

GANonymizer: 物体検出と敵対的生成を統合した 映像匿名化手法

谷村 朋樹^{1,a)} 河野 慎^{2,b)} 米澤 拓郎^{2,c)} 中澤 仁^{1,d)}

概要: 街の様子を映した映像を分析することで、街の状態を自動で把握、予測することが可能となる。しかし、街の映像には人や車などのプライバシーに関する物体が含まれているため、無加工の状態ですべてネットワークに送信・共有したり、利用することは難しい。結果、プライバシー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。本研究では、映像からプライバシーに関する物体を自動で検出し、そもそもそこに存在しなかったかごとく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ーに関する物体を検出する。そして敵対的生成ネットワークを用いて、検出した物体を取り除いた画像を生成し、元の画像に合成する。本研究では、実際の街を走る車から撮影した画像に適用し、その結果について報告する。

GANonymizer: Anonymization Method for Image Using Object Detection and Adversarial Generation

TOMOKI TANIMURA^{1,a)} MAKOTO KAWANO^{2,b)} TAKURO YONEZAWA^{2,c)} JIN NAKAZAWA^{1,d)}

1. はじめに

街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、街の様子を撮影している。また街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街のインフラ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・蓄積し共有することができれば、大規模な街の映像データセットが構築でき、街のインフラの監視や分析に活用することが可能となる。例として、街を周回する車両からの映像を解析することで、道路の劣化を自動で検出する手法等が提案されている [5][7]。しかしながら、街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ー情報が含まれており、プライバシー侵害の危険性を常に生じさせてしまう。その結果、現状ではプライバシー侵害の懸念もあり、移動カメラ映像は大規模なデータセットとして有効活用がされることなく、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。よって、今後こういった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 らは画像や映像の匿名化処理は、映像の記述粒度を荒くすることにより情報の具体性を下げることであると、記述粒度に関して映像の利用者と被写体の関係を考慮して選択すべき、実写から透明までの段階を定義した [1]。この Chinomi らの定義に基づき、本研究では街の映像に映り込む不特定多数の被写体のプライバシー侵害を極限まで抑えるため、映像の記述粒度としてもっとも抽象度が高く、映像から対象を除去する「透明」を目的とする。

映像から対象を除去する手法には映像の場合、フレーム間での背景差分を用いたパッチ処理がよく用いられる [9]。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

¹ 慶應義塾大学環境情報学部
Faculty of Information and Environment Keio University, Fujisawa, Kanagawa 252-0882, Japan

² 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

a) tanimu@ht.sfc.keio.ac.jp

b) makora@ht.sfc.keio.ac.jp

c) takuro@ht.sfc.keio.ac.jp

d) jin@ht.sfc.keio.ac.jp

ラの場合には非常に有効な手段となる。しかしながら、ドライブレコーダのような撮影している方向や走行速度が一定でない移動カメラの場合には、背景差分を獲得することは難しく、有効でない。また、複数のドライブレコーダの映像を共有しデータベース化してそこから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パッチ処理を行う方法等も考えられる。しかしながら、これらの手法もそもそも匿名化処理を施す前段階での共有や蓄積は難しいことや、エッジ側で蓄積しておく場合にはデータ量が膨大すぎるなどの問題点があり、有効でない。そのため、匿名化処理は映像をフレームごとに切り出して画像に対して行う必要がある。

そこで本研究では、画像からプライバシーに関する物体を自動で検知し、そもそもそこに存在しなかったかごとく画像中から消去する GANonymizer を提案する。GANonymizer は、プライバシー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を画像中から除去するネットワークの2つで構成されている。まず、プライバシー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等のプライバシーに関する物体の検出を行う。そして、検出された物体を除去するネットワークでは、画像の検出された物体の箇所を、画像全体を見て再構成することによって、画像中からの物体の除去を行う。

また、GANonymizer は街の画像に対しての適応を想定しているため、街の画像の特徴を考慮し、検出と除去の間に2つの前処理を導入した。一つ目は、画像の端の通行人や車両を除去するための事前パディング処理で、二つ目は高解像度な画像中の大きなマスクに対処するための2段階生成処理である。この2つの前処理によって、より自然にプライバシー情報を画像から消去することができた。

実験では、街を走る車からスマートフォンで撮影した映像に GANonymizer を適用し、その有効性について考察を行った。画像データの性質から、プライバシーに関する物体は、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。全 5246 件の画像に適用し、その出力結果から GANonymizer の有効性について、考察と議論を行った。

本稿の構成は以下の通りである。まず、2章では想定している街の利活用の例を説明し、その際に生じ得るプライバシー問題とプライバシー問題に対する他の取り組みについて述べる。3章では、2章で述べた問題に対する解決策として、GANonymizer を提案しその詳細について述べる。4章では GANonymizer を実際に街の画像に適用した結果について述べる。そして、5章では結果についての考察と議論を行い、6章で本論文をまとめる。

2. 街の映像の利活用

ドライブレコーダのような移動カメラから街の様子を撮影した映像を収集、蓄積、共有することができれば、大規

模な街の映像を利用することが可能になる。大規模な街の移動カメラからの映像を利用することにより、街のインフラを監視したり、街のインフラ分析のための大規模データを作成、公開することが可能となる。街の移動カメラからの映像を使用した街のインフラ監視や大規模データ作成の研究は盛んに行われている。

Kawano らは街の道路の画像を用いて道路の白線や道路標識のかすれの自動検出を行なっている [5]。街の画像には、定期的に街を網羅性を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用いた。そして、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自動でリアルタイムに白線や道路標識のかすれ検出を行なえるシステムを提案した。

Maeda らは街の道路の損傷の自動検出を行なった。街を走る車からドライブレコーダのように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等の道路損傷のアノテーションを行なったデータセットとして公開している。またそれらのデータセットを用いて軽量の物体検出モデルである MobileNet SSD を学習させ、損傷の自動検出を行えるスマートフォン用アプリケーションの提案も行なっている [7]。

しかしながら、個人情報保護法やヨーロッパでは一般データ保護規則 (General Data Protection Regulation: GDPR) が施行されるなど、個人情報保護の動きが加速しているため、人や車などのプライバシー情報を含む映像を扱うこれらの取り組みにも規制がかかる可能性がある。本章では、ここからまず街の画像や映像におけるプライバシーの問題について述べ、その後問題に対する他の取り組みやそこで生じている技術的課題について述べる。

2.1 プライバシー問題

現状でも、街の映像には人や車などプライバシーに関する物体が含まれているため、無加工の状態で蓄積、共有、利用することが難しく、プライバシー侵害の危険性がある映像は消去される。そのため、街の映像の利活用には映像のプライバシー侵害の危険性を極限まで下げる必要がある。

プライバシー侵害の危険性を低下させる匿名化処理には、検出した対象に対してのモザイクや除去など複数の方法が存在する。Chinomi らは画像や映像の匿名化処理は、映像の記述粒度を荒くすることにより情報の具体性を下げることであったとした。また人物映像に含まれる情報の抽象化度を代表的な画像処理による記述粒度の変化と対応させ、抽象化度の低いものから順に、実写、シースルー、モノトーン、ぼかし、モザイク、エッジ、輪郭、シルエット、ボックス、バー、ドット、透明と定義した。そして、抽象化度の高さと画像の元々の有効性はトレードオフの関係にあるため、抽象化度合いの選択には映像を利用する側と、被写体

との関係を慎重に考慮すべきであるとしている。一方街の画像における被写体は不特定多数であるため、映る人や物体の所有者を特定することや、被写体全員の意向を聞くことは難しい。そこで本研究では、街のインフラの監視が目的であることから、もっとも抽象度の高い映像の記述粒度である映像から対象を除去する「透明」を目的とした。

2.2 画像匿名化における技術的課題

画像や映像の匿名化は、防犯カメラの普及や、ドライブレコーダの普及、インターネットの発展に伴い、プライバシー問題が注目されるようになったことで、盛んに研究が行われるようになった。古典的な匿名化手法には、事前に人の顔が登録されたデータベースからのマッチングにより、人の顔を特定し、モザイクなどの処理を施す手法や、画像中の特徴量から人の顔の位置を計算し、処理を施す手法がある [12][2][1]。しかしながら、これらの手法は大規模なデータベースが必要になるため、街の画像のような不特定多数の人々が映り込む場合においては、有効でない。また Senior らや Luo らなどは画像匿名化の技術を用いた統合的なカメラ監視システムを提案しているが [12][14]、これらはカメラを含めた統合的なシステム運用を考えているため、不特定多数のカメラ映像に対して適用しようとするコストが高い。

一方、深層学習と画像処理技術の発展に伴い、画像中から人や車などの特定の物体の検出を高い精度で行うことができるようになった [6][11]。また、畳み込みニューラルネットワーク (CNN) と敵対的生成ネットワーク (GAN) により、高い精度での画像生成を行うことが可能になった [10]。GAN を用いて、画像の欠損部分を周りの情報を元に生成する画像再構成タスクにおいて、高い精度を出すことも可能となり [8][4]、Shetty らはこれを物体除去に応用している [13]。

そこで本研究では、深層学習を用いた物体検出ネットワークと敵対的生成ネットワークを統合することで、街の画像や映像から、プライバシーに関する物体を検出し除去することができる GANonymizer の提案を行う。

3. 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ーに関する物体を自動で検知し、そもそもそこに存在しなかったかごとく画像上から消去する GANonymizer を提案する。GANonymizer は匿名化対象の検出と画像再構成による匿名化処理の大きく 2 つから構成される (図 1)。また、街の画像の特徴を考慮し、2 つの事前処理を導入した。

3.1 匿名化対象の検出

匿名化対象の検出は、物体検出ネットワーク SSD (Single Shot Multibox Detector) [6] を用いて、入力画像から人や車

などのプライバシー情報の位置と大きさを検出する (図 1 の左側)。SSD は、一枚の画像から物体検出を高精度で行うことのできるネットワークで、既存の FCN から複数のスケールの特徴マップを作成し、それぞれから検出を行っている。それにより、物体の大きさに影響を受けずに検出を行うことができるため、高精度での検出が可能になっている。GANonymizer で使用した SSD は、Pascal VOC データセット [3] で学習済みのモデルを用いている。Pascal VOC データセットは人や車を含む 20 クラスの物体検出のためのアノテーションがされている。GANonymizer の主要な匿名化対象は人や車も含まれているため、Pascal VOC データセットで学習済みのモデルを使用した。SSD は高精度で検出が可能な入力画像を 512x512 に圧縮して行うモデルを使用した。

3.2 画像再構成による匿名化

検出された物体を除去するネットワークは、敵対的生成ネットワークにより画像再構成をおこなう GLCIC (Globally and Locally Consistent Image Completion) [4] を用いた (図 1 の右側)。画像再構成による匿名化では、まず検出された物体の位置と大きさをを用いて、検出された物体に該当するピクセルは 1、それ以外のピクセルは 0 で、入力画像と同サイズのマスク画像を生成する。そして、元々の入力画像と作成したマスク画像を GLCIC に入力し、入力画像のマスクに該当する箇所を敵対的生成による画像再構成を行うことにより、プライバシーに関する物体を除去した画像を生成する。GLCIC は、入力画像とマスク画像を入力として、入力画像のマスク部分を生成する Completion ネットワークと、Completion ネットワークから再構成された画像と、元々の入力画像のどちらが元々の入力画像かをマスク部分のみを見て識別する Local Discriminator と画像全体を見て識別する Global Discriminator から構成されている。学習は、Completion ネットワークは以下の式が最小化するように学習し、Discriminator 側は以下の式が最大化するように学習することで、敵対的学習を行う。

$$\min_C \max_D \mathbb{E} [L(x, M_c) + \alpha \log D(x, M_d) + \alpha \log(1 - D(C(x, M_c), M_c))]$$

その結果、Discriminator が Completion ネットワークで再構成した画像かどうかを識別することができなくなり、Completion ネットワークは元々の画像に近い自然な再構成を行うことができるようになる。また、GANonymizer では Places データセット [15] で学習済みのモデルを使用した。Places データセットは、街や風景の写真以外にも、室内の写真や人の写真、白黒画像など、多様な 1000 万件の画像データで構成されている。そのため、Places データセットを使って GLCIC を学習させることにより、画像再構成において、他のデータセットで学習させたモデルより

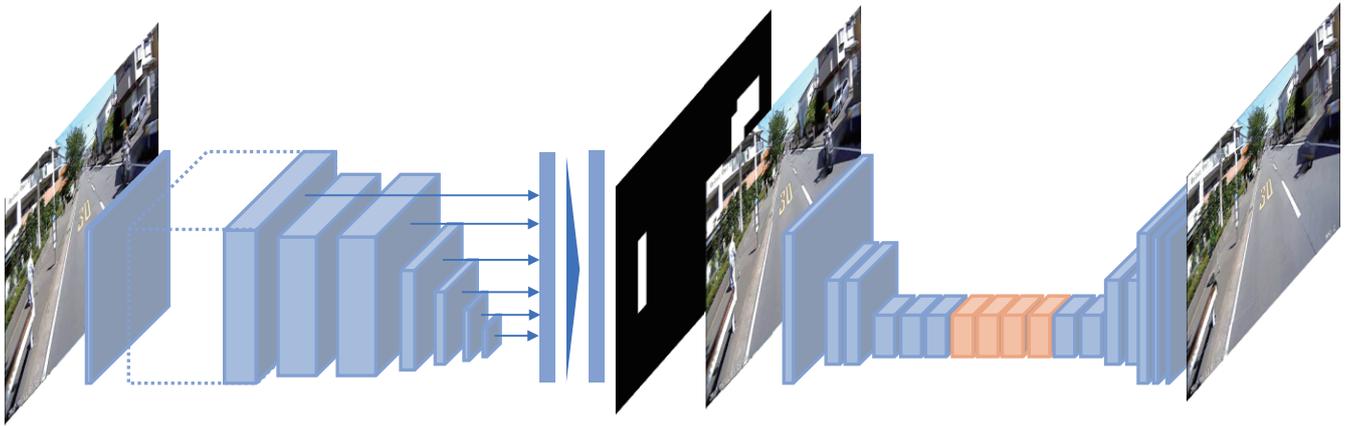


図 1 GANonymizer のネットワーク全体の構成. 中心の 2 枚の画像を境に, 左は匿名化対象検出ネットワーク. 右は検出対象の除去のための敵対的生成ネットワーク.

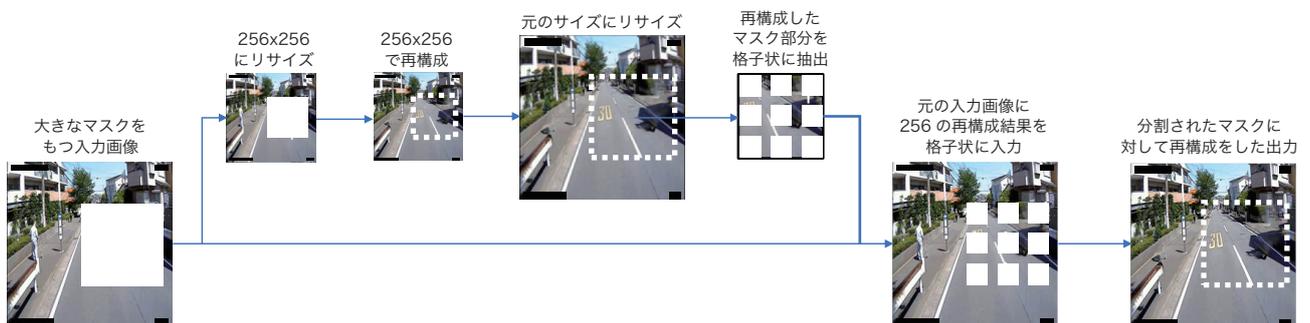


図 2 2 段階生成処理: 高解像度画像において大きなマスクを再構成する場合でも, 一度入力画像を 256×256 のサイズに圧縮してから再構成し, その出力結果を補助情報として入力することで, 擬似的なマスク分割を行う.

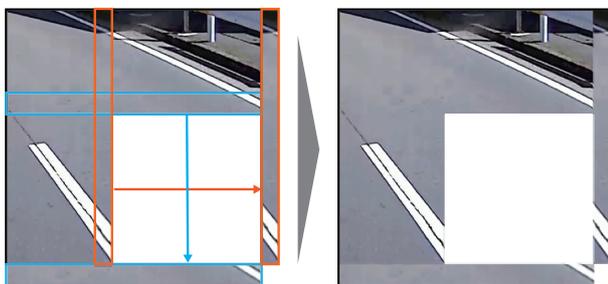


図 3 事前パディング処理: 画像の端にマスクがある場合でも, マスクの画像ない側の側面のちょうど内側のピクセル列を, 画像の端にコピーすることで一時的に画像を拡張し, マスクに対して全方向から情報が伝達される.

も多くの画像に対して, 自然な生成結果を出力することが可能な万能なモデルを作成することができる. そこで GANonymizer では, Places データセットで学習したモデルを用いた.

3.3 実用に向けた改善

街の画像には, 2 つの特徴がある. 一つ目は, 特定の対象を撮影しているわけではないため, 画像の端に物体が写り込んでいることである. 二つ目は, 物体やカメラが動くため, 物体との距離が常に一定でないことである. そのた

め, 画像の端のマスクに対して, 敵対的生成を行う必要性や, 高解像度画像において大きなマスクに対して敵対的生成を行う必要性がある. 一方で, マスクが画像の端にある場合, マスク外からの情報が少ない端は自然な生成が困難になる. また, 高解像度画像で, 大きなマスクに対して, 画像再構成を行う場合, マスクの中心までマスク外からの情報が届かず, 自然に生成できない.

そこで, マスクの端に対しては事前パディング処理を, 大きなマスクに対しては, 2 段階生成処理を取り入れることで, 自然な生成を可能にした. 事前パディング処理では, マスク外から情報が来ない側に画像中からピクセルをコピーしてパディングすることで, 一時的に入力画像を拡張し, 全方向からマスクに対して情報が行き渡るようにした. コピーする画像中のピクセルには, マスクにおいてパディングする側面と反対側面のマスクのちょうど外側のピクセルを用いた. 事前パディング処理の詳細を図 3 に示す.

2 段階生成処理では, まず高解像度入力画像とマスク画像を 256×256 にリサイズし, リサイズした画像に対して画像再構成を行い検出された物体を消去する. その後, 256 のサイズで再構成済みの出力画像を, 元の入力画像のサイズにリサイズする. その結果を, 元々の入力画像に補助情報として格子状にコピーして与え, 擬似的なマスク分割を

表 1 GANonymizer の CPU での処理速度

画像サイズ	ピクセル数	SSD(秒)	GLCIC(秒)	全体(秒)
1080x1940	2,095,200	0.503	28.196	29.998

行う。補助情報が付与された入力画像を GLCIC に入れることによって、大きなマスクに対しても自然な生成を可能にした。2 段階生成処理の詳細を図 2 に示す。

4. 実験

4.1 使用する画像と実験環境

実験にはドライブレコーダに見立てた街を走る車から iPhone7 で撮影した動画をフレームごとに切り出した画像を使用した。撮影は 2018 年 7 月末の日中と夕暮れ時に、神奈川県藤沢市で行なった。夜に関しては、物体が鮮明に映っていることが少ないため、映像のプライバシー侵害の危険性は低く、街のインフラ監視を想定した場合、暗く写りが鮮明でない画像は有用でないため、撮影の対象時間は明るい時間帯と薄暗い時間帯のみとした。使用した画像は全 5246 枚で、1 枚の画像サイズはリサイズせず撮影したままの 1080x1920 のものを用いた。

画像の特徴から、プライバシーに関する物体は、人、車、バス、自転車、バイクとし、それらを画像から除去する実験を行なった。また、2 段階生成処理は 1 辺が 200px 以上のマスクに対して適用し、マスクの外側から画像の端までが 3px 以下の場合に、事前バディング処理を適用した。

4.2 結果

図 4 と図 7 に GANonymizer を街の日中の画像と夕暮れ時の画像に適用した結果を示す。画像中から車や歩行者等を自然に除去することができた。画像の端のマスクを再構成する場合の事前バディング処理により、画像の端に映る物体の除去をより自然に行うことができた(図 5)。また、大きなマスクを再構成する場合の 2 段階生成処理では、マスクの中心部まで他の画像の箇所からの情報が伝達されず、不自然な出力になる問題の改善ができた(図 6)。

一方で、図 7 の左の画像に示すように、検出対象との距離が近く、画像のなかで大きく映りすぎている場合や、カメラから遠くにある小さい車や人を消去する場合には、検出することが難しかった。図 7 の右図のように、小さい物体の除去に比べ大きな物体を除去する場合にはマスク部分の画質だけ荒くぼやけた画像が出力されてしまっている。

実験環境は、Intel Xeon CPU E5-1680 v3 @ 3.20GHz で行なった。処理速度は大きくは入力画像の解像度に依存し、2 段階生成処理を行う場合にはより時間がかかった。平均の処理速度について表 1 に示す。

4.3 考察と課題

図 4 の通り、入力画像に対して検出した物体を画像中か

ら自然に除去することができた。しかしながら、物体との距離が極端に近い場合や遠い場合、画像から完全にプライバシーに関する物体を消去することが難しかった。物体との距離が遠いために検出できない場合は、プライバシー問題が生じ得る可能性は低いため、問題にはならない。一方、近い物体は鮮明に映っているため、検出できなかった場合、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、プライバシー侵害の危険性が高い。極端に近い物体であれば、検出は難しく、距離に関係なく 100 % の精度で検出することは不可能である。そのため、物体全体は検出できない場合でも、人であれば顔、車であればナンバープレートといった物体においてよりプライバシー侵害の危険性が高い重要な部分だけでも検出できることが重要となる。そこで、Semantic Segmentation により匿名化対象を細かく検出したり、学習データセットにナンバープレートや人の顔などの情報も入れて、重要な部分の検出を逃さないよう手法の改良が今後の課題である。

また、図 7 の右図のように、2 段階生成処理を行うと、小さいマスクに対する 1 度の再構成に比べて、出力が荒くなってしまう。しかしながら、本実験に使用した 1080x1920 等の高解像度画像における大きなマスクの再構成に関して、鮮明で周りに溶け込んだ適当な再構成結果を出力できる画像再構成モデルはない。大きなマスクに対して鮮明かつ周りに溶け込んだ再構成結果を出力することは、元々の画像の有用性を担保する上で、重要な点であるため、画像再構成モデルの改良にも今後の課題として取り組みたい。

5. まとめ

プライバシーに関する物体が含まれていることにより、街を移動なカメラで撮影した映像が消去されている問題に着目し、画像のプライバシー問題を生じうる可能性の低下に取り組んだ。画像中からプライバシーに関する物体を検出し、そもそもそこに存在しなかったかごとく画像中から消去する GANonymizer を提案した。実験では、GANonymizer を街を走行する車からスマートフォンで撮影した画像に適用し、実際に画像から人や車などのプライバシーに関する物体を消去することができた。一方で、検出に関しては重要なプライバシーの確実な対処、再構成に関しては高解像度の大きいマスクの生成に関して課題が残ったため、今後取り組んでいきたい。

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頂いた。

参考文献

- [1] Chinomi, K., Nitta, N., Ito, Y. and Babaguchi, N.: PriSur: privacy protected video surveillance system using adaptive visual abstraction, *International Conference on Multimedia Modeling*, Springer, pp. 144–154 (2008).



図 4 GANonymizer でプライバシー情報を除去した結果 (左二つ) 日中に撮影した画像と (右二つ) 夕暮れ時に撮影した画像に対して適用. (上段) 入力画像と (下段) 出力画像.



図 5 事前バディング処理の効果: (左) 通常の再構成結果と (右) 事前バディング処理を適用した場合の再構成結果



図 6 2段階生成処理の効果: (左) 通常の再構成結果と (右) 2段階生成処理を適用した場合の再構成結果



図 7 (左: 検出不可例) 物体との距離が極端に近いために大きく映り過ぎることで検出失敗している. (右: 荒い再構成結果の例) 中央の車のような大きな物体の除去は, 左や左奥の小さな物体の除去に比べ, 再構成結果が荒い. (上段: 入力, 下段: 出力)

[2] Dufaux, F. and Ebrahimi, T.: A framework for the validation of privacy protection solutions in video surveillance, *Multimedia and Expo (ICME), 2010 IEEE International Conference on, IEEE*, pp. 66–71 (2010).

- [3] 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).
- [4] Iizuka, S., Simo-Serra, E. and Ishikawa, H.: Globally and locally consistent image completion, *ACM Transactions on Graphics (TOG)*, Vol. 36, No. 4, p. 107 (2017).
- [5] Kawano, M., Mikami, K., Yokoyama, S., Yonezawa, T. and Nakazawa, J.: Road marking blur detection with drive recorder, *Big Data (Big Data), 2017 IEEE International Conference on, IEEE*, pp. 4092–4097 (2017).
- [6] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y. and Berg, A. C.: Ssd: Single shot multi-box detector, *European conference on computer vision*, Springer, pp. 21–37 (2016).
- [7] Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T. and Omata, H.: Road Damage Detection Using Deep Neural Networks with Images Captured Through a Smartphone, *arXiv preprint arXiv:1801.09454* (2018).
- [8] Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T. and Efros, A. A.: Context encoders: Feature learning by inpainting, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2536–2544 (2016).
- [9] Patwardhan, K. A., Sapiro, G. and Bertalmio, M.: Video inpainting under constrained camera motion, *IEEE Transactions on Image Processing*, Vol. 16, No. 2, pp. 545–553 (2007).
- [10] Radford, A., Metz, L. and Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks, *arXiv preprint arXiv:1511.06434* (2015).
- [11] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A.: You only look once: Unified, real-time object detection, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788 (2016).
- [12] Senior, A., Pankanti, S., Hampapur, A., Brown, L., Tian, Y.-L., Ekin, A., Connell, J., Shu, C. F. and Lu, M.: Enabling video privacy through computer vision, *IEEE Security & Privacy*, Vol. 3, No. 3, pp. 50–57 (2005).
- [13] Shetty, R., Fritz, M. and Schiele, B.: Adversarial Scene Editing: Automatic Object Removal from Weak Supervision, *arXiv preprint arXiv:1806.01911* (2018).
- [14] Winkler, T. and Rinner, B.: Security and privacy protection in visual sensor networks: A survey, *ACM Computing Surveys (CSUR)*, Vol. 47, No. 1, p. 2 (2014).
- [15] Zhou, B., Lapedriza, A., Khosla, A., Oliva, A. and Torralba, A.: Places: A 10 million image database for scene recognition, *IEEE transactions on pattern analysis and machine intelligence* (2017).