Complement Objective Training によるアテンションマップの最適化に関する研究

TR21002 岩吉 孝明　　指導教授：藤吉弘亘

1. はじめに

Attention Branch Network (ABN)[1] は，ネットワークの注視領域をアテンションマップとして可視化するだけでなく，認識処理に活用することで高精度な画像認識を実現した．一方，鳥の種類のようなサブカテゴリを分類する詳細画像分類では，クラス間の相違が小さいため，より識別に有効な領域を注視して分類する必要がある．

本研究では，識別に有効な領域をアテンションマップとして獲得するために，1. 認識に有効な領域の探索，2. 特徴空間の改善，3. 出力空間の改善の 3 つのアプローチを提案する．具体的には，ABN に有効な注視領域を探索する Attention mining branch (AMB)，特徴空間を改善する Prototype conformity loss (PC Loss) と出力空間を改善する Complement objective training (COT) を導入する．

2. Attention Branch Network

ABN は，入力画像から特徴マップを獲得する Feature extractor，アテンションマップを獲得する Attention branch，最終的な認識結果を出力する Perception branch の 3 つから構成されている．Attention branch で獲得したアテンションマップは，Attention 機構により Feature extractor で得られた特徴マップに重みづけする．これにより，重要な領域を強調した特徴が Perception branch に入力する．しかしながら，クラスの異なる物体が複数存在する場合，ABN のアテンションマップは認識対象以外の物体を注目することがあり，認識精度を劣化させる可能性がある．この問題に対し，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 (人の知見) [2] が提案された．ABN で誤認識した際のアテンションマップを人の知見を介して修正し，理想的なアテンションマップを作成して，ネットワークを再学習する．しかし，アテンションマップの修正には人的コストがかかるという問題がある．

3. 提案手法

本研究では学習によって詳細画像分類に有効な注視領域を自動で獲得することを目的とする．提案手法は，ABN に 3 つのアプローチを導入し，詳細画像分類に有効な領域の獲得を目指す．1 つ目に，ABN に Attention mining branch (AMB) を導入し，認識に有効な領域のみに注視するよう学習する．2 つ目に，ABN に Prototype conformity loss (PC Loss) を導入して特徴空間を改善する．PC Loss は，特徴空間において同じクラスの特徴量を近づけ，異なるクラスの特徴量を離すように学習することから，クラス特有の注視領域の獲得を可能とする．更に，注視領域の改善により，認識精度の向上も見込めると考える．3 つ目に，AMB に不正解クラスを平坦化する Complement objective training (COT) を導入し，不正解クラス確率を抑えることで，認識対象のみの注視領域の獲得を目指す．

3.1. Attention mining branch

提案手法は，入力画像から特徴マップを獲得する Feature extractor，アテンションマップを獲得する Attention module，最終的な認識結果を出力する Perception branch，に加え，識別に有効な領域を探索する AMB で構成されている．AMB は，認識に有効な領域のみに注視するように学習を行う．AMB は，Perception branch と同様の構造であり，アテンションが閾値より高い領域をマスク処理した特徴マップを入力し，クラス確率を出力する．このとき，対象クラスのクラス確率が低いほど，アテンション領域が認識対象に影響しているといえる．そこで，対象クラスのクラス確率を最小化するよう学習する．これにより，認識対象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．また，AMB は，Perception branch と重みを共有する．これにより，認識対象の物体のみを注視するように学習した AMB の重みを最終的な認識結果を出力する Perception branch に利用することができる．

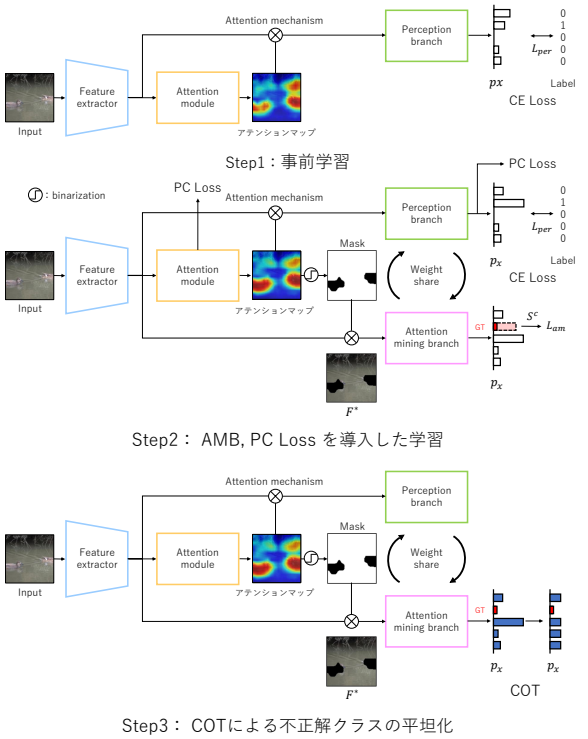

図 1: 提案手法における学習の流れ

3.2. Prototype conformity loss

特徴空間の改善として，PC Loss を Perception branch と Attention module に導入する．PC Loss は同じクラス内の特徴量を近づけ，異なるクラス間の特徴量を離す損失関数である．PC Loss L_{PC} は，サンプル数を N，クラス数を k，サンプル i の正解クラスを g，不正解クラスを j，学習可能なクラス重心 w^c と特徴量 f_i，$\alpha_i = \|f_i - w_g^c\|_2$，$\beta_i = \|f_i - w_j^c\|_2$，$\gamma_i = \|w_g^c - w_j^c\|_2$，とすると，式 (1) で表現できる．

$$L_{PC} = \sum_{i}^{N} \{\alpha_i - \frac{1}{k-1} \sum_{j \neq g} (\beta_i + \gamma_i)\} \tag{1}$$

これを最小化することで，α_i は特徴量を正解クラスの重心に集め，$(\beta_i + \gamma_i)$ はクラス間を離すように働く．これにより，クラスごとに特有の特徴が獲得され，その結果，より良いアテンションマップを獲得できると考える．

3.3. Complement objective training

AMB の出力空間の改善として，COT を導入する．COT は Complement entropy が最小になるよう重みを更新することで，不正解クラス確率を平坦化する．Complement entropy $C(\hat{y}_{\bar{C}})$ は，$\hat{y}$ を入力 x_i に対するクラス確率，g を正解クラスのインデックスとすると，式 (2) のように表される．

$$C(\hat{y}_{\bar{C}}) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1, j \neq g}^{K} \frac{\hat{y}_{ij}}{1 - \hat{y}_{ig}} \log \left(\frac{\hat{y}_{ij}}{1 - \hat{y}_{ig}} \right) \tag{2}$$

AMB に COT を導入する際，$\hat{y}_{ig} = 0$ とすることで，正解クラス確率が 0，残りの不正解クラス確率が平坦になるよう算出される．これにより，注視領域に悪影響を及ぼす突出した不正解クラス確率を削減する．

3.4. 学習手順

提案手法における学習の流れを図 1 に示す．提案手法の学習手順を以下に述べる．

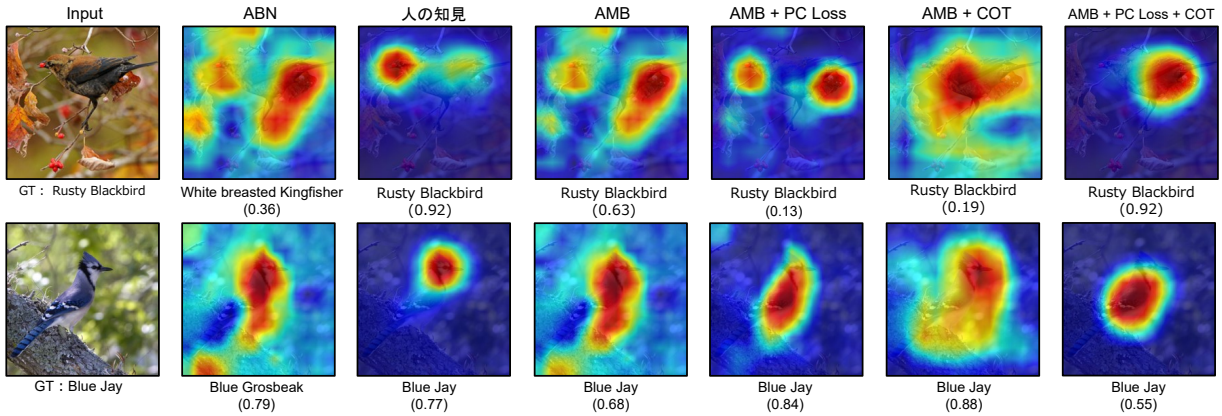

図 2: アテンションマップの例

Step1 ABN と同様のモデル構造で事前学習を行う．

Step2 ABN の損失に Attention module，Perception branch から算出した PC Loss を加え，最小になるよう学習する．COT を用いない場合は，AMB の損失も加える．

Step3 AMB の出力から Complement entropy を算出し，最小になるよう学習する．

Step2, 3 をイタレーションごとに繰り返す．

4. 評価実験

提案手法の有効性を評価するために詳細画像分類タスクにて評価実験を行う．

4.1. 実験条件

本実験では，200 クラスの鳥の画像で構成される Caltech-UCSD Birds 200-2010 (CUB-200-2010) 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．マスク処理の閾値は 0.78 とし，学習の更新回数は ABN の事前学習，提案手法それぞれで 300 エポックとする．提案手法との比較として，ABN, 人の知見を用いる．

4.2. 実験結果

各手法における分類精度の比較を表 1 に示す．表 1 より，AMB は，ABN より分類精度が向上するが，人の知見より低い．PC Loss もしくは COT を導入すると，人の知見よりも分類精度が向上することを確認した．また，PC Loss と COT の両者を AMB に組み合わせることで分類精度が更に向上しており，ABN より Top-1 の分類精度が 14.41 ポイント高い結果となることをことを確認した．

表 1: 認識精度の比較 [%]

	AMB	PC Loss	COT	Top-1 acc.	Top-5 acc.
ABN [1]	-	-	-	31.68	57.01
人の知見 [2]	-	-	-	37.42	62.08
提案手法	✓	-	-	33.33	58.56
	✓	✓	-	45.10	**71.68**
	✓	-	✓	39.76	66.57
	✓	✓	✓	**46.09**	69.24

4.3. アテンションマップの可視化

各手法におけるアテンションマップの比較を図 2 に示す．アテンションマップの下に，モデルが認識したクラスとそのクラス確率を示す．図 2 に示すように，提案手法により ABN で獲得していた認識対象以外の領域，すなわち不要な注視領域が軽減されていることが分かる．また，PC Loss を導入することで，ABN より局所的な領域に着目していることが分かる．一方で，AMB，COT は認識対象を広域に捉えつつ，不要な注視領域を軽減していることが分かる．また，COT は PC Loss より高いクラス確率を獲得していることから，PC Loss では捉えられていない重要な領域を獲得できていると考えられる．

4.4. アテンションマップの定量的評価

各手法における Insertion, Deletion 結果を表 2 に示す．提案手法により，ABN や人の知見と比べスコアが向上しており，認識に有効な領域を捉えられるといえる．また，PC Loss と COT を組み合わせることでスコアが更に高くなった．これらのことから，AMB に PC Loss と COT を組み合わせることで認識に有効な領域を獲得し，それに伴い認識精度が向上したことが分かった．

表 2: Insertion，Deletion 結果

	AMB	PC Loss	COT	Insertion↑	Deletion↓	Score↑
ABN [1]	-	-	-	0.21	0.09	0.12
人の知見 [2]	-	-	-	0.17	0.16	0.01
提案手法	✓	-	-	0.22	**0.08**	0.14
	✓	✓	-	0.29	0.13	0.17
	✓	-	✓	0.28	0.12	0.16
	✓	✓	✓	**0.33**	0.13	**0.21**

5. おわりに

本研究では，ABN に注視領域が認識に有効か考慮しながら学習する AMB，特徴空間を改善する PC Loss，出力空間を改善する COT の 3 つのアプローチにより ABN の注視領域を改善する手法を提案した．評価実験では，提案手法を用いることにより，不要な注視領域を軽減し，認識精度が向上したことを確認した．今後は，学習方法やモデル構造の検討による更なる注視領域の改善を行う．

参考文献

[1] H. Fukui, *et al.*, “Attention branch network: Learning of attention mecha-nism for visual explanation”, CVPR, 2019.

[2] M. Mitsuhara, *et al.*, “Embedding human knowledge in deep neural network via attention map”. VISAPP, 2021.

研究業績

[1] 岩吉孝明, 等,“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”, 第 24 回画像の認識・理解シンポジウム, 2021.

[2] T. Iwayoshi, *et al.*, “Attention mining branch for Optimizing Attention Map”, MVA, 2021.

[3] T. Iwayoshi, *et al.*, “Complement Objective Mining Branch for Optimizing Attention Map”, VISAPP, 2023.