JEITA

情産－23－技標－4

認識形入力方式に関する調査研究報告書

２０２４年３月

一般社団法人　電子情報技術産業協会
認識形入力方式標準化専門委員会

序　文

我が国の将来の労働力人口減少に対応するためには労働生産性の向上が急務であり、デジタル技術の利活用により効率的な社会づくりを目指すDX（Digital Transformation）や、業務の自動化を進めるRPA（Robotic Process Automation）といった概念に注目が集まっている。一方、2022年末に発表されたChatGPTを切っ掛けとして生成AIの驚くべき高機能が世の中の注目を浴びており、デジタル技術が様々な業務の質的な変革を促す可能性が話題に上っている。AI技術の急激な進歩とそれに伴うリスクへの懸念は政治的・社会的にも重要視されており、5月に開催されたG7広島サミットの首脳宣言にAIの適切利用の促進を目指す広島AIプロセスが盛り込まれるなど、デジタル技術の確立と適正利用は大きな社会的課題となりつつある。当委員会の名称となっている認識形入力方式、すなわち文字や画像などの実世界に存在するいわゆる「パターン情報」をAI（人工知能）技術によって自動的に認識し、デジタル情報システムへの入力情報に変換する枠組みは、実世界の情報をデジタル世界に取り込んで活用するための重要な基盤技術であり、まさに今後の社会変革の流れにおいて重要な役割を担うものである。

このような背景のもと、当協会では、2023 年度「認識形入力方式標準化専門委員会」を設けて、認識形入力方式を支える基盤技術に関する議論と、それを活用したシステムの現状調査を行ったので、その成果をここに報告する。

本年度、当委員会では昨年度に引き続き、認識形入力方式の性能劣化を引き起こす実環境における外乱要因について現象・要因の観点から整理を行うと共に、その成果をまとめた新たなガイドライン（『実環境における認識機器の外乱要因のガイドライン』）を作成し公開した。また、AI 技術の動向を広く調査し、研究開発を推進するための方策について議論を行った。既に実用化が進んでいる「整備された環境におけるOCR技術」については、引き続き現状の技術と装置について調査を行った。さらに、今年度開催された国内外の学術会議等の動向について調査を行い、技術の最新動向の把握と今後のあり方について議論を行った。

本報告書の作成にあたり、ご協力をいただいたユーザー、メーカー各位と、ご指導を賜った関係省庁、並びに本報告書の作成にあたって労を賜った委員各位に深く感謝の意を表すると共に、本報告書が各方面に広く利用され、我が国における情報化と産業の発展に寄与できれば幸いである。

2024年3月

一般社団法人 電子情報技術産業協会
認識形入力方式標準化専門委員会
委員長　田中　宏

認識形入力方式標準化専門委員会名簿

（敬称略、順不同）

委員長	田　中　　宏	富士通株式会社
副委員長	古　畑　彰　夫	東芝デジタルソリューションズ株式会社
監　事	山　合　敏　文	株式会社 リコー
委　員	佐　藤　雄　隆	国立研究開発法人 産業技術総合研究所
委　員	岩　田　健　司	国立研究開発法人 産業技術総合研究所
委　員	石　寺　永　記	日本電気株式会社
委　員	田　辺　吉　久	OCR エキスパート
委　員	松　村　　博	OCR, AI 技術アドバイザー
客　員	栗　田　多喜夫	広島大学
事務局	吉　田　　晃	一般社団法人 電子情報技術産業協会
事務局	塩　川　大　介	一般社団法人 電子情報技術産業協会

（２０２４年3月３１日現在）

目　次

1. はじめに

近年の AI 技術の急激な進化に伴い、認識形入力方式の性能は年々目覚ましい向上を遂げている。また、スマートフォンに代表されるモバイル端末のコモディティ化や性能向上、及び高速なモバイルネットワーク通信網の整備なども相まって、様々な人が様々な用途に認識形入力方式を利用できるようになった。例えば、レシートや手書きメモをスマートフォンで撮影することで記載内容をデジタルデータに変換したり、身の回りにある動植物を撮影するだけでその種類を検索して調べたりすることも可能になりつつある。更に ChatGPT に代表される対話型 AI により、高度な AI 技術が容易に活用できるようになった。話題になっている生成 AI は一般にはコンテンツ生成技術として捉えられているが、大規模学習による知識の活用は情報入力技術にも革新的な変革を促している。

このように認識形入力方式の活躍の場が拡大する一方で、利用環境下における外乱の影響を受けて認識精度が変化する認識形入力方式の特性と、用途の拡大に伴う外乱混入機会の増大とから、認識形入力方式が本来の性能を発揮できず性能が劣化するリスクも高まっている。

このような背景のもと、2023 年度の認識形入力方式標準化専門委員会では、認識形入力方式を支える基盤技術に関する議論と、それを活用したシステムの現状調査を行った。特に、認識形入力方式の性能劣化を引き起こす実環境における外乱要因の整理と、これを有用なガイドラインとしていくための施策について審議を行った。また、進展が著しい認識形入力方式の最新動向について、製品及び技術の観点から調査を行った。

本章では、まず、実世界のアナログ情報がデジタル化され、また、デジタル情報が必要に応じてアナログ化されて人間に利用される様子を俯瞰的に表現した。それにより、本委員会が特に課題として検討している文字情報のデジタル化変換の位置付けを明確化することを試みている。それに続き、本委員会の 2023 年度の活動状況と審議内容について述べる。

1.1　デジタル化に関する俯瞰的考察

前述のように社会のデジタル化は急速に進んでおり、会議やイベントのオンライン化が急務となると共に、紙ベースでの書類のやり取りにも支障をきたす事態となり、電子化、押印廃止などの動きが強まっている。

本委員会で従来から扱っている OCR は、主に紙に印刷された文字情報を機械で読み取り、情報をデジタル化する技術であり、アナログ情報をデジタル化して活用する試みの先駆けと考えることができる。一方、現代においては既にデジタル世界（オンライン）のみに存在し、活用される情報も急速に増加している。このような状況下において「デジタル化」というキーワードを俯瞰的視点から見た場合に、具体的に何を意味していることになるのか、OCR や RPA の文脈を中

心としながら本委員会で議論・整理を行った。特に今年度の報告では、生成 AI をどのように位置付けるかについても試案を追加した。(以降、議論の要点をまとめるが、本件は今後も継続して議論を行うこととしており、現時点までの整理の速報版であるという点に留意されたい。)

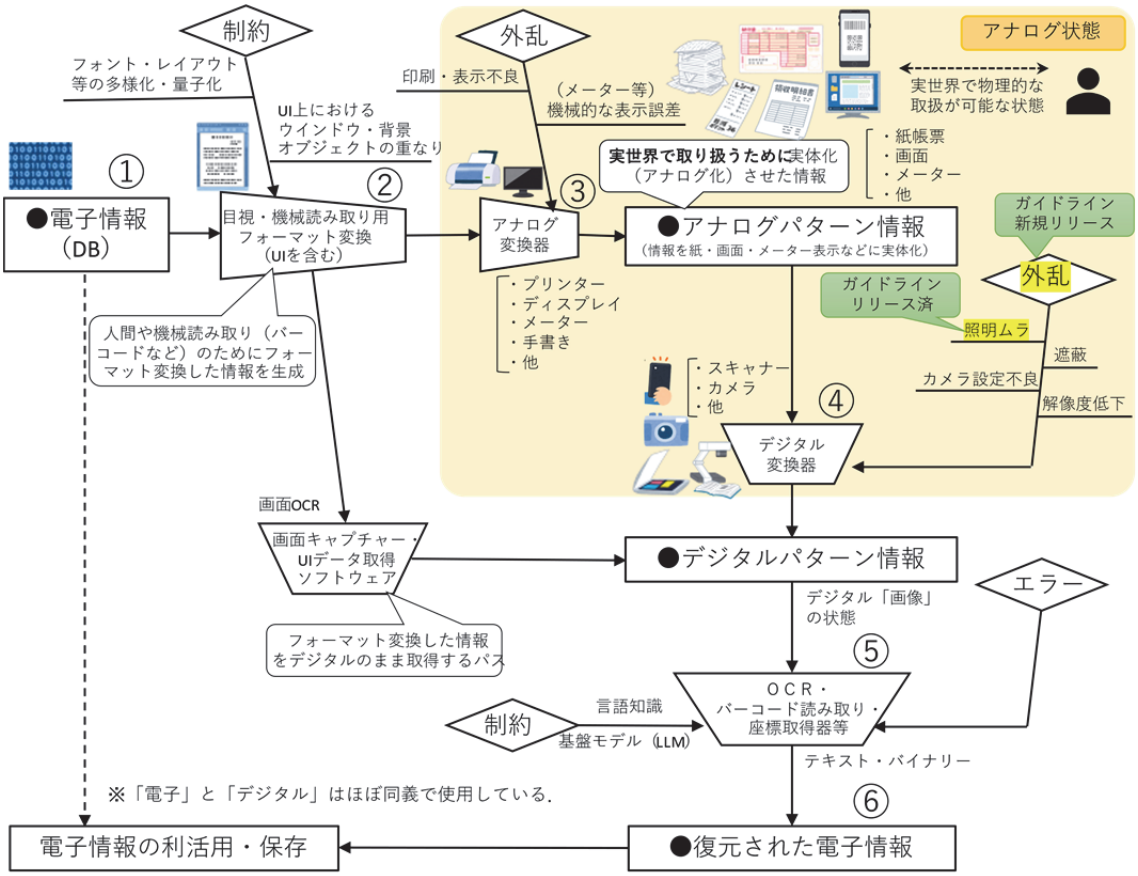


図 1.1-1　デジタル化に関する俯瞰的な考察

図 1.1-1 に考察の概略図を示す。大きな図であるが、まずは「●」印が付いた「電子情報」「アナログパターン情報」「デジタルパターン情報」「復元された電子情報」に着目し、情報がこれらの状態間を遷移する（①～⑥）流れを意識すると理解しやすい。

まず、左上に①として電子情報があり、過去に電子化された情報や、初めからデジタルデータとして作成された、いわゆるボーンデジタルの情報が存在する。現代社会では、大半の情報が電子化されてインターネットや各社のシステム上で扱われており、電子情報は既に我々の社会に不可欠なものとなっている。

この電子情報を、図中下に向かう点線矢印のように直接利活用することが最もシンプルで効率的である。しかしながら、人間が直接情報を活用できるようにするためには、人間が容易に読み取れるように情報の体裁を加工することが必要となる（②）。具体的には、画面表示や紙面生成のために文字フォントを割り当てたり、文章の配置をレイアウトしたりすることが考えられる。ま

た、データを列挙するような場合には表としてレイアウトすることも考えられる。体裁が加工されたデータはこの時点ではデジタル情報のままであるが、フォントやレイアウトの多様化や、表現の都合上画像化されるような場合には量子化による解像度低下の問題などが発生する可能性がある。

次に体裁を加工した情報を、人間が目視できるように何らかのアナログ情報に変換する（③）。例えば、プリンタで紙に印刷したり、ディスプレイ装置の画面に表示したりする。この段階では、印刷・表示不良などの外乱が加わる可能性がある。ディスプレイ上で情報を確認・操作するだけならば、ディスプレイ装置以外に実世界に実体は存在しないので、デジタル情報をそのまま扱っているとみなすこともできる。一方、紙ベースに変換された情報は近年急速に減少しているものの、依然として大量の紙の書類や帳票類などが実世界に実体を持って我々の周りに存在している。これらをどうにか生成せずにデジタルのまま扱う、またはデジタルに戻す、ということが「デジタル化」が目指す目標の一つだと考えられる。

紙の情報をデジタル情報に戻す場合、まず、デジタル変換（④）が必要となる。これを行う代表的な装置としては、スキャナーやカメラなどが考えられるが、特に近年、スマートフォンによる帳票読み取りのようにカメラを用いるケースが増加しており、撮影時に照明ムラ、遮蔽、カメラ設定不良、解像度低下などの外乱が加わることで、後の工程に支障をきたす可能性がある。なお、本委員会ではこのうち照明ムラに関して、認識機器のグレード定義を行う利用ガイドラインを JEITA ITR-4010「実世界環境における OCR の利用ガイドライン 照明ムラ版」として 2019 年度にリリースし、また、認識に影響を与える外乱を要因と影響に分けて整理したガイドラインとして、情産-23-技標-3「実環境における認識機器の外乱要因のガイドライン」を 2023 年度にリリースした。併せて参照されたい。

次に、スキャナーやカメラによって得られた「デジタル画像」から文字や数値の情報を OCR などによって「デジタル情報」として抽出する（⑤）。なお、ここでも抽出は完全ではなく、ある確率でエラーが発生することに注意が必要である。（序文でも触れた生成 AI は、事前に学習した巨大なモデルデータ（「大規模言語モデル（LLM）」や「基盤モデル」と呼ばれる）による知識によってデジタル情報への高精度な変換を可能とするものなので、本稿の俯瞰的な視点では、⑤に対する制約として位置付けた。）

このようにしてようやく復元された情報（⑥）は、②～⑤の工程において様々な外乱を受け劣化している可能性があることに注意が必要であるが、再び電子情報として利活用・保存することが可能な状態となる。なお、近年では RPA の文脈で、画面 OCR が用いられることがある。これは本来人間に提示するためのユーザーインターフェースを機械が読み取り、人間に代わって操作・入力を行うために用いられる（もともと自動操作に対応したシステムであればこのような仕組みは必要ないが、通常の人間用のソフトウェアを改修せずに、そのまま比較的容易に自動化で

きるということで注目されている技術である)。画面 OCR は、②で画面表示用に体裁が加工されたデータを取得し、⑤の工程に流すことで画面上の情報を読み取る。

以上の工程を振り返ってみると、明らかに元の電子情報（①）を直接活用することが最も効率的であることがわかる。しかしながら、情報化が既に進んでいるはずの近年においても紙媒体は重視され、証憑等、事実を証明するためのエビデンスとしてすら用いられてきた。これは、電子情報が存在しなかった時代から続く「従来からのやり方」を根本的に変えることが困難であったことが主な原因であると考えられるが、紙媒体は実世界で実体を持つため人間にとって扱いがわかりやすく、金庫などによる物理的手段によってある程度の保護も可能であるなどのメリットも重視されたのではないかと考えられる。

今後「デジタル化」のスローガンのもと、アナログ化を経ずに直接電子データとして扱われる情報の割合は大幅に増加するものと考えられる。しかしながら一方で、人間自身がアナログ世界に存在している以上、その利便性を考慮すると紙への印刷などアナログ化することが適切な例も残るだろう。そのようなアナログ情報を効率よく活用するためには、エラーの影響を考慮しつつ人手を介さずにデジタルへ変換する技術が必要であり、安心して自動変換を行うための品質保証の問題も発生する。なお、今回は OCR を中心とした考察を行ったが、近年では人・場所・システムなど、実世界の様々な対象をデジタル化して扱うデジタルツインという概念も注目を浴びており、今後はこれも踏まえた議論を本委員会で行っていく予定である。

1.2 審議過程

委員会は 2023 年 5 月から 2024 年 2 月まで計 8 回オンライン開催され、「実世界環境における認識機器の耐環境性の標準化」について審議が行われた。

1.1 節で触れたように、認識形入力方式を搭載した機器（以下、認識機器と呼ぶ）が認識対象とする情報はデジタル化の過程において多様な外乱の影響を受ける。そのため、条件によっては認識性能が劣化することが知られている。しかしながら、その条件やメカニズムは非常に複雑であるため、従来その対処は専門家の暗黙知に頼らざるを得ない面も強かった。このような知識を部分的にでも形式知化することができれば、より合理的で積極的な性能向上を図ることが検討可能になるほか、AI の学習過程にその知識を埋め込むことで認識性能を更に向上させることなども検討可能になる。

このような背景から、当委員会では約 4 年の年月をかけて、OCR に影響を及ぼす多種多様な外乱を網羅的に列挙すると共に、現象・原因の観点から整理を試みた。本年度は引き続き整理を進め、多くの開発者やユーザーに活用してもらえるようにするためのガイドラインを新たに作成し公開した。

1.3　技術の現状調査

1.3.1　認識方式と認識技術の動向調査

DL（Deep Learning；深層学習）等の技術革新を背景に、認識方式及びそれを用いた装置・ソフトウェアの、性能・機能・適用対象に大きな変化が起きている。このような背景を踏まえ、DLを中心とした最新の認識技術や文字認識・文字理解の進展、及びそれを用いた装置・ソフトウェアに関して最新動向を調査し本報告書にまとめた。また、最新動向の調査・把握のため、日本電気株式会社の小山田昌史氏にLLM（大規模言語モデル）に関する研究開発について、同じく日本電気株式会社の石寺永記委員にはご自身の研究活動の紹介を兼ねて、文字認識（郵便区分機など）や物体検知技術について、それぞれご講演を行っていただいた。

1.3.2　文字認識関連技術の調査

カメラ付き携帯デバイスによる文字認識技術等、認識形入力方式の新たな展開が期待される応用分野について調査を行った。また、文字認識装置についての現状を調査すると共に、OCR製品一覧表の更新・拡充を行った。

1.3.3　技術動向調査

今年度開催された国内外の学術会議のうち、今後認識形入力方式に大きな影響を与えると予想される最新の研究成果の調査を行い、委員会において今後の変化・発展の可能性について議論を行った。また、調査結果の要約を本報告書にまとめた。

2. 実世界環境における認識機器の耐環境性の標準化

AI 技術の急激な進化に伴い、認識形入力方式を搭載した認識機器の用途が急速に拡大している。その一方で、1.1 節で述べたように利用環境下における外乱の影響を受けて認識精度が変化するという認識形入力方式の特性と、用途の拡大に伴う外乱の混入機会増大とから、条件によっては認識機器が本来の性能を発揮できず、認識精度が劣化するリスクが高まりつつある。そして、このことが認識機器の品質管理を難しくしつつある。

このような背景から、当委員会では認識機器の代表格として特に OCR に注目し、OCR の普及を促進するための議論を進めてきた。ユーザーが OCR を正しく利用できるようにすることを目的として、外乱要因の中でも認識精度への影響が大きい照明ムラに着目し、OCR の利用ガイドラインを 2020 年度 4 月に制定した。続いて、実世界環境下で認識精度に影響を及ぼす要因とそれによって生じる画像劣化の現象を要因表をとりまとめ、実環境における認識機器の外乱要因のガイドラインとして公開した。

本章では、まず 2.1 節にて、ガイドラインとして取りまとめられた要因表と活用施策について述べる。2.2 節では、OCR の新たな活用方法として注目される RPA における UI 操作自動化について説明する。2.3 節では、急激に進歩する生成 AI 技術を中心に、実世界環境における認識機器の最新動向と今後の展望について述べる。2.4 節では、OCR の品質管理をめぐる最新動向について述べる。

2.1 実世界環境における外乱要因

ユーザーがOCRを正しく利用できるようにすることを目的として、認識形入力方式の性能劣化を引き起こす実環境における外乱要因をとそれによって生じる画像劣化の現象を要因表として整理し、その活用方法を示す。

2.1.1 外乱要因表

スタンド型のスキャナーやカメラによる文書・帳票画像のデジタル化の過程においては、図 2.1.1-1 に示すように(a)光源、(b)障害物、(c)対象物、(d)カメラ、(e)撮影者が関わり、これらに起因して外乱要因が発生すると考えられる。そこで、外乱要因をその発生源ごとにグループ化すると共に、外乱要因によって生じる画像劣化との関係を整理した。整理した結果を表 2.1.1-1 に示す。表中、L-xx（xx は数字）などの表記は、外乱要因とそれに起因して発生する画像劣化との対応関係を示す。表からわかるように、1 つの外乱要因が複数のタイプの画像劣化に関係する場合もある。また、要因表は認識対象物をスマートフォン内蔵カメラやスタンド型スキャナーで撮影して OCR 認識する際に認識精度に影響を及ぼす要因や現象を示すが、一部の要因については従来型の

スキャナーで画像を取得して OCR 認識する場合にも当てはまる。これらを区別できるよう、表 2.1.1-1 では従来型のスキャナーで画像を取得した場合にも共通して当てはまる要因については青字で記載した。

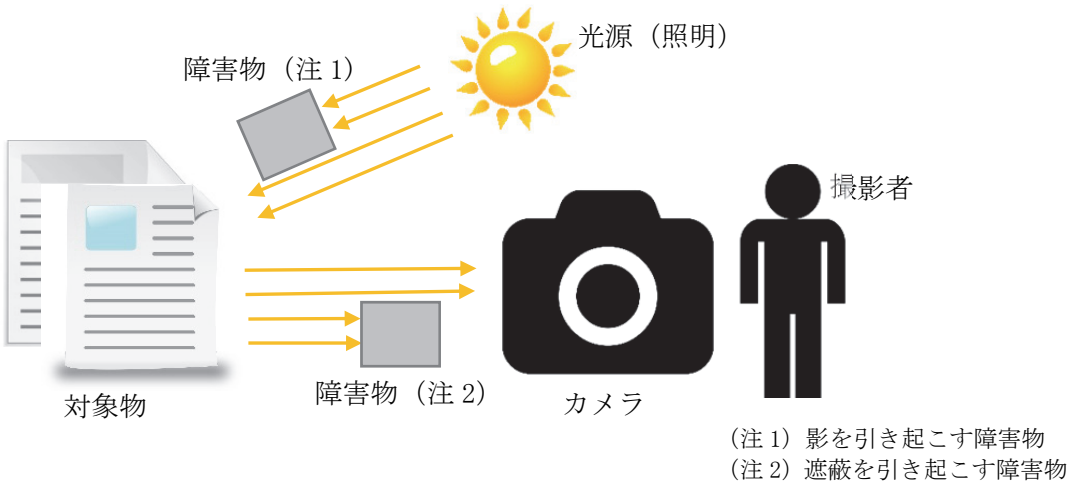


図 2.1.1-1　文書・帳票画像のデジタル化の過程

表 2.1.1-1　OCR に影響を与える要因及び現象

分類	現象	要因
光源	・黒つぶれ L-01，L-04，L-06，L-07，L-10  	・状態及び種類 －明るすぎる/暗すぎる(L-01) －ムラの出る照明(L-02) －点光源(L-03) －周波数（フリッカー）(L-04) －フラッシュ(L-05) －色がついた光(L-06) ・位置及び数 －光源から対象物までの距離が近い(L-07) －位置が局在(L-09) ・ライティング用品（レフ板、ルーバー、デフューザーなど）の不適切な使い方(L-10)

表 2.1.1-1　OCR に影響を与える要因及び現象（続き）

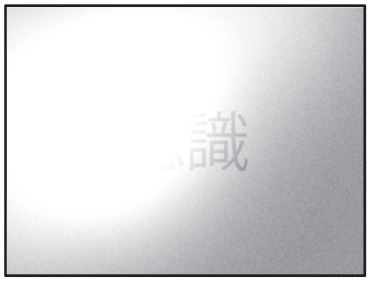
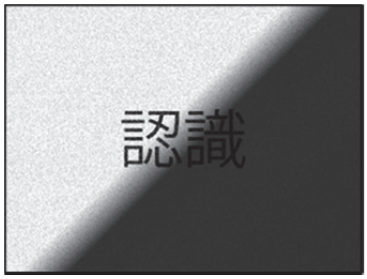
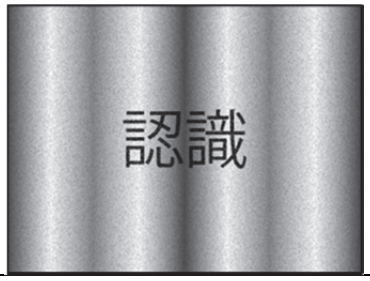
分類	現象	要因
光源	・白飛び L-01，L-05，L-07，L-10   ・シェーディング L-02，L-03，L-05，L-07，L-09，L-10  ・低コントラスト L-01，L-06，L-07，L-10  ・縞々 L-04，L-10  	

表 2.1.1-1　OCR に影響を与える要因及び現象（続き）

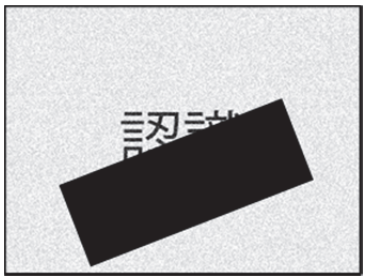
分類	現象	要因
障害物	・像がゆがむ 0-03，0-06，0-08  ・影がおきる 0-01  ・遮蔽される 0-02，0-06  ・ぼける 0-06，0-07  ・低コントラスト 0-06 	・影を引き起こす障害物（撮影者自身（頭、手）、カメラなど）(0-01) ・遮蔽を引き起こす障害物(手の指、紙面の重なり、前に立つ人の頭など)(0-02) ・媒体（水、ガラスなど）(0-03) ・媒体中の障害物（霧、雨、雪など）(0-06) ・レンズへの付着物（皮脂、結露、コーティングなど）(0-07) ・対象物への付着物（結露など）(0-08)

表 2.1.1-1　OCR に影響を与える要因及び現象（続き）

（注）青字はスキャナーベースの OCR にも共通する外乱要因を示す。

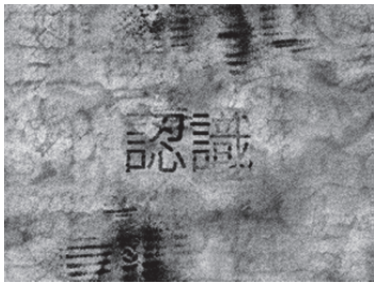
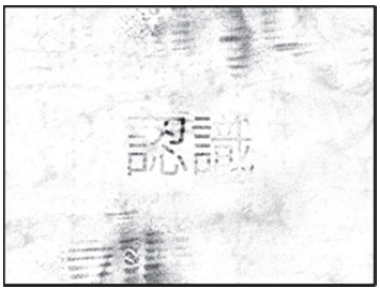
分類	現象	要因
対象物	・汚れている T-07  ・テカっている T-01，T-02，T-03，T-04，T-24，T-25 ・かすれている T-13，T-16，T-17，T-25  ・ゆがんでいる T-05，T-06，T-21，T-22，T-23 ・背景がまぎらわしい T-08，T-09，T-10，T-11，T-19，T-27，T-41，T-42，T-50，T-52，T-53  ・文字が読みにくい T-19，T-23，T-25，T-31，T-34，T-37，T-38，T-44，T-45，T-47，T-51，T-54 ・文字が劣化している T-02，T-04，T-14，T-15，T-16，T-17，T-18，T-20，T-49 ・文字を分離しづらい T-09，T-10，T-11，T-12，T-29，T-35，T-36，T-39，T-40，T-42，T-43，T-50	［画像・画質レベル］ ・ベース -物性 －光沢のある紙（コート紙、複写紙、光沢紙、トレーシングペーパーなど）（T-01） －モニター画面（T-02）、自発光（T-03） －プロジェクター画面（T-04） -空間構成 －立体（T-05） －曲面（皺、波うち、折れ）（T-06） -汚れ・模様 －汚れ（T-07） －背景模様（copy 防止文字、柄が認識対象に似ている、背景が複数色で構成され文字との分離が難しいなど）（T-08） -裏写り（T-09）、透け（T-10） ・カメラ OCR における情景（背景） -認識対象が背景と似ている（T-11） ・文字 －色が一様でない（T-12） －印字不良 －かすれ（T-13） －つぶれ（T-14） －虫食い文字（T-15） －滲み（T-16） －インクこすれ（T-17） －濃度が一様でない（T-18） －経年劣化 -インクがとび圧力跡だけ残ったもの（T-19） －感熱紙の文字劣化（T-20） －回転（T-21） -変形（T-22） -ゆがんだ手書き文字、不正確な手書き文字（T-23） －印字文字（反射）（T-24） -インク特性（うすい、てかる、メタリック）（T-25） -かくれ文字（X 線透過）（T-26） -ベースと文字のコントラストが小さい（T-27）

表 2.1.1-1　OCR に影響を与える要因及び現象（続き）

分類	現象	要因
対象物	・文字列を検出しづらい T-28，T-29，T-30，T-32，T-33，T-48 ・類似文字  T-46 ・隠れている文字 T-26	［論理レベル］ ・レイアウト -段落レベル -縦書き・横書きの混在(T-28) -図の中に文字がある(T-29) -文字列のレベル -直線状に並んでいない文字 － ルビ(T-30) － 上付き文字、下付き文字(T-31) － 二次元的に並んだ文字（化学式、数式など）(T-32) － 行内でのサイズの混在(T-33) － 字体変化(T-34) － 文字接触（合字、カーニング、重なり、重畳）(T-35) ・文字 －白黒反転文字(T-36) -文字種（サイズが小さい）(T-37) -文字装飾（文字飾り）(T-38) -記入枠からのはみ出し(T-39) ・非文字 －罫線との重畳(T-40) －判子の重畳(T-41) -網掛け(T-42) -アンダーライン(T-43) ［認識対象（カテゴリ）］ ・文字 －言語〈日本語記入欄、英語記入欄など〉(T-44) -文字種 －一般的でない字体(T-45) －異体字(T-46) －特殊なフォント(T-47) －化学式や数式などで使われる特殊文字(T-48) － 文字装飾 － ドット文字(T-49) － 彫り付け文字(T-50) － ロゴ (T-51) － 立体文字(T-52) － エンボス(T-53) － 芸術的な文字(T-54)

表 2.1.1-1　OCR に影響を与える要因及び現象（続き）

分類	現象	要因
カメラ/撮影者	・黒つぶれ C-03，C-04，C-06，C-10，C-13，C-14，C-15，C-24  ・白飛び C-03，C-10，C-11，C-13，C-14，C-15，C-24  ・ピンぼけ C-03，C-07，C-08，C-23  ・二重（多重）にみえる C-06，C-08，C-19，C-22 	・光学系の問題 －レンズの特性 －解像度が低い(C-01) －広角すぎる(C-02) －レンズの開口(C-03) －周辺劣化 －減光(C-04) －解像度低下(C-05) －内面反射(C-06) －被写界深度が不足(C-07) －収差(C-08) ・撮像素子 － 画素数が少ない(C-09) － ダイナミックレンジが狭い(C-10) － スミア(C-11) － ローリングシャッター(C-12) ・機構系 －露出が高すぎ/低すぎ（C-13) －EV 値が高すぎ/低すぎる(C-14) －絞りすぎ/絞らなさすぎる(C-15) －シャッター速度が速すぎる/遅すぎる(C-16) ・画像情報の処理 －符号化圧縮ノイズ(C-17) －高圧縮 PDF 化による文字背景の低コントラスト化(C-18) －多重露光処理の失敗(C-19) ・撮影条件 －距離が遠すぎる(C-20) －視点が被写体に対して正対していない(C-21) － カメラの保持（手ブレ）(C-22) － カメラの設定ミス － ピント(C-23) － 露出(C-24) － ズーム(C-25)

表 2.1.1-1　OCR に影響を与える要因及び現象（続き）

分類	現象	要因
カメラ/撮影	・ゆがむ C-02，~~C-10，~~C-12，C-21  ・低コントラスト C-04，C-10，C-13，C-14  ・低解像度 C-01，C-02，C-09，C-20，C-25  ・ノイズ（ホワイトノイズ、縞々ノイズ、ブロックノイズ） C-11，C-16，C-17 	

なお、要因表に挙げた要因及び現象は、文字認識のフローチャートと対応付けて捉えることができる。文字認識のフローチャートの一例を図 2.1.1-2 に示す。この図は、文字切り出し～分類までの処理について、(a)DL 不使用、(b)文字切出以降の処理について DL を使用、(c)すべての処理に DL を使用、の 3 通りの実装方式をひとまとめに描いたものである。この場合、一例として、図 2.1.1-2 に示す処理ステップには表 2.1.1-2 に示すような現象が影響を与えると考えることができる。

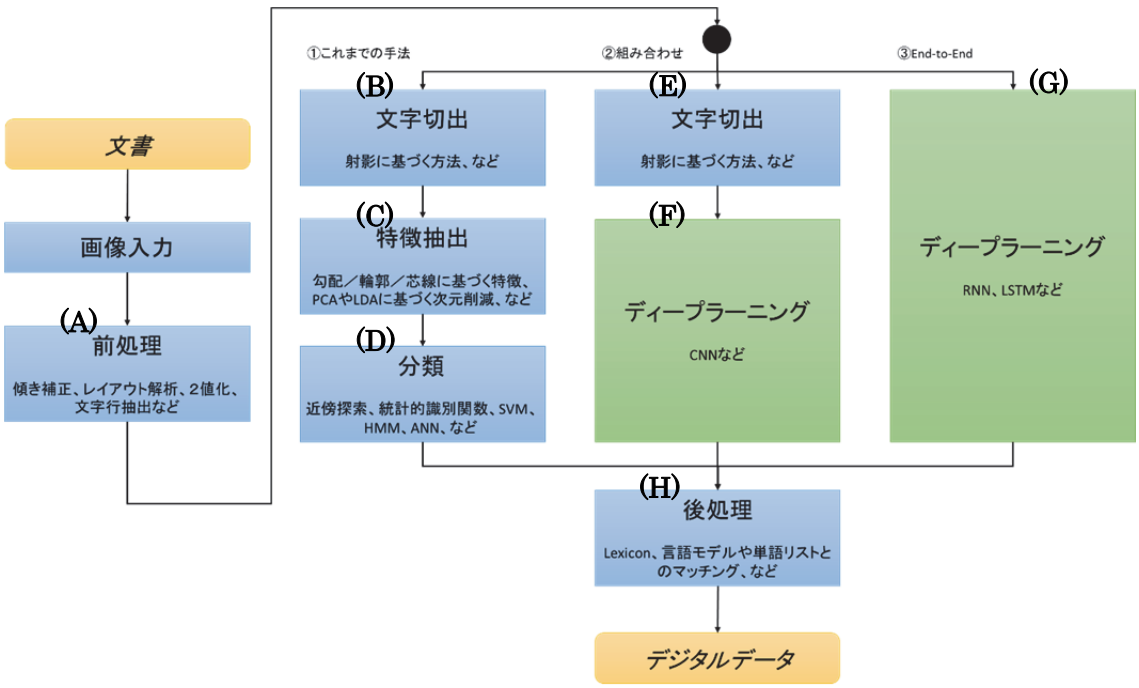


図 2.1.1-2　文字認識のフローチャートの例

表 2.1.1-2　文字認識の処理ステップに影響を与える現象の例

処理ステップ	左記処理ステップに影響を与える現象の例
(A)前処理	黒つぶれ、白飛び、シェーディング、低コントラスト、影がおきる、遮蔽される、テカっている、かすれている、文字列を検出しづらい
(B)文字切出	文字を分離しづらい
(C)特徴抽出	黒つぶれ、白飛び、ぼける、文字が劣化している、ピンぼけ、二重（多重）にみえる、低解像度、ノイズ
(D)分類	像がゆがむ、文字が読みにくい、類似文字
(E)文字切出	文字を分離しづらい

2.1.2　外乱要因表のダイジェスト版

前節で示した要因表は、主として認識形入力方式の専門家向けに、OCR に影響を与える要因及び現象をできるだけ網羅するように作成されている。そのため、通常は起こりにくい要因・現象も含んでおり、技術に明るくないユーザーにとっては難解で扱いづらい。そこで、主に技術に明

るくないユーザーに向けて、カメラで撮影した画像に対して OCR を適用する典型的なユースケースとして、オフィスの机の上に名刺・レシート・伝票などの書類を並べて、スマートフォンで撮影・OCR 認識するケースを想定し、その場合の主要な要因・現象のみを抜粋した要因表のダイジェスト版を作成した。本ダイジェスト版を表 2.1.2-1 に示す。

表 2.1.2-1　OCR に影響を与える要因及び現象

分類	現象	要因
光源	・黒つぶれ L-01，L-07，L-10 ・白飛び L-01，L-07，L-10 ・シェーディング L-02，L-07，L-10 ・低コントラスト L-01，L-07，L-10	・状態及び種類 －明るすぎる/暗すぎる(L-01) －ムラの出る照明(L-02) ・位置及び数 －光源から対象物までの距離が近い(L-07) ・ライティング用品（レフ板、ルーバー、デフューザーなど）の不適切な使い方(L-10)
障害物	・影がおきる 0-01 ・遮蔽される 0-02	・影を引き起こす障害物（撮影者自身（頭、手）、カメラなど）(0-01) ・遮蔽を引き起こす障害物(手の指、紙面の重なり、前に立つ人の頭など)(0-02)
対象物	・汚れている T-07 ・テカっている T-01 ・かすれている T-13 ・ゆがんでいる T-06，T-21 ・背景がまぎらわしい T-19，T-41，T-42 ,T-50，T-52，T-53 ・文字が読みにくい T-19，T-37，T-44，T-45，T-47，T-51，T-54	[画像・画質レベル] ・ベース － 光沢のある紙（コート紙、複写紙、光沢紙、トレーシングペーパーなど）(T-01) － 曲面（皺、波うち、折れ）(T-06) － 汚れ・模様（汚れの付着、背景模様など）(T-07) ・文字 －印字不良 －かすれ(T-13) －つぶれ(T-14) －虫食い文字(T-15) －滲み(T-16) －インクこすれ(T-17) －濃度が一様でない(T-18) －経年劣化 －インクがとび圧力跡だけ残ったもの(T-19) －感熱紙の文字劣化(T-20) －回転(T-21)

表 2.1.2-1　OCR に影響を与える要因及び現象（続き）

分類	現象	要因
	・文字を分離しづらい T-35, T-36, T-39, T-40, T-42, T-43 ・文字列を検出しづらい T-30, T-33 ・類似文字 T-46	［論理レベル］ ・レイアウト － 直線状に並んでいない文字 － ルビ(T-30) － 上付き文字、下付き文字(T-31) － 二次元的に並んだ文字（化学式、数式など）(T-32) － 文字接触（合字、カーニング、重なり、重畳）(T-35) ・文字 －白黒反転文字(T-36) － 文字種（サイズが小さい）(T-37) － 記入枠からのはみ出し (T-39) ・非文字 －罫線との重畳(T-40) －判子の重畳(T-41) － 網掛け(T-42) － アンダーライン(T-43) ［認識対象（カテゴリ）］ ・文字 －言語<日本語記入欄、英語記入欄など> (T-44) －文字種 －一般的でない字体(T-45) －異体字(T-46) －特殊なフォント(T-47) － ロゴ(T-51)
カメラ/撮影者	・ピンぼけ C-23 ・二重（多重）にみえる C-22 ・ゆがむ C-21 ・低解像度 C-20 ・ノイズ（ホワイトノイズ、縞々ノイズ、ブロックノイズ） C-17	・画像情報の処理 －符号化圧縮ノイズ(C-17) ・撮影条件 －距離が遠すぎる(C-20) －視点が被写体に対して正対していない (C-21) － カメラの保持（手ブレ）(C-22) － カメラの設定ミス ピント(C-23)

2.1.3　要因表の活用方法

上記の要因表は様々な用途に活用可能であり、例えば、以下のような使い方が考えられる。

(a)　OCR 技術者がある用途向けに OCR を開発する際に、用意すべき学習データのバリエーション（オーグメンテーション）の軸を知る

(b)　ユーザーが OCR をある用途に利用したところ認識精度が十分でなかった場合、撮影した画像の状態から再確認すべき要因を知る

(c)　ユーザーがある用途に OCR を利用しようとする際に、その利用環境に基づいて、考慮した方がよい要因を知る

本節では要因表の活用のヒントを提供するために、(a)、(b)の例をケーススタディとして取り上げ、活用方法を具体的に説明する。

(a)　活用例１：技術者がある用途向けに OCR を開発する際に、用意すべき学習データのバリエーション（オーグメンテーション）の軸を知る。

例えば、図 2.1.3-1 に示す手順に従って、要因表から用意すべき学習データのバリエーションを絞り込むことができる。以下、一例として、スーパーのチラシに印刷された金額を読み取る OCR を開発するケースを想定して、各ステップの内容について詳しく説明する。

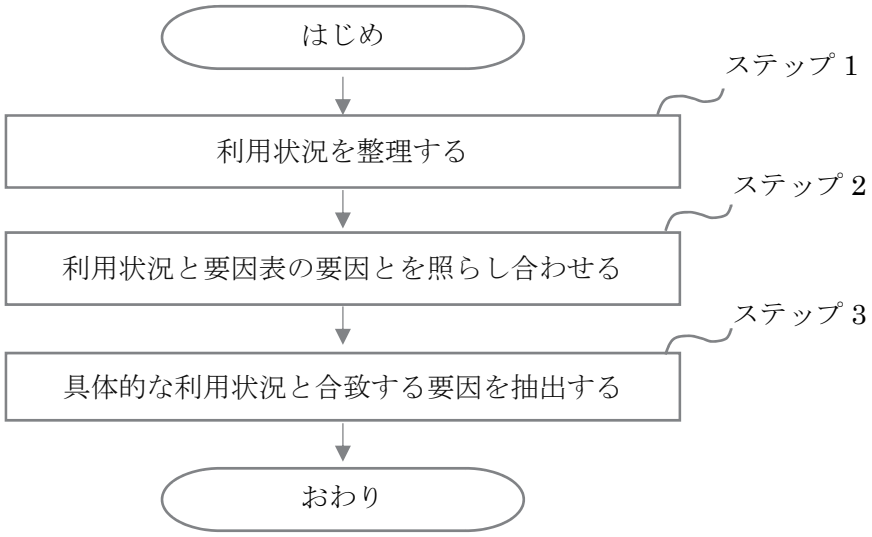


図 2.1.3-1　用意すべき学習データのバリエーションを知る際のフローチャートの例

・ステップ１：利用状況を整理する

まず、開発する OCR の利用状況を整理する。スーパーのチラシに印刷された金額を読み取る OCR を開発するといった場合であれば、例えば、利用状況として、

- チラシを事務所の机の上になるべく平らに広げる
- 机の脇に立ってスマートフォンを手に持ってチラシの画像を撮影する

などを整理する。

・ステップ２：利用状況と要因表の要因とを照らし合わせる

ステップ１で整理した利用状況と要因表の要因とを照らし合わせ、それらが合致するか否か検討する。例えば、障害物に関する要因について、表 2.1.3-1 のように整理することができる。光源、対象物、カメラ／撮影者についても同様に利用状況と要因表を照らし合わせて整理する。

表 2.1.3-1　チラシ読み取りを想定したケースでの、利用状況と要因表の要因との照合例

要因	要因№	具体的な利用状況	要因と利用状況が合致するか否か
・影を引き起こす障害物（撮影者自身（頭、手）、カメラなど）	0-01	撮影者自身などによる影が生じる可能性がある	一致
・遮蔽を引き起こす障害物（手の指、紙面の重なり、前に立つ人の頭など）	0-02	ストラップなどによる隠蔽が生じる可能性がある	一致
・媒体（水、ガラスなど）	0-03	指向性がほぼない LED 照明である	不一致
・媒体中の障害物（霧、雨、雪など）	0-06	屋内なので天候の影響はない	不一致
・レンズへの付着物（皮脂、結露、コーティングなど）	0-07	レンズはきれいにして使用する	不一致
・対象物への付着物（結露など）	0-08	チラシに付着する物はない	不一致

・ステップ３：具体的な利用状況と合致する要因を抽出する

ステップ２の整理から、利用状況が合致する要因のみを抽出する。抽出された要因が用意すべき学習データのバリエーションの軸を示す。例えば、表 2.1.3-2 からは、影を引き起こす障害物、遮蔽を引き起こす障害物、についてバリエーションを揃えればよいと知ることができる。

表 2.1.3-2　利用状況と合致する要因を抽出した場合の一例

要因	要因№	具体的な利用状況	要因と利用状況が合致するか否か
・影を引き起こす障害物（撮影者自身（頭、手）、カメラなど）	0-01	撮影者自身などによる影が生じる可能性がある	一致
・遮蔽を引き起こす障害物（手の指、紙面の重なり、前に立つ人の頭など）	0-02	ストラップなどによる隠蔽が生じる可能性がある	一致

(b) 活用例２：ユーザーが OCR を利用した時に認識精度が十分でなかった場合、撮影した画像の状態から再確認すべき要因を知る。

例えば、図 2.1.3-2 に示す手順に従って、撮影時の注意点を絞り込むことができる。

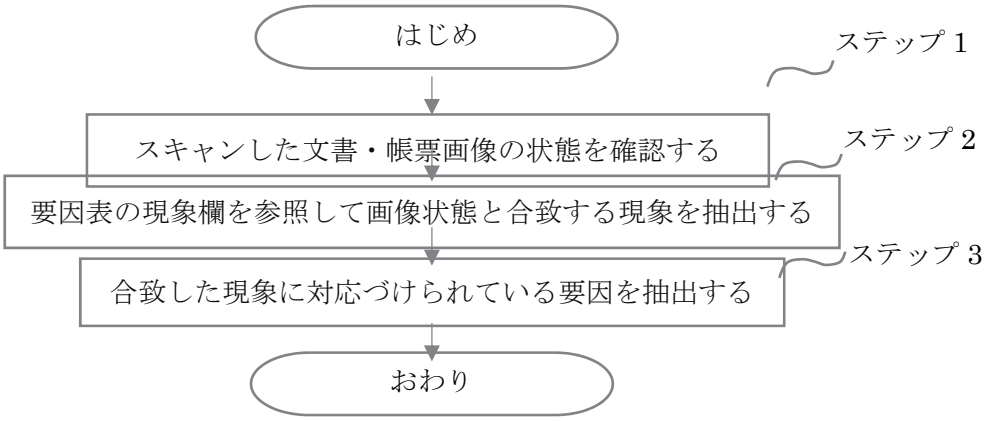


図 2.1.3-2 スキャンした文書・帳票画像の状態から撮影時の注意点を知る際のフローチャートの例

・ステップ１：スキャンした文書・帳票画像の状態を確認する

スキャンした文書・帳票画像の状態（画像中に見られる劣化の現象）を確認する。例えば、画像状態として、

- 白飛びしている
- 照明ムラ（シェーディング）がある
- 対象物にテカリがある

といった項目を抽出する。

・ステップ２：要因表の現象欄を参照して、画像状態と合致する現象を抽出する

要因表の現象欄を参照して、ステップ１で抽出した画像状態と合致する現象を抽出する。上記画像状態を想定した場合の抽出結果の例を表 2.1.3-3 に示す。

表 2.1.3-3 画像状態に合致する現象を抽出した例

分類	現象
光源	・白飛び L-01, L-05, L-07, L-10 ・シェーディング L-02, L-03, L-05, L-07, L-09, L-10
障害物	
対象物	・テカっている T-01, T-02, T-03, T-04, T-24, T-25
カメラ/ 撮影者	・白飛び C-03, C-10, C-11, C-13, C-14, C-15, C-24

・ステップ３：合致した現象に対応づけられている要因を抽出する

要因表を参照して、ステップ２で抽出した現象に対応付けられている要因を抽出する。表 2.1.3-3 に示した例に対しては、表 2.1.3-4 に記載した要因が抽出される。ここで抽出された要因が撮影時の注意点を示す。この中に現在の OCR の使用条件に合致する要因が含まれる場合は、当該要因が合致しないようにできないか検討する。例えば、撮影時にフラッシュをたいていたならば、フラッシュを使用せず撮影を行うようにする。

表 2.1.3-4　ステップ２で抽出した現象に対応付けられている要因を抽出した例

分類	現象	要因
光源	・白飛び L-01, L-05, L-07, L-10 ・シェーディング L-02, L-03, L-05, L-07, L-09, L-10	・状態及び種類 -明るすぎる/暗すぎる(L-01) -ムラの出る照明(L-02) -点光源(L-03) -フラッシュ(L-05) ・位置及び数 -光源から対象物までの距離が近い(L-07) -複数の位置関係位置が局在(L-09) ・ライティング用品（レフ板、ルーバー、デフューザーなど）の不適切な使い方(L-10)
対象物	・テカっている T-01, T-02, T-03, T-04, T-24, T-25	・ベース -物性 － 光沢のある紙（コート紙、複写紙、光沢紙、トレーシングペーパーなど）(T-01) － モニター画面(T-02)、自発光(T-03) － プロジェクター画面(T-04) ・文字 -印字文字（反射)(T-24) － インク特性（うすい、てかる、メタリック）(T-25)
カメラ／撮影者	・白飛び C-03, C-10, C-11, C-13, C-14, C-15, C-24	・光学系の問題 -レンズの特性 － レンズの開口(C-03) ・撮像素子 － ダイナミックレンジが狭い(C-10) － スミア(C-11) ・機構系 -露出が高すぎ/低すぎる(C-13) -EV 値 が高すぎ/低すぎる(C-14) -絞りすぎ/絞らなさすぎる(C-15) ・撮影条件 -カメラの設定ミス － 露出(C-24)

2.2 RPA における UI 操作自動化のための OCR の外乱要因

RPA とは、従来人手で行っていたコンピューター操作（UI 操作）を自動実行する技術である。人が目視で行っていた処理を自動化するために OCR が用いられるので、ここでは RPA において OCR に影響を与える要因について記述する。

RPA では、OCR は 2 種類の方法で利用される。

(1) 紙帳票を OCR で読み込んでデータ化し、そのデータに基づいて RPA が自動処理を実行する【帳票 OCR】

(2) 画面上の UI 操作を自動化（自動運転）することを目的として、操作座標を求めるため、または画面に表示されているテキストをデータとして取得するために、PC 画面のキャプチャー画像を対象として OCR を行う【画面 OCR】

前者(1)はデータ・エントリー業務の自動化で用いられる OCR であり、スキャナーで読み込まれた文書画像を認識対象とする。後者(2)は PC 画面上に表示されたテキストを認識する OCR であり、PC が生成した画面キャプチャー画像を認識対象とする。

前者の場合、OCR で認識した項目データを RPA が利用するが、OCR の認識結果に誤りがあると自動処理が失敗するため、目視によるデータ確認が行われることが多い（図 2.2-1）。つまり、完全な自動処理ではなく、認識結果には人が責任を持つ必要があるが、それでもキーボードでデータを手入力するよりも作業工数の大幅な削減が期待できる。また、若干の誤りならば許容できる用途の場合は、目視確認を省略することもある。

一方、後者の場合は PC 画面のキャプチャー画像を認識して文字データや座標を連続的に取得し、その座標を用いて自動実行をする。つまり、RPA 実行時に認識するため、人による確認が入る余地が無い（図 2.2-2）。これは OCR が認識誤りを起こすと誤動作に直結することを意味するので、OCR の精度に対する要求レベルは前者に比べてより高くなる。そのため後述するような OCR に影響を与える要因の精査がより重要となる。

図 2.2-1　RPA 向けデータ入力自動化のための OCR 利用

図 2.2-2　UI 操作自動化のための OCR 利用

本報告書では、これまでスキャナー及びカメラで取得した画像を対象とした OCR の外乱要因についての整理を行ってきたが、近年、UI 操作の自動化を目的とした画面 OCR の利用が一般化しており、本節において、画面 OCR の精度に影響を与える要因について簡単に述べる。（1.1 節の俯瞰的考察の図によれば、電子情報がアナログ変換を経ずにデジタルパターン情報に変換され、復元された情報が活用される、①⇒②⇒⑤⇒⑥の処理ルートに相当する）

2.2.1　画面 UI 操作の自動化

RPA における画面 UI 操作は、例えば、画面上のボタンをクリックするなど、操作対象のターゲット（ボタン）と操作内容のコマンド（クリック）を記録したシナリオに基づいて UI 操作を別環境で再生する。そのために多くの RPA ツールはユーザーが画面を操作した内容を記録するレコーディング（操作記録）機能を提供している。（図 2.2.1-1）

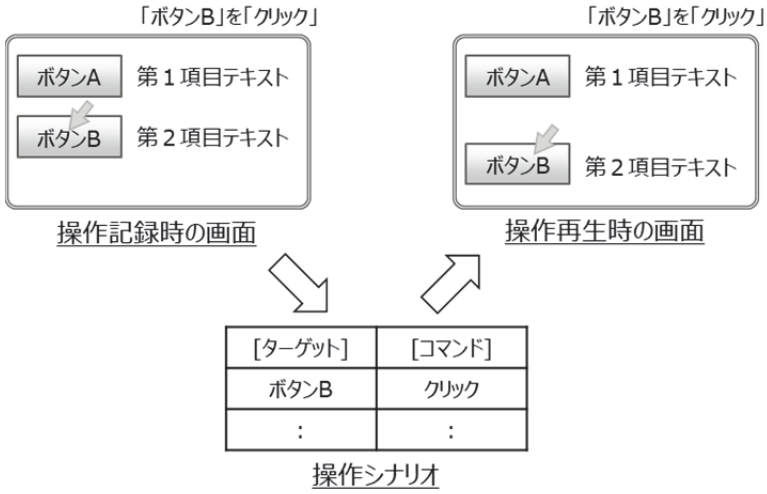


図 2.2.1-1　RPA における UI 操作の記録と再生

操作対象ターゲットは、例えば、Web ブラウザの場合には <button id=“ボタン B”> のように HTML タグで記録された対象を id や xpath で指定することが多い。また、操作対象の座標値を指定することもある。しかし、図 2.2.1-2 のように操作対象の id や表示位置が変化する場合には、操作再生時にターゲットを見つけることができない（操作ターゲットの破損）。また、Web ブラウザ以外のアプリを操作する場合には、そもそも id や xpath によるターゲット指定を行うことはできない。

このような場合の対策として画面キャプチャー画像からターゲットを検索するという手法がしばしば用いられる。例えば、クリックするボタンのアイコン画像を検索してクリック座標を求める方法や、クリック座標の近くにある文字列を検索する方法などがある。その一例を図 2.2.1-3 に示す。この例では、“第 2 項目テキスト” という文字列を検索し、その左隣にあるボタンをクリックすると記録されている。それを再生時の画面で実行すれば、操作対象ボタンの表示位置や id 等

の変化の影響を受けず記録時と同じ UI 操作を実行することができる。ここで文字列を検索するために画面 OCR 技術が使われる。

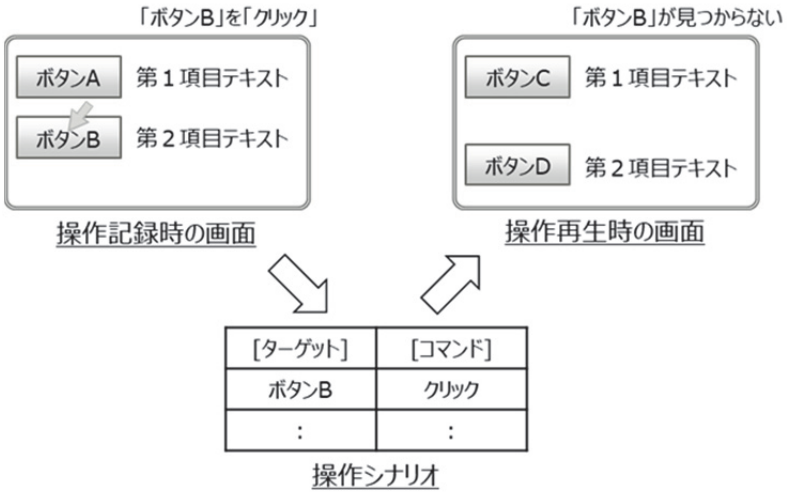


図 2.2.1-2　RPA における UI 操作の記録と再生（失敗例）

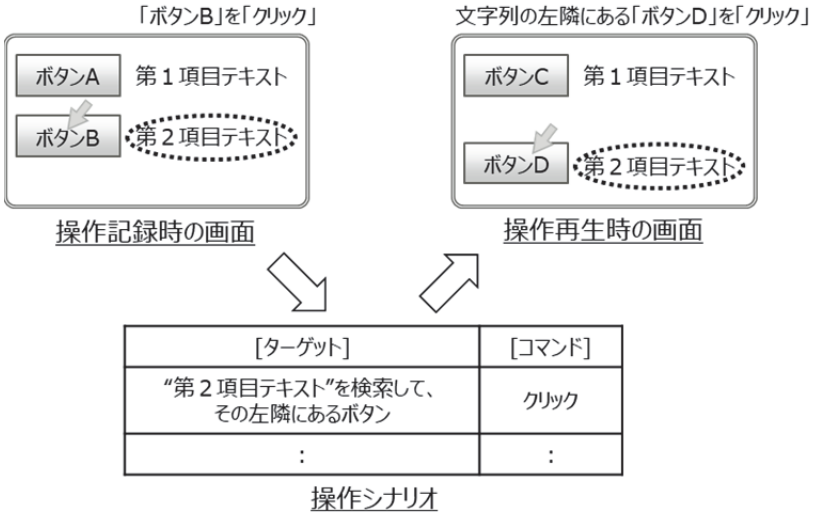


図 2.2.1-3　RPA における UI 操作の記録と再生（OCR の利用）

参考：

Web アプリケーションを対象とした自動テスト技術では、実行テスト時に行う画面操作を繰り返し自動実行するため RPA と同様の技術が用いられている。操作対象ターゲットを特定する機能は locator と呼ばれ、3 種類に分類される（OCR を用いる方法は第三世代に相当）。この研究分野では、3 種類の locator の併用や相互変換※によりターゲットの破損を修正する技術（自動修復：Self-Healing）も提案されている。

1. 第一世代 locator　　操作対象を画面上の座標値で指定する
2. 第二世代 locator　　操作対象を Tag の id や xpath などの属性値によって指定する
3. 第三世代 locator　　画面表示画像から画像処理技術によって検索することによって対象を指定する

※ “Automated Generation of Visual Web Tests from DOM-based Web Tests”, Maurizio Leotta, Andrea Stocco, Filippo Ricca, Paolo Tonella (SAC 2015, Apr. 2015)

2.2.2　画面 OCR に影響を与える要因

画面キャプチャー画像を対象とした OCR では、図 2.2.2-1 に示すような撮像過程において画像の質的変換が行われるため、OCR の精度に影響が生ずる。画面 OCR の実行条件を設定する際にはこれらの要因を考慮する必要がある。

実世界環境における OCR では文書面（紙など）に当たった照明の反射光を検出して文字画像が生成されるが、画面 OCR では文書面で生成される文字画像は電子的に生成されたものなので、文字画像には変動要因はほとんど無いように思える。しかしながら、図 2.2.2-1 に示すように、文字画像は文書データからフォント情報に基づいて生成されるため、例えば、ビットマップフォントかアウトラインフォントか、アンチエイリアスの有無、等により文字画像の品質が変化する。また、画面上には文字以外の図形も表示されており、文字を隠したり、文字に近接してオブジェクトが配置されていたりする（例：罫線が文字に接触）と、文字領域を正確に切り出すのが困難になる。更に画面サイズに応じて画像が拡大・縮小されるなど、拡大率や解像度による画像品質の変化も生じ得る。画面 OCR の精度を維持（保証）するためにはこのような変動要因を考慮する必要がある（表 2.2.2-1）。

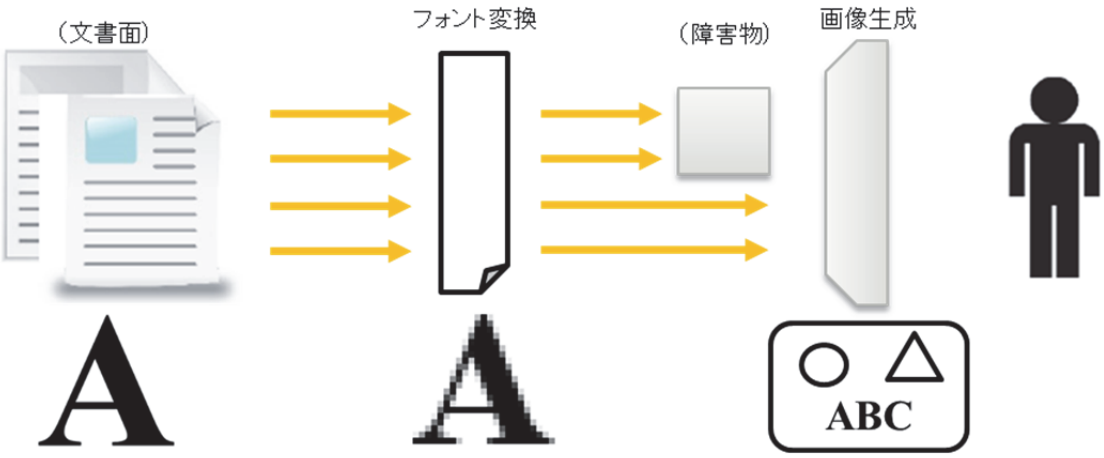


図 2.2.2-1　画面キャプチャー画像の撮像過程に影響を与える要因

表 2.2.2-1　画面 OCR に影響を与える要因

分類	要　因
フォント変換	・フォントの種類 　-アウトラインフォント、ビットマップフォント、画像 ・文字画像生成 　-変換無し、アンチエイリアス（スーパーサンプリング、マルチサンプリング、等）
障害物	・文字を隠す障害物 　-文字に重なるオブジェクト、文字間の重なり ・文字抽出の妨げになる障害物 　-文字以外の近接オブジェクト、背景模様
画像生成	・表示品質 　-表示拡大率（ディスプレイ、ブラウザ）、画面解像度

2.2.3 RPA における OCR 認識誤りの影響

OCR の認識結果を RPA 実行のための入力とする場合、認識誤りが発生すると動作にエラーが生ずるため、認識誤りを避ける対策が必要となる。そのため、OCR を事前に実行して目視確認を行うことにより、精度保証の責任を人間に任せるという手段が用いられることが多い（図 2.2-1）。しかし RPA 実行時にその場で OCR を行う UI 操作自動化においては（図 2.2.1-3）、OCR を含むシステム自身が精度保証を求められるため、①十分に実用になる程度の認識精度を保証する、②仮に認識誤りが発生してもユーザーの用途にとって致命的な影響を与えないことを保証する、のいずれかが求められる。

以上のように認識誤りの可能性を考慮して適切な品質保証を行うことは、パターン認識技術（OCR を含む）を実用化する上で不可欠である。近年では AI 品質保証の問題として議論が活発化しており、また、類似の問題として生成 AI が誤りを含む出力をするハルシネーション（幻覚）という現象も話題に上っている。詳しくは 2.4 節にて述べる。

2.3 実世界環境における認識技術の現状と今後の展望

当委員会では主に OCR を中心とした認識形入力方式について調査検討を継続している。特筆すべきトピックとして、大規模言語モデル（LLM：Large Language Model）を始めとする生成 AI 技術が、画像や音声を言語と同等に扱うことが可能な大規模マルチモーダルモデル（LMM：Large Multi-modal Model）[1]へと進化し、先端研究の発展だけでなく実務利用も始まりつつあるところであり、認識形入力方式にとって大きなパラダイムシフトとなりつつある。そこで本節では、LMM のうち画像と自然言語をつなぐ背景技術である Vision-and-Language モデル（VLM）について 2.3.1 で述べ、2.3.2 で画像生成 AI、2.3.3 で LMM への展開について現状をまとめ、2.3.4 で認識形入力方式としての今後の展望を述べる。

2.3.1 Vision-and-Language モデル（VLM）

画像と自然言語テキストの関係性に関する AI モデルは Vision-and-Language モデル(VLM)と呼ばれ、Deep Learning の性能が向上した 2010 年頃以降盛んに研究されるようになった。例えば、画像を 1 枚与えた時にその概要（キャプション）をテキストで出力する Image Captioning や、画像と質問のペアから回答をテキストで出力する Visual Question and Answering（VQA）といったタスクに取り組まれてきた。従来の VLM では、ラベルが与えられた教師データを用いて学習した物体検出やセグメンテーションの出力を RNN や Transformer 等に与えてテキスト生成を行うものが多かった。しかしながらこうした方法では、人手で付与するラベルの量の限界や、エラーデータなどの品質的な問題により性能向上に限界があるとされた。また、教師データに含まれない対象は根本的に扱うことができない。

そうした中、VLM の汎用的な能力獲得への大きな前進が OpenAI による CLIP（Contrastive Language-Image Pre-Training）[1]である。CLIP では画像と自然言語キャプションのペアのみから学習を行う。人手による正確で大量のラベルが不要であり、インターネットをクロールして収集した約 4 億セットという膨大な教師データを使うことができる。ここで問題となるのは、自然言語では例えば、「2 匹の犬が草原を楽しそうに走り回っている」というキャプションが付いていたとして、数「2 匹の」、物体クラス「犬が」、背景「草原を」、動作「走り回っている」などの情報が混在したうえで、「楽しそう」などの主観的で曖昧な表現も含んでいる。一方で、画像内の物体座標などの精緻な情報は含まれていない。物体認識であれば「犬」「猫」といったクラス分類、物体検出であれば外接矩形（Bounding Box）の座標を回帰させれば良かったが、複雑かつ曖昧で不十分な情報しかない自然言語と画像との関係をどのように学習させたらよいかという問題に対して、教師なし学習（Unsupervised Learning）の分野で研究されてきた対照学習（Contrastive

[1] Multi-modal Large Language Model (MLLM)や、Multi-modal Foundation Model (MFM)とも呼ばれている。本節では Large Multi-modal Model(LMM)で記述を統一する。

Learning）を導入している。SimCLR[2]や MoCo[3]等の対照学習は、1 枚の画像からクロッピングなどのデータ拡張で用いられる何かしらの方法で加工した 2 枚のペアの画像が同じものである、という単純なルールを学習するものである。同じように CLIP では画像とキャプションのペアが同じものであるというルールを元に、先述の自然言語の混在した情報や曖昧性なども含めて学習することを可能としている。

CLIP の特筆すべき点として、Zero-shot Learning の性能が挙げられる。事前学習済みモデルを特定のタスクに適用するには、そのタスクに合わせた学習データを用いてファインチューニングを行う Few-shot Learning が普通であったが、CLIP ではテンプレートを用いた簡単なプロンプトエンジニアリングだけで、未知のデータに対して優れた認識結果を得ることができる。CLIP は約 4 億という膨大なデータを事前学習していることから、既に汎用的な画像認識能力の一部を獲得しているのではないかと言える。

CLIP の発表を発端として、Salesforce の BLIP[4]などの多数の VLM の改良が発表され、画像生成 AI や LMM へと発展が続いている。

2.3.2 VLM から画像生成 AI へ

画像生成 AI は従来オートエンコーダーや敵対的生成ネットワーク（GAN）等を用いた研究が行われてきたが、CLIP 等の VLM と拡散モデル（Diffusion Model）[5]により、テキストから画像を高品質に生成すること（Text to Image）が可能となり、2022 年に立て続けに発表となった。OpenAI の DALL-E 2[6]は、CLIP の潜在変数空間から拡散モデルにより画像を生成するとされる。Stability AI の Stable Diffusion[7]では、VQ-GAN[8]で学習された Transformer ベースの VQ-VAE[9]の潜在変数空間に対して拡散モデルを適用し、高解像度かつ高品質の画像生成が可能となった。Stable Diffusion では CLIP のテキストエンコーダ出力を拡散モデルにおける U-Net の各層に挿入することで、テキストプロンプトを画像生成に反映させている。

Stable Diffusion の特筆すべき点としては、先行する DALL-E 2 や Midjourney がクラウドサービスとして利用できるのみであるのに対して、オープンソースとしてコードとモデルパラメータが公開されたことにある。これにより、画像生成 AI の研究が加速度的に進むこととなる。例えば、ControlNet[10]はテキストプロンプト以外に線画のラフや OpenPose のボーンモデル等を入力して生成画像を制御することができる。拡散モデルの処理は最新 GPU であっても時間が掛かるものであるが、高速化についても多数取り組まれており、Latent Consistency Model[11]では、従来数十ステップが必要なデノイズ処理を、特別なモデル蒸留により数ステップで完了させることができ、StreamDiffusion[12]では処理をパイプライン化することで 90FPS 以上のリアルタイム生成が可能となっている。

画像生成 AI には、生成された画像が他人の著作物に似てしまう、他人の著作物が学習データに

含まれる、といった権利問題があり、海外では裁判で争われている事例もある。そういった中、商用サービスとして Adobe は Photoshop で利用できる画像生成 AI である Firefly をリリースした。Firefly のモデルの学習には権利関係に問題のある著作物が含まれておらず、生成された画像は安全に利用できるとされている。

Open AI は ChatGPT の追加機能として、DALL-E 3 をリリースした。GPT-4 の高性能な自然言語生成と連動して動作するため、プロンプトに対してより的確な画像を生成できる傾向にある[13]。また、Open AI と連携する Microsoft も Bing AI にて DALL-E による画像生成サービスをリリースした。

2.3.3 VLM と LLM から大規模マルチモーダルモデル(LMM)へ

2022 年に大きな話題を呼んだ OpenAI の LLM である ChatGPT は、2023 年 3 月によりパラメータ数が多く高性能な GPT-4[14]をリリースし、有償サービスを開始した。その後 9 月には VLM を組み込み大規模マルチモーダルモデル（LMM）となった GPT-4V（ision）[15]をリリースし、順次利用できるようになった。

ChatGPT のリリース以降、LLM とその発展である LMM の開発競争が加速している。Google は 2023 年 12 月に画像や動画、音声などの様々な情報を処理できる LMM である Gemini を法人向けにリリースした。オープンソースでは Meta AI による LLM である LLaMA[16]などが公開され、それらを元に多数の LLM の研究が展開されている。CLIP 等の VLM と融合した LMM としては、BLIP 2[17]、LLaVA[18]、DeepMind の Flamingo[19]とそのオープンソース版である Open Flamingo[20]などが発表されている。

2.3.4 認識形入力方式としての今後の展望

今後は LMM を認識形入力方式として多くの場面で活用されて行くであろうことが予想される。現時点でも OpenAI の GPT-4V は、不定型フォーマットの OCR として利用することができる。例えば、図 2.3.4-1 は、SSD の写真を GPT-4V に与えたもので、型式や各種スペックはもちろん、各種の認証マークまで認識してテキストとして出力している。さらに、プロンプトを与えることで JSON フォーマットなど機械可読容易な形式での出力も可能である。しかしながら、GPT-4V の OCR に関する性能については、英語やラテン系言語の認識は良好であるが、非ラテン系言語、手書き数字、テーブル構造などの複雑なタスクでの限界が見られ、既存の最新 OCR モデルを上回る性能ではないと報告されている[21]。

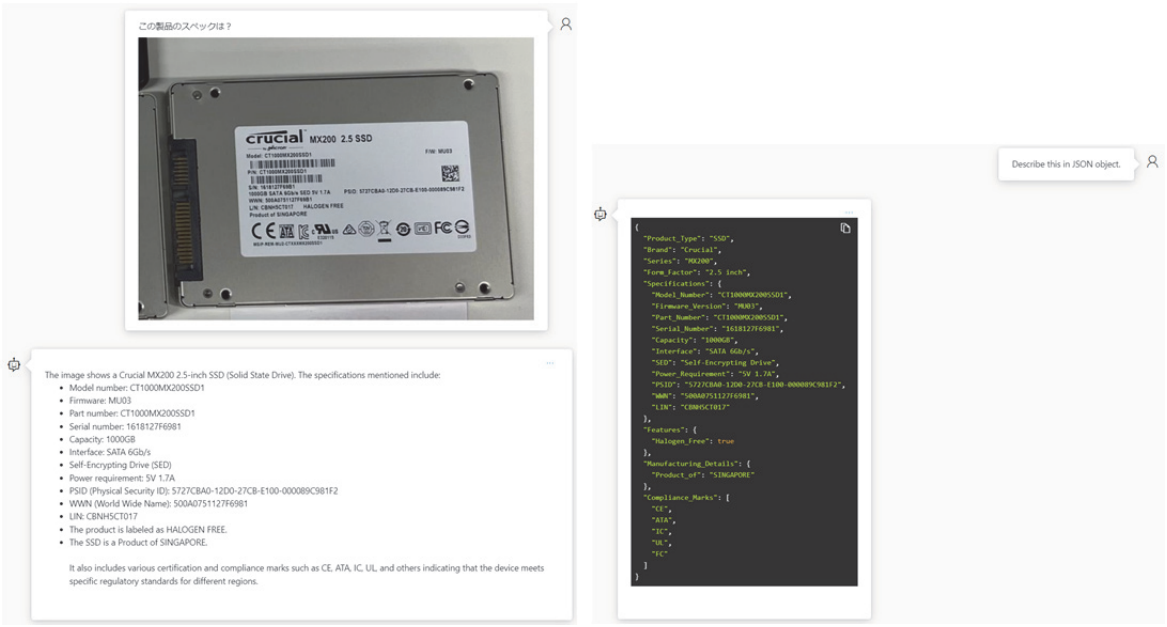


図 2.3.4-1　GPT-4V による不定型フォーマットの認識と出力

最新 OCR モデルとしては、VLM をベースに文書画像全体を入力し、機械可読容易な形式を直接出力する end-to-end なアプローチに注目したい。例えば、Donut[22]は不定型フォーマットの文書画像全体を VLM に入力し、プロンプトに従った要素を JSON フォーマットで出力をする。Nougat[23]は Donut をベースに学術論文の画像から、数式も含めて直接 TeX 形式で出力する。

VLM のニューラルネットワーク内で、文字や文章がどのように理解されているか考察する。画像生成 AI は 2.3.2 で述べたとおり、CLIP 等の VLM がネットワーク内に組み込まれていることから、生成させた画像から、テキストの関係性についてどのような情報が織り込まれているか、考察することができる。図 2.3.4-2 は画像生成 AI に "JEITA"と"Generative AI"というヘッドラインをもつ新聞の画像を生成させたものである。(a)は Stable Diffusion [7]の初期のバージョン（1.5）による出力結果である。文字をそれっぽい形状として出力しているが、人が読むことはできないものであり、文字とテキストの関係に結び付きがないと思われる。パラメータ数、学習データ数、学習計算時間が増加した(b)の Stable Diffusion XL では、"JEITA"が"JEETA"になるなど若干異なっているが、文字の理解が進みつつあるように見える。(c)の GlyphControl[24]は、ネットワークに OCR を組み込むことで文字とテキストの関係性を明確に学習しており、さらに ControlNet[10]により制御することで任意のテキストを正確に表現することができる。

このように、GPT-4V や Stable Diffusion XL のような基盤モデルは、パラメータ数、学習データ数、学習計算時間の増加により、文書認識の能力を獲得し向上を続けている。さらに OCR モデルの組み合わせや文書認識タスクへの最適化により、認識形入力方式としての性能向上と、様々な文書形式や応用分野への展開が期待できる。

(a)　Stable Diffusion 1.5

(b)　Stable Diffusion XL

(c)　GlyphControl

図 2.3.4-2　画像生成 AI におけるテキストを含む画像生成の例

（プロンプト:Newspaper with the headline "JEITA" and "Generative AI"）

文献

[1] Alec Radford, et al., Learning Transferable Visual Models From Natural Language Supervision, ICML 2021, arXiv:2103.00020

[2] Ting Chen, Simon Kornblith, Mohammad Norouzi, Geoffrey Hinton, A Simple Framework for Contrastive Learning of Visual Representations, ICML 2020, arXiv:2002.05709

[3] Kaiming He, Haoqi Fan, Yuxin Wu,Saining Xie,Ross Girshick, Momentum Contrast for Unsupervised Visual Representation Learning, CVPR 2020, arXiv:1911.05722

[4] Junnan Li, Dongxu Li, Caiming Xiong, Steven Hoi, BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation, ICML 2022, arXiv:2201.12086

[5] Jonathan Ho, Ajay Jain, Pieter Abbeel, Denoising Diffusion Probabilistic Models, NeurIPS 2020, arXiv:2006.11239

[6] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, Mark Chen, Hierarchical Text-Conditional Image Generation with CLIP Latents, arXiv:2204.06125

[7] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer, High-Resolution Image Synthesis with Latent Diffusion Models, CVPR 2022, arXiv:2112.10752

[8] Patrick Esser, Robin Rombach, Björn Ommer, Taming Transformers for High-Resolution Image Synthesis, CVPR 2020, arXiv:2012.09841

[9] Aaron van den Oord, Oriol Vinyals, Koray Kavukcuoglu, Neural Discrete Representation Learning, NIPS 2017, arXiv:1711.00937

[10] Lvmin Zhang, Anyi Rao, Maneesh Agrawala, Adding Conditional Control to Text-to-Image Diffusion Models, ICCV 2023, arXiv:2302.05543

[11] Simian Luo, Yiqin Tan, Longbo Huang, Jian Li, Hang Zhao, Latent Consistency Models: Synthesizing High-Resolution Images with Few-Step Inference, arXiv:2310.04378

[12] Akio Kodaira, et al., StreamDiffusion: A Pipeline-level Solution for Real-time Interactive Generation, arXiv:2312.12491

[13] Zhan Shi, Xu Zhou, Xipeng Qiu, Xiaodan Zhu, Improving Image Captioning with Better Use of Captions, ACL 2020, arXiv:2006.11807

[14] OpenAI. GPT-4 Technical Report, arXiv:2303.08774

[15] OpenAI, GPT-4V(ision) System Card, https://cdn.openai.com/papers/GPTV_System_Card.pdf

[16] Hugo Touvron, et al., LLaMA: Open and Efficient Foundation Language Models, arXiv:2302.13971

[17] Junnan Li, Dongxu Li, Silvio Savarese, Steven Hoi, BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models, ICML 2023, arXiv:2301.12597

[18] Haotian Liu, Chunyuan Li, Qingyang Wu, Yong Jae Lee, Visual Instruction Tuning, NeurIPS 2023, arXiv:2304.08485

[19] Jean-Baptiste Alayrac, et al., Flamingo: a Visual Language Model for Few-Shot Learning, NeurIPS) 2022, arXiv:2204.14198

[20] Anas Awadalla, et al., OpenFlamingo: An Open-Source Framework for Training Large Autoregressive Vision-Language Models, arXiv:2308.01390

[21] Yongxin Shi, et al. Exploring OCR Capabilities of GPT-4V(ision) : A Quantitative and In-depth Evaluation, arXiv:2310.16809

[22] Geewook Kim, et al. OCR-free Document Understanding Transformer, ECCV 2022, arXiv:2111.15664

[23] Lukas Blecher and Guillem Cucurull and Thomas Scialom and Robert Stojnic, Nougat: Neural Optical Understanding for Academic Documents, arXiv:2308.13418

[24] Yukang Yang, Dongnan Gui, Yuhui Yuan, Haisong Ding, Han Hu, Kai Chen, GlyphControl: Glyph Conditional Control for Visual Text Generation, arXiv:2305.18259

2.4 OCR の品質保証

2.2 節で述べたように、OCR の認識結果を自動処理に用いる場合は認識誤りによる誤動作リスクが問題となる。OCR をはじめとするパターン認識技術を実用化するためには、誤認識によるダメージを許容範囲内に収める必要があるが、どの程度のリスクなら許容可能かも含めたリスク管理の考え方はパターン認識の分野では十分に浸透していない。そのためソフトウェア開発におけるリスク管理の基準をそのまま適用し、例えば、「0.01%でも誤動作の可能性があるのならテスト不合格」、「誤認識はソフトウェアの誤動作であり、瑕疵とみなす」のような極端な保証要求が設定される場合があるなど、パターン認識技術の実用化を阻む要因となっている。

近年では深層学習の技術的な発展に伴う AI システムの普及により、信頼できる AI の実現が広く求められている。識別器としての AI は OCR やパターン認識を包含する概念であり、OCR 実用化のための品質保証問題は、そのまま「AI 品質」の問題として議論・検討することができる。AI 品質保証の問題は、更に広く AI を適正に利活用するためのガイドライン策定の動きとも絡んでおり、国内外の公的機関が様々な指針を提示している。それは単に AI システムの動作の安定性・確実性といったレベルの問題から、AI を用いることの社会的な影響など、より広い視点での議論や考察を含む。下記に主な国内のガイドライン策定プロジェクトと、AI 実用化に向けた研究開発プロジェクトを挙げる。その他にも安心して使える AI の普及を目指した動きは多く、本稿ではそれぞれの動きの詳細まで記すことはできないので、興味がある方は参考文献を参照して欲しい。(IT 各社はそれぞれ AI 利用における信頼性や責任について企業としてのスタンスを宣言していることが多い。参考文献 13 以降にてその一部を紹介する)

■機械学習品質マネジメントガイドライン（通称：産総研 AI 品質ガイドライン）[1]

https://www.digiarc.aist.go.jp/publication/aiqm/

NEDO の委託に基づいて産総研が主導して作成した、機械学習の品質マネジメントに関するガイドラインである。企業・大学などの外部有識者と共にまとめたもので、AI を用いた製品やサービスの品質を安全、安心に管理するための指針を提供する。

ガイドライン本編は 2023 年 12 月にリリースした第 4 版が最新である。英語版は 2023 年 1 月に第 3 版が公開された。また、ガイドラインの具体的な使用法の例を記した「機械学習品質マネジメントリファレンスガイド」も別途提供されており、その中に郵便番号 OCR の事例も記述されている。

第 4 版の特徴は「機械学習品質マネジメントにおける AI の新潮流への対応」という別冊（Annex）が追加されている点である。この新潮流は主に生成 AI（基盤モデル）の影響についての論点を意味する。

■AI プロダクト品質保証ガイドライン（通称：QA4AI ガイドライン）[2]

http://www.qa4ai.jp

QA4AI とは、2018 年 4 月に設立した産学コンソーシアムである。Web ページの説明によれば、“AI プロダクトの品質保証に関する調査・体系化、適用支援・応用、研究開発を推進すると共に、AI プロダクトの品質に対する適切な理解を社会に啓発する活動を行うコンソーシアム”であり、“AI 技術の活用・進化のさらなる促進と、AI プロダクトと社会との安心できる共生の実現を”目指すとされている。

QA4AI が作成・公開している AI 品質保証のためのガイドラインが「AI プロダクト品質保証ガイドライン」であり、2019 年 5 月に初版がリリースされ、現時点で 2024 年 1 月版が最新である（英語版は 2022 年 1 月版）。

■Engineerable AI プロジェクト（通称：eAI プロジェクト）[3]

https://engineerable.ai/

https://www.jst.go.jp/mirai/jp/program/super-smart/JPMJMI20B8.html

eAI プロジェクトは、国立情報学研究所の石川冬樹准教授を代表として、AI の安全性、信頼性を確立する“Engineerable AI（eAI）”という汎用的基盤技術の開発を目指した、JST 未来創造事業に採択されたプロジェクトである。2020 年の探索研究を経て、2021 年 1 月から本格プロジェクトが開始している。主に自動運転 AI と医療診断 AI をターゲットとして、信頼できる AI システムの構築技術の検討が行われている。

【参考文献】

[1] 機械学習品質マネジメントガイドライン（通称：産総研 AI 品質ガイドライン）
https://www.digiarc.aist.go.jp/publication/aiqm/

[2] AI プロダクト品質保証ガイドライン（通称：QA4AI ガイドライン） http://www.qa4ai.jp/

[3] Engineerable AI プロジェクト（通称：eAI プロジェクト） https://engineerable.ai/

[4] 我が国の AI ガバナンスの在り方 ver1.1（経産省 2021 年 7 月 30 日）
https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/2021070901_report.html

[5] AI ネットワーク社会推進会議（総務省）
https://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/index.html

[6] AI 応用システムの安全性・信頼性を確保する新世代ソフトウェア工学の確立（JST CRDS 2018 年 12 月） https://www.jst.go.jp/crds/report/CRDS-FY2018-SP-03.html

[7] 人間中心の AI 社会原則（内閣府） https://www8.cao.go.jp/cstp/aigensoku.pdf

[8] Trust and Artificial Intelligence（NIST）
https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=931087
[9] ISO/IEC JTC 1/SC 42: Artificial Intelligence
https://www.iso.org/committee/6794475.html
[10] Explainable Artificial Intelligence (XAI)
https://www.darpa.mil/program/explainable-artificial-intelligence
[11] Ethics guidelines for trustworthy AI
https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
[12] AI ガイドライン比較表（総務省） https://www.soumu.go.jp/main_content/000630126.pdf
[13] 責任ある信頼された AI（マイクロソフト）
https://learn.microsoft.com/ja-jp/azure/cloud-adoption-framework/innovate
/best-practices/trusted-ai
[14] Google と AI：私たちの基本理念（Google）
https://japan.googleblog.com/2018/06/ai-principles.html
[15] 富士通グループ AI コミットメント https://pr.fujitsu.com/jp/news/2019/03/13-1a.pdf
[16] 東芝グループ AI ガバナンスステートメント
https://www.global.toshiba/jp/technology/corporate/ai-statement.html
[17] NEC グループ AI と人権に関するポリシー
https://jpn.nec.com/press/201904/images/0201-01-01.pdf
[18] リコーグループ AI 活用基本方針
https://jp.ricoh.com/-/Media/Ricoh/Sites/jp_ricoh/security
/pdf/basic_policy_for%20_AI_technology_utilization_japanese.pdf

3. 認識技術の動向

3.1 文字認識技術の現状と今後の展望

本節では、主に企業におけるビジネス適用の観点から、文字認識／DL 技術の動向について述べる。

3.1.1 文字認識技術概説

OCR に代表される文字・文書認識は、パターン認識研究の最初期から取り組まれている研究課題である。パターン認識分野における様々な基礎的技術が文字認識研究を通して開発され、他の領域へと応用されてきた。文字認識分野では、ETL9 や MNIST データセットに代表される認識手法を評価するための標準的データセットが早期に整備されており、長年にわたりパターン認識や機械学習研究で提案される種々の認識手法のベンチマーキングに活用されている。

図 3.1.1-1 に文字認識処理の流れを示す。文字認識は書面全体、または書面に書かれた文字を各種マークアップファイルや文字コードで表現されるデジタルデータに変換する処理であり、主な構成要素は以下に示す 6 要素である。

1．画像入力：文書・書面など、文字が記された現実媒体を、スキャナやカメラなどを通じデジタル画像として計算機に取り込む処理
2．前処理：文書画像のノイズ除去、コントラスト強調、二値化などのデジタル画像に含まれる文字を見やすくし、文書のレイアウトを解析する、モデル定義に従うなどして行の切り出しを行う処理
3．文字切出：文字列切出によって切り出された文字行から 1 文字ずつの切り出しを行う処理
4．特徴抽出：文字の見かけを表現する特徴量を抽出し、特徴ベクトルを構成する処理。抽出された特徴ベクトルをより識別に有用な特徴ベクトルに変換する（例えば、計算量を削減する目的で特徴ベクトルの次元数を削減する）処理を加えることもある
5．分類：入力された特徴ベクトルに対し、あらかじめ学習済みの分類機を用いて文字ラベルを割り当てる処理
6．後処理：辞書データや言語モデルとのマッチングなどにより認識結果を補正して認識精度を向上させる処理

かつてはこれらの処理それぞれを個別に人間が考案・実装する例が多かったが、近年は少なくとも一部を DL 技術により実現するがほとんどである。典型的には特徴抽出・分類を DL 化（図 3.1.1-1 中②組み合わせ)、文字切出・特徴抽出・分類を DL 化（図 3.1.1-1 中③End-to-End）の 2 例がある。前者では CNN（Convolutional Neural Network）が、後者では RNN（Recurrent

Neural Network)、特にその拡張である LSTM（Long Short Term Memory）をベースとする手法[1]が用いられることが多い。これらの手法は、十分な学習データをすることができていれば、人手で複雑なアルゴリズムを実装する場合よりも高い認識精度が得られることに特徴がある。このため、近年の文字認識製品は基本技術として DL、特に End-to-End 型の DL を用いたものが中心となっている。

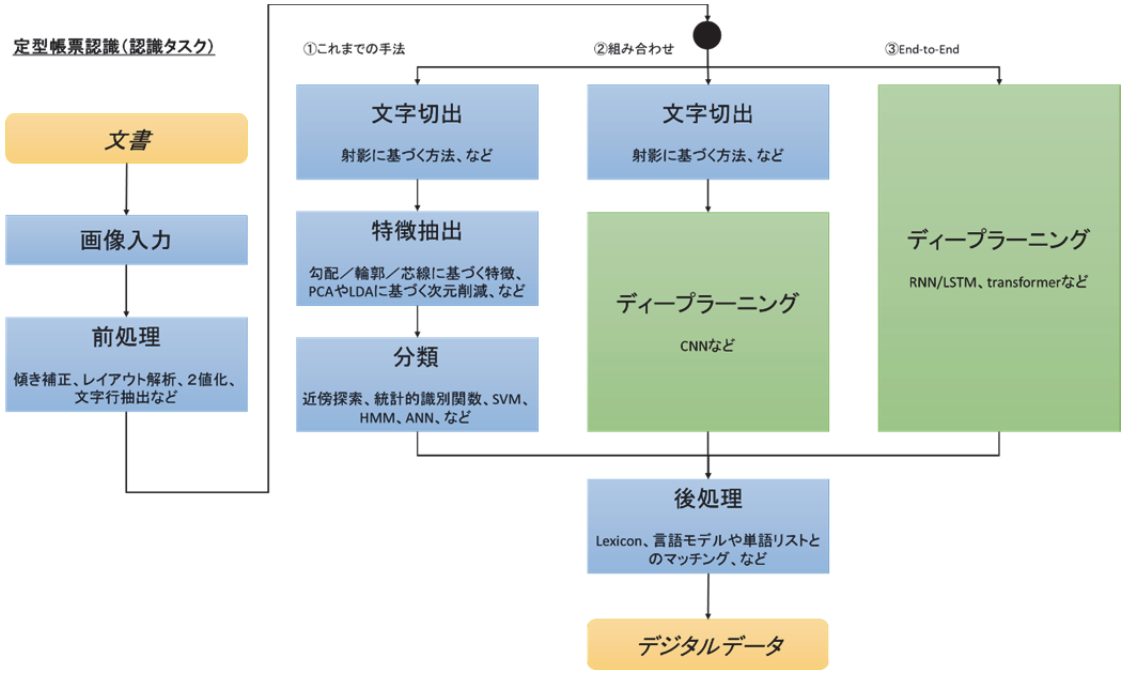


図 3.1.1-1　文字認識技術の概略

最近では、発行元などに応じた多様なデザインを持つ同種の帳票（いわゆる非定型帳票）、典型的には請求書や給与支払報告書などを対象とする OCR が目覚ましい発展を見せている。非定型帳票を対象とする OCR は、帳票内から事前に定められた情報（例えば、請求書であれば請求金額など）を抽出・出力するもので、利用者が事前に多くの読取情報定義（特に読取位置の指定）を行う必要がないために利用開始のコストが低い点に特徴がある。

図 3.1.1-2 に非定型帳票認識処理の概要を示す。現在の非定型帳票認識では、帳票内にある文字行を認識したのち、認識結果であるテキストをもとに必要な情報と不要な情報を分類し必要なもののみを出力する方式が代表的である。分類は単語リストや書式などのルールに基づくもの、AI 技術を活用した学習ベースのものと大きく２種類に分けられ、近年は技術発展によりルールベースから学習ベースへの移行が進みつつある段階である（両者を組み合わせる場合も多い）。学習ベースの代表的な技術として LayoutLM が挙げられる。この技術ではいわゆる Transformer 構造により画像とテキスト（認識結果）を融合させた情報から必要な項目を抽出しており、OCR と同様に学習データの増強で精度を向上させることができる。

今後本格化する可能性がある情報抽出の異なるアプローチとして、いわゆる大規模言語モデル（LLM：Large Language Model）を活用した項目判定への取り組みが想定される。この方法はモデル自体を学習することはなく、モデルに質問や例示を与える際の適切な与え方の開発、いわゆるプロンプトエンジニアリングによって目的を達成する。この方法の場合、学習データは例示のみに用いるため多くの数は必要なく、よって利用可能な学習データが少ない場合でも高い性能を得られる可能性がある。一方でモデルの学習によるカスタマイズ、いわゆるファインチューニングを行わないためドメイン知識の習得には限界があり、一般に広く用いられていない情報の抽出には必ずしも最適ではない可能性がある。

このため、少なくとも近い将来においては、両者が併存して使い分けられるような形で進んでいくものと思われる。

非定型帳票認識の異なるアプローチとしては、OCR を行った結果テキストを分析するのではなく、画像から直接的に求める答えを抽出する方式も考えられる。このような方式は VQA（Visual Question Answering）の一種であると考えられ、DL の内部で暗黙に OCR に相当する処理を行っているとは考えられるものの、陽に OCR を行わない点に特徴がある。非定型帳票認識はあくまで必要な情報をテキストで得ることがゴールであるため本質的には OCR を行う必要性はなく、このような方法は求められるタスクを素直に解こうとしただけ、ととらえることもできる。現在は勃興期にあるが、マルチモーダルなデータを扱う基盤モデル（Foundation Model）と呼ばれる技術が成熟していけば大きな発展を見せる可能性がある。

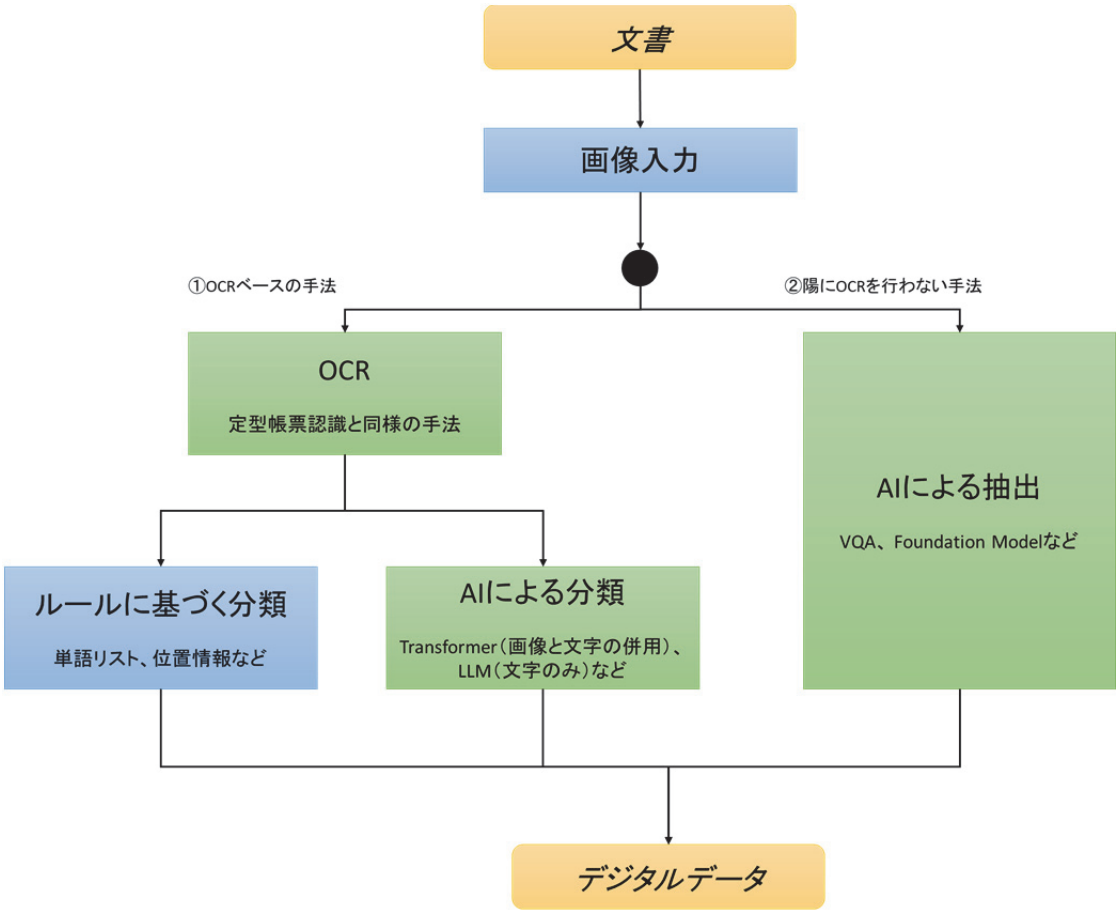


図 3.1.1-2　非定型帳票認識技術の概要

3.1.2　DL／人工知能技術を用いた OCR

DL に代表される AI 技術の進展により、一文字ごとの文字記入枠を設けないフリーエリアの認識率が大きく向上して実用水準となってきたことから、多くの会社が手書きフリーエリアの認識を含む帳票認識を主な対象としたサービスを一般に提供している。そのようなサービスの代表的な例として、AI-Inside の DX-Suite[2]や Cogent Labs の Tegaki[3]が知られている。これらの OCR はこれまで長年用いられてきた OCR と差別化するために AI-OCR と呼ばれており、DL 技術により従来と比べて非常に高い認識精度を達成したことが特徴である。例えば、Tegaki では活字文字認識製品と同程度以上と考えられる 99.22%の認識精度を達成したと謳っている。

認識精度は既にかなりの水準に達し、大きな改善は見られない状況になっている。このためソフトウェア／サービスとしての改良の焦点は利用しやすさとなってきている。例えば、PFU では、同社の文字認識ソフトである DynaEye に複数の文字認識エンジンによる認識結果を突合、認識結果が不一致の項目のみを目視で確認する方式を導入した[4]。これによりデータエントリ作業全体の作業量削減率がこれまでの 61%から 79%まで改善するとしている。ハンモックによる調査[5]にもある通り目視確認に要するコストは作業削減の上で大きな制約となっており、今後もこのような取り組みが進んでいくものと思われる。

3.1.3 多様なデザインの帳票を対象とした OCR

RPA などを活用した業務効率化ソリューションの中で、OCR が頻繁に活用されている。RPA+OCR は、インフラの維持整備コストの観点からすぐに電子システムへ完全移行することはできないが、こういった紙を用いた業務の負荷は少しでも軽減したいというニーズに応えるものである。

主要な適用事例として、他社から紙で送られてくる請求書の認識による支払処理の自動化などが挙げられる。このような用途では発行元ごとに書式は異なるが内容はほぼ同一の帳票、いわゆる非定型帳票が多く扱われている。非定型帳票認識の技術が進展していることから、これらを読取対象とした OCR が多く提供されるようになっている。これらの OCR もまた AI-OCR と呼ばれる。

代表的な例として LayerX によるバクラク請求書受取[6]がある。LayerX はブログ[7]にて関連技術に関する情報を積極的に発信しており、LayoutLM など最新技術の製品適用に向けた取り組みについても記事で公開している。このような積極的な情報発信は、情報公開・共有化が進んだ時代における取り組みの典型例として注目される。

また、2022 年末の ChatGPT の公開以来、これらを使った非定型帳票認識への取り組みが進んでいる。その代表的な例が GMO グループ研究開発本部のメンバーによる記事[8]である。本記事では Google Cloud の Document AI と ChatGPT を組み合わせ、請求書の画像データから必要なデータを構造化したうえで抽出する取り組みを行い、作成したソリューションを github で公開している。現在は「請求書認識」などのように OCR から情報抽出・構造化までがパッケージとして提供されているが、このような取り組みが広まっていくと、いわゆる非定型帳票認識のコア技術であった必要項目抽出処理が、サービスを利用する各社が自分たちにとって最適な形で実装するものとなっていくことも考えられる。一方でそのような作業がなく簡単に使えるサービスにも一定の需要はあると思われ、どちらが主流となっていくか、今後の動向が注目される。

3.1.4 AI-OCR からデータ利活用へ

旧来の OCR 製品の時代には OCR は「データ化」を中心とするサービスであったが、昨今の AI 技術の発展を受け、OCR によりデータ化するだけではなく、そのデータをいかに活用するかという点に注目がシフトしつつある。

代表的な例として AI Inside が提供する Heylix[9]がある。Heylix 自体は自律的 AI エージェントであり、OCR はこのエージェントが jpeg、tiff の一部などのファイルから情報を抽出するための手段として内部に埋め込まれている。提供される価値は OCR を含む様々な手段によって得られたデータから導かれる「支援」であり、OCR はあくまでもそれを実現するための手段にすぎない。

DataFluct でも AirLake[10]というプラットフォームで「非構造化データの構造化」と活用を謳っており、非構造化データを構造化する手段の一つとして OCR と連携している。これも新たなインサイトの理解や課題の解決に資することが提供される価値となっている。

学会レベルでも認識のタスクから情報抽出のタスクに注目が移っていっており、このような流れは今後加速していく可能性が高い。

3.1.5 AI-OCR にかかわる規制・管理動向

OCR に留まらず AI 技術の普及が進んでいく中で、品質や安全性に関する関心が高まりつつある。このため各社が AI に関するガバナンスを重視する傾向が強まっている。昨年度には大手各社が AI ガバナンスステートメントなどの形で公開していたが、今年度も継続的に取り組みが進んでいる。例えば、NEC では経済産業省のガイドラインに基づくガバナンス運用を開始している[11]。OCR では個人情報を含む画像・テキストを学習に用いる例もあり、特に安心・安全性という観点で引き続きガバナンスには注意を払うことが求められる。

また、品質についても議論が進んでいる。これに関わる代表的な事例として産総研が公開している機械学習品質マネジメントガイドライン[12]の存在が挙げられる。2023 年末にガイドライン本文第 4 版の公開にあわせて機械学習品質評価・向上技術に関する報告書も第 3 版を公開しており、品質保証についての啓蒙を積極的に進めている。QA4AI の活動も継続されており、2024 年 1 月にはこちらも AI プロダクト品質保証ガイドライン[13]が公開された。各社がこれらをもとにした独自の品質保証活動を進めており、今後もガイドラインの更新に伴って見直しが図られていくと思われる。

[参考文献]

[1] Gated Recurrent Convolution Neural Network for OCR
J. Wang, NIPS2017.

[2] DX Suite
https://inside.ai/dx-suite/

[3] Tegaki
https://www.tegaki.ai/

[4] 2種類のOCRエンジンの認識結果を突合し確認が必要な項目のみをピックアップ！確認作業を効率化
https://www.pfu.ricoh.com/dynaeye/product/dynaeye11/

[5] 【AI OCR 活用の実態調査】約 7 割が、AI OCR で読み取った文字の目視確認に 1 日 5 時間以上かかっていると回答
https://prtimes.jp/main/html/rd/p/000000224.000052725.html

[6] バクラク請求書受取
https://bakuraku.jp/invoice/
[7] LayerX Engineer Blog
https://tech.layerx.co.jp/
[8] 請求書 OCR 自動化： Document AI + ChatGPT API で非構造化データを JSON で出力させる
https://recruit.gmo.jp/engineer/jisedai/blog/automated-invoice-ocr-with-chatgpt-and-documentai/
[9] Heylix
https://service-heylix.inside.ai/
[10] AirLake
https://service.datafluct.com/airlake
[11] NEC、経済産業省の AI 原則実践のためのガバナンス・ガイドラインに基づく AI ガバナンスの運用を開始
https://jpn.nec.com/press/202304/20230403_02.html
[12] 機械学習品質マネジメントガイドライン 第 4 版
https://www.digiarc.aist.go.jp/publication/aiqm/guideline-rev4.html
[13] AI プロダクト品質保証ガイドライン 2024.01 版
https://github.com/qa4ai/Guidelines/blob/main/QA4AI_Guideline.202401.pdf

（URL 確認 2024 年 2 月 12 日）

3.2　文字認識・文書理解に関する国内学会の発表動向

文字認識・文書理解に関する研究報告の調査結果を報告する。調査対象は日本国内で開催された情報系の総合大会・学会・研究会であり、調査期間は 2023 年 1 月から 2023 年 12 月までである。

A. 文字認識

[A1：特徴抽出]

特徴抽出に関する発表が 2 件あった[10][20]。

文献[10]では、Deformable Convolution（DC）におけるカーネルの変位量を新たな特徴（変形特徴）とみなし、この特徴がフォント画像においてスタイル特有の変形を表現しているとの仮説のもとで、スタイル認識タスクにおいてその効果を検証した。変形特徴のみ、画素特徴のみ、変形特徴と画素特徴の併用、の 3 通りを比較した結果、変形特徴のみの場合と変形特徴と画素特徴を併用した場合の精度はほぼ同じであり、いずれも画素特徴のみの場合より向上していることが確認された。

文献[20]では、文字形状に対しスタイル特徴の演算可能性を考察した。実験では、スタイル特徴の演算を意識せずに学習した AutoEncoder ベースのモデルを使用してスタイル特徴を抽出し、文字の基本スタイル特徴（線の太さの変化やセリフの有無、傾きの有無など）が既に算術演算可能となっていることを確認した。一方で、基本スタイル特徴に準じない特徴（例えば、極めて装飾的なフォント）については演算可能性に限界があることも示した。

[A2：識別／学習]

識別／学習に関する発表が 1 件あった[4]。

文献[4]では、CNN を用いた低解像度ナンバープレート数字認識において、MCdropout による数字誤識別の抑制を検証した。実験では、MCdropout を用いることにより、正しい数字と識別する確率が増加することを確認した。また、Occlusion 解析及びフェイク画像の生成により識別器の特徴を視覚的に示した。

[A3：文字列認識]

文字列認識に関する発表が 12 件あった[2][3][5][7][9][12][13][16][19][22][23][24]。

文献[2]では、まず、文字列画像から Vision Transformer（ViT）によって幅、高さの二次元配置をもつ特徴量を算出した。特徴量を高さ方向に平均してから尤度推定して CTC を適用する手法と比べて、尤度推定してから高さ方向に足し合わせて CTC を適用する手法の方が高精度に文字列認識できることを示した。

文献[3]では、Transformer を用いた日本語情景文字認識において、自己回帰モデルの出力を学習済みの言語モデルの Masked Language Model の尤度で蒸留することによって、文字認識モデルに過去と未来の両方の文脈を考慮可能にし、計算コストを増加させることなく、文字誤り率を 10.22%から 9.83%に改善した。

文献[5]では、手書き文字データである近代公文書データセットと日本古典籍くずし字データセットの類似性を検証した。それぞれ片方のデータセットを用いて単文字認識モデルを学習させ、もう片方のデータセットにおける正解率を計測した。実験により両データセットには異なる特徴を持つ手書き文字が一定数存在すると結論付けた。

文献[7]では、横書きと縦書きの両方を単一のモデルで効率的かつ精緻に扱う日本語情景文字認識の手法を提案した。提案手法は書字方向に依存しないエンコーダと書字方向を考慮するデコーダから成る書字方向考慮型横書き縦書き共有モデリングと、情景認識におけるまとまりを考慮して縦書きの精度を向上する文字数同時推定で構成される。実験の結果、縦書き横書きいずれも文字認識の正解率が改善されることを確認した。

文献[9]では、手書き認識の半教師あり学習手法を提案している。まず、教師ありデータで Teacher Model を学習し、次に様々なデータ拡張を適用した教師無しデータを、Teacher Model の pseudo label を用いて Student Model を学習させる。手書き英語文字列認識の実験では、従来の半教師あり学習より高い認識率を達成できることを示した。

文献[12]では、入力の文字の系列が何らかの変換を受けた場合に所属クラスが変化し得る条件下における文字認識として、変換を回転に限定し、変換パラメータとそれが適用された時のクラスへの帰属度を用いた曖昧な認識表現と、これを用いた文脈依存型の認識法を提案した。円弧状に並んだ文字列認識と、ランダムに回転した連番数字の認識の実験の結果、提案手法が精度向上に有効であることを確認した。

文献[13]では、くずし字の文字認識において少数データ文字種の認識が困難であるという問題を解決するため、GAN をベースとした生成モデルにより学習データを生成することを提案した。生成画像を用いて学習することで認識精度の向上が見られ、特に生成画像がオリジナル画像に近い特徴を持つように促す VGG Loss を用いて学習することで高い精度を獲得できることを確認した。

文献[16]では、機械が抽出したプロトタイプを用いて、くずし字認識を行うアーキテクチャを提案した。プロトタイプは他のクラスにも共通して現れ、組み合わせることでクラスを特徴づける部位であり、エンコーダにより抽出した特徴量との類似度の計算に使用される。実験では、従来のエンコーダのみを用いた手法との比較及び、アブレーションによる提案アーキテクチャや損失関数の有効性を確認した。

文献[19]では、拡散モデルを用いた文字列認識手法を提案している。文字列認識タスクを、画

像からテキストの変換ではなく、テキストからテキストへの変換と捉え、テキスト生成のプロセスを画像条件付きの拡散過程とみなすことで実現した。実験では、提案手法が最先端の手法と比較して競合的な精度を達成することを示した。

文献[22]では、ViT を用いた文字列認識で標準的に用いられる自己回帰では、出力文字数分だけデコードをする必要があり、処理時間がかかるという問題を解決するため、一度にすべての文字をデコードする処理を低尤度の文字を中心に反復的に繰り返す Iterative ViTSTR を提案した。実験では、非公開のデータではあるものの、提案手法の有効性を確認している。

文献[23]では、既存の手書き数式認識モデル Bidirectionally Trained TRansformer（BTTR）の結果に対する、Regional Diffusion を用いた手書き数式認識結果の事後補正法を提案した。手書きと活字間、及び候補同士の局所特徴の類似度を用いる。実験の結果、大域特徴を使用した従来手法と比べて精度の向上が確認された。一方で BTTR の結果は本手法により一部は改善し、一部は改悪する事例が見られた。

文献[24]では、情景文字画像の超解像手法として、学習可能な正則化フレームワークを損失関数に組み込んだ手法を提案した。従来は固定値を使用していた正則化項の重みパラメータについても、提案手法では学習中に最適化される。提案手法により過学習が抑制され、低解像度の文字画像に対する認識精度が従来手法に比べ向上した。

[A6：その他]

文字認識に関する研究発表のうち、上記のいずれにも分類されないものが 6 件あった[6][14][15][17][18][21]。

文献[6]では、青空文庫の文書データを用いて近代言語モデルを構築する手法を検証している。具体的には、入力した文字列に続く文字を予測する分類モデルと任意の位置の文字を予測する穴埋めモデルを比較し、事前学習ありの分類モデルの予測精度が 0.701 であり最もよかったことを確認した。

文献[14]では、CLIP と微分可能レンダラーを組み合わせ、スタイル画像なしで文字画像にテキストにマッチしたスタイルを転送する手法を提案した。実験では自然言語を用いた既存のスタイル変換手法と比較を行った。結果、背景にスタイルが転送されることなく、よりテキストに沿った文字画像が生成されることを確認した。

文献[17]では、文字検出と物体検出の統合事前学習を用い、画像中に何の物体が写っているかを踏まえることで情景文字検出の精度向上手法を提案した。実験では、(1)事前学習を行わない、(2)物体検出データセットのみを用いた事前学習、(3)情景文字検出データセットと物体検出データセットを用いた事前学習、の 3 手法で比較を行い、文字検出と物体検出の統合事前学習が精度改善に貢献していることを確認した。

文献[18]では、指定した対象単語のみを消去する Selective Scene Text Removal（SSTR）と名付けた新たなタスクと、本タスクを実現するマルチモジュールモデルを提案した。マルチモジュールモデルは、背景検出モジュール、単語検出モジュール、選択的単語消去モジュール、及び再構成モジュールで構成される。実験では合成画像を用いて、マルチモジュールモデルと単一のモデル（Conditioned-U-Net、選択的単語消去モジュールでも使用しているモデル）とで比較を行い、提案手法が SSTR に有効なことを示した。

文献[21]では、(1)文字が敵対的攻撃にどの程度頑健か、(2)文字がどの程度頑健であるかを推定できるか、(3)文字をより頑健に加工することができるか、という三つの研究課題を提起し、実験と分析を行っている。実験により、(1)から、よりシンプルで標準的なフォントの文字ほど敵対的攻撃への耐性が高い傾向にあることを確認し、(2)から文字の形状と攻撃耐性の間に密接な関係があることを確認した。さらに(3)から、GAN ベースのモデルに分類器を加えることで、より攻撃耐性の高い文字画像を生成できることを確認した。

[A7：特別講演]

文字認識に関する特別講演が 1 件あった[29]。

文献[29]では、請求書における各項目の内容を推定する機能をルールベースのものから機械学習ベースに置き換えた事例について紹介している。従来のルールベースの手法で蓄積されたログからデータセットを作成し、言語モデル（RoBERTa）及びマルチモーダルモデル（LayoutLMv1-v3）の学習を行いモデルの選定を行ったことを述べている。さらに、リリース後の精度改善に向けたモニタリングやデータ基盤の改善等の工夫についても紹介した。

B. 文書理解

[B1：帳票処理・認識]

帳票処理・認識に関する発表が 4 件あった[1][25][26][28]。

文献[1]では、有価証券報告書の表を対象として、ChatGPT による PDF から JSON への自動変換手法について提案している。評価実験において、項目名とデータの一致率が 0.813 であり、項目名の親子関係の抽出正解率が 0.784 であることを確認した。

文献[25]では、重要単熟語着色機能、項目別情報抽出機能、要旨ドラフト作成機能を備わった入札説明書読込や情報整理の効率化手法 BiDding Document Analyzer（BDAA）を提案している。検証実験では、BDAA を利用しない場合に比べ、熟練者は 40%、初心者は 72%タスク達成時間が短縮できることを確認した。

文献[26]では、文書画像から 3 項関係を抽出する OpenIE タスクにおいて、OpenIE タスク専用の学習データを用いることなく、T5-base を用いた質問生成及び質問応答、及び構文解析によ

る 3 項関係の抽出を利用する手法を提案している。VQA のデータセットを用いた実験の結果、一部のデータセットでは提案手法が従来手法よりも良い精度を示した一方で、ドメインによってはOCR の誤認識や文法ミス、ハルシネーションによる精度低下が確認された。

文献[28]では、政治資金収支報告書に OCR を施し、認識結果を手作業で修正したのち項目情報と数値情報を紐付けて格納することで政治資金データベースを作成した事例について述べている。さらに、これを用いた分析手法の例として、収入に占める個人・法人・他政治団体からの寄付の割合を表す三角グラフや政治団体ごとの収入のヒストグラムを取り上げて有効性を示した。

[B4：前処理・他]

文書理解における前処理・他に関する発表が 1 件あった[11]。

文献[11]では、潜在拡散モデルにより文書画像の歪み補正と再照明を行う手法を提案している。Fine-tuning により、劣化画像を条件とした歪み補正と再照明を行った文書画像生成に特化させたモデルを使用している。従来手法に比べ、強い影を含んだ文書画像に対しても歪みや影のない正確な画像の生成が可能になり、潜在拡散モデルの有効性を確認している。

C. その他

[C0：その他]

以上のどれにも分類されない研究発表が 2 件あった[8][27]。

文献[8]では、GoogLeNet を用いて漢字筆跡 548 字種の学習を行い、異なる文字種での筆者識別を提案している。検証実験では、401 名が 5 回繰り返して筆記した筆跡標本を用いた場合の識別精度が約 7 割であり、学習に用いていない文字種でも約 6 割であることを確認した。さらに、複数文字での平均値で筆者の識別を試みたところ、99%を超える高い識別精度が得られた。

文献[27]では、画像認識に使用される事前学習済みモデルのうち、計算負荷が最も軽いものの 1 つである Alex net を転移学習し、筆跡の画像データの筆者推定の精度を検証している。実験では、「護」の字の筆者識別モデルを作成し、77.81%の正答率を得ることができた。

文献番号	掲載媒体	巻号	著者	題目
1	FIT	E-032	佐藤 栄作・木村 泰知（小樽商科大学）	TOPIX100 の有価証券報告書の表を対象とした ChatGPT による PDF から JSON への自動変換の試み
2	FIT	CF-005	Buoy Rina・Iwamura Masakazu（Osaka Metropolitan University）・Srun Sovila（Royal University of Phnom Penh）・Kise Koichi（Osaka Metropolitan University）	ViT-CTC: Vision Transformers with CTC for Scene Text Recognition
3	FIT	CH-005	折橋 翔太・山﨑 善啓・内田 美尋・高島 瑛彦・東羅 翔太郎・増村 亮（日本電信電話）	双方向言語モデルからの知識蒸留を用いた日本語情景文字認識
4	FIT	I-015	松岡 剛史・藤田 和弘（龍谷大学）・四宮 康治（兵庫県警察本部科学捜査研究所）	CNN を用いた低解像度ナンバープレート数字の識別
5	FIT	N-013	宮川 裕貴・山田 雅之・中 貴俊・兼松 篤子・宮崎 慎也・長谷川 純一（中京大学）	近代公文書データセットと日本古典籍くずし字データセットを用いた手書き文字認識に関する比較
6	FIT	N-014	亀山 京右・山田 雅之・中 貴俊・兼松 篤子・宮崎 慎也（中京大学）・長谷川 純一（中京大学人工知能高等研究所）	近代公文書 OCR に向けた近代言語モデルの構築
7	信学論 D	J106-D No.12	折橋翔太・山﨑善啓・内田美尋・高島瑛彦・増村亮	文字数推定を用いた書字方向考慮型横書き縦書き共有モデリングによる日本語情景文字認識
8	PRMU	vol. 122, no. 404	赤尾佳則（科警研）	複数文字種の筆跡を学習した深層ネットワークによる筆者識別の試み ～ 401 名の筆跡標本を用いた検証 ～
9	PRMU	vol. 122, no. 404	Masayuki Honda・Hung Tuan Nguyen・Cuong Tuan Nguyen（TUAT）・Cong Kha Nguyen・Ryosuke Odate・Takashi Kanemaru（Hitachi Ltd.）・Masaki Nakagawa（TUAT）	A Semi-Supervised Learning Framework for Handwritten Text Recognition using Mixed Augmentations and Scheduled Pseudo-Label Loss
10	PRMU	vol. 122, no. 404	北島和樹・内田誠一（九大）	Deformable Convolution による局所変形特徴抽出の試み
11	PRMU	vol. 123, no. 266	今林颯大・ハオ グオチン・飯塚里志・福井和広（筑波大）	潜在拡散モデルを用いた文書画像の歪み補正と再照明
12	PRMU	vol. 123, no. 266	木本 舟・菅間幸司・和田俊和（和歌山大）	曖昧な認識表現を用いた文脈依存型文字列認識
13	MIRU	IS1-66	阿部楓也, 岩井翔真, 宮崎智, 大町真一郎（東北大）	生成画像を利用した少数データくずし字認識に関する検討
14	MIRU	IS1-67	泉幸太, 柳井啓司（電通大）	CLIP と微分可能レンダラーを用いたフォントスタイル変換
15	MIRU	IS1-90	Jan Zdenek, Wataru Shimoda, Kota Yamaguchi (CyberAgent)	Can We Erase Japanese Text from Images Without Japanese Training Data?
16	MIRU	IS1-91	木下純哉, 宮崎智, 大町真一郎 (東北大)	パーツプロトタイプを用いたくずし字認識に関する検討
17	MIRU	IS1-92	折橋翔太, 山﨑善啓, 内田美尋, 高島瑛彦, 東羅翔太郎, 増村亮 (NTT)	文字検出と物体検出の統合事前学習を用いた情景文字検出
18	MIRU	IS2-87	三谷勇人（九大), 木村昭悟 (NTT), 内田誠一（九大)	情景内単語の選択的消去
19	MIRU	IS2-88	藤武将人（ファーストアカウンティング, FA Research)	拡散モデルを用いたシーンテキスト認識

文献番号	掲載媒体	巻号	著者	題目
20	MIRU	IS2-89	近藤徹多, 原口大地, 内田誠一 (九大)	スタイル特徴は演算可能か?
21	MIRU	IS3-89	片岡蓮太郎 (九大), 木村昭悟 (NTT), 内田誠一 (九大)	敵対的攻撃に頑健な文字のデザインをめざして
22	MIRU	IS3-90	竹長慎太朗 (筑波大), 内田奏 (Sansan)	確信度に基づいた自己修正機構を持つ高速な文字認識モデル
23	MIRU	IS3-91	ピョンケイジ (九大), Xiaomeng Wu (NTT), 内田誠一 (九大)	Regional Diffusion による手書き数式認識結果の事後補正
24	JSAI	3U5-IS-4-05	Supatta VIRIYAVISUTHISAKUL(Japan Advanced Institute of Science and Technology), Parinya SANGUANSAT(Panyapiwat Institute of Management), Teeradaj RACHARAK(Japan Advanced Institute of Science and Technology), Minh Le NGUYEN(Japan Advanced Institute of Science and Technology), Toshihiko YAMASAKI(The University of Tokyo)	Text Recognition in Low Resolution Images Using Trainable Regularization
25	JSAI	3Xin4-17	伊藤 友貴(三井物産株式会社), 中川 駿(三井物産株式会社, 三井物産スチール株式会社)	入札書説明書からの項目別情報抽出及びその業務効率化への活用
26	JSAI	3Xin4-19	山口 篤季(株式会社日立製作所), 十河 泰弘(株式会社日立製作所)	VQA データセットを活用した文書画像からのオープン情報抽出の検討
27	JSAI	3Xin4-73	菅原 滋(科学警察研究所)	Alex Net の転移学習による筆跡の筆者識別の可能性の検証
28	JSAI	2H1-OS-3a-01	山田 健太(日本経済新聞社), 青田 雅輝(日本経済新聞社), 並木 亮(日本経済新聞社), 横山 源太朗	政治資金収支報告書の OCR による政治資金データベースへの試み
29	JSAI	掲載無し	島越 直人(株式会社 LayerX), 松村 優也(株式会社 LayerX)	株式会社 LayerX「LayerX における機械学習を活用した OCR 機能の改善に関する取り組み」

3.3　文書画像認識に関する国際会議の発表動向

3.3.1　文書画像認識に関する主な国際会議

文書画像認識の研究分野は、IAPR（International Association of Pattern Recognition：国際パターン認識連盟）が国際的なコミュニティの中心となっている。その中でも、TC 10（Technical Committee Number 10：第 10 技術委員会：“Graphics Recognition”）[1]と、TC 11（Technical Committee Number 11：第 11 技術委員会：“Reading Systems”）[2]が特に文書画像認識に関係する。IAPR TC10/11 が主催する主な国際会議には ICDAR、DAS、ICFHR があり、これらが文書画像認識の国際的な研究動向を把握するために特に重要である。それぞれの国際会議の特徴を下記に示す。

●ICDAR（International Conference on Document Analysis and Recognition）

文書認識・解析の全般に関する最大の国際会議である。1991 年から隔年開催で、2023 年に第 17 回がアメリカ・サンノゼで開催された（オンラインと現地のハイブリッド開催）[3]。近年の参加者は 500 名前後である。ICDAR は 2024 年度以降毎年開催されることとなっており、次回第 18 回は 2024 年（ギリシア・アテネ）開催となる。

●DAS（International Workshop on Document Analysis Systems）

1994 年から隔年開催で単独開催されてきたが、2024 年度より ICDAR のサテライトワークショップとして開催されることとなった。

●ICFHR（International Conference on Frontiers in Handwriting Recognition）

手書き文字認識に関する国際会議である。1990 年に Workshop として（名称は IWFHR）第一回が開催され、ほぼ隔年で実施されている。規模の拡大に伴い、2008 年の第 11 回から名称が Conference となり、ICFHR に改称。2022 年 12 月にインドのハイデラバードで開催された（ハイブリッド開催）[7]。参加者は 150～200 名程度。

3.3.2　国際会議 ICDAR2023

表 3.3.2-1　ICDAR2023 の開催概要

日程	2023 年 8 月 21 日～8 月 26 日
Main Conference	8 月 21 日～8 月 23 日
Post Conference	8 月 24 日～8 月 26 日
開催地	アメリカ・サンノゼ
会場	ハイブリッド開催（現地会場は下記） Main Conf. ：San José Marriott Post Conf. ：Adobe World Headquarters
主催	IAPR（International Association of Pattern Recognition）
公式サイト	https://icdar2023.org/

表 3.3.2-2　前回開催との比較

	ICDAR2021	ICDAR2023
開催期間	6 日間	6 日間
開催時間／日	約 8 時間半（9:00～17:30）	約 8 時間半（9:00～17:30）
Tutorial	2 件（約 6 時間）	4 件（約 6 時間）
招待講演	4 件	3 件
ポスター	2 枠（1 時間半×2）	2 枠（1 時間半×2）
発表セッション	10 枠	14 枠
投稿数	340 （採択 182：oral 40, poster 142）	316 （採択 154：oral 53, poster 101）

ICDAR は文字認識の分野では長年トップカンファレンスの地位にある学会である。基本的に文書の認識・理解に特化した内容を扱う学会であり、一般的な AI 技術を扱う学会とはやや目的を異にするが、文書画像認識に関わる技術に取り組む上では極めて重要度の高い学会である。

ICDAR 2023 の論文投稿数はトータル 316 本、うち採択は 154 本（採択率 48.7%）であり、例年に比べてやや少数、低採択率となった。表 3.3.2-2 に前回（ICDAR 2021）と比較したデータを示す。開催規模は前回と大差はなかったが投稿論文数は漸減している。2021 年度は 2019 年度から比べて 100 件以上の減少となっており、減少幅は少なくなってはいる。減少の理由としては、OCR／文書理解の分野でも他分野との技術的統合が進んでいることから、AI 一般を扱う学会への投稿を優先する研究者が増えていることの影響が考えられる。

ICDAR 2023 では前半 3 日間で Main Conference、後半 3 日間で個別テーマに分かれた Workshop、Tutorial を扱う Post Conference が開催された。

◆Main Conference

Main Conference では、招待講演（Keynote Speech）が 3 件と、口頭発表とポスター発表が行われた。口頭発表はテーマ別に分かれた 14 のセッション（コンペティション結果のセッションを含む）でそれぞれ数件ずつの発表が行われた。ポスター発表は 1 時間半のセッションが 2 回行われた。また、これ以外に企業の講演とパネルディスカッションを行う Industry Panel セッションも開催された。

[Keynote Speech]

- A First Look at LLMs Applied to Scientific Documents
  Marti Hearst, UC Berkeley
- Enabling the Document Experiences of the Future

Vlad Morariu, Adobe Research

- What Are Letters?

  Seiichi Uchida, Kyushu University

[Oral Sessions]

- Oral Session 1: Graphics 1: Graphics Recognition
- Oral Session 2: D-NLP 1: Document NLP
- Oral Session 3: Graphics 2: Tables & Charts
- Oral Session 4: D-NLP 2: Information Extraction
- Oral Session 5: Applications 1: Medical, Legal, and Financial
- Oral Session 6: Handwriting 1: Online Documents
- Oral Session 7: DAR 1: Document Layout Analysis
- Oral Session 8: Handwriting 2: Historical Documents
- Oral Session 9: DAR 2: Camera Image and Schene Text
- Oral Session 10: Handwriting 3: Document Synthesis
- Oral Session 11: Competitions
- Oral Session 12: Graphics 3: Math Recognition
- Oral Session 13: DAR 3: Text and Document Recognition
- Oral Session 14: Applications 2: Document Analysis Systems

◆Post Conference

Post Conference では、Tutorial セッションが 4 テーマ、Workshop が 8 テーマ開催された。Tutorial は 24 日と 25 日に順次、Workshop は 24 日と 25 日は 2 テーマ並列、26 日は 1 テーマ単独で開催された。

[Tutorial]

各テーマの専門家による、約 3 時間の講義（とデモ）。今回は以下の内容が開催された。

- Computational Analysis of Historical Documents

  Isabelle Marthot-Santaniello(,University of Bale) ほか

- Deep Learning

  Thomas Breuel (Nvidia)

- Document Image Binarization

  Rafael Dueire Lins(Universidade Federal Rural de Pernambuco) ほか

- Unlocking the Potential of Unstructured Data in Business Documents Through Document Intelligence

Anand Mishra (IIT Jodhpur) ほか

[Workshop]

Workshop は ICDAR の本会議とは別に個別テーマで開催する小さな会議である。一般に本会議よりも査読の敷居は低いが、特定のテーマに特に関心がある参加者が多く、深い議論が期待される。今回は以下の内容が開催された。

- ICDAR 2023 Workshop on Scaling-up Document Image Understanding
- ICDAR 2023 Workshop on Computational Paleography (2nd edition) (IWCP) / 0.5 枠
- ICDAR 2023 Workshop on Camera-Based Document Analysis and Recognition (CBDAR 2023) / 0.5 枠
- ICDAR 2023 International Workshop on Graphics Recognition (15th edition) (GREC 2023) / 2 枠
- ICDAR 2023 Workshop on Automatically Domain-Adapted and Personalized Document Analysis (ADAPDA)
- ICDAR 2023 Workshop on Machine Vision and NLP for Document Analysis (1st edition) (VINALDO)
- ICDAR 2023 International Workshop on Historical Document Imaging and Processing (7th edition) (HIP'23) / 2 枠
- ICDAR 2023 International Workshop on Machine Learning (4th edition) (WML) / 2 枠

◆参加報告

投稿論文の著者は例年通り中国が多く、フランス・アメリカがその半分程度であった。今回の特徴としてはインドがフランスとほぼ同数ながら 2 位となった点である。前回も 5 位と高順位ではあったが、同国で OCR 関連が盛り上がりを見せていることがうかがえる。日本の順位は前回 6 位、今回 5 位と大きく変わっていない。

今回の参加者の傾向として、中国からの現地参加が少なかったことが挙げられる。例年、同国からの参加者は現地も含め多数いるが、今回は発表者も含めて多くがリモート参加にとどまっている。このことは、現地参加が必須条件となる賞レースにも影響した可能性がある。また、例年に比して大学からの参加比率が例年より高いようであったが、このことも例年投稿・参加の多い中国企業関係者が少なかったことによるように思われる。日本からの参加はさほど多くなく、特に企業からの参加はごく少数であった。しかし、キーノートで九州大内田先生が講演されるなど、一定のプレゼンスは維持している。

投稿論文の傾向として、マルチリンガルな認識、言語と画像のマルチモーダルを利用した文書理解などが多く投稿されていた。後者は多様なデザインで作成される文書データ、いわゆる非定

型な帳票やホワイトボードなどへの記入から情報を抽出、理解するための枠組みであり、日本でも特に非定型帳票認識の分野で活用が進んでいる。今後進展を見せるであろう、より言語処理的なアプローチも含めて、継続的な注視が必要である。なお、今年度初頭以来急速な盛り上がりを見せているにもかかわらず、言語処理（NLP : Natural Language Processing）関連の投稿数は前回と比べて横這いであった。Abstract Submission が 2 月初頭と ChatGPT の急激な盛り上がりと似たような時期であり、ChatGPT と連携したような研究が投稿に間に合わなかった可能性が高い。このため来年以降関連投稿が急増する可能性がある。このような事情を考慮してか、投稿数横這いにもかかわらずオーラルセッションが 2 枠となり、ワークショップも開催されるなど学会としても注目している分野と考えられる。

論文の投稿先は Handwritten Document Images が 1 位であった。学会の本流でもあるため例年上位にあり、ベースとして活発な投稿が維持されていることがうかがえる。今年の特徴としては、Typeset Document Images、Born-Digital といった分野への投稿数が増えていることが挙げられた。特に Born-Digital はテキストデータ埋め込みなどを含む分野で、いわゆる文字認識を行わない事例も多い。OCR と非常に関連が深い文書解析分野への興味が高まっていることを示していると思われる。

Best paper はジョンズホプキンス大学（アメリカ）による“A hybrid model for multilingual OCR”である。Transformer でエンコードし、CTC と Transformer Decoder 両方でデコード（出力統合はなし。Transoformer の方が高精度とのこと）、言語や縦書き／横書きなどを推定するサブブランチも持つというリッチな構成である。複数言語の 9000 文字程度を一括学習していることも特徴とするが、東アジア、特に漢字を常用する日本語・中国語では 9000 字種は「たかだか」というレベルでもある。文中に出てくる異言語（多くは英語）の扱いなどに優れる可能性はあるが、より詳細な評価が待たれる。

【参考文献】

[1] IAPR TC10 Homepage https://iapr-tc10.univ-lr.fr

[2] IAPR TC11 Homepage http://www.iapr-tc11.org

[3] ICDAR 2023 Homepage https://icdar2023.org

[4] [IAPR-TC10] Newsletter 147 – September 2021 https://iapr-tc10.univ-lr.fr/?p=1366

3.4　パターン認識研究の最新動向（特別講演報告）

本節では、今年度に当委員会で行われた特別講演の概略を報告する。

3.4.1　NEC の文字列認識、人物行動検知、偽薬対策画像処理の R&D 活動の紹介

NEC ではこれまで、画像認識技術のビジネス化を進めてきた。本稿では、住所文字列を認識する技術や、カメラに写る人物の行動を検知する技術、医薬品パッケージの画像を用いて偽薬対策を行うシステムの研究について述べる。

3.4.1.1　文字列認識の研究

住所文字列を読み取るためには、図 3.4.1-1 に示すように、宛名の書かれた行を抽出し、そこから文字候補を抽出し、それらの候補に文字認識を施した後、宛名住所として読み取れる結果を抽出する方法が採用されてきた。

宛名行抽出では、日本語の住所表記として可能性のある縦書きと横書きのどちらかを判断して、宛名行を抽出し後段の処理に渡す。文字候補抽出では、日本語の分かち書きの問題や、漢字の偏と旁もそれぞれで個別の漢字として読めてしまうという問題に対応するため、網羅的に文字候補を抽出する。文字認識では、すべての文字候補について文字認識を行い、最終的に住所読取りモジュールでは住所として読み取れて、かつ文字候補も過不足なく利用している組み合わせを住所読み取り結果として出力する。

特に本稿では①文字候補抽出と、②文字列読取り駆動型の行抽出と、③漢数字読み取りについて説明する。

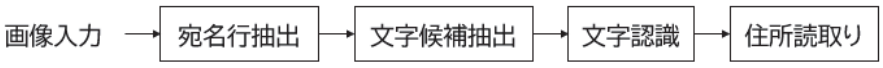


図 3.4.1-1　住所読取りの基本的な処理フロー

まず、①の文字候補抽出[1]では、図 3.4.1-2 に示すように文字候補を網羅的に抽出する。この際、黒画素の連結成分を最小単位として、連結成分の複数の組み合わせを文字候補とする。これは、図 3.4.1-2 の入力画像を「ノノノ山奇市」ではなく「ノリ山奇市」でもなく「川崎市」として認識できる組み合わせを文字候補として抽出するために重要である。

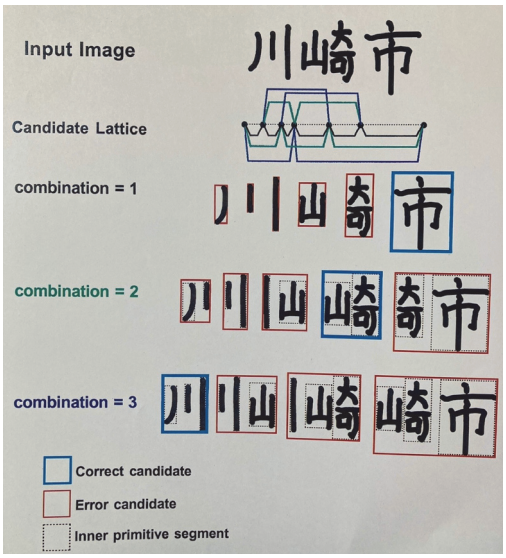


図 3.4.1-2　文字候補の網羅的な抽出

一般に、連結成分の組み合わせを増やせば、候補内には正しい文字の組み合わせが含まれる可能性が高まるが、あまりに候補が大きすぎると、逆に誤った住所として偶然読み取れてしまう可能性も高まる。図 3.4.1-3 の上のグラフは、横軸が文字多候補率（実際に切り出すべき文字数の何倍の候補を抽出したか）で、縦軸が正解候補含有率である。下のグラフは横軸が文字多候補率で、縦軸が住所読み取り正解率である。この図から、いたずらに正解候補含有率を上げても、住所読取り正解率が低下してしまうことがわかる。システムとしてのバランスは、住所読取り正解率の最大化が重要である。

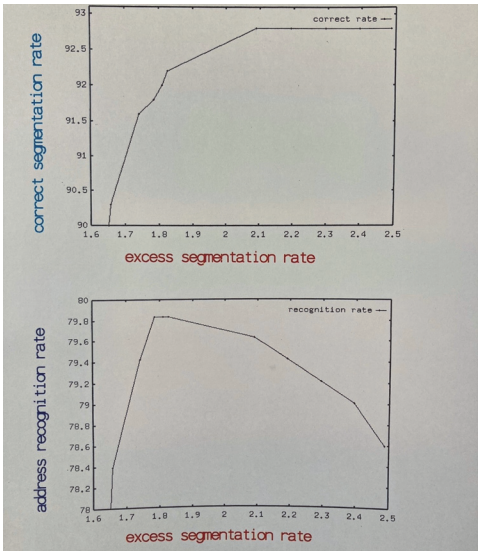


図 3.4.1-3　文字多候補率と正解含有率、住所読み取り率の関係

次に②文字列読取り駆動型の行抽出[2]について述べる。文字認識の前に画像処理だけで文字行

の抽出をしようとするタスクは、図 3.4.1-4 の例で言うと左側の矩形のように黒画素の連結成分が抽出されたときに、右側のように 1 行目から 4 行目までを記載方向も含めて抽出するタスクとなる。

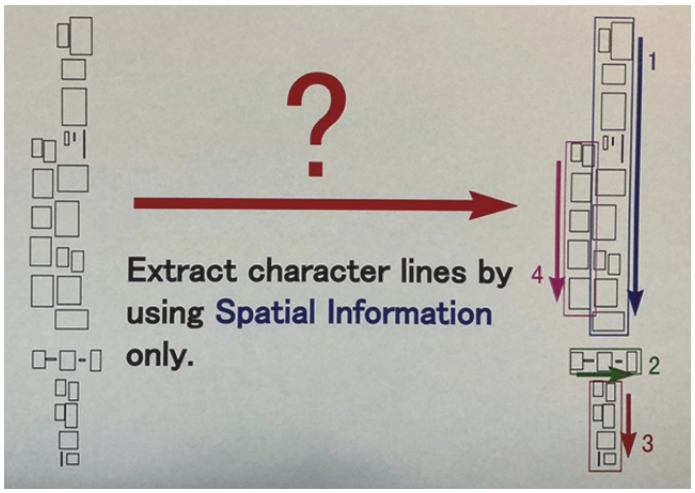


図 3.4.1-4　画像処理だけで文字行の抽出を行う困難さ

それゆえ行の抽出も、文字候補抽出と同様に、住所として正しく読める組み合わせを抽出するタスクとした。つまり、文字候補を二次元的に網羅的に抽出し、文字認識を施した後に住所として読み取れる組み合わせを抽出する処理まで行った上で、最終的に二次元的な文字配置と文字認識結果の確からしさを総合的に評価して最も良い評価値を与える組み合わせをもって行抽出の結果とする（同時に住所読み取り結果とする）方法を開発した（図 3.4.1-5）。

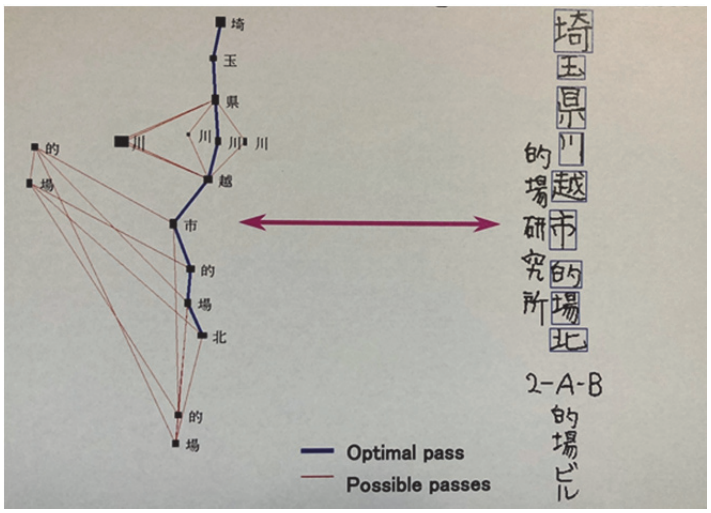


図 3.4.1-5　文字配置と文字認識信頼度と住所知識を統合評価する宛名行抽出

文字の二次元的な配置情報を用いると、③漢数字の読み取り[3]にも効果がある。図 3.4.1-6 だ

けを見ると、漢数字で「一」「二」「三」と読み取ることができる。

図 3.4.1-6　漢数字の例

では、図 3.4.1-7 の例ではどうであろう。左側の例は「一ノ四ノ三」で良いかもしれないが、右側だと、「二ノ三ノ二三」であろうか、それとも「二ノ一二ノ二三」であろうか。

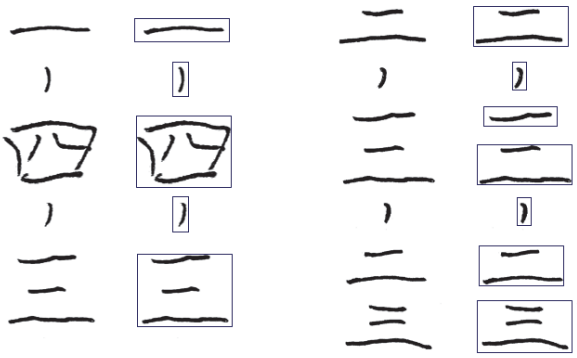


図 3.4.1-7　前後関係で漢数字の認識結果が変わる可能性のある例

左の街区住所の画像例における号を表す「三」と、右の画像例における番地を表す「一二」は、それぞれまったく同じ画像であるが、前後の関係が変わると、必ずしも認識結果が同じではなくなる場合がある。これを文字のピッチやサイズを含めた二次元的な配置を評価することで、漢数字の読み取りの確からしさを高めることができた。

これらの技術のエッセンスが OCR 製品にも活用されてきた。

3.4.1.2　人物行動検知の研究（主に監視カメラシステム向け）

本節では、監視カメラシステムのセキュリティ以外の応用について述べる。監視カメラシステムは、防犯目的として必要性は認められながらも、監視されているという意識から広く受け入れられるにはハードルがあった。しかし、防災対策や見守りのためにシステムを導入する場合は比較的受け入れやすかったと思われる。本稿では特に①防災対策のための混雑状況検知システムと、②安全操業支援のための危険エリア侵入検知システムについて述べる。

まず、①の混雑状況検知システム[4]では、様々な混雑状況に応じた学習データを事前に収集することが困難であったため、人工的に混雑状況の学習データを生成し、機械学習を行うことで、計 18 カ所の混雑状況を瞬時に数値化するシステムを世界に先駆け実現した（図 3.4.1-8）。

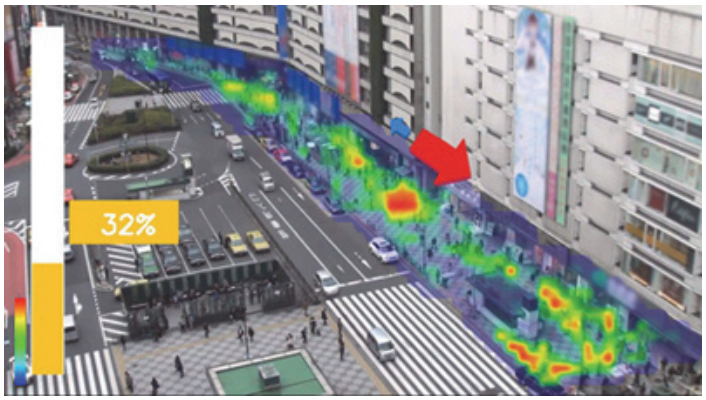


図 3.4.1-8　防災対策としての混雑状況検知システムの例

また、②危険エリア侵入検知システム[5]については、危険作業現場の例のように特殊な外乱（照明条件変化、多様な装置配置、様々な人物の姿勢、操業状態等）が多い環境下でも、予め全ての外乱の組み合わせを学習データとして収集して機械学習を行うことが困難なため、それぞれ独立して実際のデータを収集した後に、想定されるすべてのケースについて人工の学習データを網羅的に生成し、これらを学習させることで多様な外乱に対しても網羅的かつロバストに人の行動を検知できるシステムを世界に先駆けて実現した（図 3.4.1-9）。

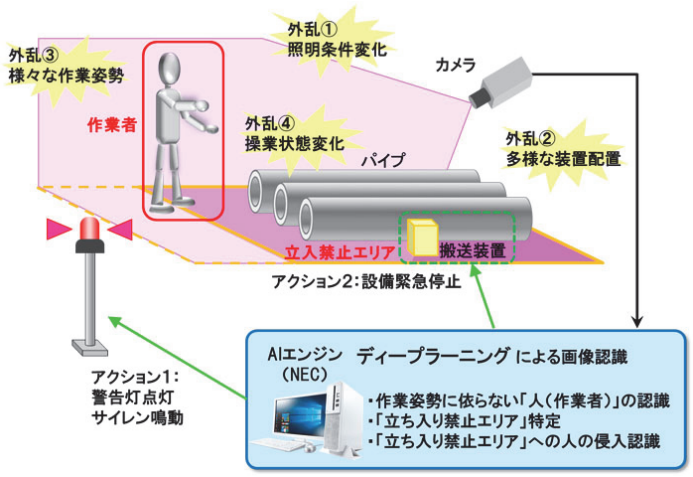


図 3.4.1-9　安全操業支援のための危険エリア侵入検知システムの例

3.4.1.3　偽薬対策システムの研究

画像認識を用いて商品の識別を行うアプリケーションも考えられる。本節では、画像による商品認識技術を応用して医薬品の真贋判定を支援するシステムについて述べる。特にアフリカでは医薬品の偽物が多く出回っており、一部地域では 3 割以上の医薬品が偽物であるとの報告もあり社会問題化している。これらの偽物の中には明らかな粗悪品も多く、画像認識技術を活用して医薬品購入者に注意喚起するシステムが考えられる。

技術課題としては、バリエーションが大量にある商品の一つ一つに対して画像をきちんとアノテーションして認識用の画像 DB として構築できるかどうかと、画像 DB を継続的にアップデートできるのかという二点がある。この課題に対処するために、ユーザーが市場でシステムを利用中に未知の画像が見つかれば、ユーザーに画像アップロードとアノテーションを依頼、その後専門家が確認し、正式に DB 登録するという情報収集とアノテーションのループをシステム内に設けた。このようにデータ収集して構築した DB を Market Intelligence 型 DB と呼ぶ（図 3.4.1-10）。

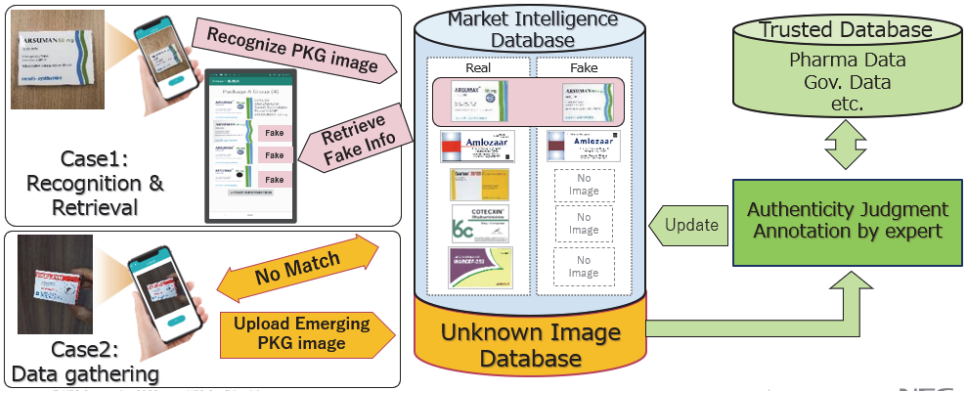


図 3.4.1-10　Market Intelligence 型の商品画像 DB 構築

この考えをさらに進め、手元に何も情報が無くても目の前の医薬品の真贋を確かめるためのヒントを提示できる枠組みも検討した。処理フローとしては、未知のパッケージ画像が入力された場合、まず汎用 OCR で商品名を読み取り、それをキーに Web をクローリングして画像を検索し、それらの画像と入力画像の比較結果をユーザーに提示するというシステムである。これを Web Intelligence と呼ぶ（図 3.4.1-11）。

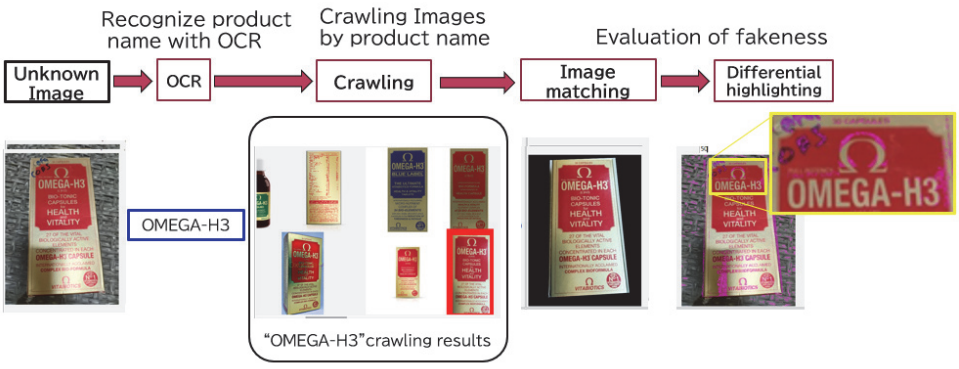


図 3.4.1-11　Web Intelligence 型の商品画像確認支援システム

本節では、これまで NEC が進めてきた画像認識技術のビジネス化について述べた。今後も画像処理技術によって社会課題を解決できるよう、研究開発を進める予定である。

[参考文献]

[1] Eiki Ishidera, Daisuke Nishiwaki, Keiji Yamada：A Candidate Reduction Method for Japanese Character Segmentation, ACCV'95

[2] Eiki Ishidera, Daisuke Nishiwaki, Keiji Yamada：Unconstrained Japanese Address Recognition Using a Combination of Spatial Information and Word Knowledge, ICDAR'97

[3] 石寺永記, 西脇大輔, 山田敬嗣：文字配置情報を用いた街区住所読み取りの一検討、1997 年信学会総合大会, D-12-96

[4] 石寺永記,鈴木哲明,宮野博義,樫原猛：豊島区総合防災システムにおける群衆行動解析, 情報処理学会 デジタルプラクティス Vol.8 No.2(Apr. 2017)

[5] 国内業界初となる AI 画像認識による安全行動サポート技術の導入について https://www.nec-solutioninnovators.co.jp/press/20181211/index.html

3.4.2 NEC の大規模言語モデル（LLM）に関する R&D 活動の紹介

NEC では、Generative AI（生成 AI）の一つとして、130 億パラメータで世界トップクラスの日本語性能を有する軽量な大規模言語モデル（LLM）を開発し、メディアでも取り上げていただいた。

本稿では、①生成 AI、②LLM（Transformer Encoder）の仕組み、③NEC の LLM、④NEC の生成 AI 事業（主に LLM)、⑤LLM を活用した他分野研究について述べる。

3.4.2.1 生成 AI について

生成 AI とは、「様々な指示」を入力すると、その指示に従って「様々なコンテンツ」に変換する AI のことである。生成 AI の入力も出力もテキストや画像、音声である。テキストを入力して、テキストを出力するような生成 AI が LLM であり、ChatGPT や LLaMA 等が知られている。LLM は、質問を入力すると答えを教えてくれたり、日本語で指示すると英語のメールを書いてくれたり、やりたいことやプログラムの要件等を日本語で伝えるとコードを書いてくれたりする。

なぜ“今”生成 AI なのかと言うと、端的に言えば生成 AI が実用レベルに達したからであろう。それを支えた技術要素は、①AI が勝手に賢くなる仕組み、②多量のデータ、③強力な計算機であり、これらが揃ったことにより、高品質なコンテンツを出力する AI が次々に登場し、実用の域に達したと思われる。

まず①では、これまで教師有り学習または弱教師有り学習をする場合はデータに対するラベリング作業（人間が介在）が必要で AI の性能向上の度合いが頭打ちになっていたが、生のデータから教師情報を機械が自動で見つけてくれる自己教師有り学習が用いられるようになった。次に②として DX 時代の昨今はデータも大量に入手可能である。そして、③これらを十分に活用できる

強力な計算機である GPU も進化しているということで、これら三つがかけ合わさって性能が高まったと言える。

NEC でも、生成 AI の研究開発とビジネス化を進めており、第一弾として LLM にフォーカスした。画像と LLM の組み合わせなどマルチモーダルも考えられるが、実用面を考えるとテキストをテキストに変換するだけでも実に色々な業務の支援が可能になる。

3.4.2.2 Large Language Model（LLM）の仕組みについて

これまで知られている古典的な Language Model として n-gram 言語モデルがある。モデルを作るにはマルコフ連鎖を使うのがポピュラーで、ある文章の次に出現する単語の確率値をモデル化する方法が知られている。例えば、「私の仕事」という文章の次には「は」が出現しやすく（P（は｜私の仕事）＝0.6)、「ぬ」は出現しにくい（P（ぬ｜私の仕事）＝0.01）という確率を計算し行列として記録したりする。これがほとんどそのまま Large Language Model のベースになっている。

LLM のアーキテクチャとしては GPT や T5 が知られている。古典的な方法と比べて LLM では実現の仕方は変わっているが、やりたいことは同じである。ある文章が与えられたときに次の文字を予測したり、全体の文章の確率推定をしたりするために、超大量のデータを用い、n-gram でなくニューラルネットでモデリングし、それを損失関数の自己教師有り学習のタスクとして解く。タスクは今主流の GPT では次の単語を予測するタスクになっている。ニューラルネットとしては RNN や LSTM、Transformer や Transformer-Decoder がある。Transformer は並列化が容易で表現力が高く、これと自己教師有り学習を組み合わせることで初めて今の GPT ができた。現在では、ネットワークが数十階層ある深いモデルが使われている。

ここで、ある文章が与えられたときに次の単語が出現する確率を予測することができるモデルがあると、なぜ色々なことができるのかについて考える。LLM は深層学習の階層が深いモデルになっていて表現力が非常に高いため、もともとの生のテキストに含まれている暗黙的な言語の構造やタスク構造が次の単語の予測を繰り返し学習するだけで暗黙的に学習され、人間がモデリングしなくても学習されてしまうと考えられる。

これを直感的に説明するために、ソースコードを出力する AI を作ることを考える。日本語でやりたいことを入力したらソースコードが出てくるタスクは、GPT をそのままソースコードに対して適応すると大体できてしまう。一般にソースコードはコメント文があってその下にインプリメンテーションがあり、例えば、パイソンコードが書いてある。GPT の学習は生のソースコードを前半と後半に分け、前半部分を所与としたときに後半部分をマスクして前から後ろを無理矢理予測させることで行う。最初はランダムな予測だが学習が進むにつれて、GPT は前半が入力された時に後半を非常に良く予測できる能力を持つようになる。これは GPT がコメント文と次に続く

ソースコードの関係性を非常にきれいに学習しているということである。

今日の GPT によると生のデータを学習させるだけで、「日本語からソースコードを書く」というタスクをはじめ、その他の様々なタスクも偶然学べてしまうようである。逆に言うとこの GPT というモデルを学習し終えたときにモデルが何をできるようになっているかは未知数である。少なくともコメント→コードの関係性を学べることはわかったが、他にも色々あると思われる。現在は色々な人が色々な使い方を発見している状況である。

他の代表的な例として質問を入れると答えがでる、辞書として使えるものがある。また、指示とデータをペアで入れるケースもあり、例えば、英語の複雑な記録に対して「日本語にした上で要点を三つ列挙せよ」と複数のタスクを一気に実行できる。GPT は複合的なタスクでもその複雑な構造を自然に解釈できる点で非常に強力なツールであると言える。

3.4.2.3 NEC の LLM

次に、NEC で LLM の性能をどうやって高めていったかを述べる。LLM の性能には Scaling-Law があることが知られており、コーパスとしてどのくらいのボリュームのデータを学習させるか、どのくらい複雑なネットワーク（モデルサイズ）を準備するかがによって性能が定まると言われている。

2018 年頃はデータ量が簡単に増やせなかったため、モデルサイズ競争が起こった。これは GPT3 がリリースされた時期に 175B というパラメータ量で一度鎮静化した。そしてもう一つの方向性として、モデルサイズを抑えめにして、学習するデータ量を増やすのが良いのではないかという考え方が最近では主流になっている。

コスト面では、GPT3 の 175B とは、約 200 万円の GPU を 8 枚並べてやっと実装できるくらいのサイズであり、これを推論させようとするとトータルのシステムとして 2000 万円くらいのコストがかかることになる。これを個別の業務に使うにはコスト面でミートしない可能性が高い。そこで、NEC では運用時のコストを最低限に抑えつつ、良い性能を得ることを目標に、性能とコストの現実的なバランスを見出すための実験を行った。

NEC では 13B のパラメータで世界トップクラスの日本語性能を有する軽量な LLM を開発した（図 3.4.2-1）。GPT3 が GPU8 枚のところ NEC は GPU 1 枚で良い。性能は、JGLUE ベンチマークで確認した。JGLUE は、ある文章と、その文章に対する質問もセットになっているので、その質問に対して文章を読解して答えを出してあげるタスクを実行することで性能評価が可能となる。このベンチマークで良好な性能を得た。

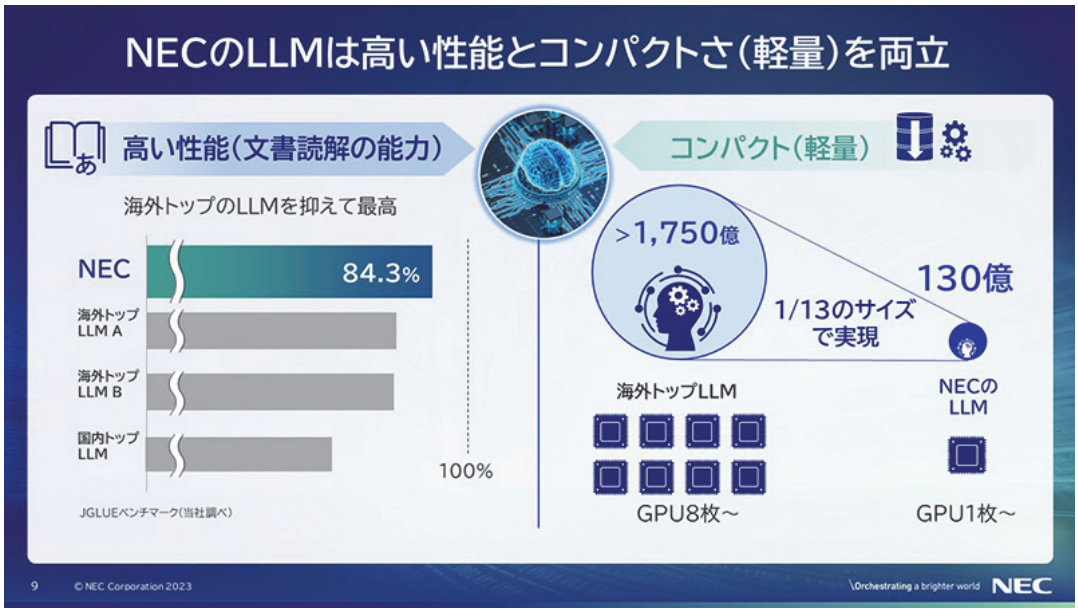


図 3.4.2-1　NEC の LLM の性能とサイズ

LLM の性能を面積で説明するならば図 3.4.2-2 のようになる。LLM の学習には非常にコストがかかるが NEC は国内トップクラスの GPU スパコンを持っておりこれを活用することで高い性能とコンパクト性を実現できた。

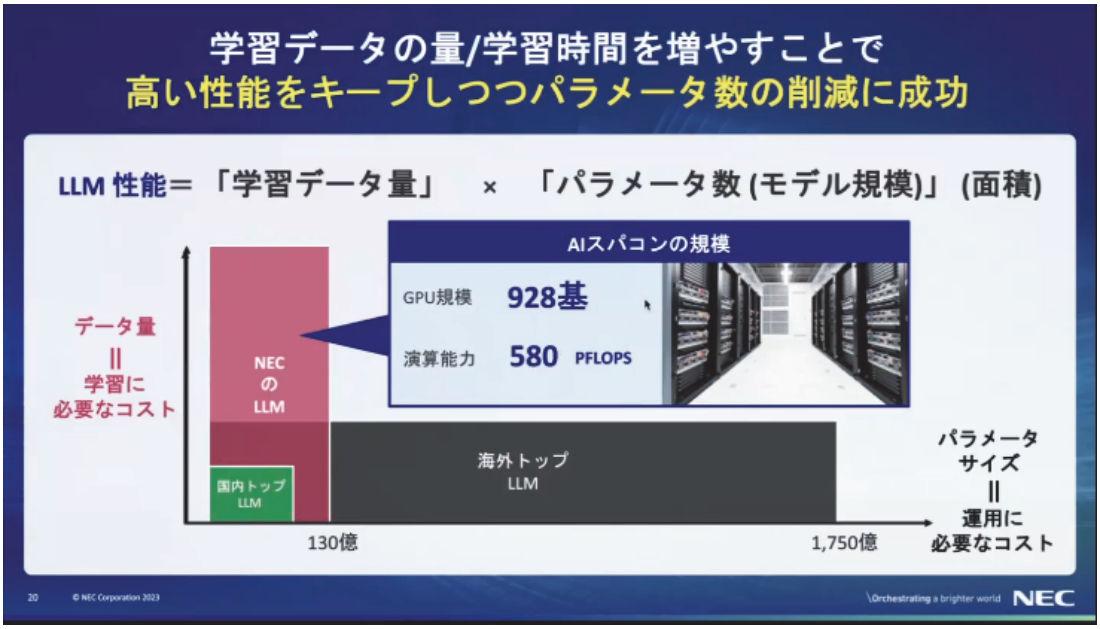


図 3.4.2-2　学習データ量とパラメータ数の LLM 性能との関係

NEC で作成したデモシステムでは、例えば、株価予測をするコードや画像二値化のコードを書くように指示を与えると、科学計算ライブラリや OpenCV などもバインドしながらほぼ正しいコードを出力できる。

NEC では性能的に先行した LLM を用いて、事業化を進めようと考え、社内でも NEC Generative AI Hub という組織を作った。技術を業務につなげていくために技術者だけでなくデータサイエンティストやコンサルタントという人材をそろえ、研究も北米や欧州にいてこれらを一気通貫でまとめ上げる組織は好評をいただいている。（図 3.4.2-3）

図 3.4.2-3　NEC Generative AI Hub

ビジネスを考えると LLM だけでは価値になりにくいため周辺サポートの色々な仕組みとして「NEC Generative AI Service Menu」も提供を開始している。また、LLM の継続的な性能改善に向けて、検索エンジンの検索と LLM を組み合わせるリトリーバル・オーグメンテッド・ジェネレーション(RAG)も検討している。文章を検索することも大事だし、検索したものを LLM にいかに賢く供給するかも大事である。他にも、出てきた結果が正しいのかどうかのファクトチェックを人間系や技術でサポートする仕組みや、答えがどこからどう出てきたのかの情報にフォーカスする仕組みも検討を進めている。

3.4.2.4　LLM を活用した他分野研究（データ統合・拡張に関する R&D 活動）について

NEC では LLM をデータ統合・拡張等の他の分野に応用することも検討している。データ分析は、過去に起こった事象に対して洞察を加えるものと、その事象をもとにして未来に何が起こるか予測することに大別される。洞察については来店者の傾向や購買された商品の全体的な傾向を知りたい場合、売り上げ情報をカテゴライズして、カテゴリごとの売り上げを集計したりする。予測については回帰分析や時系列モデリング等を使っておにぎりの売り上げ推移を予測したりする。これまでは、洞察でも予測でも分析アルゴリズムを頑張って改善しても、お客様の環境では

期待した性能が得られないことがあった。

これは、ガーベッジイン、ガーベッジアウトとして知られており、良くないデータを入れると分析アルゴリズムが良くても良好な結果が得られないというものである。例えば、POS データに商品名と価格のみ記載され、これに売り上げが追記される場合を考えると、会計上はこれで良いが、いったん需要予測をしようとすると、価格と商品名だけで売り上げを予測するような無理難題のタスクになる。それゆえに、一部では通常のオペレーションで使う情報だけでなく、商品のより詳しい属性を付与している（マスタメンテナンス）。一個一個の商品は別物に見えても、付けた属性で言えば同じ属性をもっているので機械学習、回帰分析の数式を作ることができる。このフラグ付けは非常に重要であるがお金も手間もがかかるし間違いも入り込む。NEC では自動でこれを作り出すことも検討している。これが LLM につながっており、国内検索エンジンのコンペティションでもトップの成績を獲得した（図 3.4.2-4）。

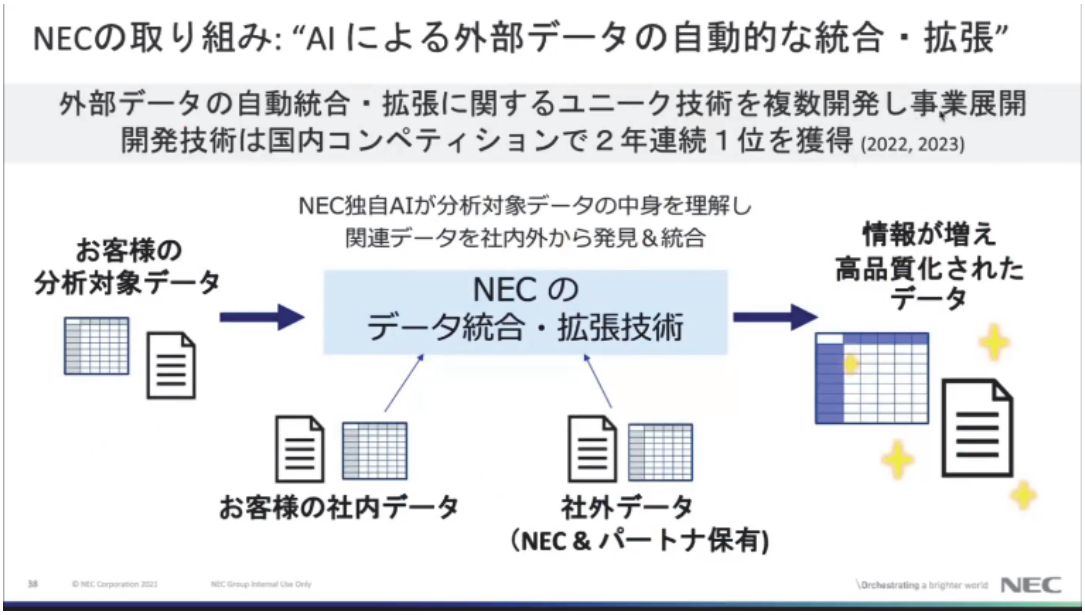


図 3.4.2-4　NEC のデータ統合・拡張技術

なんらかのデータがあったとき、そのデータに対して属性情報を付与してリッチにするというデータ拡張を下支えする LLM も作っている。属性拡張には、外部の情報源からいかに元々のデータに関するものを吸い上げて成形するかが重要になる。ある商品があったときに、商品名に関して検索をすると無数の情報が出てくるが、ここから重要な情報を吸い上げるには色々なやり方が考えられる。

この課題に対して、古典的な自然言語処理分野における「オープンドメイン質問応答」が援用でき、これが LLM の RAG 機能につながっている（図 3.4.2-5）。これは、ある質問があったとき、その質問に対して外部のドキュメントを検索してもよく、出てきた検索結果に対して質問を照ら

し合わせて質問の答えを探したり生成したりしなさい、というタスクをトランスフォーマの GPT で行っている。これが検索と LLM の組み合わせで実現されている。

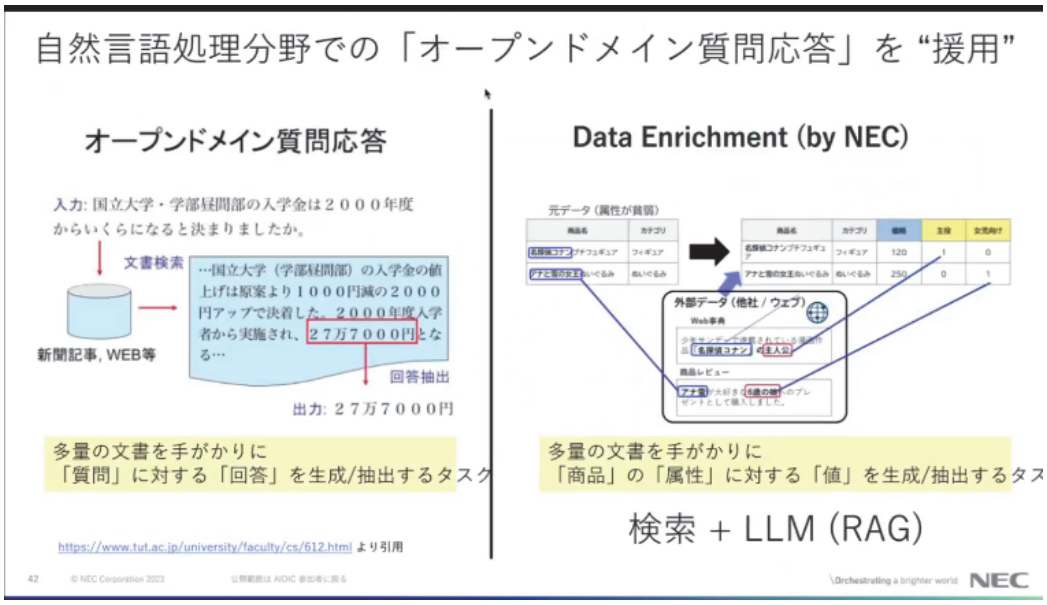


図 3.4.2-5　オープンドメイン質問応答を援用した LLM の RAG 機能開発

これをオープンドメイン質問応答にバックキャストしてみようということで、AI のクイズのコンペティションにも参加した。2023 年 9 月時点では、2 連覇していて、東大王のようなクイズ問題を出すとかなりの高率で正解できる。手掛かりになる文章を検索して文章を参考にクイズに答えるタスクになるが、検索結果の中にクイズの答えがそのまま書かれていることはまれなので、そこから答えを推論することが重要になる。この推論に LLM を使うことで精度を高めることができた。

[参考文献]

[1]　NEC、130 億パラメータで世界トップクラスの日本語性能を有する軽量な LLM を開発
https://jpn.nec.com/press/202307/20230706_02.html

[2]　大規模言語モデル（LLM）を開発
https://jpn.nec.com/rd/technologies/202308/index.html

4. 文字認識システムの技術調査

4.1 OCR の現状

本報告書の 1 章から 3 章には本委員会のミッションやパターン認識技術そして OCR 技術というように様々なレベルや視点から見た現状と未来が展望されているが、この 4.1 節ではより現場に近いところから OCR の現状を俯瞰する。

近年の AI ブームにより OCR 業界においても AI-OCR というキーワードを用いるベンダーが増えてきている。AI-OCR と旧来の OCR の違いについては提供する各社様々な定義をしているので[1][2][3]を参照されたし。

また、AI-OCR と時を同じくしてクラウドを利用した OCR サービスが増えてきている。当委員会の年度報告書における製品分類をみると、当初サービス OCR はソフトウェア OCR のカテゴリとして位置付けていたが、2019 年頃に製品数が増加したことに伴いサービス OCR をソフトウェア OCR から独立させて製品形態の 1 つとして位置付けることになった [4]。

このサービス OCR によってこれまでシステムが OCR を利用する際には SDK／ライブラリーを用いる形態が中心であったが、クラウドの API を利用する形で OCR のサービスを受けることが可能となり、サービスを受けた総量によって課金が発生する新しい業態への変化が起きたといえる。

また、固定帳票だけでなく柔軟性の高い帳票認識機能をもっている AI-OCR の特徴を活かして、請求書、領収書といった特定の文書種類に特化したサービスが増えてきているのも昨今の特徴である。IT トレンドによるとこういった特定の業務に特化したものを「業務特定非定型フォーマット型」[5]と呼んでいる。

AI-OCR を称する製品が多数上梓されてから数年経った現在において振り返ってみると、

(1) クラウドを利用した OCR サービス

(2) 固定フォームだけではない準定型・非定型といった柔軟性のある認識[6][7][8]

(3) DL などの新しい方式を利用した精度の向上

といった 3 つが AI-OCR 製品群の最大公約数的な特徴と感じる。

もちろん、RPA など他サービスへの連携や、DX を実現する上での 1 機能として OCR が前面に出てこないケース、そしてクラウドサービス以外の提供形態も存在するため、各社の定義に逆らって AI-OCR 全体を定義したいわけではないことをお断りしておく。

さて、OCR のサービス化により市場製品にどのような変化が表れているかをみていく。

サービス OCR のカテゴリ製品はそれ単体でサービスとなっており、API 駆動あるいは Web ブラウザを用いた UI で動かすことができるため、操作するクライアント PC のスペックを気にしなくてよいことが多い。そのためソフトウェアなどの製品表で記載のある OS やメモリサイズが記

載なしとなっている製品が多く存在する。また、認識速度についても記載のない製品が多い。速度については OCR 処理をするサーバー側のスペックが非公開である上に、入力画像の転送時間など処理時間に関わる要因が多数あるためやむを得ない対応であろう。さらに認識対象文字種、認識書体、文字サイズなどの項目も非公開となっている製品が多数ある。こちらの理由については調査できていない。従来のソフトウェア OCR では公開されていた項目であったために AI-OCR 化、サービス化、クラウド化のいずれかに関連する影響なのかもしれない。

AI-OCR 製品の比較記事[9]なども多数あるので製品利用を考えている方は参考になるであろう。

その他 2023 年に出てきた OCR 製品としては、凸版印刷社が公開した「ふみのは」[10]という古文書の OCR がある。それ以前にも「みを(miwo)」という AI くずし字アプリ[11]などはあったがこの分野の活性化に期待がかかる。

法律の影響という意味では電子帳簿保存法[12]とインボイス制度[13]について言及しておく。改正された電子帳簿保存法は 2022 年に施行されたものの 2023 年 12 月までは「宥恕期間」となっていた。また、インボイス制度は 2023 年 10 月から施行された。これらの法律の施行前には電子帳票やインボイスと AI-OCR を絡めた製品が多数上梓、あるいは機能的アップデートがされており、OCR 関連では一つの注目分野となっている [14][15][16]。

2022 年に ChatGPT が公開[17]されて依頼、日本国内においても生成 AI に関する話題が増えているのは周知のとおりであるが、OCR の世界でも生成 AI の利活用が始まっている。シナモン社から生成 AI と AI-OCR を組み合わせた帳票からの情報抽出を行う「Flax Scanner HUB」を 2024 年 4 月に販売開始という発表があった[18]。アライズイノベーション社からは AI-OCR（AIRead）と Microsoft 社の生成 AI を用いて金融機関の業務の高度化（金融 DX）を図る実証検証を 2024 年 1 月から実施するという発表があった[19]。注目技術である生成 AI を OCR 製品に搭載していく流れは今後も進んでいくものと思われる。

【参考文献】

[1] AI OCR とは？成功のポイントは認識精度を理解し業務全体を見直すこと、キヤノン、https://canon.jp/business/trend/ai-ocr

[2] 「AI OCR」とは～OCR との違いと 3 つのメリット、リコー、https://www.ricoh.co.jp/service/cloud-ocr/column/aiocr/

[3] AI-OCR とは？OCR との違いや種類・導入メリット・比較のポイントを解説、AI Similey、https://aismiley.co.jp/ai_news/what-is-ai-ocr/

[4] 認識形入力方式に関する調査研究報告書 2019、電子情報技術産業協会 認識形入力方式標準化専門委員会、2020 年 3 月

[5] AI-OCR とは？従来型との違いやメリット・デメリットを解説、IT トレンド、

https://it-trend.jp/ai_ocr/article/813-576

[6] ここまで読める！高精度に紙帳票を読み取る AI OCR が切り拓くデジタル化の世界、東芝デジタルソリューションズ、
https://www.global.toshiba/jp/company/digitalsolution/articles/tsoul/solution/s010.html

[7] 失敗しない AI-OCR 製品の選び方とは？メリットや比較するためのポイントをご紹介、PFU、
https://www.pfu.ricoh.com/fi/digitarakuru/column-0002.html

[8] Intelligent OCR、AI Inside、https://dx-suite.com/about/function#intelligentOcr

[9] AI OCR 製品を比較！自社に適した選び方も徹底解説、IT トレンド
https://it-trend.jp/ai_ocr/article/813-4675

[10] 古文書解読とくずし字資料の利活用サービスふみのは、凸版印刷、
https://www.toppan.com/ja/joho/fuminoha/

[11] みを（miwo）：AI くずし字認識アプリ、人文学オープンデータ共同利用センター、
http://codh.rois.ac.jp/miwo/

[12] 電子帳簿保存法の内容が改正されました、国税庁、
https://www.nta.go.jp/law/joho-zeikaishaku/sonota/jirei/pdf/0023003-082.pdf

[13] インボイス制度の概要、国税庁、
https://www.nta.go.jp/taxes/shiraberu/zeimokubetsu/shohi/keigenzeiritsu/invoice_about.htm

[14] 次世代 AI OCR「SmartRead」においてインボイス制度に対応した請求書の自動読取り機能をリリース、コージェントラボ、
https://www.cogent.co.jp/news/cogentlabs_smartread_invoice_202303/

[15] 経費精算を効率化　インボイス制度・電子帳簿保存法に対応、バクラク、
https://bakuraku.jp/expense/

[16] 電子帳簿保存クラウドサービスの DenHo（デンホー）、インフォディオ
https://www.smartocr.jp/denho/

[17] Introducing ChatGPT、OpenAI、https://openai.com/blog/chatgpt

[18] AI-OCR x 生成 AI であらゆる帳票の情報抽出を実現。企業のデータ活用を大幅に拡張する定型・非定型対応 AI-OCR「Flax Scanner HUB」を先行提供開始、シナモン
https://cinnamon.ai/news/20240119-press-ai-ocr-flax-scanner-hub/

[19] 複数の地方金融機関と共同で AI-OCR と生成 AI を用いた金融 DX の実証検証を開始、アライズイノベーション、https://ariseinnovation.co.jp/news/n20240125/

（URL 確認 2024.3.17）

4.2　製品分類について

2024年2月における主要なOCR製品を、一覧表の形で整理した。（表4.2-1～表4.2-20）

以下、一覧表はOCR製品を以下の7つに分類したものとなっている。

○　ハードOCR製品 - 帳票OCR	表4.2-1
○　ソフトOCR製品 - 帳票OCR	表4.2-2～表4.2-5
○　ソフトOCR製品 - 文書OCR	表4.2-6～表4.2-8
○　ソフトOCR製品 - 名刺OCR	表4.2-9
○　ソフトOCR製品 – 本人確認書類OCR	表4.2-10～表4.2-14
○　ソフトOCR製品 － マルチタイプOCR	表4.2-15
○　サービスOCR製品	表4.2-16～表4.2-20

製品分類の視点には、提供形態と対象文書という二つの軸を用いて分類を行った。

第一の整理軸は提供形態に関するものであり、具体的には提供形態がハードウェアかソフトウェアかクラウド上のサービスかという区分である。

ハードウェアのOCR製品とは、文字認識における主要な処理を高速で実行するための専用処理装置を備えたものをいう。OCR専用スキャナーと文字認識部と共に同時提供するデバイスタイプと呼ばれる形態が現在では主流となっている。ハードウェアOCRでは一般には通常のスキャナーより高速・高品質なものが用いられることが多く、大量のデータを高速・高精度に処理できることがメリットである。そのメリットを活かして、ハードウェアOCRは定型フォーマットの帳票を読み取り基幹系業務ソフトウェアと連携してデータ処理する用途で用いられることが多い。

一方、CD-ROMやDVDのような情報記録媒体あるいはネットワークを介して、パーソナル・コンピューター等にインストールして使用するタイプのOCR製品をソフトウェア製品と呼ぶ。スマートフォン、PDA等のモバイル機器にインストールして使用するタイプのOCR製品も含まれる。ソフトウェア実装の利点は、いわゆるモノとしての生産コストが不要であることに加え、画像入力に一般的なスキャナーやデジタルカメラ等を用いることができるため、既存の装置を流用・共用することが可能となり、その結果としてトータルな導入価格を低く抑えうることにある。

近年はユーザーのサーバーやパーソナル・コンピューターにインストールすることを不要とした、クラウド上でOCR処理を行うことで処理結果を得るサービスとしてのOCRも普及してきており、3つ目の提供形態としてサービスOCRという分類を定義した。サービスOCRは導入コストがソフトウェアOCRよりも低価格に抑える代わりに、月額基本料金や利用量に応じたランニングコストを支払う販売方法をとるものも多い。

このように提供形態の違いは、商品形態の違いを生み出し、市場セグメントや販売チャネルな

どに大きな差異をもたらしている。したがって、この整理軸は事業的な観点から特に重要な軸と考えることができる。

第二の整理軸は対象原稿に関するものであり、具体的には対象原稿が帳票であるか、一般文書であるか、名刺であるか、本人確認書類（免許証保険、保険証、マイナンバーカードなど）であるかという分類である。今回から新たに複数種類の対象原稿をユーザーの指定なしに対応できるマルチタイプという分類も定義した。

この整理軸は技術的な観点からより重要となる区分である。すなわち、対象原稿が異なることによって、読み取られる文字列の位置や文字種に関する制約に関して大きな差異が生じる。そして、まさに制約こそが情報処理（特にパターン認識系）の実用性（精度・性能）を左右する本質的要素の一つなのである。

帳票OCRが処理対象とする“帳票”は、一般には罫線によって、抽出されるべき文字領域が区分されている。あるいは罫線に代わるものとして、背景パターン・色や並びの整然さ等を仮定することができる。対象文字は、頁全体としては手書き及び印刷された活字の両方を含むことが可能だが、個々の記入領域に関しては制約が強いのが一般的である。制約は文字の種類によるものだけでなく、意味（部品名か数量か、住所か氏名かなど）によるものも含む。そのため、辞書的な知識を適用することで、実用性（精度・性能）を向上させることが可能となっている。帳票OCRの主な用途として、受発注業務やアンケート集計におけるデータ・エントリーが挙げられる。

名刺OCRが処理対象とする“名刺”は、基本の大きさは55mm×91mmの限られた小さな紙に、氏名、会社名、住所、電話番号、メールアドレスなどの項目を記したものである。一般文書とは異なり、記載されている項目は限定され、ある一定のレイアウトデザインが存在する。名刺の向きを自動的に補正し、レイアウト解析と呼ばれる処理によって、項目の異なるそれぞれの領域を抽出する必要はあるが、大容量の電話番号辞書や郵便番号辞書を搭載し、会社名や住所の誤認識を自動修正することにより、十分な実用性をもたせることが可能となっている。名刺OCRの主な用途として、登録された名刺の検索、閲覧機能だけでなく、登録された住所の地図を調べる機能、登録された住所までの経路を調べる機能、個人情報保護法の施行に伴い、名刺情報を保護するセキュリティ機能まで搭載された商品が発売されている。

本人確認書類は、2015年10月に施行されたマイナンバー制度に関連する書類、あるいは運転免許証などの本人確認書類を対象とするOCRソフトウェアを想定している。各社今後こういった認識のソリューションが増えることを想定して専用にパッケージングしていると考えられる。

文書OCRが処理対象とする“文書”は、広くは帳票、名刺、本人確認書類ではないその他一般文書という意味である。具体的には、書籍・新聞・雑誌・論文・報告書・通達文などがこれに該当する。（設計図面や建築図面などは該当しないとされる。）文書の構成には、罫線のような明確なセパレータが用いられないため、レイアウト解析によって、属性の異なるそれぞれの領域を抽

出する必要がある。文字の種類に関する制約はほとんどないといってよく、わずかに「特定言語で書かれた」「印刷（活字）文字」であるという程度である。しかしながら、この制約もそれなりに強力なものであって、十分な実用性をもたせることが可能となっている。文書OCRの主な用途として、印刷文書の再利用（テキスト化）、及び検索性の付与が上げられる。法人市場のみならず個人市場においても利用可能である点が、帳票OCRとは異なっている。

このように、事業的な観点から重要な提供形態という整理軸と、技術的な観点から重要な対象原稿という整理軸との二つの軸を使用して整理を行った。ただし近年、ハードウェアOCRは帳票分野以外存在しないので、ソフトウェアOCRとサービスOCRの2つが対象原稿の軸で分類される形になっている。それぞれの分類の動向を把握することにより、OCR市場の概略を事業的観点及び技術的観点から理解することが可能となるだろう。

表4.2-1 ハードＯＣＲ製品(帳票ＯＣＲ装置)

製品名	メーカ	処理速度(枚/分)		文字認識速度(/s)	スキャナ						機能				機構部 HxDxW mm Kg	回路部 HxDxW mm Kg	インタフェース	フォーマット指定方式	発売年月	価格(万円)[税別]
		A4 300字	A8 10字		帳票サイズ	ホッパ容量	スタッカ容量 アクセプト	スタッカ容量 リジェクト	解像度(DPI)	帳票厚(連量)	漢字読取 手書/活字	知識処理	画像出力	その他機能						
HT-4161	日　立	約60(ｵﾌﾟｼｮﾝ使用時約85)	約120		52x74～305x420	約300枚	約300枚	約100(2ｽﾀｯｶﾓﾃﾞﾙのみ)	100～600	35kg～180kg紙	OP手書/活字約4,300種	OP	あり	低電力モード(ｽﾘｰﾌﾟﾓｰﾄﾞ)，OP 異サイズ帳票混在/両面読取/カラー画像入力	1ｽﾀｯｶﾓﾃﾞﾙ 約351x約331x約410(ﾎｯﾊﾟ・ｽﾀｯｶ収納時)、約351x約533x約410(ﾎｯﾊﾟ・ｽﾀｯｶ使用時)、約18.5kg	(機構部に含む)	USB3.0/USB2.0	GUIによる帳票定義	2012/7	1ｽﾀｯｶﾓﾃﾞﾙ:131万円～2ｽﾀｯｶﾓﾃﾞﾙ:181万円～
HT-4165	日　立	約210	約380		52x74～305x458	約550枚	約550枚	約100枚	100～600	35kg～180kg紙	OP手書/活字約4,300種	OP	あり	低電力モード(ｽﾘｰﾌﾟﾓｰﾄﾞ)，OP 異サイズ帳票混在/両面読取/カラー画像入力	約407x約715x約485、約638x約980x約485(延長トレイ使用時)、約43Kg	(機構部に含む)	USB3.0/USB2.0	GUIによる帳票定義	2015/7	片面ﾓﾃﾞﾙ:241万円～両面ﾓﾃﾞﾙ:331万円～
Blinkscan	日立チャネルソリューションズ株式会社	－	－	－	216mm×297mm	－	－	－	200/240/300/400/480	被写界深度10mm	－	－	あり	カラー読取り	382x372x270 約3.5Kg	(機構部に含む)	USB2.0 (HS)	－	2011/4	ｵｰﾌﾟﾝ価格
N6370E	日本電気	120/210	300/420		74x52～420x305	55mm	55mm	10mm	100～600	40.7～209.3/㎡	OP手書約4000種	あり	あり	あり	435x545x450 49kg	(機構部に含む)	USB3.0	フォーマットプログラム	2017/9	160～
N6370M	日本電気	55	－		105x75～297x235	16mm	16mm	10mm	100～600	40.7～174.5g/㎡	OP手書約4000種	あり	あり	あり	275x420x380 24kg	(機構部に含む)	USB3.0	フォーマットプログラム	2018/2	88～

表4.2-2 ソフトＯＣＲ製品(帳票ＯＣＲ その１)

製品名	メーカ	認識対象文字種	認識書体	文字サイズ(400DPI)	認識速度(字/秒)	OS	メモリ	発売年月	価格(円)	備考
帳票OCR サーバー版 Ver.2.02	パナソニックソリューションテクノロジー	手書き：英字、数字、ひらがなの一部、カタカナの一部、記号の一部、第1水準漢字、第2水準漢字の一部(518字) 活字：約6800字、数字、英大、カナ、ひらがな、JIS記号（一部）168字、ギリシャ文字（一部）32字、第1水準漢字、第2水準漢字	手書き、活字、バーコード(NW-7/CODE39/GS1-128/CODE128/JAN-8/JAN-13/ITF-6/ITF-14/ITF-16）、QRコード、チェックマーク（レ点、塗りつぶし、オーバーライト）	6～60ポイント(400dpiの場合)		Windows Server 2016 / 2012 R2/ 2012 / 2008 R2 SP1以上/ 2008 SP2以上 いずれも日本語版	お使いのOSが推奨する環境以上	2017/2	サーバー版 2,000,000［税別］	
AI帳票OCR Ver.9 (WisOCR)	パナソニックソリューションテクノロジー	手書き：英字、数字、ひらがなの一部、カタカナの一部、記号の一部、第1水準漢字、第2水準漢字の一部 活字：約6800字、数字、英大、カナ、ひらがな、JIS記号（一部）168字、ギリシャ文字（一部）32字、第1水準漢字、第2水準漢字	AI手書き、マルチフォント（明朝体、ゴシック体、教科書体、ワープロ体、新聞文字など）、バーコード、QRコード、チェックマーク	6～60ポイント(400dpiの場合)		Windows 11 version 21H2/22H2 Windows 10 version 22H2	お使いのOSが推奨する環境以上		要問合せ［税別］	Basic/Standard/Pro 月額基本利用料 PC上で動作するクライアントアプリとクラウドのAI-OCRエンジンで構成される
帳票認識ライブラリー Ver.8.51	パナソニックソリューションテクノロジー	手書き（英字、数字、ひらがなの一部、カタカナの一部、記号の一部、第1水準漢字、第2水準漢字の一部(518字)）、活字（約6800字、数字、英字、カナ、ひらがな、JIS記号（一部)168字、ギリシャ文字（一部）32字、JIS第1水準漢字、JIS第2水準漢字)	手書き、活字、バーコード(NW-7/CODE39/GS1-128/CODE128/JAN-8/JAN-13/ITF-6/ITF-14/ITF-16）、チェックマーク（レ点、塗りつぶし、オーバーライト）	6～60ポイント(400dpiの場合)		Windows 10/8.1/Windows Server 2019/2016/2012 R2/2012	対応OSが必要とする最低メモリーに加えて256MB以上(512MB以上を推奨)	2021/2	550,000［税別］	AI手書き文字認識オプションあり（価格は別途ご相談）
imageWARE Scan Manager DS V1.1	キヤノン	手書き（数字、英大文字、カナ、ひらがな、漢字、記号）、活字(日本語、英語)	手書き、OCR-B、ＭＩＣＲ、チェックライター文字、明朝、ゴシック、Helvetica、Courier New、Times New Roman	6ポイント～48ポイント相当		Windows 8/8.1/10 Windows Server 2008/2012	1GB以上	2012/5	400,000［税別］	
MELFOS	三菱電機ITソリューションズ	ー	-	-	-	-	-		要問合せ	サービスも有

表4.2-3 ソフトＯＣＲ製品(帳票ＯＣＲ その２)

製品名	メーカ	認識対象文字種	認識書体	文字サイズ(400DPI)	認識速度(字/秒)	OS	メモリ	発売年月	価格(円)	備考
WinReader Hand S v.6.0	NTTデータNJK(メディアドライブ)	手書(英数字,カタカナ,ひらがな,漢字4419文字,記号)	手書(OCR-HN,OCR-HK,OCR-HS,OCR-B,OCR-Kなど)	4x6mm～12x12mm		Windows 11/10	2GB以上	2013/3	300,000円［税別］	
FormOCR v.8.0	NTTデータNJK(メディアドライブ)	手書(英数字,カタカナ,ひらがな,漢字4419文字,記号) 活字（英数字,カタカナ,ひらが,漢字6355文字,記号) バーコード/QRコード	手書(OCR-HN,OCR-HK,OCR-HS,OCR-B,OCR-Kなど) 活字（マルチフォント,IBM407,12F)	手書（4x6mm～12x12mm) 活字（3～15mm)		Windows 11/10	8GB以上	2022/8	600,000円［税別］	AI-OCR機能（フリーピッチ手書き文字認識機能、他搭載）ServerOS(Windows Server2022/2019/2016）対応版あり、色指2値化オプション、カラー分離帳票認識オプションあり
帳票認識ライブラリ v.9.5	NTTデータNJK(メディアドライブ)	手書(英数字,カタカナ,ひらがな,漢字4419文字,記号) 活字（英数字,カタカナ,ひらが,漢字6355文字,記号) バーコード/QRコード	手書(OCR-HN,OCR-HK,OCR-HS,OCR-B,OCR-Kなど) 活字（マルチフォント,IBM407,12F)	手書（4x6mm～12x12mm) 活字（3～15mm)		Windows 11/10	4GB以上	2021/6	1,500,000円［税別］	AI-OCR機能（フリーピッチ手書き文字認識機能、他搭載）ServerOS(Windows Server2022/2019/2016）対応版あり、色指2値化オプション、カラー分離帳票認識オプションあり
FUJITSU AI-OCR「Keyword Capture Client Edition」	富士通フロンテック					Windows 10 (64bit)	4GB以上（推奨8GB以上)	2019/10	個別見積	帳票定義を用いる「定義型」と、キーワード登録により汎用帳票に対応する「定義レス型」の両方のタイプの帳票認識機能を提供

表4.2-4 ソフトＯＣＲ製品(帳票ＯＣＲ その３)

製品名	メーカ	認識対象文字種	認識書体	文字サイズ(400DPI)	認識速度(字/秒)	OS	メモリ	発売年月	価格(円)	備考
DynaEye 10	PFU	活字、手書き文字（数字、日本語）、手書きマーク、QRコード、バーコード	手書き，OCR-B,OCR-K，MS明朝，MSゴシック,手書きフリーピッチ（オプション）			Windows 10/11（AMR版除く）	2GB以上（4GB推奨）	2019/09	要問合せ	標準アプリケーション、帳票OCR Entry、SDK/ランタイムなどの提供形態あり
DynaEye 11	PFU	活字、手書き文字、フリーピッチ手書き文字、QRコード、バーコード	手書き，活字（オムニフォント），手書きフリーピッチ			Windows 10/11（AMR版除く）	2GB以上（4GB推奨）AI-OCRは8GB以上（12GB推奨）	2022/07	Entry:初期1,008,000円[税別]・継続168,000円/年[税別] AI-OCR：初期2,016,000円[税別]・継続336,000円/年[税別]等	DynaEye 11 Entry、DynaEye 11 Entry AI-OCR、DynaEye 11 Entry マルチステーション、SDK/ランタイムなどの提供形態あり
FAX/OCR SYSTEM 伝匠 V9	リコージャパン	手書き（数字、英大文字、カナ、漢字、ひらがな、記号）、活字（数字、英字、カナ、漢字、ひらがな、記号）	手書き、明朝、ゴシック、（OCR-B,Original）	手書き(5mm以上推奨)、活字(6～60ポイント推奨)		Windows 10/11 Server2012R2/2016	OCRサーバ：1GB以上 結果修正：512MB以上推奨	2012/04	680,000～[税別]	FAXサブシステム無。自動方向判別機能。自動帳票識別機能。手書き住所辞書、氏名辞書標準添付。帳票定義ツール付き
FAX/OCR SYSTEM 伝匠 V10	リコージャパン	手書き（数字、英大文字、カナ、漢字、ひらがな、記号）、活字（数字、英字、カナ、漢字、ひらがな、記号）	手書き、明朝、ゴシック、（OCR-B,Original）	手書き(5mm以上推奨)、活字(6～60ポイント推奨)		OCR処理側：Windows Server/2016/2019 認識結果修正側：Edge/Chrome	サーバ：8GB以上 結果修正：2GB以上	2017/12	880,000～[税別]	バーコード、QRコードの認識。TIFFに加えて、カラー画像、PDFファイルも処理可能。OCR結果確認修正をWebブラウザを使用（クライアント配布不要）

表4.2-5 ソフトＯＣＲ製品(帳票ＯＣＲ その４)

製品名	メーカ	認識対象文字種	認識書体	文字サイズ(400DPI)	識速度(字/秒	OS	メモリ	発売年月	価格(円)	備考
GAZOMAGIC	日立チャネルソリューションズ	ー	ー	ー	ー	ー	ー		要問合せ	TWAIN対応、ﾊｰｺｰﾄﾞ認識、印鑑照合、帳票識別機能等。住所、氏名知識辞書、帳票定義ﾕｰﾃｨﾘﾃｨあり。 手書きフリーピッチ(住所,氏名,数字列等)
帳票マスタ SE V4	日立ソリューションズ・テクノロジー	活字、手書き	英数字、ひらがな、カタカナ、JIS第1水準漢字、JIS第2水準漢字、記号			MS Windows/MS Windows Server 詳細は要問合せ	1GB以上(3GB推奨)		580,000［税別］	
帳票マスタ LE V4	日立ソリューションズ・テクノロジー	活字、手書き	英数字、ひらがな、カタカナ、JIS第1水準漢字、JIS第2水準漢字、記号			Windows7(32/64ビット版)	1GB以上(3GB推奨)		380,000［税別］	
TeleForm	Hammock	ー	-	-	-	-	-		要問合せ	
AnyForm OCR	Hammock							2019/7/3	オンプレ版2,000,000円[税別]～※製品構成により価格が異なります。	クラウド版もあり
AI Read	アライズイノベーション	活字、手書き	ー	-	-	-	-		オンプレミススタンドアロン版1,320,000円［税別］	事例：注文書、請求書、決算書、アンケート用紙、チェーンストア共通伝票、特許関連書類 クラウド版もあり

表4.2-6 ソフトＯＣＲ製品(文書ＯＣＲ　その１)

製品名	メーカ	認識対象 文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	TWAIN 対応	メモリ	発売年月	価格(円)	備考
活字OCRライブラリ v.10.0	NTTデータNJK (メディアドライブ)	約6700+英語+中韓語 <以下オプション> タイ語＋ベトナム語＋マレー語＋インドネシア語	明朝、ゴシック、教科書体などのマルチフォント	5～45ポイント		Windows 11/10			2023/3	500,000円 [税別]	ServerOS(Windows Server 2022/2019/2016) 対応版あり
活字文書OCRライブラリ for Linux	NTTデータNJK (メディアドライブ)	約6700+英語	明朝、ゴシック、教科書体などのマルチフォント	5～45ポイント		Red Hat Enterprise Linux 7/8/9 CentOS 7 Rocky Linux 8/9			2012/8	750,000円 [税別]	中韓語認識オプションあり
活字文書OCRライブラリ for iOS/Android	NTTデータNJK (メディアドライブ)	約3700 (英語含む)	明朝、ゴシック、教科書体などのマルチフォント	7.5～45ポイント		iOS 16～17 Android 10～14			2012/6	550,000円 [税別]	中韓語認識オプションあり オプション認識対象文字種： 中国語（簡体字）6763 中国語（繁体字）13053 韓国語Hangul 2350 韓国語Hanja 4888
OCRパッケージ4	日立	－	－	－	－	－	－			要問合せ	
Mobile OmCR	オムロン ソフトウエア	辞書サイズによる	辞書サイズによるドット文字OCR有り			32bit/64bit CPU マルチOS対応 (Android™, iOS Linux, Symbian, WindowsMobile , etc.)				要問合せ	

表4.2-7 ソフトＯＣＲ製品(文書ＯＣＲ　その２)

製品名	メーカ	認識対象 文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	TWAIN 対応	メモリ	発売年月	価格(円)	備考
活字認識ライブラリー Ver.15.7	パナソニックソリューションテクノロジー	約6,800字(英字、数字、ひらがな、カタカナ、JIS記号（一部）168字、ギリシャ文字（一部）32字、JIS第1水準漢字、JIS第2水準漢字)	マルチフォント(明朝体、ゴシック体、教科書体、ワープロ体、新聞文字など)	6～60ポイント(400dpiの場合)		Windows 11 21H2/22H2 Windows 10 21H2/22H2 Windows Server 2022 21H2/2019 version1809/2016 version1607/2012 R2 SPなし		OSが必要とする最低メモリーに加えて256MB以上（512MB以上推奨)	2021/10	550,000円 ［税別］	
活字認識ライブラリー for iOS/Android V15.6	パナソニックソリューションテクノロジー	約6,800字(英字、数字、ひらがな、カタカナ、JIS記号（一部）168字、ギリシャ文字（一部）32字、JIS第1水準漢字、JIS第2水準漢字)	マルチフォント(明朝体、ゴシック体、教科書体、ワープロ体、新聞文字など)	6～60ポイント(400dpiの場合)		iOS 11/12/14 Android 9/10/11		OSが必要とする最低メモリーに加えて256MB以上　(512MB以上を推奨)	2021/10	550,000円 ［税別］	
読取革命Ver.16	ソースネクスト	約6800	マルチフォント(明朝体、ゴシック体、教科書体、ワープロ体、新聞文字、斜体など)、英語もマルチフォント(Century、Helvetica、Courierなど)	6～60ポイント(400dpi)		Windows 11 Windows 10（32ビット/64ビット版)	○	お使いのOSが推奨する環境以上	2020/10	12,980円 (税込)	「簡単！PDF変換」「簡単！PDF for Office」「クリップボードOCR」「フォルダーウォッチャー」同梱
本格読取5	ソースネクスト	約4000	明朝、ゴシック、教科書体、正楷書体、ワープロ書体、新聞文字など、英語はマルチフォント対応	6～60ポイント(400dpi)		Windows 11 Windows 10（32ビット/64ビット版)	○	-	2016/11	ダウンロード版 4,378円 (税込)	「瞬間テキスト2」「おまかせ名刺管理3」同梱

表4.2-8 ソフトＯＣＲ製品(文書ＯＣＲ　その３)

製品名	メーカ	認識対象文字種	認識書体	文字サイズ(400DPI)	認識速度(字/秒)	OS	TWAIN対応	メモリ	発売年月	価格(円)［税別］	備考
e.Typist v.15.0	NTTデータNJK(メディアドライブ)	約6700+欧米55言語+中韓語	明朝、ゴシック、教科書体などのマルチフォント	5～45ポイント		Windows 11/10	○	1GB以上	2013/9	19,800円［税別］	
WinReader PRO v.15.0	NTTデータNJK(メディアドライブ)	約6700+欧米55言語+中韓語	明朝、ゴシック、教科書体などのマルチフォント	5～45ポイント		Windows 11/10	○	2GB以上	2014/11	198,000円［税別］	ServerOS(Windows Server 2022/2019/2016) 対応版あり、OLEオートメーション開発キットオプションあり
ドキュメントリーダ ExpressReader Pro V4.5	東芝デジタルソリューションズ	約4000	オムニフォント	6～40ポイント	1,200以上	Windows10 x64(1909) Enterprise/Standard	○	512MB以上		195,000円	
Rosetta-Stone-Components V1.71	キヤノンマーケティングジャパン	手書き（数字、英大文字、カナ、ひらがな、漢字、記号）、活字(日本語、英語)	手書き、OCR-B、ＭＩＣＲ、チェックライター文字、明朝、ゴシック、Helvetica、Courier New、Times New Roman	6ポイント～48ポイント相当		Windows 10 Windows Server 2012 R2/2016		256MB以上推奨	2016/10	オープン	前処理（適応的二値化，方向判別他），帳票認識/登録，印影抽出，チェックマーク、丸囲み判定、固有名詞，住所，氏名知識辞書あり

表4.2-9 ソフトＯＣＲ製品(名刺ＯＣＲ)

製品名	メーカ	認識対象文字種	認識書体	文字サイズ(400DPI)	認識速度(字/秒)	OS	TWAIN対応	メモリ	発売年月	価格(円)	備考
名刺認識ライブラリー Ver.3.20	パナソニックソリューションテクノロジー	約6,800字	マルチフォント(明朝体、ゴシック体、教科書体、ワープロ体、新聞文字など)	6～60ポイント(400dpi)	－	Windows 10/8.1/7/ Windows Server 2016/2012 R2/2012/2008 R2 SP1/2008 SP1 SP2		128MB(推奨256MB)以上	2018/2	要問合せ	
名刺認識ライブラリー Ver.3.10 for iOS / Android	パナソニックソリューションテクノロジー	約6,800字	マルチフォント(明朝体、ゴシック体、教科書体、ワープロ体、新聞文字など)	6～60ポイント(400dpi)	－	iOS 6.0 / 6.1 / 7.0 / 7.1/ 8.0/ 8.1 Android 4.0 / 4.1 / 4.2 / 4.3 / 4.4		OSが必要とする最低メモリーに加えて128MB以上(256MB以上を推奨)	2015/3	要問合せ	
本格読取おまかせ名刺管理3	ソースネクスト	－	－	－	－	Windows 10 (32ビット/64ビット版)	○	1GB以上	2016/11	ダウンロード版 2,178(税込)	
名刺認識ライブラリ v.8.0	NTTデータNJK(メディアドライブ)	約3700+英語+中韓国語	明朝、ゴシックほかマルチフォント	5～36ポイント		Windows 11/10			2016/1	500,000［税別］	ServerOS(Windows Server2022/2019/2016) 対応版あり
名刺認識ライブラリ v.3.0 for Linux	NTTデータNJK(メディアドライブ)	約3700+英語+中韓国語	明朝、ゴシックほかマルチフォント	5～36ポイント		Red Hat Enterprise Linux 7/8/9 CentOS 7 Rocky Linux 8/9			2015/9	500,000［税別］	
やさしく名刺ファイリング PRO v.15.0	NTTデータNJK(メディアドライブ)	約3700+英語+中韓国語	明朝、ゴシックほかマルチフォント	5～36ポイント		Windows 11/10	○	1GB以上	2017/11	7,800［税別］	

表4.2-10 ソフトＯＣＲ製品(本人確認書類　その１)

製品名	メーカ	認識対象 文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	TWAIN 対応	メモリ	発売年月	価格(円)	備考
免許証認識ライブラリ v.6.0	NTTデータNJK (メディアドライブ)	漢字(JIS 第1 水準2965 文字、第2 水準3390 文字、機種依存文字111 文字)、ひらがな、カタカナ、アルファベット、数字、記号9 文字（－（）［］／・々ー（長音））	明朝体、ゴシック体、他マルチフォント対応	42.5 ～ 236 ピクセル角程度 （500万画素カメラでカード全体がちょうど収まるサイズの場合、1.8 ～ 10 mm 角程度）		Windows 11/10 Windows Server 2022/2019/2016	ー		2023/9	要問合せ	
免許証認識ライブラリ v.6.0 for Linux	NTTデータNJK (メディアドライブ)	漢字(JIS 第1 水準2965 文字、第2 水準3390 文字、機種依存文字111 文字)、ひらがな、カタカナ、アルファベット、数字、記号9 文字（－（）［］／・々ー（長音））	明朝体、ゴシック体、他マルチフォント対応	42.5 ～ 236 ピクセル角程度 （500万画素カメラでカード全体がちょうど収まるサイズの場合、1.8 ～ 10 mm 角程度）		Red Hat Enterprise Linux 7/8/9 CentOS 7 Rocky Linux 8/9	ー		2023/9	要問合せ	
免許証認識ライブラリ v.5.0 for iOS	NTTデータNJK (メディアドライブ)	漢字(JIS 第1 水準2965 文字、第2 水準3390 文字)、ひらがな、カタカナ、アルファベット、数字、記号（－（）［］／・々）	明朝体、ゴシック体、他マルチフォント対応	42.5 ～ 236 ピクセル角程度 （500万画素カメラでカード全体がちょうど収まるサイズの場合、1.8 ～ 10 mm 角程度）		iOS 16～17	ー		2020/5	要問合せ	
免許証認識ライブラリ v.5.0 for Android	NTTデータNJK (メディアドライブ)	漢字(JIS 第1 水準2965 文字、第2 水準3390 文字)、ひらがな、カタカナ、アルファベット、数字、記号（－（）［］／・々）	明朝体、ゴシック体、他マルチフォント対応	42.5 ～ 236 ピクセル角程度 （500万画素カメラでカード全体がちょうど収まるサイズの場合、1.8 ～ 10 mm 角程度）		Android 10～14	ー		2020/5	要問合せ	
在留カードOCRライブラリ for Linux	NTTデータNJK (メディアドライブ)	漢字(JIS 第1 水準2965 文字、第2 水準3390 文字)、ひらがな、カタカナ、アルファベット、数字、記号（－（）・々／）	明朝体、ゴシック体、他マルチフォント対応	42.5 ～ 236 ピクセル角程度 （500万画素カメラでカード全体がちょうど収まるサイズの場合、1.8 ～ 10 mm 角程度）		Red Hat Enterprise Linux 7/8/9 CentOS 7 Rocky Linux 8/9	ー		2021/6	要問合せ	

表4.2-11 ソフトＯＣＲ製品(本人確認書類　その２)

製品名	メーカ	認識対象 文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	TWAIN 対応	メモリ	発売年月	価格(円)	備考
マイナンバーカードOCRライブラリ v.2.0 for Windows	NTTデータNJK(メディアドライブ)	漢字(JIS 第1 水準2965 文字、第2 水準3390 文字、機種依存文字132 文字)、ひらがな、カタカナ、アルファベット、数字、記号7 文字（－（ ）／・々ー（長音））	明朝体、ゴシック体、他マルチフォント対応	42.5 ～ 236 ピクセル角程度（500万画素カメラでカード全体がちょうど収まるサイズの場合、1.8 ～ 10 mm 角程度）		Windows 11/10 Windows Server 2022/2019/2016	ー		2023/6	要問合せ	
マイナンバーカードOCRライブラリ v.2.0 for Linux	NTTデータNJK(メディアドライブ)	漢字(JIS 第1 水準2965 文字、第2 水準3390 文字、機種依存文字132 文字)、ひらがな、カタカナ、アルファベット、数字、記号7 文字（－（ ）／・々ー（長音））	明朝体、ゴシック体、他マルチフォント対応	42.5 ～ 236 ピクセル角程度（500万画素カメラでカード全体がちょうど収まるサイズの場合、1.8 ～ 10 mm 角程度）		Red Hat Enterprise Linux 7/8/9 CentOS 7 Rocky Linux 8/9	ー		2023/6	要問合せ	
マイナンバーカードOCRライブラリ v.2.0 for iOS	NTTデータNJK(メディアドライブ)	漢字(JIS 第1 水準2965 文字、第2 水準3390 文字)、ひらがな、カタカナ、アルファベット、数字、記号（－（ ）・々／）	明朝体、ゴシック体、他マルチフォント対応	42.5 ～ 236 ピクセル角程度（500万画素カメラでカード全体がちょうど収まるサイズの場合、1.8 ～ 10 mm 角程度）		iOS 16～17	ー		2023/9	要問合せ	
マイナンバーカードOCRライブラリ v.2.0 for Android	NTTデータNJK(メディアドライブ)	漢字(JIS 第1 水準2965 文字、第2 水準3390 文字)、ひらがな、カタカナ、アルファベット、数字、記号（－（ ）・々／）	明朝体、ゴシック体、他マルチフォント対応	42.5 ～ 236 ピクセル角程度（500万画素カメラでカード全体がちょうど収まるサイズの場合、1.8 ～ 10 mm 角程度）		Android 10～14	ー		2023/9	要問合せ	

表4.2-12 ソフトＯＣＲ製品(本人確認書類　その３)

製品名	メーカ	認識対象 文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	TWAIN 対応	メモリ	発売年月	価格(円)	備考
マイナンバーカード認識ライブラリー　V3.0 Windowsサーバー、PC対応	パナソニックソリューションテクノロジー	約6,800字	－	6～60ポイント(400dpi)		Windows 11 21H2/22H2 Windows 10 22H2 Windows Server 2022 21H2 2019/2016/2012 R2/2012	－	OSが必要とする最低メモリーに加えて256MB以上(512MB以上を推奨)		550,000円 [税抜]	
マイナンバーカード認識ライブラリー　for Linux　V3.0	パナソニックソリューションテクノロジー	約6,800字	－	6～60ポイント(400dpi)		Red Hat Enterprise Linux 7/8 Ubuntu 18.04 LTS /20.04 LTS /22.04 LTS	－	CPU版:OSが必要とする最低メモリーに加えて4GB以上（6GB以上を推奨） GPU版:OSが必要とする最低メモリーに加えて2GB以上（3GB以上を推奨）		550,000円 [税抜]	
マイナンバーカード認識ライブラリー for iOS V2.0	パナソニックソリューションテクノロジー	約6,800字	－	6～60ポイント(400dpi)		iOS 13/14/15	－	OSが必要とする最低メモリーに加えて256MB以上(512MB以上を推奨)		550,000円 [税抜]	
マイナンバーカード認識ライブラリー for Android　V3.1	パナソニックソリューションテクノロジー	約6,800字	－	6～60ポイント(400dpi)		Android 10/11/12	－	OSが必要とする最低メモリーに加えて256MB以上(512MB以上を推奨)		550,000円 [税抜]	
在留カード認識ライブラリー	パナソニックソリューションテクノロジー	－	－	6～60ポイント(400dpi)		Windows/Linux	－			要問合せ	
免許証認識ライブラリー Ver4.0	パナソニックソリューションテクノロジー	約6,800字	マルチフォント(明朝体、ゴシック体、教科書体、ワープロ体、新聞文字など)	6～60ポイント(400dpi)		Windows 11 version 21H2/22H2 Windows 10 version 21H1/21H2 Windows Server 2022 version 21H2/ 2019/2016 / 2012 R2/ 2012	－	OSが必要とする最低メモリーに加えて、256MB以上（512MB以上を推奨）		550,000円 [税抜]	

表4.2-13 ソフトＯＣＲ製品(本人確認書類　その４)

製品名	メーカ	認識対象 文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	TWAIN 対応	メモリ	発売年月	価格(円)	備考
免許証認識ライブラリー Ver3.20 for iOS	パナソニックソリューションテクノロジー	約6,800字	マルチフォント (明朝体、ゴシック体、教科書体、ワープロ体、新聞文字など)	6～60ポイント (400dpi)		iOS 14/15/16	－	OSが必要とする最低メモリーに加えて128MB以上(256MB以上を推奨)	2021/10	550,000円 ［税抜］	
免許証認識ライブラリー Ver3.0 for Android	パナソニックソリューションテクノロジー	約6,800字	マルチフォント (明朝体、ゴシック体、教科書体、ワープロ体、新聞文字など)	6～60ポイント (400dpi)		Android 9/10/11	－	OSが必要とする最低メモリーに加えて128MB以上(256MB以上を推奨)	2021/10	550,000円 ［税抜］	
DynaEye 運転免許証OCR	PFU					Windows 11/10	－	以上 Windows 10 32bit ： 1GB以上, 64bit ： 2GB以上	2015/12	202,000円 ［税別］	A6コンパクトフラットベッドスキャナfi-65F/fi-60Fを使用
DynaEye マイナンバーOCR	PFU					Windows 11/10	－	32bitOS:1GB以上 64bitOS:2GB以上	2015/12	52,000円 ～ ［税別］	対応スキャナ：fi-65F、fi-60F、fi-800R、ScanSnap iX100、ScanSnap S1100 対象カード：通知カード表面、個人番号カード表面/裏面
DynaEye 本人確認カメラOCR V5.0	PFU					Windows 11/10 Android 7.1.1/10.0 iOS 12.0～12.4, iPadOS 13.1～17.0	－		2017/12	要相談	Arrows Tab、Xperia、iPad等、タブレット端末のカメラで読み取った本人確認書類（運転免許証/マイナンバーカード/在留カード）を認識

表4.2-14 ソフトＯＣＲ製品(本人確認書類　その５)

製品名	メーカ	認識対象 文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	TWAIN 対応	メモリ	発売年月	価格(円)	備考
自動車運転免許解析ライブラリ	アイエスピー			サポート画像サイズ　写真：推奨画素数　500万画素（カード部分が300万画素を要求）		サーバー上での画像解析及びスマートフォンアプリ内での解析	ー			要相談	
マイナンバーカード解析ライブラリ	アイエスピー			サポート画像サイズ　写真：推奨画素数　500万画素（カード部分が300万画素を要求）		サーバー上での画像解析及びスマートフォンアプリ内での解析	ー			要相談	
領収書解析ライブラリ	アイエスピー			サポート画像サイズ　写真：推奨画素数　500万画素（カード部分が300万画素を要求）		サーバー上での画像解析及びスマートフォンアプリ内での解析	ー			要相談	
名刺解析ライブラリ	アイエスピー			サポート画像サイズ　写真：推奨画素数 300万画素～		サーバー上での画像解析及びスマートフォンアプリ内での解析	ー			要相談	
在留カードライブラリ	アイエスピー			サポート画像サイズ　写真：推奨画素数 300万画素～		サーバー上での画像解析及びスマートフォンアプリ内での解析	ー			要相談	

表4.2-15 ソフトＯＣＲ製品(マルチタイプ)

製品名	メーカ	認識対象 文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	TWAIN 対応	メモリ	発売年月	価格(円)	備考
本人確認書類読取ソフトウェア	東芝デジタルソリューションズ	約4000	オムニフォント	6～40ポイント	5秒以下/1枚	Windows Server 2012 R2 Windows Server 2016 RedHat Enterprise Linux 8.3	-	1GB以上	2020/10	月額利用料金150万円～（免許証読取5万枚。開発キット、メディア別途）	運転免許証、マイナンバーカード、在留カード、保険証、パスポートが読取対象。顔認証機能あり動作に必要なソフトウェア（JAVA）等あり
AI OCR文字認識サービス	東芝デジタルソリューションズ	約4000	オムニフォント	6～40ポイント	5秒以下/1枚	Red Hat Enterprise Linux 8　64ビット版(x86_64)	-	システム構成による	2021/5/1	要問合せ	
OCR Multi Entry Stage	NTTデータNJK(メディアドライブ)	手書(英数字,カタカナ,ひらがな,漢字4419文字,記号) 活字（英数字,カタカナ,ひらがな,漢字6355文字,記号) バーコード/QRコード	手書(OCR-HN,OCR-HK,OCR-HS,OCR-B,OCR-Kなど) 活字（マルチフォント,IBM407,12F)	手書（4x6mm～12x12mm) 活字（3～15mm)		OCRサーバ：Windows Server 2019/2016 クライアント：Windows 11/10	―	OCRサーバ：8GB以上 クライアント：4GB以上	2019/5	要問合せ	※オプション 活字文書、名刺、運転免許証、健康保険証

表4.2-16 サービスＯＣＲ製品（その１）

製品名	メーカ	認識対象文字種	認識書体	文字サイズ (400DPI)	認識速度(字/秒)	OS	メモリ	発売年月	価格(円)	備考
Cloud OCR API (領収書)	NTTデータNJK (メディアドライブ)	約3700	明朝、ゴシックなどのマルチフォント	5～45ポイント		依存なし(JSON フォーマットインターフェース)	－	2018/11	要相談	領収書OCR Cloud商品 サーバー上での画像解析
Cloud OCR API (免許証)	NTTデータNJK (メディアドライブ)	漢字(JIS 第1 水準 2965 文字、第2水準3390 文字、機種依存文字111 文字)、ひらがな、カタカナ、アルファベット、数字、記号9 文字（－（）［］/ ・ヽー（長音））	明朝体、ゴシック体、他マルチフォント対応	42.5 ～ 236 ピクセル角程度 (500万画素カメラでカード全体がちょうど収まるサイズの場合、1.8 ～ 10 ㎜角程度)		依存なし(JSON フォーマットインターフェース)	－	2018/11	要問合せ	免許証OCR Cloud商品 サーバー上での画像解析
Cloud OCR API (マイナンバーカード)	NTTデータNJK (メディアドライブ)	漢字(JIS 第1 水準 2965 文字、第2水準3390 文字、機種依存文字132 文字)、ひらがな、カタカナ、アルファベット、数字、記号7 文字（－（）／・ヽー（長音））	明朝体、ゴシック体、他マルチフォント対応	42.5 ～ 236 ピクセル角程度 (500万画素カメラでカード全体がちょうど収まるサイズの場合、1.8 ～ 10 ㎜角程度)		依存なし(JSON フォーマットインターフェース)	－	2023/6	要問合せ	マイナンバーカードOCR Cloud商品 サーバー上での画像解析
Cloud OCR API (保険証)	NTTデータNJK (メディアドライブ)	漢字(JIS 第1 水準 2965 文字、第2水準3390 文字)、ひらがな、カタカナ、アルファベット、数字、記号	明朝体、ゴシック体、他マルチフォント対応	42.5 ～ 236 ピクセル角程度 (500万画素カメラでカード全体がちょうど収まるサイズの場合、1.8 ～ 10 ㎜角程度)		依存なし(JSON フォーマットインターフェース)	－	2018/11	要相談	保険証OCR Cloud商品 サーバー上での画像解析
Cloud OCR API (名刺)	NTTデータNJK (メディアドライブ)	約3700+英語+中韓国語	明朝、ゴシックほかマルチフォント	5～36ポイント		依存なし(JSON フォーマットインターフェース)	－	2018/11	要相談	名刺OCR Cloud商品 サーバー上での画像解析
THE 名刺管理 Business	NTTデータNJK (メディアドライブ)	約3700+英語+中韓国語	明朝、ゴシックほかマルチフォント	5～36ポイント		－	－	2018/4	800 ［税別］	Cloud商品

表4.2-17 サービスＯＣＲ製品（その２）

製品名	メーカ	認識対象文字種	認識書体	文字サイズ(400DPI)	認識速度(字/秒)	OS	メモリ	発売年月	価格(円)	備考
クラウドOCRサービス Ver. 2.40 (AI手書き文字)	パナソニックソリューションテクノロジー	・英数字 英字、数字、記号の一部 ・住所 英字、数字、ひらがなの一部、カタカナ、ギリシャ文字の一部、記号の一部、第1水準漢字の一部（2,015字）、第2水準漢字の一部（366字） ・氏名 ひらがなの一部、カタカナの一部、記号の一部、第1水準漢字の一部（2,368字）、第2水準漢字の一部（466字）、第3水準漢字の一部（8字）、JIS規格外漢字（1字） ・カテゴリフリー 英字、数字、ひらがな、カタカナ、ギリシャ文字の一部、記号の一部、第1水準漢字の一部（2,965字）、第2水準漢字の一部（1,210字）、第3水準漢字の一部（24字）、第4水準漢字の一部（2字）、JIS規格外漢字の一部（3字）			－	－	－	2021/7	要問合せ	クラウドOCRサービス
クラウドOCRサービス Ver. 2.40 (名刺)	パナソニックソリューションテクノロジー	約6,800字 英字、数字、ひらがな、カタカナ、JIS記号（一部） 168字、ギリシャ文字（一部） 32字、JIS第1水準漢字、JIS第2水準漢字	マルチフォント(明朝体、ゴシック体、教科書体、ワープロ体、新聞文字など)	6～60ポイント(400dpi)	－	－	－	2021/7	要問合せ	クラウドOCRサービス
クラウドOCRサービス Ver. 2.40 (免許証)	パナソニックソリューションテクノロジー	約6,800字 英字、数字、ひらがな、カタカナ、JIS記号（一部） 168字、ギリシャ文字（一部） 32字、JIS第1水準漢字、JIS第2水準漢字	マルチフォント(明朝体、ゴシック体、教科書体、ワープロ体、新聞文字など)	6～60ポイント(400dpi)	－	－	－	2021/7	初期登録料 100,000 月額利用料 50,000～ [税別]	クラウドOCRサービス
WisOCR for 注文書・請求書	パナソニックソリューションテクノロジー	手書き、活字			－	Windows 11/10	お使いのOSが推奨する環境以上	2021/12	Basic 初期登録料なし 月額基本利用料 300枚 30,000円[税別] Pro 初期登録料 100,000円[税別] 月額 2500枚 100,000円[税別]	PC上で動作するクライアントアプリとクラウドのAI-OCRエンジンで構成

表4.2-18 サービスOCR製品（その3）

製品名	メーカ	認識対象文字種	認識書体	文字サイズ(400DPI)	認識速度(字/秒)	OS	メモリ	発売年月	価格(円)	備考
RICOH 受領請求書サービス	リコー	漢字、ひらがな、英大文字、英小文字、数字、カタカナ、記号、手書き	－	－	－	－	－	－	ライトコース ベーシックコース BPOも選択可 要問合せ	
RICOH 受領納品書サービス	リコー	漢字、ひらがな、英大文字、英小文字、数字、カタカナ、記号、手書き	－	－	－	－	－	－	ライトコース ベーシックコース BPOも選択可 要問合せ	
PFU Smart Capture Service	PFU Limited	活字、手書き	－	－	－	－	－	2018/4		アップロードされた文書画像の種類を自動的に仕分け、「特定帳票OCR」「汎用帳票OCR」「日本語手書きOCR」を自動選択して呼び出すクラウドサービス。
AI Read on Cloud	アライズイノベーション	活字、手書き	－	－	－	－	－	－	要問合せ	事例：注文書、請求書、決算書、アンケート用紙、チェーンストア共通伝票、特許関連書類 オンプレミスもあり
Flax Scanner ぜんぶもむもん	cinnamon	活字、手書き	－	－	－	－	－	－	要相談	
Flax Scanner HUB	cinnamon	対象言語：日本語、英語	－	－	－	－	－	－	要相談	2024年4月より提供開始予定

表4.2-19 サービスOCR製品（その4）

製品名	メーカ	認識対象文字種	認識書体	文字サイズ(400DPI)	認識速度(字/秒)	OS	メモリ	発売年月	価格(円)	備考
AI OCR文字認識サービス	東芝デジタルソリューションズ	約4000	オムニフォント	6～40ポイント	5秒以下/1枚	－	－	2019/1	月額利用料金30万円～（初期構築別途）	請求書、受発注伝票等の帳票を読取対象としたクラウドサービス提供
マイナンバー収集用カード読取サービス	東芝デジタルソリューションズ	約4000	オムニフォント	6～40ポイント	4秒以下/1枚	－	－	2015/12	月額利用料金45万円～（初期構築別途）	マイナンバーカードを読取対象としたクラウドサービス提供
本人確認書類読取サービス	東芝デジタルソリューションズ	約4000	オムニフォント	6～40ポイント	4秒以下/1枚	－	－	2015/12	月額利用料金45万円～（初期構築別途）	運転免許証、マイナンバーカード、在留カード、保険証、パスポートを読取対象としたクラウドサービス提供
SmartRead	コージェントラボ	活字、手書き	－	－	－	－	－	－	スモールプラン　36万円/年 1.2万枚前後 スタンダードプラン　96万円/年 6万枚前後 エンタープライズプラン　240万円/年 26万枚前後 オンプレミスプラン すべて税抜き価格	
Tegaki	コージェントラボ	手書き	－	－	－	－	－	－	要問合せ	事例：　各種申込書類やアンケートをはじめ医療機関での問診票など
スマートOCR	インフォディオ	-	－	－	－	－	－	－	要問合せ	事例　帳票種類：決算書、健康診断書、健康保険書、勤務管理表　等 オンプレミスもあり
DenHo（デンホー）	インフォディオ	-	－	－	－	－	－	－	要問合せ	電子帳票保存法の対応

表4.2-20 サービスOCR製品（その5）

製品名	メーカ	認識対象文字種	認識書体	文字サイズ (400DPI)	認識速度(字/秒)	OS	メモリ	発売年月	価格(円)	備考
LINE WORKS OCR	Softbank	-	-	-	-	-	-	-	Template 月額　55,000円(税込) General 月額　55,000円(税込) 特化型（請求書） 月額　88,000円(税込)	事例：伝票登録OCR、領収書OCR、身分証明書OCR 詳しい利用条件は要問合せ
DX Suite	AI Inside	活字、手書き	-	-	-	-	-		Lite［税別］ 初期費用 0円 月額　30,000円～ 無料枠　18,000円分 Standard［税別］ 初期費用 200,000円 月額　100,000円～ 無料枠　50,00円分 Pro［税別］ 初期費用 200,000円 月額　200,000円～ 無料枠　200,000円分	Intelligent OCR Elastic Sorter Multi Form
AI よみと～る	NTT東日本	-	-	-	-	-	-		プラン1（小型） 初期費用なし 通常版　33,000円/月(税込)読み取り箇所　月6千まで プラン2（中型） 初期費用なし 通常版　110,000円/月(税込)読み取り箇所　月6万まで プラン3（大型） 初期費用なし 通常版　220,000円/月(税込)読み取り箇所　月20万まで	プラン3（大型）に自動帳票仕分けオプションあり
AnyForm OCR	Hammock	-	-	-	-	-	-	2019/7/3	クラウド版 1,200,000円(年)～ ※製品構成により価格が異なります。	オンプレ版もあり
バクラク経費精算	LayerX	-	-	-	-	-	-		要問合せ 月額2万円～（税抜き）	契約は年間契約 領収書・請求書

4.3　ペン入力文字認識システム

ペン入力文字認識とは、専用のペン（または指）とタブレット装置（または専用の用紙）を用いて手書き筆跡をコンピューターが読み取り、書かれた文字を認識する手法・技術である。スキャナーやカメラで撮影した文字画像を認識する OCR に対して、筆跡の座標系列を認識するもので、オンライン文字認識と呼ばれることもある。本稿では簡単のため「ペン入力」と記すが、指によるタッチ入力も含める。

近年、スマートフォンやタブレット型コンピューターの普及によってペン入力文字認識はコモディティ化（汎用品化）しており、製品の標準機能に組み込まれていることが多い。各社のスマートフォンとペン入力技術の動向を以下に記述する。

4.3.1　スマートフォン／タブレットの製品動向

・GALAXY シリーズ

Samsung は専用ペン（S ペン）を使う手書き入力を前面に出した GALAXY Note を以前から発売している。その最新の後継機種が 2023 年 4 月に発表された「Galaxy S23 Ultra」[1]である。手書き文字認識機能が標準搭載されており、また、「Galaxy Notes アプリ」によって手軽なメモ機能が実現されている。2020 年に発売された「Galaxy Note20 Ultra」以降、デバイス製品名に Galaxy Note という名称は使わなくなっている。2024 年 2 月には S23 Ultra の廉価版「Galaxy S23 FE」が発売[2]され、「Galaxy S24 Ultra」も 2024 年夏頃に発売される予定である。その他に画面の折り畳みが特長の「Galaxy Z Fold4」[3]でも S ペンが搭載されるように、ペン入力機能は Galaxy シリーズでは標準機能となっている。Galaxy の特徴である S ペンはワコム製の電磁誘導式のペン入力機能であり、「Samsung Note」等の専用アプリを用いて簡単に手書きメモが取れる操作性が売りである。

・Arrows シリーズ

富士通の Arrows シリーズにはスマートフォンとタブレット端末がある。スマートフォンには独自開発の手書き入力機能が搭載されている。ATOK ソフトキーボード上で筆跡入力が可能なシームレス入力に加え、一つの文字枠内に文字を重ね書きできる機能も搭載されている。2023 年モデルの Arrows N F-51C 等にも踏襲されているが、スペック等で特筆されているわけではなく、2020 年頃より変更はされていないようである[4]。タブレットの最新機種は Arrows Tab F-02K である。2018 年のモデルだがそれ以降の更新は行われていない[5]。

・HUAWEI タブレット

ファーウェイ・ジャパンは付属の専用ペン「M-Pen」による筆記が可能なタブレットを継続的に発売している。「MediaPad M3」（2017 年 8 月）の発売から、「MediaPad T5」（2019 年 8 月）、「MatePad」（2020 年 6 月）など様々な製品を発売しており、「MatePad 11」（2021

年 7 月）が最新機種である。HUAWEI M-Pencil（第 2 世代）を採用。圧力・傾斜を 4096 レベルで感知し、約 2 ミリ秒の遅延を実現している[6, 7]。

2022 年 6 月に、E-Ink 社の電子ペーパーを採用した「HUAWEI MetaPad Paper」が発売された。独自の「HarmonyOS 2」を搭載し、独自の電子書籍ストアアプリ「HUAWEI Books」も導入された電子書籍端末だが、スタイラスペンで手書きが可能な点が特徴である。電子書籍への手書き機能は充実していないとの報告があるが、手書きメモ機能としては十分に使え、高精度な手書き文字認識機能も搭載されているようである。なお、ペン入力機能は第 2 世代の M-Pencil を採用している[8]。

・QUADERNO（電子ペーパー）

富士通は 2021 年から電子ペーパー端末「QUADERNO」を発売している[9][10]。E-Ink 社の電子ペーパーを採用し、10.3 インチと 13.3 インチの 2 機種が存在する。Wacom 社製の電磁誘導スタイラスペンと、静電誘導式のタッチ入力の両方をサポートしている。

4.3.2 スマートフォン／タブレット向け手書きソフトウェア

・OneNote[11, 12, 13]

OneNote は、マイクロソフト社が提供する無料の Windows／Mac／iOS／Android／Web アプリケーション向けのデジタルノートアプリケーションである。オフライン手書き文字認識機能があり、ペン入力した手書き文字や、カメラで撮影した手書き文字などを認識する。OneNote 各 OS 用のアプリストアから無料でダウンロードできる。Windows 版は 32bit 版が提供されており、Windows11 及び Microsoft 365 に同梱されている。

・Neo Studio[14]

Android、iOS のスマートデバイス用アプリケーション Noe Studio は専用の微細なコードが印刷された紙のノートと専用のネオスマートペン（M1、N2、Dimo の 3 タイプ）を使うことで、手書きのリプレイやテキストへの変換などができるアプリである。テキスト変換は、英語、日本語、中国語等 15 か国の言語をサポートしている。

・MyScript Nebo[15, 16]

MyScript Nebo は、iOS、Windows、Android 上で使えるノートアプリである。基本機能は無料で使用できるが、ノート数が 5 ページまでと制限される。無制限とするには有料プランが必要である。手書き変換機能があり、リアルタイムに手書き文字が文字認識されていく。書いた後に文字をダブルタップするとテキストに変換する機能や、PDF ファイルをインポートして注釈を付ける機能もある。2023 年 12 月に最新の Nebo 5.8 がリリースされた。

・mazec[17]

サードパーティー製のソフトキーボードには「mazec」がある。Windows、Android、iOS

に対応しており、個人向けアプリに加えて法人向けの mazec for Business や業種向け（医療、建設等）、SDK などもリリースされている。最新辞書を同期できるクラウドサービス mazec Plus が iOS 向けに存在する。

4.3.3 ペン入力デバイス

スマートフォンやタブレット筆記デバイスは指とペンの併用が可能な静電誘導式タブレットが使われることが多いが、筆跡データ入力を重視した用途が広がるにつれ、専用ペンを搭載するケースも次第に増えている。

近年の製品に使われている専用ペンは主に 4 種類の方式（プロトコル）が使われている[18]。Wacom には EMR 方式と AES 方式があり、Microsoft は MPP（Microsoft Pen Protocol）、Apple も独自方式を採用している。ASUS 等が採用していた Synaptic 方式は、最近はあまり見ないようである。

各プロトコルの概要を下記に記す。どの方式が良いかについてはいくつかのレビュー記事が存在するが、概ねプロ用（イラスト等）の書き味としては Wacom EMR と Apple Pencil を推す声が多く、筆圧や傾き検出などの機能では Wacom EMR と MPP が優っているという声が多い。

・Wacom EMR 方式[19]

ワコムが特許を取得している電磁誘導方式（EMR: Electro-Magnetic Resonance）であり、同社の MobileStudio Pro 等で使われているプロ仕様の規格である。Wacom Feel IT Technologies という名称で呼ばれることもあるが、これは後述の Wacom AES 方式でも使われることがあり、呼称や規格名は必ずしも統一されていない。EMR 方式の中にも 4 種類の規格があるなど、互換性については個別に確認する必要がある[20]。

・Wacom AES 方式[21, 22]

ワコムのアクティブ静電結合方式（AES: Active Electrostatic）である。他社 PC に採用されているのは多くがこの方式であり、公式の Wacom「Bamboo Ink」だけでなく、互換性のあるペンが他社から発売されている。

・Microsoft MPP 方式

マイクロソフトの Surface Pro 3 で採用された「Surface Pen」が用いている方式である。マイクロソフトは Surface Pro 2 まではワコムの AES 方式を用いていたが、イスラエルの N-Trig 社を買収し、その技術を導入した方式を MPP（Microsoft Pen Protocol）と称している。

・Apple 方式

Apple Pencil が採用しているアップル社独自の方式である。

続いて、専用ペンの主な製品を記す。Wacom EMR や Apple Pen はメーカーが固定されるため上記の方式と分類が重なる部分が多い。両方の説明を併せて参考にして欲しい。

・Surface Pen

Microsoft Surface Pro 3 で採用された「Surface Pen」は、前身の Surface Pro 2 で「Pro Pen」と呼ばれていた専用ペン（Wacom 製）のバージョンアップ版であり、電池を用い Bluetooth 機能を内蔵している（イスラエルの N-Trig 社製）[23, 24]。2017 年 8 月発売の「Surface Pro」（Surface Pro 4 の次機種なので「Pro 5」と呼ばれる）は、筆圧検知が 1024 段階から 4096 段階に拡張され、傾き検知にも対応した新「Surface Pen」が採用された[25]。Surface Pen は PC 本体とは別売りであり、同じ PC で旧ペンも新ペンも使える[26]。

Surface Pen は使用する通信プロトコル MPP（Microsoft Pen Protocol）のバージョンが PC と一致している必要がある。例えば、最新機種である Surface Pro 9 は「Surface スリム ペン 2 専用」と仕様に記されている。スリムペン 2 は MMP v2.6 であり、4,096 段階の筆圧検知、傾き検知などの機能をサポート対応している[27]。MPP のペンはいわゆる静電容量方式で、ペン側で電磁波を発生させることで、パネル側がペンの位置を検知する。このためペンの側には Apple Pencil と同じようにバッテリーが必要になるが、ペン先に電磁波を発生させるだけでいいので、小さな電池（単 6 形）で半年といった長期間利用ができる[28]。

・Apple Pencil

Apple Pencil は iPad Pro の入力デバイスとして 2015 年 11 月に発売され、2018 年 10 月に第二世代が発表された。2018 年以降の iPad や iPad mini では第一世代 Apple Pencil が使えるが、第二世代 Apple Pencil は iPad Pro のみ対応である。第二世代の Apple Pencil は形状や充電方法などの使い勝手が第一世代より改善されている[29]。[29]は 2021 年の記事だが、2024 年 2 月現在、Apple Pencil 第三世代は発表・発売されていない。しかし、2024 年秋の iPad Pro 新版と共に第三世代が発売されるのでは？と噂されている[30]。

・Wacom Pro Pen 2

ワコムは、2016 年 11 月に発売した「Wacom MobileStudio Pro」において、筆圧 8192 段階の「Wacom Pro Pen2」を採用した[31]。これは 2017 年 1 月発売のタブレット「Intuos Pro」を始め、多くのワコムタブレットに採用されている。2019 年 2 月には同じ性能でスリムタイプの「Wacom Pro Pen slim」も発売された[32, 33]。2023 年 5 月には更に滑らかな書き味を追求した「Wacom Pro Pen 3」が発売されたが、過去のモデルとの互換性は無い[34]。

・GALAXY Note「S ペン」[1, 2]

GALAXY Note には「S ペン」という電磁誘導式の専用ペンが添付されている。これはワコムの技術を採用しており、筆圧も感知できるため自然な筆跡の入力（太さの変化まで表現）が

可能である。S ペンは Galaxy シリーズの強力な特長となっており、新機種が発表されるごとに改善が加えられている。

4.3.4 その他の製品・サービス

・SkyCom 社 SkyPDF Touch Ink for win 7 [35]

同社製品「SkyPDF Professional 7」（PDF 作成・編集・加工を行う）のオプション製品であり、Windows タブレット上に表示した PDF に直接ペンで手書き文字入力するアプリケーションである。紙に文字を書くような自然な書き味を再現し、手書きした文字は手書き文字認識機能によりテキストデータに変換し、用途に合わせて置換や埋込が簡単に行える。また、手書き筆跡に電子署名技術を融合させ、証跡情報を暗号化して埋め込むことで、ペーパーレスを推進するソリューションも提供している。iOS 用（iPad 専用）の製品として「SkyPDF Sign for iPad」[36]がある。こちらは帳票への署名入力が主な目的の製品である。

4.3.5 主な日本語オンライン文字認識エンジン

・iLabo 手書き文字認識エンジン[37, 38]

東京農工大学中川研究室が開発したオンライン手書き文字認識技術を事業化するために設立された大学発ベンチャー「アイラボ株式会社」が販売している認識エンジンである。中川研究室のオンラインマッチング（高速な非線形伸縮マッチング）とオフライン文字認識（OCR）を統合したハイブリッド型文字認識を用い、CRF 等を用いた学習技術も含む。MetaMoji 社の mazec や 7note 等に採用されている。2023 年 4 月にはベトナム語認識機能も加わった。

・MyScript[39]

多言語に対応したオンライン文字認識エンジンを展開しているフランスの会社である。従来は MyScript 社が Vision Object というブランドでソフトを開発していたが、現在は会社名もブランドもすべて MyScript に統一されている。

・Ink シリーズ（ポトス株式会社）

ポトス株式会社は 1997 年に設立された、ペン入力インターフェース技術を中心にライブラリーやソリューション開発を行っている企業である[40]。InkTool（ペン入力ソフト開発ツール）や InkFep（日本語手書き文字認識ツール）などが様々な企業で利用されている。例えば、富士通は自社ソリューションにおいて手書き文字入力機能の実現をサポートする開発ライブラリーとして InkTool 及び InkFep を活用している[41]。ポトス株式会社の製品は、文字認識機能そのものは東京農工大学（中川研究室）の技術を利用している。

【参考文献】

[1] 「Galaxy S23 Ultra」、『撮る』『観る』『描く』で楽しめる史上最強の Galaxy
https://k-tai.watch.impress.co.jp/docs/column/mobile_catchup/1499855.html

[2] サムスンが「Galaxy S23 FE」を発表、au から 2 月 9 日発売
https://k-tai.watch.impress.co.jp/docs/news/1565032.html

[3] Galaxy Z Fold4 公式ページ
https://www.samsung.com/jp/smartphones/galaxy-z-fold4/

[4] arrows N F-51C 取扱説明書
https://www.docomo.ne.jp/binary/pdf/support/manual/F-51C_J_syousai_13.pdf

[5] Arrows Tab F-02K
https://www.fmworld.net/product/phone/f-02k/

[6] HUAWEI MatePad 11 公式ページ
https://consumer.huawei.com/jp/tablets/matepad-11/

[7] HUAWEI M-Pencil（第 2 世代）
https://consumer.huawei.com/jp/accessories/m-pencil-2nd-generation/

[8] 「HUAWEI MatePad Paper」の使い勝手はどう？　最新 E Ink 搭載タブレットを試す
https://www.itmedia.co.jp/pcuser/articles/2206/29/news060.html

[9] QUADERNO 公式ページ
https://www.fmworld.net/digital-paper/top.html

[10] この感覚が欲しかった！レビュー後に結局買った手書き電子ペーパー「クアデルノ」
https://xtech.nikkei.com/atcl/nxt/column/18/01736/080400006/

[11] OneNote デジタルのノートブック
https://www.microsoft.com/ja-jp/microsoft-365/onenote/digital-note-taking-app

[12] 無料で使えるマイクロソフトの OneNote とは？ https://allabout.co.jp/gm/gc/453716/

[13] Windows 版「OneNote」アプリが Microsoft Store に登場
https://forest.watch.impress.co.jp/docs/news/1450976.html

[14] Neo Studio　https://neosmartpen.jp/

[15] Nebo 公式ページ　https://www.nebo.app/ja/

[16] MyScript Nebo とは？使い方や価格・評判まで解説
https://www.stock-app.info/media/myscript-nebo/

[17] mazec 公式ページ　http://mazec.jp/

[18] お絵かきペンは“Wacom＞N-trig＞Synaptic”だと思う！―デジタイザーと Windows タブについて想うコト　https://mupon.net/digitizer-protocol-best

[19] ワコムのテクノロジー Electro-magnetic Resonance
https://www.wacom.com/ja-jp/for-business/technologies/emr
[20] ワコムの電磁誘導ペン（EMR）は現在 4 種類もある上に全部互換性がなくてめちゃ分かりづらいのでまとめておく
https://29udon.com/6143.html#Wacom_Feel_IT_TechnologiesEMR
[21] ワコムのテクノロジー Active Electrostatic
https://www.wacom.com/ja-jp/for-business/technologies/aes
[22] Wacom AES 対応デジタイザーペンの“最強”はどれだ？
https://mupon.net/wacom-aes-no1-pen/
[23] Microsoft、マルチタッチスクリーン技術の N-trig に出資
http://www.itmedia.co.jp/news/articles/0901/13/news029.html
[24] Surface Pro 3 と Pro 2 のスタイラスペン比較（N-Trig vs Wacom）
http://tabkul.com/?p=64938
[25] 新しい「Surface Pro」は「4」よりも完成度が高まった！
https://kakakumag.com/pc-smartphone/?id=10536
[26] ［新 surface pen 描き心地レビュー］絵・イラストの描き味は旧ペンとどう変わった？
https://maekoart.net/new-surface-pen
[27] 別の PC で Surface ペンを Windows する
https://support.microsoft.com/ja-jp/surface/7e1861d0-d6fa-4ba5-a9e3-fe210806b211
[28] 同じ 10 型級の Surface Go と iPad Pro はどちらが使いやすいのか?多方面から実機で検証
https://pc.watch.impress.co.jp/docs/column/ubiq/1140203.html
[29] [比較] Apple Pencil の第 1 世代・第 2 世代のどちらを買えばいいの?
https://yossense.com/comparing-apple-pencils/
[30] 【最新情報まとめ】Apple Pencil 第 3 世代 発売日・スペック
https://motifyublog.com/new-apple-pencil-spec/
[31] ワコム、思い通りの制作フローを可能にする Wacom Intuos Pro を発売
http://www.wacom.com/ja-jp/about-wacom/news-and-events/2017/1213
[32] 「ペンが走ってお絵描きが楽しい」ワコムの液タブ・ペンタブ用のスリムペン「Wacom Pro Pen slim」レビュー https://gigazine.net/news/20190227-wacom-cintiq-pro-pen-slim/
[33] くらべてみました！Wacom Pro Pen
https://tablet.wacom.co.jp/article/choice-wacompropen
[34] Wacom Pro Pen 3 公式
https://estore.wacom.jp/ja-JP/products/accessories/acp50000dz.html

[35] Windows タブレットを使って PDF に手書き文字入力 / 電子サイン「SkyPDF Touch Ink for win 7」 https://www.skycom.jp/product/skypdf/touchink_for_win_7/

[36] iPad で PDF ファイルに手書き文字入力／電子サイン「SkyPDF Sign for iPad」 https://www.skycom.jp/product/skypdf/sign_for_ipad/

[37] 日本語文字列認識エンジン https://ilabo.biz/ilabo_japanese_engine/

[38] 世界最高精度の手書き文字認識技術を実用化! https://www.jst.go.jp/pr/jst-news/backnumber/2012/201205/pdf/2012_05_p12.pdf

[39] AI、ニューラルネットワーク、手書き認識 https://www.myscript.com/ja/ai/

[40] ポトス株式会社 http://pothos.to/

[41] セキュリティ Solution ラインナップ（その他） https://www.fujitsu.com/jp/solutions/business-technology/security/secure/lineup/solutions-2/

（URL 確認 2024.2.3）

5. 今後の展望

認識形入力方式は、我が国で急務となっているデジタル化を推進するための重要な基盤技術であり、今後も急速な勢いで進化し続けるものと予想される。その動きは生成AIに代表される新しい技術（新潮流）の進展にも関連し、人々の業務や暮らしの質そのものを変革する可能性を秘めている。技術進化に伴いその活躍の場が拡大することで、利用環境及び利用者が多様化し、環境に起因する様々な外乱の影響を受けやすくなっている。認識形入力方式を搭載した機器を様々な利用者が活用できるようにするためには、外乱によるシステムの性能低下を事前に予測して対策する必要がある。そのために我々は外乱要因とその影響を整理し、検討の経過を年次報告書にて公開してきた。今回、その成果として要因表及びその活用方法をガイドラインとして公開することができた。このガイドラインが認識形入力方式を搭載したシステムの技術者や利用者に気付きを与え、その利活用の促進に貢献することを期待する。

既に実用化が進んでいる整備された環境におけるOCR 装置については、引き続き最新情報のアップデートを行い、我が国におけるこの分野の信頼できる情報発信拠点として公開を行っていく。また、関連する規格について各規格の利用状況を考慮しつつ、見直し等を進める。

DL等の技術革新を背景とした認識形入力方式の最新技術動向や、これを搭載したシステム及びサービスについても動向調査を継続すると共に、最新技術の実用化に向けた課題、及び急速に進化する技術と拡大する市場の要求レベルを調和させる各種規格化について議論を深めていく予定である。

認識形入力方式に関する調査研究報告書

発　行　月　　2024 年 3 月
編集・発行　　一般社団法人 電子情報技術産業協会
認識形入力方式標準化専門委員会
〒100-0004
東京都千代田区大手町 1-1-3 大手センタービル
TEL (03)5218-1050(代表)