

# 画像特徴量に基づくデータクレンジングが ひび割れセグメンテーションの精度に与える影響の検討

Verification of learning accuracy in crack segmentation using data cleansing based on image features

吉田 龍人\*<sup>1</sup> 藤井 純一郎\*<sup>1</sup> 大久保 順一\*<sup>1</sup> 天方 匡純\*<sup>1</sup>  
Ryuto Yoshida Junichiro Fujii Junichi Okubo Masazumi Amakata

\*<sup>1</sup> 八千代エンジニアリング株式会社  
Yachiyo Engineering Co.,Ltd.

In deep learning, data cleansing is effective in improving the accuracy of the model. On the other hand, the number of data is also an important factor for proper training. Therefore, when performing data cleansing, it is necessary to apply an effective method. Based on this problem, this study verified the effect of data cleansing on the crack segmentation for revetment. In the verification, various datasets was created based on the features of training images. And training results was compared for each dataset.

## 1. はじめに

### 1.1 背景

令和2年度の国土交通白書[国土交通省 2020]では、既存のインフラの老朽化に加えて、老朽化したインフラ点検に携わる技術者の減少といった課題が示されており、インフラの維持管理の効率化が建設業界における目下の課題となっている。

現在の河川の護岸の点検は、点検・評価要領[国土交通省 2019]に従って、技術者の目視を基本としていることから、何百、何千 km といった単位で河川沿いを歩いて点検する必要がある。中でも護岸の点検項目の一つであるひび割れは、同要領により幅や長さを評価指標とすることが定められているが、個別のひび割れをクラックゲージ等によって計測するのは労力が大きく、ひび割れ評価の自動化は急務である。

### 1.2 護岸ひび割れ検出に関する既往研究

前節の課題を受けて、筆者らは河川護岸のひび割れセグメンテーション手法と、それを活用したひび割れ評価手法をこれまでに研究、開発した。

[齋藤 2019]では、FusionNet をベースとしたモデルを、224×224の小領域に分割した護岸の画像(以下、クロップ画像と称する)で学習させることで、高い精度でひび割れが検出可能なことを示した。さらに[吉田 2020]は、護岸画像 50 枚を教師画像とした学習を行うことで、同じ意匠の護岸における、未学習画像のひび割れが精度よく検出可能となることを確認し、検出結果からひび割れ幅や延長を計測する手法を示した。

### 1.3 課題

前節で示した先行研究では、デジタル画像からひび割れを検出し、評価する手法が検討されたが、デジタル画像による点検を行う際は、背景や護岸を覆う植物などといった解析対象外の要素(以下、ノイズと称する)も同時に写り込んでしまうため、ノイズの除去を行うための前処理が必須となる。これに対し、[齋藤 2019]では Faster-RCNN で作成した護岸検知器によって護

岸領域を検出し、検出範囲内をクロップ画像の分割対象とすることで、背景のたまかな削除を行った。

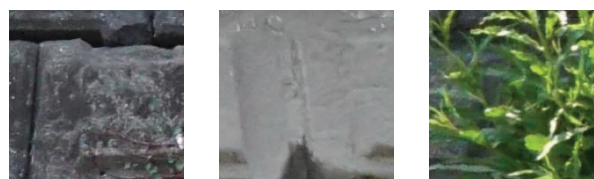
護岸検知によって背景を取り除いた画像の中には、図-1のように、ひび割れのある護岸を写したポジティブ画像、ひび割れない護岸を写したネガティブ画像だけでなく、護岸検出では除ききれないノイズ画像が多数存在するといった問題がある。したがってより高精度なモデルを作成するためには、ノイズ画像が学習に与える影響を明らかにし、場合によっては学習データセットから画像を除去する必要がある。

### 1.4 データクレンジングの既往研究と本研究の取り組み

ノイズ画像の取り扱いに関して、深層学習の教師データと精度に関する既往研究はいくつか存在する。[Jeatrakul 2010]では、ニューラルネットワークを用いた画像分類の精度が、クレンジングによって向上することが示されている。一方で、[Sun 2017]では教師データの量によって学習精度が向上することが示されている。

ノイズ画像をクレンジングすることで、教師データのばらつきが減り、モデルの精度が向上する可能性もあるが、安易なクレンジングはデータ数の減少を招くため、かえって精度低下を誘発する恐れがあることが推察される。

上述の課題を受けて、本研究では、[吉田 2020]で使用された河川の護岸のひび割れセグメンテーションの画像データセットを題材に、データクレンジングが学習時に与える影響を検討する。クレンジングは、いかなるデータに対しても簡易的に適用できるよう、画像の基礎的特徴量に閾値を設けて、半自動的に処理を行う。あらゆる特徴量によってクレンジングしたデータセットを複数作成し、各データセットでのモデルの学習結果を評価する。



(a) ポジティブ (b) ネガティブ (c) ノイズ

図-1 クロップ画像例

連絡先: 吉田龍人, 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 技術創発研究所,  
TEL: 03-5822-6843, Mail: ry-yoshida@yachiyo-eng.co.jp

## 2. 特徴量に基づく教師画像のクレンジング

### 2.1 対象画像の概要

本研究の検出対象は図-2のような護岸画像のひび割れである。検出対象となる画像は、ドローン自動航行によって護岸に正対するように撮影しており、護岸の空間解像度は 2mm/px を満足している。

教師画像は同様の距離・画角から撮影した 6000px×4000px の 50 枚の画像を用いており、対象護岸の意匠はすべて同じである。これら 50 枚の画像はすべて同日に撮影されたものだが、数 km に及ぶ河川の左右岸を連続的に撮影した結果からひび割れの多く写った画像 50 枚を抽出しているため、日照条件の影響で画像ごとに明るさの違いが見られる。



図-2 解析対象の護岸画像例

### 2.2 画像特徴量の整理

データセットの作成にあたって、護岸検知領域内のクローズアップ画像ごとに画素の平均値と標準偏差を算出した。図-3 に ch ごとの画素値のヒストグラム、図-4 に標準偏差のヒストグラムを示す。

平均画素値のヒストグラムより、RGB それぞれの ch の頻度である R, G, B と 3ch 平均の頻度である RGB の頻度が同程度であり、コンクリートの色である灰色が多く占める対象画像の性質が示されている。次に標準偏差のヒストグラムを見ると、護岸ブロックのクローズアップ画像は基本的に画像全体が一般的な色合いをしているため、比較的低い値に集中しており、ある ch のヒストグラムが他のチャンネルを卓越する例もほとんど見られないことが明らかとなった。

以上より、ヒストグラムの頻度が低いものや、RGB のうち一つの ch の値が突出しているデータは、ノイズ画像である可能性が高く、データセット作成時に取り除くべき画像であると判断した。

### 2.3 データセットの作成と学習

特徴量の整理結果を受けて、様々な条件によってクローズアップ画像を抽出し、8 対 2 の割合で train と validation (以下、val と称する) に分けたデータセットを作成した。表-1 はデータセットの作成条件とその時のデータ数をまとめたものである。ここで ch 突出度とは RGB の各 ch の平均画素値に対して、3ch の平均画素値で割った値であり、RGB すべての ch において 0.9 から 1.1 の範囲に収まるものだけを取り出している。さらに、同程度の教師枚数での学習状況を明らかにするため、train と val を 7 対 3, 6 対 4 に分けたデータセットも作成した。

表-1 のデータセットによって学習を行い、モデルの性能を比較する。学習にあたって、モデルは[齋藤 2019]で検討されたモデルを使用し、ハイパーパラメータも同論文に準拠した。学習の終了条件はいずれも 50 エポックとし、val\_loss が最も低くなったものをベストモデルとして、次章での推論時に採用した。

## 3. 実験結果

### 3.1 学習曲線の比較

表-1 のデータセットで学習を行った際の学習曲線を図-5 に示す。データセット内で loss, val\_loss ともに最も低くなったのは、標準偏差 10-20 のデータセットで、その次はクレンジングを行っていない train 80%データセットであった。その他のフィルターを加えたデータセットの結果は train 80%の結果より低い位置を推移していた。

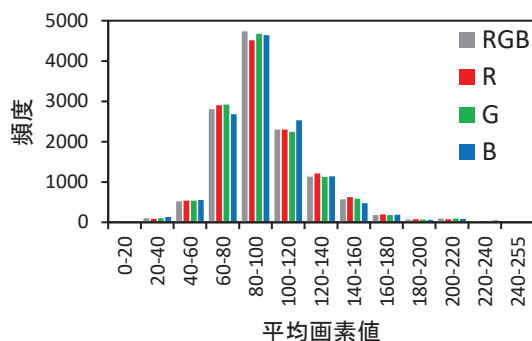


図-3 平均画素値のヒストグラム

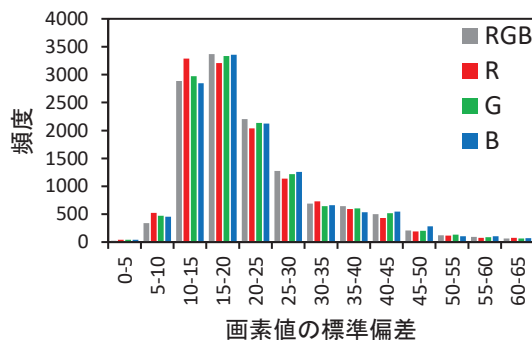


図-4 画素値の標準偏差のヒストグラム

表-1 データセット概要

クレンジング概要	総枚数	train	val
クレンジングなし train 60%	12558	7534	5024
クレンジングなし train 70%	12558	8790	3768
クレンジングなし train 80%	12558	10046	2512
標準偏差 10-20	6252	5001	1251
標準偏差 10-40	11066	8848	2218
平均画素値 65-140	10627	8501	2126
ch 突出度 0.9-1.1	11532	9225	2307

標準偏差 10-20 という強いクレンジングによって作成したデータセットは、入力画像のバリエーションが抑制され、モデルが学習しやすいデータとなったのだと推察した。一方で、弱いクレンジングを行ったその他の3ケースの場合、ノイズ画像が除去されることから、データ全体のポジティブデータの比率が増え、accuracy と loss が train 80%モデルの結果より悪化したのだと推察した。

図-5 では、train の比率を 70%以下にした2ケースにおいて、accuracy が 0.990 程度、loss が 0.3 以上と他ケースに比べて遥かに悪い結果となったため、グラフから取り除いている。train に含まれる画像をランダムに分けてしまうと、学習の安定性が損なわれることが示唆された。

### 3.2 推論結果の評価

作成したモデルにより、表-1 の train 60%, train70%を除くデータセットに対して推論を行い、precision, recall, mIoU (mean Intersection over Union) を算出した。その結果が図-6 である。

precision と recall を比較すると、recall の方が全体的に高く、ひび割れが過剰に検出される傾向にあることが分かった。個々のモデル間で性能を比べると、precision や mIoU に比べて、recall の差は小さかった。mIoU が最も高くなったのは train 80% のモデルで、他のデータセットのモデルと比べても、明らかに高い値となった。クレンジングを行ったデータセットによって作成したモデルは、各自のクレンジングデータへの推論に特化することが期待されたが、それぞれの性能指標の算出結果に大きな変化はなかった。

### 3.3 中間層の出力値の比較

推論をかけた際の中間層を出力し、ポジティブ、ネガティブ、ノイズ画像が持つ特徴量の違いを比較した。比較にあたって、ポジティブ画像 300 枚、ネガティブ画像 300 枚、ノイズ画像 300 枚を抽出した。この時ポジティブ画像は、学習に使用したデータセットからアノテーションのピクセル数別に 60 枚ずつを取り出し、ネガティブとノイズ画像は無作為に取り出した。比較時に使用したモデルは mIoU が最も高い値を示した train 80% のモデルで、最も抽象化した特徴量で比較すべく、最初のアップサンプリング層の手前の  $7 \times 7 \times 1024$  の出力値を抽出した。さらに、その出力値の違いを視覚的に表現するため、出力値を PCA (Principal Component Analysis) および t-SNE (t-Distributed Stochastic Neighbor Embedding) によって 2 次元に削減した。

図-7 は中間層の出力を 2 次元に削減した結果を示している。グレーのプロットはノイズ画像、最も濃い青のプロットはネガティブ画像、それ以外のプロットはポジティブ画像の出力で、色別にアノテーションしたピクセルの量を示している。図より、ポジティブ、ネガティブ、ノイズのプロットの分布の違いが明確であり、モデルのエンコーダーがそれぞれの画像が持つ特徴の違いを判別できていると判断した。また、ひび割れのアノテーション量と分布に相関性はなく、ひび割れの位置や形状と言った別の要素が、プロット位置に影響している可能性が高いと推察した。さらにネガティブ画像のプロットは比較的まとまりを持って分布しているのに対して、ポジティブ画像とノイズ画像のプロットにはばらつきがあることが分かった。

以上のことから、train 80% のモデルでは、ネガティブ画像を入力すると、エンコーダーはほとんど発火せず、ポジティブやノイズ画像を入力すると、発火して何らかの特徴量を出力していると判断した。

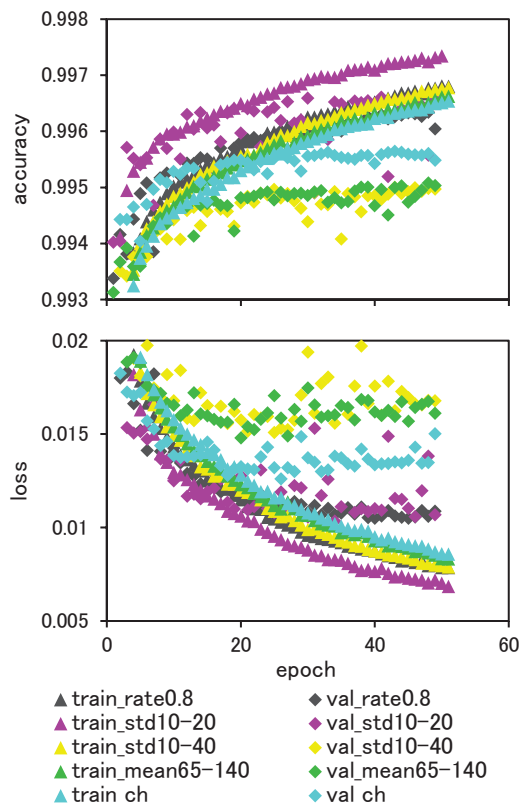


図-5 学習曲線

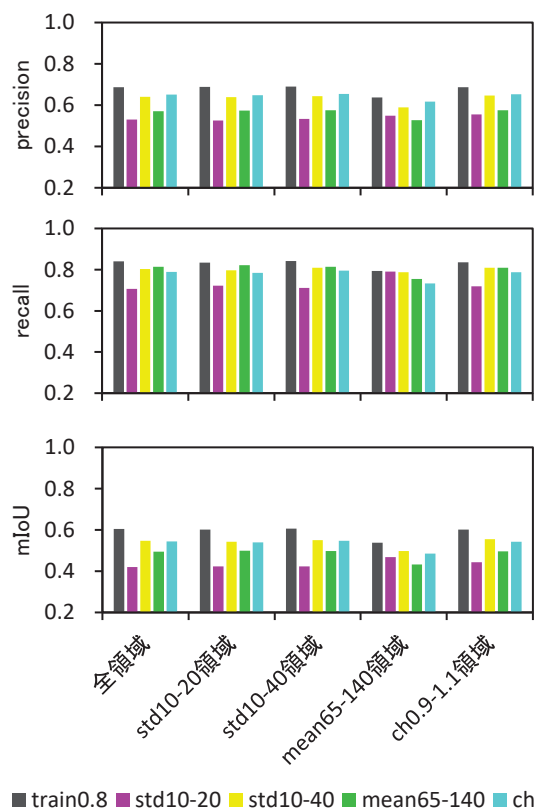


図-6 モデルごとの推論結果評価



### 3.4 データクレンジングに関する考察

これまでの結果により得られた知見を整理する。まず図-6より、train 80%モデルと、標準偏差 10-40、平均画素値、ch 突出度といった弱いクレンジングのデータセットで作成した3つのモデルの結果を比較すると、recall に比べて precision に大きな差があることが分かった。さらに、図-7より、train 80%モデルにネガティブ画像を入力すると0に近い値を、ポジティブ・ノイズ画像を入力すると0ではない何らかの数値を出力するよう学習されており、さらにエンコーダーが出力したポジティブ・ノイズ画像の特徴量空間に違いがあることが分かった。

以上の2点より、クレンジングしたデータセットから作成したモデルは、正常な護岸画像(ネガティブ画像)以外をエンコーダーに入力すると、正常な護岸画像とは明らかに異なる特徴量空間に落とし込むことができるが、その空間内でひび割れの特徴と、それ以外の特徴を識別する能力に欠けたため、推論精度が悪化したのだと推察した。よって本検証の教師画像に含まれるノイズ画像は、学習において不要なものではなく、正常な護岸以外の特徴量空間に内包される、ひび割れの特徴量空間の境界条件をより明確にするために必要であったと考える。

本研究で検討した画素に着目した基礎的な画像特徴によるデータクレンジングは、ある1種の意匠の護岸ブロックにおけるひび割れセグメンテーションにおいて、学習精度を低下させると判断した。

## 4. 最後に

### 4.1 本研究の成果

本研究では河川の護岸ひび割れを対象としたセグメンテーションタスクにおける、データクレンジング、特にノイズ画像の除去がモデルの精度に与える影響を検討した。

その結果、ある1種の意匠の護岸ブロックにおけるひび割れセグメンテーションにおいて、ノイズ画像はひび割れの特徴量空間の境界条件を定める上で重要で、精度向上に欠かせないデータであることが分かった。

### 4.2 今後の課題

本論文で示した結果は、1種の意匠の護岸でのセグメンテーションタスクでの検証結果であるため、複数の意匠の護岸に対応した汎用性のあるモデルを作成する場合は、データクレンジングに効果がある可能性がある。特に1種の意匠を対象としたモデルでは、ネガティブ画像のバリエーションが極端に少ないため、バリエーションに富んだネガティブ画像がデータセットに含まれる場合、ネガティブ画像の特徴量分布にバリエーションが生まれ、ひび割れを表す特徴量空間にも変化が生じることが推察される。したがってネガティブ画像のバリエーションが増えた場合に、ノイズ画像がもたらす効果について、改めて調査を行うべきと考える。

データセット内のポジティブ、ネガティブ、ノイズ画像の比率も学習精度に関わる重要な要素であると考えられる。護岸のひび割れは発生頻度が低く、全クロープ画像のpx数とアノテーションしたpx数との割合を取ると、老朽化の激しい河川であっても1%以下とかなりインバランスな状態にある。本検証で用いたデータセットにおいても8000枚弱はアノテーションしたpxが0であり、ネガティブ過多の状態にある。よってネガティブデータのダウンサンプリングの効果の検証を行い、より高精度なモデルを作成するために必要なポジティブ、ネガティブ、ノイズ画像の比率を明らかにする必要があると考える。

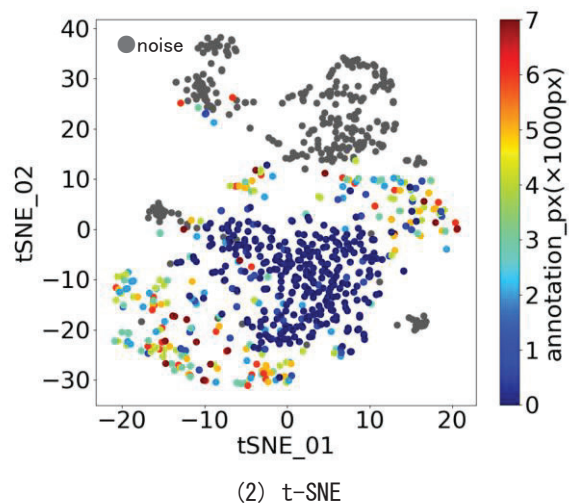
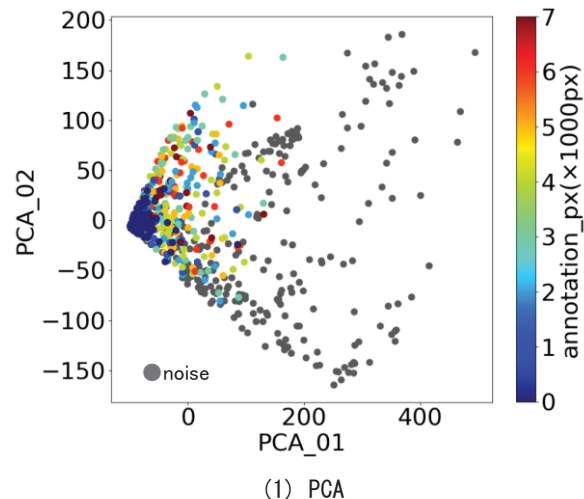


図-7 中間層の次元削減結果

## 参考文献

- [国土交通省 2020] 国土交通省: 令和2年版 国土交通白書, p.141-148.
- [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.5, p.9, p.19, 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.
- [吉田 2020] 吉田龍人, 藤井純一郎, 大久保順一, 天方匡純: 護岸のひび割れ検出 AI の作成とオルソ画像を活用したひび割れ幅の評価, AI・データサイエンス論文集 1 巻 J1 号, 土木学会, 2019.
- [Jeatrakul 2010] Piyasak Jeatrakul, Kok Wai Wong, and Chun Che Fung: Data Cleaning for Classification Using Misclassification Analysis, Journal of Advanced Computational Intelligence and Intelligent Informatics, 14 (3), pp.297-302, 2010.
- [Sun 2017] Chen Sun, Abhinav Shrivastava, Saurabh Singh, Abhinav Gupta: Revisiting Unreasonable Effectiveness of Data in Deep Learning Era, International Conference on Computer Vision, IEEE, 2017.