

Focal Tversky Loss による護岸のひび割れセグメンテーションの精度向上

Improving the accuracy of crack segmentation of revetment using Focal Tversky Loss

都築 幸乃*¹ 吉田 龍人*¹ 大久保 順一*¹ 藤井 純一郎*¹ 山下 隆義*²
Yukino Tsuzuki Ryuto Yoshida Junichi Okubo Junichiro Fujii Takayoshi Yamashita

*¹ 八千代エンジニアリング株式会社 Yachiyo Engineering Co.,Ltd.
*² 中部大学 Chubu University.

A technique for detecting cracks in river concrete revetments using image segmentation technology is being studied. However, the area of the crack to be classified is very small compared to the background. Models trained on such imbalanced data are known to be unstable in prediction accuracy due to optimization that is more influenced by classes with more data than classes with fewer data. Therefore, in order to improve the crack detection accuracy, this paper tries to adopt Focal Tversky Loss, which is a loss function robust to imbalanced data. The model using Focal Tversky Loss showed higher crack detection accuracy than the commonly used Binary Cross Entropy and Dice Loss. In addition, by introducing an attention mechanism into the segmentation model, a visualized image representing a judgment reason as to which part of the image was focused for segmentation was generated.

1. はじめに

日本には数多くの河川があり、その附帯施設としてコンクリート護岸が設置されている。これら河川管理施設は老朽化が進行し、維持管理が大きな課題となっている。それに対して現在の河川護岸の点検は、堤防等河川管理施設及び河道の点検・評価要領[国土交通省 2019]に従い、技術者による目視点検が主体とされている。しかし、長距離を徒歩で移動し、変状を発見する度に計測や撮影を行う現在の点検手法は非常に非効率的である。この従来手法を機械的な手法に置き換えることができれば、技術者の作業負担の低減や生産性の向上が期待できる。すでに道路橋やトンネルでは、2019年に改定された点検要領により、近接目視と同等の情報が得られると判断された場合には代替手法の採用が可能とされている。このことから、河川護岸においても代替手法が法令として容認されることが予測される。

このような背景から近年、AIを活用した維持管理が検討されている。筆者らはこれまで画像セグメンテーション技術を用いて河川コンクリート護岸のひび割れを自動的に検出するツール開発を行ってきた。これは、画素ごとにクラス分類を行うセマンティックセグメンテーションを採用している。先行研究[齋藤 2019]では、医療画像分析用に開発された FusionNet をベースに、護岸のひび割れ検出に適したセグメンテーションモデルを構築した。

あるクラスの領域が別のクラスの領域に比べて非常に小さいデータを不均衡データ(imbalanced data)と呼ぶ。この不均衡データでモデルを訓練した場合、データ数の少ないクラスよりもデータ数の多いクラスの影響を強く受けた最適化が行われるため、予測精度が低下することが知られている。セグメント化の対象である護岸のひび割れも健全な背景領域に比べて画像全体に占める割合が非常に小さいため、不均衡データであると考えられる。しかし、従来のひび割れ検出モデルでは不均衡データへの対応が十分でないため、性能向上を図る必要がある。

従来研究では、不均衡データに対処するために、各クラスにおける誤差が損失関数に与える影響を調整した様々な損失関数が提案されている[Jadon 2020]。これにより、背景領域よりも対

象領域に注視してモデルを学習させることが可能となった。そこで本研究では、不均衡データにロバストな損失関数とされる Focal Tversky Loss[Abraham 2019]を用いることで、モデルの性能向上を試みる。

また、不均衡データに対する別のアプローチとして、Attention U-Net[Oktay 2018]が考案されている。これは、U-Netのスキップ接続上に Attention 機構[Vaswani 2017]を組みこむことで、対象領域の特徴量を強調し、背景領域の特徴量の活性化を抑制している。本研究では Attention 機構を用いて、従来のセグメンテーションモデルをベースモデルとした Attention Gate モデルを新たに構築する。また、Attention 機構によって着目した領域を可視化することで、画像のどの部分に着目してセグメンテーションが行われたのかという判断根拠を表す可視化画像(Attention マスク)の生成を行う。

ここでは、Focal Tversky Loss と従来の学習で使用していた Binary Cross Entropy をそれぞれ用いたベースモデルと Attention Gate モデルの合計 4 つのモデルを作成し、ひび割れ検出精度について検討する。

2. 損失関数

損失関数とは、予測値が真値に対してどれだけ近いかを表す関数であり、モデルの学習は損失関数を最小化することを目的としている。セグメンテーションにおける損失関数は、一般的に Binary Cross Entropy や Dice Loss が用いられる。しかし、不均衡データを学習する場合は各クラスの損失関数への寄与率を調整した損失関数を選択することが有効であるとされている。

本章では、Binary Cross Entropy と Dice Loss, Focal Tversky Loss の 3 種類の損失関数について述べる。

2.1 Binary Cross Entropy

Binary Cross Entropy は、与えられた確率変数または事象集合に対する 2 つの確率分布間の差の尺度として定義され、次式で表される。

$$BCE(p, \hat{p}) = -(p \log(\hat{p}) + (1 - p) \log(1 - \hat{p})) \quad (1)$$

ここで、 $p \in \{0,1\}$ は真値の確率変数、 $\hat{p} \in [0,1]$ は予測値の確率変数である。

連絡先: 都築幸乃, 八千代エンジニアリング株式会社, 東京都台東区浅草橋 5-20-8 CS タワー3F 技術開発研究所, TEL: 03-5822-7032, Mail: yk-tsuzuki@yachiyo-eng.co.jp

不均衡データを学習する場合、Binary Cross Entropy では、すべてのデータをデータ数の多いクラスに予測した場合にも損失が下がってしまうおそれがある。

2.2 Dice Loss

Dice Loss は Dice 係数を利用した損失関数である。Dice 係数とは、2 つの集合の平均要素数と共通要素数の割合を示した指標であり、以下の式で表される。

$$DSC(p, \hat{p}) = \frac{2 \sum p\hat{p} + 1}{\sum p + \sum \hat{p} + 1} \quad (2)$$

ここで、分子と分母に加えた 1 はゼロ除算対策である。この Dice 係数は、値が大きいほど 2 つの集合の類似度が高いことを表す。Dice Loss は式(2)を用いて以下の式で表される。

$$DL(p, \hat{p}) = \sum 1 - DSC(p, \hat{p}) \quad (3)$$

Dice Loss は precision(適合率)と recall(再現率)の調和平均で定義されているため、不均衡データで起こり得る一部のクラスに過剰適合しやすいといった問題などが回避できる。しかし、FP(偽陽性)と FN(偽陰性)における重みが等しいため、非常に小さな領域をセグメント化することは困難な場合がある。

2.3 Focal Tversky loss

前節で述べた問題を解決するために、Dice 係数における FP と FN に重み付けをして一般化した指標を Tversky Index と呼ぶ。この Tversky Index を用いた損失関数が Tversky Loss であり、それぞれ以下の式で表される。

$$TI(p, \hat{p}) = \frac{\sum p\hat{p} + 1}{\sum p\hat{p} + \alpha \sum p(1 - \hat{p}) + \beta \sum (1 - p)\hat{p} + 1} \quad (4)$$

$$TL(p, \hat{p}) = \sum 1 - TI(p, \hat{p}) \quad (5)$$

ここで、 α は FN、 $\beta (= 1 - \alpha)$ は FP の重みを表し、 $\alpha = 1/2$ のとき Tversky Index は Dice 係数と等しい値になる。このパラメータを調整することで、Tversky Loss は Dice Loss に比べて precision と recall のバランスを柔軟にとることが可能となった。

さらに、Tversky Loss を Focal Loss に似た形に派生させたものを Focal Tversky Loss と呼び、以下の式で表される。

$$FTL(p, \hat{p}) = \sum (1 - TI(p, \hat{p}))^\gamma \quad (6)$$

ここで、 $\gamma \in [1, 3]$ であり、 $\gamma = 1$ のとき Focal Tversky Loss は Tversky Loss と等しい値になる。式(6)のように指数関数として定義することで、分類が容易な背景領域が損失関数に与える影響を抑制し、分類が困難な領域に比重が置かれるようになった。図-1 に示すように、Tversky Index が 0 に近い場合、Focal Tversky Loss の値は Tversky Loss の値とほぼ変わらないが、Tversky Index が 1 に近い場合は Tversky Loss に比べて値が大きく減少する。また、 γ の値が大きいほどその減衰が大きくなる。

3. Attention 機構

Attention 機構とは、文章や画像の特定の部分に注意を向けるようにモデルを学習させる仕組みである。これは、入力値と予測値の関連度が高い特徴量に対して重み付けを行い、その特徴量を重視した予測モデルを構築することができる。Attention 機構は元々自然言語処理の分野で考案された技術であるが、現在は画像処理の分野でも応用されている。画像認識における Attention 機構には空間方向とチャンネル方向の Attention があり、本研究では空間方向の Attention を採用する。これは、畳

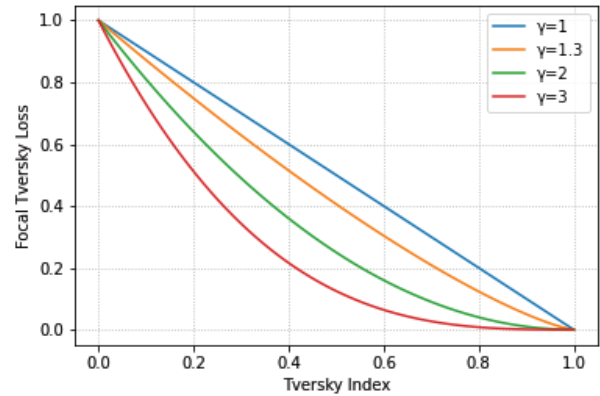


図-1 Focal Tversky Loss

み込み処理によって抽出した特徴量マップのどの領域に注意を向けるかを重み付けする手法である。

本章ではまずベースモデルとする従来モデルについて述べた後、本研究で構築した Attention Gate(AG)モデルについて述べる。

3.1 ベースモデル

本研究では、[齋藤 2019]によって提案されたモデルをベースモデルとする。これは、Encoder-Decoder 型のセマンティックセグメンテーションモデルである FusionNet に対していくつかの変更を行ったネットワークである。FusionNet における Encoder と Decoder は、畳み込み層と残差ブロック、プーリング層またはアップサンプリング層で構成されており、ここではこのまわりを 1 つの層として数える。FusionNet からベースモデルへの主な変更点は、ネットワークを 4 層から 5 層に増やし、Encoder から Decoder へのスキップ接続を残差結合から U-Net と同様の連結結合にした点である。これらの変更は、対象とするひび割れの特徴量の取得を促進させるためのものであり、細かいひび割れの形状を捉えることを可能にした。このことは、[天方 2021]の t-SNE を用いた分析により立証されている。

3.2 Attention Gate (AG) モデル

図-2 に AG モデルの概要を示す。ベースモデルとの変更点は、Encoder の特徴量を Decoder に連結する前に AG へ通すことである。AG への入力、Encoder の特徴量 x^l と Decoder の特徴量 g である。まず、それぞれに対して畳み込み処理を行い、特徴量マップを計算する。次に、これらの特徴量マップの和を ReLU 関数 σ_1 にかけた後、再度畳み込み処理を行い、シグモイド関数 σ_2 にかける。こうして各画素 i における Attention 係数 $\alpha_i^l \in [0, 1]$ を取得し、重み付けされた特徴量マップ \hat{x}_i^l を出力する。それぞれの計算を次式に表す。

$$q_{att}^l = \psi^T \left(\sigma_1 (W_x^T x_i^l + W_g^T g_i + b_g) \right) + b_\psi \quad (7)$$

$$\alpha_i^l = \sigma_2(q_{att}^l) \quad (8)$$

$$\hat{x}_i^l = x_i^l \cdot \alpha_i^l \quad (9)$$

ここで、 b_g と b_ψ はバイアス項である。

Attention U-Net では AG へ入力する g は 1 つ前の層の出力であるが、本研究の AG モデルでは、すべての AG に 5 層目の出力 g を入力する。

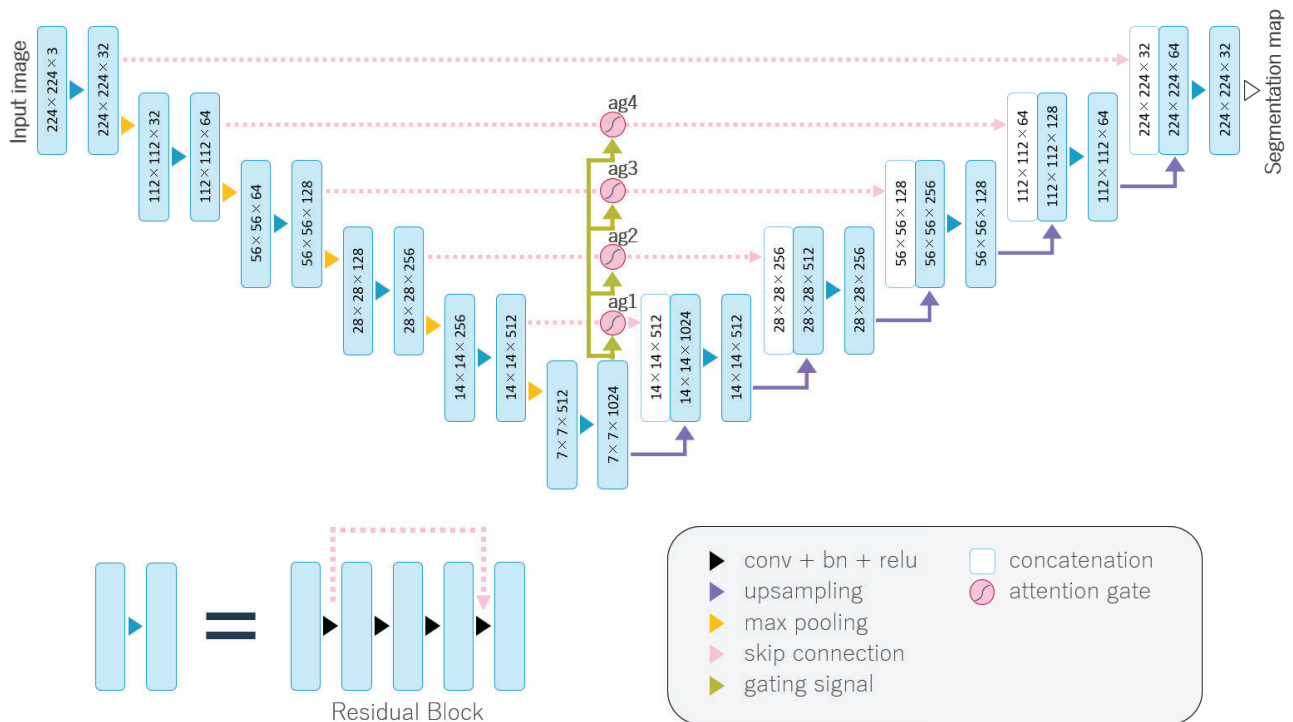


図-2 Attention Gate (AG) モデル

4. 実験

本章では、Binary Cross Entropy(BCE)と Focal Tversky Loss(FTL)をそれぞれ用いたベースモデルと AG モデルの合計 4 パターンのモデルを作成し、ひび割れ検出精度を比較する。また、本実験の前に FTL のパラメータを決定するための小規模な実験を行った。

4.1 学習データ

護岸画像の例を図-3 に示す。画像サイズは 6000×4000 ピクセルで、256 階調の RGB 画像である。前処理として、この元画像をモデルに入力するサイズに分割しておく。ここでは、Faster R-CNN で護岸領域を取得した後、 224×224 ピクセルの小領域画像に分割した。この分割した画像を crop 画像と呼ぶ。

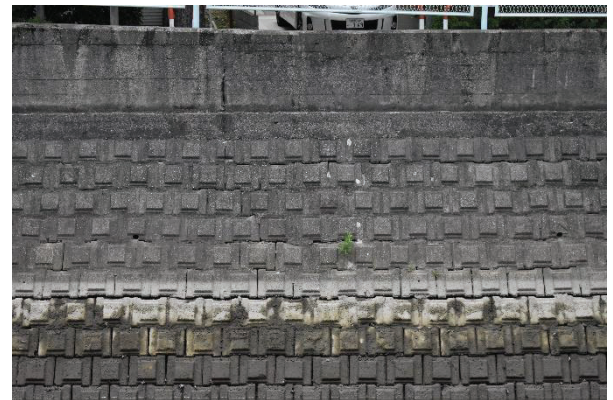


図-3 護岸画像例

4.2 FTL パラメータの決定

FTL のパラメータ(α , β , γ)を決定するために、小規模のデータセット(ひび割れを含む crop 画像 200 枚、ひび割れを含まない crop 画像 100 枚)を用意した。

まず BCE, Dice Loss(DL), FTL($\alpha = 0.1 \sim 0.9$)の合計 11 パターンのモデルを作成した。このとき γ は $4/3$ で統一している。各モデルで推論を行い、precisionとrecall, F 値を算出した結果を表-1 に示す。FTL モデルでは α が大きくなるほど precision が減少し、recall が増加する傾向が見られた。また、 $\alpha = 0.1$ と $\alpha = 0.9$ の場合を除き、BCE モデルや DL モデルよりも FTL モデルの F 値が高い結果となった。本研究では検出精度の向上を目的としているが、従来モデルで検出できていたひび割れを見落とすことは望ましくない。そのため、recall が BCE モデルと同程度であり、F 値が最も高い FTL モデルのパラメータ $\alpha=0.7$ ($\beta = 0.3$)を採用する。

次に $\alpha=0.7$ に固定し、 $\gamma = 4/3 \sim 3$ の合計 6 パターンのモデルを作成した。各モデルで推論を行い、precisionとrecall, F 値を算出した結果を表-2 に示す。 γ と性能評価結果との明確な相関

関係は見られなかったが、 $\gamma = 4/3$ のとき F 値が最も高い結果となった。これより、 γ の値は $4/3$ に決定する。

4.3 学習

護岸画像 17 枚分の crop 画像のうち、ひび割れを含む crop 画像 2274 枚、ひび割れを含まない crop 画像 1000 枚で学習データを構築した。また、その 80% を training data, 20% を validation data に用いた。ここでは、すべてのモデルに対して 5 分割交差検証を実施する。

4.4 推論

学習データとは別の護岸画像 3 枚を用いて推論を行った。precisionとrecall, F 値の平均値を算出した結果を表-3 に示す。まず損失関数の影響を比較すると、FTL を用いたベースモデルが BCE を用いたベースモデルよりも F 値が 6.22 ポイント高い結果となった。AG モデルでも同様に、FTL を用いた場合の F 値が高い。このことから、FTL を用いた学習が護岸のひび割れセグメンテーションに有効であることが分かった。さらに Attention 機

構の影響を比較すると、FTL を用いたベースモデルと AG モデルでは、AG モデルの F 値が 0.22 ポイント高く、最も精度が高い結果となった。一方で、BCE を用いた AG モデルは、精度が最も低い結果となった。このことから、本研究における実験では Attention 機構は精度向上にほとんど寄与せず、また、損失関数との組み合わせによっては Attention 機構がうまく機能する場合としない場合があることが分かった。

4.5 Attention マスクの可視化

図-2 に示したネットワークのうち、出力層に最も近い AG(ag4) における Attention マスクをヒートマップとして可視化した(図-4)。ここでは、FTL を用いた AG モデルの結果を示している。Attention マスクで強く反応を示している領域では、推論結果でも正しくセグメント化されていることが確認できる。一方、正しくセグメント化が行われていない領域は、Attention の値も小さい。このことから、本研究では Attention 機構は精度の向上にはほとんど寄与しなかったが、モデルの判断根拠を可視化することができ、ひび割れ領域に着目していることが確認できた。

5. おわりに

本研究では、河川コンクリート護岸のひび割れセグメンテーションにおいて、セグメント化の対象領域が非常に小さい不均衡データに対処するために、損失関数の変更と Attention 機構の導入を試みた。ここでは、損失関数に Focal Tversky Loss を採用し、従来の学習で使用していた Binary Cross Entropy に比べて検出精度が向上することを示した。一方、Attention 機構を導入したことによる検出精度への影響はほとんど見られなかったが、モデルの判断根拠を可視化することができ、ひび割れ領域に着目したセグメンテーションが行われていることが確認できた。

参考文献

- [国土交通省 2019] 国土交通省 水管理・国土保全局河川環境課: 堤防等河川管理施設及び河道の点検・評価要領, p.2, 2019.
- [齋藤 2019] 齋藤彰儀, 上総虎智, 平木悠太, 天方匡純, 吉田武司: 深層学習によるコンクリート護岸劣化領域検出システムの開発, デジタルプラクティス Vol.10 No.2, 情報処理学会, 2019.
- [Jadon 2020] Jadon, S.: A survey of loss functions for semantic segmentation, arXiv:2006.14822v4, 2020.
- [Abraham 2019] Abraham, N. and Khan, N. M.: A NOVEL FOCAL TVERSKY LOSS FUNCTION WITH IMPROVED ATTENTION U-NET FOR LESION SEGMENTATION, 2019 IEEE International Symposium on Biomedical Imaging (ISBI), 2019.
- [Oktay 2018] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B. and Rueckert, D.: Attention U-Net: Learning Where to Look for the Pancreas, arXiv: 1804.03999, 2018.
- [Vaswani 2017] Vaswani, A., Shazeer, N., Parmar, Niki., Uszkoreit, J., Jones, L., Gomez, A. N., Ksidrt, L. and Polodukhin, I.: Attention Is All You Need, arXiv: 1706.03762, 2017.
- [天方 2021] 天方匡純, 吉田龍人, 藤井純一郎, 高橋悠太, 大久保順一: 深層学習ネットワークによるひび割れ検出精度の違い, 人工知能学会全国大会論文集, 2021.

表-1 α および β を変更したモデルの性能評価結果

Loss	α	β	γ	precision	recall	F値
BCE	-	-	-	0.4548	0.8340	0.5886
DL	-	-	-	0.5460	0.7222	0.6218
FTL	0.1	0.9	4/3	0.8151	0.2790	0.4157
	0.2	0.8		0.7213	0.6087	0.6602
	0.3	0.7		0.6885	0.5966	0.6393
	0.4	0.6		0.7165	0.6026	0.6546
	0.5	0.5		0.6401	0.7242	0.6795
	0.6	0.4		0.5908	0.7674	0.6677
	0.7	0.3		0.5618	0.8242	0.6682
	0.8	0.2		0.5261	0.8187	0.6405
	0.9	0.1		0.4055	0.8867	0.5565

表-2 γ を変更したモデルの性能評価結果

Loss	α	β	γ	precision	recall	F値
BCE	-	-	-	0.4548	0.8340	0.5886
DL	-	-	-	0.5460	0.7222	0.6218
FTL	0.7	0.3	4/3	0.5618	0.8242	0.6682
			5/3	0.4988	0.8645	0.6326
			2	0.5410	0.8224	0.6527
			7/3	0.5081	0.8580	0.6382
			8/3	0.5636	0.7557	0.6457
			3	0.5451	0.8009	0.6487

表-3 モデルの性能評価結果

モデル	Loss	precision	recall	F値
base	BCE	0.5059	0.8343	0.6293
AG	BCE	0.5083	0.7916	0.6190
base	FTL ($\alpha 0.7, \beta 0.3, \gamma 4/3$)	0.5717	0.8756	0.6915
AG	FTL ($\alpha 0.7, \beta 0.3, \gamma 4/3$)	0.5735	0.8794	0.6937

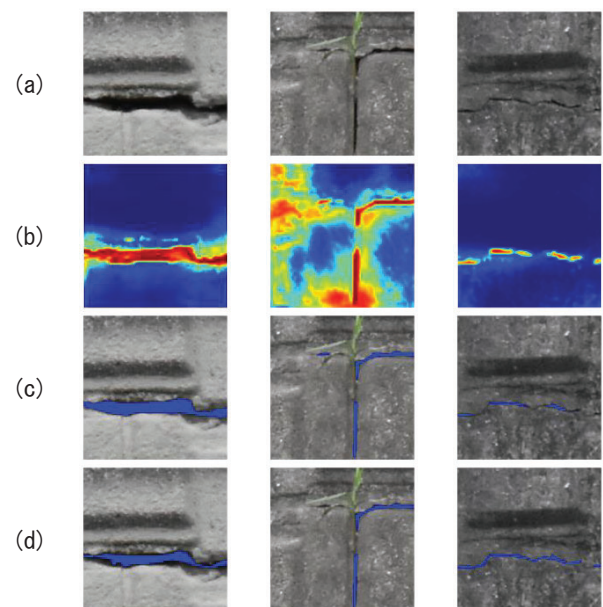


図-4 (a)入力画像, (b)Attention マスク, (c)推論結果, (d)正解画像