Vol.2021-HCI-195 No.8
Vol.2021-UBI-72 No.8
2021/11/30

物体検出を用いた道路陥没箇所検出モデルにおける合成画像を使用した学習の有効性の検討

坂内 理人[1,a)]　別宮 広朗[2,b)]　河崎 隆文[3,c)]　檀上 誠[4,d)]　中澤 仁[2,e)]

概要：日本全国に敷設された道路はおよそ 120 万キロにおよび，地球 30 周分にもあたる．そのうちの約 7 割以上が市町村によって管理されているが，国によって 5 年に 1 度の専門技能を持った人間による近接目視での定期点検作業が義務付けられており，地方公共団体にとって大きな負担となっている．また道路においては突発的な損傷が発生した場合，大きな事故につながる場合もあることから，5 年に 1 度の定期点検だけでなく，より頻繁に点検作業を行うことが望ましい．しかし，現在そのような突発的な道路損傷に関して，費用や労力的問題から多くの地方公共団体では市民からの通報に頼っていることが多いのが現状である．この問題を解決するための一つのアプローチとして，ドライブレコーダー画像から，ディープラーニングを用いた物体検出モデルを用いて自動検出する方法がある．しかし，検出できる損傷箇所の種類が不十分であることなど，まだまだ改善の余地が多い．高い精度を持ち，十分な種類の道路損傷箇所を検出できる物体検出モデルを作成することが難しい理由の一つとして，学習に使用できる道路損傷箇所の画像データ数の確保が難しいことが挙げられる．本研究では，道路損傷箇所が含まれる合成画像を作成し，学習に使用することで，データ数の少なさを補う手法を提案する．特に発生した際の往来への影響が大きいことや合成画像での再現が比較的容易であることから，道路上に空いた比較的深い陥没穴を検出対象として，合成画像を用いて学習を行った場合とそうでない場合とで比較実験を行い，合成画像を用いて学習を行うことの有効性について検証した．実験の結果から，合成画像を学習させた物体検出モデルは，合成画像に対して高い精度を出したが，実空間で定義された陥没穴と合成画像で定義した陥没穴とに乖離があることがわかった．しかし，作成した合成画像を学習したモデルは同様に作成した合成画像を高い精度で検出できたことから，合成画像を実空間上の陥没穴に近づけた場合，合成画像を用いることで実空間上の陥没穴に対するモデルの精度を補える可能性が示唆された．

1. はじめに

日本全国に敷設された道路はおよそ 120 万キロにおよび，地球 30 周分にもあたる．そのうちの約 7 割以上が市町村によって管理されているが，平成 24 年 12 月 2 日に起きた笹子トンネル天井板落下事故を契機に，道路法 [1] が改正され，それに伴い専門技能を持った人間による，全ての道路構造物に対する 5 年に 1 度の近接目視での定期点検が義務化されたことにより，地方公共団体は大きな負担を強いられている．

また道路においては多くの人々が徒歩だけでなく，車やバイク，自転車等で利用しており，突発的な損傷が発生した場合，大きな事故につながる場合もあることから，5 年に 1 度の定期点検だけでなく，より頻繁に点検作業を行うことが望ましい．しかし，現在そのような突発的な道路損傷に関して，費用や労力的な問題から多くの地方公共団体では市民からの通報に頼っていることが多い．突発的に発生した損傷を通報によって地方公共団体が把握することができるかどうかは人々の行動に依存する部分が大きいことから，より確実にそれらを捉えられる仕組みが必要である．

この問題を解決するための一つのアプローチとして，ディープラーニング [2] を用いた物体検出モデルを作成し，道路の損傷箇所をドライブレコーダー画像から自動検出す

1 慶應義塾大学法学部法律学科
Faculty of Law, Keio University, Yokohama, Kanagawa 223-8521, Japan
2 慶應義塾大学環境情報学部
Faculty of Environment and Information Studies, Keio University, Fujisawa, Kanagawa 252-0882, Japan
3 慶應義塾大学大学院政策・メディア研究科
Graduate School of Media and Governance, Keio University, Fujisawa, Kanagawa 252-0882, Japan
4 埼玉工業大学人間社会学部情報社会学科
Department of Informational Society Studies, Saitama Institute of Technology, Fukaya, Saitama 369-0293, Japan
a) rb@keio.jp
b) t18727hb@sfc.keio.ac.jp
c) drgnman@ht.sfc.keio.ac.jp
d) danjo@sit.ac.jp
e) jin@sfc.keio.ac.jp

る試みが一部ではなされている [10]．しかし，検出できる損傷箇所の種類が不十分であることなど，まだまだ改善の余地が多い．十分な種類の道路損傷箇所を高い精度を持って検出できる物体検出モデルを作成することが難しい理由の一つとして，物体検出モデルの学習に使用できるデータ数の少なさが挙げられる．一般的に，ディープラーニングを使用した物体検出モデルの精度をあげるにはより多くの画像データが必要であるが，大規模な道路損傷箇所に関するデータセットは [10] で作成されたもの以外，ほとんど存在しておらず，また日本の道路は前述した定期点検や補修工事によって良好な状態に保たれていることが多いため損傷が発生しづらく，損傷が発生しても市民からの通報等によって地方公共団体が迅速に補修するため，道路損傷箇所の画像データ数の確保が難しい．

本研究では，データ数の少ない道路損傷箇所を検出するモデルを作成するにあたって，その道路損傷箇所が含まれる合成画像を作成し，物体検出モデルの学習に使用することで，データ数の少なさを補う手法を提案する．道路損傷箇所には道路上のひび割れや白線の擦れ，ガードレールや標識の歪みなどさまざまな種類があるが，その中でも道路上にできた比較的深さのある陥没穴は発生した際の往来の安全に与える影響が大きく，他の道路損傷箇所に比べて危急性が高いと言える．また，陥没穴は場所によって形状に大きな変化がなく，合成画像を用いた再現が比較的容易なため，本研究では検出対象を道路上に発生した比較的深さのある陥没穴に限定することとした．また，道路損傷箇所検出モデルにおいては，地方公共団体の費用面，労力面での負担を軽減し，将来的にはゴミ収集車等の公用車を走行中にドライブレコーダーによって道路損傷箇所を自動検出することを想定しているため，今回使用する画像については走行中の車のドライブレコーダーからの画角に近いものを使用することとした．これらを用いた実験の結果から，合成画像を学習させた物体検出モデルは，合成画像に対して高い精度を出したが，実空間で定義された陥没穴と合成画像で定義した陥没穴に乖離があった．しかし，作成した合成画像を学習したモデルは同様に作成した合成画像を高い精度で検出できたことから，合成画像を実空間上の陥没穴に近づけた場合，合成画像を用いることで実空間上の陥没穴に対するモデルの精度を補える可能性が示唆された．

本稿では，1 章で本研究における背景及び概要を述べた．また，2 章では関連研究を示し，3 章では合成画像の作成と，実験の概要とその評価を，4 章では本研究の考察を，5 章では本研究の結論を述べる．

2. 関連研究

前田ら [10] は，各自治体から集めた画像から 15,435 の損傷箇所を含む 9053 枚の道路損傷箇所の大規模データセットが作成されており，そのデータセットを使って SSD[9] を用いて道路損傷検出モデルの作成を行なっている．この研究ではタイヤ痕，縦状の施工継ぎ目，横状の施工継ぎ目，横状の線状ひび割れ，亀甲状ひび割れ，轍や段差や剥離や陥没穴，横断歩道の掠れ，白線の掠れの 8 つの道路損傷箇所を検出対象としており，いずれにおいても高い検出精度が出ているが，検出対象のうち陥没穴として定義しているものは轍や段差，道路上の剥離も含んでいることからわかるように比較的浅いものが多い．これに対して本研究ではより深い陥没穴を検出対象としている．本研究で定義した陥没穴とこちらの研究で定義されている陥没穴の違いについては後述する．

工藤ら [8] は，視覚的に把握することが困難である機械製品の製造工場のバルクコンテナ内の在庫量をその画像からディープラーニングを用いて推定する試みを行なっており，稼働中の工場では学習データとなる実写画像を大量に蓄積することが困難であることを問題点として挙げており，部品数量の範囲が広いため，多クラス分類モデルではなく，深層学習の回帰モデルを用いた方が効率的であることを述べている．これらのことから実写画像だけでなく，自動生成した CG 画像を用いて回帰モデルを作るという手法を提案している．CG 画像のみを用いて回帰モデルの学習を行なった場合，実写画像のみを用いて学習を行なった場合に比べて，実写画像に対する推定精度が悪化したが，CG 画像に実写画像を混ぜることによって大幅な推定精度の向上が見られたことを示している．

筆者らはこれまで，白線などの道路上の路面標示の欠損を，ゴミ収集車のドライブレコーダー画像を用いてディープラーニングにより検出する技術を提案してきた [6], [7]．これらの技術では，画像認識技術に基づいて欠損を高い精度で検出できることを示した．しかし，白線の掠れをはじめとした路面標示の欠損は実世界に遍在することから，学習データの収集が容易である．これに対して稀有な事象については学習データの収集が困難であり，その不足を補う手法を確立する必要がある．

また，本研究で扱う道路の損傷箇所と同様に，医療用の画像データセットも一般的に不足している．このデータ数の不足は医師の診断を助けるコンピュータを使った診断システムを作成する際に，大きな障害となる．Han ら [4] は，データ数の確保が難しい事象の一つである肺結節の自動検出を行うモデルを作成する際に，GAN（Generative Adversarial Network）[3] に基づいたアプローチを用いて，専門医でも本物の肺結節と区別のつかないほど精巧な画像を作成するデータ補強手法を提案している．その作成した画像を用いることで，3 次元畳み込みニューラルネットワークを用いた検出では高い精度を実現し，肺結節におけるデータ数の少なさを克服できることを示している．

これらの研究成果から，人工的に作成した画像を用いることで実写画像だけでは用意できなかったパターンの学習

画像を大量に作成できることが明らかである．また人工的に作成した画像を実写画像と組み合わせることによって，学習データ数の少なさを補い，検出モデルの汎化性能を向上させられることが示されている．従って，陥没穴においても同様に，人工的に生成した画像を用いて既存データを補強することにより，検出精度の向上を図ることが可能だと考えられる．

3. 実験

実空間で撮影された実写画像と，CG 空間で生成した合成画像をそれぞれ用いてディープラーニングを用いた物体検出モデルを作成し，同モデルを用いて陥没穴を検出する際の精度を比較する．本章では実験の方法と結果を示し，次章でその考察をする．

3.1 CG を用いた合成画像の作成

CG を用いることによって，陥没穴の大きさや深さ，場所を任意に変更した合成画像を作成することが可能になるため，多種多様な陥没穴画像データを作成することができる．本実験では，三次元空間モデルを生成し，そこにドライブレコーダで撮影された実写画像をテクスチャマップした上で，大きさと位置の異なる陥没穴を路面上のみに作成する．実際の路面は樹木等の影や道路標示等により明るさや色が一定ではなく，実際の陥没はそのどこにでも起こり得る．そこで，そうした影や道路標示等をまたがる位置にも陥没穴を作成した．合成画像の例は事項に示す．

3.2 本研究で用いるデータセット

本研究では，実際に撮影された道路上の陥没穴の画像を 1155 枚，CG を用いて作成した道路上の陥没穴の合成画像を 120 枚使用した．以下では，実際に撮影された道路上の陥没穴の画像を「実写画像」，作成した道路上の陥没穴の合成画像を「合成画像」と省略する．

3.2.1 実写画像データセット

実写画像には [10] で公開されているデータセットの画像を使用した．本データセットに含まれる実写画像は全て日中に撮影されたもので，また車のダッシュボードに固定したスマートフォンで撮影されており**図 1** のような画角である．また，図 1（左）のように晴天のものや，図 1（中央）のような曇天のものがある．また図 1（左）のような比較的閑散としている場所のものや図 1（中央）のような山林部などで背景に多く緑が入っているもの，図 1（右）のように撮影場所が住宅街などで建物が密集している背景が写っているものなど，様々な背景を持つ画像が含まれている．

データセットの画像の中に映っている陥没穴は図 1（左）のようにほぼひび割れに近いようなものや，図 1（中央）の中央左下にあるような小さいもの，図 1（右）の左端中央にあるような表面が剥がれているに過ぎないものなど，深さが比較的浅いものがほとんどである．以上のように [10] で公開されているデータセットの画像に含まれる陥没穴は，本研究で検出対象とした陥没穴と比べて比較的浅いものが多いという問題点があるが，より適当なデータセットが見つからなかったことからこちらのデータセットを使用することとした．

3.2.2 合成画像データセット

本研究で作成した合成画像では実写画像と同じ画角の画像を用い，**図 2**（左）のように比較的小さな陥没穴から図 2（中央）のような比較的大きな陥没穴など様々な大きさの陥没穴を再現した．また，図 2（左），図 2（中央）のように道路の中央寄りにある陥没穴や図 2（右）のような道路の端にあるような陥没穴など様々な場所に陥没穴を作成している．

3.3 物体検出モデル

上記のデータを用いて物体検出モデルを作成した．物体検出モデルの作成には現在，研究等で幅広く使われている YOLOv5[5] を用いた．

実写画像，合成画像ともに 100 枚をそれぞれ物体検出モデルの学習用とし，残りの画像については作成した物体検出モデルの検証用とした．（以下，学習に使用していない画像をテスト画像と呼ぶ．）物体検出モデルは，実写画像 100 枚のみを学習に使用したモデル，合成画像 100 枚のみを学習に使用したモデル，実写画像と合成画像の計 200 枚を学習に使用したモデルの 3 つを作成し，それぞれ実写画像 1055 枚，合成画像 20 枚のそれぞれのテスト画像に対する精度を検証し，比較した．以下では，実写画像 100 枚のみを学習に使用したモデルを実写画像モデル，合成画像 100 枚のみを学習に使用したモデルを合成画像モデル，実写画像と合成画像の計 200 枚を学習に使用したモデルを実写画像/合成画像モデルと省略する．

3.4 結果

本実験の結果を表 1 に示す．AP@[.5] は $IoU = 0.5$ のときの AP を指し，AP@[.5:.95] はステップサイズ 0.05 で $0.5 <= IoU <= 0.95$ のときの各 IoU の値における AP の平均を意味する．実写画像モデルの実写画像に対する精度と，実写画像/合成画像モデルの実写画像に対する精度が約 0.28 とほぼ同じであるのに対し，合成画像モデルの実写画像に対する精度は極端に低い．また，合成画像モデルの合成画像に対する精度と実写画像/合成画像モデルの合成画像に対する精度が 0.995 とほぼ同じであるのに対し，実写画像モデルの合成画像に対する精度は極端に低い．

実写画像モデルを用いてテスト用実写画像に対して推論を行なった際の検出結果画像の例を，**図 3** に示す．実写画像の陥没穴をいくつか検出しているが，アンテナのように黒く細長い物体や，マンホールの蓋などの形状が似ており，

図 1 実写画像の例

図 2 合成画像の例

表 1 各モデルのテスト画像に対する検出精度

モデル	実写画像		合成画像	
	AP@[.5]	AP@[.5:.95]	AP@[.5]	AP@[.5:.95]
実写画像モデル	0.287	0.0991	0.0396	0.0109
実写画像/合成画像モデル	0.284	0.0871	0.995	0.743
合成画像モデル	0.00345	0.000844	0.995	0.776

光沢のある黒色らしい物体を誤検出する例が目立った．

実写画像モデルを用いてテスト用合成画像に対して推論を行なった際の検出結果画像の例を，**図 4** に示す．作成した合成画像を全く検出できておらず，影や実写画像を検出した時と同様にアンテナを誤検出する例が目立った．

合成画像モデルを用いてテスト用実写画像に対して推論を行なった際の検出結果画像の例を，**図 5** に示す．実写画像の陥没穴をほとんど検出できていないのがわかる．また，マンホールやアンテナといった実写画像モデルで誤って検出したものの検出もしてない．

合成画像モデルを用いてテスト用合成画像に対して推論を行なった際の検出結果画像の例を，**図 6** に示す．この場合，合成画像の陥没穴をほとんど検出している．また，誤検出もほとんど生じていない．

実写画像/合成画像モデルを用いてテスト用実写画像に対して推論を行なった際の検出結果画像の例を，**図 7** に示す．実写画像モデルを用いてテスト用実写画像に対して行なった推論結果とほとんど同じ検出結果となっているのがわかる．

実写画像/合成画像モデルを用いてテスト用合成画像に対して推論を行なった際の検出結果画像の例を，**図 8** に示す．合成画像モデルを用いてテスト用合成画像に対して行なった推論結果とほとんど同じ検出結果となっている．

4. 考察

本研究の実験やその結果の考察をまとめる．本研究の実験では CG を用いて本研究で定義した陥没穴を三次元空間上に作成し，その三次元空間モデルにドライブレコーダー画像をテクスチャマップすることによって合成画像を作成した．そして，YOLOv5[5] を用いて実写画像モデル，実写画像/合成画像モデル，合成画像モデルを作り，実写画像，合成画像のそれぞれのテスト画像に対する陥没穴の検出精度を比較した．

その結果，実写画像/合成画像モデルにおいて，実写画像と合成画像を混ぜて学習させたが，表 1 からわかるように，実写画像と合成画像をそれぞれ単体で学習させたときと同等の検出精度を得ることができた．しかし，実写画像/合成画像モデルにおいて，合成画像と実写画像と合成画像を混ぜて学習させたのにもかかわらず，実写画像と合成画像のそれぞれのテスト画像に対する検出精度が，実写画像モデルの実写画像に対する検出精度と，合成画像モデルの合成画像に対する検出精度に，それぞれほぼ同等であったことから，実写画像と合成画像が物体検出モデルの学習において，互いにほとんど影響を与え合っていないことがわかった．これは，それぞれの誤検出の対象の比較でも確認できる．実写画像モデルではアンテナのように細長く黒い物体やマンホールの蓋を誤検出したが，合成画像モデルではこれらの物体の誤検出は生じていない．このことからも，それぞれのモデルが学習した特徴量が全く異なること

図 3 実写画像モデルの実写画像に対する検出例

図 4 実写画像モデルの合成画像に対する検出例

Vol.2021-HCI-195 No.8
Vol.2021-UBI-72 No.8
2021/11/30

図 5 合成画像モデルの実写画像に対する検出例

図 6 合成画像モデルの合成画像に対する検出例

Vol.2021-HCI-195 No.8
Vol.2021-UBI-72 No.8
2021/11/30

図 7 実写画像/合成画像モデルの実写画像に対する検出例

図 8 実写画像/合成画像モデルの合成画像に対する検出例

がわかる．

以上のことから，物体検出モデルは今回使用した実写画像と合成画像をほぼ別個のものとして学習していると言える．本研究で検出対象とした陥没穴と，今回使用した実写画像で定義されている陥没穴は図 1 の実写画像と図 2 の合成画像からもわかるように，その深さや大きさがかなり異なるものであったことから，このような乖離が生じたと考えられる．

5. おわりに

本研究では，道路上の深い陥没穴に代表される，画像データ数が少なく，また発生した際の危急性が高い道路損傷箇所を検出対象とし，道路損傷箇所検出モデルの学習に合成画像を使用することの有効性を検証した．本研究では，比較的大きく深い穴を陥没穴と定義し，その定義に基づいて作成した合成画像を用いて実験を行なった．実験の結果から，合成画像を学習させた物体検出モデルは，合成画像に対して高い精度を示したが，実空間で定義された陥没穴と合成画像で定義した陥没穴に乖離があったため，実空間上の陥没穴を検出する物体検出モデルにおける合成画像を使用した学習の有効性について十分に検証できたとは言えない．以上の原因として本研究で定義した陥没穴に近い実写画像を十分に入手することができなかったことが挙げられる．

この問題を解決するために，今後は，合成画像で定義されているそれに近い陥没穴を実空間上で人為的に再現するなどして，適切な検証を行える学習データセットを構築していく必要がある．また，近年では GAN（Generative Adversarial Network）[3] に代表されるように，ディープラーニングを用いて本物に近い画像を自動生成する生成型深層学習が提案されていることから，調整する要素を指定できれば、より実空間上に発生する陥没穴に近い合成画像を作成可能と考えられる．このことから，本研究で合成画像として定義した陥没穴を実空間上で再現した後，その陥没穴を撮影して少数の実写画像データを作成し，それらの画像を参考に実空間上で発生した陥没穴に近いものを GAN を用いてより精巧に作成することにより，実写画像データ数の不足を補うことができる可能性が考えられ，今後はそのような検討を進めていく．

謝辞

本研究の一部は東京都江戸川区役所にご協力及び支援頂いた．

参考文献

[1] 道路法. https://elaws.e-gov.go.jp/document?lawid=327AC1000000180.
[2] Li Deng and Dong Yu. Deep learning: methods and applications. *Foundations and trends in signal processing*, Vol. 7, No. 3–4, pp. 197–387, 2014.
[3] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, Vol. 27, , 2014.
[4] Changhee Han, Yoshiro Kitamura, Akira Kudo, Akimichi Ichinose, Leonardo Rundo, Yujiro Furukawa, Kazuki Umemoto, Yuanzhong Li, and Hideki Nakayama. Synthesizing diverse lung nodules wherever massively: 3d multi-conditional gan-based ct image augmentation for object detection. In *2019 International Conference on 3D Vision (3DV)*, pp. 729–737. IEEE, 2019.
[5] Glenn Jocher, Alex Stoken, Ayush Chaurasia, Jirka Borovec, NanoCode012, TaoXie, Yonghye Kwon, Kalen Michael, Liu Changyu, Jiacong Fang, Abhiram V, Laughing, tkianai, yxNONG, Piotr Skalski, Adam Hogan, Jebastin Nadar, imyhxy, Lorenzo Mammana, AlexWang1900, Cristi Fati, Diego Montes, Jan Hajek, Laurentiu Diaconu, Mai Thanh Minh, Marc, albinxavi, fatih, oleg, and wanghaoyang0106. ultralytics/yolov5: v6.0 - YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCV DNN support, October 2021.
[6] Makoto Kawano, Kazuhiro Mikami, Satoshi Yokoyama, Takuro Yonezawa, and Jin Nakazawa. Road marking blur detection with drive recorder. In *2017 IEEE International Conference on Big Data (Big Data)*, pp. 4092–4097, 2017.
[7] Takafumi Kawasaki, Makoto Kawano, Takeshi Iwamoto, Michito Matsumoto, Takuro Yonezawa, Jin Nakazawa, and Hideyuki Tokuda. C's: Sensing the quality of traffic markings using camera-attached cars. *SICE Journal of Control, Measurement, and System Integration*, Vol. 10, No. 5, pp. 393–401, 2017.
[8] Tsukasa Kudo and Ren Takimoto. Cg utilization for creation of regression model training data in deep learning. *Procedia Computer Science*, Vol. 159, pp. 832–841, 2019.
[9] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pp. 21–37. Springer, 2016.
[10] Hiroya Maeda, Yoshihide Sekimoto, Toshikazu Seto, Takehiro Kashiyama, and Hiroshi Omata. Road damage detection using deep neural networks with images captured through a smartphone. *arXiv preprint arXiv:1801.09454*, 2018.