

# コミックにおける読者依存性の高い地雷表現の VLMを用いた自動検出の検討

新畠 道大<sup>1, a)</sup> 櫻井 翼<sup>1</sup> 中川 由貴<sup>1</sup> 中村 聡史<sup>1</sup>

**概要:** コミックは多種多様な表現があるため、読者が大好きな作品であっても、そこに苦手とする描写が登場することがある。こうした表現の苦手意識は読者依存性が高いものであるため、規制するべきではないが、そこさえ読み飛ばせば問題なく読めるという読者も珍しくない。我々はこうした問題を踏まえ、読者が苦手な描写に怯えることなくコミックを鑑賞するための手法を実現してきた。しかし、これまでの手法では読者が一度は地雷と遭遇する必要があるなどの問題があった。そこで本研究では、VLMを活用した読者依存性の高い地雷表現の検出手法の検討を行う。ここでは特に、人体の異常にまつわる表現の自動検出にむけて、カテゴリごとの検出率について調査する。本研究では、コマ単位の画像をテキスト化し、そのテキストをもとに人体の異常にまつわる表現の有無や種類を判定・分類した。その結果、前のコマやページの情報を考慮した手法は、単一のコマのみを用いた手法と比較し、人体の異常にまつわる表現の検出における再現率が向上することが明らかになった。

**キーワード:** コミック, 読者依存性の高い地雷表現, VLM, LLM

## 1. はじめに

コミックは日本を代表する文化の一つであり、インプレス総研によると2023年度電子書籍市場のコミックのシェア率は87.6%で、市場規模は前年度から448億円増加の5647億円となっている<sup>\*1</sup>。このように市場拡大を続け、読者の多いコミック作品であり、その表現力により様々な体験・知識を得ることができるが、一方でそのコミックの表現力により、現実ではあまり遭遇したくないような表現や状況に遭遇することも珍しくない。

ここで、ある作品が大好きで読み続けているときに、先端恐怖や高所恐怖、共感性羞恥やグロテスク、動物がひどい目に遭うようなものや虫といったように、その読者が苦手とする表現が含まれているとき、その表現を恐れて読書意欲が低下することがある。日本ではこうしたものに対し「地雷」という表現が日常的に用いられ、不快に感じる描写を回避するための概念として定着しつつある。しかし、これらの苦手な表現は読者の個人的な苦手意識に基づく読者依存性の高い地雷といえるものであり、作品内で意図的に描かれたものであることも多く、その表現を規制することは適切ではない。

こうした問題を踏まえ、伊藤ら [1] はコミックにおける読者依存性の高い地雷を「不快に感じたり、苛立ちを覚えたりして受け入れられず、読むのを避けてしまう描写」と広く定義し、コミックをよく読む人々を対象にアンケート調査を実施してきた。その結果、回答者の約4割がコミック内に自身の地雷となる表現があると回答し、特にグロテスクなものや血液など、人体にまつわるもので嫌悪感を抱くとされる描写を苦手とする読者が多いことが明らかになった。また、読者同士が読者依存性の高い地雷表現に対するアノテーションを共有し、その表現が含まれるページの直前で警告を出すような手法を実現してきた [2]。しかし、これまでの手法では同一の読者依存性の高い地雷をもつ他者が、その作品を読み、地雷としてのフラグを付与してくれないと回避することができないという問題があった。そのため、Vision API を用いた地雷表現の自動検出の試みもあるが、そのデータセット自体に曖昧性があり、また手法として十分に考慮でなかったため詳細な描写の判断には誤判定が多く、より高度な分析手法が求められることが明らかになった [3]。

そこで本研究では、コミックのコマに読者依存性の高い地雷が含まれるかどうかを自動判定する手法の実現を目指す。ここでは、各コマをVLMを用いてテキスト化し、そのテキストをもとに読者依存性の高い表現であるかの判定を行う。この手法では、従来の視覚的特徴だけでは依存せ

<sup>1</sup> 明治大学

<sup>a)</sup> michimaru007.com@gmail.com

<sup>\*1</sup> <https://research.impress.co.jp/report/list/ebook/502008>

ず、画像から抽出されたテキスト情報を用いることで、関係性も考慮しながらコンテキスト依存の表現を識別する。ここでは特に、単一のコマのテキストのみを用いて判定を行うベースライン手法と、比較対象として、現在のコマの判定に前の数コマのテキストをコンテキストとして与える手法、前ページの全てのコマのテキストをコンテキストとして与える手法を提案して比較を行う。また、対象とする読者依存性の高い表現を、主に人体の異常にまつわるものに限定し、コンテキストを考慮することでそうした表現の検出精度が向上するかどうかを実験により検証する。

## 2. 関連研究

### 2.1 コミックに関する研究

コミックの大規模データセットである Manga109 データセット [4] [5] などのように、コミックは人々を楽しませるだけでなく研究の対象としても注目されている。Sakurai ら [6] はコミックに登場するセリフの話者推定などに向けて、セリフと発話者を対応付けるデータセットの拡張を行った。また、このデータセットを用いて各セリフの評価の分散度合いなどを分析し、アノテーション付与者の人数が2名の場合に比べ、5名の場合には約10%一致率が低下することを明らかにした。樋口ら [7] は、コミックの説明文やレビュー文などの情報源によって内容が異なることに着目し、出現単語を調査することでその特性を分析した。その結果、説明文はキャラクターの特徴などが多いためコミックの内容に関する分析に用いることが有用であり、レビュー文は作風などのメタ的情報が多いためトピック分類などに用いることが有用である可能性を示した。

本研究で検出する人体の異常にまつわるような読者依存性の高い地雷表現のデータは、地雷を事前に回避できるだけでなく、こうしたデータセット構築にも貢献できる可能性がある。

### 2.2 グロテスクな表現に関する研究

松尾 [8] は、裁判においてグロテスクな証拠提示が陪審員の判断に及ぼす影響について研究を行った。研究では、被害者の死体写真や傷の詳細な描写が陪審員に与える心理的影響を分析し、グロテスクな視覚情報が陪審員の不快感情を喚起し、有罪判断の確率を高める可能性が示された。また、カラー写真と白黒写真では影響に差があり、カラー写真の方がより強い感情を引き起こすことが分かった。さらに、陪審員自身は写真が判断に影響を与えたことを認識していないことが示唆され、グロテスクな情報が無意識のうちに判断を歪める可能性が指摘された。巴 [9] は、グロテスク描写を「死体の描写及び致命的な人体の損傷を描いたもの」と定義し、そのような描写の多いコミック作品を分析した。

本研究においても、Manga109 の作品をもとにグロテス

クの一部に分類される、人体の異常にまつわる読者依存性の高い地雷表現を定義しており、「暴力」「血」「人体損壊」「死体」の4つのカテゴリに分類している。

### 2.3 コミックの理解・解析に関する研究

Sachdeva ら [10] は、コミックのコマの自動転写生成に関する研究を行い、視覚情報をもとにセリフやキャラクターの識別を自動化する AI モデル「Magi」を提案した。コミックのページを入力とし、コマ、テキストブロック、キャラクターボックスを検出し、それらを正しい順序で並べてセリフを話者に関連付ける手法を開発した。また、Magi は OCR 技術を用いてテキストを抽出し、セリフを適切に整理する機能を備えている。さらに、コミック特有のレイアウトの影響を考慮し、ページ内の情報を適切に処理するアルゴリズムを開発した。これはコミックのテキスト情報を自動的に抽出・整理する技術の発展に寄与するものであり、コミック解析やデジタルアーカイブの分野において重要な知見を提供する。Nguyen ら [11] はコミックの画像解析を効率的に行うために、コマ検出、キャラクター識別、吹き出しのセグメンテーション、吹き出しとキャラクターの関連付けといった複数のタスクを統合したマルチタスク学習手法を提案した。DCM772 および eBDtheque というデータセットを用いた評価を行い、従来手法よりも高精度かつ効率的にコミックの構造を解析できることを示した。

本研究では、コマ単位でのテキスト化を行うが、コミックのレイアウト解析やキャラクターの話者推定といった技術は、人体の異常の自動検出にも応用可能である。特に、コマ内のセリフやテキストブロックを正確に抽出し、キャラクターの発話と結びつけることで、コンテキストを適切に考慮した判定が可能になると考えられる。

Guo ら [12] は、コミックのストーリーを補完するため、視覚情報とテキスト情報を統合し、文脈を考慮した内容を生成する「Multimodal Manga Complement」というタスクを提案した。特に LLM を活用し、画像とテキストの統合的な解析を行う FVP-M2 手法を導入することで、不足している情報を推定し、コミックの内容をより自然に補完できることを示した。また Vivoli ら [13] は、コミックのテキスト補完に焦点を当てた手法を提案した。コミックの吹き出しテキストを予測するタスクを設定し、従来のリカレントニューラルネットワークを用いた手法と比較して、LLM を用いたアプローチが10%以上の精度向上を達成できることを示した。

これらの文脈やテキスト補完を組み合わせることで、前後のコマの文脈を考慮しつつ、適切なテキスト化を行い、人体の異常にまつわる表現の判定精度を向上させることが期待できる。

表 1: 各コミックのコマ数, 人体の異常にまつわる画像の割合および分類

	総コマ数	割合	暴力	血	人体損壊	死体
あっけら貫刃帖	975	2.4 %	20	3	0	0
ベルモンド Le VisiteuR	909	5.1 %	10	1	35	0
ぶらり鉄扇捕物帳	1335	3.4 %	16	22	0	7
ドールガン	864	3.9 %	29	5	0	0
黒井戸眼科	1029	3.3 %	8	15	10	1
マッド・ストーン	618	9.9 %	23	38	0	0
魔童タイガ	670	2.4 %	10	3	3	0
征神記ヴァルナス	942	7.0 %	28	28	5	5
藤太参ります!	891	3.9 %	30	5	0	0
花影戦記 妖魔降臨	950	4.2 %	19	10	5	5
平均	918	4.5 %	19.3	13	5.8	1.8

### 3. 人体の異常にまつわる画像データセット構築

本研究では, コミックにおける人体の異常にまつわる表現の自動検出を行うため, 109 冊のコミックについて, アノテーションとともに公開されている Manga109 データセット [4] [5] を用いる. また, Manga109 データセットの中から対象とする要素を含む作品を著者が選定し, 画像データを用いてアノテーションを行った.

#### 3.1 データセットの構築

本研究では, Manga109 データセットの中から, 人体の異常にまつわる表現を持つコミック 10 作品を選定した. 選定作品は以下の通りである.

- 小林ゆき「あっけら貫刃帖」
- 石岡ショウエイ「ベルモンド Le VisiteuR」
- 佐々木あつし「ぶらり鉄扇捕物帳」
- 出口竜正「ドールガン」
- 平雅巳「黒井戸眼科」
- 正木秀尚「マッド・ストーン」
- 井萩寿一「魔童タイガ」
- 島崎譲「征神記ヴァルナス」
- 西条真二「藤太参ります!」
- 花影戦記「花影戦記 妖魔降臨」

選定基準として, 戦闘・バトル要素が強い作品, 怪奇的な要素を含む作品, 医療や人体改造など身体的異常を描く作品とした.

#### 3.2 人体の異常にまつわる表現の定義

本研究では, Manga109 の作品に含まれる読者依存性の高い人体の異常にまつわる表現を体系的に分類し, 以下の 4 つのカテゴリとして定義した.

- (1) **暴力**: 暴力行為が具体的に描写されており, それによってダメージが発生している (例: 素手, 武器, 火, 光線, 投石などが人体に命中する).

- (2) **血**: 血が人体から出ている, または体に付着している描写がある (ただし鼻血は除く). かつそれが明確に「血」であると判断できる.

- (3) **人体損壊**: 現実ではほぼ起こりえない人体の変化が描写されている (例: 目が飛び出る, 体が溶ける, 関節がおかしくなる, 体がバラバラになる, 身体が異常に変形するなど).

- (4) **死体**: 描写から明らかに死体であると判断できる状態が含まれている.

上記の 4 つのカテゴリに分類した理由は, Manga109 の作品を精査した結果, これらの表現が人体の異常にまつわるものとして描かれることが多く, かつ明確な分類が可能であったためである.

本研究では, これらの定義と分類をもとに, 各コマ単位で人体の異常にまつわる表現の有無とカテゴリをアノテーションし, 正解データを作成した. このデータセットを使用し, VLM を用いた自動検出の評価を行う. 表 1 に作品ごとの集計結果を示す.

### 4. 人体の異常にまつわる画像判定手法

人体の異常にまつわる表現は, 視覚的な要素だけでなく, 物語の文脈や登場キャラクターの発言が影響を与えることが多いため, 画像認識のみでは適切に判定することが困難である. また, Vision API を用いた地雷表現の自動検出 [3] では誤判定が多く, 精度が低かった. そこで本研究では, コマ内の内容を VLM を用いて, 言語情報としてテキスト化し, そのテキスト情報を LLM を用いて周辺も含め解析することで, 人体の異常にまつわる表現を判定する手法を提案する. 手法の全体図を図 1 に示す. テキスト化の手法として, コマ単位でテキスト化を行う手法を採用する. まず, コミックのページ画像から各コマを順番にトリミングし, コマ単位の画像データを作成する [14]. これにより, 各コマ画像を入力として用いてコマごとの詳細な情報を取得できるようにする.

次にトリミングしたコマ画像を VLM を用いてテキスト

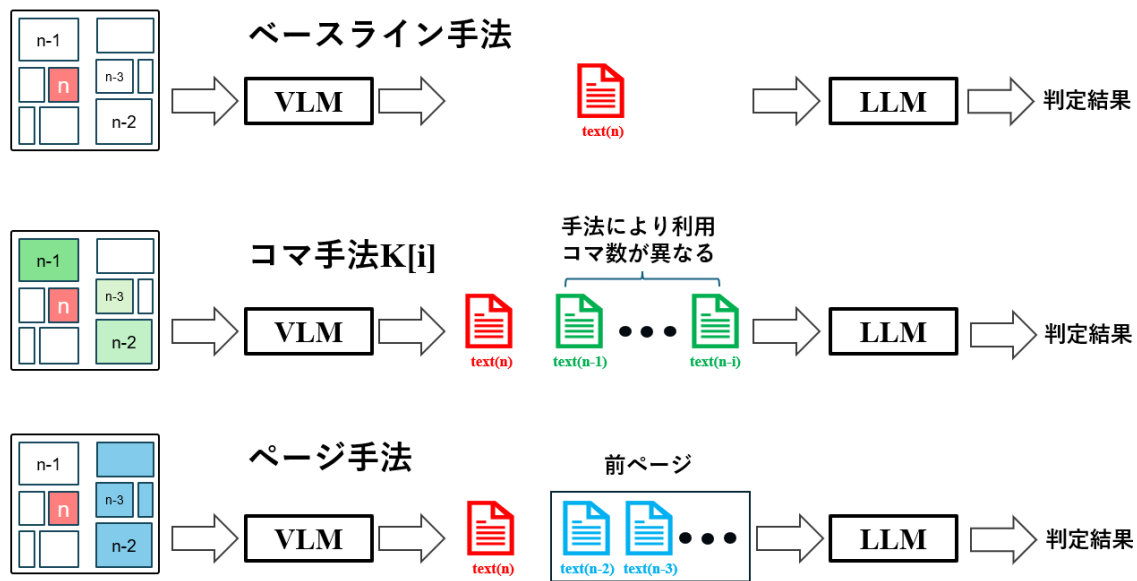


図 1: 手法の全体図

化する。プロンプトは特定の表現に偏ることなく、コミックのすべてのコマに用いることができるものを用いる。テキスト化されたコマの情報をもとに以下に示す3つの手法で、LLMを用いた自動判定を行う。

**ベースライン手法：** 対象とするコマのテキストのみをもとに、その表現の有無を判定する。

**コマ手法：** 対象のコマのテキストに加えて、そのコマの直前の数コマのテキストをコンテキストとして与える手法。

**ページ手法：** 対象のコマのテキストに加えて、1つ前のページに含まれるすべてのコマのテキストをコンテキストとして与える手法。

コマ手法とページ手法では、前後の関係を理解しやすくすることで、単なる静止画像としてではなく、物語の流れを考慮した検出が可能になると考えた。コマ手法では1～5個前までのコマをコンテキストとして与える計5パターンの検出を行い、それぞれコマ手法 K1～K5 とする。

また、人体の異常にまつわる表現の判定において、3章で述べたように「暴力」「血」「人体損壊」「死体」の4つのカテゴリに分類し、各コマのテキストがこれらのいずれかに該当するかを判定する。人体の異常にまつわる表現の判定には、LLM に対して定義した基準に基づく判定プロンプトを与え、そうした表現であると判断された場合には該当するカテゴリを出力し、複数のカテゴリに該当する場合は、それらを全て出力させる。LLM および VLM には GPT-4o API [15] を使用した。

VLM および LLM に与えたプロンプトを表 2 に示す。LLM のプロンプト内には VLM で出力した結果を挿入して利用している。

表 2: VLM と LLM のプロンプト

<b>VLM</b>	この画像は漫画の一コマです。内容を詳細に説明してください。キャラクターが何をしているか？何をされているか？そして何が映っているかを詳しく出力してください。またオノマトベがあれば出力してください。
<b>LLM</b>	<p>あなたはグロテスクの専門家です。以下の基準に基づいて、現在のコマのテキストにグロテスクな要素が含まれているかどうかを判定してください。グロテスクな描写が含まれている場合は 1、そうでない場合は 0 と出力し、該当する要素も出力してください。</p> <p>またグロテスクな要素が含まれる場合は該当する要素を出力してください。その時、グロテスクの要素が重複しても大丈夫です。また他の事は一切出力しないでください。</p> <p><b>判定基準</b></p> <p><b>暴力描写：</b> 以下条件を満たす場合、「グロテスク」と判断してください。暴力行為が具体的に描写され、ダメージが発生している（例：素手、武器、火、光線、投石など）。</p> <p><b>血の描写：</b> 以下の条件を満たす場合、「グロテスク」と判断してください。血が人体から出ている、または体に付着している描写がある（ただし鼻血は除く）。血が明確に「血液」であると判断できる。</p> <p><b>人体の異常：</b> 以下の条件を満たす場合、「グロテスク」と判断してください。現実ではほぼ起こりえない人体の変化が描写されている（例：目が飛び出る、体が溶ける、関節がおかしくなる、体がバラバラになる、身体が異常に変形するなど）。</p> <p><b>死体の描写：</b> 以下の条件を満たす場合、「グロテスク」と判断してください。描写から明らかに死体であると判断できる状態が含まれている。</p> <p>この基準に基づき、以下の【過去のコマ】の情報を考慮した上で【現在のコマ】の内容を分析してください。</p> <p>【現在のコマ】 現在のコマに関する VLM で出力した説明を挿入</p> <p>【過去のコマ】 過去のコマに関する VLM で出力した説明の羅列を挿入</p>

## 5. 評価実験

### 5.1 実験の方法

構築したコミックの人体の異常にまつわる表現データセットを用いて、ベースライン手法、コマ手法、およびページ手法の判定精度を比較した。ここでは、評価指標として正答率、適合率、再現率、F値を用いた。なお、正答率は「入力されたコマが人体の異常にまつわるかそうでないかを正しく判定できた割合」、適合率は「人体の異常にまつわると判定したもののうち、それが正解である割合」、再現率は「正解ラベルが人体の異常にまつわるものであるもののうち、それを正しく判定できた割合」を表す。

読者依存性の高い地雷の自動検出においては、できる限りその読者依存性の高い地雷を網羅的に検出し、読者が事前に回避できるようにすることが求められる。そのため、本研究では評価指標のうち再現率を特に重視する。この再現率を高めるつつも、読者が避けたいと考えるその表現を見逃さないようにするため、適合率とのバランスを考慮した評価を行う。

### 5.2 実験の結果

人体の異常にまつわる画像データセットの全データを用いてベースライン手法、コマ手法、ページ手法のそれぞれで判定を行った結果を図2に示す。図の縦軸は各評価指標の平均値を示している。図2より、再現率はページ手法が最も高くなっている一方で、適合率はページ手法が最も低い値となっていた。また、前のコマを使わないベースライン手法が最も精度が悪い一方で、1つ前のコマを使うコマ手法K1がもっとも精度が高いという結果となっている。

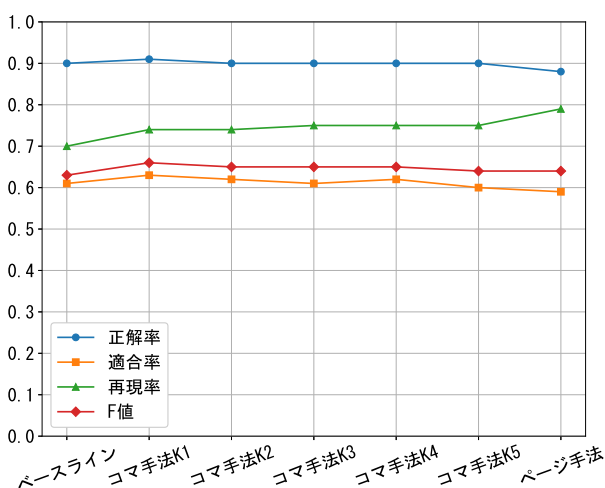


図2: 全データに対する手法ごとの精度比較

次にカテゴリごとの再現率を図3、判定結果がカテゴリまで正しく一致した場合の再現率を図4に示す。図3より、

暴力はカテゴリのなかで最も高い値になった。また図4より、カテゴリ厳密一致時の再現率は図3と比べ低くなっており、カテゴリの判定が容易でないことがわかる。

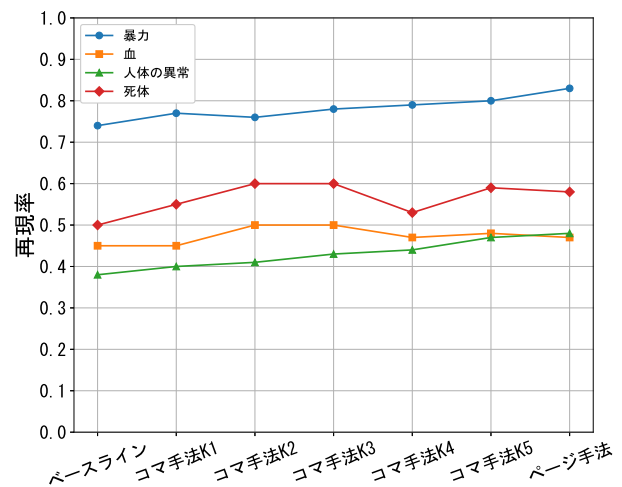


図3: 各カテゴリの再現率の比較

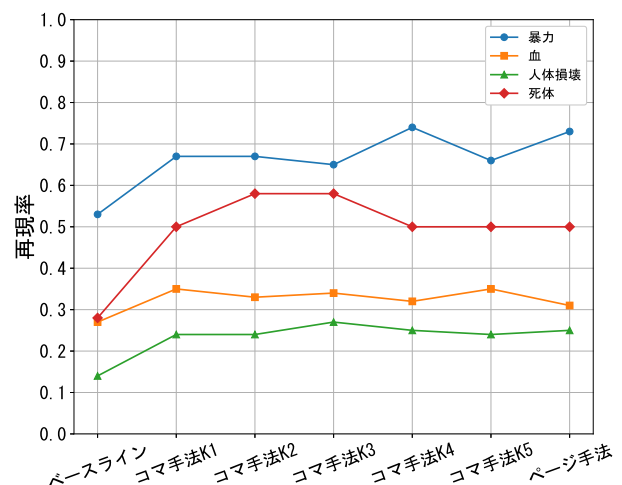


図4: カテゴリ厳密一致時の再現率の比較

## 6. 考察

### 6.1 手法の違いにおける精度差について

ベースラインでは、正答率が高いが、再現率は最も低い結果になった(図2)。これは、本手法が単一のコマのみを対象として判定を行うため、文脈情報を考慮できないことが要因と考えられる。特に、人体の異常にまつわる表現が前後のコマの流れの中で明確になる場合において、ベースライン手法では正しく判断できず、誤判定するケースが多く見られた。

コマ手法では、前後のコマ情報を考慮することで、ベースライン手法と比較して再現率、F値が向上した(図2)。

特に、コマ手法 K1 では、適合率と再現率のバランスが取れた結果となり、F 値が最も高くなった。これは、直前の 1 コマの情報を付加することで、対象とする人体の異常にまつわる表現が前後の流れの中でより正確に把握されるようになったためと考えられる。コマ手法 K2, コマ手法 K3, コマ手法 K4, コマ手法 K5 では再現率がさらに向上したものの、適合率の低下が見られ、F 値はコマ手法 K1 と比較して低くなった。特に K5 では誤検出が増加し、適合率が大きく低下した。これは、直前の 5 コマまでの情報を考慮することで、関連性の低い情報が含まれ、誤った判定が増加した可能性がある。

ページ手法では再現率が最も高い値を示した (図 2)。その一方で、適合率が低く、F 値もベースライン手法の次に低くなった。これは、ページ全体の情報を用いることで、人体の異常にまつわる表現が前後の流れの中でより検出されやすくなる一方で、それ自体が余計な情報となってしまふことにより、誤検出が増加したものと考えられる。特に、ページ単位での解析では、複数のコマが同時に考慮されるため、本来そうした表現ではないコマも周囲の描写に影響を受け、誤判定されやすくなっていた。そのため、ページ手法では、人体の異常にまつわる表現を検出する上では有効であるが、誤判定を抑制する適切なフィルタリングが必要であると考えられる。

## 6.2 カテゴリ別の検出精度の差について

4 つの人体の異常にまつわる表現のうち、暴力表現が最も検出結果が高かった。また、コマ手法、ページ手法の両方で検出精度がベースラインと比べて高い結果となった。これは前後の情報を考慮したことによって、攻撃動作をより正しく識別できたからであると考えられる。その一方で、問題のない暴力描写が誤判定される傾向があった。図 5 に暴力と誤判定されたコマの例を示す。キャラクターにナイフが向けられているコマは、まだ攻撃が命中しておらず、ダメージの描写は無い。しかし、LLM の判定では武器を向けていることが暴力表現とみなされ、誤判定されていた。また、キャラクターの前に剣が刺されているコマも、威嚇的な描写であり、本来特に問題のないものであるが、このような状況でも誤判定されていた。このように、攻撃動作や威嚇的な描写が、LLM の判定において人体の異常にまつわる表現であると誤認識される傾向があった。

血の表現は、カテゴリの中で適合率が高かった。これは、血の描写が比較的明確な視覚的特徴を持つため、明らかな血の描写を正しく判定できたからであると考えられる。一方で、再現率がかなり低く、血の見逃しが多く存在していることがわかる。図 6 に検出されなかった血のカテゴリの例を示す。女の子の顔に血が飛び散っているコマでは、LLM によるテキスト化において黒いインクのようなものと解釈され、血であると正しく認識されなかった。また血

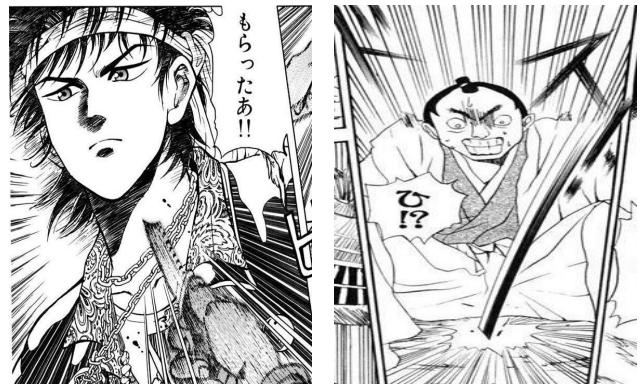


図 5: 暴力表現と誤判定されたコマ

© 正木秀尚「マッド・ストーン」

© 小林ゆき「あっけら貫刃帖」

ではないものが、血と解釈されたケースもあった (図 7)。これは、LLM が特にモノクロの絵において、血と他の液体や汚れを区別することが難しいためと考えられる。血の表現は手法間で精度の差が見られなかったことから、コンテキストは意味をなさない可能性が考えられる。



図 6: 検出できなかった血の表現

© 佐々木あつし「ぶらり鉄扇捕物帳」

© 島崎譲「征神記ヴァルナス」

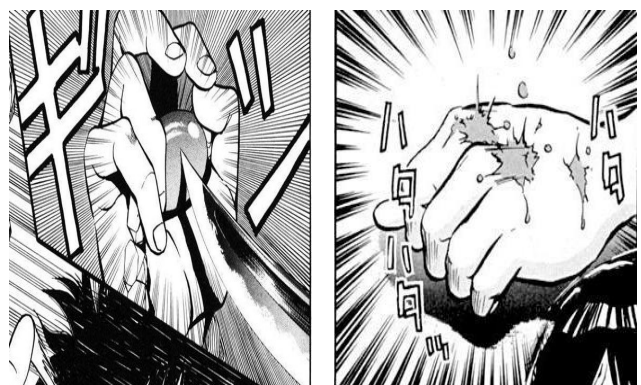


図 7: 誤って血の表現と検出されたコマ

© 正木秀尚「マッド・ストーン」

© 西条真二「藤太参ります!」

カテゴリの中で、人体損壊が再現率が最も低い結果となった。これは人体損壊という特殊な描写が、うまくテキスト化されていないことが原因として考えられる。例えば、図8の「ベルモンド Le VisiteuR」では体がバラバラになる描写が多くあるが、その描写を正しくテキスト化しているものは少なかった。また図8の右側のコマのテキスト化では、「この漫画の一コマには、キャラクターが足と体をロープで縛られている様子が描かれています。キャラクターは壁に寄りかかりながら、驚いた表情で叫んでいます。背景にはレンガの壁と木の床が見えます。」と出力され、体がバラバラになっていることがテキストとして表れていなかった。この人体損壊のテキスト化がうまくいかなかった理由として、VLMが現実では起こりえないコミックならではの表現を、正確に理解することができないからであると考えられる。

死体も適合率が低かったが、これは単なる倒れているキャラクターと区別がつかなかったことが原因として考えられる。一方、コマ手法、ページ手法ともにベースラインと比べて検出精度が向上していた。このことから、コンテキストを考慮することで、死体と単なる倒れているキャラクターとの区別がしやすくなったと考えられる。

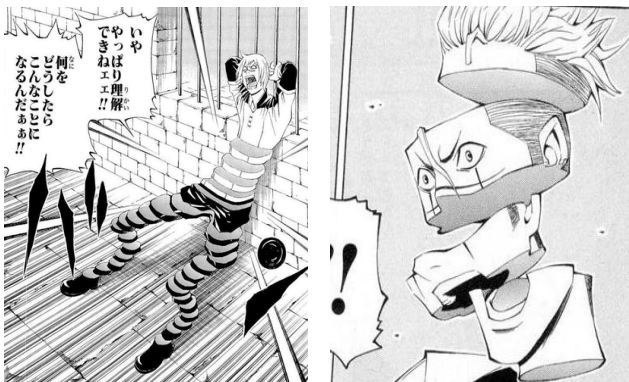


図8: 検出できなかった人体の損壊の表現

© 石岡ショウエイ「ベルモンド Le VisiteuR」

### 6.3 テキスト判定における誤判定の問題

本研究では、コミックのコマをVLMを用いてテキスト化し、そのテキスト情報をもとに人体の異常にまつわる表現の有無を判定する手法を行った。しかし、評価実験の結果テキスト化されたコマの説明文に「血」や「死体」といった単語が含まれる場合、それだけで人体の異常にまつわる表現と判断され誤判定となるケースが確認された(図9)。例えば、キャラクターが「滴たる血を眺めていたかった」と発言しているコマでは、視覚的には血液の描写は含まれておらず、単なるセリフとして使われている。しかしLLMによる判定では、この「血」という単語が含まれるために誤って分類されていた。また、「血のシミが残っている」と

いうセリフがあるコマでは、実際には血の描写は無く、テキスト化された情報のみに基づいて判定されていた。



図9: テキストにより誤判定されたコマ

© 石岡ショウエイ「ベルモンド Le VisiteuR」

© 佐々木あつし「ぶらり鉄扇捕物帳」

以上のように、事件の説明や状況描写に関するセリフの中で「死体」や「死んだ」といった単語が含まれているだけで、LLMが人体の異常にまつわる表現であると誤判定していたのは興味深い。ミステリー作品やサスペンス作品では、事件に関する説明として「死体」や「血」などのテキストが頻繁に登場する。そのため、視覚的には問題のないコマであってもLLMによる誤検出が発生しやすいことが考えられる。この問題の要因として、LLMが単語に過度に依存した判定を行っていることが挙げられる。この問題を解決するために、人体の異常にまつわる表現判定のプロンプトにおいて、セリフや内言、ナレーションとしてのテキストと、LLMがコマの内容を説明したテキストを明確に区別する必要があると考えられる。

### 6.4 テキスト化の際のVLMの問題

VLMを用いて、コミックのコマをテキスト化する際に、VLMが過度にグロテスクまたは性的な描写を含むコマに対して、適切なテキスト化を行わない問題が確認された。特に暴力的なシーンにおいて流血や切断といった描写が顕著なコマ、または性的な要素を含むコマでは、VLMがテキスト化を回避する、もしくは内容を大幅に省略するケー

スが多発した。この問題は、VLM がコンテンツの安全性に配慮する設計になっているため、極端な表現をさけるような動作をすることが要因と考えられる。これにより、人体の異常にまつわる表現の検出精度に影響を与えた可能性があり、テキスト化の段階で重要な視覚情報が省略されると、その後の判定において誤検出や見逃しが生じてしまうことになった。

## 6.5 今後の展望

本研究では、人体の異常にまつわる表現の自動判定において、文脈情報を活用するテキスト化手法の有効性を示したが、テキストからの判定において「血」や「死体」といった単語を含むと誤判定される場合があることが明らかとなった。この問題点を改善するために、LLM の判定時のプロンプト設計において、コマの描写の説明とコマ内の文字情報を明確に区別する方策を検討する。また、今回の手法では、判定対象とするコマの前方のみの情報をコンテキストとして与えたが、前後両方のコマの情報を利用することで、さらなる検出精度向上の可能性が期待される。

テキストベースの地雷表現検出手法は、人体の異常にまつわる表現に限らず、他の地雷表現の自動検出にも応用可能であると考えている。各種地雷表現は、作品ごとのジャンルや文脈によりその意味が大きく異なるため、同様にテキスト情報を詳細に抽出し、文脈を踏まえて判定する手法が有効であると期待される。

## 7. おわりに

本研究では、LLM を用いたコマ単位のテキスト化により、物語の文脈やコンテキストを踏まえ、人体の異常にまつわる表現の自動検出を行った。実験結果からは、前後のコマやページ全体の文脈を考慮する手法（コマ手法およびページ手法）が、単一コマのみを用いたベースライン手法に比べて再現率および全体の検出精度を向上させることが確認された。

しかし、一部の誤判定や見逃しも課題として残されており、今後はプロンプト設計のさらなる最適化や、視覚情報とテキスト情報の統合的解析による誤検出の抑制に取り組む必要がある。そして、これらの自動検出技術を活用して、地雷を事前に回避できるシステムに組み込み、地雷表現の登録を人間が行う手間を省く仕組みを構築することを目指す。

## 謝辞

本研究の一部は JSPS 科研費 JP22K12338 の助成を受けたものです。

## 参考文献

- [1] 伊藤理紗, 中村聡史: コミックにおける読者依存の地雷表現に関する基礎検討と軽減手法の検討, *DEIM2021*, Vol. 11, No. D13-1, pp. 1–8 (2021).
- [2] 伊藤理紗, 中村聡史: コミックにおける読者依存性の高い地雷表現回避手法の実現, *HCI*, Vol. 7, No. 39, pp. 1–7 (2022).
- [3] Nakagawa, Y., Ito, R. and Nakamura, S.: A Study on Anxiety Reduction of Reader-dependent “Jirai” Expressions in Comics, *Procedia Computer Science*, Vol. 246, pp. 3918–3927 (2024).
- [4] Fujimoto, A., Ogawa, T., Yamamoto, K., Matsui, Y., Yamasaki, T. and Aizawa, K.: Manga109 dataset and creation of metadata, *Proceedings of the 1st international workshop on comics analysis, processing and understanding*, pp. 1–5 (2016).
- [5] Aizawa, K., Fujimoto, A., Otsubo, A., Ogawa, T., Matsui, Y., Tsubota, K. and Ikuta, H.: Building a manga dataset “manga109” with annotations for multimedia applications, *IEEE multimedia*, Vol. 27, No. 2, pp. 8–18 (2020).
- [6] Sakurai, T., Ito, R., Abe, K. and Nakamura, S.: A Method to Annotate Who Speaks a Text Line in Manga and Speaker-Line Dataset for Manga109, *International Conference on Pattern Recognition*, Springer, pp. 22–33 (2022).
- [7] 樋口亮太, 山西良典, 松下光範: 単語の頻度と意味に基づいたコミックに関するテキスト情報源の特性分析, *コミック工学*, Vol. 5, No. E21-2, pp. 23–27 (2022).
- [8] 松尾加代: 感情を喚起する証拠提示が陪審員の判断に及ぼす影響: グロテスクな写真と被害影響陳述 (海外での心理学研究の紹介), *法と心理*, Vol. 10, No. 1, pp. 148–152 (2011).
- [9] 巴亜沙美: 漫画におけるグロ描写の比較, 筑波大学情報学群卒業論文 (2012).
- [10] Sachdeva, R. and Zisserman, A.: The manga whisperer: Automatically generating transcriptions for comics, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12967–12976 (2024).
- [11] Nguyen, N.-V., Rigaud, C. and Burie, J.-C.: Comic MTL: optimized multi-task learning for comic book image analysis, *International Journal on Document Analysis and Recognition (IJDAR)*, Vol. 22, pp. 265–284 (2019).
- [12] Guo, H., Wang, B., Bai, J., Liu, J., Yang, J. and Li, Z.: M2C: towards automatic multimodal manga complement, *arXiv preprint arXiv:2310.17130* (2023).
- [13] Vivoli, E., Lafuente Baeza, J., Valveny Llobet, E. and Karatzas, D.: Multimodal transformer for comics text-cloze, *International Conference on Document Analysis and Recognition*, Springer, pp. 128–145 (2024).
- [14] Ikuta: Panel-order-estimator, オンライン (2022). 入手先 <<https://github.com/manga109/panel-order-estimator/>> (参照 2025-02-19) .
- [15] OpenAI: GPT-4o API Documentation, オンライン (2024). 入手先 <<https://platform.openai.com/docs/>> (参照 2025-02-19) .