Attention mining branch を導入した ABN によるアテンションマップの最適化

ER17009 岩吉 孝明　　指導教授：藤吉 弘亘

1. はじめに

Attention Branch Network (ABN)[1] は，認識時におけるネットワークの注視領域を可視化し，認識処理に活用することで，高精度化を実現している．認識対象の物体以外に注視領域が発生すると，誤認識を誘発することがある．これを解決する手法として，人の知見により修正したアテンションマップを用いて ABN を再学習する手法 [2] が提案されている．しかしながら，本手法はアテンションマップを人が修正するため人的コストが掛かる．本研究では，ABN に Attention mining branch を導入し，アテンションマップを自動で最適化する手法を提案する．

2. Attention Branch Network

ABN は，Feature extractor，Attention branch，Perception branch から構成される．Feature extractor は入力画像から特徴マップを抽出する．これを Attention branch に与え，アテンションマップを獲得する．アテンションマップを特徴マップに乗算し，Perception branch に与えることで認識結果を出力する．ABN を応用した研究として，人の知見を導入する手法がある．この手法では，ABN で誤認識が発生した画像のアテンションマップを人の知見に基づいてすべて修正する．その後，それらを学習データに追加して再学習することより認識精度が向上する．

3. 提案手法

本研究では，ABN に Attention mining branch を導入したアテンションマップの最適化手法を提案する．提案手法のネットワーク構造を図 1 に示す．

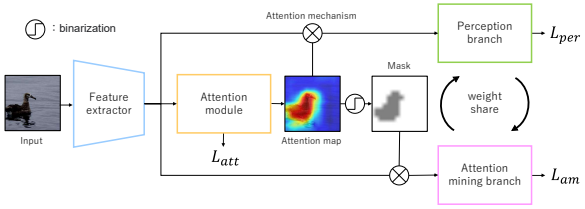

図 1：提案手法

提案手法では，ABN に Perception branch と重みを共有した Attention mining branch を追加する．Attention mining branch には，Mask を適用した特徴マップを入力し，クラス確率を出力する．ここで，Mask は Attention module で獲得したアテンションマップを 2 値化して生成する．この Mask を特徴マップに乗算することで注視した領域を隠した特徴マップを獲得できる．Mask 適用後の特徴マップ F^{*c} は，特徴マップを F，Mask を $T(A^c)$ とすると式 (1) のように表される．

$$F^{*c} = F - (T(A^c) \odot F) \tag{1}$$

提案手法では，Attention mining branch から出力した各サンプルのクラス確率の総和を新たな損失とする．これにより，Mask を適用していない領域のクラス確率を最小化するように学習し，認識対象の物体のみを注視するようにアテンションマップを最適化する．Attention mining branch の損失 L_{am} は，クラス確率を $S_i^c(F^{*c})$，サンプル数を n とすると式 (2) のように表される．

$$L_{am} = \sum_{i=1}^{n} S_i^c(F^{*c}) \tag{2}$$

Attention module の出力と正解クラスのクロスエントロピー誤差を L_{att}，Perception branch の出力と正解クラスのクロスエントロピー誤差を L_{per}，L_{am} の重みを α とすると，提案手法の損失関数 L は式 (3) のように表される．

$$L = L_{att} + L_{per} + \alpha L_{am} \tag{3}$$

提案手法では，段階的に学習する．まず ABN のみを事前学習する．そして，Attention mining branch を追加してアテンションマップの最適化となる再学習を行う．

4. 評価実験

提案手法の有効性を評価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 200-2010 (CUB-200-2010) データセットと，Stanford Dogs データセットを用いる．ベースネットワークとして ResNet-50 を使用し，バッチサイズは 16 とする．Mask の閾値は CUB-200-2010 データセットで 0.83，Stanford Dogs データセットで 0.40 とする．L_{am} の係数 α は 0.0001 に設定する．学習の更新回数は ABN の事前学習，提案手法それぞれで 300epoch とする．

4.2 実験結果

各データセットにおける認識精度の比較を表 1 に示す．表 1 より，CUB-200-2010 データセットにおいて，提案手法は，人の知見を導入する手法よりも Top-1 の認識精度が低くなっているが，ABN よりも向上していることが確認できる．また，Stanford Dogs データセットにおいて，提案手法が ABN よりも，Top-1 の認識精度が向上していることが確認できる．

表 1：認識精度の比較 [%]

Model	CUB-200-2010		Stanford Dogs	
	Top-1 acc.	Top-5 acc.	Top-1 acc.	Top-5 acc.
ABN	31.68	57.01	71.81	**93.02**
提案手法	**33.33**	**58.56**	**71.99**	92.80
人の知見	37.42	62.08	-	-

4.3 アテンションマップの可視化

ABN，提案手法，人の知見を導入する手法の Attention map の比較を図 2 に示す．図 2 (a)，(b) より，人の知見を導入する手法は，より局所的な領域に着目してクラス識別ができている．一方で，図 2 (c) のように，人の知見を反映させすぎると誤認識を誘発する事がある．これに対し，提案手法は広範囲を注視できているため正しく認識している．

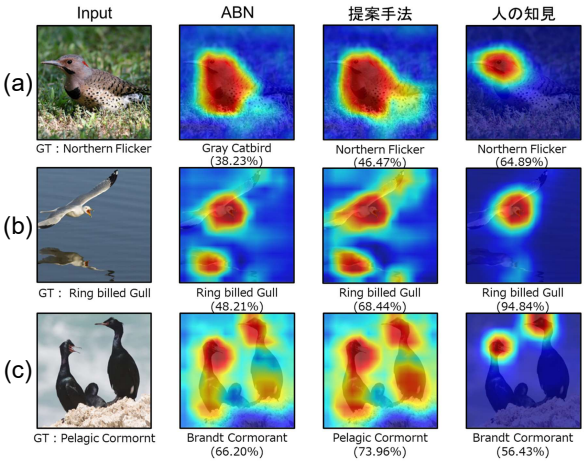

図 2：アテンションマップの可視化例

5. おわりに

本研究では，Attention mining branch を追加した ABN を提案した．評価実験では，提案手法を用いることにより，認識精度が向上した．また，注視領域についても改善されていることが確認できた．今後は，他のデータセットに対して提案手法の評価を行う．

参考文献

[1] H. Fukui, *et al.*, “Attention Branch Network:Learning of Attention Mechanism for Visual Explanation”, CVPR, 2019.

[2] M. Mitsuhara, *et al.*, “Embedding Human Knowledge into Deep Neural Network via Attention Map”, VISAPP, 2021.