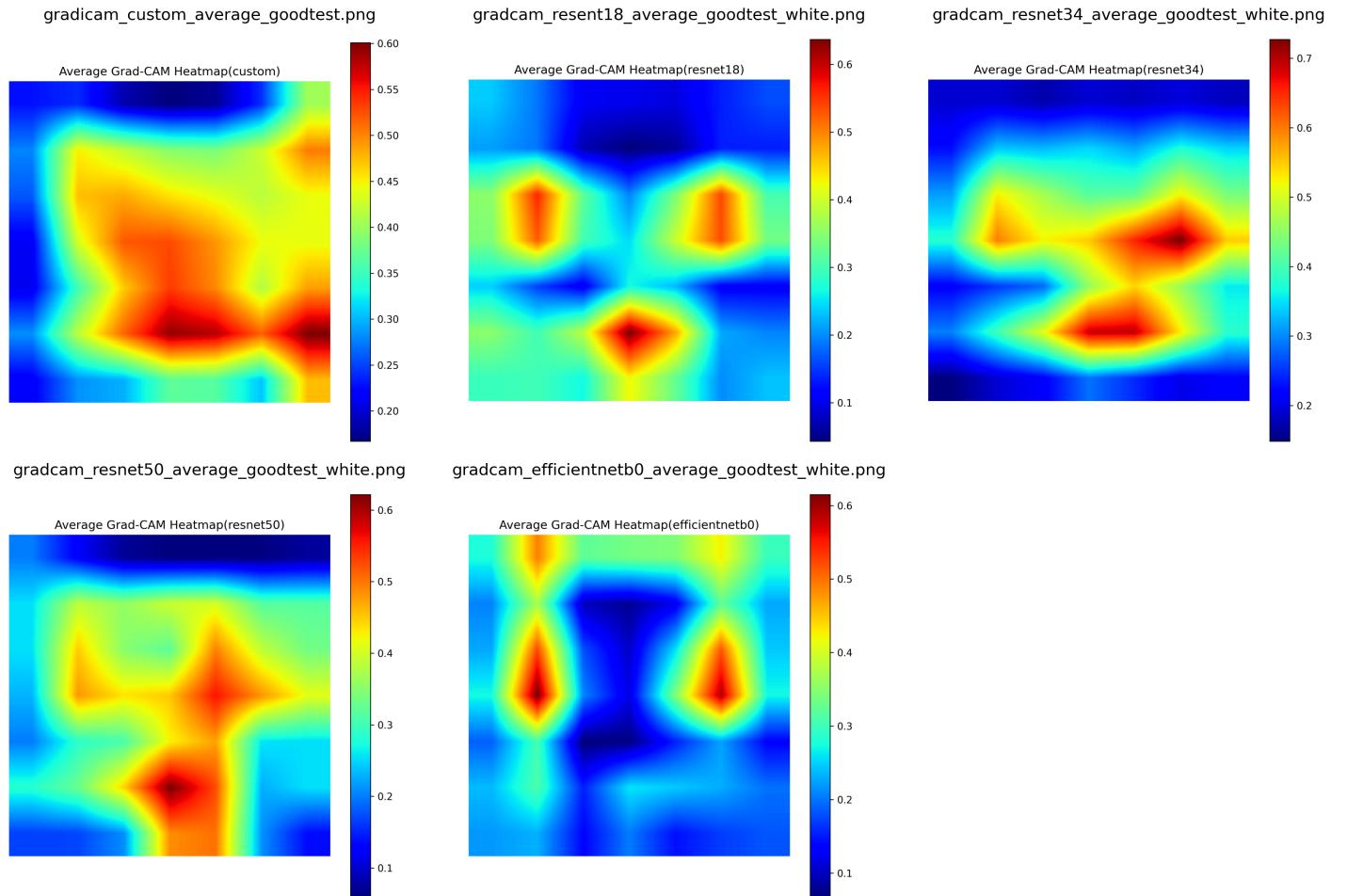


図 8: 白背景での Grad-CAM を用いた美男美女の判断根拠の可視化

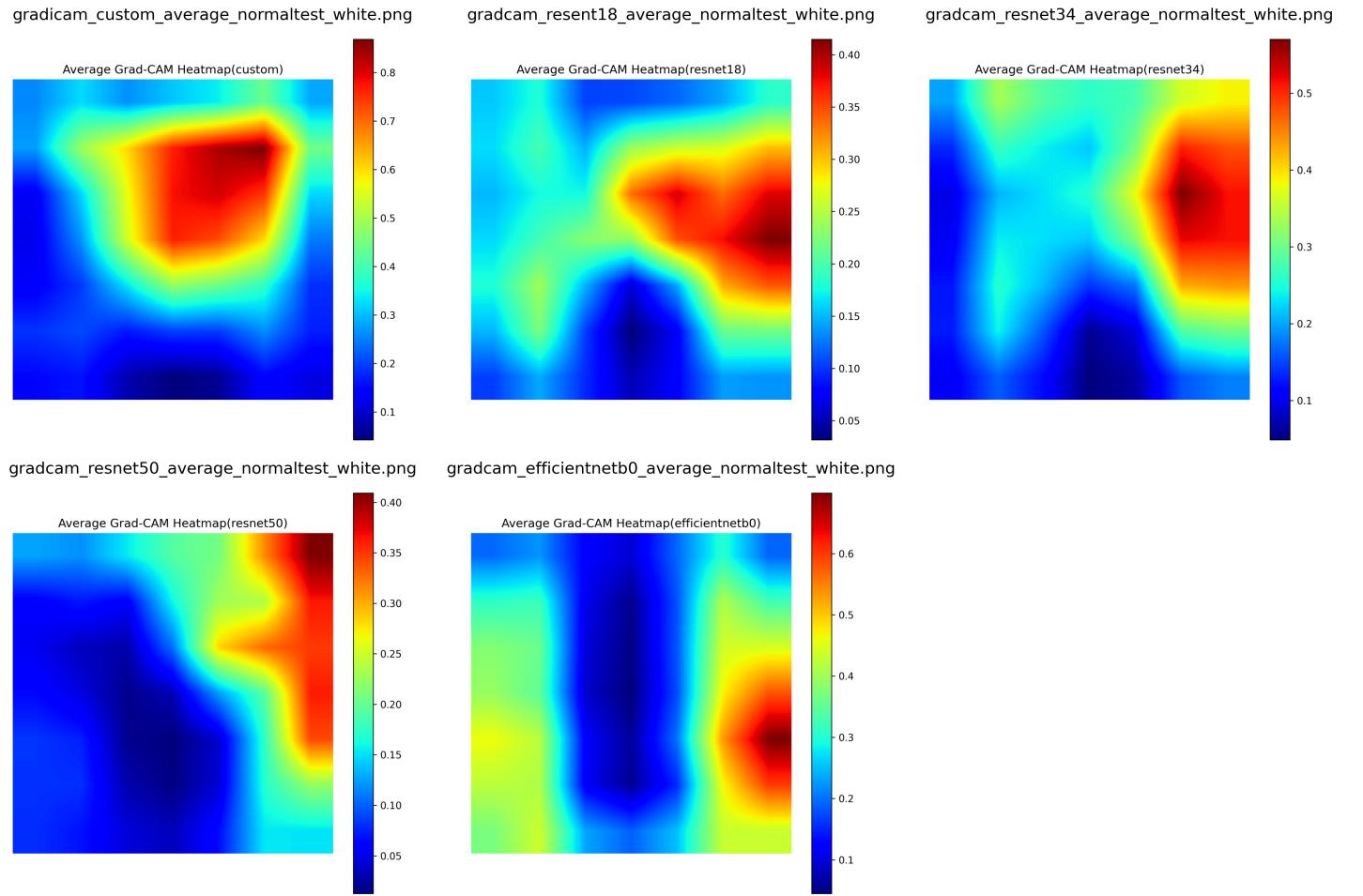


図 9: 白背景での Grad-CAM を用いた一般人の判断根拠の可視化

#### 6.4 検証結果・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、白背景の有無に関わらず右背景を注視する傾向にあり、人物には注目できていないことがわかる。

この原因については、データに内在していたバイアスの影響が考えられる。

図 3、図 4 に示したように美男美女の画像の全体的な傾向として、記者会見などで用いられるバックパネルやスタジオで撮影されていると思われる画像など、背景がシンプルなものが多い。また人物の写りがよくなるように画像加工されていたり、美男美女の写真是いわゆる一眼レフなどの良いカメラが使われていたりといった可能性も考えられる。美男美女・一般人ともに  $300 \times 300$  にリサイズ処理を施したため、解像度については統一されているが、色やボケ感・特定の構図についてはデータセット構築時に考慮していない。

一方で、一般人の画像は、Flickr 画像と Twitter やオンライン新聞の画像などがデータセットに使われており、人物の横に他の人の体の一部が写っているなどの背景が乱雑なものやカメラを向いていなかったりするものが多い傾向にある。そういう構築したデータセットの画像の特徴から、一般人の分類において背景で判断されたと考えられる。

しかし、それを踏まえたとしても 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 が一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

gradcam の実行結果を見ると、口と目が美男美女と一般人で異なることが示唆される。

## 8 反省・今後の課題

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。

### 8.1 時間の都合上省いた項目

#### 8.1.1 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。

#### 8.1.2 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。

## 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/](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>