

Fully Convolutional Networkを用いた インフラ点検におけるひび割れの自動検出

木村 宇任¹ アランニャ・クラウス² 櫻井 鉄也²

概要：インフラの老朽化に伴うひび割れは重大な事故の要因となるため、継続的な点検によってひび割れを発見し適切な修繕を行う必要がある。現在、主に人による目視検査や打音検査が行われているが、時間や人員コストがかかるため、画像解析によるひび割れの自動検出が期待されている。ひび割れの自動検出に関する従来手法の課題として、検出精度が十分でないことや、計算時間がかかるといったことが挙げられる。このような課題に対して本論文では、画像の領域分割問題において高い性能を示している Fully Convolutional Network を用いた手法を提案する。ひび割れ画像のデータセットを用いた評価実験によって、提案手法が従来手法より優れていることを示した。

キーワード：ひび割れ、画像認識、Fully Convolutional Network, Deep Learning, 未然防止

Automatic Crack Detection in Infrastructure Inspection Using Fully Convolutional Network

TAKATO KIMURA¹ ARANHA CLAUS² TETSUYA SAKURAI²

Abstract: It is necessary to detect cracks of infrastructure caused by aging by periodic inspection and make appropriate repairs, since they become a factor of serious accidents. Currently, manual visual inspection and hammering tests are mainly used to detect cracks, but it takes time and labor cost. For that reason, automatic detection of cracks by image analysis is necessary. Conventional methods of automatic crack detection have problems with detection accuracy and computation cost. We propose a novel crack detection method using Fully Convolutional Network which shows high performance in semantic segmentation of image. We show that the proposed method is superior to conventional methods by evaluation experiments using dataset of crack images.

Keywords: Crack Detection, Image Recognition, Fully Convolutional Network, Deep Learning, Prevention

1. はじめに

高度経済成長期に集中的に整備された道路やトンネル、橋などのインフラの老朽化が急速に進んでおり、それに伴いひび割れや剥離などの変状が増加している。老朽化による事故を未然に防ぐために、継続的な点検によって変状を

発見し、適切な修繕を行う必要がある。インフラの変状の中で重要なものの一つが図1のようなひび割れである。現在、ひび割れの検出のために人による目視検査や打音検査が行われているが、人手のみで膨大な数のインフラ構造物を全て点検することは困難である。そのため、点検用のロボットなどによって撮影したインフラ構造物の表面画像から、画像解析によりひび割れを検出することで、点検作業を自動化することが期待されている。

画像解析によるひび割れの自動検出について、現在に至るまで数多くの手法が提案されている [1]。初期の研究ではしきい値処理やモルフォロジー演算、エッジ検出といった

¹ 筑波大学大学院システム情報工学研究科コンピュータサイエンス専攻

Department of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba

² 筑波大学システム情報系情報工学域

Division of Information Engineering, Faculty of Engineering, Information and Systems, University of Tsukuba

た基本的な画像処理を用いた手法が提案されている [2–9]. しかし, これらの手法にはノイズが多く, ひび割れが鮮明に写っていない画像については検出が難しいといった問題がある.

これに対して近年は, 機械学習を用いることでより精度を高めた手法が提案されている. Shi ら [10] は, ランダムフォレスト [11] によるエッジ検出手法である Structured Random Forest (SRF) [12] を用いた手法を提案しており, SRF を用いることで精度の向上を図るとともに, 近傍画素の画素値や勾配といった求める際の計算コストが低い特徴量のみを用いることで, 検出の高速化を図っている. Zhang ら [13] は, 畳み込みニューラルネットワーク (CNN: Convolutional Neural Network) を用いた手法を提案しており, CNN を用いることで, 分類器に加えて特徴抽出器も学習によって得ることで精度の向上を図っている.

Shi や Zhang らの手法では, 着目画素を中心とした一定サイズの小領域 (ウィンドウ) を入力画像として着目画素のクラス認識を行い, 1 画素ごとにウィンドウをずらすことで画像全体からひび割れを検出する. これはスライディングウィンドウ法と呼ばれる方法であるが, 以下の欠点を持つ.

欠点 1 画素数と同じ回数のクラス認識を行う必要があるため膨大な計算コストがかかる.

欠点 2 学習と予測に用いることができるのはウィンドウ内の局所的な情報のみである.

欠点 3 領域の境界を画素単位で正確に検出することができない.

これらの欠点はひび割れ検出において以下のような問題点となる.

問題点 1 ひび割れの検出を必要とするインフラの数は膨大であることから, 計算コストを削減する必要がある.

問題点 2 ひび割れの多くが細長い形状をしているため, 検出において画像の局所的な情報のみならず大域的な情報も活用することが必要である.

問題点 3 ひび割れの太さなどの情報から対処方法を判断するため, ひび割れを画素単位で正確に検出する必要がある.

これらの従来手法における問題点を改善するために, 本研究では Fully Convolutional Network (FCN) [14, 15] を用いたひび割れ検出手法を提案する. FCN は予測クラスではなく入力画像と同じスケールの画像を出力する CNN である. スライディングウィンドウ法の持つ欠点 1 については, FCN は検出対象の画像を直接入力画像とすることができるため, 画像中の同じ箇所を複数回解析する必要がなくなり計算コストがより少なくて済む. 欠点 2 については, FCN はウィンドウではなく画像全体に対して解析を行うため, より大域的な情報を学習と予測に用いることができる. また, 欠点 3 については, 出力が画像であることから正解

の領域分割画像を直接教師データとして用いることができるため, 画像間の対応関係を直接モデルに学習させることが可能であり, 領域の境界をより正確に検出することができる. このように FCN を用いることで, スライディングウィンドウ法を用いた従来のひび割れ検出における問題を改善することができる. 提案手法では, 医療画像の領域分割において高い性能を持つ U-Net [15] をベースとした FCN を用いる.

本論文では 2 章で FCN について述べ, 3 章では FCN を用いたひび割れの検出手法を示す. 4 章では道路のひび割れ画像のデータセットに対して提案手法を適用し, 検出精度と実行時間の評価を行うとともに, 従来手法との比較を行うことで提案手法の有用性を示す.



図 1 ひび割れの例 [10, 16]

2. FCN

FCN は特殊な構造を持つ CNN であるため, まず CNN について述べた後に FCN について述べる.

2.1 CNN

CNN は, カーネルによる畳み込み演算を用いた順伝搬型ニューラルネットワークである. ニューラルネットワークが隣接層のユニット全てが結合された全結合層のみで構成されているのに対して, CNN は全結合層に加えて, 隣接層間の特定のユニットのみが結合された畳み込み層とプーリング層を持つ. 畳み込み層はカーネルによる畳み込み演算によって入力データの特徴を抽出する働きをする. 畳み込み層の出力画像は特徴マップと呼ばれる. 多様な特徴を抽出するために複数種類のカーネルを用いることができ, カーネルを K 種類使用した場合, 同一の入力から K 枚の異なる特徴マップが得られる. 特徴マップの種類はチャンネルとも呼ばれ, 特徴マップは幅, 高さ, チャンネルの 3 次元構造を持つ. プーリング層は特徴マップの局所領域の統計量を計算することで, 微小な歪みや平行移動に対する頑健性を持たせる働きをする. また, プーリング層によって特徴マップのダウンサンプリングを行うことで, 小さい

カーネルでも大域的な特徴を抽出することが可能になる。よく用いられる統計量は最大値であり最大プーリングと呼ばれる。

CNN は畳み込み層とプーリング層の繰り返しによる特徴マップの抽出部分と、特徴マップからクラス分類や回帰を行う全結合層部分で構成される。このように分類器に加えて、特徴抽出器を学習によって得ることができる点がCNNの利点であり、画像認識分野における物体カテゴリ認識などでハンドクラフト特徴量を用いた手法よりも高い精度を示している [17]。

2.2 FCN

画像単位でのクラス分類タスクに対して、画像内の物体のクラスとその存在領域まで推定するタスクはセマンティックセグメンテーションと呼ばれる。CNNを用いてセマンティックセグメンテーションを行う手法として、Longら [14] によって全結合層を持たないCNNであるFCNが提案されている。畳み込み層で保持されている空間的な情報は全結合層において捨てられるが、CNNの全結合層を畳み込み層によって置き換えることで、クラス確率のヒートマップを出力するネットワークに変換することができる。しかし、畳み込み層とプーリング層の繰り返しを経た特徴マップは低解像度になっており、空間的な情報が少なくなっているため良好な領域分割結果が得られない。そこで、FCNでは特徴マップのアップサンプリングを行う畳み込み層であるDeconvolution層によって、低解像度の特徴マップを入力画像と同じ解像度まで拡大したものを出力とする。これにより、領域分割の正解画像を直接教師データとして用いることができるため、物体が画像中のどこにあるのかという情報を学習させることが可能になり、セマンティックセグメンテーションのベンチマークにおいて全結合層を持つCNNを用いた手法よりも高い精度を示している [14]。また、CNNとスライディングウィンドウ法を用いたセマンティックセグメンテーションを行う場合、畳み込み層のカーネルは画像中の同じ箇所に対して何回も適用されることになるが、FCNは全結合層を持たないため入力画像サイズが任意であるため、画像全体を入力とすることで同じ箇所に同じカーネルが何回も適用されることがなくなり、より高速にセマンティックセグメンテーションを行うことができる。

FCNを生体医学画像のセマンティックセグメンテーションに適用した手法として、U-Netが提案されている [15]。U-NetはContracting path, Expansive path, Shortcut connectionの3つの経路によって構成されている。Contracting pathは通常のCNNと同様な特徴マップの抽出部分であり、畳み込み層による特徴の抽出とプーリング層によるダウンサンプリングを行う。Expansive pathは畳み込み層による特徴の抽出とDeconvolution層によるアップサンプリングを行

うことで、低解像度のContracting pathの出力マップを入力画像と同じ解像度まで拡大する。これら2つの経路は他のFCNと同じ構造であるが、U-Netの大きな特徴はShortcut connectionを持つ点である。Shortcut connectionは、同一スケールのContracting pathの特徴マップとExpansive pathの特徴マップを、チャンネルで連結する構造である。Shortcut connectionによって、プーリング層によるダウンサンプリングを繰り返すうちに失われる局所的な特徴を保持することが可能であり、層を深くした場合でも画像の大域的特徴と局所的特徴の両方を伝搬することが可能となる。U-Netは生物組織画像内の細胞のセマンティックセグメンテーションにおいて高い精度を示している。

3. FCNを用いたひび割れの検出

本研究では、U-NetをベースにしたFCNによるひび割れの自動検出手法を提案する。

3.1 FCNの構造

本手法で用いるFCNの構造について説明する。構造の概略を図2に示す。U-Netと同様に各ブロックでは、小さいサイズのカーネルによる畳み込みが2回繰り返された後に最大プーリング層もしくはDeconvolution層が配置される。この構造はVGG-16 [18] と呼ばれるCNNをベースにしており、小さいサイズのカーネルによる畳み込みと非線形な活性化関数を多く繰り返すことで、メモリを削減しつつモデルの非線形性を高めることを目的としている。

Contracting pathでは以下のブロックが4回繰り返される。

- カーネルサイズが 3×3 の畳み込み層
- カーネルサイズが 3×3 の畳み込み層
- カーネルサイズが 2×2 の最大プーリング層

最大プーリング層によって特徴マップを $1/2$ 倍のサイズにダウンサンプリングする。最終的に画像は $1/16$ のサイズになる。Expansive pathにおいても同様に、以下のブロックが4回繰り返される。

- カーネルサイズが 3×3 の畳み込み層
- カーネルサイズが 3×3 の畳み込み層
- カーネルサイズが 3×3 のDeconvolution層

Deconvolution層によって特徴マップを2倍のサイズにアップサンプリングする。そして、最後に以下の出力ブロックによって特徴マップを一つのチャンネルに統合する。

- カーネルサイズが 3×3 の畳み込み層
- カーネルサイズが 3×3 の畳み込み層
- カーネルサイズが 1×1 の畳み込み層

出力層の活性化関数にはシグモイド関数を用い、出力層以外の活性化関数にはReLU関数を用いる。出力層以外の畳み込み層とDeconvolution層において、活性化関数への入力の前にBatch normalization [19]を行う。Batch normalizationはバッチごとに特徴マップの平均を0、標準偏差を1に

正規化する処理であり、学習の収束速度の向上や正則化による汎化性の向上、パラメータの初期値依存性の緩和といった効果を持つ。Shortcut connection では同一スケールの Contracting path ブロックの出力マップと、一つ前の Expansive path ブロックの出力マップをチャンネルで連結したものを、Expansive path ブロックの入力マップとする。

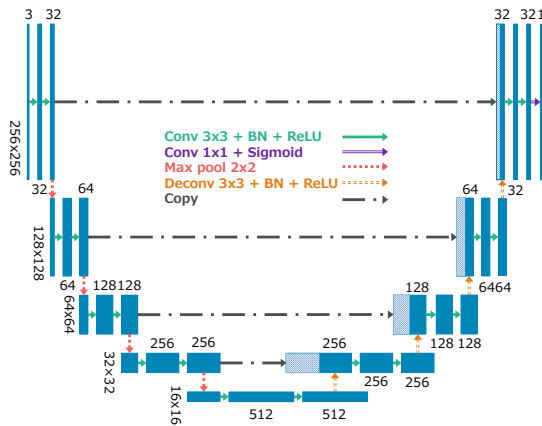


図2 FCNの構造（入力画像のサイズは任意であり、ここでは便宜上256×256とした。）

3.2 モデルの学習

重みとバイアスの学習には最適化手法である Adam [20] を用いたミニバッチ学習を行う。損失関数には以下の式で表されるバイナリクロスエントロピー関数を用い、これを最小化することで重みとバイアスを学習する。

$$L(\mathbf{w}) = - \sum_i^N [g_i \log p_i + (1 - g_i) \log (1 - p_i)].$$

ここで N はミニバッチにおける総画素数、 p_i は i 番目の画素に対する FCN の出力値、 g_i は i 番目の画素がひび割れの場合は 1、そうでない場合は 0 をとる。学習におけるハイパーパラメータは表 1 の通りである。

表 1 学習におけるハイパーパラメータ

学習係数	10^{-4}
ミニバッチサイズ	16
重みの初期値	Glorot の正規分布 [21] に従う乱数
バイアスの初期値	0

バリデーションデータに対する損失関数が 10epoch（反復回数）連続で減少しなくなったときに学習を早期終了し、バリデーションデータに対する損失関数が最も小さい値を示した epoch のモデルを最終的なモデルとする。バリデーションデータに対する損失関数の値が 3epoch 連続で減少しない場合には、次の epoch における学習係数を 1/2 に減らすことでより最適なパラメータを探索する。その際に学習係数の最小値は 10^{-5} とする。

3.3 Data Augmentation

本手法では、学習の際に Data Augmentation を行う。Data Augmentation とは、学習データに対して変換を施すことで学習データ数を擬似的に増やす手法であり、CNN の学習に用いることでモデルの汎化性能が向上することが知られている。各 epoch ごとに学習用画像に対して、回転、拡大・縮小、ガンマ補正の変換をランダムなパラメータで施す。各変換のパラメータの範囲は表 2 の通りである。また、学習用画像としてデータセットの画像をそのまま用いるのではなく、画像中から一定サイズの矩形領域をランダムにサンプリングしたものを用いる。これによって、データセットの画像に対してそのまま Data Augmentation を行う場合よりも多様な変換を施すことが可能になる。

表 2 Data Augmentation のパラメータ (θ は回転角、 σ はスケールファクタ、 γ はガンマ補正のガンマ値を表す。)

変換方法	パラメータの範囲
回転	$\theta = [0, 2\pi)$
拡大・縮小	$\sigma = [0.9, 1.1]$
ガンマ補正	$\gamma = [0.8, 1.2]$

4. 評価実験

提案手法の性能について道路のひび割れ画像のデータセットを用いて評価する。従来手法の SRF を用いた手法 [10] と CNN を用いた手法 [13] との比較を行う。それぞれを従来手法 1、従来手法 2 と呼ぶ。各手法の性能について妥当な比較を行うために、 K -fold 交差検証を用いた評価を行う。本評価実験では $K = 3$ とする。

提案手法と従来手法 2 については Python3 及びライブラリの Keras と TensorFlow, scikit-image を用いて実装した。また、従来手法 1 については公開されている Matlab のコードを用いた。学習時とテスト時の計算環境を表 3 に示す。

表 3 計算環境

	学習	テスト
CPU	Intel Core i5 3.2 GHz	Intel Core i5 2.7 GHz
メモリ	16GB	16GB
GPU	NVIDIA GeForce GTX 1060 6GB	

4.1 データセット

用いるデータセットは CrackForest Dataset [10, 16] であり、サイズが 480×320 のひび割れ画像 118 枚と、各画像のひび割れ位置を示した二値画像で構成されている。また、学習データとバリデーションデータについては、各画像から 256×256 の矩形領域をランダムに 50 枚ずつサンプリングしたものを用いる。

4.2 評価指標

評価指標として Dice の類似度係数 (Dice Similarity Coefficient, DSC) を用いる。DSC は二つの集合間の類似性を表す指標である。二つの集合が一致しているほど DSC は 1 に近い値を示す。DSC は以下の式で求められる。

$$DSC = \frac{2|P \cap G|}{|P| + |G|}.$$

ここで、 P は手法がひび割れと予測した画素の集合、 G は真のひび割れの画素の集合、 $|\cdot|$ は集合の要素数を表す。比較する 3 つの手法では各画素がひび割れである確率が $[0, 1]$ の範囲で出力されるため、DSC を求める場合しきい値を設定する必要がある。従来手法については元論文において示されているしきい値を用い、提案手法については 0.5 をしきい値とする。

4.3 実験結果

各手法をテストデータに適用した際の評価指標と実行時間の平均と標準偏差を表 4 に示す。表 4 を見ると DSC について提案手法が最も高い値を示しており、従来手法と比較して画素単位でより正確にひび割れを検出していることがわかる。また、提案手法の実行時間は従来手法 1 の約 2 倍となっているが、従来手法 2 に対しては約 0.015 倍になっており、CNN とスライディングウィンドウ法を用いた場合と比較して大幅に高速化されている。CNN とスライディングウィンドウ法を用いた場合に対して、FCN では画像中の同じ箇所に何回も同じカーネルを適用することがないために、高速化されたと考えられる。

表 4 各手法の評価指標と実行時間

手法	DSC	実行時間 (秒)
従来手法 1	0.3872±0.03629	2.269±0.3615
従来手法 2	0.3205±0.03001	387.8±22.75
提案手法	0.7122±0.007473	5.933±0.6280

各手法の結果画像を図 3 に示す。図 3 から以下のことがわかる。

- スライディングウィンドウ法を用いた従来手法では、ひび割れが正解画像と比較して太く検出されているが、提案手法では正解画像と同じ太さで検出している。そのため、複数のひび割れが複雑に入り組んだ形状のひび割れについても正確に検出できている。
- 従来手法ではひび割れではない領域をひび割れとして検出してしまったものが多く見られるが、提案手法ではそのような誤って検出した領域はみられない。
- 従来手法では検出できていない細いひび割れを提案手法は検出できている。

このように提案手法は従来手法と比較して、ひび割れ自体の検出精度が高いのに加えて、ひび割れの細かい形状まで

画素単位で正確に検出することに成功している。これは FCN の利点である大域的な情報と局所的な情報を学習と予測に利用できるという点と、画像間の対応関係を直接モデルに学習させることができるという点が有効に働いたためであると考えられる。

5. おわりに

本研究では FCN を用いたひび割れの検出手法を提案し、評価実験において検出精度と実行時間について、従来手法との比較を行なった。それによって提案手法が従来手法と比較して、より高精度かつ高速なひび割れ検出が可能であることを示した。

提案手法をより有用なものとするために、以下のような改良が考えられる。

(1) 検出したひび割れの分類

インフラの点検においてはひび割れを検出し、ひび割れの太さなどの情報からひび割れに対する対処方法を判断する必要がある。提案手法ではひび割れの検出のみを行うため対処方法の判断は人が行わなければならない。そのため、FCN を用いてひび割れを検出し、さらに検出したひび割れに対処方法ごとに自動で分類する手法を組み合わせることで、提案手法はより有用なものになると考えられる。

(2) ひび割れの歪みに対する頑健性

実際に点検用のロボットなどにおいてインフラ構造物を撮影する際、撮影角度や構造物との距離はばらついたものになると考えられるため、ひび割れが歪んだ形で画像に写ってしまう。今回用いたデータセットには、撮影条件によってひび割れが歪んだ形で写った画像は含まれていない。そのため、提案手法を実際に点検で使用することを想定した場合、そのような画像を含むデータセットを用いて、提案手法が歪みに対してどれ程の頑健性を持つのか検証する必要がある。提案手法では被写体が歪んだ形で写ることを考慮していないため、歪みに対する頑健性は低いと予想される。しかし、Data Augmentation においてシアー変形などを施すことで、擬似的に歪んだ画像を生成することにより、頑健性を高めることができると考えられる。

参考文献

- [1] Mohan, A. and Poobal, S.: Crack detection using image processing: A critical review and analysis, *Alexandria Engineering Journal* (2017).
- [2] Nishikawa, T., Yoshida, J., Sugiyama, T. and Fujino, Y.: Concrete crack detection by multiple sequential image filtering, *Computer-Aided Civil and Infrastructure Engineering*, Vol. 27, No. 1, pp. 29–47 (2012).
- [3] Zhu, Z., German, S. and Brilakis, I.: Visual retrieval of concrete crack properties for automated post-earthquake structural safety evaluation, *Automation in Construction*, Vol. 20, No. 7, pp. 874–883 (2011).
- [4] Yamaguchi, T. and Hashimoto, S.: Fast crack detection method for large-size concrete surface images using

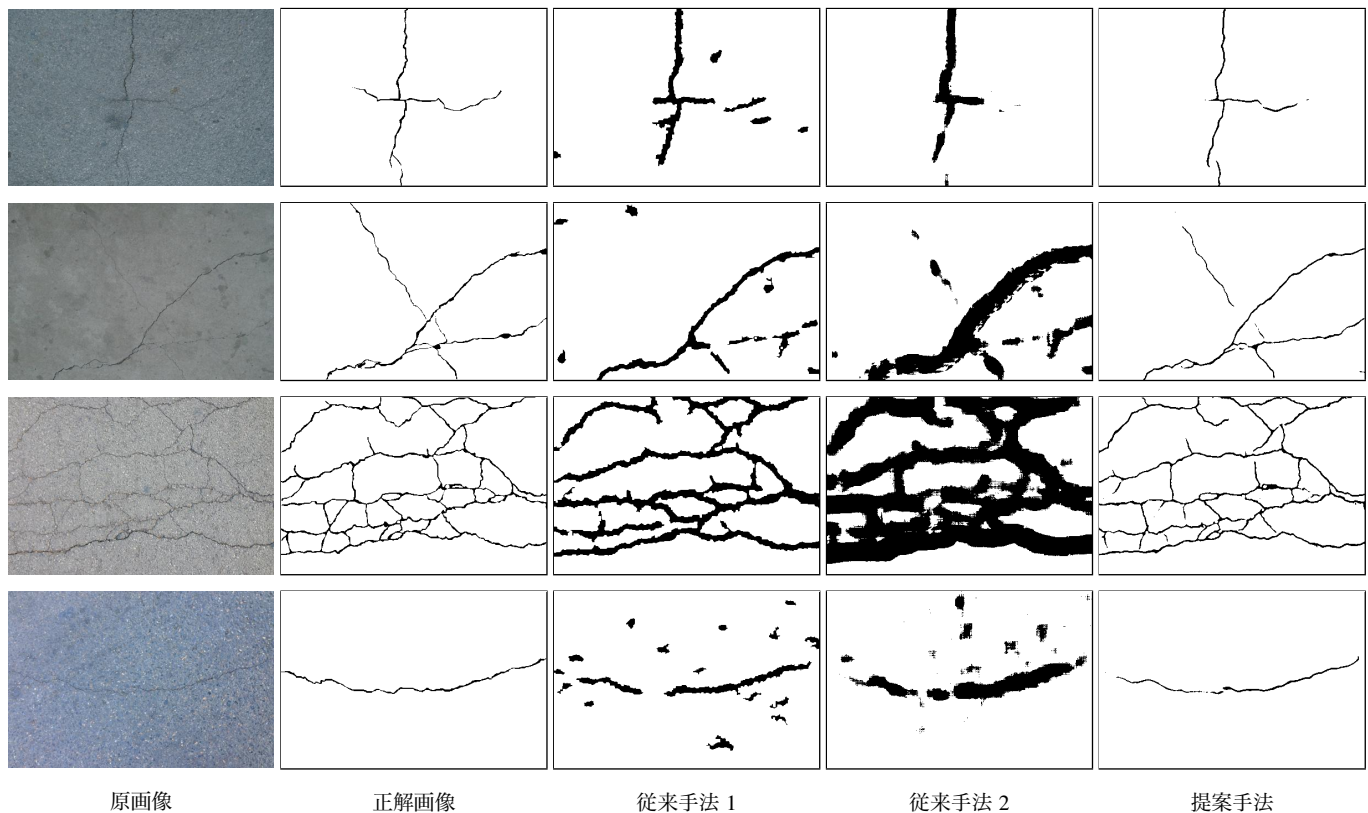


図3 テストデータに対する各手法の適用結果

- percolation-based image processing, *Machine Vision and Applications*, Vol. 21, No. 5, pp. 797–809 (2010).
- [5] Yamaguchi, T., Nakamura, S. and Hashimoto, S.: An efficient crack detection method using percolation-based image processing, *Industrial Electronics and Applications, 2008. ICIEA 2008. 3rd IEEE Conference on*, IEEE, pp. 1875–1880 (2008).
- [6] Tong, X., Guo, J., Ling, Y. and Yin, Z.: A new image-based method for concrete bridge bottom crack detection, *Image Analysis and Signal Processing (IASP), 2011 International Conference on*, IEEE, pp. 568–571 (2011).
- [7] Nguyen, H.-N., Kam, T.-Y. and Cheng, P.-Y.: A novel automatic concrete surface crack identification using isotropic undecimated wavelet transform, *Intelligent Signal Processing and Communications Systems (ISPACS), 2012 International Symposium on*, IEEE, pp. 766–771 (2012).
- [8] Adhikari, R., Moselhi, O. and Bagchi, A.: Image-based retrieval of concrete crack properties for bridge inspection, *Automation in construction*, Vol. 39, pp. 180–194 (2014).
- [9] Lim, R. S., La, H. M., Shan, Z. and Sheng, W.: Developing a crack inspection robot for bridge maintenance, *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, IEEE, pp. 6288–6293 (2011).
- [10] Shi, Y., Cui, L., Qi, Z., Meng, F. and Chen, Z.: Automatic road crack detection using random structured forests, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 17, No. 12, pp. 3434–3445 (2016).
- [11] Breiman, L.: Random forests, *Machine learning*, Vol. 45, No. 1, pp. 5–32 (2001).
- [12] Dollar, P. and Zitnick, C. L.: Fast edge detection using structured forests, *IEEE transactions on pattern analysis and machine intelligence*, Vol. 37, No. 8, pp. 1558–1570 (2015).
- [13] Zhang, L., Yang, F., Zhang, Y. D. and Zhu, Y. J.: Road crack detection using deep convolutional neural network, *Image Processing (ICIP), 2016 IEEE International Conference on*, IEEE, pp. 3708–3712 (2016).
- [14] Long, J., Shelhamer, E. and Darrell, T.: Fully convolutional networks for semantic segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440 (2015).
- [15] Ronneberger, O., Fischer, P. and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, *arXiv preprint arXiv:1505.04597* (2015).
- [16] Cui, L., Qi, Z., Chen, Z., Meng, F. and Shi, Y.: Pavement Distress Detection Using Random Decision Forests, *International Conference on Data Science*, Springer, pp. 95–102 (2015).
- [17] Krizhevsky, A., Sutskever, I. and Hinton, G. E.: Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, pp. 1097–1105 (2012).
- [18] Simonyan, K. and Zisserman, A.: Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* (2014).
- [19] Ioffe, S. and Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift, *International Conference on Machine Learning*, pp. 448–456 (2015).
- [20] Kingma, D. and Ba, J.: Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).
- [21] Glorot, X. and Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 249–256 (2010).