GANと転移学習を併用した仮想都市への高解像度自動マッピング

九州大学　学生会員　○柴田　洋佑
九州大学　学生会員　町田　禎弥
九州大学　非会員　西村　和也
九州大学　非会員　備瀬　竜馬
九州大学　正会員　浅井　光輝

１．緒言

2011 年の東北地方大震災は，観測史上最大の地震として広域的かつ破壊的な被害をもたらしたことは記憶に新しい．この震災による死者・行方不明者は，2 万人を超えており，既存の防災教育の不十分さが改めて指摘されることとなった．

また防災意識の醸造には自然災害をバーチャルに体験することが可能な VR が，非常に効果的であり，昨今のシミュレーション技術及びコンピューターグラフィックスの精度向上に伴い実用化が進められている．しかし，多大なコストを要する構造物のテクスチャマッピングを手動で実施して広範な領域の仮想空間を創造することは，実際は非常に困難である．

他方，Ian らが提唱した，データ分布を捉える生成モデル（以降，生成器）と，対象画像が学習データか生成画像かを判断する識別モデル（以降，識別器）を，相互に学習させることで画像生成を行う Generative Adversarial Networks[1)]（以降，GAN）は，今日まで多種多様に拡張されてきた．中でも，pix2pixHD[2)]は，高解像度への拡張を目的とした生成器，ネットワーク容量低減及びオーバーフィッティング抑制を目的としたマルチスケール識別器，特徴量のマッチングなどを加味することで，従来の GAN では不可能であった高解像度の画像合成を実施できる．本研究では，この pix2pixHD と，類似した他の領域によって学習されたパラメータを転用して学習を行う転移学習[3)]を用いることで，日本の画像データが限られた都市域でのテクスチャマッピングを行うことを検証した．

２．解析手法

本研究では，仮想都市へ写実的なマッピングを実施するため，高解像度の出力を行うことが出来るフレームワークである pix2pixHD 使用した．加えて，出力画像をより写実的にするため，転移学習を用いることとした．

２．１．pix2pixHD

pix2pixHD は，GAN を元に設計されたフレームワークであり，画像の生成を行う生成器と，入力値が生成器由来か否かを判断する識別器を学習させることによって画像の合成を行う．以下に目的関数$\mathcal{L}_{\mathrm{GAN}}$を示す．

$$\min_G \max_D \mathcal{L}_{\mathrm{GAN}}(G,D) = \mathbb{E}_{\mathbf{x}}[\log D(\mathbf{x};\phi)] + \mathbb{E}_{\mathbf{y}}[\ln(1 - D(G(\mathbf{y};\theta);\phi)] \quad (1)$$

ここで，$\boldsymbol{x}$は学習データ，$\boldsymbol{y}$は学習データに対して意味クラスによる分類を実施したインスタンス画像，ϕ,θは各モデルのパラメータである．また，入力$\boldsymbol{x}$に対して画像を正しく識別出来る確率分布関数を$D(\boldsymbol{x};\phi)$，入力$\boldsymbol{y}$に対して合成画像を出力する確率分布関数を$G(\boldsymbol{y};\theta)$とする．なお，$\mathbb{E}_{\mathbf{p}}[\ln q]$は$p$における$q$の交差エントロピーである．

また，解像度の異なる複数の識別器を用いて，ネットワーク容量及び過学習の抑制を行うマルチスケーリング識別器を採用した．以下に目的関数$\mathcal{L}_{\mathrm{FM}}$を示す．

$$\mathcal{L}_{\mathrm{FM}}(G,D_k) = \mathbb{E}_{(\mathbf{x},\mathbf{y})} \sum_{i=1}^{T} \frac{1}{N_i} \left[\left\| D_k^{(i)}(\mathbf{x};\phi) - D_k^{(i)}(G(\mathbf{y};\theta);\phi) \right\|_1 \right] \quad (2)$$

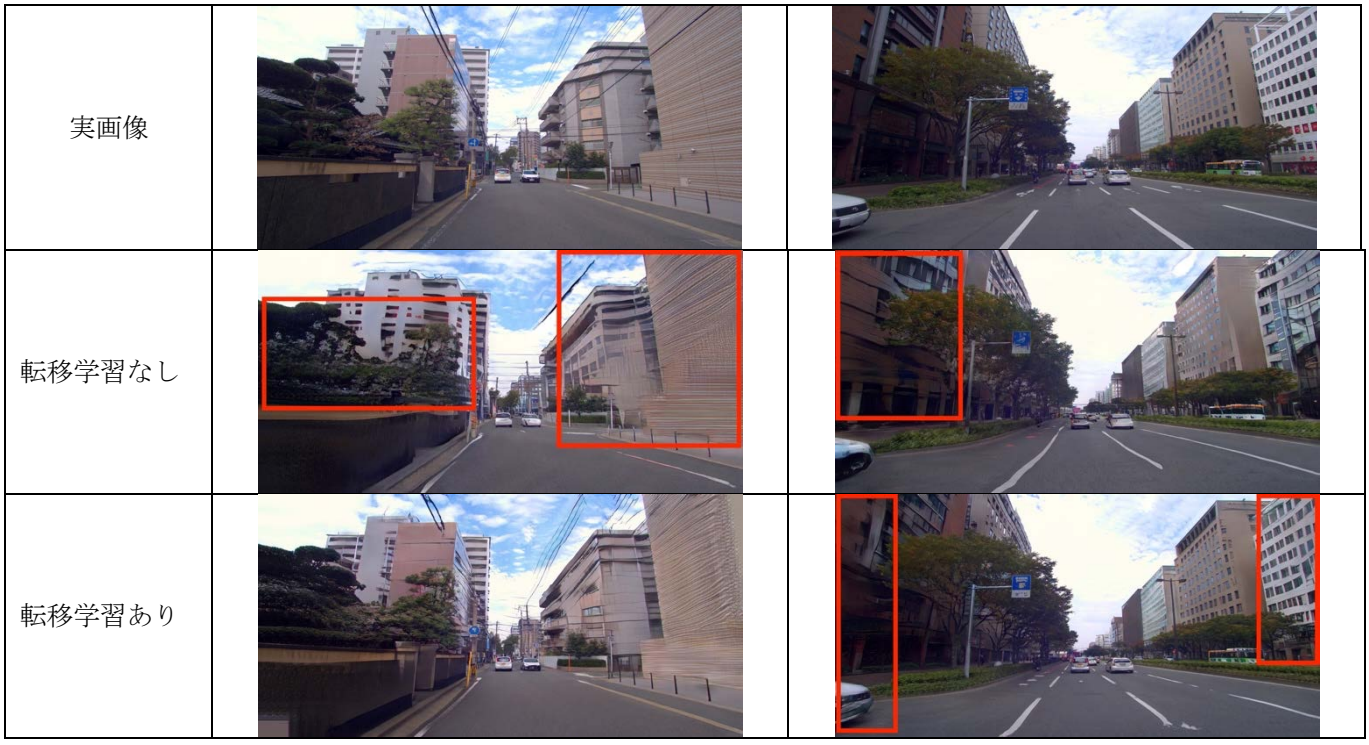
ここでTはマルチスケール識別器の層数，N_iは各層の解像度を示している．なお，マルチスケールの各層の解像度は$\frac{1}{T^2}$である．本研究では，$T = 3, N_1 = 1024 \times 2048$とした．

なお，G, D 共に最適化アルゴリズムには Adam を使用し，$\beta_1 = 0,\ \beta_2 = 0.999$，学習率$10^{-4}$と設定した．なお，イテレーション数は$10^9$，エポック数は$200$である．

２．２．転移学習

転移学習は，ソースドメイン$\mathcal{D}_S$，ターゲットドメイン$\mathcal{D}_T$が与えられたときに，$\mathcal{D}_S$の学習によって得られたパラメータを$\mathcal{D}_T$のパラメータの初期値として設定するこ

キーワード　機械学習，深層学習，データサイエンス，GAN，pix2pixHD，転移学習
連絡先　〒819-0395 福岡県福岡市西区元岡 744 九州大学 構造解析学研究室 TEL: 092-802-3370



とで，$\mathcal{D}_T$の学習改善を支援することを目的としている

$$\mathcal{D}_S = \{Y_S, X_S\}$$
$$\mathcal{D}_T = \{Y_T, X_T\} \quad (3)$$

ここで，X_S, X_Tはそれぞれソースドメイン，ターゲットドメインの実画像群Y_S, Y_Tはソースドメイン，ターゲットドメインのインスタンス画像群を示している．また以降は，転移学習と比較してターゲットドメインを学習することを本学習と呼ぶ．

３．例題設定

福岡県博多市の画像合成を目的とし，特に転移学習の有効性を議論した．

転移学習を行わないモデルでは，同地区の車載画像 1000 枚のみを使って pix2pixHD により本学習させた．一方で転移学習を行うモデルでは，citycapes[4)]として公開されているドイツ・フランクフルトの市街地の車載画像 3500 枚を転移学習した後，博多区の画像データのうち 3500 枚を更に転移学習させ出力した重みを初期値とし，同区の画像 1000 枚で本学習させた．上記 2 モデルを比較することで，転移学習の有効性を議論する．

４．結果

まずは転移学習なしモデル（表-1 中段）について述べる．左右列とも概ねの特徴を捉えられてはいるが，赤枠に示す部分で解釈不可能な部分が多く見られている．これらは単純な学習不足が原因であると考えられる．

次に，転移学習ありモデル（表-1 下段）について述べる．左右列とも全体の特徴を良く掴めており，自然な合成画像の再構築に成功した．しかし，右列赤枠内に歪が多少残されているため，より高精度な学習が必要となると考えている．

５．結言

本研究では，pix2pixHD と転移学習を併用した都市域への高解像度自動マッピングを検証した．特に，転移学習を用いた際の出力画像は，実画像と遜色がないほどの精度を出すことが可能となった．今度は，視点移動に伴う物理的な整合性を担保に Wang らが提案した vid2vid[5)]の導入を検討している．

参考文献

1) I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al.:Neural Information Processing Systems (NIPS), 2014.
2) Ting-Chun. Wang, Ming-Yu. Liu, Jun-Yan. Zhu, et al.: High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
3) Tan C., Sun F., Kong T., et al.: A Survey on Deep Transfer Learning. In Artificial Neural Networks and Machine Learning (ICANN) 2018.
4) M. Cordts, et al: The Cityscapes dataset for semantic urban scene understanding, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
5) T. Wang, M. Liu, J. Zhu, et al: Video-to-Video Synthesis, arXiv preprint, arXiv:1808.06601, 2018.