

令和 5 年 5 月 25 日現在

機関番号：12601

研究種目：若手研究

研究期間：2021～2022

課題番号：21K17759

研究課題名（和文）漫画におけるシーングラフを用いたシーン理解と生成

研究課題名（英文）Scene Understanding and Generation for Manga using Scene Graphs

研究代表者

古田 諒佑（Furuta, Ryosuke）

東京大学・生産技術研究所・助教

研究者番号：20843535

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）：2021年度は漫画におけるシーン理解という包括的な問題を個々の基礎的なタスクに分割して取り組んだ。これらの内容は国内学会では画像の認識・理解シンポジウム(MIRU)のポスター発表と映像表現・芸術科学フォーラムの口頭発表を行い、3件の優秀発表賞を受賞した。また、画像処理分野における査読付き旗艦国際会議ICIPとCG分野における国際会議SIGGRAPH ASIAのポスターに採択され発表を行った。2022年度はドメイン（作風）に依存しない物体検出器の学習に取り組んだ。本研究の成果は国内で最大規模の査読有りコンピュータビジョンの会議であるMIRU2023にショートオーラル発表として採択された。

研究成果の学術的意義や社会的意義

本研究の漫画のシーン理解に向けた取り組みにより、これまでよりも詳細かつ高精度なシーン理解が期待される。応用先としては、類似シーンを含む漫画の検索が可能となり、例えば販売サイトにおいてタグやタイトル・作者名による検索以外の新たな検索方法として利用可能である。別の応用先としては漫画の自動翻訳が挙げられ、自動翻訳の精度向上という産業応用に対する貢献だけでなく、高精度な自動翻訳により日本語版と同時に（遅延なく）正規の翻訳版を海外で出版できるようになり、海外における日本の漫画の海賊版普及の抑制という社会的貢献の可能性も秘めている。

研究成果の概要（英文）：In FY 2021, I split the scene understanding on Manga into some fundamental tasks on computer vision and worked on them. These works were presented in MIRU2021 as poster presentations and in Expressive Japan 2022 as oral presentations, and three of them won Outstanding Presentation Award in Expressive Japan 2022. Also, one of these works was accepted to the International Conference on Image Processing (ICIP) 2021, which is a flagship conference in the field of image processing, and SIGGRAPH ASIA 2021 Posters, which is an international conference on computer graphics.

In FY 2022, I worked on domain generalization for object detection, where the objective is to train a domain-invariant (i.e., robust to the change of image styles) detector. This work was accepted to MIRU2023 as a short oral presentation, which is the biggest conference on computer vision in Japan.

研究分野：コンピュータビジョン

キーワード：物体検出 半自動着色 クラスタリング インペインティング 話者推定 ドメイン汎化

1. 研究開始当初の背景

漫画(特に電子コミック)は今や日本のみならず世界的に人気な一大コンテンツの1つである。公益社団法人全国出版協会・出版科学研究によると2019年度の電子コミック市場は前年度比29.5%増の2,593億円であり、その急速な成長と市場規模の大きさが窺える。「漫画村」に代表される海賊版サイトが社会問題として話題になったのも記憶に新しい。また、漫画投稿サービスが世界中(日本: Pixiv, 中国: Kuaikan, 韓国: Naver等)で急速に拡大しているのに伴い、特殊な訓練を受けたプロ作家以外が創作活動をする機会が増えている。

漫画はキャラクターや場面の描写が線で描かれており、その表現は高度に抽象化されている。その上、時系列情報も限られたコマ数で表現されており、非常にスパースである。これらは自然画像や通常の動画像とは大きく異なる点である。それでも我々人間は漫画を読んだ際にはキャラクターの動きや場面の遷移を自然に理解することができる。コンピュータビジョンの分野で自然画像認識や動画像認識に取り組んだ研究は数多く存在するが、抽象化された表現である漫画に対してシーン理解に取り組んだ研究はほとんどない。また、漫画を読んだ際にシーンを理解すること自体は誰でも可能であるが、専門家ではない素人が描いた絵ではキャラクターの躍動感や場面の緊迫感などを表現できない可能性が大きい。

2. 研究の目的

申請当初の検討としては、(i)漫画の詳細なシーン理解に取り組み、またその応用として(ii)シーン表現からのコマ画像の生成を目的としていた。ここで、シーン表現からのコマの画像生成とは、例えば「キャラクターAがキャラクターBと会話している。キャラクターAは剣を持っている。」等のシーンを表現する簡単な入力からコマ画像を生成することである。この入力は自然言語ではなくグラフ表現を用いる。本研究の(i)により漫画のシーンがより詳細に理解できるようになれば、例えば自動翻訳に応用することができる。日本の漫画を海外に輸出する際には翻訳が必要であるが、誰がどのテキストを誰に向かって話しているのかを認識することができれば、自動翻訳において3人称(he, she等)を適切に選択できたり、あるキャラクターの特徴的な言い回し等の翻訳の精度向上に繋がる。別の応用としては、類似シーンを含む漫画の検索が可能となり、例えば販売サイトにおいてタグやタイトル・作者名による検索以外の新たな検索方法として利用可能である。これらの応用により漫画業界の更なる発展に貢献することができる。それだけでなく、(ii)によりコマ画像の生成が可能になれば素人の創作支援につながり、漫画文化の裾野を広げることができる。また本研究の要素技術を応用することで、映画やアニメなど他のコンテンツの理解と生成にも応用できると考えている。そのため、プロの作家のみならずあらゆる人々の創作活動支援に役立つことが期待される。しかしながら、通常の動画像認識とは異なり、漫画は絵による表現だけでなく時系列(コマ間)情報も高度に抽象化されているため、漫画のシーン理解と生成はチャレンジングなタスクである。

3. 研究の方法

申請時に検討していた方法は以下の通りである。

(i) 漫画の詳細なシーン理解

まず初めにシーングラフ生成器を学習する。シーングラフとは、入力画像中の被写体(キャラクターやテキスト、オブジェクト)とそれらの間の関係性をグラフで表現したものである。例えば、「character A」というノードから「hold」というノードにエッジが張っており、さらにそこから「racket」のノードにエッジを張ることで「キャラクターAがラケットを持っている」という関係性を表現している。キャラクターがセリフを喋っているという関係性も同様にグラフで表現できる。シーングラフ表現を用いることで、通常の物体検出等よりもより詳細なシーン理解に繋がると考えている。自然画像からのシーングラフ生成は[Xu, CVPR17]を初めとして他にも研究が行われているが、漫画においてシーングラフ生成に取り組んだ研究はこれまでにない。そのため既存技術そのまま用いるのではなく、抽象的表現である漫画の性質を考慮したより高精度な認識(シーングラフ生成)技術の開発に取り組む。

(ii) シーングラフからのコマ画像生成

次に、コマ画像生成器を学習する。学習時にはコマ画像を入力し、シーングラフ生成器でシーングラフを生成すると同時に、画像特徴抽出ネットワークを用いてそのコマ画像を表現する特徴ベクトルを抽出する。そして、Variational AutoEncoder (VAE)のようにLatent spaceへのマッピングを学習する。コマ画像生成器は、シーングラフとLatent spaceからサンプリングしたベクトルを入力として受け取り、元のコマ画像を復元するように学習する。一般的に漫画のコマ画像はバリエーションが大きく、特徴ベクトルのみからの復元は難しいと予想されるが、シーングラフを補助情報として用いることで効果的に学習可能であると予想される。推論時には、シーングラフ生成器と画像特徴抽出ネットワークは用いず、Latent spaceからサンプリングしたベクトルとユーザーが頭の中で思い描くシーンを表現したグラフをコマ画像生成器に入力することで、ユーザーが所望するコマの生成が可能になると期待される。また推論時にシーングラフはそ

のままに、別の作品のコマ画像から抽出した画像特徴ベクトルを入力することで、シーンの構成はそのままで絵柄だけを別の作品風に転写する等の応用も期待される。

4. 研究成果

2021 年度は漫画におけるシーン理解に向けた取り組みとして、シーン理解という包括的な問題をコンピュータビジョン分野における個々の基礎的なタスクに分割して取り組んだ。具体的には、図 1 に示すキャラクター顔画像のクラスタリング、話者推定、インペインティング、描き文字検出、半自動着色という 5 つのコンピュータビジョン分野における基礎的なタスクに切り分けて取り組んだ。個々のタスクにおいて、既存手法の課題を発見しそれらを改善する手法を提案し、精度評価にてその効果の確認を行った。具体的には、キャラクター顔画像のクラスタリングでは作品に依らない特徴表現の学習方法を提案しクラスタリング精度を向上させ、話者推定ではルールベースと BERT による機械学習の識別を組み合わせた高精度な推定方法を提案した。インペインティングでは作者の画風に注目して各画風の特徴を捉えて補間を学習する方法を提案した。描き文字検出では合成データを使用した効率的な学習方法と線画化を用いた高精度な検出方法を、半自動着色ではスクリーントーン画像とベタ塗り（下塗り）画像を用いた高精度な自動着色手法をそれぞれ提案した。これらの内容は国内学会では画像の認識・理解シンポジウム (MIRU) のポスター発表と映像表現・芸術科学フォーラムの口頭発表を行い、映像表現・芸術科学フォーラムでは 3 件の優秀発表賞を受賞した。また、半自動着色の内容は画像処理分野における査読付き旗艦国際会議 ICIP とコンピュータグラフィックス分野における国際会議 SIGGRAPH ASIA のポスターに採択され発表を行った。



図 1 取り組んだ基礎的なタスクとその実際の結果

2022 年度は主に物体検出におけるドメイン汎化という問題に取り組んだ。漫画における物体検出は、キャラクターや内容に基づく推薦や自動翻訳などの様々な応用を持つ重要なタスクである。しかしながら漫画では作品や作者が変わるごとに作風（画風やテキストのフォントや配置）が大きくことなり（これはドメインの違いと呼ばれる）、ある特定の作品群で学習した検出器を別の作品でテストした場合に大きく性能が低下することが知られている。そこで本研究ではドメイン（作風）に依存しない物体検出器の学習に取り組んだ。具体的には、ある一つのラベルありドメインと複数のラベル無し（もしくは弱ラベル付き）ドメインのデータを学習に使用することでドメインに依存しない検出器を学習する“半教師無し（もしくは弱教師あり）ドメイン汎化物体検出”という新たなタスクを提案し（図 2）、これを統一的に解く学習方法も示した。これによりラベル付けのアノテーションの手間をほとんど増やすことなくドメインに依存しない物体検出器の学習が可能になる。本研究の成果は国内で最大規模の査読有りコンピュータビジョンの会議である MIRU2023 にショートオーラル発表として採択された。


















	Single-DGOD	Semi-supervised DGOD	Weakly-supervised DGOD	DGOD	UDAOD
Training	 natural	 natural  unlabeled comic  unlabeled clipart	 natural  weakly-labeled comic  weakly-labeled clipart	 natural  comic  clipart	 natural  unlabeled watercolor
Testing	 watercolor	 watercolor	 watercolor	 watercolor	 watercolor

図 2 提案する半教師ありおよび弱教師ありドメイン汎化物体検出とその関連タスク

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計12件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 古田諒佑, 佐藤洋一
2. 発表標題 物体検出の半教師有りおよび弱教師有りドメイン汎化
3. 学会等名 第26回画像の認識・理解シンポジウム (MIRU2023)
4. 発表年 2023年

1. 発表者名 小松俊太, 古田諒佑, 谷口行信, 日並遼太, 石渡祥之佑
2. 発表標題 漫画におけるコマと作風を考慮したキャラクター顔クラスタリング
3. 学会等名 映像表現・芸術科学フォーラム
4. 発表年 2022年

1. 発表者名 仲偉博, 古田諒佑, 谷口行信, 日並遼太, 石渡祥之佑
2. 発表標題 深層学習に基づく漫画における描き文字の検出
3. 学会等名 映像表現・芸術科学フォーラム
4. 発表年 2022年

1. 発表者名 清水友悟, 古田諒佑, 欧陽徳龍, 谷口行信, 日並遼太, 石渡祥之佑
2. 発表標題 ベタ塗り画像を利用した半自動漫画彩色手法
3. 学会等名 映像表現・芸術科学フォーラム
4. 発表年 2022年

1. 発表者名 欧陽徳龍, 古田諒佑, 清水友悟, 谷口行信, 日並遼太, 石渡祥之佑
2. 発表標題 Watershed法を用いたインタラクティブ漫画彩色システム
3. 学会等名 映像表現・芸術科学フォーラム
4. 発表年 2022年

1. 発表者名 山口理哉, 古田諒佑, 谷口行信, 日並遼太, 石渡祥之佑
2. 発表標題 会話の流れを考慮した漫画における自動話者推定
3. 学会等名 映像表現・芸術科学フォーラム
4. 発表年 2022年

1. 発表者名 横田広之, 古田諒佑, 谷口行信, 日並遼太, 石渡祥之佑
2. 発表標題 作者分類と線画のヒントを用いた漫画の画像修復
3. 学会等名 映像表現・芸術科学フォーラム
4. 発表年 2022年

1. 発表者名 小松俊太, 古田諒佑, 谷口行信, 日並遼太, 石渡祥之佑
2. 発表標題 線画画像を用いた漫画キャラクターの顔画像クラスタリング
3. 学会等名 画像の認識・理解シンポジウム (MIRU), poster
4. 発表年 2021年

1. 発表者名 横田広之, 古田諒佑, 谷口行信, 日並遼太, 石渡祥之佑
2. 発表標題 GANと作品分類を用いた漫画の画像修復
3. 学会等名 画像の認識・理解シンポジウム (MIRU), poster
4. 発表年 2021年

1. 発表者名 山口理哉, 古田諒佑, 谷口行信, 日並遼太, 石渡祥之佑
2. 発表標題 セリフの内容を考慮した漫画における自動話者推定
3. 学会等名 画像の認識・理解シンポジウム (MIRU), poster
4. 発表年 2021年

1. 発表者名 Yugo Shimizu, Ryosuke Furuta, Delong Ouyang, Yukinobu Taniguchi, Ryota Hinami, and Shonosuke Ishiwatari
2. 発表標題 Painting Style-Aware Manga Colorization Based on Generative Adversarial Networks
3. 学会等名 IEEE International Conference on Image Processing (ICIP) (国際学会)
4. 発表年 2021年

1. 発表者名 Delong Ouyang, Ryosuke Furuta, Yugo Shimizu, Yukinobu Taniguchi, Ryota Hinami, and Shonosuke Ishiwatari
2. 発表標題 Interactive Manga Colorization with Fast Flat Coloring
3. 学会等名 SIGGRAPH ASIA Posters (国際学会)
4. 発表年 2021年

〔図書〕 計0件

〔出願〕 計1件

産業財産権の名称 画像着色方法及び装置、着色画像生成モデルの生成装置、並びにプログラム	発明者 清水友悟、古田諒 佑、欧陽徳龍、谷口 行信、日並遼太、石	権利者 同左
産業財産権の種類、番号 特許、特願2021-11521	出願年 2021年	国内・外国の別 国内

〔取得〕 計0件

〔その他〕

-

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
---------	---------