# MAIA: 画像認識と文字認識を用いた漫画背景の自動生成システムの検討

木村峻輔[1,a)]　大井翔[1,b)]

**概要**：漫画の制作には長時間の作業が必要となる．週刊連載を行っている漫画家だと 1 日に 10 時間以上の作業を有する．この作業工程のペン入れという作業には漫画の背景作画というものがある．この背景を作成するのに 1 日以上の作業を必要とする場合がある．そのため，背景作画にはアシスタントに手伝ってもらう方法が一般的な作業時間の短縮方法である．しかし，現在連載を行っていない漫画家やこれから漫画を描こうと考えている人はアシスタントを雇うことが難しいので作者自ら描かなければならない．本研究では漫画の背景を自動で制作するシステムを開発することで背景作画の作業時間を行うことが可能かの調査を行う．具体的には，画像認識，文字認識を用いて，場所推定を行い，生成 AI で背景画像を出力するシステムを提案する．

## 1. はじめに

2017 年のアミューズメントメディア総合学院が投稿した記事 [1] によると出版社への漫画の持ち込み，投稿を行う漫画家志望者が約 1000 名いると言われており，その中から新人賞等を受賞することができる漫画家が 10 人程となっている．さらに賞を受賞した志望者から数名がプロとしてデビューすることができる．このようにプロの漫画家になれる割合は非常に低いとされている．プロとしてデビューした漫画家の最初の課題として期限内に漫画を制作するということが挙げられる．その理由として漫画の制作には長時間の作業が必要であることが考えられる．売れている週刊連載の漫画家を基準として説明すると 1 日に最低で 10 時間以上の労働を行わなければ締め切りに間に合わなくなると考えられる．

図 1 に漫画家の作業スケジュールを示す．

図 1 を参考に漫画家の主なスケジュールを説明すると編集者とのストーリー等の打ち合わせを行い，その打ち合わせを基にコマ割りや絵を構成するネーム制作を行う．ネーム制作を終えた後編集者の確認を行い，問題なければ原稿作成に移行する．下書き，ペン入れを行い，原稿を完成させ提出することで作業の全てが終了となる．以上が漫画家の主な作業スケジュールとなっている．

原稿作成時に行ったペン入れという作業は細かく分ける

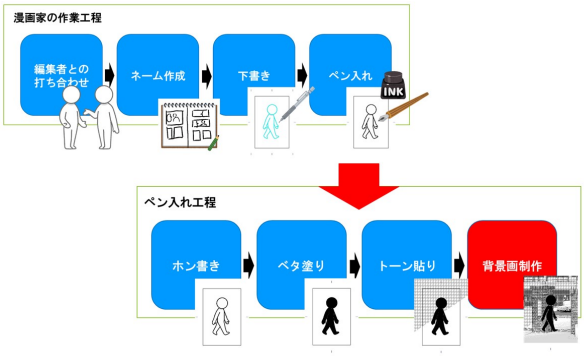


**図 1**　漫画家の作業工程

ことができ，ホン書き，ベタ塗り，トーン貼り，背景制作で構成されている．この作業工程の 1 つである背景制作には最低でも 3 時間，直線が多く細かい背景だと 1 日以上かかる場合もある．そのため多くの漫画家は背景制作をアシスタントに任せる場合が多い．しかし現在連載を行っていない漫画家やこれから漫画を描こうと考えている人はアシスタントを雇うことが難しいので作者自ら描かなければならない．

本研究では漫画の背景の自動生成を行うシステムを開発することで背景制作に有する時間を短縮することができ作業の効率を向上させることを目的とする．

## 2. 関連研究

関連研究として漫画の画像認識，文字認識及び背景等の生成に関する論文に注目し調査を行った．中山らは LinDA: 漫画向けの建造物の背景線画半自動生成 [2] という研究を

[1] 大阪工業大学
[a)] shunsuke.kimura@mix-lab.net
[b)] sho.ooi@outlook.jp

行った．この研究の背景としては，近年漫画制作はデジタルで行う事が増えており，背景の制作等もデジタルの背景のフリー素材や写真を線画化し使用するケースが増えてきている．しかしフリー素材や写真の線画化した素材は汎用性を重視したものとなっているため自分の作品と馴染まない等といった問題点が挙げられる．この原因の 1 つとして作者の線画とフリー素材の線画の手癖が異なると考えられる．

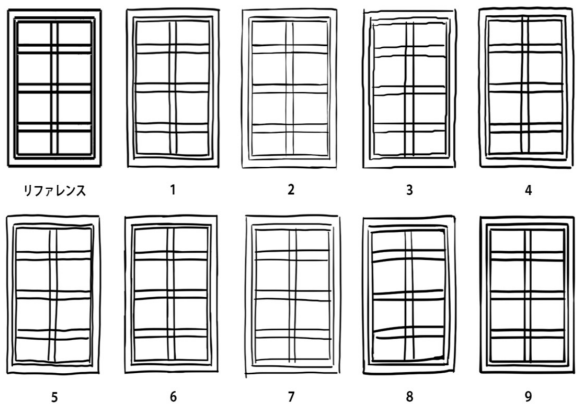


**図 2**　サンプルデータ

この研究では作者の手癖と近い背景を生成するために図 2 のような作者が描いた手書きの線画をサンプルデータとし，線の特徴を抽出し建造物の背景画像を生成するシステムを構築する．これにより研究背景で挙げられた問題点の解決を目指すものとする．本研究との類似点は漫画の背景の制作を自動化といった点が類似しているので研究に参考になると考えられる．本研究に活用が可能な部分として挙げられるのは，作者の手癖に近い背景を生成することができるという点で本研究の背景生成のクオリティ向上に繋がると考えられる．

青木らが行った Vision Transformer と BART を用いた漫画のマルチモーダル識別 [3] という研究がある．この研究の背景は AI による小説やイラストといった創作物の自動生成は近年盛んに行われている．しかしこうした創作物の理解を計算機で行うのは困難であるとされている．この研究では，数ある創作物の中で絵と文字の要素からなる漫画を用いて，AI に作品を理解させ，作品ごとの識別を行う．この研究で使用された技術として Vision Transformer と BART が挙げられた．Vision Transformer とは主に自然言語処理で利用される Transformer という技術を画像処理の分野に適応させた画像認識モデルのことを指す．この研究では漫画のキャラクターの顔画像等の認識に使用される．BART(Bidirectional Encoder Representations from Transformer) とは文章分類や質疑応答等様々な分野で活用することが可能な自然言語処理モデルを指す．この技術を使用し漫画内に出現したセリフの文字認識を行う．

**表 1**　実験結果比較

| モデル | 結果 |
|---|---|
| 青木らの研究手法 | 0.872 ± 0.008 |
| Vision Transformer | 0.862 ± 0.007 |
| BART | 0.418 ± 0.002 |

この研究の実験として Manga109[4] の画像データを活用しその漫画画像の作品識別を行い，その精度を結果として示す．


**図 3**　漫画の識別画像

図 3 はこの実験で得られた結果を可視化したものである．図 3 の実験結果の可視化の方法として用いられたのは Grad-CAM(Gradient-weighted Class Activation Mapping) という可視化技術である．これによりカラーマップ可視化を可能とする．色分けは注目度が高い部分が赤色に，注目度が低い部分が青色に近い色となる．

表 1 がこの研究の実験結果を示したものである．
この実験で制作された提案手法の比較として Vision Transformer，BART それぞれ単体で行った結果を比較対象としている．実験の結果から提案手法での作品識別がそれぞれの単体モデルで行った際の識別結果より優れていることが分かる．

本研究との類似点として画像認識，文字認識技術を用いて，キャラクターやセリフを認識しているという部分が類似している点であると考えられる．また，この研究で使用された BART や Transformer は本研究の各認識の精度向上に繋がる可能性があると考えられる．

## 3.　MAIA システム

本研究では画像認識及び文字認識を使用し，各認識から得た情報を基に背景の場所を推定，生成を行う MAIA(Manga AI Assistant) システムの開発を行う．

MAIA システムの主な流れとしては，初めに背景以外のキャラクターの絵やセリフ等が描かれた漫画画像を準備する．その画像を MAIA システムに入力し，その画像内にある情報から背景を出力する．以上が MAIA システムと流れとなっている．

MAIA システムのフローチャートを図 4 に示す．

図 4 を参考にシステムの概要を説明すると，初めに漫画の画像を入力し，その画像から画像認識，文字認識を用いて背景の場所を推定するために必要な情報の収集を行う．

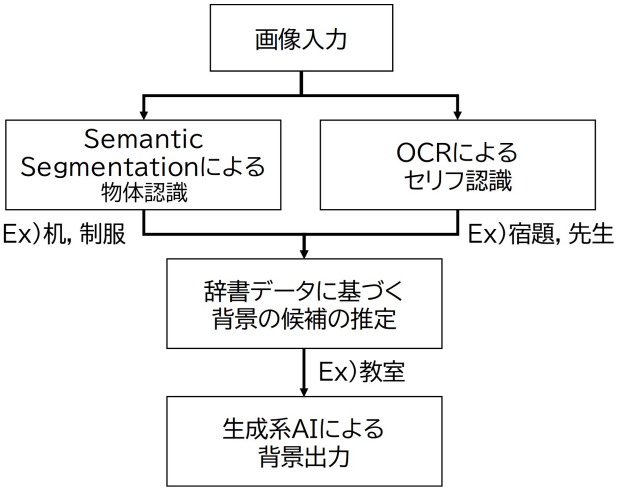


**図 4** MAIA システムフローチャート

各認識から情報を収集し終えた後，その情報から背景の場所の推定を行う．最後に推定した場所をテキスト形式で出力し，出力テキストを基に生成 AI を用いて背景画像を出力させる．

図 4 に示した MAIA システムの各要素の詳細な内容の説明を行う．尚, 本論文で使用される漫画画像は Manga109[4] のデータセットを許諾の元，使用，掲載を行っている．

### 3.1 画像認識

本研究で画像認識に使用する技術として semantic segmentation を用いる．semantic segmentation とは深層学習の 1 つで画像を指定のカテゴリに分類し，そのカテゴリをピクセル単位ごとに出力を行う手法のことを指す．semantic segmentation の学習データとして Manga109[4] の漫画画像をアノテーション [5] したものを画像データとする．semantic segmentation で得られる結果の例として机や椅子等の人力で可動可能な物体，人物やその人物が着用している服装等の情報を得ることが可能である．
semantic segmentation を用いた漫画画像の認識結果例を図 5 に示す

©ありさ 2（ありさのじじょう） 八神健

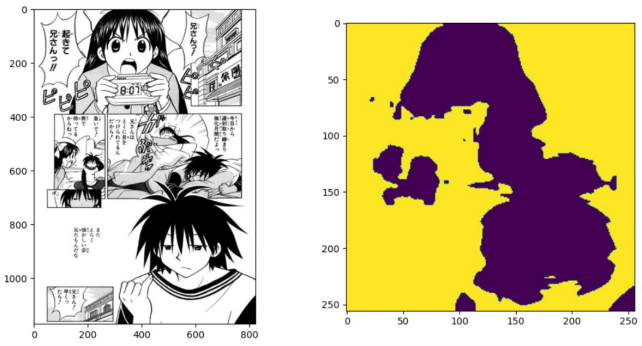


**図 5** semantic segmentation 実行結果

### 3.2 文字認識

文字認識だが，今回は OCR を使用し文字認識を行う．しかし OCR を使用する際漫画の絵が混入すると文字認識が誤認するといった問題点があると考えられる．このような問題点を解決するために次のような前準備を行う．

©蜜・リターンズ 八神健

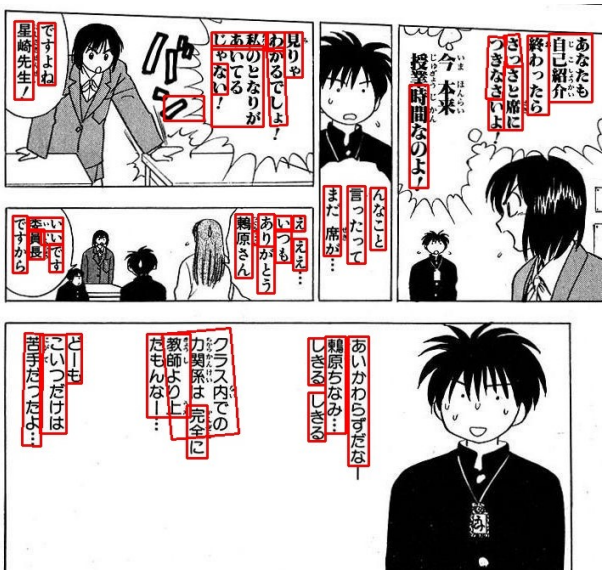

**図 6** CRAFT 実行後画像

初めに入力した漫画画像からセリフや効果音の出現する部分の座標位置を抽出する．抽出する方法として本研究では CRAFT[6] という文字部分の座標位置の取得を行うプログラムを活用する．図 6 画像にある赤枠の座標を取得を行う．座標位置を取得した後，その座標を基に画像の文字部分を切り取る ROI 抽出を行う．これらの前準備を終えた後 OCR を用いて，切り取り後の画像内の文字の認識を行う．文字認識を行う際の注意点として漫画のセリフは基本的に縦書きで構成されているため縦書きに対応した技術を用いる必要がある．このことから本研究では縦書きの日本語に対応している文字認識である Tesseract OCR[7] というものを使用している．

### 3.3 背景の場所推定

画像認識，文字認識をテキスト形式で出力し，その結果基に背景の場所推定を行う．場所推定の手法は TF-IDF を用いた特徴語辞書を作成し，辞書に記入されている特徴語と各認識で出力されたテキストが該当した単語を識別し，最も割合が高い場所の推定を行う．場所推定で作成した特徴語辞書は藤川らが行ったイベント特徴語に基づくコミックのイベント推定 [8] という研究に使用されている特徴語辞書を参考として制作している．

**表 2** 実験結果

| 結果 | 数値 |
|---|---|
| 元画像の場所と一致した背景画像 | 0.45 |
| 元画像の場所と一部一致した背景画像 | 0.09 |
| 元画像の場所と一致しなかった背景画像 | 0.45 |

### 3.4 生成 AI を用いた背景画像生成

背景の場所推定が完了した後，テキスト形式で出力し，出力された場所を基に生成 AI を用いて画像生成を行う．本研究では Bing Image Creator[9] という生成 AI を用いて背景画像の作成を行った．Bing Image Creator とは Microsoft 社が提供している画像生成 AI ツールである．作成する画像の特徴を記した指示文を入力し画像の生成を行う．図 7 は指示文を「日本で出版されるマンガの背景画像の生成をお願いします。画像は、線画でお願いします。作成してほしい背景のキーワードは、”教室”です」と入力した場合に出力された画像である．このように指示文に沿った画像を出力することができる．この生成 AI を Bing Image Creator API[10] で繋ぎ使用し入力した漫画画像に合う背景を出力を行う．

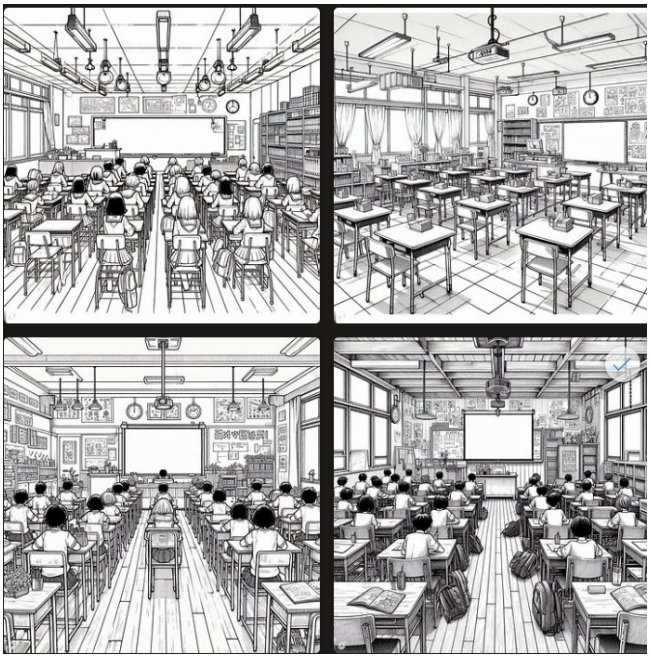

**図 7** Bing Image Creator 実行例

## 4. 実験

MAIA システムの出力画像の精度を調査するための実験を行った．MAIA システムに入力する漫画画像として Manga109[4] の画像を使用し，出力した背景画像が元の画像の背景の一致度でシステムの精度を測るものとする．

また本実験の実施に辺り, 大阪工業大学における倫理委員会の審査 (2023-35) に基づき実施した.

図 8 が入力画像の背景と出力画像の場所が一致した場合

©ジョバレ 白井三二郎


**図 8** 実験結果成功例：(左) 入力画像 (右) 出力画像

©ボクはしたたか君 新沢基栄

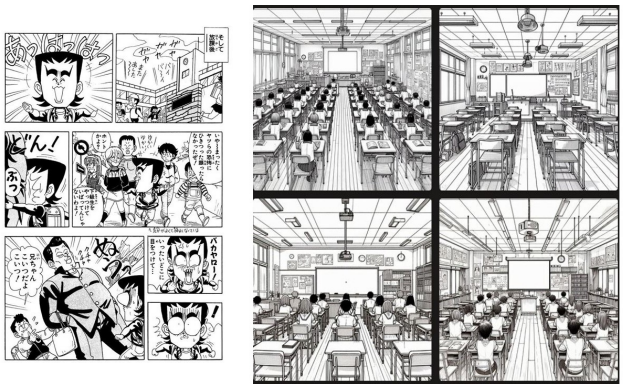


**図 9** 実験結果失敗例：(左) 入力画像 (右) 出力画像

の結果である．この場合，入力画像の背景に体育館のバスケットコートが描かれおり，出力画像にも似た体育館のバスケットコートの漫画画像が出力されている．反対に図 9 が入力画像の背景と出力画像の場所が一致しなかった場合の結果である．この場合，入力画像が歩道が描かれているのに対し，出力画像は学校の教室となっている．

表 2 の実験の結果から元の画像に描かれていた背景と一致した画像を出力した割合は約 4〜5 割．元の画像の背景と一致しなかった画像が出力された割合は 4〜5 割という結果となった．また元の画像の背景と一部一致したという結果も得られた．例えば，元画像が学校の廊下であったのに対し出力された画像が学校の教室出会った場合等を指す．

## 5. 考察

今回の実験結果から元画像と出力された画像の一致が成功したを割合は 4〜5 割程となり，精度に関して不安定であるという結果となった．この理由としては TF-IDF を用いた特徴語辞書内の単語や semantic segmentation の学習データの量が少ないことが関係していると考えられる．本実験に使用した semantic segmentation 学習データの枚数 121 枚，TF-IDF 特徴語辞書の単語の量は 778 単語となっ

ている．よってこれ以上の学習データが必要であると考えられる．

## 6. 今後の予定

今回の実験で挙げられた問題点の改善として，第一に semantic segmentation 及び TF-IDF 特徴語辞書のデータの増量を行い精度を向上具合を確認する．また他の案として TF-IDF を用いた場所推定の際，割合が高い場所を複数推定を行う．その推定した複数の場所の内 1 つを被験者が選択し，選択した場所の画像の生成を行うプログラムの作成を行う．これにより，被験者の希望に合う画像の生成を行うことが可能だと考えられる．

## 7. おわりに

本研究では漫画の自動生成システムである MAIA システムを制作することで漫画の作業の一つである背景作成の時間短縮に繋がると考え，実験から有用な情報を得ることができた．

**謝辞** 本研究の一部は, JSPS KAKENHI Grant Number JP19K20750 の支援を受けた．

### 参考文献

[1] アミューズメントメディア総合学院. 漫画家デビュー後の実態. マンガ業界情報局, (2017) https://www.amgakuin.co.jp/contents/comic/column/become/mangaka/debut-life/
[2] 中山雅紀, 野村芽久美, 藤代一成. Linda：漫画向けの建造物の背景線画半自動生成. 芸術科学会論文誌, Vol. 20, p. 210 – 218, 2021.
[3] 青木尚人, 森直樹, 岡田真. Vision transformer と bert を用いた漫画のマルチモーダル識別. 人工知能学会全国大会論文集, Vol. 36, 2022.
[4] MANGA109 Japanese Manga Dataset, (2022.08) http://www.manga109.org/ja/
[5] Labelme:Image Polygonal Annotation with Python (2023.5.17) https://github.com/wkentaro/labelme
[6] Clovaai/CRAFT-pytorch:Character-Region Awareness For Text detection (2019.12.4) https://github.com/clovaai/CRAFT-pytorch
[7] UB-Mannheim/tesseract. Tesseract OCR. (2022) https://github.com/UB-Mannheim/tesseract
[8] 藤川雄翔, 松下光範. イベント特徴語に基づくコミックのイベント推定. 電子情報通信学会第 9 回コミック工学研究会. p. 13 - 20, 2022.
[9] Microsoft Bing. Image Creator from Microsoft Designer https://www.bing.com/images/create
[10] acheong08/BingImageCreator. BingImageCreator. (2023.8.10) https://github.com/acheong08/BingImageCreator