

令和6年度
静岡大学 工学部数理システム工学科
卒業論文

キャラクター顔検索と背景情報提示機能
を持つ漫画読者支援システム

指導教員

岡部 誠 准教授

提出者

50116038 中村 斗耶

提出日

2025年2月13日

目次

第1章	はじめに	4
第2章	関連研究	5
2.1	物体検出モデル	5
2.2	顔埋め込み技術	5
2.3	自然言語生成技術	6
2.4	マルチモーダル技術	6
第3章	提案手法	8
3.1	ユーザインタフェース	8
3.2	システムの概要	9
3.3	キャラクター顔検出	10
3.4	キャラクター識別	10
3.5	背景情報生成	11
第4章	結果と考察	13
4.1	キャラクター識別精度の評価	13
4.2	キャラクター識別の結果と考察	13

4.3 背景情報の評価	15
第5章 まとめと今後の展望.....	16
謝辞.....	17
参考文献.....	18

第1章 はじめに

漫画作品は、読者に魅力的なストーリーと豊かな視覚表現を提供する文化的・商業的価値の高いコンテンツである。しかし、登場人物が多く、複雑なストーリーを持つ長期連載作品では、いくつかの課題が生じる。一つは、キャラクターの名前や役割を忘れてしまい、物語の理解が難しくなる点である。特に、時間を空けて再読する際には、前回の内容やキャラクターの背景情報を思い出すのが困難になり、物語への没入感が損なわれる。もう一つは、情報を得ようとしてインターネット検索を行う際、意図せずネタバレに遭遇するリスクが高いことである。ストーリーの展開を事前に知ってしまうことで、読書体験が大きく損なわれる可能性がある。

このような課題の背景には、読者がストーリーやキャラクターに関する情報を迅速かつ安全に得るための仕組みが十分に整備されていない現状がある。必要な情報を収集するには、多くの時間と労力がかかり、さらに検索中に予期せぬ形で物語の核心部分に触れる危険性もある。このような状況は、作品本来の楽しみを損なうだけでなく、情報収集の効率を大きく低下させる要因となる。

そこで、本研究では、これらの課題を解決し、読者がより快適に漫画作品を楽しめるよう支援するシステムを提案する。本システムは、以下の二つの主要な機能を備えている。一つ目はキャラクター認識機能である。この機能では、入力された画像から登場キャラクターを自動的に識別し、対応するキャラクター名を出力する。本機能の実現には、Danbooru API[1]を用いて特定のキャラクターに関連する画像を収集し、YOLO[2]による顔検出技術とFaceNet[3]による顔埋め込み生成技術を組み合わせることで、高精度な識別を行う。二つ目は背景情報提示機能である。この機能では、識別されたキャラクターに関連する簡潔なあらすじや背景情報を自動生成し、提示する。ここではOpenAI GPT[4]を用いて、キャラクターごとに適切な背景情報を生成する仕組みを導入する。

また、本システムはWebアプリケーションとして提供されるため、専門知識がなくても直感的に利用できる点が特徴である。読者は画像をアップロードするだけで、キャラクターの識別と背景情報の提示を受けられる。さらに、Flask[5]を基盤とした軽量なWebフレームワークを採用することで、システムの利便性を向上させた。この仕組みにより、情報収集の手間を省き、読者がスムーズに物語の世界に没入できる環境を提供することを目指している。

第2章 関連研究

2.1 物体検出モデル

画像内の対象物を検出する技術は、コンピュータビジョンの重要な研究分野であり、近年、多くのアルゴリズムが提案されてきた。初期の物体検出手法としては、Haar-like 特徴[6]や HOG (Histogram of Oriented Gradients) [7]を用いたものが挙げられる。これらの手法は、手動で設計された特徴量に依存しており、対象物の検出精度や柔軟性に限界があった。

その後、ディープラーニング技術の進展に伴い、畳み込みニューラルネットワーク (CNN) を用いた手法が主流となった。特に、R-CNN (Region-based Convolutional Neural Networks) [8]やその改良版である Fast R-CNN[9], Faster R-CNN[10]が提案され、検出精度と処理速度が大幅に向上した。しかし、これらの手法は計算コストが高く、リアルタイム処理には適していなかった。

これを解決するために、YOLO (You Only Look Once) が提案され、物体検出を1回の推論で完了するアプローチにより、高速かつ高精度な検出が可能となった。その後、YOLOv4[11], YOLOv5[12], そして最新バージョンである YOLOv8[13]が登場し、さらなる性能向上が実現している。

本研究では、漫画キャラクター特有の顔や多様な描写スタイルにも対応するため、YOLOv8 を基盤としたアニメキャラクター専用のモデルである yolov8_animeface[14]を採用する。このモデルは、Safebooru のアニメ画像を使用してトレーニングされており、アニメ特有のデータに最適化されている。この結果、デフォルメや複雑な背景に影響されることなく、アニメキャラクターの顔を高精度で検出することが可能となっている。

2.2 顔埋め込み技術

顔埋め込み技術は、画像データを低次元の特徴ベクトルに変換することで、顔画像の類似性を計算可能にする技術である。この技術は、顔認識、顔クラスタリング、特徴抽出など、幅広い応用分野において重要な役割を果たしている。

初期の顔認識技術では、主成分分析や線形判別分析などの次元削減手法が広く使用されてきた。しかし、これらの手法は顔の外観変化や照明条件の影響を受けやすく、精度面での課題があった。その後、HOG 特徴や SIFT[15]などの手法が提案され、より堅牢な特徴抽出が可能となったが、依然として限界が存在していた。

ディープラーニング技術の進展に伴い、畳み込みニューラルネットワークを用いた顔埋め込み技術が登場し、大きな飛躍を遂げた。特に、DeepFace[16]は、ディープラーニングを利用した最初の顔認識モデルの一つであり、顔画像を 128 次元の埋め込みベクトルに変

換する技術を導入した。その後、VGG-Face[17]や OpenFace[18]が開発され、埋め込みベクトルを用いた顔認識の精度がさらに向上した。中でも、FaceNet は、トリプレットロスを採用することで、顔埋め込みベクトルの分離性能を劇的に向上させたモデルである。FaceNet は、顔画像を 512 次元の特徴空間にマッピングし、同一人物の顔ベクトルが近接し、異なる人物の顔ベクトルが遠く離れるように最適化されている。この手法は、顔認識だけでなく、顔クラスタリングや類似度計算にも広く応用されている。

2.3 自然言語生成技術

自然言語生成は、ユーザからの入力やデータに基づいて人間が理解可能なテキストを生成する技術であり、対話システム、機械翻訳、文章要約、クリエイティブなコンテンツ生成など幅広い分野で活用されている。

初期の自然言語生成技術では、ルールベースのアプローチが一般的であった。これらの手法は、あらかじめ設計されたテンプレートに基づいてテキストを生成するものであり、言語的な制御が可能である一方で、柔軟性や多様性に欠けるという課題があった。その後、統計的手法が登場し、例えば n-gram モデル[19]が普及したが、これも文脈理解や長文生成には限界があった。

ディープラーニング技術の発展により、RNN (Recurrent Neural Networks) [20]や LSTM (Long Short-Term Memory) [21]といったモデルを用いた手法が登場し、より高度な文脈理解とテキスト生成が可能になった。特に、seq2seq モデル[22]の導入により、機械翻訳や文章生成タスクにおいて大きな進展が見られた。しかし、これらのモデルは長文の生成や文脈整合性の維持に限界があり、大規模データセットを必要とすることが課題であった。

これらの課題を解決するために、自己注意機構を用いた Transformer モデル[23]が提案され、自然言語処理の分野で革新的な進歩を遂げた。特に、GPT (Generative Pre-trained Transformer) シリーズは、事前学習とファインチューニングを組み合わせたアプローチにより、高精度かつ多様なテキスト生成を実現している。最新の大規模言語モデルである GPT-4[24]は、文脈理解能力や推論能力が大幅に向上しており、複雑なタスクや応用分野にも対応可能である。本研究では、この GPT-4 を採用する。

2.4 マルチモーダル技術

近年、画像やテキストといった異なるモダリティを統合的に処理するマルチモーダル技術が注目されており、これらの技術は画像認識や自然言語生成を組み合わせたタスクにおいて有効性を示している。特に、大規模マルチモーダルモデル (LMMs) は、異なるデータ形式間の情報の相互利用を可能にし、これまで個別に行われてきた画像認識とテキスト生成のプロセスを統合する新しいアプローチを提供している。

その中でも、Anim-Director[25]は、LMMs を活用してアニメーション動画を自律的に生成する最先端のフレームワークとして注目されている。Anim-Director は、GPT-4 を中核に据え、テキスト、画像、動画を統合的に処理することで、物語の文脈に基づいたスクリプト生成、シーン画像生成、さらには動画生成までを完全に自動化している。また、MidJourney を用いた画像生成と Pika を用いた動画生成を組み合わせることで、各シーン間のキャラクターや背景の一貫性を維持しつつ、高品質なアニメーション制作を実現している。さらに、自己反射的な評価メカニズムを導入することで、生成コンテンツの文脈的および視覚的一貫性を向上させている。

本研究では、これらの先行研究を踏まえ、YOLO と FaceNet を用いた高精度な顔検出および識別技術と、GPT-4 を活用した動的な背景情報生成技術を統合することで、漫画キャラクターの認識と関連情報の提示を一体化した支援システムを構築する。

第3章 提案手法

3.1 ユーザインターフェース

本研究では、漫画キャラクターの認識結果や背景情報を直感的に確認できるユーザインターフェースを提供するため、軽量な Web フレームワークである Flask を採用して Web アプリケーションを構築した。このインターフェースは、専門的な知識がなくても誰でも簡単に操作できるよう設計されており、ユーザがスムーズにシステムの機能を利用できる環境を実現している。



図 1 ユーザインターフェースの操作手順である。(a)は最初の画面で、(b)はキャラクターの顔の選択画面で、(c)はキャラクター識別結果の選択画面で、(d)は背景情報の表示画面である。

Web アプリケーションでは、最初に「画像を選択してください」のボックスから、情報を知りたいキャラクターを含む画像を選択する。この画像ファイルはユーザが用意するものであり、アップロード後に「顔を検出」ボタンを押すことで、キャラクターの顔検出が実行される (図 1(a))。

顔検出が終了すると、図 1(b)のように選択した画像内に存在するキャラクターの顔がすべて表示される。ユーザは、情報を知りたいキャラクターの顔を選択し、「この顔を選択」の左の丸が青になっている状態が選択済みであることを示す。そして、「次へ」ボタンを押すことで、選択が確定する。

顔の選択が完了すると、図 1(c)のように選択した顔画像に対してキャラクター識別が実行される。識別結果として、最も可能性が高いキャラクター名とその確率が表示されるほか、他の候補キャラクターも提示される。ユーザは、提示された候補の中から最も適切なキャラクター名を選択し、「このキャラクターを選択」の左の丸が青になっている状態が選択済みであることを示す。また、キャラクターの背景情報を取得するために、ユーザが既に読了している巻数を入力する必要がある。その後、「次へ」ボタンを押すことで、次のステップへ進むことができる。

キャラクター名と巻数が確定すると、図 1(d)のように、識別されたキャラクターに関する背景情報が生成される。本システムでは、GPT-4 を活用し、キャラクターの設定やストーリーの要点を生成し、ユーザが読み進めている巻数までの情報のみを提供することで、意図しないネタバレの発生を防ぐよう工夫している。この背景情報は、キャラクターの基本的な説明に加えて、直前の物語の流れや関連するイベントについても補足し、読者が物語を理解しやすくすることを目的としている。

本 Web アプリケーションは、直感的なユーザインタフェースを備えており、専門知識がなくても容易に操作可能である。ユーザは画像をアップロードし、数回の選択操作を行うだけで、キャラクターの識別結果と背景情報を取得できる。これにより、読者は作品の内容をスムーズに把握し、物語への没入感を高めることができる。

3.2 システムの概要

本研究では、漫画キャラクターの認識と背景情報提示を統合した支援システムを構築し、誰でも直感的かつ簡単に利用可能なユーザインタフェースを提供することを目的とする。従来のキャラクター認識手法は、主にターミナル上で実行されるものが一般的であり、操作に専門的な知識を必要とするだけでなく、認識結果の確認や追加情報の取得にも多くの手間がかかるという課題があった。さらに、これらの手法はキャラクターの識別までしか行えず、識別結果をもとに背景情報を提示する機能を備えていなかった。

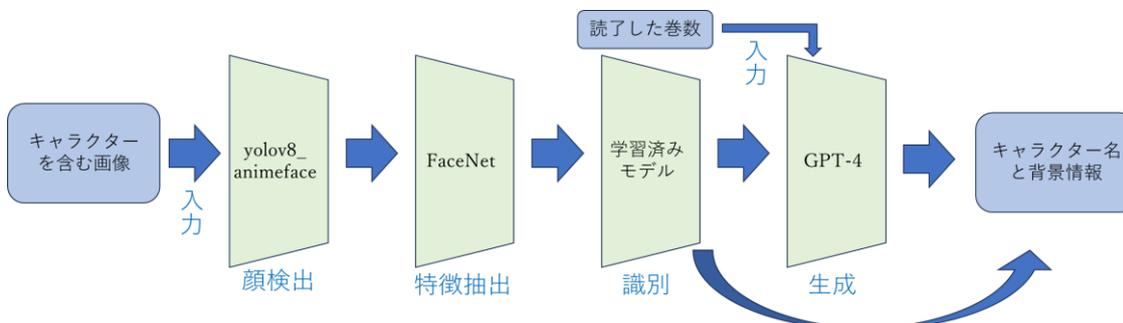


図2 システムの構造

これらの課題を解決するために、本研究では Web アプリケーションを開発し、前述のユーザインタフェースを通じて、ユーザが画像と読了巻数を入力するだけで、キャラクターの認識結果や背景情報を簡単に確認できる環境を提供する。具体的には、図 2 のように yolo8_animeface によるキャラクター顔検出、FaceNet を用いた識別、および GPT-4 を活用した背景情報生成を一連のプロセスとして統合し、画像とテキストという異なるモダリティを組み合わせることで、効率的かつ一貫性のある情報提供を実現した。また、認識されたキャラクターに基づき、背景情報や物語の要約を動的に生成する仕組みを導入することで、ユーザの意図に応じた柔軟な情報提供を可能にする。

さらに、Flask を基盤とした軽量な Web フレームワークを採用することで、システムの

開発・運用の効率化を図り、いつでもどこでも簡単に利用可能な環境を提供する。本システムは、これらの要素を組み合わせることで、従来のキャラクター認識手法の限界を克服し、漫画読者の体験を向上させる新たなアプローチを提案する。

3.3 キャラクター顔検出

本システムでは、漫画画像内のキャラクターの顔を高精度かつ効率的に検出するため、`yolov8_animeface` を使用している。関連研究で述べた通り、このモデルはアニメ特有のデータセットでトレーニングされており、デフォルメされた顔や複雑な背景にも対応可能である。本研究では、このモデルの特徴を最大限に活用し、顔領域の検出精度を向上させるための追加処理を施す。

具体的には、入力画像に対して `yolov8_animeface` を適用し、検出された顔領域の座標を取得する。その際、キャラクターの髪や装飾品など、識別に寄与する可能性のある要素を含めるため、検出範囲を一定のマージンで拡張して切り出しを行う。こうして抽出された顔領域は、次のキャラクター識別プロセスにおいて特徴ベクトルに変換され、識別処理に利用される。

3.4 キャラクター識別

キャラクター識別は、漫画画像内で検出された顔領域を基に各キャラクターを特定するプロセスであり、本研究のシステムにおいて重要な役割を果たす。本研究では、FaceNet を用いて高次元の画像データを低次元の埋め込みベクトルに変換し、効率的かつ高精度なキャラクター識別を実現する。

まず、FaceNet を用いて入力画像を 128 次元の埋め込みベクトルに変換する。FaceNet は、画像を低次元の特徴空間にマッピングし、ユークリッド距離を用いて類似性を計算する手法である。しかし、FaceNet 単体では主に距離計算による照合に用いられ、分類を直接行うことは難しい。そこで、本研究では、この埋め込みベクトルを、収集したキャラクター画像で事前に学習させたニューラルネットワークに入力し、キャラクターを識別する。

学習データの収集には、Danbooru API を利用した。Danbooru は、イラストを集積・管理するオンライン画像共有プラットフォームの一つであり、各画像に詳細なタグ付けが施されているため、特定のキャラクターに対応する画像を効率的に収集できる。本研究では、このタグ情報を活用し、各キャラクターに関連する画像を取得した。各キャラクターにつき最大 1000 枚の画像を収集し、枚数が不足する場合にはデータ拡張（回転、拡大縮小、輝度変更）を適用して 1000 枚に統一した。これにより、データの均一化を図り、識別の安定性を向上させた。さらに、異なるポーズや表情、影のつけ方などのバリエーションに対しても高い識別精度を維持できるモデルを構築した。

ニューラルネットワークには、特徴ベクトルを入力とし、各キャラクターに対応する確率を出力するソフトマックス層を持つ多層パーセプトロンを採用した。これにより、高次元の特徴を効率的に処理し、キャラクターの識別精度を向上させた。

本研究のキャラクター識別手法は、漫画特有の一貫した描写スタイルの中でも、異なるポーズや表情を持つキャラクターを適切に識別できる柔軟性を備えている。また、高精度な識別結果を出力することで、次の背景情報生成ステップへスムーズに接続できるよう設計した。

3.5 背景情報生成

背景情報生成は、キャラクター識別の結果を基に、そのキャラクターに関連する物語の背景情報やあらすじを生成するプロセスであり、本研究のシステムにおけるもう一つの重要な機能である。従来の背景情報提示手法は、固定されたデータベースに依存しており、事前に登録された情報しか提供できないという制約があった。このため、情報の更新が困難であり、文脈に応じた柔軟な情報提供ができないという課題があった。本研究では、これを克服するために、自然言語生成技術として GPT-4 を活用し、文脈に適応した背景情報を生成する仕組みを導入した。

背景情報生成プロセスでは、FaceNet を用いたキャラクター識別結果（キャラクター名）と、ユーザが読み終えた漫画の巻数を GPT-4 にプロンプトとして入力し、そのキャラクターに関する背景情報を生成する。プロンプトには、図 3 および図 4 の内容を採用し、GPT-4 に対して明確な指示を与えることで、一貫性のある情報生成を実現した。図 3 では、キャラクター名と読了巻数を指定し、そのキャラクターの背景情報を 500 字以内で出力するよう指示している。さらに、ユーザがまだ読んでいない巻以降のネタバレを含めないように制御することで、意図しないネタバレを防ぐ工夫を施した。

しかし、この段階では誤情報や不適切な表現が含まれる可能性があるため、生成された背景情報の信頼性を向上させるための修正プロセスを導入した。具体的には、図 4 に示すように、初めに出力された背景情報を再度 GPT-4 に入力し、ネタバレや誤情報の有無をチェックした上で、必要に応じて修正を行う。修正の必要がない場合はそのまま出力し、修正が必要な場合は正しい情報のみを返すよう指示している。この修正プロセスにより、誤った情報や不適切な内容を排除し、信頼性の高い情報を提供できるようになった。

本研究の背景情報生成は、固定データベースに依存しない新しい情報提供の形を提案し、ユーザの入力に基づいて柔軟に情報を生成できる仕組みを構築している。キャラクター識別結果と連動した情報生成を通じて、ユーザは必要な情報を迅速かつ正確に取得できる。これにより、読者は物語への没入感を損なうことなく、キャラクターに関する補足的な情報を容易に得ることが可能となる。

『{character_name}』というキャラクターについて、{volume}巻までの背景情報と直前の行動を500字以内で教えてください。
ただし、{int(volume)+1}巻以降の内容やネタバレを含めないでください。

図3 GPT-4 への初回入力プロンプト

以下の文章は、『{character_name}』の背景情報です。ただし、{int(volume)+1}巻以降のネタバレや間違いが含まれている可能性があります。必要に応じて修正し、正しい背景情報のみを出力してください。
【背景情報】 {summary_initial}
注意:
- 修正が不要な場合は、そのまま出力してください。
- 修正する場合は、修正後の背景情報のみを出力し、修正点の説明や注釈は不要です。

図4 GPT-4 による修正プロンプト

第4章 結果と考察

4.1 キャラクター識別精度の評価

提案システムのキャラクター識別精度を評価するため、学習キャラクター数の異なる3種類のデータセットを構築し、それぞれの識別精度を測定した。具体的には、1つの漫画から3キャラクターのみを学習する3キャラクター設定、3つの漫画から各3キャラクターずつ合計9キャラクターを学習する9キャラクター設定、10の漫画から各3キャラクターずつ合計30キャラクターを学習する30キャラクター設定の3種類を用意した。

各設定において、学習したキャラクターごとに10枚の評価用画像を準備し、識別精度を測定した。評価用画像は各漫画から手動で収集し、事前に `yolov8_animeface` を用いて顔領域を検出し、正しく顔が検出された画像のみを識別対象とした。これにより、顔検出の影響を排除し、純粋なキャラクター識別の性能を評価できるようにした。識別精度の評価には、最も高い確率で識別されたキャラクターが正解である場合の Top-1 精度を用いた。また、9キャラクター設定および30キャラクター設定では、識別対象が増えることで誤識別の可能性が高まることを考慮し、上位3候補のいずれかに正解が含まれる場合の Top-3 精度も算出した。

学習データの均等化を図るため、各キャラクターにつき最大1000枚の画像を収集し、枚数が不足する場合にはデータ拡張を適用した。データ拡張には、回転 (-15度~15度)、拡大縮小 (0.9倍~1.1倍)、輝度変更 (0.8~1.2倍) を用い、学習データの多様性を確保するとともに、異なる環境下でも安定した識別が可能となるようにした。

識別モデルは、入力層の後に3層の全結合層を配置し、それぞれ512、256、128ノードの構成とした。各層には Batch Normalization[26]を適用し、学習の安定性を向上させるとともに、Dropout を導入して過学習を防止した。最終出力層にはソフトマックス関数を用い、各キャラクターに対する確率分布を出力することで識別を行う。学習アルゴリズムには Adam オプティマイザ[27]を採用し、学習率は0.001に設定した。また、損失関数には sparse categorical crossentropy を適用し、識別タスクに適した最適化を行った。また、L2正則化を適用し、モデルの一般化能力を向上させた。

本実験により、提案システムが異なるキャラクター数に対してどの程度の識別精度を持つかを明らかにし、汎用性や拡張性を検証する。

4.2 キャラクター識別の結果と考察

表1に示すように、3キャラクター設定の Top-1 精度は0.800であり、9キャラクター設定では0.705、30キャラクター設定では0.500となった。キャラクター数が増加するにつ

れ、Top-1 精度は低下する傾向が確認された。一方、Top-3 精度は 9 キャラクター設定で 0.898、30 キャラクター設定で 0.761 となり、Top-1 精度と比較して高い値を示した。これは、キャラクター数が増加することで識別の難易度が上がるものの、上位 3 候補に正解が含まれる確率は依然として高いことを示している。

設定	Top-1 精度	Top-3 精度
3 キャラクター設定	0.800	-
9 キャラクター設定	0.705	0.898
30 キャラクター設定	0.500	0.761

表 1 キャラクター識別結果

9 キャラクター設定におけるキャラクターごとの識別精度を表 2 に示す。キャラクターごとの識別精度にはばらつきがあり、特定のキャラクターは高い精度で識別される一方で、識別が難しいキャラクターも存在することが分かった。例えば、キャラクター 1 の Top-1 精度は 1.000 であり、すべての評価画像において正しく識別されていた。一方で、キャラクター 2 の Top-1 精度は 0.300 にとどまり、誤識別が多く発生していることが確認された。また、Top-3 精度においては、すべてのキャラクターが 0.700 以上の値を示しており、上位 3 候補の中に正解が含まれる確率は高いことが分かる。

キャラクター	Top-1 精度	Top-3 精度
キャラクター 1	1.000	1.000
キャラクター 2	0.300	0.800
キャラクター 3	0.400	0.900
キャラクター 4	0.889	1.000
キャラクター 5	0.900	0.900
キャラクター 6	0.900	1.000
キャラクター 7	0.556	0.889
キャラクター 8	0.800	0.900
キャラクター 9	0.600	0.700

表 2 9 キャラクター設定における識別精度

本研究の結果から、提案システムはキャラクター数が少ない場合には高い識別精度を示すものの、キャラクター数が増加すると識別精度が低下する傾向が確認された。この要因として、キャラクター間の特徴の類似性が増すことや、学習データにおける各キャラクターの個体差が識別精度に影響を与えている可能性が考えられる。特に、9 キャラクター設定の識別結果を詳細に分析したところ、高い識別精度を維持できるキャラクターも存在する一方で、識別が困難なキャラクターも確認された。この識別精度の差異は、キャラクターの特徴の明確さや、学習データに含まれるバリエーションの違いが影響していると考えられる。

また、上位 3 候補まで考慮することで識別精度を大きく向上できることが確認された。これは、キャラクター識別の際に完全一致を求めるのではなく、上位の候補を提示することで、実用的な精度を確保できる可能性を示唆している。特に、類似するキャラクター同士が誤識別されるケースでは、Top-3 の候補提示が有効であると考えられる。

以上の結果から、提案システムは比較的少数のキャラクターを対象とする場合には高精度な識別が可能であることが確認された。しかし、学習するキャラクター数が増加すると識別精度が低下する傾向が見られた。このため、今後の課題として、識別が困難なキャラクターに対する特徴抽出の改善、学習データの多様性を確保するためのデータ拡張の強化、および学習キャラクター数の増加への対応が挙げられる。これらの課題に取り組むことで、より多くのキャラクターを正確に識別できるシステムの構築が可能となり、本研究の成果をさらに発展させることができると考えられる。

4.3 背景情報の評価

背景情報の生成精度を評価するため、ユーザテストを実施した。具体的には、生成された情報の正確性やネタバレ防止機能の有効性について、5段階評価のアンケートを作成し、回答を収集した。評価は「非常にそう思う (5)」から「全くそう思わない (1)」までの数値スケールで行い、各項目の平均スコアを算出した。

その結果、「キャラクターやストーリーの説明の正確性」に関する評価は平均 4.2 となり、ほとんどの回答者が生成された情報を正確であると評価した。また、「ネタバレ防止機能の有効性」についても平均 4.0 となり、ネタバレを回避しつつ適切な情報提供が行われていることが確認された。一方で、一部の回答者からは「特定の表現に違和感があった」との指摘があり、さらなる表現の自然さの向上が求められることが示された。

加えて、自由記述のフィードバックでは、「単行本の発売間隔が長く、キャラクターを忘れることがあるため、このようなシステムが役立つ」や「登場人物が多い作品や設定が複雑な作品では、情報整理の補助として有用」といった肯定的な意見が寄せられた。これにより、本システムが読者の理解を助け、物語への没入感を維持するための有効な手段となる可能性が示された。

以上の結果から、背景情報の修正プロセスは一定の効果を発揮し、正確性やネタバレ防止の面で実用的なレベルに達していることが確認された。しかし、表現の自然さや情報の詳細度に関する改善の余地があることも明らかとなった。今後は、キャラクターごとの説明のバランス調整や、より自然な文章生成のためのプロンプト設計の最適化を進めることで、システムの精度向上を図る。

第5章 まとめと今後の展望

本研究では、漫画のキャラクターを自動で識別し、それに基づいた背景情報を提供するシステムを提案した。システムは、yolov8_animeface を用いたキャラクター顔検出、FaceNet を用いた識別、GPT-4 による背景情報生成という3つの主要な機能を統合し、ユーザが直感的に操作可能な Web アプリケーションとして実装した。実験の結果、少数のキャラクターを対象とする場合には高い識別精度を達成した一方で、キャラクター数が増加するにつれて識別精度が低下する傾向が確認された。また、Top-3 精度を考慮することで、識別が困難なキャラクターを含む場合でも有効に機能する可能性が示された。さらに、GPT-4 による背景情報生成に関する評価では、生成された情報の正確性やネタバレ防止機能の有効性が確認され、読者の物語理解を支援する手段としての有用性が示された。

今後の展望としては、識別精度の向上を目的として、特徴抽出の改善やデータ拡張の強化を進めるとともに、学習キャラクター数を増やし、より多様なキャラクターに対応可能なシステムの構築を目指す。また、背景情報生成においては、より自然で正確な文章を生成するためのプロンプト設計の最適化を行い、さらなる精度向上を図る。これらの改良により、より多くの漫画作品に対応可能な汎用的な読者支援システムの実現を目指す。

謝辞

本研究及び論文の作成にあたり、研究の着想や論文執筆等、多くのご指導、ご助言を頂きました静岡大学工学部の岡部誠准教授に心から感謝申し上げます。また、ご助力頂いた修士課程学生及び学部生の皆様に深く感謝致します。

参考文献

- [1] Danbooru, “help:api”, https://danbooru.donmai.us/wiki_pages/help:api, (参照 2025-01-31).
- [2] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, “You Only Look Once: Unified, Real-Time Object Detection”, in CVPR 2016.
- [3] Florian Schroff, Dmitry Kalenichenko, James Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering”, in CVPR 2015.
- [4] OpenAI, “Models”, OpenAI, <https://platform.openai.com/docs/models>, (参照 2025-01-31).
- [5] Pallets Projects, “Flask: Web development, one drop at a time”, <https://flask.palletsprojects.com/>, (参照 2025-01-31).
- [6] Paul Viola, Michael Jones, “Rapid object detection using a boosted cascade of simple features”, in CVPR 2001.
- [7] Navneet Dalal, Bill Triggs, “Histograms of oriented gradients for human detection”, in CVPR 2005.
- [8] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation”, in CVPR 2014.
- [9] Ross Girshick, “Fast R-CNN”, in ICCV 2015.
- [10] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, in NeurIPS 2015.
- [11] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection”, arXiv:2004.10934.
- [12] ultralytics, “yolov5”, GitHub repository, <https://github.com/ultralytics/yolov5>, (参照 2025-01-31).
- [13] ultralytics, “Ultralytics YOLOv8”, ultralytics, <https://docs.ultralytics.com/ja/models/yolov8/>, (参照 2025-01-31).
- [14] Fuyucchi, “yolov8_animeface”, GitHub repository, https://github.com/Fuyucch1/yolov8_animeface, (参照 2025-01-31).
- [15] DAVID G. LOWE, “Distinctive Image Features from Scale-Invariant Keypoints”, in International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.
- [16] Yaniv Taigman, Ming Yang, Marc’Aurelio Ranzato, Lior Wolf, “DeepFace: Closing the Gap to Human-Level Performance in Face Verification”, in CVPR 2014.
- [17] Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman, “Deep Face Recognition”, in BMVC 2015.

- [18] Brandon Amos, Bartosz Ludwiczuk, Mahadev Satyanarayanan, “OpenFace: A general-purpose face recognition library with mobile applications”, in CMU-CS-16-118, 2016.
- [19] Chris Manning, Hinrich Schütze, “Foundations of Statistical Natural Language Processing”, in MIT Press, 1999.
- [20] Jeffrey L. Elman, “Finding structure in time”, in Cognitive Science, vol. 14, no. 2, pp. 179–211, 1990.
- [21] Sepp Hochreiter, Jürgen Schmidhuber, “Long short-term memory”, in Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
- [22] Ilya Sutskever, Oriol Vinyals, Quoc V. Le, “Sequence to sequence learning with neural networks”, in NeurIPS 2014.
- [23] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, “Attention Is All You Need”, in NeurIPS 2017.
- [24] OpenAI, “GPT-4 Technical Report”, OpenAI, <https://openai.com/research/gpt-4>, (参照 2025-01-31).
- [25] YUNXIN LI, HAOYUANSI, BAOTIAN HU, LONGYUEWANG, JIASHUN ZHU and JINYI XU, ZHENZHAO, MINZHANG, “Anim-Director: A Large Multimodal Model Powered Agent for Controllable Animation Video Generation”, in SIGGRAPH Asia 2024.
- [26] Sergey Ioffe, Christian Szegedy, “Batch Normalization: Accelerating deep network training by reducing internal covariate shift”, in ICML 2015.
- [27] Diederik P. Kingma, Jimmy Ba, “Adam: A method for stochastic optimization”, in ICLR 2015.