大規模視覚基盤モデルからの特徴量からの カメラ情報の予測可能性の検証

大阪大学 産業科学研究所 氏名 中島悠太

目的　CLIPと呼ばれる大規模視覚基盤モデルは、画像の意味内容をエンコードし、画像検索等に利用するものである一方で、撮影時のカメラに関する情報も併せてエンコードする可能性があり、これは情報漏洩や意味内容によらない検索につながる可能性がある。

内容　CLIPに限らず、種々の大規模視覚基盤モデルからの特徴量からカメラ情報の予測可能性を検証した。

結果　CLIP特徴量はカメラの種類等に関する情報を多く含むことを明らかにした。一方で、最近よく使われるDINO等の特徴量では、カメラ情報の予測精度は低いことがわかった。

利用した計算機
SQUID GPUノード群

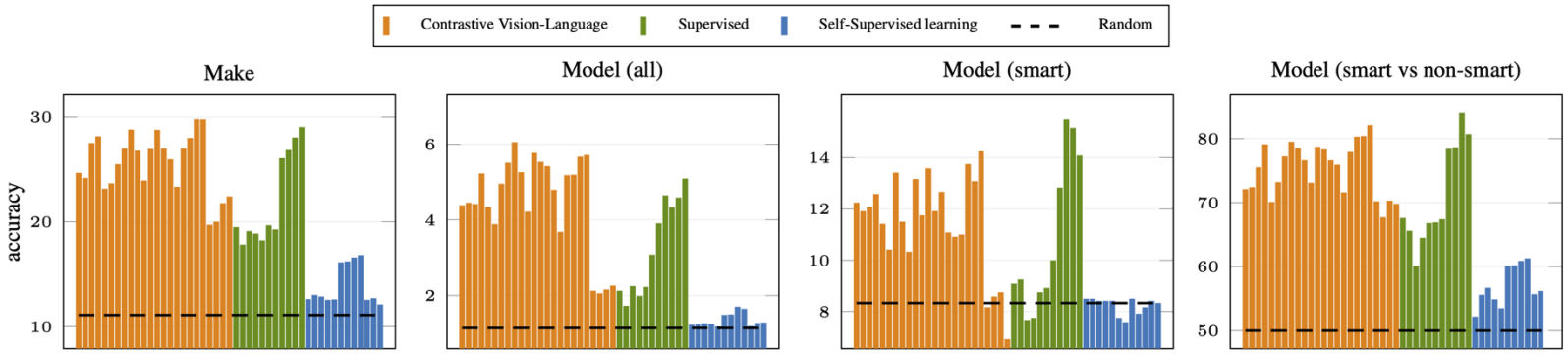

図: カメラ情報の予測精度。高いほどカメラの情報が予測できていることを表す。オレンジ色はCLIPベースのモデル、緑色は教師あり学習によるモデル、青色は自己教師あり学習によるモデル。