1U-08

CNNに対する可視化手法の計算機実験による比較評価

下村 真生† 中村 和幸†‡

†明治大学 ‡国立研究開発法人科学技術振興機構，さきがけ

1 はじめに

本稿では，深層学習を用いた画像分類における判断根拠の可視化技術である，Grad-CAM等の可視化手法を比較評価する指標として筆者が提案したBlack Average Dropについて，その効果検証を行う．

CNNを用いた画像分類で，その判断根拠を可視化する手法としてGrad-CAM[1]やGrad-CAM++[2]などがある．可視化手法は増加傾向にあり，今後はデータセットに応じて選択する必要がある．画像セグメンテーションで使われるmean Intersection over Union (mIoU)等の評価指標はピクセル単位のラベルが必要だが，医用画像等，ラベル付けが難しい分野もあるため，ピクセル単位のラベルなしで評価できる指標が必要である．筆者が提案したBlack Average Drop (BAD)は先行指標であるAverage Drop (AD)[2]と同じく分類スコアベースで評価するが，評価時に使用する加工画像についてアルファ値を調整せずに等積のマスクで覆うように定義を変更することで，医用画像など適用可能なデータセットを増やすことが可能となった．

2 先行研究

2.1 可視化手法：CAMs

Class Activation Mapping（CAM）[3]は，CNNの最後の畳み込み層の出力として得られる特徴量マップA^kと，それにGlobal Average Poolingを施し，そのk番目のノードからクラスcへの重みw_k^cから顕著量マップL^cを作る:

$$L^c = \sum_k w_k^c A^k. \tag{1}$$

このL^cを入力画像のサイズまで拡大し，入力画像と重ね合わせることにより可視化を行う．ここで，w_k^cを出力層への入力から特徴量マップへの勾配

$$w_k^c = \frac{1}{Z}\sum_x\sum_y \frac{\partial y^c}{\partial A_{xy}^c}, \tag{2}$$

として，適用可能なCNNの制限を緩和したものがGrad-CAMである．しかし，これは勾配を特徴量マップの面積Zで割るため，小さい特徴を見逃す可能性がある．これに対し，Grad-CAM++では，特徴量マップに含まれる特徴の面積に関する値αで重み付けを行う．

$$w_k^c = \sum_x\sum_y \alpha_{xy}^c \mathrm{relu}\left(\frac{\partial Y^c}{\partial A_{xy}^c}\right), \tag{3}$$

$$\alpha_{xy}^c = \frac{\frac{\partial^2 Y^c}{\left(\partial A_{xy}^k\right)^2}}{2\frac{\partial^2 Y^c}{\left(\partial A_{xy}^k\right)^2} + \sum_a\sum_b A_{ab}^k\left\{\frac{\partial^c Y^c}{\left(\partial A_{xy}^k\right)^3}\right\}}. \tag{4}$$

これにより，全特徴量マップに現れる様々な特徴を平等に評価し，可視化に反映することができる．特に医用画像の例では小病変も重要であるためGrad-CAM++を適用すべきであることが定義から示唆される．

2.2 評価指標

画像データセットに対して，可視化手法がどれほど適するかを評価するAD，BADの評価式は次である：

$$\frac{1}{N}\sum_i^N \frac{\max\left(0, Y^c - O^c\right)}{Y^c}. \tag{5}$$

但しN：テストデータ数，Y^c，O^c：それぞれテスト画像Iと後述の加工画像BをCNNに通した際の分類スコアである．ADでの加工画像Eは顕著量マップをテスト画像のサイズまで拡大した$\tilde{L}^c$とテスト画像Iから$E^c = \tilde{L}^c \circ I$により定義される．

一方，BADでの加工画像Bは次のフローで得られる．

1. $\tilde{L}^c$の各ピクセル値を降順にソートする．
2. 上位β%となる値を閾値ηとする．
3. 加工画像Bの各ピクセル値を以下で決める．

$$B_{xy}^c = \begin{cases} I_{xy} & (L_{xy}^c > \eta) \\ 0 & (L_{xy}^c \leq \eta) \end{cases} \tag{6}$$

4. 加工画像Bを1.と同じCNNに入れた時の出力スコアをO^cとし，評価式で評価する．

但し，βは比較する可視化手法間のそれぞれに対して作る加工画像の差異が最大となるように定義し，可視化結果の差が最大の状態で評価を行う．

"Comparative Evaluation of Visualization for CNN by Computer Experiments"
by Masaki Shimomura†and Kazuyuki Nakamura†‡
†Meiji University, ‡JST, PRESTO

表 1: 高精度モデルに対する可視化評価比較

	Flowers			ChestX-rays		
	閾値	GCAM	GCAM++	閾値	GCAM	GCAM++
AD	-	**.43**	.60	-	**.8299**	.8304
BAD	.57	.79	**.30**	.30	.907	**.905**
mIoU	.71	.26	**.52**	.08	**.8906**	.8905

3 実験と考察

17 種の花画像 17 Category Flower Dataset(Flowers) [4] と 14 種の疾病を含む胸部 X 線画像 ChestX-ray14[5] から “肺炎” と “疾病の特徴なし” の画像 (ChestX-rays) を利用する．Flowers は 1360 枚の画像とその全てのピクセル情報があり，ChestX-rays は 1431 枚の肺炎画像とその一部の疾病位置が矩形で付されている．本節ではそれぞれ学習したモデルを基に以下の検証を行う．

3.1 可視化評価

Grad-CAM と Grad-CAM++のいずれが可視化に適するか，2 つデータセットに対して AD と BAD で評価した結果は表 1 上段である．なお最適と判断されるものを太字で表した．値が小さい程良く，Flowers と ChestX-rays のいずれも AD では Grad-CAM，BAD では Grad-CAM++が適するとの評価を得た．

3.2 mIoU との比較

BAD の評価結果が正しいものであることを確認するために mIoU:

$$\texttt{mIoU} = \frac{1}{N}\sum_{i=1}^{N}\frac{TP^{(i)}}{TP^{(i)}+TN^{(i)}+FP^{(i)}}, \qquad (7)$$

(それぞれ i 番目の画像における $TP^{(i)}$:真陽性，$TN^{(i)}$:真陰性, $FP^{(i)}$:偽陽性のピクセル数を表す) を用いた評価結果との比較を行う．予測の陰陽については，それぞれのピクセルラベルにおいて真に特徴を表す部分の平均の面積 M を求め，次に CAMs における $\tilde{L}^c$ のピクセル値を降順にソートした後に，上位 M ピクセルを陽性, それ以外を陰性とした．この定義による評価結果は表 1 下段である．値が大きい程良く，両データセットに対して Grad-CAM++が良いとの評価を得た．

Flowers に対しては，0.25 程度の差が付いており，Grad-CAM++が適することが示唆できたが，ChestX-rays に対してはほぼ差がつかなかった．これは，ChestX-rays に付されたピクセルラベルが矩形で疾病部位よりも大きく，CAMs 間の細かな違いを評価するには不十分なものであったことに起因する．

3.3 CNN 精度との関係

本指標では，CNN モデルの分類スコアを基に評価をするため，分類精度が十分高いことを要請する．図 2〜

図 1: 元画像

図 2: 精度 5%モデル

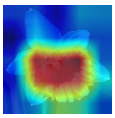

図 3: 精度 56%モデル

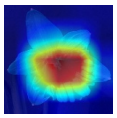

図 4: 精度 95%モデル

図 4 には，図 1 に対してモデル精度がそれぞれ 4.71%，55.98%と 95.30%のモデルを用いて Grad-CAM++で可視化をした結果を掲載した．低精度モデルを利用した場合, 特徴を捉えていないため正確な判断ができない．

4 おわりに

本稿では，筆者が提案した BAD についての効果を検証するために一般画像として Flowers，医用画像として ChestX-rays を利用して可視化の評価を行った．Flowers に対しては，mIoU の結果から Grad-CAM++が適するという示唆に対し，BAD は既存指標である AD と異なり，正確に判定をすることができた．一方で医用画像については，Grad-CAM より小病変を逃しにくいという Grad-CAM++の定義から得られた示唆に対して，BAD は正確な示唆を与えたものの，mIoU との比較について課題が残ったため，ピクセル単位のラベルを用いて再評価を行う必要がある．

謝辞 本研究は JST，さきがけ，JPMJPR1774，ならびに JSPS 科研費 JP19H04186 の助成を受けた．

参考文献

[1] R.R. Selvaraju *et al.*, Grad-cam: Visual explanations from deep networks via gradient-based localization, In *Proceedings of the IEEE International Conference on Computer Vision,* pp.618-626, 2017.

[2] A. Chattopadhay *et al.*, Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks, In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, pp.839-847, 2018.

[3] B. Zhou *et al.*,Learning deep features for discriminative localization, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.2921-2929, 2016.

[4] M. Nilsback and A. Zisserman, 17 Category Flower Dataset, `http://www.robots.ox.ac.uk/~vgg/data/flowers/17/`

[5] X. Wang *et al.*, Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.3462-3471, 2017.