H-030

非厳密な領域アノテーションによる 畳み込みニューラルネットワークの一般画像分類精度の向上
Improvement of Generic Image Classification Accuracy via Convolutional Neural Networks with Non-strict Attentional Region Annotation

荒井 敏† 白川 真一† 長尾 智晴†
Satoshi Arai Shinichi Shirakawa Tomoharu Nagao

1. はじめに

機械学習による画像分類では一般に訓練データ数が多いほど高精度となるが，実際には入手性やコストの観点からデータ数を増やすことが難しい場合がある．限られたデータ数でより高精度を得るため，画像データに対してカテゴリラベル以外のアノテーションを追加し，モデルに複数のタスクを同時に学習させることで分類精度を向上させる方法が知られている．追加のアノテーションとしては物体認識を想定した外接矩形や領域分割のための物体輪郭などが用いられるが，これらを付与するための作業コスト自体が大きく，より簡便な手法が求められている．

筆者らはその様なアノテーション手法として Non-strict Attentional Region Annotation (NARA) [1] を提案し，STL-10 データセット [2] において畳み込みニューラルネットワーク(CNN)の分類精度が向上することを確認している．本研究ではより一般的な被写体を対象とする画像データセットに NARA を適用しその効果を検証した．

2. NARA

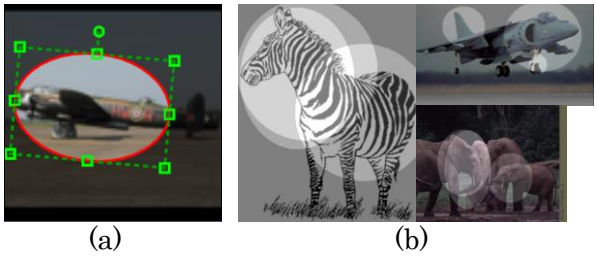
2.1 領域アノテーション

NARA [1] では人間が画像のカテゴリを判断する際に最も注目した領域に対し，単純な図形マーカを用いてアノテーションを付与する．どの領域に注目するかは個人の主観に依存するため，作業者によって異なる領域となり得る．複数のアノテーション結果が得られる場合は画素毎に選択回数に応じた重みを与えることで，より選択されやすい領域を重視する様にする．被写体の外形や輪郭を正確に囲まなくともよいため，作業者の負担が軽減される．これらの点において NARA は従来用いられてきた外接矩形や物体輪郭などの厳密な領域アノテーションとは思想が異なる．

アノテーション作業は一般的なドローツールでも可能だが，筆者らは自作の描画ツールを使用している(図 1(a))．図形マーカとして単一の楕円を用いた場合の例を図 1(b)に示す．各画像に対して異なる 3 名の作業者によるアノテーション結果を明領域として示している．

2.2 領域アノテーションを用いた学習

領域アノテーションの結果を CNN の学習に利用する．ここで対象とする CNN は最終層として線形層(Linear)，その直前に大域平均値プーリング層(global average pooling, GAP)を持つ構造を想定する(図 2(a))．Linear と GAP はいずれも線形演算であるため順番を入れ替えても等価な処理となる(図 2(b))．但し，入れ替え後に Linear はフィルタサイズ 1×1 の畳み込み層(Conv)に置き換える．これらの操作は



(a) (b)

図 1 NARA による領域アノテーション

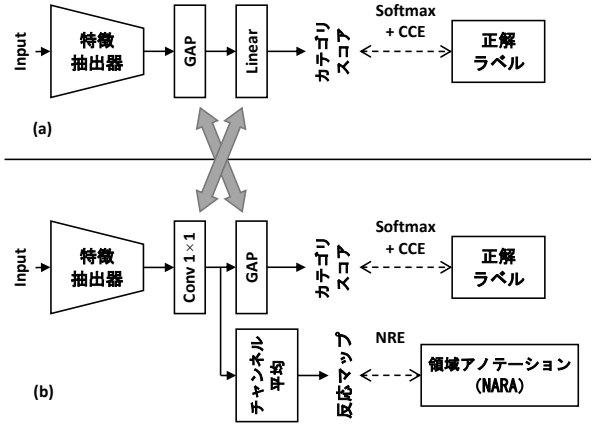


図 2 モデル構造の変換

学習後に等価な関係を維持したまま元に戻すことができる．

領域アノテーションを学習に反映させるため，通常のカテゴリスコアを出力し正解ラベルと照合する経路とは別に，領域アノテーションとの類似性を評価する経路を設ける．この経路は GAP の直前で分岐し，空間サイズを維持したままチャンネル方向に平均することで反応マップを得る．反応マップと領域アノテーションのノルムをそれぞれ正規化し，それらの平均二乗誤差を指標として算出する．この評価指標を正規化反応誤差(normalized response error, NRE)と称する．カテゴリスコアのためのクロスエントロピー損失 (CCE)と NRE を加算した値を系全体の損失としてモデルを学習する．NRE を与えることで，単に分類カテゴリを学習するだけではなく，人間の注目領域に類似した反応を模倣するモデルが獲得される．

3. 実験と考察

3.1 データセットの作成

一般画像を用いた公開データセットである Harvesting Image Databases [3] (以下 Harvesting DB)に対して，NARA の考え方に基づく領域アノテーションを付与した．

† 横浜国立大学 Yokohama National University

Harvesting DB は 18 カテゴリの画像分類用データセットであり，Web からキーワードベースで収集した画像を基にカテゴリに適合しない画像を目視で除外したものである．各カテゴリには最小 236 枚から最大 1,267 枚の画像が含まれ，全体で 14,391 枚の画像から構成されている．

アノテーション作業には Amazon Mechanical Turk を使用し，訓練されていない非専門家が担当した．具体的な作業内容として，先ず被写体のカテゴリを判断して選択し，更に判断に際して最も注目した領域にマーカを描画するよう指示した．但し，マーカが画像全体に広がることがない様，楕円マーカの長径が画像の対角線長の半分を超えないようにツール側で制限している．事前に描画ツールの習熟時間は設けていない．画像 1 枚当たりのカテゴリ判断とアノテーションに要した合計時間の中央値は 25 秒，最頻値は 14 秒であり，負荷的に軽い作業であることがわかる．各画像に 3 人の作業者がアノテーションを付与し，延べ 43,173 枚のアノテーション結果を得た． 図 1(b)に結果の例を示す．これらのアノテーションデータは本プロジェクトの Web ページからダウンロードできる[1]．

3.2 画像分類精度の評価

NARA の効果を確認するため，NARA を使用した場合(NARA あり)と使用しなかった場合(NARA なし)についてそれぞれモデルを学習し，分類精度を比較した．評価用の CNN モデルには ResNet18 [4] を使用した．

画像を入力する際，前処理およびデータ拡張として，①入力画像を zero padding によって正方形化，②画像の 1/2 から 1 倍までのサイズの領域をランダムに切り出し 224×224 画素にリサイズ，③確率 0.5 で左右反転，の処理を順に行っている．

学習率の初期値は 0.1 とし，60 epoch と 90 epoch でそれぞれ 0.1 倍に切り下げながら確率的勾配降下法を用いて 120 epoch まで学習した．

5 分割交差検証により平均精度を求めたところ，NARA なしと NARA ありの分類精度はそれぞれ 72.52%と 76.65%であった．NARA を用いることで分類精度が 4%以上向上していることがわかる．

カテゴリごとの分類精度を図 3 に示す．折れ線は NARA なしを基準とした NARA ありとの分類精度の差(向上幅)を表す．カテゴリによって効果に差はあるが，全てのカテゴリおいて分類精度が向上している．但し，このグラフからは向上幅の大小についてカテゴリの影響は読み取れない．最も向上幅が大きいカテゴリは camel であるが，同様に四足歩行動物のカテゴリである zebra では反対に向上幅が最も小さかった． カテゴリ以外の影響が大きいと推測される．

アノテーション領域の面積と分類精度の関係を図 4 に示す．横軸はアノテーション領域の面積(但し画像の面積で正規化済み)を表し，各区間に属するデータ群の NARA あり／なしでの分類精度を表す．折れ線は図 3 と同様に分類精度の向上幅を示している．アノテーション面積が小さいデータにおいて向上幅がより大きい傾向が見られる． これらは人間がより狭い範囲に注目した画像であり，どこに注目するかという情報がより重要な画像であると考えられる．画像中の被写体が相対的に小さい場合やカテゴリにとって

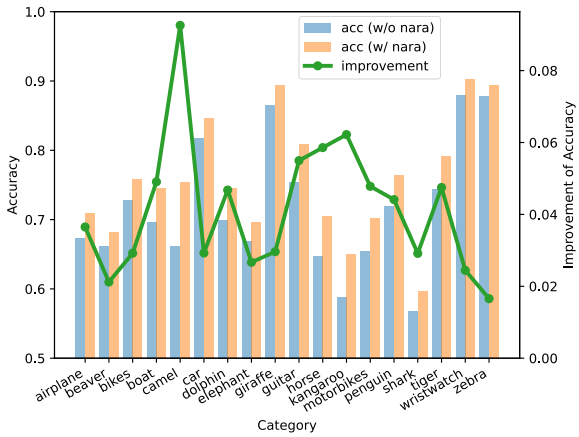


図 3　カテゴリごとの分類精度

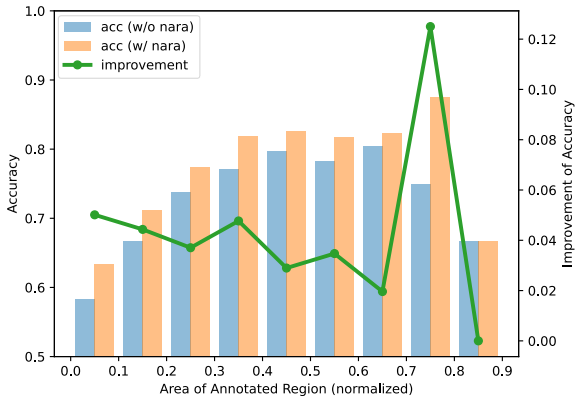


図 4　アノテーション面積ごとの分類精度

特徴的な部位が狭い範囲に集約されている場合などがこれに相当する．例外的に区間[0.7, 0.8]における向上幅が大きいが，これはこの区間のデータ数が数例程度と少なく，偶発的な偏りが生じたためと考えられる．

4. おわりに

STL-10 より一般的な被写体の画像データセットである Harvesting DB の分類においても NARA は有効であり，物体サイズが小さい場合に精度向上幅がより大きいことを示した．この知見を活かして小物体の場合のみアノテーションを付与するなどの更なる省力化が期待される．

謝辞

この成果は，国立研究開発法人新エネルギー・産業技術総合開発機構(NEDO)の委託業務(JPNP20001221-0)の結果得られたものです．

参考文献

[1] S. Arai, S. Shirakawa, and T. Nagao, “Non-strict Attentional Region Annotation to Improve Image Classification Accuracy”, IEEE SMC, 2021. (https://github.com/EMI-XAI/stl10_nara)
[2] A. Coates, H. Lee, and A. Y. Ng, “An Analysis of Single Layer Networks in Unsupervised Feature Learning”, AISTATS, 2011.
[3] F. Schroff, A. Criminisi, and A. Zisserman, “Harvesting Image Databases from the Web”, IEEE TPAMI, Vol.33, No.4, pp.754-766, 2011. (https://www.robots.ox.ac.uk/~vgg/data/mkdb/index.html)
[4] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition”, Proc. IEEE CVPR, pp.770-778, 2016.

[1] https://github.com/EMI-XAI/harvestingdb_nara