様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月　 ６ 日現在

機関番号：１２１０２
研究種目：若手研究
研究期間：2020～2022
課題番号：２０Ｋ１９８１６

研究課題名（和文）限られた教師データを用いた意味的画像合成モデルの開発

研究課題名（英文）Developing Semantic Image Synthesis Model Using Limited Training Data

研究代表者
遠藤　結城（Endo, Yuki）

筑波大学・システム情報系・助教

研究者番号：００７９０３９６

交付決定額（研究期間全体）：（直接経費）　　3,200,000円

研究成果の概要（和文）：意味的画像合成とは、画素単位で建物や木々などの意味情報をラベル付けしたレイアウトのマップから画像のを生成する技術である。本研究では、少量のラベル付き教師データだけを用いて、高品質かつ多様な意味的画像合成が可能なアルゴリズムを開発した。さらに、ラベル付き教師データを全く用いずに、画像生成のレイアウトを制御する手法も開発した。研究成果として、意味的画像合成の多様化技術(国内発表2件、雑誌論文・国際会議論文2件)、few-shot意味的画像合成技術(国内発表1件、雑誌論文・国際会議論文1件)、zero-shot画像生成制御技術(国内発表1件、雑誌論文・国際会議論文1件)などの論文成果が得られた。

研究成果の学術的意義や社会的意義
本研究成果は、ここ数年で急速に発展している画像生成モデルにおいて、ユーザが介入可能な方法を開拓し、意味ラベルマップなど用いて出力を従来よりも低コストで、柔軟かつ多様に制御可能な方法を示したことに学術的な意義がある。社会的には、コンテンツ産業における創作活動の促進だけでなく、自動運転や医用画像解析の画像認識モデルの精度向上のための訓練データの構築など、本技術の広範な応用が期待できる。

研究成果の概要（英文）：Semantic image synthesis is a technique that can generate images from a semantic map annotated with pixel-level labels, such as buildings and trees. In this research, we developed an algorithm that can perform high-quality and diverse semantic image synthesis using only a small amount of labeled training data. Furthermore, we also developed a method for controlling the layout of generated images without using any labeled training data. We obtained research outcomes containing semantic image synthesis diversification (two domestic meetings and two journals/international conferences), few-shot semantic image synthesis (one domestic meeting and one journal/international conference), and zero-shot control of image generation (one domestic meeting and one journal/international conference).

研究分野：コンピュータグラフィックス

キーワード：深層学習　画像生成　GAN

様　式　C－１９、F－１９－１、Z－１９（共通）

１．研究開始当初の背景

画像対画像のマッピングを学習する条件付き画像生成の技術は、深層学習の急速な発展と共に、ここ数年で精力的に研究されてきた。特に画素単位で建物や木々などの意味情報をラベル付けしたマップ（意味ラベルマップ）から画像を生成する技術（意味的画像合成；Semantic Image Synthesis）は、コンピュータグラフィックス（CG）による複雑なジオメトリ情報やレンダリング設定なしに、直感的に任意のシーンを再現できる。生成画像は広告などのコンテンツ用途だけでなく、コンピュータビジョン（CV）分野における自動運転や医用画像解析のための画像認識モデルの性能を向上させる訓練データの用途も期待できる。

意味的画像合成を扱う多くの従来技術[1]は、意味ラベルマップとそのシーンの画像のペアから成る教師データを利用することが前提である。しかし、画素単位のラベリングは人的コストが高く、対象のドメインによっては教師データも十分に整備されていない。逆に画像から意味ラベルマップを推定するタスク（意味的領域分割；Semantic Segmentation）は、CV 分野において古くから取り組まれている技術[2] であり、これを用いて教師データを作る対処法も考えられる。しかし、もとよりこの方法も教師デ－タの利用が前提にあるため、教師データ不足に対処するという問題の本質的な解決には至らない。ラベル付き教師データが少数しか得られない、あるいは全くない状況において、コンピュータによる画像認識と生成を同時に解決し、それによる高品質な意味的画像合成の実現することは未解決の問題である。加えて、入力の意味ラベルマップはレイアウトのみを指定したものであるので、出力の色やテクスチャなどに曖が残るが、従来手法はこの曖昧性を考慮して多様な出力を得ることが難しい。

参考文献：

[1] P. Isola et al., Image-to-Image Translation with Conditional Adversarial Networks, CVPR2017.

[2] F. Lateef and Y. Ruichek, Survey on semantic segmentation using deep learning techniques, Neurocomputing, 338 (21), 2019.

２．研究の目的

本研究の目的は、限られたラベル付き教師データしかない状況において、高品質な意味的画像合成が可能なモデルおよび学習アルゴリズムを構築することである。そこでラベルの付与されていない画像を訓練データとして利用する、半教師あり学習や教師なし学習の枠組みでこの目的の達成を図る。特に、意味的領域分割と意味的画像生成を同時に学習可能な深層学習モデルと学習アルゴリズムによって、ラベルなしデータを有効活用するアプローチを検討する。加えて、出力画像の曖昧性に対処するために、シーンのレイアウトに依存する意味ラベルマップなどのユーザ入力と、それ以外の外見要素の特徴とのもつれをほどいた表現を学習し、出力画像を多様化する手法を開発する。

３．研究の方法

上記の目的をふまえ、本研究では 1)意味的画像合成の多様化、2)few-shot　意味的画像合成、3)zero-shot　な画像生成モデルのレイアウト制御の三つの課題に取り組んだ。以下に詳細を説明する。

(1) 意味的画像合成の多様化

これまでの手法では、単一の潜在空間を学習することで出力画像の大域的な外観を制御してきた。しかし、物体の見え方は複数の要因に依存するため、単一の潜在空間では様々な物体の見え方を捉えることができない場合が多い。そこで、クラスとレイヤごとに拡張した変分オートエン

図１：単一の意味マスクから多様な画像を生成できる。

コーダ(VAE)を提案し、複数の潜在空間を学習することで、局所的なレベルから大域的なレベルまで柔軟に各物体クラスの見た目を制御することが可能になった（図 1 参照）。三つの異なるドメインの実データと合成データを用いた広範な実験により、本手法が最先端の手法と比較して、写実的かつより多様な画像を生成することを実証した。

(2) few-shot 意味的画像合成

本プロジェクトでは、画素単位にアノテーションされた訓練画像が少量しか手に入らない few-shot 環境において、レイアウトを表す意味スクリブルから写実的な画像を生成する新たなタスクを開拓した（図 2 参照）。これを実現するため、StyleGAN Prior を活用した擬似ラベリングに基づいて、意味的画像合成モデルを学習する手法を提案した。提案手法のアイデアは、少量の訓練データに定義された意味クラスと、StyleGAN の特徴量とのマッピングを構築する点にある。これにより生成された無数の擬似的な意味スクリブルを利用して、StyleGAN を制御するためのエンコーダを学習している。評価実験により 1 ショットや 5 ショットの設定で、レイアウトの忠実性と視覚的品質の観点から、提案手法は既存手法よりも良好な結果が得られることを実証した。

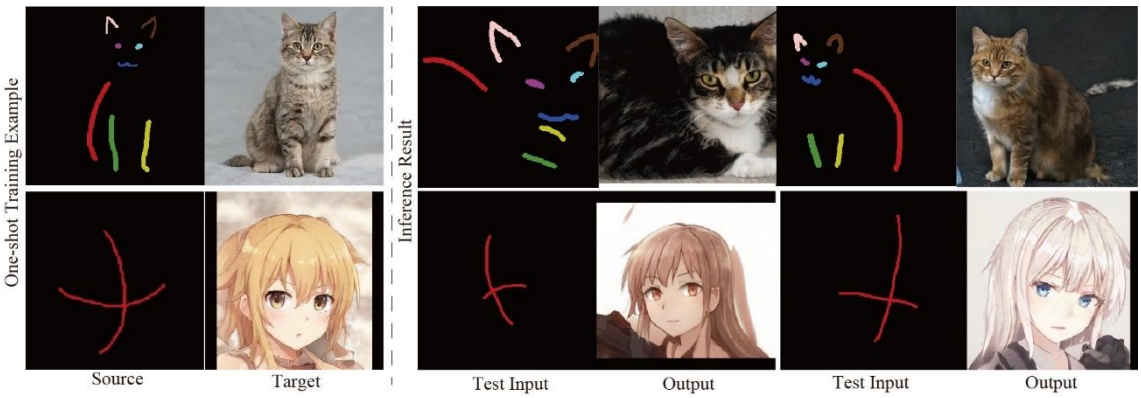

図 2：左のように訓練のペア画像が少量でも右のように多様なレイアウトに対応する画像を生成できる。

(3) zero-shot な画像生成モデルのレイアウト制御

Latent space exploration は、generative adversarial network (GAN) において解釈可能な潜在変数の操作方向を発見し、生成画像に対する様々な属性を編集できる技術である。既存研究では、空間的な制御が平行移動や回転など単純な変換に限られるとともに、適切な潜在変数の操作方向を特定し調整するのに手間がかかるのが問題であった。そこでこのプロジェクトでは、画像に直接アノテーションを施すことで、StyleGAN 画像のレイアウトを対話的に編集するフレームワークを開発した。本フレームワークでは、ユーザが StyleGAN 画像に対して、動かしたい場所や固定したい場所をアノテーションし、マウスドラッグにより移動方向を指定する（図 3 参照）。これら可変長のユーザ入力に応じて潜在変数を適切に変換するために、transformer に基づく潜在変数変換器を提案する。学習には、学習済みの StyleGAN とオプティカルフローモデルによって生成された合成データを利用するので、追加の教師データを必要としない。定量的・定性的な評価により本手法の有効性を示した。

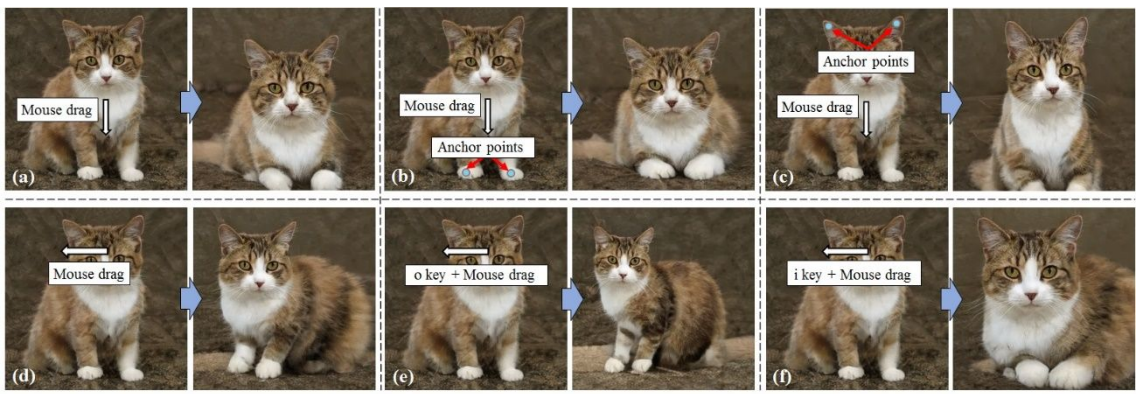

図 3：(a)～(f)に示すアノテーションによって右隣の各画像のように生成画像のレイアウトを制御できる。

４．研究成果

上述した各プロジェクトについて、当該研究分野における貢献および成果を以下に述べる。

(1) 意味的画像合成の多様化

これまでも意味的画像合成において出力を多様化するマルチモーダル画像合成の手法は数多く研究されてきた。しかし、これらの手法は出力画像中の全体的な色味やテクスチャなど、大域的な要素しか編集できなかった。これに対して本研究のポイントは、1)画像中の物体ごとの見た

目の変化を分解して、異なる潜在空間でモデリングすることで、物体ごとに色やテクスチャを変えるなど、柔軟な出力画像の制御を可能にしており、これが貢献の一つといえる。また、本手法を用いることで、2) 物体ごとの Style の転写や、材質の編集など、画像合成・編集作業への幅広い応用が可能であることも示している。

本成果は CG 分野において SIGGRAPH や SIGGRAPH Asia に次ぐ難関国際会議である Pacific Graphics 2020 において発表した。会議に採択された論文は CG 分野において権威ある論文誌 Computer Graphics Forum にも採録されている。

(2) few-shot 意味的画像合成

従来研究では、意味スクリブルやマスクのようなアノテーションから画像を生成するために、画素単位のアノテーションと実画像からなる大量のペアデータを訓練する必要があった。本研究では、1)わずかなペアデータさえあれば、対話的に高品質な画像を合成できる、few-shot 意味的画像合成という新たなタスクを開拓した。技術的には、2) StyleGAN の中間層から識別性の高い特徴量を得られることを発見し、この特徴量を使った擬似ラベリング手法を提案した。また、擬似ラベリングの精度を補うために、3) 最適化ベースの GAN inversion による後処理手法も提案している。

本成果は CG 分野において歴史ある国際会議である Computer Graphics International 2022 で発表した。会議に採択された論文は論文誌 Computer Animation and Virtual Worlds にも採録されている。また、国内の CGVI 研究発表会において優秀研究発表賞を受賞した。

(3) zero-shot な画像生成モデルのレイアウト制御

GAN による画像生成において、年齢や性別など特定の属性やテキスト入力によって出力を制御しようとする研究は数多く存在するが、特定の箇所のレイアウトを編集するには適していない。本研究の主な貢献は、1)画像上のユーザ入力に合わせて StyleGAN 画像のレイアウトを制御するフレームワークを世界で初めて提案したこと、2)transformer に基づく潜在変数変換器の提案、3)合成データを使った教師なしの学習方法の三つである。

本成果は CG 分野において SIGGRAPH や SIGGRAPH Asia に次ぐ難関国際会議である Pacific Graphics 2020 において発表した。会議に採択された論文は CG 分野において権威ある論文誌 Computer Graphics Forum にも採録されている。また、Best Paper Honorable Mention Award を受賞し、国内のシンポジウムにおいて VC 論文賞、CGVI 研究会優秀研究発表賞を受賞した。

５．主な発表論文等

〔雑誌論文〕　計4件（うち査読付論文　4件／うち国際共著　0件／うちオープンアクセス　0件）

１．著者名 Endo Y.	４．巻 41
２．論文標題 User‐Controllable Latent Transformer for StyleGAN Image Layout Editing	５．発行年 2022年
３．雑誌名 Computer Graphics Forum	６．最初と最後の頁 395～406
掲載論文のDOI（デジタルオブジェクト識別子） 10.1111/cgf.14686	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Endo Yuki、Kanamori Yoshihiro	４．巻 33
２．論文標題 Controlling StyleGANs using rough scribbles via one‐shot learning	５．発行年 2022年
３．雑誌名 Computer Animation and Virtual Worlds	６．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/cav.2102	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Yoshikawa Takato、Endo Yuki、Kanamori Yoshihiro	４．巻 38
２．論文標題 Diversifying detail and appearance in sketch-based face image synthesis	５．発行年 2022年
３．雑誌名 The Visual Computer	６．最初と最後の頁 3121～3133
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s00371-022-02538-7	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Endo Y.、Kanamori Y.	４．巻 39
２．論文標題 Diversifying Semantic Image Synthesis and Editing via Class‐ and Layer‐wise VAEs	５．発行年 2020年
３．雑誌名 Computer Graphics Forum	６．最初と最後の頁 519～530
掲載論文のDOI（デジタルオブジェクト識別子） 10.1111/cgf.14164	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計9件（うち招待講演　0件／うち国際学会　4件）

1．発表者名
Yuki Endo

2．発表標題
User-Controllable Latent Transformer for StyleGAN Image Layout Editing

3．学会等名
Pacific Graphics 2022（国際学会）

4．発表年
2022年

1．発表者名
Takato Yoshikawa, Yuki Endo, Yoshihiro Kanamori

2．発表標題
Diversifying detail and appearance in sketch-based face image synthesis

3．学会等名
Computer Graphics International 2022（国際学会）

4．発表年
2022年

1．発表者名
Yuki Endo, Yoshihiro Kanamori

2．発表標題
Controlling StyleGANs using rough scribbles via one-shot learning

3．学会等名
Computer Graphics International 2022（国際学会）

4．発表年
2022年

1．発表者名
Yuki Endo, Yoshihiro Kanamori

2．発表標題
Diversifying Semantic Image Synthesis and Editing via Class- and Layer-wise VAEs

3．学会等名
Pacific Graphics 2020（国際学会）

4．発表年
2020年

１．発表者名
吉川 天斗，遠藤 結城，金森由博

２．発表標題
StyleGAN を用いたテキストによる人物画像の服装編集手法

３．学会等名
第 189 回コンピュータグラフィックスとビジュアル情報学研究発表会

４．発表年
2022年

１．発表者名
遠藤 結城

２．発表標題
ユーザ制御可能なLatent Transformer を用いたStyleGAN 画像のレイアウト編集

３．学会等名
Visual Computing 2022

４．発表年
2022年

１．発表者名
遠藤結城、金森由博

２．発表標題
StyleGAN Prior を用いたFew-shot 意味的画像合成

３．学会等名
Visual Computing 2021

４．発表年
2021年

１．発表者名
吉川 天斗，遠藤 結城，金森 由博，三谷純

２．発表標題
詳細とスタイルを制御可能にしたスケッチからの顔画像生成手法

３．学会等名
第 184 回コンピュータグラフィックスとビジュアル情報学研究発表会

４．発表年
2021年

１．発表者名
吉川 天斗，遠藤 結城，金森 由博

２．発表標題
多様性を考慮したスケッチからの画像生成手法

３．学会等名
Visual Computing 2021（ポスター）

４．発表年
2021年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

研究プロジェクトページ
http://www.cgg.cs.tsukuba.ac.jp/~endo/projects/clVAE/
http://www.cgg.cs.tsukuba.ac.jp/~yoshikawa/pub/sketch_to_diverse_image/
http://www.cgg.cs.tsukuba.ac.jp/~endo/projects/StyleGANSparseControl/
http://www.cgg.cs.tsukuba.ac.jp/~endo/projects/UserControllableLT/

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関