視線情報を用いたアンサンブル推論による詳細画像識別の高精度化

ER17078 谷中 亮佑　　指導教授：藤吉弘亘

1. はじめに

詳細画像識別は，同一カテゴリ内において，視覚的・意味的に類似している物体の種類を識別するタスクである．詳細画像識別の対象クラスは，人間が認識することさえ困難なほど視覚的差異がない場合がある．そのため，詳細画像識別では，対象物体の全体ではなく，特徴的な領域を捉える必要がある．そこで，本研究では，視線情報を用いたアンサンブル推論法について提案する．

2. アンサンブル推論

アンサンブル推論は複数の出力を加算し平均することにより汎化性能を向上させる手法である．推論時に複数の画像を入力し，それらの結果を統合して最終結果を出力する手法は Test Time Augmentation(TTA) と呼ばれる．TTA によるアンサンブル推論は式 (1) のようになる．

$$y = \frac{1}{M}\sum_{m=1}^{M} p_m(x) \tag{1}$$

ここで，M はデータ数，p_m はネットワークの事後確率，x は入力データを表す．一般物体認識では，ランダムに画像をクロップしたりコントラストを変えたりして，TTA を行う．しかし，詳細画像識別の場合，異なるクラス間で特徴が類似していることが多く，ランダムクロップによる TTA は効果が低いという問題がある．

3. 提案手法

本研究では，アンサンブル推論時に人間の視線情報を利用した TTA を導入した詳細画像識別手法を提案する．

3.1 視線情報によるサンプリング

詳細物体識別において，人の知見を用いて重要となる特徴領域を与えることは有効であると考える．そこで本手法では人間の視線情報を利用する．あらかじめアイトラッカーを用いてデータセットの各画像に対し，視線情報を取得する．次に，得られた視線情報からサッケードを除外し，対象物体に対する画像上の注目座標データを求める．学習時及び推論時，注目座標データに対し Mean Shift クラスタリングを行い注視点クラスタを求め，クラスタの極大位置をクロップの中心位置とする．これらのクロップ画像を用いることで詳細物体識別に重要となる特徴領域を捉えた学習及び推論を可能にする．画像のクロップ時に入力画像外にはみ出す場合，領域外を [0,255] の値で埋める erasing，及び領域内をにリサイズする resize の 2 種類のいずれかを行う．これにより，画像クロップ時に発生するゼロパディングによる精度低下を抑える．視線情報を用いた固視点の密度に応じた画像のクロップを図 1 に示す．

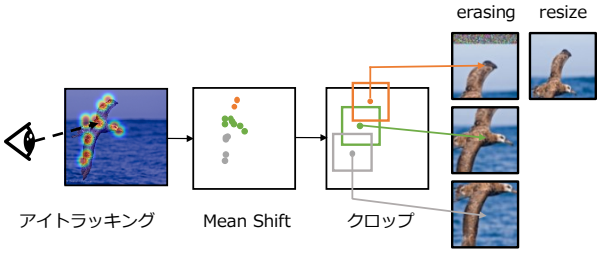


図 1: 固視点の密度に応じた画像のクロップ

3.2 Mean Shift クラスタリング

Mean Shift とは，近傍の点群の平均位置に移動を繰り返し，集合の極大点を求める方法である．繰り返しステップにおける極大点探索を式 (2) に示す．

$$s(x) = m(x) - x = \frac{\sum_{i=1}^{n} K(x_i; x, h)x_i}{\sum_{i=1}^{n} K(x_i; x, h)} - x \tag{2}$$

ここで，K はカーネル関数，x_i は点の集合，h はバンド幅を表し．$s(x)$ は極大点に近づくほど値が小さくなる．各点 x_i を収束先の極大点ごとにラベル付けすることでクラスタに分割し，それらのクラスタの極大位置を取得する．

3.3 視線情報を用いた学習と TTA

本手法では，視線情報を用いてクロップした画像を学習データとする．そして，推論では図 2 に示すように，視線情報を用いて TTA を行う．本手法でのクロップ枚数は 4 枚とする．アンサンブル推論は，式 (1) に従い CNN の出力から平均クラス確率を求める．

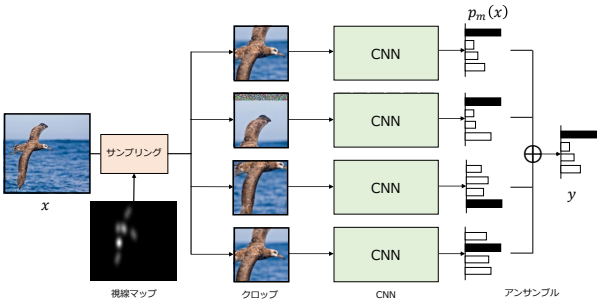


図 2: 視線情報を用いた TTA の概要

4. 評価実験

提案手法の有効性を検証するために評価実験を行う．

4.1 実験条件

本実験では，Caltech-UCSD Birds(CUB-200-2011)[1] データセットに，視線情報を付与し，50 クラス，2,880 枚を用いて実験を行う．学習用に 2,015 枚，評価用に 865 枚を用いる．画像サイズは 256×256 画素にリサイズした後，視線情報をもとに 112×112 画素にクロップする．学習モデルには ResNet18 を使用する．

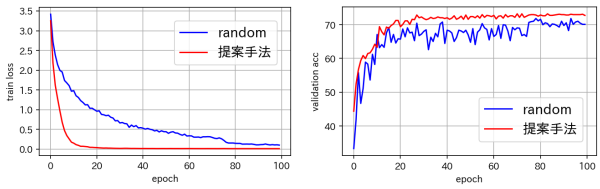
4.2 実験結果

表 1 に視線情報とランダムな TTA を比較対象とした結果を示す．TTA の各結果は 5 回試行したときの正解率の平均値と標準偏差である．表 1 より提案手法は，ランダムな TTA と比較して，1.71 ポイント精度が向上した．

表 1: クロップサイズ変更による認識精度 [%]

TTA なし	TTA あり		
	ランダム	視線情報	
		erasing	resize
71.52	72.85 ± 0.93	**74.56 ± 0.73**	73.86 ± 0.71

また，ランダムと視線情報を用いた場合の学習曲線を図 3 に示す．図 3 より，視線情報を用いた提案手法は学習の早期で収束している．これより，提案手法は少ないエポックでクラス特有の詳細な特徴を効率的に学習できていることがわかる．



(a) 損失の推移　　**(b)** 認識率の推移

図 3: 学習曲線

5. おわりに

本研究では，視線情報を用いたアンサンブルによる詳細物体識別法を提案した．今後の展望として，視線推定モデルを用いた一貫学習による TTA を検討する．

参考文献

[1] Wah. C, *et al.* “The Caltech-UCSD Birds-200-2011 Dataset”, Technical Report CNS-TR-2011-001, Caltech, 2011.