

Deep Learning を用いた画像の回転補正の方法

日大生産工(院) ○荻上 隼,
日大生産工 伊藤 浩

1 まえがき

スマートフォンなどのカメラで人物や風景を撮影するとき、被写体に合わせてカメラを横向きに構えたり縦向きに構えたりすることで、アスペクト比が縦長、横長の画像を保存することができる。しかし、撮影時に傾いて撮影してしまったり、意図した方向に撮影できず、正しい方向に画像が保存されない場合がある。また、一般的にカメラで撮影した画像はEXIF (Exchangeable digicame-exif file format) により画像の向きなどの情報を保存している。このような画像をコンピュータなどで取り扱う場合、EXIFデータによって向きの補正を自動で行うことができるが、編集ソフトの種類などによって正しく向きが補正できない場合がある。近年、脳の神経回路を人工的に模したニューラルネットワークを多層からなる構成にし、Deep Learning[1] で学習した分類器が高い認識精度で注目されている。この分類器を用いて、EXIFデータなどに寄らず、画像情報のみで傾いてしまった画像を自動補正する手法を提案する。

2 原理

一般的な分類器は、学習段階で正しい方向、つまり傾きのない画像で学習をしている。そのため、画像を傾けて分類器に入力すると、出力が小さくなる。これを利用して、分類器に入力する画像として、入力時の向きとこれに対して傾いた向きの画像を用意し、出力の最大値を比較することで、正しい方向の画像を判断することができる。この手法は入力画像が分類器に学習されたデータラベルにない場合でも有効である。本実験に用いる Deep Learning の分類器は、Omkar M. Parkhi らによる CNN (Convolutional Neural Network) である学習済みのモデル VGGface[2] を用いる。VGGface は、俳優や女優などの人物を分類する分類器である。VGGface に人物の画像を入力すると、データ

ラベル内の人物のうち出力が最大の人物を出力する。しかし、人物の画像を傾けて入力した場合、分類器は正しい方向で学習を行っているため、それを人物と認識できず出力が全体的に小さくなる。また、傾きが大きいほどその出力は小さくなる。これにより、出力の大きさを比較すれば、どの程度傾いているか区別できるため、正しい方向に補正することができる。M 個のラベルをもつ分類器に、人物画像を傾けながら入力するとき、 θ 度回転させたときの j 個目のラベルの出力を $x_j(\theta)$ とする。 $x_j(\theta)$ が最大となるときの θ を m 度とすれば

$$m = \arg \max_{\theta, j} x_j(\theta) \quad (1)$$

となる。入力画像を m 度回転させると正しい方向の画像となる。

3 実験

本実験では、回転角度を $\theta = 0, 180, \pm 90$ とした 4 枚の画像を用いる。まず、Fig.1 の画像について実験を行った。この人物は VGGface に学習されている人物である。Fig.1 の 4 枚の画像をそれぞれ

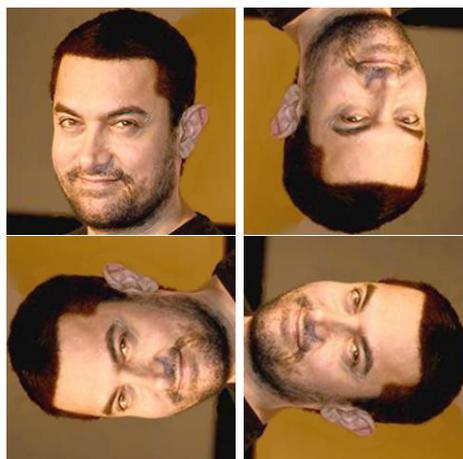


Fig. 1: θ 度回転した学習された入力画像 (左上: $\theta = 0$ 右上: $\theta = 180$ 左下: $\theta = -90$ 右下: $\theta = 90$)

VGGfaceに入力した。Fig.2はその結果であり、横軸をアルファベット順のデータラベル、縦軸を最終層のニューロンの出力としている。VGGfaceのデータラベルの数は $M = 2622$ である。

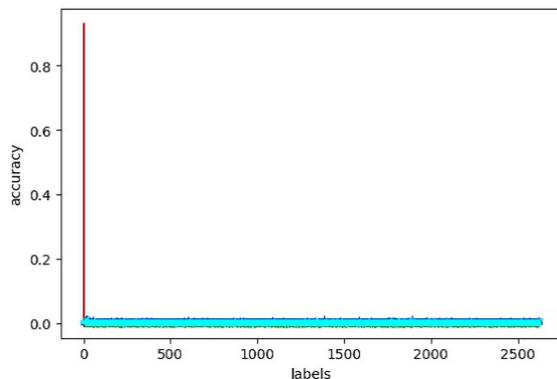


Fig. 2: 入力画像の傾きに対する最終層のニューロンの出力 (赤: $\theta = 0$ 、青: $\theta = 180$ 、緑: $\theta = -90$ 、シアン: $\theta = 90$)

Fig.2 より、出力の最大値は正しい方向である $\theta = 0$ のとき約 0.9 であり、そうでないときは全て 0.1 以下である。このことから、正しい向きと傾いた向きの画像を区別することができる。



Fig. 3: θ 度回転した学習されていない入力画像 (左上: $\theta = 0$ 右上: $\theta = 180$ 左下: $\theta = -90$ 右下: $\theta = 90$)

次に、VGGface に学習されていない人物である Fig.3 の画像を入力する。Fig.4 は Fig.2 と同様の条件で、VGGface に Fig.3 の 4 枚の画像を入力したときの出力結果である。Fig.4 より、Fig.2 と同様に正しい向き ($\theta = 0$) と傾いた向きは区別できるが、正しい向きにおける出力は小さく、学習さ

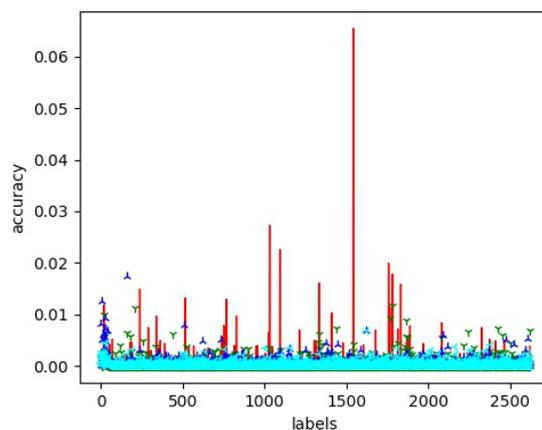


Fig. 4: 入力画像の傾きに対する最終層のニューロンの出力 (赤: $\theta = 0$ 、青: $\theta = 180$ 、緑: $\theta = -90$ 、シアン: $\theta = 90$)

れた画像 Fig.1 のときより傾きの区別が難しいことが分かる。

4 まとめ

学習された入力画像、学習されていない入力画像ともに、回転角度 ± 90 度および 180 度において、画像の傾きを検出することができたが、学習されていない画像では補正の精度が低下した。今後の課題として、より汎用性を高めるために、実際のカメラから取得した画像について提案手法の評価を行い、どのような条件で傾きを検出できなくなるのか、またどの程度の傾きまで補正できるのか検討する。さらに、動物や風景に対して補正できるように、A.Krizhevsky らによる学習済みモデル ImageNet[3] を用いて同様の実験を行う。

参考文献

- [1] 岡谷貴之, 深層学習 Deep Learning, 講談社.
- [2] M. Parkhi, A. Vedaldi, A. Zisserman, "Deep Face Recognition," British Machine Vision Conference, 2015.
- [3] A. Krizhevsky, I. Sutskever, G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Neural Information Processing Systems Conference, 2012.