卒 業 論 文 概 要 書

Summary of Bachelor's Thesis

Date of submission: (02/02/2020)

学科名 Department	情報通信	氏名 Name	森 勇綺	指導教員 Advisor	渡辺 裕 印
研究指導名 Research guidance	オーディオビジュアル情報処理研究	学籍番号 Student ID number	1w172324-5 CD		
研究題目 Title	Shape Matching GAN と ST-GAN を用いた 2 画像の写実的合成 Photorealistic Image Synthesis by Using Shape Matching GAN and ST-GAN				

1. まえがき

近年, 画像や映像による媒体が増えていく中で, 画像の合成技術がビジネスの場でも多く使われている. 広告や不動産のイメージ画像などを作成する際には, 写真画像を合成することで写真撮影の省略を行う. また自動車メーカーにおいては新車開発におけるモックアップ製作などを省く事で, 大幅なコスト削減や期間短縮を実現している[1]. さらに, 故障・崩壊を未然に防止するために重要な, 製品や構造物の亀裂検出技術の改善などにも用いられている[2].

本研究では, より写実的に画像を合成する技術を実現するため, 背景画像に合わせた前景画像のスタイル変換と, Spatial-transformer-networks (STN) を用いた空間変換の二つの機構を合わせた手法を検討する. また, 本研究では前景画像としてテキスト画像を扱うことで, 道路の路面標識や石像などに彫り込まれる文字の再現を可能にすることを目的とする.

2. 関連技術

2.1. ST-GAN

ST-GAN は, Lin らが提案した, STN を利用した空間変換を反復的に学習させることによって, より自然に見えるように前景画像を幾何学的補正する手法である[3]. STN を用いて前景画像を空間変換し, その前景画像の情報をパラメータ化する. その後, そのパラメータをさらに STN に入力して反復的な学習を行う. これにより背景に自然に合成できる形での空間変換を行う.

2.2. Shape Matching GAN

Shape Matching GAN は, Yang らが提案したグリフのスタイルの程度を制御できるスタイル変換の手法である[4]. パラメータ$l \in [0,1]$で制御された異なる変形度のもとで, 芸術的なテキストをレンダリングするためのフィードフォワード型モデルGを設計する. さらに, 伝達プロセスを構造伝達とテクスチャ伝達の二つの連続した段階に分解して, G_SとG_Tの別々の形にモデル化する. これにより, テクスチャの影響を切り離し重要な形状変形問題に焦点を当てる.

3. 提案手法

3.1. Shape Matching GAN と ST-GAN の組み合わせによる写実的合成

本節では, ST-GAN と Shape Matching GAN を組み合わせ, 写実的に合成する手法を提案する. 前景オブジェクトとなるテキスト画像を, Shape Matching GAN でスタイル変換を行う. その後, スタイル変換された前景オブジェクトを ST-GAN を用いて空間変換し, 合成画像を生成する. この手法の概要を図 1 に示す.

前景オブジェクトのスタイル変換を行う上で, スタイル画像として背景画像に近い質感を持った実画像を用いる.

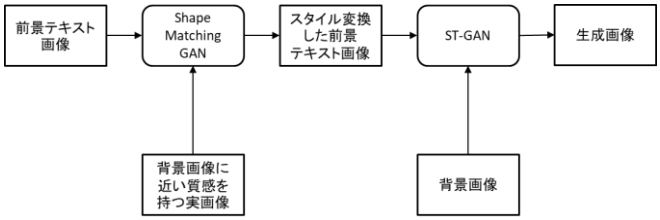

図 1　提案手法 1 の概要図

3.2. 背景画像からスタイル抽出も行う写実的合成

3.1 節で提案した手法では, スタイル変換を行う際, 背景画像に近い質感を持つ実画像を利用した. これは, 背景画像からスタイルも抽出しようとした場合, 合成領域が確保できない場合があるためである. また, 形状マッチングを行う上で必要なスタイル領域が不足してしまっている可能性も考えられる. しかし, スタイルの抽出を背景画像とは別の画像にする事で, スタイル変換の質が落ちてしまう可能性も考えられる. 本節ではスタイル変換を行うための質感

も併せ持つ画像を背景画像として設定し，同一の画像でスタイル変換，空間変換の両方を行う手法を提案する．この手法の概要を図 2 に示す．

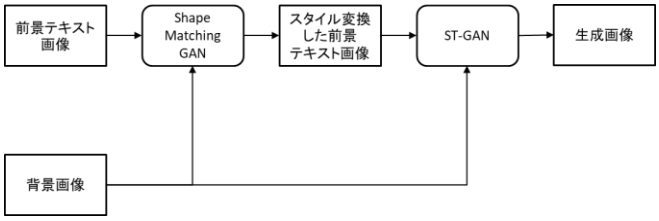

図 2　提案手法 2 の概要図

4. 実験

4.1. スタイル画像に背景画像とは別の実画像を用いた写実的合成

入力画像には, スタイル領域を抽出し，α チャンネルを削除したスタイル画像のマスク画像を追加で作成し，スタイル変換を行う．さらに，ST-GAN では，前景オブジェクトの入力として，生成されたスタイル変換後の前景画像のマスクを追加作成する．その後, 変換および合成を行う．本手法での画像の生成結果を以下の図 3 に示す．また，今回 ST-GAN は 3 次元情報を得るため, SUNCG dataset[5]で学習済みのモデルを用いた．

図 3　手法 1 での生成結果(左: 成功例，右: 失敗例)

4.2. 背景画像からスタイル抽出も行う写実的合成

ST-GAN における空間変換および合成の機構では，入力画像として 4.1 節で述べたものに加え，背景画像のスタイル領域のマスク画像および背景画像からスタイル領域を除いた画像を追加で作成する．その後, 変換および合成を行う．本手法での生成結果を以下の図 4 に示す．

図 4　手法 2 での生成結果(左: 成功例，右: 失敗例)

4.3. 評価実験

手法 1 と手法 2 の実験について，生成結果の自然さを評価するための主観評価実験を行った．それぞれの手法における生成結果を 5 枚ずつ用意し，50 人に対して実施した．生成結果の合成画像を見て，「テキストと背景が自然に合成できているか」という問いに対して 5 段階 (5: とても自然, 4: 少し自然, 3: どちらとも言えない, 2: 違和感が強い, 1: とても違和感が強い）で評価した．その 5 段階での評価の平均値を以下の表 1 に示す．

表 1　評価実験の結果

手法 1	手法 2
2.332	2.72

これより，手法 1 より手法 2 の方が評価の平均値が 0.388 ポイント高く，より自然な合成結果を得ることが分かる．

5. むすび

本論文では，より写実的に画像を合成するために, Shape Matching GAN と ST-GAN を用いて，前景オブジェクトのスタイル変換と空間変換の両方を行う手法を提案した．

実験結果より, スタイル画像にも同一の背景画像を用いた場合のほうが良い結果が得られることが分かった．しかし，背景画像と大きくずれる空間変換がされた生成結果も得られた．

参考文献

[1] 長尾健作: “CAD データから写真画像を合成する技術とビジネス -CG 技術によるデジタル画像制作ビジネスの実践-”，公益社団法人日本印刷技術協会 テキスト＆グラフィック研究会 tech Seminar, June. 2009.

[2] 富山真一, 大平倫宏: “画像合成を用いた亀裂検出システムの開発(ディペンダブルコンピューティング)”, 電子情報通信学会技術研究報告: 信学技報 113(353), p41-44, December. 2013.

[3] Chen-Hsuan Lin, Ersin Yumer, Oliver Wang, Eli Shechtman, and Simon Lucey, “ST-GAN: Spatial Transformer Generative Adversarial Networks for Image Compositing,” arXiv preprint arXiv:1803.01837v1, Mar. 2018.

[4] Shuai Yang, Zhangyang Wang, Zhaowen Wang, Ning Xu, Jiaying Liu, and Zongming Guo, “Controllable Artistic Text Style Transfer via Shape-Matching GAN,” arXiv preprint arXiv:1905.01354v2, Aug. 2019.

[5] Shuran Song, Fisher Yu, Andy Zeng, Angel X Chang, Manolis Savva, and Thomas Funkhouser, “Semantic Scene Completion from a Single Depth Image,” arXiv preprint arXiv:1611.08974v1, Nov. 2016.

2020 年度　卒業論文

Shape Matching GAN と ST-GAN を用いた 2 画像の写実的合成

Photorealistic Image Synthesis by Using Shape Matching GAN and ST-GAN

提出日　2020 年 2 月 2 日

指導教員　渡辺 裕 教授

早稲田大学 基幹理工学部 情報通信学科

1W172324-5

森 勇綺

目次

第1章　序論

1.1　研究背景

近年，画像や映像による媒体が増えていく中で，画像の合成技術がビジネスの場でも多く使われている．広告や不動産のイメージ画像などを作成する際には，写真画像を合成することで写真撮影の省略を行う．また自動車メーカーにおいては新車開発におけるモックアップ製作などを省く事で，大幅なコスト削減や期間短縮を実現している[1]．さらに，故障・崩壊を未然に防止するために重要な，製品や構造物の亀裂検出技術の改善などにも用いられている[2]．これは，複数の画像から 1 枚のパノラマ画像を作成する画像合成技術を活用している．本研究では，これらの様々な技術のさらなる向上のため，画像の合成技術の向上の手法を提案する．

1.2 研究目的

本研究では，より写実的に画像を合成する技術を実現するため，背景画像に合わせた前景画像のスタイル変換と，Spatial-transformer-networks（STN）を用いた空間変換の二つの機構を合わせた手法を検討する．また，本研究では前景画像としてテキスト画像を扱うことで，道路の路面標識や石像などに彫り込まれる文字の再現を可能にすることを目的とする．

1.3 関連研究

1.3.1 Spatial Fusion GAN for Image Synthesis

Zhan らは，幾何学的合成と外観的合成を実現する二つの機構を合わせる事で，より写実的に画像を合成する手法を提案している[3]．まず，幾何学的合成を実現するための機構において，背景画像に合わせるように STN で前景画像を幾何学変換して合成する．その後，外観的合成用の機構において，先ほど作成した合成画像に近い形状を持った実画像を入力して要素を抽出する．これにより写実的に前景画像を変換することを実現している．また，外観的合成の機構においてガイドフィルタという前景画像のエッジを保持するためのフィルタを利用することで，合成した際の前景画像のぼやけを防ぐ．

1.4 本論文の構成

本論文の構成を以下に示す．
第 1 章は本章であり，本研究の背景，目的，関連研究について述べる．
第 2 章では本研究で用いる関連技術ついて述べる．

第 3 章では本研究で提案する手法について述べる.
第 4 章では実験結果および考察について述べる.
第 5 章では本論文の結論と今後の課題について述べる.

第2章 関連技術

2.1 まえがき

本章では，本研究の関連技術について詳細に述べる．まず，ニューラルネットワーク，畳み込みニューラルネットワークである Convolutional Neural Networks（CNN）および CNN をより頑健性の高いモデルにするためのモジュールである STN について述べる．また，本研究における空間変換の点で参照した Spatial Transformer GAN およびスタイル変換の点で参照した Shape Matching GAN について述べる．

2.2 ニューラルネットワーク

ニューラルネットワークとは，複数のノードをつなぎ合わせた入力層，中間層および出力層から構成されるネットワークである[4]．ニューラルネットワークの構造例を図 2.1 に示す．ニューラルネットワークは複数の入力と出力を持つ．また，各ノードは値，ノード間を結ぶ各エッジは重みを持つ．入力データに対して，重みパラメータを最適化していく事で学習を行う．

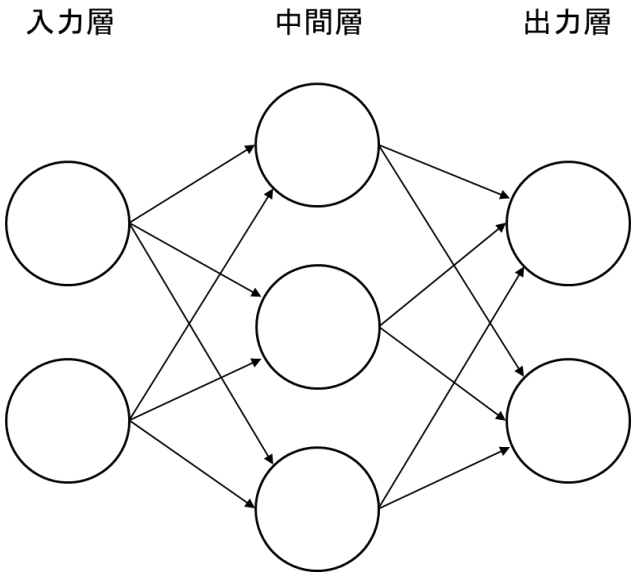

図 2.1　ニューラルネットワークの構造例

2.3 CNN

CNN とは，画像認識などによく用いられるニューラルネットワークの構造で，何段もの深い層を持つ[4][5][6]．構造は入力層，畳み込み層，プーリング層，全結合層および出力層からなる．CNN の構造例を図 2.2 に示す．

畳み込み層での処理は，畳み込み演算である．元の画像に一定間隔でスライドさせながらフィルタをかけて特徴マップを抽出する．その結果を出力先へ格納する．

プーリング層では特徴の情報を残しながら元の画像を縮小する処理を行う．畳み込み演算によって得られた結果を小領域に分割する．その中から最大のものを選択し，画像の情報を圧縮する．これにより画像の移動や回転など微小な位置変化に対して頑健性を持つ．さらに計算コストも下げることが可能となる．

畳み込み層では，処理が膨大にならないよう全てのノードを結合しない．一方，全結合層においてはノード間をすべて結合する．畳み込み層とプーリング層を通して圧縮された画像データの特徴部分の情報を，全結合層に渡す．その結果，大幅に計算を削減し，全結合計算を行った結果を出力層へ出力する事が可能となる．

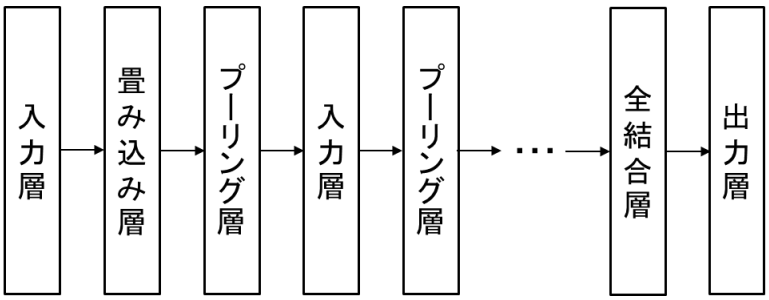

図 2.2　CNN の構造例

2.4 STN

STN は，Jaderberg らによって提案された空間変換の手法である[7]．従来の CNN では浅い層での画像の大きな空間的変換に対して不変性を持たない．そのため，STN を用いてニューラルネットワークに空間的な歪みを解消する能力を与える．これにより空間的普遍性を解消する．

STN はローカリゼーションネットワーク，グリッドジェネレータおよびイメージサンプラーの三つの機構からなる．STN の概要を図 2.3 に示す．

ローカリゼーションネットワークは，幅W，高さH，チャンネルCを持つ入力特徴空間$U \in \mathbb{R}^{H\times W\times C}$から変数パラメータ$\theta$を出力するネットワークである．$\theta$を求めるための変換を式(2.1)に示す．

$$\theta = f_{loc}(U) \tag{2.1}$$

ローカリゼーションネットワークの関数$f_{loc}()$はθを生成するための最終回帰層を含むものであれば，層の種類は制限されない．

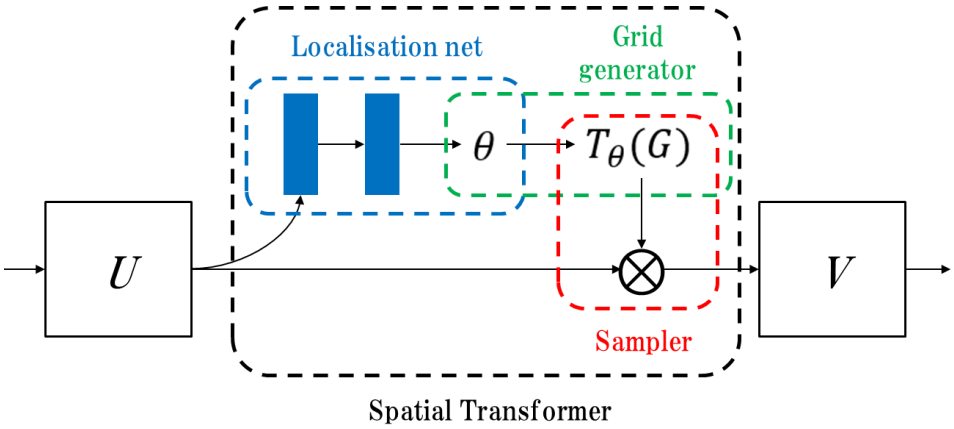

図 2.3　STN の概要図

グリッドジェネレータは，最終的な出力空間Vを生成するためのグリッドGを求める．出力は，$G_i = (x_i^t,\ y_i^t)$からなる規則的なグリッド$G = \{G_i\}$上にあるように定義されている．また，入力特徴空間の特定の位置$U_i = (x_i^s,\ y_i^s)$を中心としてサンプリングカーネルを適用して，出力特徴空間$V \in \mathbb{R}^{H' \times W' \times C}$を形成する．ここで，$H'$，$W'$はグリッド$G$の高さ，幅であり，$C$は入力，出力で同じチャンネル数である．サンプリング中心点を算出するための変換を式(2.2)に示す．

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = T_\theta(G_i) \tag{2.2}$$

パラメータθが微分可能な形であれば，変換T_θは平面射影変換やアフィン変換など様々な変換を用いることができる．θが微分可能であることで，サンプル点$T_\theta(G_i)$からローカリゼーションネットワークの出力θまで勾配を逆伝播させることができる．

イメージサンプラーはグリッドジェネレーターによって得たグリッドGを用いて，入力特徴空間Uとサンプリング点の集合$T_\theta(G)$から出力特徴空間Vを生成する．サンプリングを行うための一般式を式(2.3)に示す．

$$V_i^c = \sum_n^H \sum_m^W U_{nm}^c\, k(x_i^s - m; \Phi_x) k(y_i^s - n; \Phi_y), \forall i \in [1 \dots H'W'], \forall c \in [1 \dots C] \tag{2.3}$$

ここで，Φ_x，Φ_yはサンプリングカーネル$k()$のパラメータ，U_{nm}^cは入力特徴空間のチャンネルCでの点(n,m)の値，V_i^cはチャンネルCにおける点$(x_i^t,\ y_i^t)$の出力値である．

2.5 Spatial Transformer GAN (ST-GAN)

2.5.1 GAN

GAN は，Goodfellow らによって提案された，Generator と Discriminator の二つのネットワークを敵対的に学習する新しい画像の生成モデル推定手法である[8]．Generator と Discriminator を互いに競争させながら学習する事で，最終的に Discriminator に本物と識別されるような精度の高い生成結果を Generator で生成する．GAN の構造例を図 2.4 に示す．

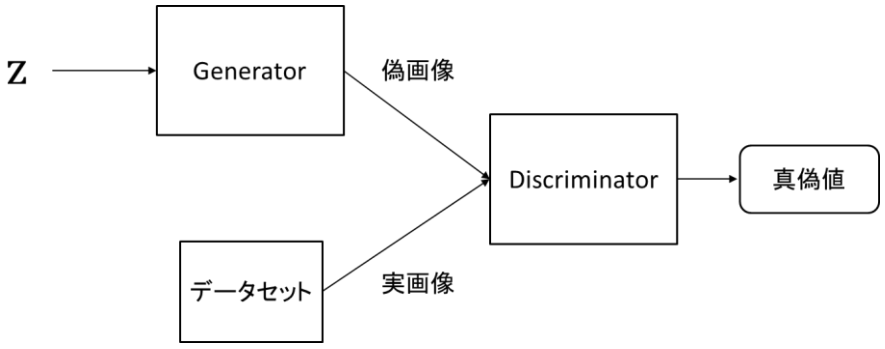

図 2.4　GAN の構造例

Generator は入力となる潜在変数zから，画像を出力するニューラルネットワークである．Discriminator は入力された画像がデータセット由来か，Generator によって生成された偽画像かを識別しその判定を出力するニューラルネットワークである．Generator は出力画像が Discriminator にデータセットからの実画像と判別させるように学習する．一方，Discriminator は実画像と偽画像の判別を正しく行えるように学習する．この学習を敵対的に進めることで，それぞれのネットワークの能力をより高精度のものにしていく．このフレームワークは Generator と Discriminator の二つにおけるミニマックス法を用いつ．その時の価値関数$V(G,D)$を求めるミニマックス法の式を式(2.4)に示す．

$$\min_G \max_D V(G,D) = \mathbb{E}_{x\sim p_{data}(x)}[logD(x)] + \mathbb{E}_{z\sim p_z(z)}[\log\big(1-D\big(G(z)\big)\big] \tag{2.4}$$

ここで，$p_{data}(x)$は入力データx上のデータセットの確率分布，$p_z(z)$は入力ノイズzの確率分布，$D(x)$はxがデータセットのものである確率，$G(z)$は Generator がzから生成したデータである．

2.5.2 ST-GAN

ST-GAN は，Lin らが提案した，STN を利用した空間変換を反復的に学習させることによって，より自然に見えるように前景画像を幾何学的補正する手法である[9]．STN を用いて前景画像を空間変換し，その前景画像の情報をパラメータ化する．その後，そのパラメータをさらに STN に入力して反復的な学習を行う．これにより背景に自然に合成できる形での空間変換を行う．

背景画像I_{BG}と前景オブジェクトI_{FG}の合成プロセスを式(2.5)に示す（I_{FG}は対応するマスクM_{FG}を含むものである）．

$$I_{comp}(p_0) = I_{FG}(p_0) \oplus I_{BG} \tag{2.5}$$

ここで，p_0はI_{FG}の初期ワープ状態での合成パラメータであり，前景オブジェクトの画像はワープパラメータの関数として記述される．この演算子を図 2.5(a)に示す．

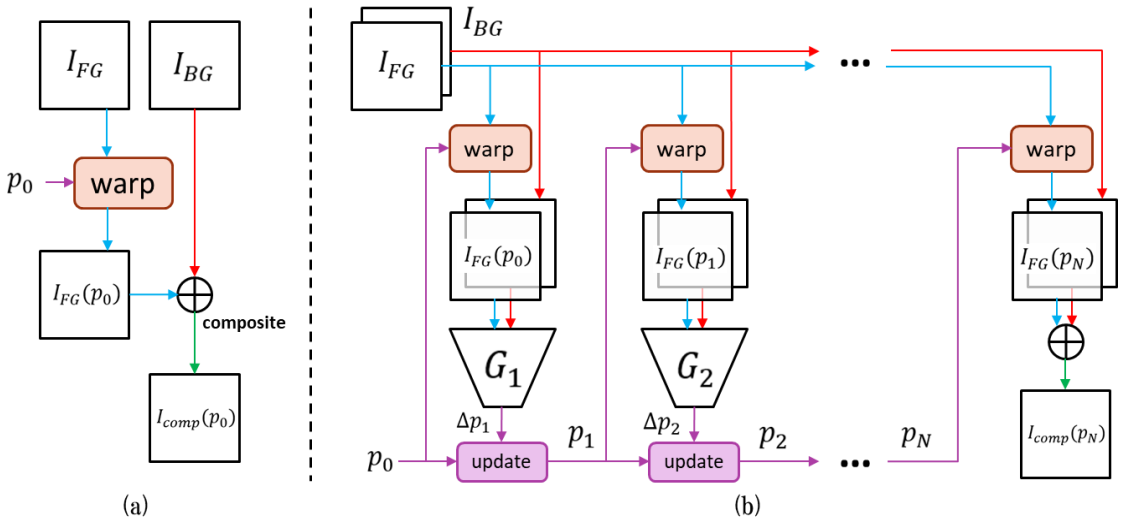

図 2.5　ST-GAN で用いられる機構の概要図
(a)前景オブジェクトを背景画像に合わせて空間変換させ，合成する演算子
(b)前景オブジェクト上の一連のワープ更新を予測するための反復 STN

画像ピクセルから大きな変位ワープパラメータを予測することは非常に困難である．そこで，図 2.5(a)に示した演算子に STN を適用させ，反復的に処理を行う．これにより局所的な幾何学的変換を予測する．背景画像に適合するように前景オブジェクトがどのように

変換されるべきかの相互作用の情報として，Generator で補正ワープ更新Δp_1を予測する．その後，図 2.5(b)に示すように一連の流れを反復的に行う．ここで，入力画像Iと前回のワープ状態p_{i-1}が与えられた時の補正ワープ更新Δp_iおよび新しいワープ状態p_iを以下の式(2.6)に示す．

$$\Delta p_i = G_i\left(I_{FG}(p_{i-1}),\ I_{BG}\right)$$
$$p_i = p_{i-1} \circ \Delta p_i \tag{2.6}$$

ここで，$G_i()$はワープパラメータの構成を表す．

STN は，画像を自然画像多様体に近づける幾何学的ワープを学習する．そこで，学習においては反復的 STN のための逐次敵対学習戦略を採用する．逐次敵対学習戦略の概要を図 2.6 に示す．単一のG_1を学習するところから始め，全ての前の生成器$\{G_j\}_{j=1\ldots i-1}$の重みを固定する．その後，後続の新しいG_iを追加および学習する．さらに，結果として得られたワープ状態$I_{comp}(p_i)$での合成画像をDに送り込み実データ分布と照合する．これにより，G_iとDのみを学習させる．

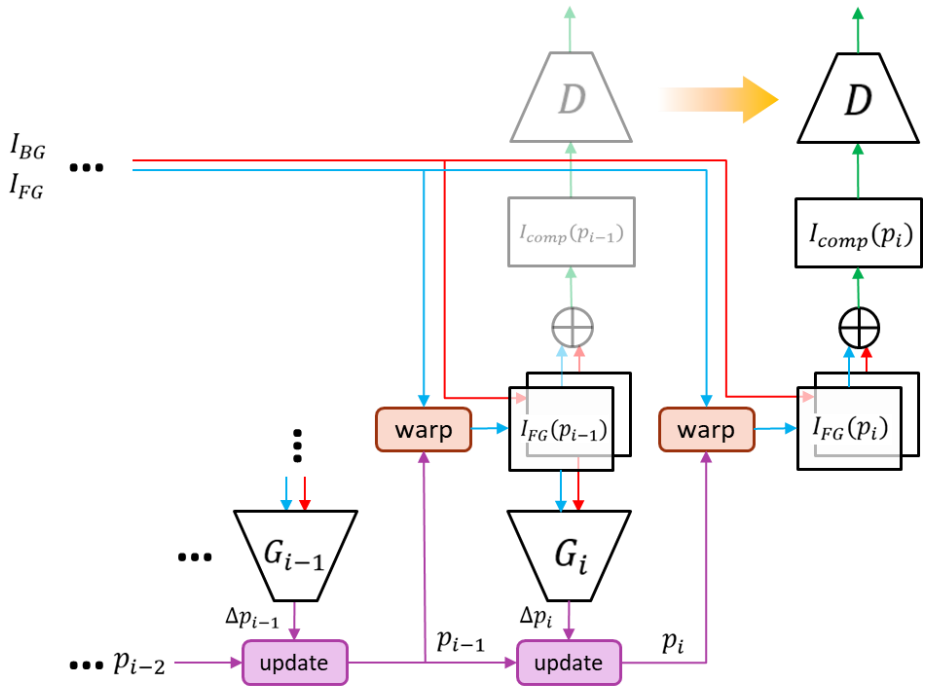

図 2.6　ST-GAN の逐次敵対学習

2.6 Shape Matching GAN

2.6.1 Shape Matching GAN の概要

Shape Matching GAN は，Yang らが提案したグリフのスタイルの程度を制御できるスタイル変換の手法である[10]．パラメータ$l \in [0,1]$で制御された異なる変形度のもとで，芸術的なテキストをレンダリングするためのフィードフォワード型モデルGを設計する．さらに，伝達プロセスを構造伝達とテクスチャ伝達の二つの連続した段階に分解して，G_SとG_Tの別々の形にモデル化する．これにより，テクスチャの影響を切り離し重要な形状変形問題に焦点を当てる．$G = G_T \circ G_S$とし，スタイル変換プロセスを定式化したものを以下の式(2.7)に示す．

$$I_l^Y = G_T\big(G_S(I,l)\big), I_l^Y \sim p(I_l^Y | I, Y, l), \tag{2.7}$$

ここで，変換された出力画像I_l^Yの目標統計量$p(I_l^Y)$は，テキスト画像I，スタイル画像Yおよび制御可能なパラメータlによって特徴づけられる．

構造伝達には双方向形状マッチングを用いる．スタイル画像の構造伝達の段階では，G_Sの学習ペアとなる$\{\tilde{X}_l, X\}$を得るためにXの前処理を行う．ここで，$\tilde{X}_l$はテキスト画像の形状特性を持つXの粗いスケッチであり，lは粗さのレベルを制御するパラメータである．また，テキスト画像の構造伝達の段階では，G_Sは$\{\tilde{X}_l, X\}$を学習して様々な変形度のグリフをスタイリングする．G_SとG_Tの二つの主要な構成要素からなる全体的なフレームワークをまとめたものを図 2.7 に示す．以下では，2.6.2 節でストラクチャー転送ネットワークG_Sおよびその学習用のモジュールG_Bを用いた双方向構造変換について述べる．続いて，2.6.3 節でテクスチャ転送ネットワークG_Tについて述べる．

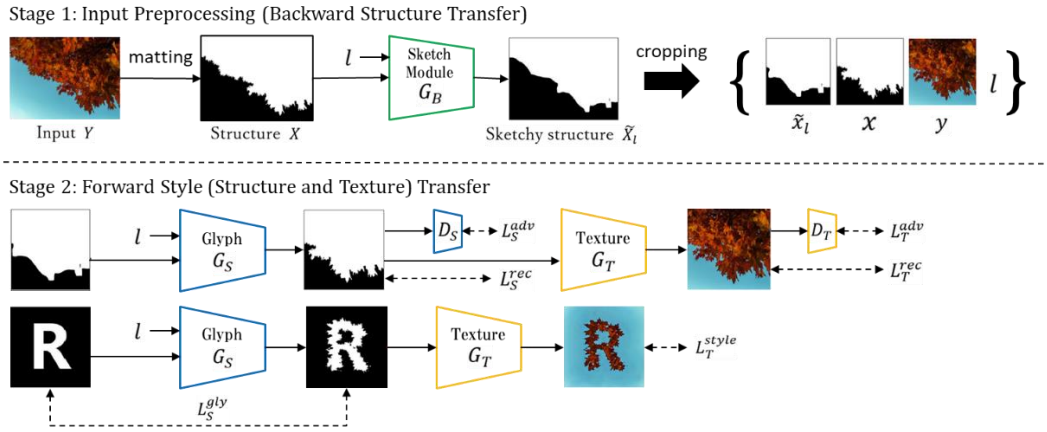

図 2.7　双方向形状マッチングフレームワークの概要図

2.6.2 Bidirectional Structure Transfer (G_S)

グリフの特性を異なる粗さのレベルでXに伝達するために，平滑化ブロックおよび変換ブロックからなるスケッチモジュールG_Bを用いる．G_Bの概要を図 2.8 に示す．

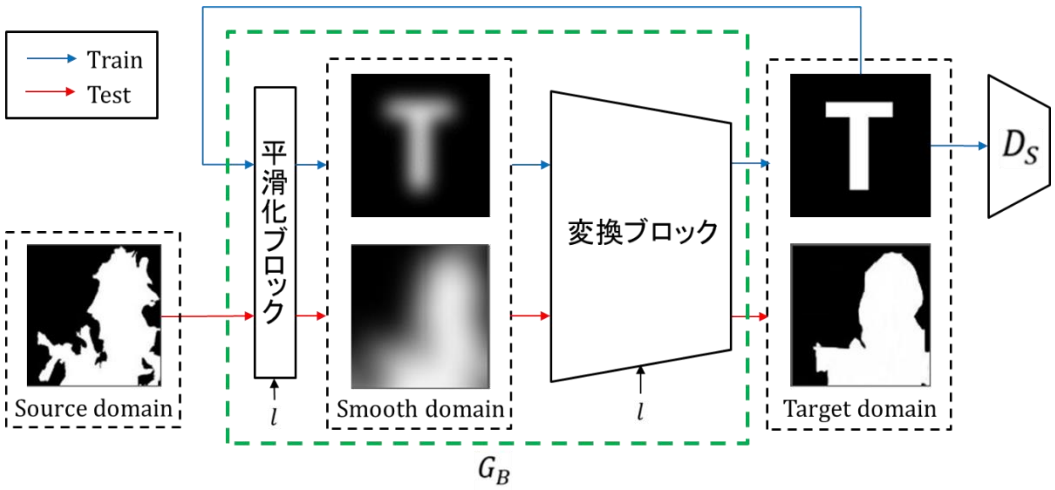

図 2.8　スケッチモジュール$\boldsymbol{G_B}$の概要図

平滑化ブロックは，ガウシアンカーネルを持つ固定畳み込み層として設定され，その標準偏差$\sigma = f(l)$と線形関数$f(\cdot)$によって制御される．この平滑化ブロックを利用して，同様の平滑性を持たせた輪郭で，スタイル画像の領域とテキスト画像の領域を対応させる．その後，変換ブロックを学習させることで構造転写を可能にする．最後に，学習したG_Bを様々なレベルlを持つXに適用する．これにより，対応するスケッチ形状$\tilde{X}_l = G_B(X, l)$を得る．

G_Bによりスケッチ形状$\{\tilde{X}_l\}$とパラメータ$l \in [0,1]$の対応する組を得た．ここで，G_Sを訓練して，対応する組をXにマッピングする．これにより，G_SがXの形状特徴をターゲットとなるテキスト画像に伝達できるようにする．

G_Sでは，データ増強と制御可能な ResBlock を採用している．これは，単にXを記憶してモデル崩壊を起こさないための戦略である．まず，Xと$\tilde{X}_l$をサブ画像ペア$\{x, \tilde{x}_l\}$にランダムにクロッピングする．次に，トレーニングセットとして収集する事でデータ増強を行う．さらに，StyleNet[11]のアーキテクチャをベースにG_Sを構築し，ResBlock[12]に対して，図 2.9(b)に示すように，lで重み付けをしたものを直線的に組み合わせる．これにより，制御可能な ResBlock を実現する．

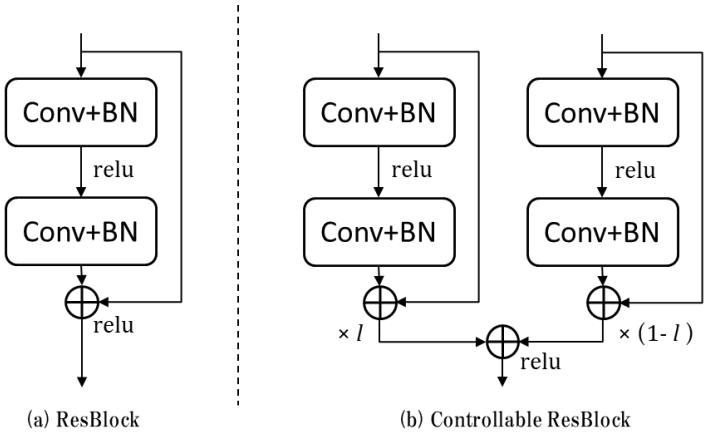

図 2.9　制御可能な ResBlock

損失関数については，D_Sを混乱させることを目的とする．

$$L_S^{rec} = \mathbb{E}_{x,l}\left[\left|\left|G_S(\tilde{x}_l, l) - x\right|\right|_1\right] \tag{2.8}$$

$$L_S^{adv} = \mathbb{E}_x[\log D_S(x)] + \mathbb{E}_{x,l}\left[\log\left(1 - D_S\left(G_S(\tilde{x}_l, l)\right)\right)\right] \tag{2.9}$$

また，大きなパラメータlをもつスタイルにおいては，テキストtを認識できない変形をしてしまう可能性がある．そこで，構造の転送結果$G_S(t, l)$に強制的にtのストロークを強制的に保持させるため，任意のグリフの可読性に関する損失も採用する．

$$L_S^{gly} = \mathbb{E}_{t,l}\left[\left|\left|(G_S(t, l) - t) \otimes M(t)\right|\right|_1\right] \tag{2.10}$$

ここで，$\otimes$は要素ごとの乗算演算子であり，$M(t)$はtの輪郭のうち最も近い点との距離に応じて画素値が増加する，距離ベースの重み付けマップである．G_Sの全体的な損失を以下の式(2.11)に示す．

$$\min_{G_S} \max_{D_S} \lambda_S^{adv} L_S^{adv} + \lambda_S^{rec} L_S^{rec} + \lambda_S^{gly} L_S^{gly} \tag{2.11}$$

2.6.3 Texture Transfer (G_T)

G_Sの学習と同様に，ランダムクロッピングを用いてXとYから適切な訓練ペア$\{x, y\}$を取得する．さらに，以下の式(2.12)，(2.13)に示す再構成用の損失関数と条件付き敵対損失関数を用いてG_Tを学習する．

$$L_T^{rec} = \mathbb{E}_{x,y}\left[\left|\left|G_T(x) - y\right|\right|_1\right] \tag{2.12}$$

$$L_T^{adv} = \mathbb{E}_{x,y}[\log D_T(x, y)] + \mathbb{E}_{x,y}\left[\log\left(1 - D_T\left(x, G_T(x)\right)\right)\right] \tag{2.13}$$

また，Neural Style Transfer[13]で提案されているスタイル損失L_T^{style}を加える．これにより，サンプリングされたテキスト画像tの全体的なスタイルのレンダリング性能を考慮する．G_Tの全体的な損失を以下の式(2.14)に示す．

$$\underset{G_T}{\text{Min}}\ \underset{D_T}{\max}\ \lambda_T^{adv} L_T^{adv} + \lambda_T^{rec} L_T^{rec} + \lambda_T^{style} L_T^{style} \tag{2.14}$$

2.7 むすび

本章では，本研究で用いるニューラルネットワーク，CNN，STN，ST-GAN および Shape Matching GAN の技術について述べた．

第3章 提案手法

3.1 まえがき

本章では，本研究で提案する手法について述べる．

3.2 Shape Matching GAN と ST-GAN の組み合わせによる写実的合成

本節では，ST-GAN と Shape Matching GAN を組み合わせ，空間変換とスタイル変換の両方を行い，写実的に合成する手法を提案する．前景オブジェクトとなるテキスト画像を，Shape Matching GAN でスタイル変換を行う．その後，スタイル変換された前景オブジェクトを ST-GAN を用いて空間変換し，合成画像を生成する．この手法の概要を図 3.1 に示す．

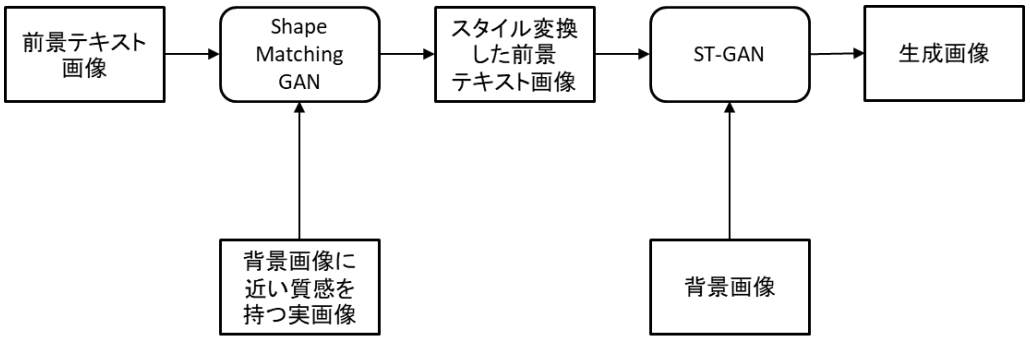

図 3.1　Shape Matching GAN と ST-GAN を用いた写実的合成の概要図

前景オブジェクトのスタイル変換を行う上で，スタイル画像として背景画像に近い質感を持った実画像を用いる．スタイル変換された前景オブジェクトを ST-GAN を用いて背景画像に合う形で空間変換を行い，合成画像を生成する．テキスト画像には α チャンネルを消去したものを用いる．

3.3 背景画像からスタイル抽出も行う写実的合成

3.2 節で提案した手法では，スタイル変換を行う際，背景画像に近い質感を持つ実画像を利用した．これは，背景画像からスタイルも抽出しようとした場合，合成領域が確保できない場合があるためである．また，形状マッチングを行う上で必要なスタイル領域が不足してしまっている可能性も考えられる．さらに，背景画像にスタイル領域が存在しない事も考えられる．しかし，スタイルの抽出を背景画像とは別の画像にする事で，スタイル変換の質が

落ちてしまう可能性も考えられる．そこで，本節ではスタイル変換を行うための質感も併せ持つ画像を背景画像として設定し，同一の画像でスタイル変換，空間変換の両方を行う手法を提案する．これにより，スタイル変換の面で品質を向上させ，さらに入力の画像を二つに削減する．この手法の概要を図 3.2 に示す．

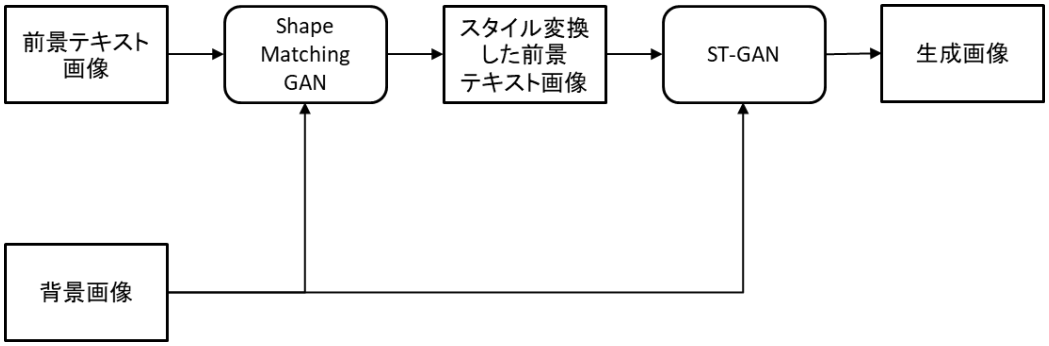

図 3.2　背景画像からスタイルも抽出する写実的合成の概要図

ST-GAN への入力画像として，背景画像のスタイル領域のマスク画像およびスタイル領域を除去した画像を用いる．これは，背景画像のスタイル部分に前景オブジェクトを代わりとして合成する事を目的としている．合成領域を確保し，スタイル領域の位置情報を得ることで，前景オブジェクトの空間変換をより自然に合成できる形に行うことを狙いとする．

3.4 むすび

本章では，写実的な合成を行う上で，スタイル画像として別の実画像を用いる手法および背景画像をスタイル画像としても利用する二つの手法について説明した．

第4章 実験

4.1 まえがき

本章では，第 3 章で提案した手法を用いた実験を行い，その結果および考察について述べる．

4.2 スタイル画像に背景画像とは別の実画像を用いた写実的合成

4.2.1 実験概要

入力画像には，主に前景オブジェクトとなるテキストのマスク画像，スタイル画像となる実画像および合成先の背景画像を用いる．テキストのマスク画像は画像形式がグレースケール，画像サイズは縦横比が同じものを用いる．その他は RGB の画像形式で，画像サイズは任意のものを用いる．

Shape Matching GAN の機構においては，スタイル領域を抽出し，α チャンネルを削除したスタイル画像のマスク画像を追加で作成する．さらに，テキストとスタイルのマスク画像をそれぞれ距離ベースに重み付けする．その後，スタイル画像の元画像とマスク画像を連結したもの用いて学習および変換を行う．

ST-GAN では，まず前景オブジェクトの入力として，Shape Matching GAN から生成されたスタイル変換後の前景画像のマスクを追加作成する．次に，背景画像，前景画像および前景マスク画像のそれぞれを ST-GAN のデフォルトの入力画像サイズである 120×160 画素に調整する．その後，それぞれの画像に対して変換および合成を行う．また，今回 ST-GAN は 3 次元情報を得るため，SUNCG dataset[14]で学習済みのモデルを用いる．

4.2.2 生成結果

本手法での画像の生成結果の成功例を以下の図 4.1 に示す．

図 4.1　手法 1 での生成結果の成功例

また，本手法での画像の生成結果の失敗例も以下の図 4.2 に示す.

図 4.2　手法 1 での生成結果の失敗例

図 4.1 に示した成功例では, 背景に合う形で前景のテキストがテクスチャも位置も変換できている．しかし，少し右に傾いた形となった．また，図 4.2 に示した失敗例は，背景に合わせた傾きがほとんどなく，前景の大きさも合成先をはみ出している．その結果，前景がただ拡大されただけで合成されたような生成結果となった.

4.3 背景画像からスタイル抽出も行う写実的合成

4.3.1 実験概要

スタイル変換のプロセスまでは，4.2.1 節で述べた手法 1 のものと同じである．ST-GAN における空間変換および合成の機構では，入力画像として 4.2.1 節で述べたものに加え，背景画像のスタイル領域のマスク画像および背景画像からスタイル領域を除いた画像を追加で作成する．その後，変換および合成を行う.

ここで, 背景画像全体に対するスタイル領域の占める割合が小さい場合, スタイルのテクスチャを正常に抽出できない可能性が考えられる．そこで，Shape Matching GAN での入力は，背景画像のスタイル領域の部分をトリミングして調整を行う.

4.3.2 生成結果

本手法での画像の生成結果の成功例を図 4.3 に示す.

図 4.3　手法 2 での生成結果の成功例

また，本手法での画像の生成結果の失敗例も以下の図 4.4 に示す.

図 4.4　手法 2 での生成結果の失敗例

図 4.3 で示した成功例はテクスチャ, 位置ともに背景に合う形で変換および合成ができている．図 4.4 に示した失敗例も，位置や傾きなどは変換ができている．しかし，テキストがはみ出す形となった.

4.4 評価実験

手法 1 と手法 2 の実験について，生成結果の自然さを評価するための主観評価実験を行った．それぞれの手法における生成結果を 5 枚ずつ用意し，50 人に対して実施した．生成結果の合成画像を見て,「テキストと背景が自然に合成できているか」という問いに対して 5 段階（5: とても自然，4: 少し自然，3: どちらとも言えない，2: 違和感が強い，1: とても違和感が強い）で評価した．その 5 段階での評価の平均値を以下の表 4.1 に示す.

表 4.1　評価実験の結果

手法 1	手法 2
2.332	2.72

また，評価実験で使用した生成結果の内，最も評価が高かったものは手法 2 を用いて生成したもので，評価の平均値は 3.78 であった．

これより，評価は背景からスタイル抽出も行った手法 2 の方が高く，より自然な合成結果を得ることが分か．

4.5 考察

4.5.1 スタイル画像に背景画像とは別の実画像を用いた写実的合成

4.2 節で示した生成結果と表 4.1 で示した評価実験の結果より，手法 1 における合成結果は自然さの低いものであることが分かった．この原因を考察する．

まず，スムーズに処理を行うため，スタイルの抽出先に背景画像と質感の似た別の実画像を用いた．しかし，これにより前景と背景のテクスチャに少くなからず違いが生じ，違和感のあるものとなったと考えられる．

また，入力に使用した背景画像，特に図 4.2 に示した失敗例の背景画像では，背景画像全体に対して合成したい箇所以外の領域が広かった．これにより，うまく 3 次元情報が取得できなかったことが推察される．

4.5.2 背景画像からスタイル抽出も行う写実的合成

4.3 節で示した生成結果と表 4.1 で示した評価実験の結果より，手法 2 での結果について考察する．背景画像からスタイル抽出も行った手法 2 の方が，手法 1 よりも良い結果となった．これについては，スタイル抽出を直接背景画像から行っているため，テクスチャの自然さは別画像からスタイルを抽出する手法 1 に比べ高くなる事が考えられる．また，ST-GAN にて背景画像のスタイル領域を合成先の位置情報として取得している．これにより自然に見えるように前景オブジェクトの空間変換も行えていると考えらえられる．

しかし，図 4.4 に示した失敗例のように，手法 2 における失敗結果では合成したい領域よりも，前景のテキストが大きくなる結果が多くあった．このことより，ST-GAN 内の反復 STN では前景オブジェクトの大きさの極端な調整は難しいことが推察される．

4.5.3 評価実験

4.4 節で示した評価実験の結果について考察する．評価実験の結果より，手法 2 の方が本提案手法内では自然度の高い生成結果を得られる手法であることが分かった．しかし，それぞれの手法を総合的に見ると，評価の最大値が 5 なのに対して，どちらの手法も評価の平

均値が 3 を超えない結果となった．これより，どちらの手法も総合的には精度の低い手法であると考えられる．

4.6 むすび

本章では，提案手法についての実験を行った．それぞれの手法での生成結果の成功例および失敗例，また評価実験およびその結果について述べた．さらに，それぞれの提案手法に対する考察を述べた．

第5章 結論と今後の課題

5.1 結論

本論文では，より写実的に画像を合成するために，Shape Matching GAN と ST-GAN を用いて，前景オブジェクトのスタイル変換と空間変換の両方を行う手法を提案した．

合成画像の生成には，スタイル抽出に背景画像に似た質感を持った実画像を用いた場合と，スタイル画像にも同一の背景画像を用いた場合の 2 通りの手法を行った．主観評価実験により，スタイル画像にも同一の背景画像を用いた場合のほうが良い結果が得られることが分かった．しかし，背景画像と大きくずれる空間変換がされた生成結果も得られた．

5.2 今後の課題

本論文では主に石彫りの文字を再現する実験を行った．しかし，空間変換の際に ST-GAN は家具の位置情報を得るための室内画像である，SUNCG dataset をデータセットとして学習したモデルを使用した．これにより，目的とする生成画像とデータセットの間に不整合さが生じていた．そのため，より自然な合成結果を得るため，データセットを目的とする生成画像に合わせたものにして学習を行う必要がある．

また，本論文では手法として入力画像のパターンを変えた場合しか提案をしていない．そのため，スタイル変換，空間変換のそれぞれの機構において別の手法を用い，さらに比較および検討を行う必要があると考えられる．

謝辞

本研究を行うに際して，研究の方向性や手法について丁寧にご指導くださった渡辺裕教授に深く感謝申し上げます.

また，ゼミやメール等で日頃からご協力やアドバイスをくださった研究室の皆様にお礼申し上げます.

最後に，ここまで常に支えてくださった家族に心より感謝いたします.

参考文献

[1] 長尾健作: “CAD データから写真画像を合成する技術とビジネス -CG 技術によるデジタル画像制作ビジネスの実践-”, 公益社団法人日本印刷技術協会 テキスト＆グラフィック研究会 tech Seminar, June. 2009.

[2] 富山真一, 大平倫宏: “画像合成を用いた亀裂検出システムの開発(ディペンダブルコンピューティング)”, 電子情報通信学会技術研究報告: 信学技報 113(353), p41-44, December. 2013.

[3] Fangneng Zhan, Hongyuan Zhu, and Shijian Lu, “Spatial Fusion GAN for Image Synthesis”, arXiv preprint arXiv:1812.05840v3, Apr. 2019.

[4] 斎藤康毅, ゼロから作る Deep Learning－Python で学ぶディープラーニングの理論と実装, pp.39-44, 205-220, オライリー・ジャパン, 2016.

[5] “畳み込みニューラルネットワーク（CNN）, ”MathWorks, https://jp.mathworks.com/discovery/convolutional-neural-network.html. 2021 年 1 月閲覧

[6] システムインテグレータ, “畳み込みニューラルネットワーク_CNN(Vol.16),” AISIA, 2018 年 5 月 11 日. https://products.sint.co.jp/aisia/blog/vol1-16. 2020 年 1 月閲覧.

[7] Max Jaderberg, Karen Simonyan, Andrew Xisserman, and Koray Kavukcuoglu, “Spatial Transformer Networks”, Neural Information Processing Systems (NIPS), 2015.

[8] Ian Goodfellow, Jean Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, ”Generative Adversarial Networks,” Neural Information Processing Systems (NIPS), 2014.

[9] Chen-Hsuan Lin, Ersin Yumer, Oliver Wang, Eli Shechtman, and Simon Lucey, “ST-GAN: Spatial Transformer Generative Adversarial Networks for Image Compositing,” arXiv preprint arXiv:1803.01837v1, Mar. 2018.

[10] Shuai Yang, Zhangyang Wang, Zhaowen Wang, Ning Xu, Jiaying Liu, and Zongming Guo, “Controllable Artistic Text Style Transfer via Shape-Matching GAN,” arXiv preprint arXiv:1905.01354v2, Aug. 2019.

[11] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, “Perceptual Losses for Real-Time Style Transfer and Super-Resolution,” arXiv preprint arXiv:1603.08155v1, Mar. 2016.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun, “Deep Residual Learning for Image Recognition,” arXiv preprint arXiv:1512.03385v1, Dec. 2015.

[13] Leon A Gatys, Alexander S Ecker, and Matthias Bethge, “Image Style Transfer Using Convolutional Neural Networks,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.2414-2423, 2016.

[14] Shuran Song, Fisher Yu, Andy Zeng, Angel X Chang, Manolis Savva, and Thomas Funkhouser, “Semantic Scene Completion from a Single Depth Image,” arXiv preprint arXiv:1611.08974v1, Nov. 2016.

図一覧

表一覧