画像変換による背景イラストの昼画像から夜画像の生成

持田　海希†　　白井　匡人††

† 島根大学 自然科学研究科　〒 690-8504 島根県松江市西川津町 1060
†† 島根大学 学術研究院理工学系　〒 690-8504 島根県松江市西川津町 1060
E-mail: †n22m112@matsu.shimane-u.ac.jp, ††shirai@cis.shimane-u.ac.jp

あらまし　ゲーム制作において素材として背景イラストが用いられることがある．背景イラストを使用する場合，ゲーム内での時間経過により，昼画像だけでなく同じ背景の夜画像が必要となる．夜画像の差分を作成する場合，イラストを始めから描画するのに比べると楽ではあるが，作成にコストがかかる．このため，画像投稿サイト等でゲーム制作に使用可能な背景イラストは多数存在しているが，夜画像の差分まで用意されているものはほとんどない．そこで本研究では，画像変換に基づいた昼画像から夜画像への変換手法を提案する．先行研究による変換では，オブジェクトの形状を保持することが難しく，一部に靄のかかった画像が生成されてしまう問題がある．この問題を解決するために本研究では，昼画像とセグメンテーション画像の 2 つを入力として，夜画像の生成を行う．評価実験により提案手法の有効性を示し，今後の課題について検討する．

キーワード　GAN，画像変換，pix2pix，昼夜変換，ペア画像

1 はじめに

近年，家庭用ゲーム機を必要としない，スーマートフォンや PC 向けのダウンロードゲームの需要が高まり，企業だけでなく個人でのゲーム制作が増えている．また，プログラミングすることなくゲームを作成可能なツールなどもあり，ゲーム開発に取り組むハードルは下がっている．しかし，開発の際に問題となるのがゲームに利用する素材であり，背景イラストがその例に挙げられる．

背景イラストを使用する場合，ゲーム内での時間が経過することにより，昼画像だけでなく同じ背景の夜画像が必要となる．昼画像に対して夜差分の画像を作成する場合，イラストを始めから描画するのに比べると楽ではあるが，作成にコストがかかる．特に市街地のイラストは作画コストが自然背景に比べると高く，ただ暗くするだけでなく建物の色調変換や，街灯や窓などの光源を光らせる作業を行うことがある．このため，画像投稿サイト等でゲーム制作に使用可能な背景イラストは多数存在しているが，夜画像の差分まで用意されているものはほとんどない．

本研究では，昼画像から夜差分を生成し，作画コストを削減することを目的とする．画像変換に基づいた昼画像から夜画像への変換手法を提案する．まず，昼画像を建物，窓など，それぞれのラベルに分類したセマンティックセグメンテーション画像を作成する．次に，昼画像を暗くするために暗い青の画像を乗算で合成する．pix2pix[1] をベースとした 2 入力に対応したモデルを作成し，セグメンテーション画像と乗算画像の 2 つを入力することで，夜画像を自動的に生成することを可能となる．

2 関連研究

2.1 ペア画像を用いた手法

ペア画像とは学習に用いるデータセットの内，入力データと教師データなど 2 枚の画像間に関係があるものである．ペア画像の例として，ドイツを中心に 50 の都市の道路を専用車から撮影した Cityscapse[8] データセットが挙げられる．Cityscapse は道路の画像と，画像全体を空，車，建物，人など 30 種類のラベルにごとに色分けしたセマンティックセグメンテーション画像のペアとなっている．

ペア画像を用いた画像変換手法の例として，pix2pix[1] が挙げられる．pix2pix は GAN[5] から派生した画像変換手法である．ペア画像から画像間の関係を学習し，変換したい画像を入力すると，関係を考慮したうえで別の画像へと変換を行う．pix2pix はペア画像があれば，その対応関係を学習することができることから，航空写真からマップの作成，自動着色，セマンティックセグメンテーションなど，その用途は多岐にわたる．しかし，少ないデータセットで多くの対応関係を学習することは難しく，また，生成される画像が 256 × 256 と低解像度となってしまう点が問題である．

そこで，高解像度な画像に対応した pix2pixHD[2] という手法がある．pix2pixHD でも pix2pix と同様に，ペア画像から画像間の関係を学習し，入力画像から画像関係を考慮した対となる画像を生成する．pix2pix と違う点として，解像度の異なる 2 つの生成器をもつため，高解像度な画像が生成可能となっている．そのため，pix2pixHD ではより高解像度な 2048x1024 までの，画像変換が行える．

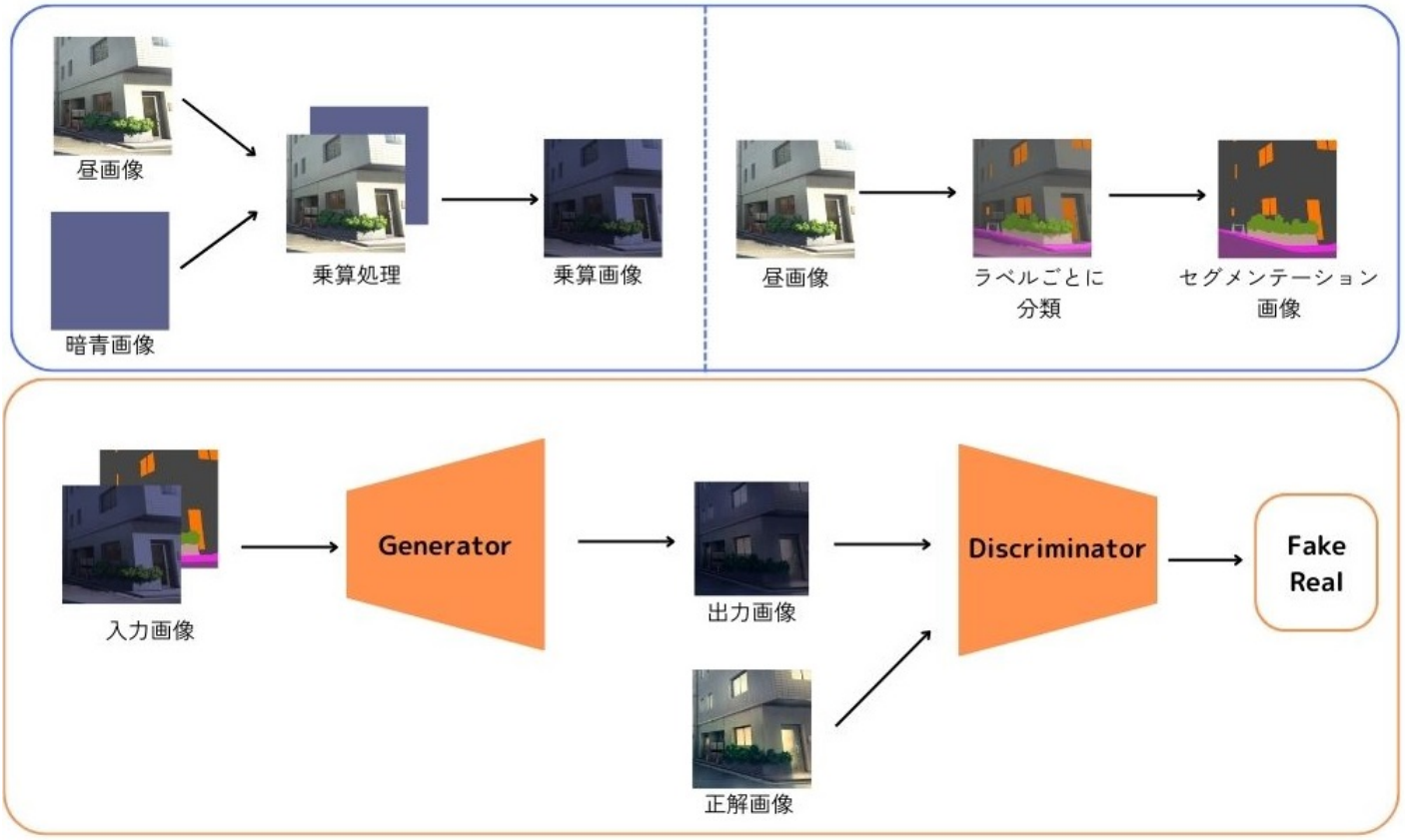

図 1　提案手法の流れ．昼画像と暗青画像から乗算画像，昼画像からセグメンテーション画像を作成する．乗算画像とセグメンテーション画像の 2 つから夜画像を生成器で生成し，識別器で本物か偽物か判定する．

2.2　ペアでない画像を用いた手法

データセットとしてペア画像が用意できない場合も考えられる．そのためペアでない画像をデータセットとして学習可能な手法も提案されている．その代表的な例として CycleGAN[6] が挙げられる．CycleGAN はデータセットから外見的特徴を捉え，入力した画像を特徴に対応させて変換する手法であり，馬からシマウマへの変換や，夏の風景から冬の風景への変換などが行われる．CycleGAN では 2 組の生成器と識別器をもつため，一方向の変換だけでなく，逆方向の変換が可能となっている．しかし，サイクル変換が可能であるがために，多くのメモリを使用し，学習にかかる時間が長いという問題が挙げられる．

そこで，その問題を解決する新たな手法として CUT[3] が提案されている．CUT は画像内の一部で入力画像に似るように変換を行い，別の一部では，対照学習を用いて入力画像とは似ないように変換を行う．これにより，入力画像の輪郭や背景が大きく改変されることなく，自然なスタイル変換を行える．また，生成器が 1 つであるため，モデルが軽量化され，短い時間で学習が可能である．

3　提 案 手 法

本研究では，昼画像にヒントとなるような画像を加えたうえで夜画像への変換を目指す．昼画像に暗く青い画像を乗算し，セグメンテーション画像と乗算画像を入力として夜画像の生成を行う．このように，昼画像から夜画像への一度での変換ではなく，段階的な変換手法を提案する．本手法のおおまかな流れを図 1 に示す．

3.1　セグメンテーション

pix2pix を用いて主に空，建物，道路，その他で図 2 のように分類する．ラベルは Cityscapes データセットと同様に設定する．Cityscapes では，光源である窓や，街灯はラベルに分類されておらず建物の一部となっている．本研究では光源を光らせる必要があるため，新しく窓ラベルを追加し，識別色をオレンジ色 (R，G，B：255，117，0) に設定する．

3.2　乗　　算

昼画像から夜画像へ変換するために，画像全体を暗くする必要があるため，昼画像に図 3 のような暗い青色 (R，G，B：93，97，142) を乗算で合成する．この乗算に用いる画像の色を変更することにより，生成される夜画像の明るさを変更することが可能になる．

3.3　2 入力アーキテクチャ

pix2pix をベースラインとした，2 つの入力画像から 1 つの生成画像を出力するようモデルの設計を行う．乗算画像を変換元の基本の画像として扱いセグメンテーション画像と乗算画像

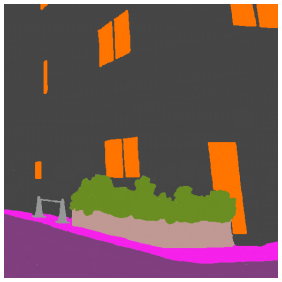

図 2　セグメンテーション画像

図 3　暗い青色の画像

図 4　昼　画　像

図 5　夜　画　像

の 2 つの入力から生成器で出力画像を生成する．

生成器では，U-Net[7] のような Encoder-Decoder 構造を取る．Encoder 部分で畳み込みを行い，徐々にボトルネック層までダウンサンプリングしていく．Decoder 部分で前層のデータと，それに対応した同サイズの Encoder のデータをもとにアップサンプリングを行う．このように，Decoder で同サイズの Encoder の出力を用いる手法をスキップ接続と呼ぶ．

識別器には生成器で生成された画像か，正解画像のどちらかが入力される．ここで画像は N × N サイズのパッチごとに分割し，入力される．識別器では分割されたパッチ単位で本物か偽物かの分類を行い，すべての判定を平均化したものが出力となる．このようにパッチごとに分割し，判定を行うものを PatchGAN と呼ぶ．

GAN の損失関数には pix2pix と同様に絶対値の足し算となる L1 損失を用いる．また，目的関数は生成器を G，識別器を D，L1 損失を $\mathcal{L}_{L1}$ とすると，

$$G^* = arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G) \qquad (1)$$

のように表される．ここで λ は L1 損失の重みを示すパラメータとなる．生成器は識別器に本物と間違えさせるような画像生成をし，識別器は正確に本物偽物の判定を行えるよう，互いに敵対的に学習を行う．

4　実　　験

4.1　データセット

33 枚の背景イラストの画像と，46 枚の Cityscapes データセットの計 79 枚の昼画像を使用する．また，それに対応させた乗算画像，セグメンテーション画像と夜画像を用意し，これら 79 枚，3 セットを学習データセットとする．夜画像では建物の部分は赤紫色に色調補正され，窓の部分で光源処理が行われている．学習データセットのもととなる昼画像と夜画像をそれぞれ，図 4 と図 5 に示す．

テストデータとして 5 枚の背景イラストの昼画像を使用する．背景イラストは，みんちりえ [9] の画像を 512 × 512 のサイズで切り取り作成する．

提案手法について，入力画像として乗算画像とセグメンテーション画像を用意する必要がある．そこで，テスト時に使用するセグメンテーション画像について，昼画像の建物部分を灰色 (R，G，B：70，70，70) で塗り，光源となる窓部分をオレンジ色 (R，G，B：255，117，0) で塗った画像を作成する．

4.2　比 較 手 法

比較手法として，pix2pix，pix2pixHD，CUT，DCLGAN[4] を用いる．各比較手法に乗算画像と夜画像から画像間の関係を学習させ，乗算画像を入力して出力画像を生成する．それぞれ 200 エポックで学習を行う．

4.3　評 価 方 法

評価方法としてピクセルでの評価と Frechet Inception Distance(FID)[10] での評価を行う．ピクセルでの評価では，各モデルで出力された画像と正解画像との RGB 値の平均二乗誤差の総和を 3 で割り，ピクセル数で割った値で評価する．これにより，画像全域での正解画像との類似度を測り，スコアが低いほど評価が高い．

FID は生成画像と正解画像の特徴距離を測定する手法であり 2 つの画像の分布間の距離を計算する．スコアが低いほど評価が高い．

本研究と既存手法に対し，正解の夜画像がある昼画像を入力し，出力された画像を各評価指標に沿って計算し比較する．評価計算時には，画像を 256 × 256 で入力する．

表 1　各手法のピクセルスコア

手法	ピクセルスコア
提案手法	339.142
pix2pix	1417.827
pix2pixHD	1297.307
CUT	2341.193
DCLGAN	2445.769

表 2　各手法の FID スコア

手法	FID スコア
提案手法	72.121
pix2pix	158.517
pix2pixHD	138.688
CUT	180.932
DCLGAN	190.76

4.4　実 験 結 果

提案手法と比較手法におけるピクセルスコアを表 1 に，FID スコアの計算結果を表 2 に示す．ピクセルスコアと FID スコアともに提案手法が最も低く，次いで pix2pixHD，pix2pix，CUT，DCLGAN となっている．ピクセルスコアでペア画像を用いた手法である pix2pix とペアでない画像を用いた手法の CUT との間で大きく差がある．

4.5　考　　察

提案手法と比較手法における出力結果の画像が図 6 である．

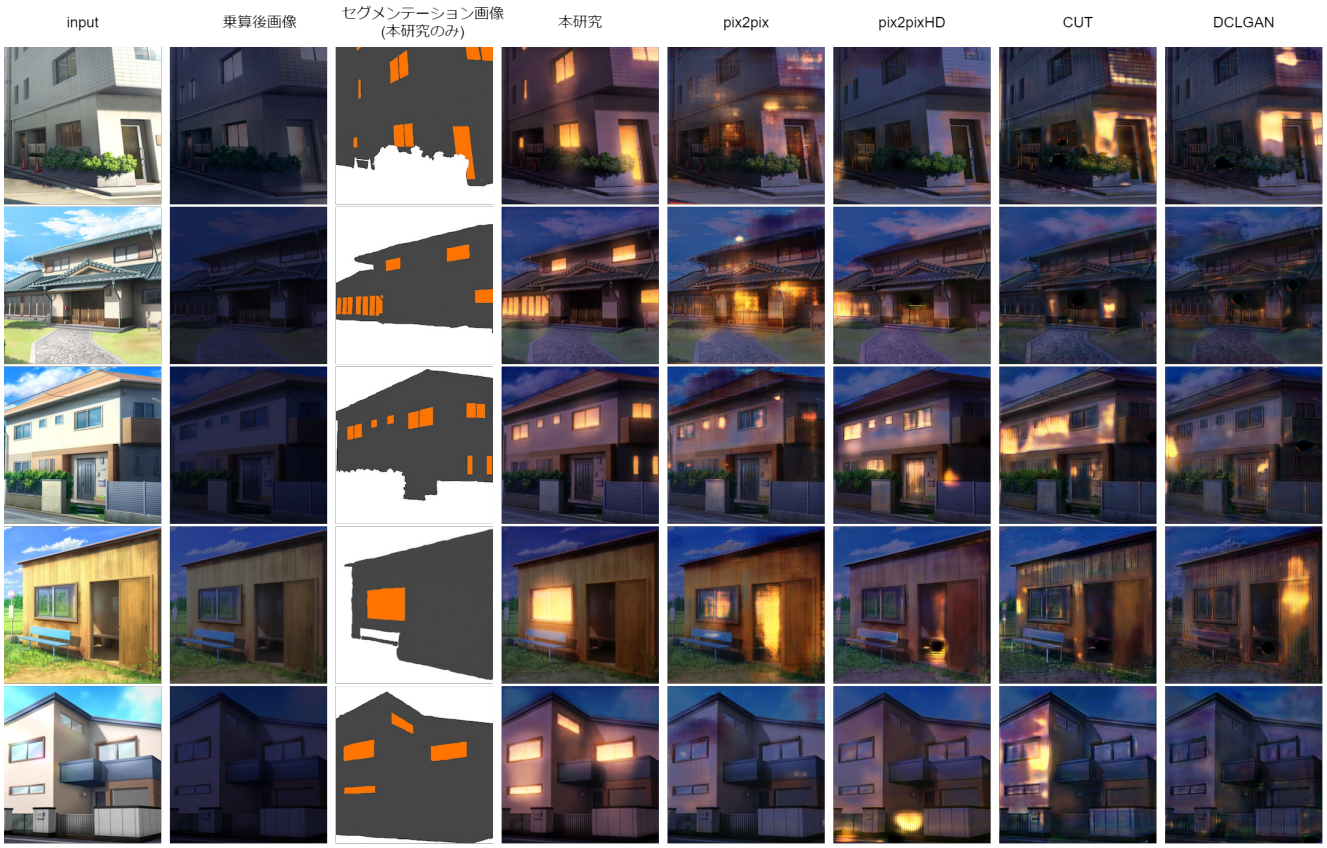

図 6　提案手法と比較手法における出力結果．input が入力画像であり，提案手法のみセグメンテーション画像も入力する．一番上の画像を 1 枚目とし，下に 2，3，4，5 とする．

比較手法では建物の色調変換が部分的であり，光源処理については本来窓でない部分を光らせるような処理が見られる．3 枚目の画像については窓の位置を比較的予測できているが，5 枚目の画像ではどの手法も窓の位置を予測できていない．

提案手法ではセグメンテーション画像の建物部分で色調変換が行われ，窓部分で光源処理が行われている．このことから，画像変換を行う際にセグメンテーション画像がどの部分で変換を行うべきかのヒントとして有効であると考えられる．また，セグメンテーション画像を用いることにより 79 枚という少ないデータセットでも学習が可能である．

定量的評価ではピクセルスコア，FID スコアともに最も低い数値となっているため，提案手法の有効性が示されているといえる．次いで pix2pixHD，pix2pix となっているため，少ないデータセットで画像変換する場合はペア画像を用いる手法が有効と考えられる．ペア画像を用いない手法で数値が高かった理由として本来窓のない場所に広い範囲で光源処理を行っていることが挙げられる．

5　結　　論

本研究では，昼画像から夜画像に変換する際に，セグメンテーション画像をヒントとして用いることの有効性を示した．また，セグメンテーション画像により，少ないデータセットでも精度の高い画像が生成可能である．しかし，昼画像からセグメンテーション画像の生成をすることができておらず，昼画像から夜画像への完全な自動変換には至らなかった．現時点でも，本来の目的である作画コストの削減は図れるが，さらなる効率化や，生成される画像が低画質である点など課題解決が求められる．

文　　献

[1] Isola, Phillip, et al. ”Image-to-image translation with conditional adversarial networks.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[2] Wang, Ting-Chun, et al. ”High-resolution image synthesis and semantic manipulation with conditional gans.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[3] Park, Taesung, et al. ”Contrastive learning for unpaired image-to-image translation.” European conference on computer vision. Springer, Cham, 2020.

[4] Han, Junlin, et al. ”Dual contrastive learning for unsupervised image-to-image translation.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[5] Goodfellow, Ian, et al. ”Generative adversarial networks.” Communications of the ACM 63.11 (2020): 139-144.

[6] Zhu, Jun-Yan, et al. ”Unpaired image-to-image translation using cycle-consistent adversarial networks.” Proceedings of the IEEE international conference on computer vision. 2017.

[7] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. ”U-net: Convolutional networks for biomedical image segmentation.” International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.

[8] Cordts, Marius, et al. ”The cityscapes dataset for semantic

urban scene understanding." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[9] https://min-chi.material.jp/

[10] Heusel, Martin, et al. "Gans trained by a two time-scale update rule converge to a local nash equilibrium." Advances in neural information processing systems 30 (2017).

[11] Zheng, Chuanxia, Tat-Jen Cham, and Jianfei Cai. "The spatially-correlative loss for various image translation tasks." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[12] Lee, Hsin-Ying, et al. "Diverse image-to-image translation via disentangled representations." Proceedings of the European conference on computer vision (ECCV). 2018.

[13] Zhu, Jun-Yan, et al. "Toward multimodal image-to-image translation." Advances in neural information processing systems 30 (2017).

[14] Liu, Rui, et al. "Divco: Diverse conditional image synthesis via contrastive generative adversarial network." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.