小説読書中の視線移動に基づく挿絵自動提示システム

西ノ原司瑳[†1]　河野恭之[†1]

概要：本研究では，ディスプレイを用いて小説を読む際の読者の視線移動に基づき，読者が読んでいる小説の場面に適切な挿絵を自動で提示するシステムを提案する．近年，人々の読書離れが進んでいる．読書が嫌いな理由は「文字が多い」「読んでいくにつれて集中力がないので続かない」というものがある．本研究では，読書中に挿絵を提示することで読書支援が可能であり，この課題を解決することが可能であると考え，画像生成 AI を利用して小説の挿絵を生成し，小説の場面が切り替わるたびに挿絵を提示するシステムを開発し，読書中の集中力の向上や読者への没入感の向上を目指す．

1. はじめに

本研究では，ディスプレイを用いて小説を読む際の読者の視線移動に基づき，読者が読んでいる小説の場面に適切な挿絵を自動で提示するシステムを提案する．近年，人々の読書離れが進んでいる傾向が指摘されている．平成 30 年度に文化庁が全国の 16 歳以上の男女 3590 人に行った調査では，1 ヶ月に本を 1 冊も読まないと回答した者が 47.3%に上がることが示された[1]．また，読書が嫌いな理由は「文字が多い」「読んでいくにつれて集中力がないので続かない」というものがあり[2]，特に文字のみで構成されたコンテンツは読者にとって心理的な負担となることが明らかになっている．小説は物語の情景やキャラクターを詳細に読むことが求められるため，このような負担が特に大きい．この課題に対し，読書中に挿絵を提示することで読書支援が可能であり，課題解決に寄与すると考える．和田[3]は心理学の授業を受講している生徒に対して実験を行い，挿絵が読者に対して与える効果は存在しており，挿絵が読書体験において重要な役割を果たすことを示した．しかし Web 上の投稿作品を含む全ての小説に適切な挿絵を手作業で用意することは，時間的・経済的に非効率で困難である．

この問題を解決するために本研究では画像生成 AI を活用する．画像生成 AI とは，与えられたテキスト(プロンプト)を基に内容を視覚的に表示した画像を生成する技術である．図 1 に提案システムの概要を示す．この画像生成 AI を活用し，読者が読んでいる場面に合わせて自動的に小説の挿絵を生成・提示し，読者が読んでいる小説の場面が切り替わるたびに挿絵を切り替えるシステムを開発することで読書中の集中力の向上や読者への没入感の向上を目指す．

本研究の対象は英語で記述された小説とする．英語は国際的に広く使用されている言語であるため，広く国際的な利用を想定した基盤を構築することが可能である．さらに，英語の小説を対象とすることで，自然言語処理技術や画像生成 AI の活用が容易となり，解析精度の向上が見込まれる．GPT-4[4]や spaCy[5] などの大規模言語モデルおよび自然言語処理ツールは英語テキストを高精度で解析できる．これにより，小説本文の文脈解析や場面変化の抽出を高精度に行うことが可能である．

図 1　提案システム

2. 関連研究

これまでに読者に対する読書支援を目的としたさまざまな研究が行われている．今井[6]は電子書籍を読んでいる最中の視線を追跡し，読者が読書を中断した場合に読者に対して読み始め地点の文字の色を変化させることによって情報提示する手法を提案している．この研究により読書という題材において視線追跡インターフェースの有用性・可視性が確認できた．

上野ら[7]は，自然言語処理を用いて日本語の文章を解析し，小説文中の単語や感情に対応した映像や背景色を VR 画面上で提示するシステムを提案している．しかし，この研究では小説テキストから抽出した名詞を検索した画像を読者に提示しており，小説の内容に不適切な画像が提示さ

†1 関西学院大学

れる可能性がある．

村山ら[8]は Word2Vec[9]を用いた挿絵自動挿入手法を提案している．この手法では，ロイヤリティーフリー素材を用いて画像と説明文のデータベースを構築し，入力文章に対して Word2Vec を利用して類似度を計算し，最適な画像を検索・提示している．しかし，小説の内容と適合しない画像が含まる場合があり，精緻化効果を阻害する可能性が指摘されている．

齋藤[10]は小説の要約文を基に挿絵を生成する手法を提案しているが，要約文の生成精度が挿絵の品質に直接影響を与える点が挙げられる．要約文が小説の文脈や内容を十分に反映していない場合，不適切な挿絵が生成される可能性が増大する．特に，登場人物の詳細な外見や場面特有の情景描写など、要約では省略される情報が欠落し，挿絵が小説の内容と乖離する可能性がある．

西川ら[11]は読者が文字を視認するタイミングで効果音を提示することで特に短い文や会話文に限り効果があることを示した．しかしこの研究で文章の理解度と没入感を調査した結果，優位な差は見られなかった．よって，本研究では視覚情報から読書支援を行う．小説中の場面が切り替わるごとに挿絵を生成し読書に反映し，読者が読書中に挿絵を見ることで読書への没入感を高め，読書効率の向上を目指す．

3. 提案システム

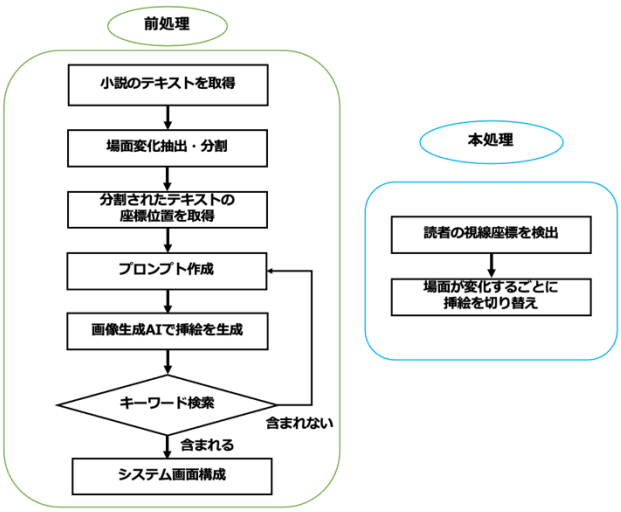

図 2　本システムの流れ

本システムの流れを図 2 に示す．本研究では，読者の小説読書中の視線移動を計測し，計測した視線先のテキスト情報を基に画像生成 AI を用いて挿絵を自動的に生成する．挿絵はディスプレイ上で小説テキストの右側に表示され，小説の場面に応じて適切に更新される．読書中の読者の視線をリアルタイムで計測することで読者が読んでいる場面に応じて挿絵を切り替える．視線入力機器はディスプレイ下部に装着することでユーザの視線を計測する．また，挿絵の生成を行う画像生成 AI は Stable Diffusion[12]を使用する．Stable Diffusion は，入力したテキストを基に画像を生成する Text2Image モデルである．

3.1 場面変化抽出

読者が読む小説のテキストを取得し，吉原ら[13]の場面変化パターンを参考に，自然言語処理を活用した場面変化の抽出を行う．吉原らが定義した場面変化の抽出パターンは，文章の係り受け関係に基づき，以下の 4 種類に分類される．

- 場面変化，場面属性ともに抽出に条件がないパターン
- 場面変化の抽出に条件はないが，場面属性の抽出には条件があるパターン
- 場面変化の抽出に条件はあるが，場面属性の抽出には条件がないパターン
- 場面変化，場面属性ともに抽出に条件があるパターン

ここで，場面変化とは場所や状況が切り替わることを指し，一枚の挿絵で情景を表す際の単位として捉える．また，場所属性とは場所の様子 (人・物など) や，場所同士のつながり (位置関係や包含関係など) や場所自体の様子を表す要素であり，吉原らはこれを場面変化に付随する重要な情報として定義している．この 4 種類の抽出パターンを活用して場面変化の抽出を行う際には自然言語処理ツールである spaCy を使用する．1 ページ中に複数の場面転換が存在する場合があり，これらを適切に識別・分割する必要がある．文章の係り受け関係に基づいて抽出パターンを作成し，小説テキストに適用することで場面変化の抽出を行う．検出された場面変化箇所でテキストを分割し，場面ごとの挿絵をディスプレイ上で視覚的に提示するために分割されたテキスト群の座標を取得する．

3.2 プロンプト作成

画像生成 AI におけるプロンプト設定では，生成する要素を明示的に指定するポジティブプロンプトと，生成から除外する要素を指定するネガティブプロンプトを活用している．ポジティブプロンプトとネガティブプロンプトを組み合わせることで，生成画像が小説本文と整合性を持ちつつ，高品質な挿絵として提示されることを目指している．

本研究では，小説本文の内容を正確に反映した画像生成を目指し，ポジティブプロンプトとして小説本文の内容から「小説のジャンル・スタイル」「舞台となる国・時代」「小説場面の要約」「小説場面に出てくる人物の外見的特徴」を取得し，プロンプトに組み込んでいる．ポジティブプロンプトの設定項目を表 1 に示す．「小説のジャンル・スタイ

ル」にはミステリー，ファンタジー，歴史小説などの小説のジャンルや小説のスタイルが含まれ，「舞台となる国・時代」には 18 世紀のフランスや現代の日本などの時代や場所の設定が該当する．「小説場面の要約」はその場面の内容を端的に説明したものである．「小説場面に出てくる人物の外見的特徴」は，登場人物の具体的な外見描写を含み，髪の色，服装，表情などの特徴要素が対象となる．これらの要素は，それぞれ異なる手法を用いて抽出する．

「小説のジャンル・スタイル」「舞台となる国・時代」「小説場面の要約」の取得には GPT-4 を使用して小説本文の文脈を解析し，要素ごとに情報を抽出している． GPT-4 は，大規模な事前学習データを基にした高い文脈理解能力を持ち，小説本文の複雑な内容や曖昧な表現を解析することが可能である．これにより小説のジャンルや場面要約を抽出することができる．一方「小説場面に出てくる人物の外見的特徴」は spaCy を用いて固有名詞を抽出し，係り受け関係を解析することで人物名に関連する外見的属性情報を取得している．このようにして取得した情報をテンプレート化し，ポジティブプロンプトとして画像生成 AI に入力することで，画像生成 AI が小説本文の内容を忠実に視覚化することを目指す．

ネガティブプロンプトは，生成画像の品質を向上させるために不可欠な要素であり，生成プロセスにおいて除外すべき特徴や不適切な要素を明示的に指定する役割を果たす．ネガティブプロンプトの設定項目を表 2 に示す．ネガティブプロンプトは「画像品質」，「形状および構造」，「不要または不適切な内容」の 3 項目を設定している．「画像品質」には，画質が低くぼやけていて粗く見える画像を除外する“low quality”や，鮮明さに欠け，焦点が合っていない画像を除外する“blurry”，物体や人物が不自然に歪んでいる画像を除外する“distorted”など，全般的な画質の低下を示す要素を排除する項目を設定している．「形状および構造」には，人物や動物などの体の構造が不自然な画像を除外する“anatomically incorrect”，頭や胴体，手足などのパーツのサイズ比が不自然な画像を除外する“inaccurate proportions”，人物や動物に本来存在しない余分な手足がついている画像を除外する“extra limbs”，必要な手や足が欠けている画像を排除する“missing limbs”などの記述を通じて，人物や動物の形状や構造における自然な表現が維持されるよう設定をしている．これらの設定は，特に登場人物や具体的な場面描写を視覚化する際に，観る者に違和感を与えない正確な画像を生成するための重要な要素である．「不要または不適切な内容」には，著作権保護のための透かしやロゴが含まれている画像を排除する“watermark”や不要なテキストが含まれている画像を除外する“text”など生成画像に不要な要素が含まれることを防ぐ設定も追加している．これにより，生成画像が純粋に視覚的なコンテンツとして機能し，余計な情報が提示されるリスクを回避している．また，暴力的，性的，差別的，または文化的・倫理的に問題のある内容を含む画像を除外する“inappropriate content”を設定することで，文化的・倫理的に問題のある画像が生成される可能性を低減し，生成システムの安全性と信頼性を確保している．このようなネガティブプロンプトの設定は，生成画像の一貫性を保ちながら，望ましくない特徴や誤った内容が含まれる画像が生成されるリスクを最小限に抑える効果を発揮する．

表 1　ポジティブプロンプトの設定項目

小説のジャンル・スタイル
舞台となる国・時代
小説場面の要約
小説場面に出てくる人物の外見的特徴

表 2　ネガティブプロンプトの設定項目

画像品質
形状および構造
不要または不適切な内容

3.3 画像生成

プロンプト作成工程で得られたポジティブプロンプトとネガティブプロンプトを Stable Diffusion に入力し，画像生成を行う．

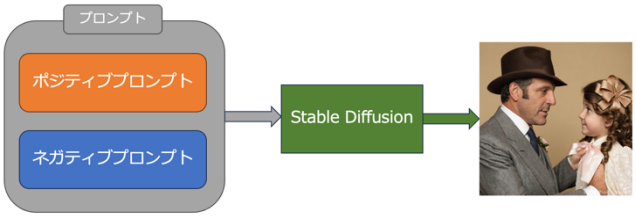

図 3　画像生成の流れ

3.4 プロンプトの最適化

小説本文に記載されている重要な情報が画像生成時に欠落する場合がある．この情報の欠落を防ぐために生成された画像の内容を解析し，必要に応じてプロンプトを修正，再生成し，場面描写における情報の欠落を最小限に抑える．Mina ら[14]の研究を参考に，BLIP-2 を用いて画像生成 AI が生成した画像から視覚的特徴を抽出しテキストを生成する．BLIP-2 は，画像から視覚的特徴を取得し，取得した視覚的特徴を自然言語形式で表現することが可能なマルチモーダル AI モデルである．まず，小説本文に対して GPT-4 を用いてキーワード抽出を行い，場面描写の重要な要素を特定する．抽出したキーワードリストを基に BLIP-2 に対して"Is (keyword) included in the image?"と個別にプロンプト

で質問を行い，Yes または No で応答を得る形式を採用する．(keyword)には小説本文から抽出された具体的な要素が挿入される．全キーワードについて Yes が返された場合，生成画像が適切に小説の場面を反映していると判断し，プロセスを終了する．一方，いずれかのキーワードに No が返された場合，欠落しているキーワードをプロンプトに追加する形でプロンプトを再生成し，再度画像を生成する．このプロセスは，すべてのキーワードが生成画像に反映されるまで繰り返し実行される．この手法により小説内の場面描写における重要な情報が画像生成時に欠落することを防ぐことが可能である．小説の文脈に基づいた適切な挿絵を生成するためには本文と挿絵の間で意味的な一貫性が保たれることが重要である．本研究では，BLIP-2 の視覚的特徴抽出とテキスト生成を活用することで，本文に含まれる描写内容が適切に視覚化されるようプロンプトを最適化している．プロンプト最適化後の画像生成を「Gutenberg[1]」のサイトより，以下の小説を使用して行なった．

- Gutenberg 「MARGERY DAW」著者名：Bertha M. Clay

生成時に使用したポジティブプロンプトとネガティブプロンプトをそれぞれ図 4 および図 5 に示し，出力した画像例を図 6 に示す．また，対応する場面のテキスト情報および小説タイトルを表 3 に示す．この場面は埃をかぶったグレー色のコートを着た女性が登場する場面であり，ポジティブプロンプトには「青い目」や，「赤金色の髪」などの小説内の女性の外見的特徴が反映されている．また，女性が着ている埃をかぶったグレー色のコートも反映されており，プロンプトが忠実に場面を再現していることが確認できる．これにより，画像生成 AI がテキスト情報に基づいた視覚的表現を適切に行なったことが示される．

表 3　小説タイトル・小説場面テキスト

小説タイトル	小説場面テキスト
MARGERY DAW	Mardie was dressed in a little gray coat, all covered now with dust.

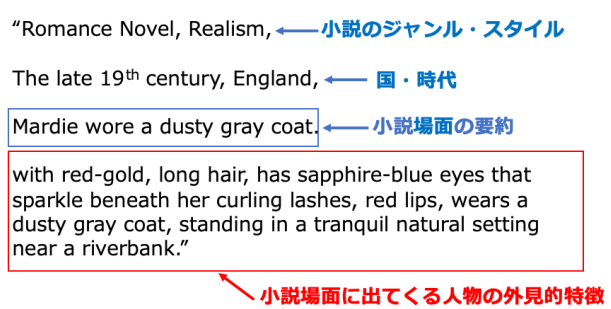

図 4　ポジティブプロンプト

“Low quality, blurry, distorted, damaged, anatomically incorrect, inaccurate proportions, extra limbs, missing limbs, watermark, text, inappropriate content.”

図 5　ネガティブプロンプト

図 6　生成画像

3.5 システム画面

図 7　システム画面

本システムは，読者がディスプレイ上で小説を読む際の視線移動に基づいて場面に応じた挿絵を提示するものである．システム画面を図 7 に示す．システム画面は，ゲームエンジンである Unity を用いて作成し，主に小説本文領域と挿絵提示領域の 2 つの領域から構成されている．ディスプレイ中央には小説本文が表示される領域を設けている．小説本文の見開きの両隣には視線位置および本文内容に基づいて生成された挿絵を提示する領域を設ける．この領域では，読者の視線移動による場面の切り替わりが検出されるたびに，プロンプトを基に画像生成 AI によって生成された挿絵が自動的に表示される．

4. おわりに

本研究では，ディスプレイを用いて小説を読む際の読者

1: Gutenberg: https://www.gutenberg.org

の視線移動に基づき，読者が読んでいる小説の場面に適切な挿絵を自動で提示するシステムを提案した．本システムの導入によって，読者は場面ごとに挿絵を見ることで物語への没入感を向上させ，文字情報だけでは得られない視覚的な補完効果が期待される．また，生成された挿絵の精度や一貫性のさらなる向上が求められる．現段階では，プロンプトの設定や最適化によってある程度の改善が見られたが，本文中の微細なニュアンスや抽象的な表現を完全に反映するには技術的な課題が残されている．特に，小説本文における登場人物の心情や場面の雰囲気など，視覚化が困難な要素に対する適切な画像生成手法の開発が必要である．さらに，対象とする小説の範囲を拡大し，英語以外の言語のテキストへの対応や，登場人物の感情や歴史的背景を含む複雑な文脈を持つ文学作品への適用も検討する必要がある．多言語対応や長編小説，歴史小説など様々なジャンルの作品への実装が進めば，本システムの汎用性が高まり，さまざまな読者層に対して有益な読書支援を提供する可能性がある．

参考文献

[1] 文化庁:平成 30 年度「国語に関する世論調査」の結果について(2018).
[2] 日本財団:18 歳意識調査「第 30 回―読む・書く―」(2020).
[3] 和田裕一:挿絵が物語文の読解における状況モデルの構築に及ぼす影響．心理学研究,90(4),368-377(2019).
[4] OpenAI, "GPT-4 Technical Report," 2023. [Online]. Available: https://openai.com/research/gpt-4.
[5] Honnibal, M., & Montani, I. spaCy 2: Natural Language Understanding with Bloom Embeddings, Convolutional Neural Networks and Incremental Parsing (2017).
[6] 今井真:視線追跡による読書支援に関する研究～亜アンビエントインタフェースの試み～,早稲田大学理工学術院基幹理工学部表現工学科 卒業論文,(2016).
[7] 上野浩平, 内田知巳, 羽岡浩二, 中島智晴, 吉田典弘:人工知能の自然言語処理を利用した VR による本読解アプリの提案, 情報システム学会 (2019).
[8] 村山貴志, 入江英嗣, 坂井修一: Word2Vec を用いた挿絵自動挿入手法の提案, 情報処理学会 インタラクション 2020 論文集, 377-380 (2020) .
[9] Mikolov, T., Chen, K., Corrado, G., & Dean, J. Efficient estimation of word representations in vectorspace. arXiv preprint arXiv:1301.3781 (2013).
[10] 齋藤優也: 要約文章に基づいた画像生成による挿絵自動生成システム, 法政大学大学院紀要．情報科学研究科編, 17, 1-6 (2022).
[11] 西川尚志, 橋本直他:文字表示に同期した音の提示が読書体験に与える影響, 研究報告エンタテインメントコンピューティング (EC), Vol. 2020, No. 27, pp. 1–6 (2020).
[12] Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B.: High-resolution image synthesis with latent diffusion models, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10684–10695 (2022).
[13] 吉原亮, 前田修作,五十嵐俊介,韓東力:小説における場面変化の情報抽出,言語処理学会第 15 回年次大会講演論文集(2009) .
[14] Mina Huh, Yi-Hao Peng, Amy Pavel. GenAssist: Making Image Generation Accessible. In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (*UIST'23*). CA, USA. 1-17 (2023).