

# カメラ俯角変動に頑健な人物対応付けにおける 画像特徴と視覚言語特徴の比較検討

佐久高太郎 神谷卓也 米田駿介 西山正志（鳥取大学）

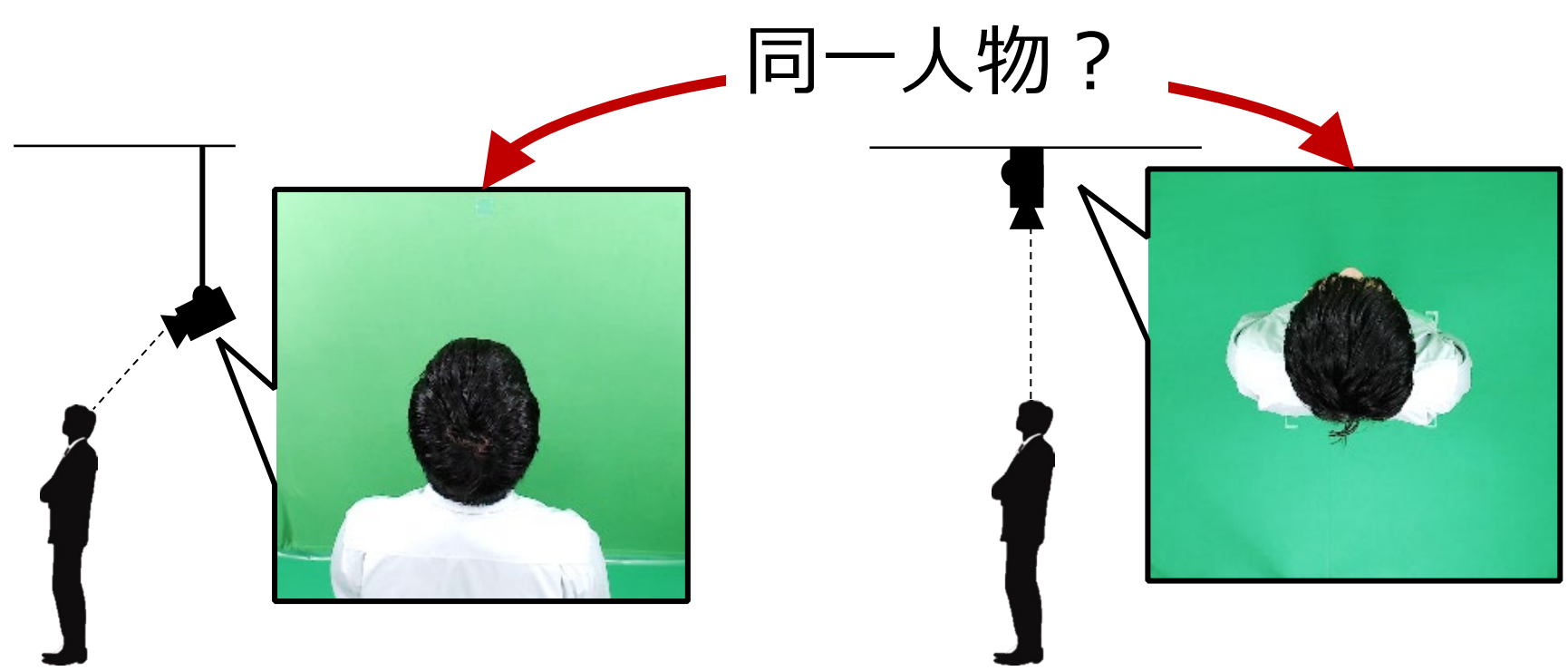


## 目的

### カメラ俯角変動への頑健性向上に向けて画像特徴のみを用いる場合の対応付け精度と 視覚言語特徴を用いる場合の対応付け精度とを比較

**問題設定：** 頭上カメラを用いた人物対応付け

カメラ俯角変動により人物の見え方が大きく変化



人物対応付けの精度が低下

- エレベータ内など人物が立ち止まっており顔などの有効な特徴が得られないシーンを想定
- 本人の画像は1枚のみ辞書として利用可能

**画像特徴のみを用いる場合**

背景：古くから活用されている

狙い：**見え方の分布**に基づきカメラ俯角に不変な特徴を抽出

現状：カメラ俯角変動への頑健性向上に向けた手法としても**有用性が確かめられている**

[Yan+, ICCV2021]

**視覚言語特徴を用いる場合**

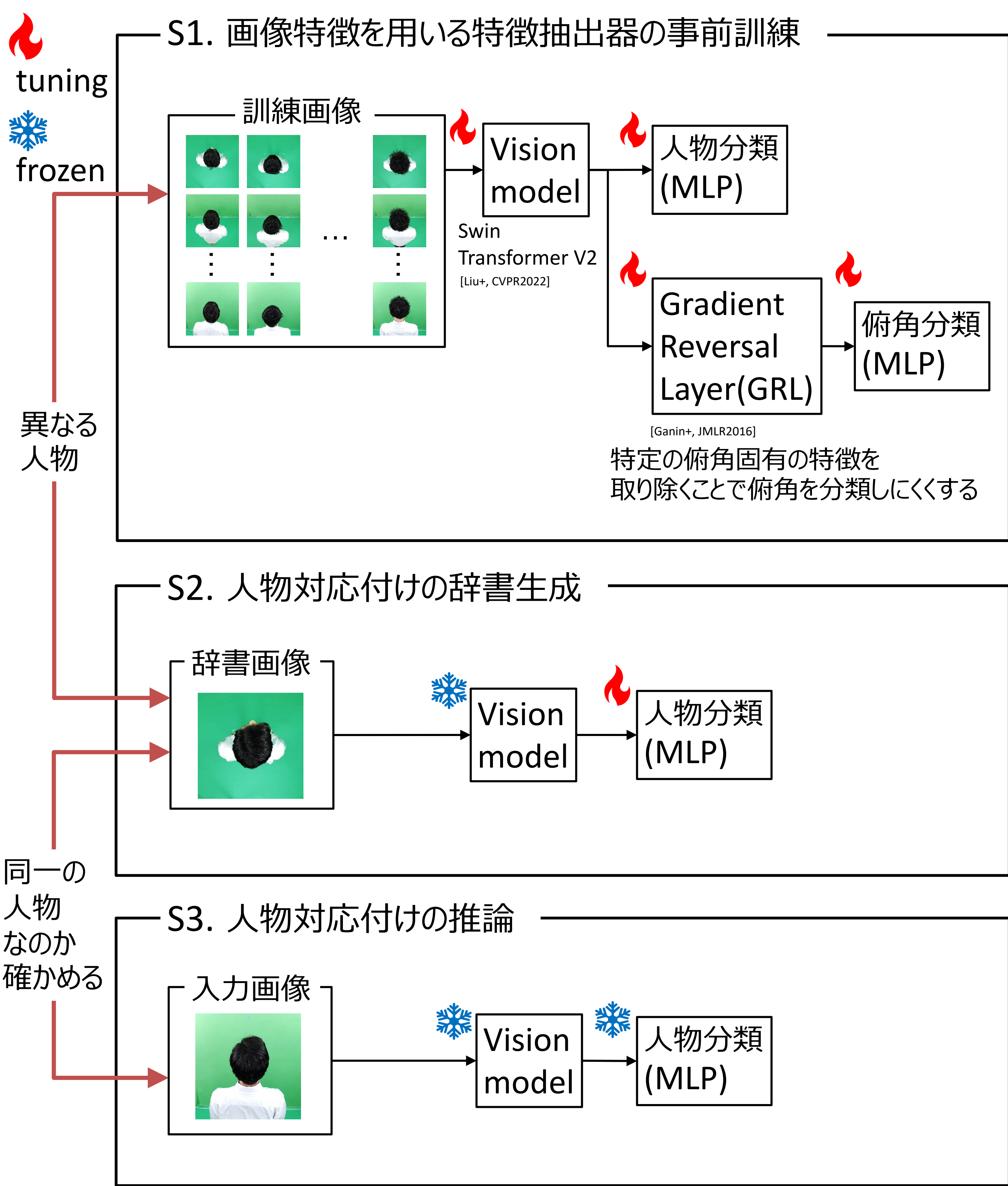
背景：近年注目されている

狙い：**大規模視覚言語モデルに内在する**カメラ俯角に不変な特徴を抽出

現状：カメラ俯角変動への頑健性向上については**有用性が確かめられていない**

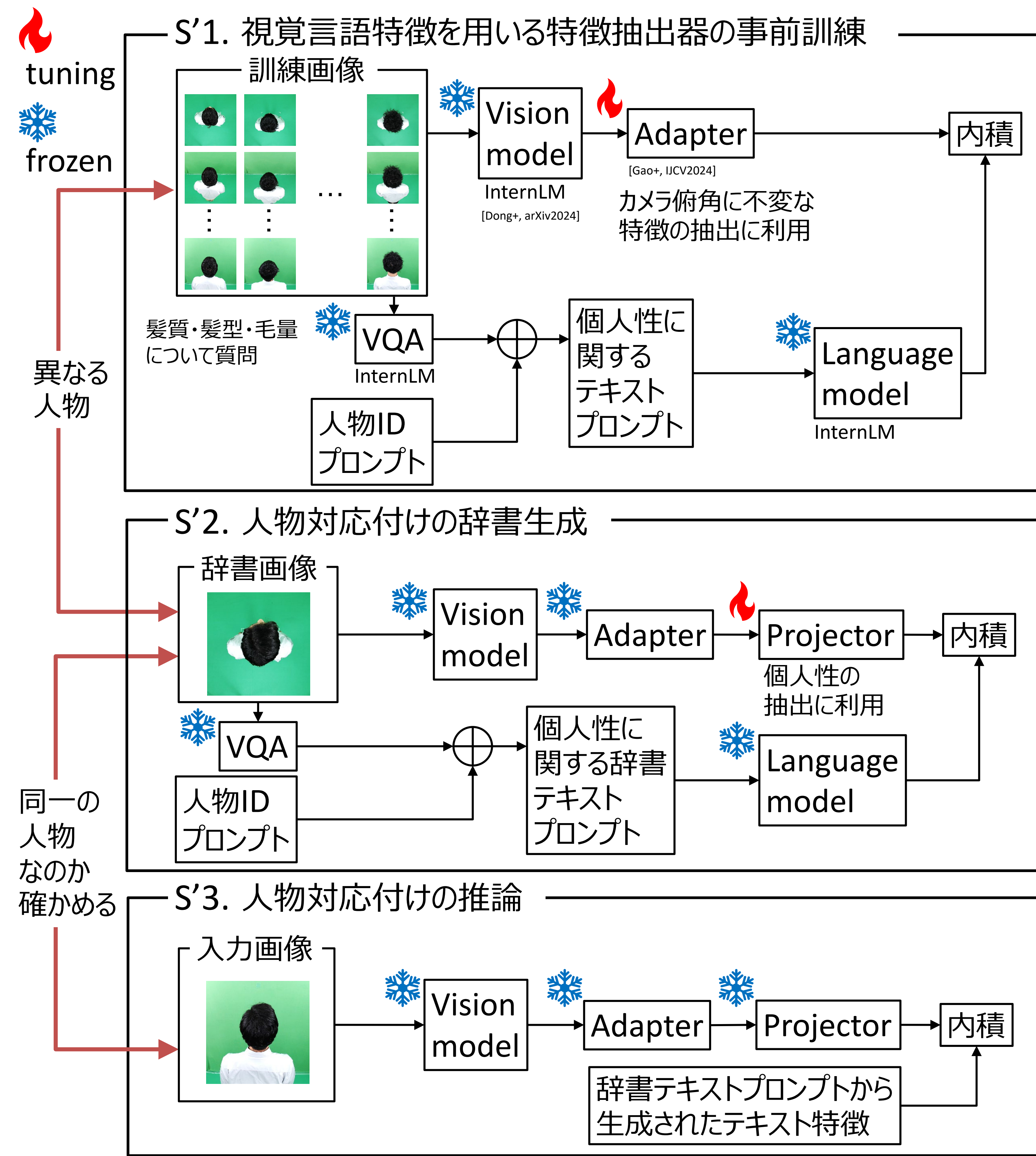
## 画像特徴のみを用いる場合

人物内変動を抑えるため勾配反転を用いて学習



## 視覚言語特徴を用いる場合

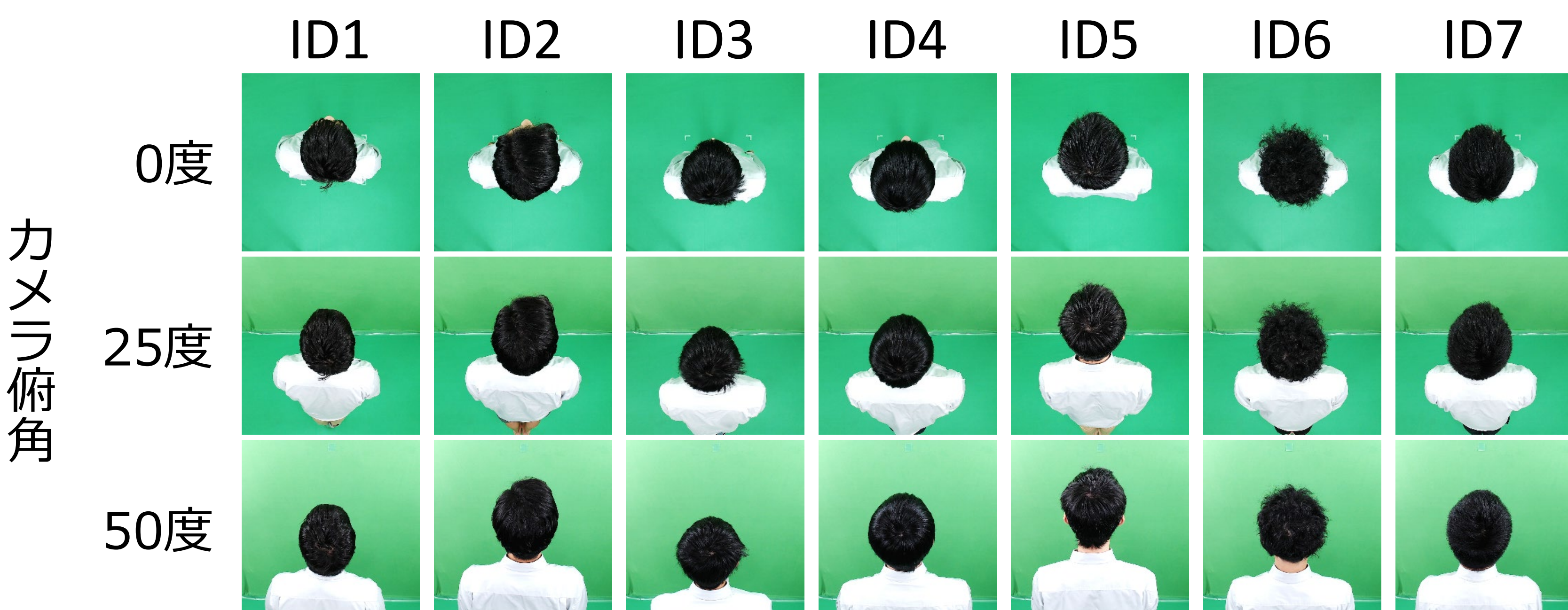
画像特徴とテキスト特徴とが類似するように対照学習



## 実験

**データセット**

- 被験者数：50人
- カメラ俯角：0度・25度・50度
- 特徴抽出器の事前訓練：25人をサンプリング
- 評価指標：カメラ俯角の異なる2枚の正解率
- 本人の辞書画像:1枚のみ
- 推論時の入力画像：辞書とはカメラ俯角の異なる1枚
- 推論時の登録人数：25人



**人物対応付けの精度**

(%)

|      | 画像特徴のみ           | 視覚言語特徴    |
|------|------------------|-----------|
| 精度 ↑ | <b>73.7±13.2</b> | 39.8±12.7 |

- 画像特徴のみを用いる場合の精度と比較して視覚言語特徴を用いる場合の精度が低下したことからテキスト特徴が影響していると考えられる

**視覚言語特徴を用いる場合の精度低下の考察**

- 頭上カメラから撮影した画像での人物対応付けを行う場合個人性に関する微細な視覚的差異を識別するためのテキストプロンプトの表現能力が足りない可能性がある
- Vision modelとLanguage modelの両方においてファインチューニングを行っていない