情報処理学会論文誌 Vol.60 No.10 1829–1844 (Oct. 2019)

GANonymizer：物体検出と敵対的生成を用いた映像匿名化手法

谷村 朋樹[1,a)] 河野 慎[2,b)] 米澤 拓郎[3,c)] 中澤 仁[1,d)]

受付日 2018年12月25日, 採録日 2019年7月3日

概要：都市の様子を記録した映像を分析することで，都市の状態を自動で把握，予測することが可能となる．しかし，都市の映像には人や車などのプライバシに関する物体が含まれているため，無加工の状態でネットワークに送信・共有したり，利用したりすることは難しい．結果，プライバシ情報が含まれる可能性のある映像は有効活用されないまま，消去されている場合も多い．本研究では，映像からプライバシに関する物体を自動で検出し，映像上から消去する GANonymizer を提案する．提案手法では，まず入力画像から深層学習を用いた物体検出技術を用いて，人や車などのプライバシに関する物体を検出する．そして敵対的学習によりトレーニングされたネットワークで，検出した物体部分の背景を生成し，元の画像に合成する．さらに，自然な背景生成が困難なケースに対応するため，2 つのネットワークの接続点に新たなパディング処理を施す手法を提案する．本研究では，実際に記録された都市映像の匿名化実験を行い，複数の指標で映像の自然さを定量的に評価するとともに，匿名化された箇所をマーキングしてもらうユーザ評価実験を行うことで，提案手法の有効性を評価した．

キーワード：プライバシ保護，都市画像・動画，匿名化，深層学習

GANonymizer: Image Anonymization Method Using Obejct Detection and Generative Adversarial Network

Tomoki Tanimura[1,a)] Makoto Kawano[2,b)] Takuro Yonezawa[3,c)] Jin Nakazawa[1,d)]

Received: December 25, 2018, Accepted: July 3, 2019

Abstract: Sharing and analyzing image data from ubiquitous urban cameras must enable us to understand and predict various contexts of the city. Meanwhile, since such image data always contains privacy data such as people and cars, we cannot easily share and analyze the data through the Internet for the viewpoint of privacy protection. As a result, most of the urban image data are only kept/shared within the camera owners or even discarded to reduce risks of privacy data leakage. To solve the privacy problem and accelerate sharing of urban image data, we propose GANonymizer that automatically detects and removes objects related to privacy from the urban images. GANonymizer combines two neural networks: 1) a network which detects objects related to privacy such as persons and cars in an input image using object detection network, and 2) a network that removes the detected objects naturally as though they do not exist originally. In addition, we propose two padding layers for removing the detected objects more naturally. Through our experiment of applying GANonymizer to urban video images, we confirmed that GANonymizer partially achieved natural removal of objects related to privacy.

Keywords: privacy protection, urban image anonymization, DNN

[1] 慶應義塾大学環境情報学部
Faculty of Information and Enviroment, Keio University, Fujisawa, Kanagawa 252–0882, Japan
[2] 東京大学大学院工学系研究科
Graduate School of Engineering, University of Tokyo, Bunkyo, Tokyo 113–0033, Japan
[3] 名古屋大学大学院工学研究科
Graduate School of Engineering, Nagoya University, Nagoya, Aichi 464–0861, Japan
[a)] tanimu@ht.sfc.keio.ac.jp
[b)] makora@ht.sfc.keio.ac.jp
[c)] takuro@nagoya-u.jp
[d)] jin@ht.sfc.keio.ac.jp

1. はじめに

都市には防犯カメラなどの固定カメラとドライブレコーダなどの移動カメラが存在し，都市の様子を撮影している．都市の移動カメラは空間網羅性が高いため，時間的変化がゆるやかな都市のインフラ（例：道路や街灯，街路樹など）を監視するには有効である．複数の移動カメラで撮影した映像を収集・蓄積し共有すれば，都市全体のインフラを監視できる．また，蓄積した映像をもとに，大規模な都市の映像データセットを構築することも可能となり，機械学習を用いた都市のインフラの自動監視や分析に活用することが可能となる．例として，都市を周回する車両からの映像を解析して，道路の劣化を自動で検出する研究 [13], [16] や，街路樹の桜の開花状況を測定する研究が行われている [17]．

しかし，都市の移動カメラ映像には通行人や車両のナンバープレートなどのプライバシ情報が含まれており，プライバシ侵害の危険性をつねに生じさせてしまう．その結果，現状ではプライバシ侵害の懸念もあり，移動カメラ映像は都市のインフラ監視や分析に利用されることなく，組織内の保有にとどまっているか，個人情報保護の観点から消去される場合も少なくない．一方，Geiger らが公開している KITTI データセット [8] のように共有されている都市画像のデータセットもある．しかし，これらの公開されている都市データセットは使用用途が学術目的に限定されていることが多く，都市インフラ監視などの目的で使用することは難しい．さらに，Geiger らも，データにプライバシに関する物体が写っている可能性があるため問題であると述べており，報告があればそのデータをデータセット中から削除する対策をとっている [8]．よって，今後これらのカメラ映像の利活用を推進していくためには，カメラ映像の匿名化手法が求められる．

Chinomi らは様々な画像や映像の匿名化処理を匿名化レベルを定義することによって整理した [5]．Chinomi らが定義した匿名化レベルは，匿名化対象の物体の情報の損失量に基づいて定義されているため，映像の利用者と被写体の意見を考慮して選択すべきであるとした．この Chinomi らの定義に基づき，本研究では都市の映像に写り込む不特定多数の被写体のプライバシ侵害を極限まで抑えるため，最も高い匿名化レベルである物体除去を目的として，匿名化手法を実現する．

映像から特定の物体を除去する手法としては，映像のフレーム間での背景差分を用いた手法が一般的である．これらの手法は，前景と背景を分離することで，移動する対象を前景として抽出し，前後のフレームを用いて背景をパッチする手法である [20]．このような背景差分を用いたパッチ処理は防犯カメラなどの固定カメラ映像の物体除去には有効である．しかし，ドライブレコーダのようなカメラの方向や走行速度が一定でない移動カメラ映像に対しては，機能しない．また，背景パッチの手法で，複数のドライブレコーダの映像を共有，データベース化し，パッチ処理を行う方法や，エッジ側で映像を保持しておき，そこから必要な背景パッチを取得してパッチ処理を行う方法も考えられる．しかし，これらのパッチを保持する手法は匿名化処理を施す前段階での共有や蓄積が必要であるため個人情報保護の観点から難しい．また，エッジ側でそれらのパッチを蓄積しておくにはデータ量が膨大である．以上より，映像に対して直接物体除去を行うことが難しいため，フレームごとに切り出して画像単位で処理する必要がある．

そこで本研究では，画像からプライバシに関する物体を自動で検出し，画像中から消去する GANonymizer を提案する．GANonymizer は，プライバシに関する物体を検出するネットワークと検出された物体の背景を再構成するネットワークの 2 つで構成されている．まず，プライバシに関する物体を検出するネットワークでは，入力された画像の中から人や車などのプライバシに関する物体の検出を行う．そして，検出された物体の背景を再構成するネットワークでは，検出された物体の箇所の背景を，画像全体をもとに生成する．これら 2 つのネットワークによって，プライバシに関する物体が除去された画像を出力することができる．一方で，プライバシに関する物体が画像の端に位置している場合や画像中で大きく写っている場合，ネットワークの性質から背景再構成に必要な情報を伝播させることが難しく，その背景を自然な形で再構成することは難しい．そこで，背景再構成を補助する 2 つのパディング層，Edge Shift Padding Layer（ESP）と Global Feature Padding Layer（GFP）を提案する．ESP は画像の端に写る物体の背景の再構成を補助するパディング層で，GFP は大きく写る物体の背景の再構成を補助するパディング層である．

ESP と GFP では，それぞれ挙動を決定するパラメータを用意し，最適なパラメータを決定する実験を行った．本研究では，まず提案する 2 つのパディング層の有効性を検証するため，実際の道路画像と白色画像を用いて，定性評価と Peak Signal-to-Noise Ratio（PSNR）と Structural Similarity（SSIM）による定量評価を行った．また，ユーザテストにより客観的に ESP と GFP の有効性を検証する実験もあわせて行った．さらに，GANonymizer 全体の有効性を評価をするにあたり，車載カメラで撮影した都市の画像に GANonymizer を適用し，出力結果をもとにした著者らによる定性評価と，ユーザテストによる客観的な定性評価を行った．実験に使用した画像は全 5,246 件で，画像データの性質からプライバシに関する物体は，人，車，バス，バイク，自転車とし，それらを画像中から除去した．結果として，GANonymizer は，画像中に存在するプライバシに関する物体の 6 割以上を人間が除去された箇所を特定不可能な精度で，除去できることを示した．

本論文の構成は以下のとおりである．まず，2 章では想定している都市映像の利活用の例を示す．その後それらの利活用の際に生じうるプライバシ問題と本研究の目的を述べる．3 章では，関連する映像匿名化の取り組みと，画像処理分野における物体除去研究について述べる．次に，4 章では提案する GANonymizer の全体構造と背景再構成を補助する 2 つのパディング処理の詳細について述べる．5 章では ESP と GFP のパラメータを決定するにあたり行った実験について述べる．そして，6 章では実際に撮影した都市画像に対して GANonymizer を適用した際の，ESP と GFP の有効性の検証と GANonymizer 全体における評価とその結果について述べる．最後に，7 章で本論文をまとめる．

2. 都市映像活用における問題

2.1 都市映像の利活用

都市の移動カメラから撮影された映像を収集，蓄積，共有することができれば，都市インフラなどの都市全体の静的な様子を監視することができる．都市には，ドライブレコーダのような移動カメラがいたるところに存在し，都市の状態をつねに撮影している．それら都市中の移動カメラの映像を蓄積し利用することで，都市インフラの監視に加えて，都市インフラの分析のための大規模データセットの作成も可能となる．

Kawano らは都市の道路画像を用いた道路上の標識劣化の自動検出を行った [13]．Kawano らは，都市中を網羅的に走行しているゴミ清掃車に着目し，ゴミ清掃車に取り付けられたドライブレコーダ映像をデータとして用いた．物体検出ネットワークで白線や道路標識のかすれを学習し，ドライブレコーダ映像から白線や道路標識のかすれを自動かつリアルタイムで検出するシステムの提案を行った．

Maeda らは Kawano らの道路の標識劣化に加えて，道路のひび割れなどの損傷の自動検出も行った [16]．Maeda らの場合は，スマートフォンアプリでの道路損傷検出を可能にすることを目的としているため，物体検出ネットワークの中でも計算コストが比較的少ない MobileNet SSD を使用して自動検出を実現した．また，都市を走る車からスマートフォンで撮影した画像に，白線のかすれや道路のひび割れなどの道路損傷のアノテーションを行ったデータセットを公開している．

Morishita らは移動カメラの映像を用いて季節変化を都市映像から取得し共有するシステムを提案している [17]．走行する車からスマートフォンカメラで街路樹を撮影，その映像から桜を検出し共有するシステム SakuraSensor を提案した [17]．SakuraSensor を使用することで，全国の桜の開花情報や満開情報を細かい時空間の粒度で取得することが可能となる．これらの研究から，移動カメラからの都市映像は，広範囲における都市の静的な状態を把握するのに有効であり，それらを共有，分析するなど，有効活用できることが分かる．

2.2 都市映像の利活用におけるプライバシ問題

都市映像を実社会で活用をする場合，映像を保存，蓄積，共有することが不可欠となる．たとえば，都市映像からの道路損傷の自動検出によって，道路修復業務の効率化を図ろうとする場合，道路損傷の位置情報の共有だけではなく，損傷状況を示す実際の画像の共有も求められる．我々は，神奈川県藤沢市の道路状況の監視を行っている職員に対して，道路損傷の自動検出技術 [13], [16] を業務に実際に組み込む際の問題点の聞き取り調査を実施した．結果，自動で道路損傷の場所を特定することができる技術は実際の業務でも有効であることが確認できた．加えて，道路損傷が検出された箇所を修復すべきかという最終判断は，少なくとも画像で実際の状況を見て行いたいということが分かった．そのため，実社会での活用を想定する場合，都市の道路映像を収集・蓄積・共有する需要は大きい．

都市映像を収集・蓄積・共有する際，障害となるのはプライバシ侵害の問題である．個人情報保護の観点から，都市映像を保有する企業や行政が，人や車などのプライバシに関する物体が含まれている都市映像を公開・共有することができない．さらに，公開や共有の規制だけでなく，それらの映像は一定期間保持した後，プライバシ侵害のリスク軽減のために消去されることが多い．ヨーロッパでも一般データ保護規則（General Data Protection Regulation：GDPR）が施行されるなど，個人情報を含め大量の情報が流動する現代において，プライバシ保護に関する関心は世界的に高まっている．Park らの研究でもドライブレコーダ映像の共有の際に，人々が最も関心を寄せるのはプライバシ保護であると報告されている [19]．そのため，都市映像を共有する際のプライバシ保護は必要不可欠となる．

映像共有の際のプライバシ保護には，大きく分けて 2 つの方法が存在する．1 つ目は，映像自体に含まれる人や車などのプライバシ情報に匿名化処理を施す方法である．たとえば，プライバシ情報が含まれるフレームの削除やプライバシに関する物体へのモザイクなどのマスク処理がこれにあたる．2 つ目は，映像をネットワーク上で安全に共有する方法である．これには，共有の際の暗号化通信やデータのトラッキング精度を高めるなどの方法がある．本研究では映像中のプライバシ情報に匿名化処理を施すことで，映像自体のプライバシ侵害の可能性を低下させ，映像の共有方法に依存しないプライバシ保護方法の構築を目的とする．

2.3 本研究の目的

映像の匿名化にはモザイクやシルエットのみを残すなど，多様な手法が存在するため，それらの中から都市映像のプ

ライバシ保護に最適な手法を検討することが必要となる．本研究では，Chinomi らが定義した匿名化レベル [5] を参照して，最適な匿名化手法の検討を行う．Chinomi らは，代表的な匿名化手法の匿名化度合いを表す匿名化レベルによって，匿名化手法の整理を行った．匿名化レベルは，何も匿名化処理を施さない無修正を最低レベルとして，その上に物体のシルエットのみを残す手法，物体にモザイクをかける手法などがある．最も匿名化レベルが高い手法は物体除去による透明化と定義されている．さらに，Chinomi らは，匿名化処理は画像中の個人に関わる物体に関する情報量を損失させる処理のことであり，画像の利用価値とトレードオフの関係にあるとしている．たとえば，モザイク処理であれば，物体の詳細な情報を失うが存在情報は残る．一方，物体除去による透明化を行った場合，匿名化された映像は物体の存在情報そのものを失うことになる．そのため，映像の匿名化処理を検討する際には，映像の利用者と被写体の両者の要望を考慮して，最適な方法を選択するべきであるとしている．

本研究では，Chinomi らが定義した匿名化レベルから，本研究の目的と被写体を考慮して，最も高い匿名化レベルの物体除去を選択した．まず，本研究が想定している応用先は，収集・蓄積した都市映像を用いて，道路の劣化状態や道路標識・ガードレールなどの破損状態，植樹帯の状態などの都市インフラの監視を行うことである．そのため，人や車などのプライバシに関する物体の情報は必ずしも必要ではない．また，市区町村などの小規模の自治体が，毎日や毎週など定期的に同地域の映像を収集し，監視する場合を想定すると，物体自体の存在を映像中から除去しないモザイクなどの匿名化処理では，不十分であると考える．たとえば，ある家の前に駐車されている車の有無などから，同宅における住人の外出パターンなどを抽出できる可能性がある．このように，プライバシに関する物体の詳細を隠すだけでは防げない，2 次的なプライバシ侵害の対処には物体除去による匿名化が必要となる．

一方，画像から物体を除去する際に，除去対象物体の裏にある背景を 100%の精度で推定することができなければ，画像としての利用価値は落ちる．この場合，物体除去による匿名化がされた画像を監視する際に，画像において物体が除去された箇所が，破損していたり劣化したりしていると判断されてしまう可能性もある．しかし，この画像の利用価値と匿名化はトレードオフの関係にある [5]．このトレードオフに対して，本研究ではプライバシ物体を除去することで都市映像の利用可能性を高めることを重視する．また，都市全体を撮影した映像には不特定多数の被写体が含まれるため，都市映像に写る全被写体の要望を考慮することはできない．そこで，人や車などの匿名化対象となる被写体を完全に除去できる物体除去を本研究の目的として設定することで，これらの問題を回避する．

3. 関連研究

3.1 映像匿名化

カメラ映像の匿名化は監視システムにおける必須要素として，さかんに研究が行われている [4], [21]．Cheung らは，動画に写る人物が所有する Radio-Frequency Identification（RFID）タグによって，匿名化が必要な人物を特定し，前後フレームの背景で埋めることでその人物を除去する手法と，撮影された動画を保持する統合的なシステムの提案を行った [4]．Cheung らの手法では，人物ごとに匿名化の必要性を判断することができ，RFID タグによる追跡を行うことで，正確に人物を特定・追跡し除去することができる．一方で，カメラに写る人物が限定されているような閉じられた空間における固定カメラでの運用を想定しているため，本研究で利用するような都市の移動カメラへの適用は難しい．

また，Yu らはドライブレコーダや監視カメラなどの小型の IoT デバイスでの動作を想定したプライバシ保護システム Pinto を提案した [26]．Pinto は匿名化の処理を 2 つに分割することで，画像処理の計算コストを抑え，IoT カメラに導入可能なシステムである．1 つ目の処理では，リアルタイムにモザイク処理を各フレームに施すことによって，動画の詳細を認識できないよう変換し，この映像をリアルタイムにユーザに提供する．そして，2 つ目の処理で撮影された動画から物体検出手法によって，人の顔やナンバプレートなどのプライバシ侵害の危険性が高い箇所を特定し，特定された箇所だけにモザイク処理を施した動画を作成する．2 つ目の処理で作成された動画はリアルタイムに取得することはできないが，後から取得することができ．動画の管理や利用ににおける認証システムまで提案されている．Pinto は本研究で対象とするドライブレコーダ映像にも適用可能である．しかし，プライバシに関する物体の一部に対してモザイクをかけるだけであるため，都市映像を公開することを想定した匿名化レベルとしては，不十分である．このように，カメラによる監視システムの多くはカメラ映像を保持することが目的で，本研究のようにカメラ映像を 2 次利用のために公開することは想定していない．そのため，本研究で求められる匿名化レベルの処理が行える手法は存在しない．

3.2 物体除去

画像や動画から物体を除去する方法に関する研究は，画像処理研究におけるタスクの 1 つとして取り組まれている．動画を対象とした自動の物体除去手法には，前景と背景を分離する手法が存在する [10], [25]．しかし，これらの手法の多くは固定カメラからの映像などの静的な背景の動画を対象としており，本研究で扱う背景が大きく変化する動画においては，適用が難しい．

画像を対象とした自動物体除去では，Shetty らが Generave Adversarial Networks（GAN）[9] を用いた手法を提案している [22]．Shetty らの手法では，画像とその画像に写っている物体のクラスラベルを 1 つ入力することによって，そのクラスに属する物体を画像から自動で除去することができる．このモデルは，クラスラベルをもとにマスク画像を自動で予測，作成する Mask Generator と，作成されたマスク画像を利用して，マスク部分を再構成することで物体除去を行う Image Inpainter から構成されている．そして，Mask Generator によって作成されたマスクを Mask Discriminator によって評価し，Image Inpainter によって生成された画像は画像中に物体が写っているか判断する Object classifier と，画像の質を評価する Real/Fake classifier によって評価される．これら 5 つのネットワークを GAN で学習することによって，画像とクラスラベルを入力として自動で物体除去が行えるモデルが学習されている．一方で，この手法はモデルの学習自体の難しさから，現状では検出に失敗する例も多い．本研究では匿名化が主な目的であるため，プライバシに関する物体を高い精度で検出することが求められる．

そこで，本研究では深層学習を用いた高精度の物体検出手法と背景を再構成する手法を統合することで，物体を自動で除去する手法を検討する．動画を対象とした再構成手法では，フレームごとにその前後の数フレームからマスクに該当する箇所の背景を取得し変換し貼り付けることで，再構成を行う [6], [14], [18]．そのため，前景と背景を分離する手法とは異なり，動的な背景の動画に対しても適用が可能である．しかし，これらの手法では再構成対象の背景を他のフレームから取得可能であるという前提に基づいている．そのため，動画中で停止しているような物体を除去することはできず，都市の動画において停止している車や人を除去することができない．

画像を対象とした再構成手法は，基本的に 3 つの種類に分類することができる．1 つ目は，古典的な手法で，マスク周辺のピクセルから計算する [1], [3]．この手法では，周辺のピクセルになじませることによって，画像の修復を行う．そのため，大きいマスクの再構成や顔などの複雑な物体の修復は難しい．

2 つ目は，パッチベースの手法で画像内もしくは，あらかじめ用意されている画像データベースからマスク周辺のピクセルと類似性が高いパッチを検索，取得し，マスク部分に貼り付ける手法である [2], [11]．画像内から検索する代表的な手法として，Barnes らが提案した PatchMatch がある [2]．PatchMatch では，マスクに該当する箇所の類似パッチを検索する際，確率的な挙動を用いることで，計算時間を大幅に削減しつつ，高精度な再構成を行うことを可能にした．また，再構成だけでなく多様な画像編集に応用が可能である．PatchMatch などの手法では，マスク部分として考えられるパッチが画像内もしくは，画像データベースに存在することが前提であるため，画像の性質やマスクの位置に再構成結果の精度が大きく依存する．たとえば，風景画などの単調な模様が続いている画像においては，類似パッチを検索し再構成することは容易である．一方で，複雑かつ画像内や画像データベースにおいて，出現確率の低いパッチが必要となる場合，自然な再構成は難しい．

3 つ目は，深層学習と GAN の学習手法で画像の再構成方法をモデルに学習させるラーニングベースの手法である [12], [27]．ラーニングベースの手法では，パッチベースとそれ以前の手法でおきていた問題を解決することができる．ラーニングベースの手法では，膨大な種類の大量の画像データセットかつ，様々なマスクでモデルを学習させることにより，モデルは画像全体のピクセル情報を使用して，任意のマスクを再構成することが可能となる．そのため，マスクの大きさや画像の種類，画像中におけるマスクの位置などの違いによる影響が少なく，自然な再構成ができる．そこで本研究では深層学習を用いた高精度の物体検出手法と任意のマスクに対して自然な再構成が可能なラーニングベースの画像再構成手法を用いて，画像から自動で物体除去による匿名化を行う手法を提案する．

4. GANonymizer

本研究では入力された画像から人や車などのプライバシに関する物体を自動で検出し，画像上から消去する GANonymizer を提案する．GANonymizer はプライバシに関する物体の検出とその除去を行う 2 つのニューラルネットワークから構成される．また，物体除去の際に検出箇所の背景再構成を補助する ESP と GFP を提案する．本章では，まず GANonymizer の全体像として使用した 2 つのニューラルネットワークについて説明し，その後再構成を補助するパディング処理の ESP と GFP について説明する．

4.1 ネットワーク構造

GANonymizer の全体の構成を**図 1** に示す．GANonymizer はプライバシに関する物体を検出するフェーズと検出箇所の背景を再構成するフェーズの 2 つから構成される．検出のフェーズでは，画像のプライバシ侵害の危険性を最小化するために，物体のスケールや向きによらず，多くの物体を検出することが求められる．そこで，検出には高精度な物体検出ネットワークである Single Shot Multibox Detecter（SSD）[15] の入力画像を 512 px に圧縮するモデルを使用した．SSD は入力画像から複数のスケールの特徴マップを作成し，それぞれの特徴マップから検出を行うため，同じ種類の物体でも異なるスケールであらわれる可能性のある移動カメラからの画像などに有効な手法である．また，本研究の匿名化対象である人や車

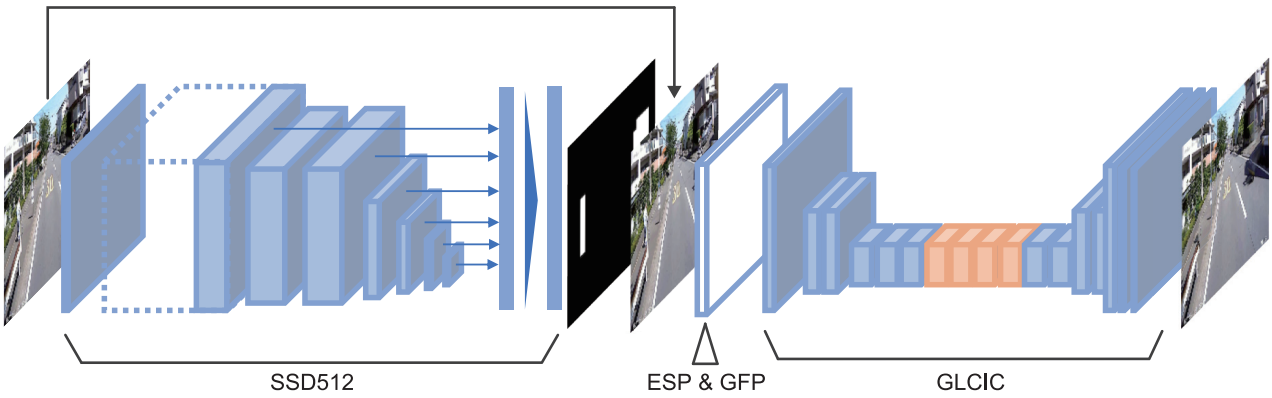

図 1 GANonymizer のネットワーク構造の全体図．SSD512 [15] ではプライバシに関する物体を検出し，物体の箇所を表すマスク画像を作成する．その後，端にあるマスクの再構成を補助する ESP と大きなマスクの再構成を補助する GFP によって，画像とマスクが変換される．最後に，GLCIC [12] を使用して検出した物体の背景を推定し再構成する

Fig. 1 Overview of the architecture of GANonymizer. SSD512 [15] detect objects related to privacy, and we then create the mask image which represents the objects' position. After that, ESP and GFP process the input image and the mask image, which are the layers that help the reconstruction of edge parts in the mask and large parts in the mask respectively. Finally, GLCIC [12] reconstructs a background of the objects.

など都市映像に含まれるプライバシに関する物体は，一般物体検出タスクにおける既存の画像データセットにラベルとして存在する．そのため，これらの物体が含まれている Pascal VOC dataset [7] で訓練された SSD のモデルをプライバシに関する物体の検出器として使用した．

次に，GANonymizer は検出された物体の背景を推定し検出箇所に合成することで再構成を行う．まず，画像中のプライバシに関する物体として検出された箇所にマスクをかけ，マスク部分を再構成する画像再構成手法を適用することで，背景の再構成を行った．画像再構成における代表的な手法には PriSurv [5] や PatchMatch [2] などがある．しかし，これらの手法は再構成された結果がぼやけていたり，周りの背景情報との一貫性が低い結果が得られることが多い．一方で，Deep Neural Network（DNN）を使用した画像再構成手法は自然な背景の推定に成功している．そこで，本手法では，DNN を使用して自然な背景生成が可能なモデルの 1 つである Globally and Locally Consistent Image Completion（GLCIC）[12] を使用した．

GLCIC は，マスク部分の再構成を行う Completion Network と再構成された箇所の評価を行う Global Discriminator，Local Discriminator の 2 つの Discriminator によって GAN で学習する．Completion Network は入力画像と対応するマスク画像を入力とし，入力画像のマスクに該当する箇所を周辺ピクセルをもとに推定し再構成する．再構成された画像は，2 つの Discriminator によって入力画像と比較され，もっともらしい再構成結果が得られているか評価される．その際，GLCIC では Global Discriminator は画像全体を評価し，Local Discriminator は再構成された箇所のみを評価することで，画像全体として一貫性を担保した再構成結果が得られているかを評価することができる．最終的に，2 つの Discriminator が元々の入力画像と再構成された画像を識別できなくなったとき，学習は終了し Completion Network は自然な画像再構成が可能なネットワークとなる．本研究における画像再構成は，都市画像に存在する除去対象物体の背景を再構成することが目的となる．そこで，世界中の多様な場所や環境における大量の画像を含む Places2 Dataset [28] で学習されたモデルを使用する．

4.2 ESP と GFP

GLCIC における背景生成を補助するためのレイヤとして，ESP と GFP を提案する．GLCIC では，任意のサイズ・位置のマスクに対して自然な再構成が行えるわけではなく，自然な背景再構成が困難なマスクが 2 つある．1 つ目は，再構成対象となるマスクが画像の端に位置している場合で，画像の端に隣接する側のマスクの周辺の画像情報が不足するために，再構成が困難になる．たとえば，マスクが画像の右下の角に隣接している場合，マスクはマスクの右と下から画像情報を得られず，マスクの右下における再構成が不自然になる．2 つ目は，マスクが画像の端にある場合である．GLCIC は Convolutional Neural Network（CNN）でマスクの周りの情報をマスク内に伝達することによって，再構成を行っている．そのため，マスクが大きい場合，マスクの外側の情報をマスクの中心部まで伝達できず，画像の中心部の再構成結果が不自然になる．そこで，それら再構成が困難な場合に GLCIC の再構成を補助

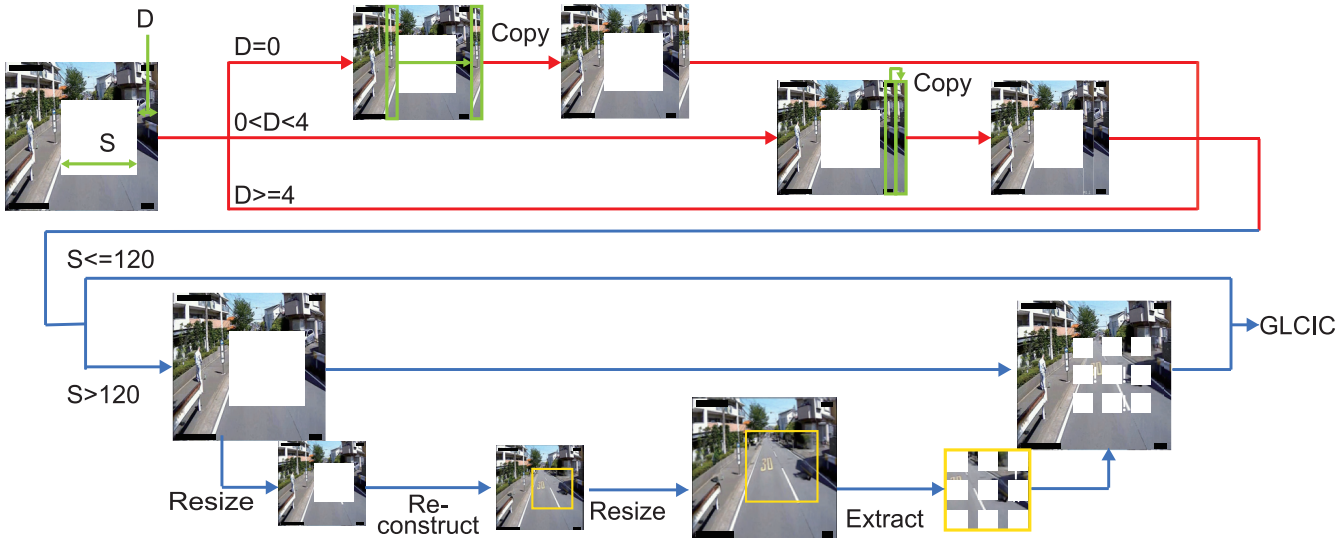

図 2 ESP と GFP の全体図．ESP は画像の端に位置するマスクの再構成を補助し（赤のラインフロー），GFP は大きなマスクの再構成を補助する（青のラインフロー）．図中の D は，マスクと画像の端の距離を表しており，S はマスクのサイズである．ESP では，画像中のピクセルを用いてパディングを行うことで，マスクの外側からの情報を増やす．GFP は小さくリサイズした入力画像の再構成結果を使用して，大きなマスクを擬似的に小さな複数のマスクに分割する

Fig. 2 Overview of ESP and GFP. ESP (red-line flow) helps GLCIC to reconstruct edge parts in the mask. GFP (blue-line flow) helps large parts in the mask. D is the distance between a mask and the edge of the image. S is the mask size. ESP pad the pixels in the image to the outside of the image to increase information from outside of the image. GFP divides a large mask into small masks using the reconstruction results of the small-resized input image.

する Edge Shift Padding Layer（ESP）と Global Feature Padding Layer（GFP）を提案する．

4.2.1 Edge Shift Padding Layer（ESP）

ESP はマスクが画像の端にある場合に GLCIC の背景再構成を補助する（**図 2** の赤いフロー）．ESP は画像内のピクセル列をマスク外からの情報が少ない側の画像の外側にコピーすることで，マスクの端に伝わる情報を擬似的に増加させる．外側にコピーするのに使用されるピクセル列は，マスクの端と画像の端の距離によって異なる．マスクと画像の端の距離が 1 px 以上の場合，画像の最も外側のピクセル列が使用される．一方，マスクと画像の端の距離が 0 px の場合，すなわち画像の端とマスクの端が隣接している場合，マスクの端側にあたる辺と反対側の辺の隣のピクセル列が使用される．また，マスクと画像の端の距離が十分に大きい場合，ESP は適用されない．画像内のピクセルを利用して画像を一時的に拡張することで，マスクが端にある場合でもマスクが全方向から画像情報を取得でき，GLCIC による再構成結果が改善される．

4.2.2 Global Feature Padding Layer（GFP）

GFP はマスクのサイズが大きい場合に GLCIC の背景再構成を補助するレイヤである（図 2 の青いフロー）．GFP は，大きなマスクを持つ入力画像を小さくリサイズした画像に対しての再構成結果を利用して，擬似的に大きなマスクを小さく分割する．GFP の処理は次の 3 つの手順で行われる．まず，GLCIC のみで自然な再構成が可能な画像サイズに入力画像をリサイズし，リサイズされた画像に対して GLCIC を適用し，再構成結果を得る．その後，リサイズされた画像での再構成結果を元のサイズに拡大し，格子状に抽出する．最後に，抽出した格子状の再構成結果を元の入力画像のマスク部分に合成する．これにより，入力画像の大きなマスクを擬似的に複数の小さなマスクに分割し，大きなマスクに対しても自然な背景再構成を実現する．

5. パラメータ設定のための事前実験

本章では ESP と GFP に関するパラメータの決定の際に行った実験とその結果について述べる．実験では，その後の GANonymizer 全体の精度検証で使用する都市画像の一部と，再構成結果をはっきり表示するための単純な白色画像を使用して，パラメータの変化にともなう再構成結果の変化を検証した．

5.1 ESP

ESP はマスクが画像の端にあるために，画像の端に接しているマスク辺からの情報が減少し，結果的に自然な再

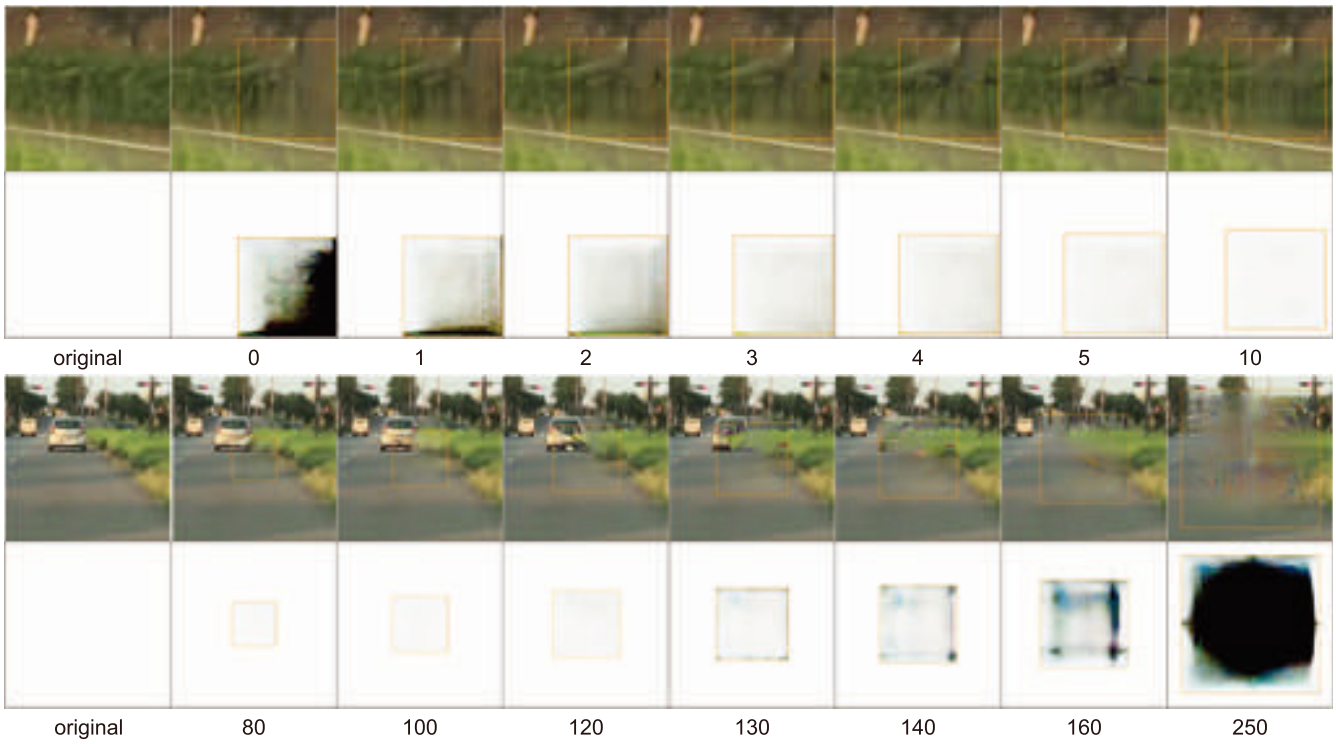

図 3 マスクと画像の端の距離（上段），マスクサイズ（下段）のそれぞれと GLCIC の再構成精度の関係についての検証結果．各検証結果について，上段は都市画像に対して，下段は白色画像に対して，オレンジで示した枠に GLCIC を適用した結果ある．下段の白色画像に対する結果では可視化のために，実際の再構成結果の画素値を 3 倍したものを表示している

Fig. 3 The influence of the distance between the mask and the image (top row) and the mask size (bottom row) for the GLCIC's reconstruction result. For each result, the top row is the reconstruction result for an urban image and the bottom row is the reconstruction result for a simple white image. Note that the pixel values of the reconstruction result for the white image is multiplied by a factor of 3 for visualization.

構成が困難になるという問題を緩和するためのパディング層である．ESP では画像の外にパディングを行うことで，マスクの端の辺からの情報を増やす．理想的なパディングは，パディングするピクセル列に入力される画像外のシーンを利用して，画像として切り取られるシーンを延長することであるが，これは非現実的な操作である．そのため，パディングするピクセル列と入力画像との一貫性を完璧に保つことはできない．したがって，パディングするピクセル列によって，マスクに伝達されるピクセルの情報量が増える一方で，パディングするピクセル列の影響が強すぎる場合，パディングしたピクセル列と実際の画像との一貫性が低いことにより，再構成精度を低下させるような影響を及ぼしかねない．そこで，マスクを自然に再構成するのに十分かつ，マスクの再構成に悪影響を及ぼさない，適切なピクセルと幅によるパディングが求められる．また，可能な限り ESP を適用せずに再構成するため，マスクの隣に幅何ピクセルの情報が必要であるかを調査し，ESP を適用する閾値に関しても検証を行った．

以上をふまえ，ESP の挙動を決定するためのパラメータを以下の 3 つで定義した．

(1) ESP を適用するマスクと画像の端の距離の閾値
(2) パディングの幅
(3) パディングに利用するピクセル列

上記のパラメータに関して最適値を探索するため，以下の 2 つの実験を行った．まず，(1) を決定する実験では，両辺 120 px の正方形のマスクで，マスクと画像の端の距離を 1 px から 10 px まで変化させながら，GLCIC による再構成を行った．実験における代表的な結果を**図 3** 上段に示す．図 3 上段は，マスクと画像の端の距離が 0 px から 5 px までと，10 px のマスクに対する再構成結果で，各画像において再構成された箇所はオレンジの枠で示している．

下段の白色画像において，マスクと画像の端との距離が 0 px の場合では，マスクの右下までマスク周辺の画素値が適切に伝播されず，本来は白色の画素値で再構成されるべき箇所に，黒色の画素値が多く含まれている．この黒色の部分は，GLCIC のネットワークの初期値に依存するもの

である．GLCIC のネットワークを構成する CNN で，情報をマスクの右下まで伝播しきれなかったことにより，ネットワークの初期値が再構成結果に大きく影響を与えている．同様の現象がマスクと画像の端の距離が 2 px 以下の結果においても生じており，3 px の場合においても，マスク下方において同様の現象が起きていることが確認できる．一方，マスクと画像の端の距離が 4 px 以上の再構成結果では，マスクの下端や右端においても白色の画素値で再構成されている．

上段の画像は，都市画像の左端に配置したマスクに対する再構成結果を示している．ここで，最左端に示す入力画像では再構成されるオレンジの枠の箇所の大部分は植樹帯であることが分かる．最右端におけるマスクと画像の端の距離が 10 px ある場合の再構成結果では，マスクの右端に植樹帯と同じ緑色の画素値が生成されている．一方，上段左から 2 番目のマスクと画像の端の距離が 0 px のマスクの再構成結果では，マスクの右端に情報が伝播されていないために，植樹帯の緑色の画素値ではなく，植樹帯の上部の壁の影響を強く受けた再構成結果になっている．これは，マスクの左側にある植樹帯の情報がマスクの右端まで伝播されないことにより，マスク右端では，マスクに対して垂直方向の情報のみを頼りに再構成されており，マスクに対して水平方向の情報が考慮されていないことに起因している．同様の現象がマスクと画像の端の距離が 1 px から 3 px の場合でも起こっていることが分かる．一方，マスクと画像の端の距離が 4 px 以上のマスクになると，再構成結果の右側が全体的に植樹帯の緑色の画素値に近づいており，マスクに対して水平方向の情報を考慮することができている．

以上 2 つの検証結果から，ESP を適用する閾値には 4 px を採用し，マスクと画像の端の距離が 4 px 未満の場合には，ESP を適用してマスクの端側からの情報を増やすようにした．同時に，この結果からマスクを自然に再構成する際に十分なピクセル幅が 4 px であることが分かった．したがって，画像として切り取られるシーンを延長したわけではないパディングによって，4 px 以上増やすことは，かえって再構成結果に悪影響を与える可能性がある．よって，2 のパディング幅もマスクと画像の端の距離が 4 px になるように，パディング幅を調整することにした．

次に，3 つ目パラメータである画像の外側にコピーする際のピクセル列を決定するための実験を行った．実験では，ESP を適用して再構成結果を比較する対象として，乱数で構成されたピクセル列（Random），画像中からランダムに抽出したピクセル列（Random Pick），画像の最も外側にあるピクセル列（Edge），画像の端から遠い方のマスク辺の隣のピクセル列（Opposite）を用いた（**図 4**）．パラメータ (1) と (2) に関する実験の結果から，マスクの端側にパディングすることで，マスクに対して水平方向の情報を再構成結果に反映させることが求められる．そのため，Random ではマスクの右端に植樹帯の緑色の画素値が生成されていないことから，パディングに使用するピクセル列として適切でないことが分かる．一方，Random 以外の Edge，Opposite，Random Pick の再構成結果では，マスクの右端部分において，植樹帯の緑色の画素値が生成されていることから，パディングによって水平方向の情報をマスクの再構成に反映させることができている．これら 3 つの中でも Edge では最右端のピクセル列をコピーしているため，画像として切り取るシーンを拡張することに最も近いピクセル列によるパディングであると考えられる．一方の Opposite では，マスクを挟んで端側とは反対側のピクセル列を使用してパディングを行うため，マスクが大きくマスクの右側と左側では，写っているシーンが大きく異なるような場合，画像として切り取るシーンを延長したピクセル列とは大きく異なる可能性がある．Random Pick では，ランダムでピクセル列を選択するため，Opposite 以上にマスクの右端から離れた位置のピクセル列が選択される．以上より，マスクと画像の端との距離が 1 px 以上ある場合には，Edge によるパディングを採用し，0 px の場合 Edge は存在しないため，Opposite によるパディングを採用した．

図 4 ESP におけるパディングに使用するピクセル列を変化させて，再構成を行った結果．左から，乱数で構成されたピクセル列（Random），画像中からランダムに抽出したピクセル列（Random Pick），画像の最も外側にあるピクセル列（Edge），画像の端から遠い方のマスク辺の隣のピクセル列（Opposite）を使用した場合の再構成結果である

Fig. 4 The influence of the pixel using padding in ESP for GLCIC's reconstruction result. From left to right, the reconstruction result using the most outside pixel in the image (Edge), the pixel next to the mask's edge of the opposite side of the image edge (Opposite), the random pixel (Random), and the pixel extracted from the image randomly (Random Pick).

5.2 GFP

GFP では，挙動を決定するパラメータとして以下の 3 つを定義した．

(1) GFP を適用するマスクサイズの閾値
(2) パディングによる分割の粒度
(3) パディングする際の格子の幅

GFP では小さくリサイズされた入力画像の再構成結果から格子状に抽出したものを入力画像に合成することで，

マスクを擬似的に分割する．そのため，分割が不要な小さいマスクに関してはGLCICで直接再構成する方が自然な結果を得ることができる．そこで，はじめにGLCICで自然な背景再構成が直接的に行えるマスクサイズの限界について調査し，GFPを適用する閾値を探索する．また，GFPでは分割に用いる格子の情報が再構成結果に過度な影響を与えると，低解像度のぼやけた再構成結果になる．一方，格子の情報が弱すぎる場合，分割しても再構成結果が改善されないため，適切な格子の情報の強度を設定する必要がある．この格子の情報の強度は，マスクの分割数と格子の幅として定義できる．そこで，これら2つに関して複数の値で実験を行い最も効果が高いものを採用する．

まず，(1)の閾値を設定するにあたって，マスクを画像の中央に配置し，マスクサイズを50 pxから200 pxまで10 px間隔で変化させたマスクに対してGLCICを直接適用した．代表的な結果を図3下段にオレンジの枠で示した．白色画像の再構成結果において，マスクサイズが80 pxから120 pxまでは再構成されたマスク部分全体が白くなっているため，マスクの全方向から適切に情報が伝播されていることが分かる．一方，マスクサイズが130 pxのときは，マスクの端が黒くなっていることが確認できる．これは，マスクの端においては反対側からの情報が伝達されにくいために，全方向からのマスク周辺の情報を統合し，再構成結果に適切に反映できないことに起因する．同様の傾向が140 px以上のマスクサイズにおいても見られる．また，各辺から130 pxの距離より遠い領域では，最も近い辺の方向から伝播される情報のみを反映して再構成されているため，白色の画素値で再構成できている．白色画像における再構成結果の最右端の画像では，中央付近において，すべての方向における情報を再構成結果に反映させることができないために，黒色の画素値が生成されている．以上より，120 px以下のサイズのマスクに対しては，GLCICは全方向からの情報を反映した再構成を行うことができるが，130 px以上のサイズのマスクに対しては難しいことが分かる．

上段の都市画像に対する再構成結果でも同様の現象を確認することができる．120 px以下のマスクサイズの再構成結果では，車の一部を生成しているために，写真としては不自然に見える箇所があるものの，マスク周辺の画素とは馴染んだ再構成結果となっている．一方で，130 pxのマスクサイズの再構成結果では，マスクの右端部分に赤と黒の横線が生成されているように，明らかにマスク周辺の情報とは異なる画素値が生成されている箇所がある．この現象は，140 pxのマスクサイズの再構成結果においても，特に中央と右端部分に確認することができ，マスクサイズが160 px以上の再構成結果では特にマスクの中央部において，マスクの全辺からの情報が伝達されないために，GLCICのネットワークの初期値に依存した画素値が再構成結果に

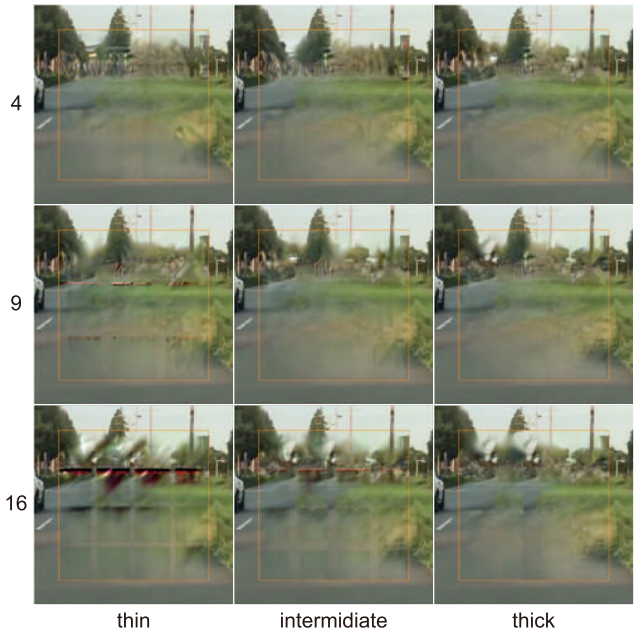

図 5 GFPの分割粒度と分割に用いる格子の幅を変化させて，再構成を行った結果．画像の左にマスクの分割数を，下に分割に用いた格子の幅を示す

Fig. 5 The influence of the division-granularity and the lattice width used in a mask-division for GLCIC's reconstruction result. The left numbers are the number of mask-division. The lattice width is shown at the bottom of the image.

表れている．以上より，GFPはマスクサイズが120 pxより大きい場合に適用するように閾値を設定した．

次に，2つ目の分割の粒度と3つ目の格子の幅を決定するため，4分割，9分割，16分割の3種類の分割方法を適用し，それぞれの分割方法で最適な格子の幅を探索した．各分割方法と格子の幅の組合せにおける検証結果を**図 5**に示す．格子の幅は分割数に合わせて適応的に変化させ，検証を行った．4分割の場合，格子の幅のthin，intermediate，thickはそれぞれマスクサイズの$\frac{1}{8}$，$\frac{1}{4}$，$\frac{1}{2}$とし，9分割の場合はそれぞれ$\frac{1}{16}$，$\frac{1}{8}$，$\frac{1}{4}$とし，16分割ではそれぞれ$\frac{1}{32}$，$\frac{1}{16}$，$\frac{1}{8}$として，再構成結果を比較した．図5において，分割数によらず，thinの場合は格子の情報の強度が低いため，分割されたマスクが再構成される際にマスクの周辺の情報の影響を強く受けて生成されている．そのため，分割後のマスクの再構成結果の解像度と格子の解像度の差が大きくなり，結果として再構成結果において格子部分が目立ってしまっている．16分割では，intermediateでもthinと同じように格子部分が目立っていることが分かる．また，thickでは格子の情報の強度が強く，再構成に与える影響が大きすぎるために，再構成結果の解像度が全体的に低い．一方で，4分割と9分割のintermediateでは，マスクを分割するための格子とマスク周辺の情報の解像度の両方に馴染むような画像を生成することができている．この結果と，より大きなマスクサイズを分割することも想定し，9分割で

格子の幅が $\frac{1}{8}$（intermediate）の分割方法を採用した．

6. 実験

本章ではまず，ESP と GFP の有効性の検証のために行った定性評価と定量評価について述べる．次に，GANonymizer 全体の有効性を検証するため，実際の道路画像に適用しその結果について考察を行った．本実験のユーザテストにおける実験参加者は，筆者が所属する大学において，筆者が授業補助を行っているクラスの学生（10 代：機械学習や画像解析に関する知識をほとんど有さない）5 人と，同研究室の学生（20 代：機械学習や画像解析に関する知識をある程度有する）5 人である．

6.1 使用する都市画像

実験では，神奈川県藤沢市を走行する車から iPhone7 で撮影した画像を使用した．撮影時間帯は，実際の都市インフラ監視の運用を想定し，道路の詳細が鮮明に写っており都市のインフラ監視が可能な日中と夕方に限定した．使用した画像の総数は 5,246 枚で，1 枚の画像サイズは 1,080 × 1,920 である．また，撮影した都市画像の特徴から，検出対象のプライバシに関する物体は，人，車，バス，自転車，バイクとし，それらの物体を画像から除去する実験を行った．

6.2 物体検出の精度評価

本実験で使用する画像と，プライバシに関する物体を検出する SSD の学習に使用されている Pascal VOC データセットでは，画像のドメインが大きく異なるため，本実験で使用する都市画像に対しての検出精度を測定した．実験で使用する画像の中から，50 件の画像を無作為に抽出しそれらの画像におけるプライバシに関する物体の検出成功割合を入力画像と GANonymizer を適用した後の画像を比較して，測定した．実験の結果，すべてのプライバシに関する物体のうち，81.9%を検出することに成功した．検出に失敗した物体には，物陰に物体の大部分が隠れている物体や，小さく写る歩行者などが多く見られた．

6.3 ESP と GFP の評価

6.3.1 定性評価

ESP と GFP の有効性を検証するために，ESP と GFP を適用した場合と適用しない場合の GLCIC の再構成結果の画像を比較した．実際に比較した図の例を**図 6** に示す．ここでは，画像サイズが大きいため，プライバシに関する物体が検出され作成されるマスクは，すべて 120 px 以上になっている．ESP と GFP を適用していない左側の再構成結果では，除去されたすべての箇所がマスクが大きかったり，画像の端に位置したりしており，マスク全体に周辺の画素情報を伝播しきれていないことが分かる．一方，右側の ESP と GFP を適用した場合の再構成結果では，ESP と GFP を適用することで自然な再構成を実現している．

図 6 ESP と GFP を適用しない場合の再構成結果（左）と適用した場合の再構成結果（右）

Fig. 6 Effect of ESP and GFP. The left images are the reconstruction results without ESP and GFP, while the right ones are that with ESP and GFP.

6.3.2 定量評価

6.3.2.1 マスク作成方法

ESP と GFP を適用した場合と適用しない場合，それぞれの GANonymizer による出力画像と入力画像における PSNR と SSIM を比較し，定量的な評価を行った．PSNR と SSIM は，一般的に画像の質を評価する際に使用されることが多い [24], [27]．PSNR と SSIM は 2 つの画像の輝度の差分をもとに計算されており，PSNR ではピクセル単位で計算が行われ，SSIM では周囲のピクセルとの相関も考慮して計算されており，どちらの指標も値が高い方が精度が高いことを意味している．実験では，画像中においてプライバシに関する物体の箇所ではなく，背景となっている箇所にあえてマスクを作成し，入力画像を PSNR と SSIM の基準画像として使用した．マスクの作成は，マスクを決定するための複数のパラメータを用意し，それらをランダムサンプリングによって取得することで，ランダムなマスク生成を行った．

また，ESP と GFP の評価には，それぞれ必要なマスクの位置やサイズが異なる．そのため，まず ESP の評価の際は次の 3 つのパラメータを定め，各パラメータに対してランダムサンプリングを実行し，マスクを作成した．1 つ目は，マスクの配置場所に関するパラメータで，左上，中央上，右上，中央右，右下，中央下，左下，中央左の 8 つを用意した．2 つ目は，マスクのサイズで 50 px から 200 px と定めた．3 つ目は，マスクと画像の端の距離で，0 px から 3 px までとした．GFP の評価の際のマスクは，マスクの配置場所を画像中からランダムサンプリングし，サイズ

は 120 px から 400 px の中からランダムサンプリングすることで，2 つのパラメータを用意した．そして，これらのパラメータによって作成されたマスクが，プライバシに関する物体と重なった場合は，ランダムサンプリングのやり直しを繰り返すことで，評価用のマスクを作成した．

6.3.2.2 結果

上記の方法によって行った実験で得られた PSNR と SSIM の値を**表 1** に示す．GFP は適用することで，PSNR，SSIM どちらの指標も上がっていることから，GFP により再構成の精度が改善されていることが分かる．しかし，ESP を適用した場合，適用しなかった場合に比べて大幅に精度が低下している．したがって，PSNR と SSIM による定量的な評価の結果，GFP は GLCIC による再構成の結果を改善できるが，ESP は GLCIC による再構成の精度を低下させることが分かった．

表 1 PSNR と SSIM による ESP と GFP の評価
Table 1 The quantitative comparison for the effect of ESP and GFP.

	ESP なし	ESP あり	GFP なし	GFP あり
PSNR	**35.71**	26.26	31.16	**32.10**
SSIM	**0.966**	0.765	0.954	**0.956**

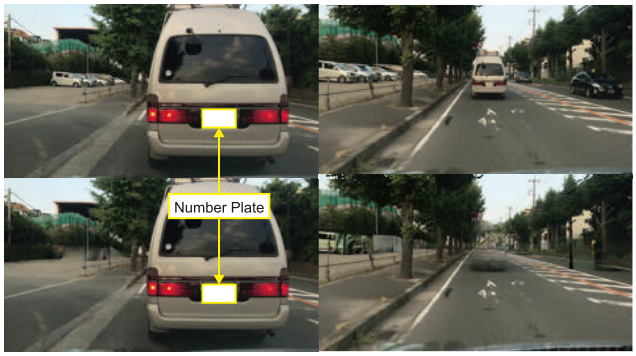

図 7 プライバシに関する物体を自然に除去することに失敗した例．左の画像は物体がカメラと近すぎるために物体が大きく写り込み，検出に失敗している例である．右側はマスク部分に中央の車の影が含まれていないことによって，再構成結果が影の影響を受けてしまっている

Fig. 7 The examples that GANonymizer fails to remove objects related to privacy naturally. In the left result, GANonymizer fails to detect the objects because of too close and too large. In the right result,

6.3.3 ユーザ評価

ESP と GFP を適用し匿名化された画像と，適用せずに匿名化した画像のどちらがより自然かを，実験参加者に判定してもらう評価実験を実施した．実験参加者数は，1 人につき 6 枚の画像ペアを比較してもらった．また，テスト時には画像は撮影した都市画像のデータセットからランダムで選択され，実験参加者はその画像を GFP と ESP ありで除去した結果と GFP と ESP なしで除去した結果を比較し，どちらが自然であるかを判定する．ユーザ評価で使用した画像ペアは定性評価に用いたものと同じものを使用した（図 6）．全 10 人 × 6 回の評価のうち，54 枚で提案手法の GFP と ESP を適用した結果の方が，より自然であるという回答が得られた．

6.4 GANonymizer 全体の評価

6.4.1 定性評価

図 7 と**図 8** に GANonymizer を移動カメラからの都市画像に適用した結果を示す．図 8 では，画像中から車や歩行者などのプライバシに関する物体を自然に除去できていることが分かる．特に下段の GANonymizer の出力画像を見ただけでは，どこが再構成された箇所かを判断することが困難な結果もある．一方で，図 7 の左の画像に示すように，物体との距離が近く，画像中で大きく写りすぎているものや，遠くに写る小さい車や人は検出することができないケースもあった．また，図 8 の右図の中央の車の除去結

図 8 GANonymizer を都市画像に適用した結果．左の 2 つの画像は正午に撮影された画像で，右の 2 つの画像は日没に撮影された画像である．それぞれ上段が入力画像で下段が入力に対する GANonymizer の出力画像である

Fig. 8 The result of GANonymizer for urban images. The left two-column images are taken in the daytime, while the right two-column images are taken in the evening.

果には，車の影が残っているために不自然になっていることが分かる．

6.4.2 ユーザ評価

GANonymizer の出力画像を人が監視することを想定し，GANonymizer の出力画像からプライバシに関する物体が除去されたと思われる箇所をマーキングするユーザ評価を実施した．実験は，実験参加者に事前に何も伝えずにマーキングしてもらう事前情報なしのテストと，実験参加者に事前に GANonymizer によって除去されている物体の種類と個数を伝えてからマーキングしてもらう事前情報ありのテストの 2 つを実施した．また，テストでは実験に使用したすべての都市画像の中から，事前情報ありとなし，それぞれのテストごとにランダムに 10 件選択して行った．

マーキングテストによって得られたマーキング画像を**図 9** に示す．図 9 では，左側の画像が入力画像で，右側の画像が入力画像に対して GANonymizer を適用して得られた結果の画像に実験参加者が黒色のペンでマーキングを行った結果である．実験参加者に，事前情報あり 10 件，事前情報なし 10 件をそれぞれを行った結果，事前情報なしでは 36.0%，事前情報ありでは 44.9%であった．結果として，GANonymizer は画像中に写るプライバシに関する物体の 6 割以上を，人間に認識不可能な精度で除去できた．また，事前に写っている物体の情報を伝えることで，マーキングが容易になることも分かった．

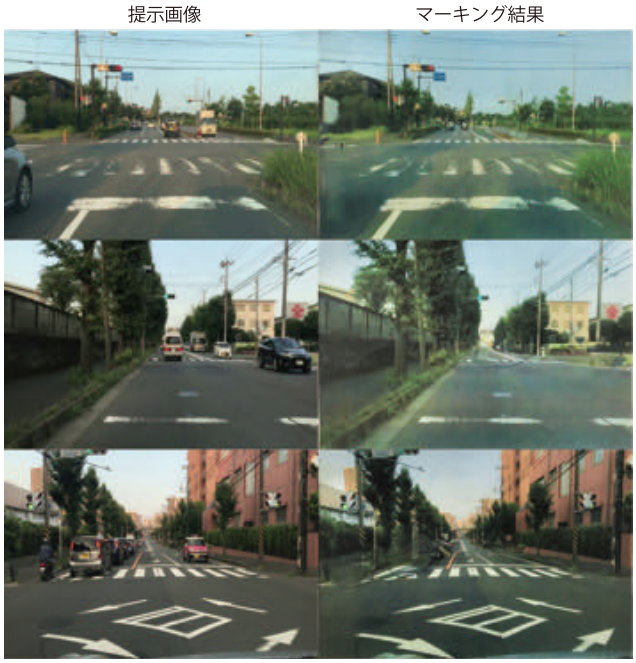

図 9 ユーザ評価であるマーキングテストの結果．左がプライバシに関する物体を含む入力画像で，右が左の画像に対して GANonymizer でプライバシに関する物体を除去した結果に対しての実験参加者によるマーキング結果である

Fig. 9 The result of the user marking test. The left column images are the input image including objects related to privacy. The right ones are the output of GANonymizer with user-marking.

6.5 考察

提案した GFP に関しては，定量的評価と定性的な評価の両面で有効性を確認することができた．一方で，ESP の定量評価では ESP の適用によって，再構成の精度を下げていることが確認された．この違いは PSNR と SSIM の計算方法によるところが大きいと考えられる．ESP を適用しない場合，図 6 に示すように，マスクの端部分における情報が少ないために，再構成結果は砂嵐のような単調な模様になっていることが分かる．一方，ESP を適用した場合，元々のマスク周辺における画素情報と ESP でパディングされた画素情報を利用して，それらに馴染むような背景を再構成する．そのため，砂嵐のような単調な画像ではなく，道路や建物などの構造的情報を持つ画像が生成される．しかし，ESP によってパディングされるピクセル列は，入力画像との一貫性が完璧にとれていないため，それらを利用して再構成される画像も元画像の背景と完全に一致させることはできない．PSNR と SSIM の計算方法の性質上，砂嵐のような平均的なピクセル値の画像の方が元画像の背景とのピクセル単位での差分は小さくなり，精度が高く出ることが考えられる．また PSNR，SSIM を含め，画像生成や画像再構成のタスクでは，人間が自然であると感じたり，綺麗であると感じたりする感覚と一致する定量評価指標は存在しないといわれている [27]．これらのことから，最終的に画像から人間がプライバシ物体の存在を特定できないように除去して匿名化するという観点においては，ESP と GFP による再構成結果の改善は良い方向に働いていると考えられる．

GANonymizer 全体の評価においては，定性評価に近い結果がユーザテストでも得られた．物体を除去する手法を定量的に評価する場合，同じ背景シーンにおいて除去対象が写っている画像と写っていない画像が必要となる．しかし，これを作成することは難しく，近年の研究でもシュミレータ環境を利用して人工的に作り出すことにとどまっている．そこで本研究では匿名化のための物体除去手法としての精度を客観的に測るために，出力画像から除去された箇所を特定するタスクを解かせるユーザテストを実施した．結果として，多くのプライバシに関する物体を高い確率で完全に除去できた．また，事前に除去された物体の種類と数を知らせることで検出が容易になったことから，事前情報がない場合では除去された箇所か否かの判別が困難で，実験参加者にとって不確実であった箇所が，数や種類の情報によって不確実性が排除されマーキングに至ったと考えられる．そのため，実験参加者がランダムにマーキングをしている可能性は低いと考えられる．

これらの結果は，GANonymizer を構成する 2 つのニューラルネットワークの効果が大きい．検出のフェーズにおい

ては，SSD の高精度モデルを使用したことで，実験で使用したほとんどの都市画像で様々なスケールの物体を検出することに成功した．一方，物体との距離が極端に近い場合や遠い場合には検出ができないケースがあった（図 7）．物体との距離が遠いために検出できない場合は，物体が鮮明に写っていないためプライバシ侵害の危険性は低い．一方，カメラに近い物体は鮮明に写っているため，検出できなかった場合，プライバシ侵害の危険性が高い．たとえば，図 7 の左側のようにナンバープレートが無加工のままになってしまうなどの危険性がある．

本研究では，GANonymizer によって除去された部分が除去結果から特定可能なケースを想定したときに，物体の Semantic Segmenation などの出力結果では，物体の形状が分かってしまうためプライバシ侵害のリスクが高いと考え，物体の位置と大きさがバウンディングボックスとして出力される物体検出手法をプライバシに関する物体の検出に使用した．一方で，Semantic Segmentation はピクセル単位でのクラス分類するため，物体が部分的に画像中に写っている場合でも検出できる可能性が高い．そのため物体が極端に近いために，物体全体が写り込んでいない場合でも検出が可能であるため，今後 Semantic Segmentation を用いた検出も検討したい．

背景の再構成に関しても，画像再構成手法の中でも高い精度を達成している GLCIC と，予備実験で有効性を示した ESP と GFP を適用したことで，人間が再構成された箇所を特定することが困難な精度で再構成することができた．一方で，図 7 の右図のように，道路に写った物体の影が残っている場合，GLCIC による再構成結果がこの影の影響を大きく受けてしまい，全体的に不自然な結果になってしまう場合がある．これを解決する方法として検出されたバウンディングボックスを下に延長する方法が考えられるが，撮影時間によって，影の向きや大きさが異なるため根本的な解決には至らない可能性が高い．そのため，影を検出するネットワーク [23] を Semantic Segmentation のネットワークに統合して影をマスクに含める手法の開発も今後検討したい．

また，GANonymizer によって生成された箇所があたかも本物の道路であるかのように写っているため，都市インフラの監視の際に，生成された箇所が道路の損傷として検出されてしまうというリスクもある．一方で，このリスクを減らすため生成した箇所に，印を残してしまうと，物体を除去することによってプライバシに関する物体の情報を完全に除去するという目的から逸れる．このトレードオフ関係についても今後検討していく必要がある．

7. まとめ

本研究では，都市の移動カメラで撮影した画像からプライバシに関する物体を除去することが可能な画像匿名化手法 GANonymizer の提案を行った．GANonymizer は，SSD512 によってプライバシに関する物体を検出し，検出された箇所の背景を GLCIC とそれを補助する 2 つのパディング層（ESP および GFP）によって推定し，それを合成することで画像からプライバシに関する物体の除去を行う．実験では，まず背景再構成を補助する ESP と GFP に関しての有効性を示し，次に撮影した実際の移動カメラからの道路画像を用いて，GANonymizer 全体の評価を行った．結果，画像に含まれるプライバシに関する物体の 6 割以上に対し，人間が除去された箇所の識別困難な精度で，除去できることが確認された．本提案手法により，都市カメラで撮影された多種多様な映像の共有や利活用が進み，都市のスマート化に寄与することが期待される．

謝辞 本研究の一部は国立研究開発法人情報通信研究機構に支援をいただいた．

参考文献

[1] Ballester, C., Bertalmio, M., Caselles, V., Sapiro, G. and Verdera, J.: Filling-in by joint interpolation of vector fields and gray levels (2000).
[2] Barnes, C., Shechtman, E., Finkelstein, A. and Goldman, D.B.: PatchMatch: A randomized correspondence algorithm for structural image editing, *ACM Trans. Graphics (ToG)*, Vol.28, No.3, p.24, ACM (2009).
[3] Bertalmio, M., Sapiro, G., Caselles, V. and Ballester, C.: Image inpainting, *Proc. 27th Annual Conference on Computer Graphics and Interactive Techniques*, pp.417–424, ACM Press/Addison-Wesley Publishing Co. (2000).
[4] Cheung, S.-C., Venkatesh, M.V., Paruchuri, J., Zhao, J. and Nguyen, T.: Protecting and managing privacy information in video surveillance systems, *Protecting Privacy in Video Surveillance*, pp.11–33, Springer (2009).
[5] Chinomi, K., Nitta, N., Ito, Y. and Babaguchi, N.: PriSurv: Privacy protected video surveillance system using adaptive visual abstraction, *International Conference on Multimedia Modeling*, pp.144–154, Springer (2008).
[6] Ebdelli, M., Le Meur, O. and Guillemot, C.: Video inpainting with short-term windows: Application to object removal and error concealment, *IEEE Trans. Image Processing*, Vol.24, No.10, pp.3034–3047 (2015).
[7] Everingham, M., Van Gool, L., Williams, C.K., Winn, J. and Zisserman, A.: The pascal visual object classes (voc) challenge, *International Journal of Computer Vision*, Vol.88, No.2, pp.303–338 (2010).
[8] Geiger, A., Lenz, P. and Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite, *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp.3354–3361, IEEE (2012).
[9] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y.: Generative adversarial nets, *Advances in Neural Information Processing Systems*, pp.2672–2680 (2014).
[10] Grosek, J. and Kutz, J.N.: Dynamic mode decomposition for real-time background/foreground separation in video, arXiv preprint arXiv:1404.7592 (2014).
[11] Hays, J. and Efros, A.A.: Scene completion using mil-

lions of photographs, *ACM Trans. Graphics* (*TOG*), Vol.26, No.3, p.4 (2007).
[12] Iizuka, S., Simo-Serra, E. and Ishikawa, H.: Globally and locally consistent image completion, *ACM Trans. Graphics* (*ToG*), Vol.36, No.4, p.107 (2017).
[13] Kawano, M., Mikami, K., Yokoyama, S., Yonezawa, T. and Nakazawa, J.: Road marking blur detection with drive recorder, *2017 IEEE International Conference on Big Data* (*Big Data*), pp.4092–4097, IEEE (2017).
[14] Le, T.T., Almansa, A., Gousseau, Y. and Masnou, S.: Motion-consistent video inpainting, *2017 IEEE International Conference on Image Processing* (*ICIP*), pp.2094–2098, IEEE (2017).
[15] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y. and Berg, A.C.: Ssd: Single shot multibox detector, *European Conference on Computer Vision*, pp.21–37, Springer (2016).
[16] Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T. and Omata, H.: Road damage detection and classification using deep neural networks with smartphone images, *Computer-Aided Civil and Infrastructure Engineering*, Vol.33, No.12, pp.1127–1141 (2018).
[17] Morishita, S., Maenaka, S., Nagata, D., Tamai, M., Yasumoto, K., Fukukura, T. and Sato, K.: Sakurasensor: Quasi-realtime cherry-lined roads detection through participatory video sensing by cars, *Proc. 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp.695–705, ACM (2015).
[18] Newson, A., Almansa, A., Fradet, M., Gousseau, Y. and Pérez, P.: Video inpainting of complex scenes, *SIAM Journal on Imaging Sciences*, Vol.7, No.4, pp.1993–2019 (2014).
[19] Park, S., Kim, J., Mizouni, R. and Lee, U.: Motives and concerns of dashcam video sharing, *Proc. 2016 CHI Conference on Human Factors in Computing Systems*, pp.4758–4769, ACM (2016).
[20] Patwardhan, K.A., Sapiro, G. and Bertalmío, M.: Video inpainting under constrained camera motion, *IEEE Transactions on Image Processing*, Vol.16, No.2, pp.545–553 (2007).
[21] Senior, A.: Privacy protection in a video surveillance system, *Protecting Privacy in Video Surveillance*, pp.35–47, Springer (2009).
[22] Shetty, R.R., Fritz, M. and Schiele, B.: Adversarial scene editing: Automatic object removal from weak supervision, *Advances in Neural Information Processing Systems*, pp.7706–7716 (2018).
[23] Wang, J., Li, X. and Yang, J.: Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.1788–1797 (2018).
[24] Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J. and Catanzaro, B.: High-resolution image synthesis and semantic manipulation with conditional gans, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.8798–8807 (2018).
[25] Ye, X., Yang, J., Sun, X., Li, K., Hou, C. and Wang, Y.: Foreground–background separation from video clips via motion-assisted matrix restoration, *IEEE Trans. Circuits and Systems for Video Technology*, Vol.25, No.11, pp.1721–1734 (2015).
[26] Yu, H., Lim, J., Kim, K. and Lee, S.-B.: Pinto: Enabling Video Privacy for Commodity IoT Cameras, *Proc. 2018 ACM SIGSAC Conference on Computer and Communications Security*, pp.1089–1101, ACM (2018).
[27] Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X. and Huang, T.S.: Generative image inpainting with contextual attention, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.5505–5514 (2018).
[28] Zhou, B., Lapedriza, A., Khosla, A., Oliva, A. and Torralba, A.: Places: A 10 million image database for scene recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.40, No.6, pp.1452–1464 (2017).

谷村 朋樹

慶應義塾大学環境情報学部在学中．主に画像処理，深層学習，機械学習，ユビキタスコンピューティングシステムの研究に従事．

河野 慎

2016 年東京大学大学院学際情報学府修士課程修了．2019 年慶應義塾大学大学院政策・メディア研究科博士課程修了．学術振興会特別研究員（DC1）．2019 年より東京大学大学院工学系研究科技術経営戦略学専攻特任研究員．博士（政策・メディア）．主に深層学習，機械学習，ユビキタスコンピューティングシステム，サイバーフィジカルシステムの研究に従事．人工知能学会学生会員．

米澤 拓郎 （正会員）

2010 年慶應義塾大学大学院政策・メディア研究科後期課程博士号取得後，同大学院特任助教，特任講師，特任准教授を経て，2019 年より名古屋大学大学院工学研究科准教授．主に，ユビキタスコンピューティングシステム，ヒューマンコンピュータインタラクション，センサネットワーク等の研究に従事．ACM 会員．

中澤 仁 （正会員）

1975 年生．1998 年慶應義塾大学総合政策学部卒業．2001 年同大学大学院政策・メディア研究科修士課程修了．2001 年同大学院同研究科博士課程修了．現在，慶應義塾大学環境情報学部教授．博士（政策・メディア）．ミドルウェア，システムソフトウェア，ユビキタスコンピューティング，センサネットワーク等の研究に従事．電子情報通信学会，ACM，IEEE 各会員．