Lambda Attention Branch Networksによる視覚的説明生成

Visual Explanation Generation Using Lambda Attention Branch Networks

飯田 紡 *1 Tsumugi Iida　兼田 寛大 *1 Kanta Kaneda　平川 翼 *2 Tsubasa Hirakawa　山下 隆義 *2 Takayoshi Yamashita　藤吉 弘亘 *2 Hironobu Fujiyoshi　杉浦 孔明 *1 Komei Sugiura

*1慶應義塾大学 Keio University　*2中部大学 Chubu University

Explanation generation for transformers enhances accountability for their predictions. However, there have been few studies on generating visual explanations for the transformers that use multidimensional context, such as LambdaNetworks. In this paper, we propose the Lambda Attention Branch Networks, which attend to important regions in detail and generate easily interpretable visual explanations. We also propose the Patch Insertion-Deletion score, an extension of the Insertion-Deletion score, as an effective evaluation metric for images with sparse important regions. Experimental results on two public datasets indicate that the proposed method successfully generates visual explanations.

1. はじめに

深層学習が幅広い分野に応用されている現代において，モデルの説明性は重要である．例えば，理論が未解明な自然現象の予測に深層学習を用いた場合，視覚的説明による重要な部分の可視化を通して，理論への洞察を与えることができる．

畳み込みニューラルネットワークにおける説明生成は盛んに研究されてきた [Fukui 19, Selvaraju 17, Wang 20]．一方，transformer，特に Lambda [Bello 21] に基づく transformer に対する説明生成を行った研究は少ない．また，視覚的説明のための標準的な評価指標（Insertion-Deletion score [Petsiuk 18]）は，スパースな重要領域を有する画像に対しては不適切な場合がある．

このような背景から，本論文では Lambda に基づく transformer に対して，容易に解釈可能な説明を生成できる，Lambda Attention Branch Networks (LABN) を提案する．本論文の概要図を図 1 に示す．提案手法は，Lambda Feature Extractor (LFE)，Lambda Attention Branch (LAB)，Lambda Perception Branch (LPB) の 3 モジュールから構成される．Lambda 層 [Bello 21] の注意機構を用いると，明瞭ではない説明が生成されることがある．一方で，説明生成に特化したブランチ構造を導入することで，LABN は明瞭な説明を獲得できる．

さらに，スパースな重要領域を有する画像に対して有効な評価指標として，Insertion-Deletion score を拡張した Patch Insertion-Deletion (PID) score を提案する．PID score は，Insertion-Deletion score とは異なり，パッチ単位で視覚的な説明を評価する．本研究の貢献を以下に示す．

- Lambda 層の注意機構による説明よりも明瞭な説明を獲得するため，ブランチ構造を持つ LABN を提案する．
- スパースな重要領域を有する画像に有効な評価指標として，Insertion-Deletion score を拡張した PID score を提案する．

2. 問題設定

本論文では，分類問題におけるモデルの判断根拠の視覚的説明生成として，画像中の重要な領域を可視化するタスクを扱う．特に，Lambda [Bello 21] に基づく transformer モデルにおける説明に焦点をあてる．本タスクでは，モデルが正しく予測するために貢献した画素を重要とする説明が望ましい．

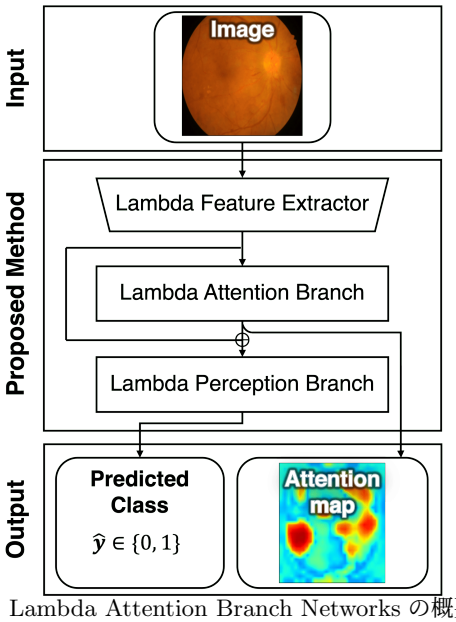


図 1: Lambda Attention Branch Networks の概要図

本論文では，以下の入出力を想定する．

- **入力:** 分類対象の画像 $\boldsymbol{x} \in \mathbb{R}^{c_1 \times w_1 \times h_1}$
- **出力:** 画像が各クラスに属する確率の予測値 $\hat{\boldsymbol{y}} \in \mathbb{R}^{C}$

ここで，c_1, w_1, h_1, C はそれぞれ入力画像のチャンネル数，横幅，縦幅，クラス数を表す．出力に加えて，モデル中の attention map $\boldsymbol{\alpha} \in \mathbb{R}^{w_1 \times h_1}$ として，画像の各画素における重要度が獲得できる．この attention map を視覚的説明として使用する．

本タスクの評価尺度には，PID score を用いる．PID score により，attention map とモデルの判断に寄与した領域の一致度を評価することができる．本論文では，Lambda に基づく transformer モデルを前提とする．また，attention map がクラスに依存しないことを仮定する．

連絡先: 飯田紡，慶應義塾大学，神奈川県横浜市港北区日吉 3-14-1，tiida@keio.jp

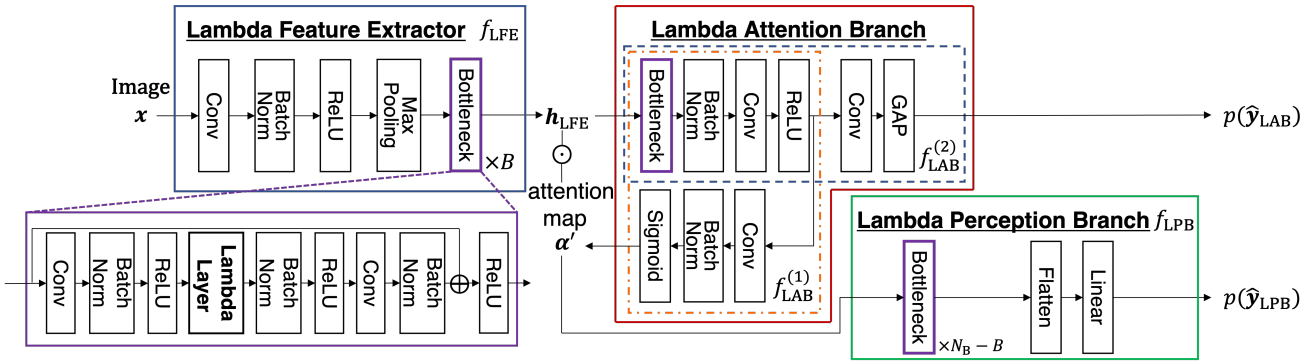


図 2: LABN のネットワーク図．“Conv” および “GAP” はそれぞれ畳み込み層と global average pooling 層を表す．

3. 提案手法

3.1 構造

ネットワークの構造を図 2 に示す．本ネットワークは，LFE, LAB, LPB の 3 モジュールから構成される．バックボーンネットワークが N_B 個の bottleneck 層を含むと仮定する．まず，B 番目のボトルネック層で，バックボーンネットワークを分割して，LFE と LPB とする．次に，LFE と LPB の間に並列に LAB を導入する．

LFE の入力は画像 $\boldsymbol{x}$ である．LFE は B 個の bottleneck 層やバッチ正規化層を含み，画像の特徴量を抽出する．Bottleneck 層は後述する Lambda 層，畳み込み層，バッチ正規化層，ReLU 活性化関数を含む．LFE の出力を $\boldsymbol{h}_{\mathrm{LFE}} \in \mathbb{R}^{c_2 \times w_2 \times h_2}$ とする．ここで，c_2, w_2, h_2 はそれぞれ LFE の出力のチャンネル数，横幅，縦幅を表す．

LAB は，$f^{(1)}_{\mathrm{LAB}}$ と $f^{(2)}_{\mathrm{LAB}}$ の 2 つに分かれる．$f^{(1)}_{\mathrm{LAB}}$ は attention map を生成する．$f^{(1)}_{\mathrm{LAB}}$ は bottleneck 層と global average pooling を含む．$f^{(1)}_{\mathrm{LAB}}$ の入力と出力は，それぞれ $\boldsymbol{h}_{\mathrm{LFE}}$ と $\tilde{\boldsymbol{\alpha}} \in \mathbb{R}^{w_2 \times h_2}$ である．$\tilde{\boldsymbol{\alpha}} \in \mathbb{R}^{w_2 \times h_2}$ を拡大して，視覚的説明として用いる $\boldsymbol{\alpha} \in \mathbb{R}^{w_1 \times h_1}$ を得る．$\tilde{\boldsymbol{\alpha}}$ のうち，$\theta_{\boldsymbol{\alpha}}$ より小さな値を 0 として最終的な $\boldsymbol{\alpha}' \in \mathbb{R}^{w_2 \times h_2}$ とする．ここで，$\theta_{\boldsymbol{\alpha}}$ は attention map のしきい値である．

$$\tilde{\boldsymbol{\alpha}} = f^{(1)}_{\mathrm{LAB}}\left(\boldsymbol{h}_{\mathrm{LFE}}\right) \tag{1}$$

$$\boldsymbol{\alpha}'_{ij} = \begin{cases} \tilde{\boldsymbol{\alpha}}_{ij} & (\theta_{\boldsymbol{\alpha}} < \tilde{\boldsymbol{\alpha}}_{ij}) \\ 0 & (\text{otherwise}) \end{cases} \tag{2}$$

$f^{(2)}_{\mathrm{LAB}}$ の入力と出力は，それぞれ $\boldsymbol{h}_{\mathrm{LFE}}$ と $p(\hat{\boldsymbol{y}}_{\mathrm{LAB}})$ である．損失関数に $\boldsymbol{h}_{\mathrm{LFE}}$ と $p(\hat{\boldsymbol{y}}_{\mathrm{LAB}})$ を用いることで，LAB を分類に直接関連付けて学習させることができる．その結果，分類結果と強く関連するアテンションマップを生成できる．

次に，LPB は LFE と LAB の出力を元に分類を行う．LPB の入力と出力は，それぞれ $\boldsymbol{\alpha}' \odot \boldsymbol{h}_{\mathrm{LFE}}$ と $p(\hat{\boldsymbol{y}}_{\mathrm{LPB}})$ である．ここで，$\odot$ はアダマール積を表す．

Bottleneck 層の中で，[Bello 21] で提案された Lambda 層を用いた．Lambda Layer の入力を $\boldsymbol{h} \in \mathbb{R}^{c_3 \times w_3 \times h_3}$ とする．ここで，c_3, w_3, h_3 はそれぞれ Lambda 層の入力のチャンネル数，横幅，縦幅を表す．

まず，$\boldsymbol{h}$ からクエリ Q，キー K，バリュー V を生成する．

$$Q = \mathrm{Conv}(\boldsymbol{h}),\ K = \mathrm{Softmax}(\mathrm{Conv}(\boldsymbol{h})),\ V = \mathrm{Conv}(\boldsymbol{h}) \tag{3}$$

次に，K, V から content lambda $\boldsymbol{\lambda}_{\mathrm{c}}$ を，V から position lambdas $\boldsymbol{\lambda}_{\mathrm{p}}$ を生成する．

$$\boldsymbol{\lambda}_{\mathrm{c}} = K^{\top} V,\ \boldsymbol{\lambda}_{\mathrm{p}} = \mathrm{Conv}(V) \tag{4}$$

Lambda 層の出力 $\boldsymbol{h}_{\mathrm{L}} \in \mathbb{R}^{c_3 \times w_3 \times h_3}$ は，$\boldsymbol{\lambda}_{\mathrm{c}}$，$\boldsymbol{\lambda}_{\mathrm{p}}$ を用いて以下のように表される．

$$\boldsymbol{h}_{\mathrm{L}} = (\boldsymbol{\lambda}_{\mathrm{c}} + \boldsymbol{\lambda}_{\mathrm{p}})^{\top} Q \tag{5}$$

損失関数として，以下を使用する．

$$\mathcal{L} = \mathrm{CE}(\hat{\boldsymbol{y}}_{\mathrm{LPB}}, \boldsymbol{y}) + \lambda \mathrm{CE}(\hat{\boldsymbol{y}}_{\mathrm{LAB}}, \boldsymbol{y}) \tag{6}$$

ここで，$\boldsymbol{y}$，CE，λ はそれぞれ正解ラベル，交差エントロピー誤差関数，損失関数の重みを表す．

3.2 PID score

Insertion-Deletion score [Petsiuk 18] は，視覚的説明の標準的な評価指標である．Insertion-Deletion score は，説明生成手法で与えられた重要度に従って画素を挿入したときに，正解クラスの確率がどのように変化するかを測定する．しかし，Insertion-Deletion score は，スパースな重要領域を有する画像に対して，粗い説明と細かい説明を区別することができない．

そこで，スパースな重要領域を有する画像に対して有効な評価指標として，Insertion-Deletion score を拡張した PID score を提案する．PID score は，パッチ単位で挿入と削除を行うことで，このような画像を適切に評価することができる．PID score は以下のように定義される．

$$\mathrm{PID} = \mathrm{AUC}(\text{patch-insertion}) - \mathrm{AUC}(\text{patch-deletion}) \tag{7}$$

ここで，AUC は Area Under the Curve を表す．

Patch-insertion と patch-deletion 曲線は，以下の手順で得られる．まず，入力 $\boldsymbol{x}$ をパッチ（部分行列）$\boldsymbol{p}_{ij} \in \mathbb{R}^{c_1 \times m^2}$ に分割する．ここで，m はパッチの大きさ，i, j は縦横方向のインデックスである．$m = 1$ のとき，PID score は Insertion-Deletion score と一致する．

次に，attention map $\boldsymbol{\alpha}$ に max-pooling を適用して，各パッチにおける attention map $\boldsymbol{\alpha}_{\mathrm{p}} \in \mathbb{R}^{m^2}$ を作成する．$\boldsymbol{\alpha}_{\mathrm{p}}$ の要素を，値が大きい順に $\alpha_{i_1 j_1}$,$\alpha_{i_2 j_2}$,$\alpha_{i_3 j_3}$, $\cdots$,$\alpha_{i_m j_m}$ として，集合 A_n を次のように定義する．

$$A_n = \{(i_k, j_k) | k \leqq n\} \tag{8}$$

ここで，n は挿入・削除したパッチの数を表す．A_n を用いて，patch-insertion，patch-deletion の入力 $\boldsymbol{i}_n$, $\boldsymbol{d}_n$ はそれぞれ次のように表される．

$$(\boldsymbol{i}_n, \boldsymbol{d}_n) = \begin{cases} (\boldsymbol{p}_{ij}, \boldsymbol{0}) & (i, j) \in A_n \\ (\boldsymbol{0}, \boldsymbol{p}_{ij}) & (\text{otherwise}) \end{cases} \tag{9}$$

最後に，n と $\boldsymbol{y}_{\mathrm{c}}^{(\mathrm{ins},\ n)}$，$\boldsymbol{y}_{\mathrm{c}}^{(\mathrm{del},\ n)}$ をそれぞれプロットして，patch-insertion，patch-deletion がそれぞれ得られる．ここ

表 1: 提案手法のパラメータ設定

		IDRiD	DeFN magnetograms
Optimizer		AdamW	AdamW
Learning rate	LAB, Linear	1.0×10^{-3}	1.0×10^{-3}
	LFE, LPB	1.0×10^{-4}	1.0×10^{-4}
Weight decay		0.09	0.09
Batch size		32	32
Loss weights	Negative	2	1
	Positive	1	1
$\theta_{\boldsymbol{\alpha}}$		0.75	0.5
N_{B}		16	1
B		7	0

で，$\boldsymbol{y}^{(\mathrm{ins},\,n)}$，$\boldsymbol{y}^{(\mathrm{del},\,n)}$，$c$ はそれぞれ $\boldsymbol{i}_n$，$\boldsymbol{d}_n$ を入力したときの出力，$\boldsymbol{x}$ が属するクラスを示す．

4. 実験

4.1 実験設定

評価実験のデータセットには The Indian Diabetic Retinopathy Image Dataset (IDRiD) [Porwal 20] および DeFN magnetograms dataset [Nishizuka 18] を使用した．IDRiD は視覚的説明生成タスクにおける標準データセットであるため使用した．また，原理が未解明な太陽フレアに対して，理論への洞察を与える視覚的説明を生成することが重要になるため，DeFN magnetograms dataset を使用した．

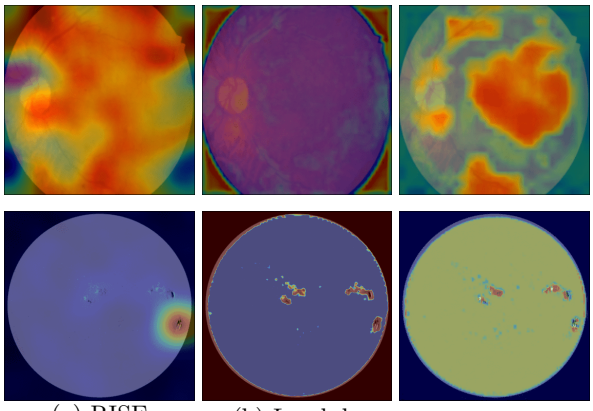
本実験では，IDRID の grade 0〜4 の 5 段階のアノテーションに基づき，grade 0 を陰性，grade1〜4 を陽性とする 2 段階に変換して $\boldsymbol{y}$ を作成した．$\boldsymbol{x}$ は 224×224 にリサイズした．

DeFN magnetograms dataset は，Solar Dynamic Observatory [Pesnell 12] のウェブアーカイブより収集した，Helioseismic and Magnetic Imager [Scherrer 12] で撮影された 1 時間間隔の太陽画像を含み，24 時間以内に発生する最大の太陽フレアクラスをラベルとして用いた．DeFN magnetograms dataset は，2010 年 6 月から 2017 年 12 月までの合計 61315 サンプルを含む．このとき，O，C，M，X の 4 段階の太陽フレアクラスから，O・C の 5237 サンプルを “< M”，M・X の 56078 サンプルを “≥ M” とする 2 段階に変換した．$\boldsymbol{x}$ は 512×512 にリサイズした．2010-2015 年の 45530 サンプルを訓練集合に，2016 年の 7795 サンプルを検証集合に，2017 年の 7990 サンプルをテスト集合に割り当てた．

訓練集合，検証集合，テスト集合はそれぞれパラメータの学習，ハイパーパラメータの検証，性能の評価に使用した．検証集合における損失関数の値が 6 回連続で改善しなかった場合に Early Stopping を行った．このとき，検証集合における損失関数の値が最も低いときのテスト集合における精度を最終的な精度とした．また，訓練時には画像の反転・回転・切り抜き・輝度変化・RandomErasing [Zhong 20] によるデータ拡張を行った．提案手法のパラメータ設定を表 1 に示す．ここで，loss weights は損失関数におけるクラスごとの重みである．学習にはメモリ 11GB 搭載の GeForce RTX 2080 および Intel Core i9-9900K を使用した．IDRiD，DeFN magnetograms dataset の訓練にはそれぞれ 40 分，1 日かかった．推論はどちらも 0.1 秒程度であった．

4.2 定量的結果

ベースライン手法として，RISE [Petsiuk 18] と Lambda attention を用いた．Lambda attention は，transformer の注意機構を説明として利用した [Vig 19] を基にして，Lambda 層における $\boldsymbol{\lambda}_c^{\top} Q$ をチャネル次元方向に平均して生成した．Lambda attention は transformer に対する説明生成の標準的な手法で，RISE は汎用的なモデルに適用できる標準的な手法のため使用した．IDRiD と DeFN magnetograms dataset には重要領域がスパースな画像が含まれているため，PID score を主要評価尺度として使用した．$m = 1$ における PID score は，視覚的説明の標準的な評価指標である Insertion-Deletion score と一致する．



(a) RISE [Petsiuk 18]　(b) Lambda attention　(c) Ours

図 3: 定性的結果．上段と下段はそれぞれ IDRiD と DeFN magnetograms dataset の結果を表す．

表 2 上表，下表にそれぞれ IDRiD，DeFN magnetograms dataset における定量的結果を示す．各手法につき 5 回の実験を行い，表にはその平均値及び標準偏差を示した．また，IDRiD，DeFN magnetograms dataset はそれぞれ陽性，“≥ M” のデータのみを PID の計算に使用した．これは，陰性や “< M” のデータは，説明として適切な病変や太陽フレアに影響する領域を含まないためである．

表 2 より，IDRiD における $m = 16$ の PID score は，RISE，Lambda attention，LABN でそれぞれ 0.182，-0.093，0.230 ポイントであった．したがって，LABN は RISE よりも PID score で 0.048 ポイント上回った．同様に，$m = 4$ と 8 のとき，PID score はそれぞれ 0.020 ポイントと 0.047 ポイント向上した．m が小さいときにベースライン手法の PID score が過大評価される原因を次節で分析する．また，DeFN magnetograms データセットでは，$m = 32, 64, 128$ における PID score が向上した．これらの結果から，LABN が重要領域がスパースな画像に対しても適切な説明を生成したことがわかる．

4.3 Ablation Studies

Ablation Studies として，4 つの条件を設定した．条件 (i)〜(iii) において，LAB の入力を抽出する層を変化させた場合に，性能にどの程度の差が生じるかを調査した．条件 (i)，(ii)，(iii) はそれぞれ $B = 3$，7，13 に対応する．条件 (iv) では，attention map のしきい値 $\theta_{\boldsymbol{\alpha}}$ を 0 に設定して画素のマスクを行わない場合について，性能への影響を調べた．

表 3 に Ablation Studies の定量的結果を示す．(i) と (iii) より，$B = 3$ と 13 の場合，$m = 16$ での PID スコアは (ii) と比較してそれぞれ 0.305 ポイントと 0.093 ポイント減少した．また，(iv) より，画素のマスクを行わない場合，(ii) と比較して $m = 16$ における PID スコアが 0.141 ポイント減少した．これらの結果から，中間層からの LAB 入力の取得と画素をマ

表 2: IDRiD(上表）および DeFN magnetograms dataset（下表）の定量的結果.

Method	PID				
	$m=1$	$m=2$	$m=4$	$m=8$	$m=16$
RISE [Petsiuk 18]	$\mathbf{0.319 \pm 0.015}$	$\mathbf{0.179 \pm 0.080}$	0.130 ± 0.045	0.136 ± 0.050	0.101 ± 0.033
Lambda attention	-0.101 ± 0.074	-0.105 ± 0.073	-0.116 ± 0.081	-0.123 ± 0.078	-0.093 ± 0.054
Ours (LABN)	0.111 ± 0.273	0.084 ± 0.111	$\mathbf{0.150 \pm 0.183}$	$\mathbf{0.183 \pm 0.253}$	$\mathbf{0.230 \pm 0.329}$

Method	PID				
	$m=1$	$m=16$	$m=32$	$m=64$	$m=128$
RISE [Petsiuk 18]	0.235 ± 0.145	0.261 ± 0.217	0.296 ± 0.199	0.379 ± 0.172	0.461 ± 0.164
Lambda attention	$\mathbf{0.374 \pm 0.080}$	$\mathbf{0.414 \pm 0.129}$	0.403 ± 0.138	0.378 ± 0.162	0.291 ± 0.216
Ours (LABN)	0.044 ± 0.055	0.311 ± 0.269	$\mathbf{0.489 \pm 0.207}$	$\mathbf{0.523 \pm 0.132}$	$\mathbf{0.556 \pm 0.135}$

表 3: Ablation Studies の定量的結果

Condition	PID				
	$m=1$	$m=2$	$m=4$	$m=8$	$m=16$
(i) $B=3$	-0.079	-0.079	-0.053	-0.067	-0.075
(ii) $B=7$	**0.111**	**0.084**	**0.150**	**0.183**	**0.230**
(iii) $B=13$	0.061	0.062	0.094	0.150	0.137
(iv) w/o pixel masking	0.045	0.020	0.023	0.091	0.089

スクすることが，モデルの性能向上に寄与していると考えられる.

4.4 定性的結果

図3 上段と下段に，それぞれ IDRiD と DeFN magnetograms dataset における定性的な結果を示す．上段左図から，RISE が広い範囲に注目し，重要領域にフォーカスしていないことがわかる．また，上段中央図は，Lambda attention が注目した領域のほとんどが背景であり，不適切であることを示している．一方，上段右図から LABN は精度に寄与する適切な領域に注目したことがわかる.

DeFN magnetograms dataset においては，下段左・中央図より，RISE と Lambda attention は円周や背景などの無関係な領域に注目していることがわかる．一方で，下段右図は LABN が太陽フレア予測に重要な黒点に適切に注目していることを示す.

最後に，Insertion-Deletion score（$m=1$ における PID score）が上段左図の不適切な視覚的説明を過大評価した理由を分析した．上段左図の画像の Insertion-Deletion score は 0.173 ポイントで，上段右図の Insertion-Deletion score の 0.034 ポイントより高い．Insertion-Deletion score の過大評価は，AUC(patch-deletion) が低いことに起因していると考えられる．式 (7) より，AUC(patch-deletion) が低いほうが Insertion-Deletion score は向上する．例えば，畳み込み層において，粗い説明では受容野がまとまって削除されやすいため AUC（patch-deletion）は低下する一方で，詳細な説明では入力が部分的に保持されるため，低下しにくい．$m=2$ においても，同様の問題がある．また，DeFN magnetograms dataset でも同様の問題があると考えられる．したがって，粗い説明と詳細な説明の両方を適切に評価できる，適切な m による PID score を用いるべきである.

5. おわりに

本論文では，画像分類モデルにおける判断根拠の視覚的説明を生成するタスクを扱った．提案手法による貢献は以下である.

- Lambda attention を用いた説明よりも明瞭な説明を獲得できる，ブランチ構造を導入した LABN を提案した.
- 重要領域がスパースな画像の説明に有効な評価指標として，PID score を提案した.
- パッチサイズが大きいとき，PID score で LABN はベースライン手法を上回った.

謝辞

本研究の一部は，JSPS 科研費 20H04269，JST ムーンショット，NEDO の助成を受けて実施されたものである.

参考文献

[Bello 21] Bello, I.: LambdaNetworks: Modeling Long-Range Interactions without Attention, in *ICLR* (2021)

[Fukui 19] Fukui, H., Hirakawa, T., et al.: Attention Branch Network: Learning of Attention Mechanism for Visual Explanation, in *CVPR*, pp. 10705–10714 (2019)

[Nishizuka 18] Nishizuka, N., Sugiura, K., et al.: Deep Flare Net (DeFN) Model for Solar Flare Prediction, *The Astrophysical Journal*, Vol. 858, No. 2, p. 113 (8pp) (2018)

[Pesnell 12] Pesnell, W., Thompson, B., and Chamberlin, P.: The Solar Dynamics Observatory (SDO), *Solar Physics*, Vol. 275, No. 1–2, pp. 3–15 (2012)

[Petsiuk 18] Petsiuk, V., Das, A., and Saenko, K.: RISE: Randomized Input Sampling for Explanation of Black-box Models, in *BMVC*, p. 151(13pp) (2018)

[Porwal 20] Porwal, P., et al.: IDRiD: Diabetic Retinopathy – Segmentation and Grading Challenge, *Medical Image Analysis*, Vol. 59, No. 101561 (2020)

[Scherrer 12] Scherrer, P., Schou, J., Bush, R., et al.: The Helioseismic and Magnetic Imager (HMI) Investigation for the Solar Dynamics Observatory (SDO), *Solar Physics*, Vol. 275, pp. 207–227 (2012)

[Selvaraju 17] Selvaraju, R., et al.: Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization, in *ICCV*, pp. 618–626 (2017)

[Vig 19] Vig, J.: A Multiscale Visualization of Attention in the Transformer Model, in *ACL*, pp. 37–42 (2019)

[Wang 20] Wang, H., Wang, Z., Du, M., et al.: Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks, in *CVPR*, pp. 24–25 (2020)

[Zhong 20] Zhong, Z., Zheng, L., Kang, G., Li, S., and Yang, Y.: Random Erasing Data Augmentation, in *AAAI*, Vol. 34, pp. 13001–13008 (2020)