The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

多様な仮想空間を構築するための画像モダリティ変換

Image Modality Translation for Enriching Virtual Space

益田慎太 *1 Shinta Masuda　町田貴史 *2 Takashi Machida　松原崇 *1 Takashi Matsubara　上原邦昭 *1 Kuniaki Uehara

*1神戸大学 大学院システム情報学研究科 Graduate School of System Informatics, Kobe University　*2(株) 豊田中央研究所 Toyota Central R&D Labs., Inc.

Following great successes of machine learning in various benchmarks, its practical use is attracting attention. The machine learning system has to be trained using a wide variety of data samples and to be tested under various conditions, but collecting numerous data samples is very costly. Here, a demand for data augmentation arises. In this paper, we tackle the augmentation of real images by translating their modality to another modality such as daytime vs. night-time. This data augmentation enables us to train and test the machine learning system in various modality. We first demonstrate that existing approaches, pix2pix and cycle-GAN have some difficulties of applying data augmentation; pix2pix requires paired samples in both modalities or cannot overcome the difference in the modalities, and cycle-GAN sometimes fails in keeping consistency in both modalities. We propose modifications of these methods, which improve the consistency in image modality translation.

1. はじめに

深層学習を始めとした機械学習が様々なベンチマーク問題で目覚ましい成果を上げたのに伴い，機械学習システムを自動運転やロボット制御のような実問題に応用する研究が盛んに行われている．実問題への応用には機械学習システムの訓練や検証が必要であり，そのためには物体を検出するための画像データと正解データのような，現実空間でのデータを大量に収集する必要がある．また充分な訓練と検証のためには単なるデータ量だけでなく，データの多様性が必要である．例えば，昼の晴天時で十分な精度を示す物体認識システムが，夜の雨天時に同等の性能を示す保証はない．そして雨天のデータを収集するには，実際に夜に雨が降るのを待つしかない．また大雨や雷雨，霧など様々なケースが考えられ，そのようなデータ収集しつくすためには莫大な金銭的・人的コストが要求される．

一つの解決策として，環境シミュレータを構築し，その中で機械学習システムの訓練や検証することが考えられる．しかし環境シミュレータと現実空間の差異によって，訓練の性能や検証の妥当性は大きく損なわれる．もう一つの解決策として，ドメイン適応を用いる手法が盛んに研究されている [Shrivastava 17]．ドメイン適応とはあるドメイン（たとえば環境シミュレータ）で得られたデータが持つドメイン独自の情報に手を加えし，他のドメイン（たとえば現実空間）から得られたデータと同じように扱えるようにする手法である．この場合は，シミュレータで大量にデータを収集し，得られた画像を事後処理で現実空間で得られたデータのように加工することを意味する．ただやはり高性能なシミュレータを開発する必要があり，そのコストは無視し難い．もう一つの解決策として，現実空間で得られた 1 つの画像から，様々な環境下の画像を複数生成することが考えれる．例えば，昼と夜という画像のモダリティ変換をドメイン変換問題として捉え，高い性能を発揮している pix2pix[Isola 17] という手法が存在する．これは変換前と変換後の両ドメインにデータ対（ペア画像）が必要であり，定点カメラには適用可能だが車載カメラでは適用できないなど，範囲が限られている．一方，絵画と写真という画像のモダリティ変換をペア画像なしで行える Cycle-GAN[Zhu 17] という手法が提案されている．

本研究では，自動運転用の物体検出システムを想定し，車載カメラから見た道路画像のデータセットである KAIST Multispectral Pedestrian Detection Benchmark[Hwang 15] をベンチマークとし，画像の夜と昼というモダリティを相互に変換するタスクを論じる．既存手法である pix2pix と Cycle-GAN の問題点を指摘した上で，それらを解決する手法を提案する．

連絡先: 松原崇，神戸大学 大学院システム情報学研究科，matsubara@phoenix.kobe-u.ac.jp

2. pix2pix によるモダリティ変換

pix2pix[Isola 17] とは，条件付き生成的敵対ネットワーク (conditional GAN)[Mirza 14] の手法を用い，ペア画像から教師あり学習によりドメイン変換を行う手法である．変換前後のドメインのデータ $x \in X$ と $y \in Y$ について，変換器 G_Y を作成し，変換後のデータ $\hat{y} = G_Y(x)$ を得る．そして条件付き分布 $p(y|x)$ と $p(\hat{y}|x)$ の分布間のダイバージェンスを生成的敵対ネットワーク (GAN) の手法で最小化する．ペア画像には被写体が寸分変わらず同じ位置同じ角度で存在している必要がある．その為，参考文献 [Isola 17] では建物を写した定点カメラ画像を用いて，昼と夜のモダリティ変換を行っていた．しかし道路事情は刻一刻と変化するため，車載カメラでペア画像を入手することは困難である．そこで元のデータとペアで集める事が容易な赤外線データを用いることにより間接的に昼と夜のモダリティ変換を行う．方法として，ドメイン X（昼または夜）の赤外線のドメインを X_{Inf}，ドメイン Y（X が昼ならば夜）の赤外線のドメインを Y_{Inf} とおいた時，X_{Inf} と Y_{Inf} があまり形が変わらないことを仮定し，同じドメインとして扱う．X から Y の変換の場合，まず Y_{Inf} から Y への変換を学習する．そのモデルを用いる事で X_{Inf} から Y への変換を行う事で間接的に X から Y へのドメイン変換を行った．

モダリティ変換結果を図 1 にまとめた．上段が昼から夜，下段が夜から昼への変換であり，左列が実画像，中列が変換後の画像である．昼から夜の変換では車が白飛びしてしまい，夜から昼の変換ではボケたような画像になっている．これは，夜の赤外線画像は強度が低くノイズが大きいため，夜の赤外線画像

と同一ドメインとして扱えず，不適切な変換をもたらしたと考えられる．

昼夜の赤外線の強度差を補正すべく，赤外線画像を一画像ごとの平均が揃うように正規化した．結果を図 1 右列にまとめた．昼から夜の変換を見ると，白飛びが抑えられ，自動車の外観が判別可能になっているものの，依然として異常に明るく，またノイズのような乱れがある．夜から昼については大幅な改善が見られるが，看板などが判別不可能なほどボケてしまっている．結果として，十分な結果とは言い難い．

3. Cycle-GAN によるモダリティ変換

Cycle-GAN[Zhu 17] はペア画像のないデータセットについて，教師無し学習によりドメイン変換を行う手法である．pix2pix の変換が一方向であったのに対し，Cycle-GAN は両方向の変換を同時に学習する．この時，ペア画像がないため，GAN の手法では $p(y)$ の $p(\hat{y})$ の分布間のダイバージェンスしか最小化しない．変換前後のデータ x と $\hat{y}$ がペアになることを保証するため，変換後のデータ $\hat{y} = G_Y(x)$ を更に逆方向の変換器 G_X に入れ，再変換後のデータ $\hat{\hat{x}}$ を得て，x と $\hat{\hat{x}}$ の距離を最小化する．再変換可能であれば変換後データ $\hat{y}$ においても変換前のドメインのデータ x の特徴が保持されるだろうという仮定を用いている (cycle consistency)．

モダリティ変換結果を図 2 にまとめた．予想に反して，両ドメインにおける一貫性が保たれていない．昼から夜への変換では被写体の自動車が消失してしまっている．また夜から昼の変換では，木のあった場所に建物のようなものが生成されている．このように変換元ドメインの物体と変換後ドメインの物体が適切に合致しない結果が得られた．変換前後のドメイン X と Y で一貫性を保つため，cycle consistency という拘束を課しているが，たとえば昼の自動車と夜の植木のように全く異なる物体を紐付けして変換しても，cycle consistency は満たされる．参考文献 [Zhu 17] では主な被写体が一つであるか，絵画の画風のような全体的な印象のみを変換していたため大きな問題にならなかったと考えられる．しかし，車載カメラ映像のように多様な物体が写る場合においては，大きな問題である．

そこで，GAN の手法でダイバージェンスを最小化する分布を同時分布である $p(x, \hat{y})$ の $p(\hat{x}, y)$ とした．同時分布の最適化は ALI[Dumoulin 17] の手法の応用と言える．ALI を用いても，誤った紐付けが行われる可能性は残る．しかし，同時分布のダイバージェンスの計算には，ニューラルネットワークを用いて変換前後の二枚の画像を同時に局所的な部分から判定していく．そのため，大局的な変化が必要な誤った紐付けに比べ，正しい紐付けがなされる状況に収束しやすいと予想された．図 2 の結果から，自動車の消失が起こらず，木も木のまま変換されていることが確認でき，提案手法の有効性を示すことが出きた．

4. まとめと今後の課題

本稿では車載カメラ映像を用いて，昼と夜の画像モダリティ変換の手法について提案した．今後の課題として，自動車のライトや街頭のように昼と夜で明らかな変化のある物体や，影のような日照条件に関わる要素は，ペア画像なしで学習することが原理的に不可能である．そのため少ないながらもペア画像用意し，pix2pix と cycle-GAN を組み合わせて半教師あり学習を構築する必要がある．本研究は科研費 (16K12487) の支援を受け開発した技術をもとに，神戸大学上原研究室と株式会社豊田中央研究所の共同研究として実施された．

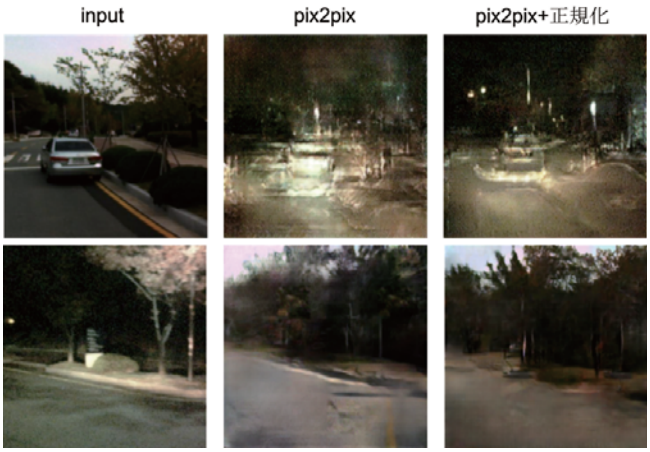

図 1: pix2pix による昼から夜への画像モダリティ変換．

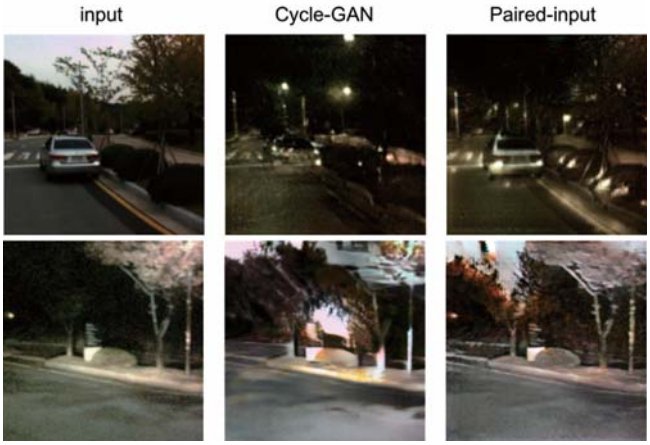

図 2: Cycle-GAN による昼から夜への画像モダリティ変換．

参考文献

[Dumoulin 17] Dumoulin, V. *et al.*: Adversarially Learned Inference, *International Conference on Learning Representations (ICLR)*, (2017)

[Hwang 15] Hwang, S. *et al.*: Multispectral pedestrian detection: Benchmark dataset and baseline, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, (2015)

[Isola 17] Isola, P. *et al*: Image-to-Image Translation with Conditional Adversarial Networks, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017)

[Mirza 14] Mirza, M. and Osindero, S.: Conditional Generative Adversarial Nets, *arXiv*, 1411.1784 (2014)

[Shrivastava 17] Shrivastava, A. *et al*: Learning from Simulated and Unsupervised Images through Adversarial Training, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017)

[Zhu 17] Zhu, J.-Y. *et al*: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, in *IEEE International Conference on Computer Vision (ICCV)*, (2017)