

LINE WORKS、文書解析と認識に関する国際会議 「ICDAR2024」にて新技術「CREPE」に関する論文が採択。

従来のOCRでは難しかった課題を解決。

End-to-Endドキュメント解析モデルを活用し、OCRを使用せず、
テキスト情報と座標情報を高精度に同時抽出が可能に。

ビジネス現場のコミュニケーションツール「LINE WORKS」を提供するLINE WORKS株式会社(本社:東京都渋谷区、代表取締役社長:島岡 岳史)は、文書解析と認識に関する国際会議「ICDAR2024」にて、従来のOCRモデルを使わずに、単一モデル(End-to-End)で高精度に画像ドキュメントからテキスト情報及び座標情報を同時に抽出する新技術「CREPE」を開発した論文が採択されたことをお知らせいたします。

「ICDAR」(International Conference on Document Analysis and Recognition)は、文書解析と認識の分野でも権威のある国際会議の一つで、世界トップレベルの研究者や開発者が最新の技術や研究成果の発表、交流をする権威のある場です。今回採択された論文は2024年8月30日～9月4日にかけて開催される「ICDAR2024」(ギリシャ・アテネ)にて発表されました。

本論文はLINE WORKS株式会社 リサーチエンジニアのYoungmin Baek、中尾亮太らによる研究成果です。LINE WORKS株式会社は、AIの研究開発に積極的に取り組んでおりますが、今回の「ICDAR2024」での論文採択は、画像ドキュメント解析性能の向上により、文書やテキスト資料のデジタル化が進みにくい業界・業種のペーパーレスやデジタル化に期待が高まる有望な成果となりました。

▼論文の詳細については下記をご参照ください。

Y. Okamoto, Y. Baek, G. Kim, R. Nakao, D. Kim, M. Yim, S. Park, B. Lee "CREPE: Coordinate-Aware End-to-End Document Parser"

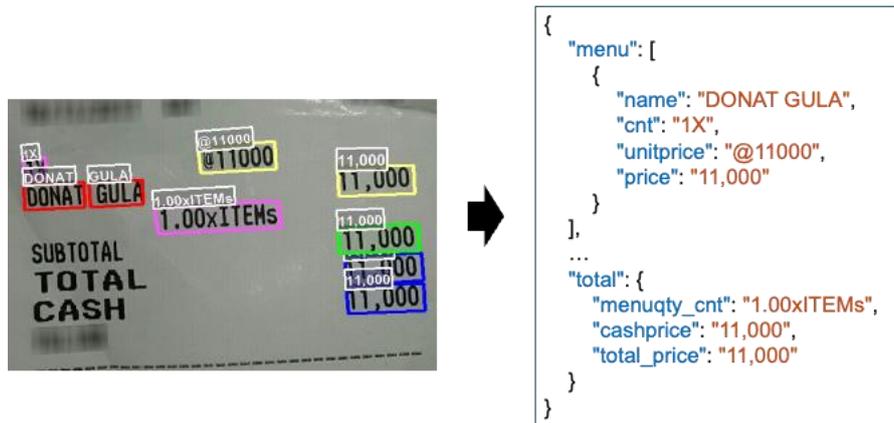
<https://arxiv.org/abs/2405.00260>

■ 論文の概要

従来のOCR依存型モデルを使用せず、単一モデルでの処理を実現。OCR依存型モデルの課題であった複雑な処理や認識精度の低下などを解消し、効率的で高精度な画像ドキュメント処理が可能に。

本論文の「CREPE」は、広く利用されている従来のOCR依存型モデルに起因した性能劣化問題を解消するもので、単一モデル(End-to-End)で画像ドキュメントからテキスト情報と座標情報を同時に抽出し、高精度で認識・解析を行う新技術として提案しました。

図1:「CREPE」の画像ドキュメント解析の一例
(レシート画像(左)から有用な情報を抽出した結果(右))

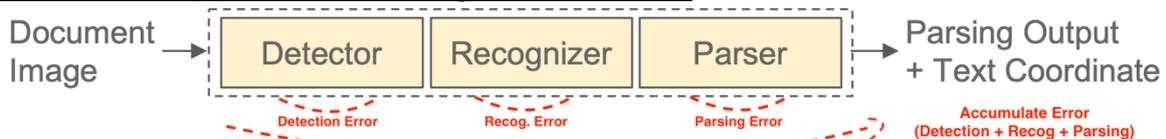


■新技術開発の背景<従来のOCR依存型モデルが抱える課題>

従来の画像ドキュメントの解析には、画像からテキストを抽出する技術 OCR(Optical Character Recognition) が広く活用されてきました。しかし、OCRを使用すると、さまざまなモデル(検出、認識、解析モデルなど)を組み合わせたことが必要で、別々の工程が段階的に処理されるため、システムの複雑度が増し、各ステップでの誤差も蓄積されるため、結果として文書の解析精度が低下するという課題がありました。

図2:OCR依存型アプローチの課題

End-to-end approach including OCR modules



<図2の説明>

前段モジュールで発生した誤差の影響を各モジュールでも受けてしまうため、誤差の累積により性能劣化につながります。

この課題を解決するため、これまででもコンピュータビジョン分野における世界三大国際会議「ECCV2022」で採択された論文^{※1}において、単一モデルで情報抽出を行うEnd-to-Endモデルを提案してきました。End-to-Endは、画像全体を入力として取り込み、直接的に解析結果を生成するアプローチを取るため、複雑なシステムは不要で、かつ認識や解析の誤差の蓄積を防ぐことが可能となります。

※1: 論文詳細「G. Kim, et al. "Ocr-free document understanding transformer", ECCV 2022」

LINE WORKS

■論文の詳細<新技術「CREPE」でEnd-to-Endモデルの課題を解決>

End-to-Endモデルは、複雑なシステムが不要かつ認識や解析の誤差の蓄積を防ぐことができる一方で、OCRを使用しないため、重要な情報である”テキストの座標”を抽出できないという新たな課題がありました。

そこで、End-to-Endモデルの利点を活かしつつ、この課題を解決する新技術「CREPE」を開発しました。「CREPE」は、テキスト情報だけでなくその座標情報も同時に抽出することが可能で、特にドキュメントのレイアウトやフォーマットに依存するタスクでの性能が向上しました。さらに弱教師あり学習^{※2}という手法を採用し、少ないデータで学習できるように設計されています。これにより、従来のEnd-to-Endモデルより柔軟性が向上し、レイアウト解析など座標情報を必要とするタスクへの適用も期待できるようになりました。

※2: 弱教師あり学習は、AIモデルのトレーニングにおいて、不完全なラベル情報を活用して効率的に学習する手法です。ラベル付けコストを抑えつつ、高精度なモデルの構築が可能です。

また、LINE WORKS株式会社が提供する「LINE WORKS OCR (AI-OCRサービス)^{※3}」に「CREPE」を搭載することで、さまざまなドメインに特化したさらに高精度な解析モデルへ適応できる技術的に高い成果となっています。

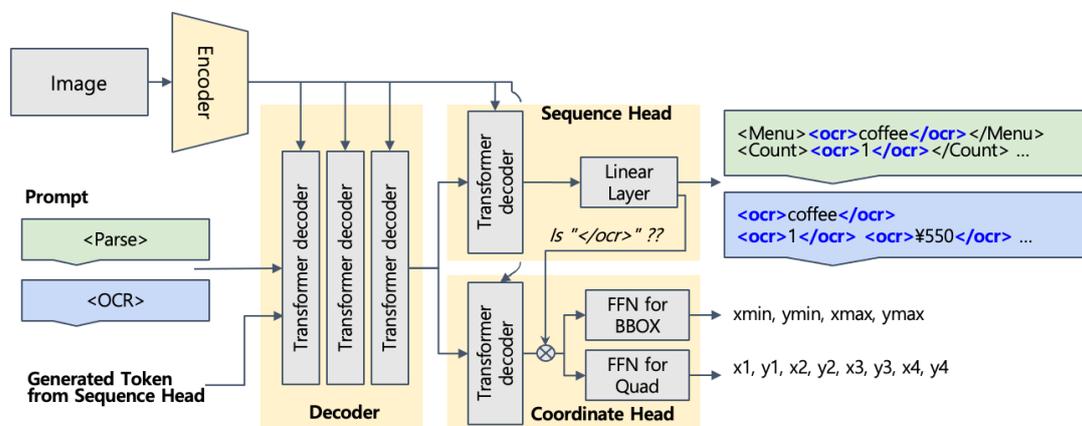
※3: LINE WORKS OCR 公式サイト <https://line-works.com/ai-product/ocr/>

あらゆる文書や画像・PDFの文字情報を読み取り素早くデータ化するAI-OCR(文字認識)サービス。世界的なコンペティションで計6分野で世界No.1の認識率を獲得(2022年9月28日時点の実績)

<「CREPE」の手法>

- ①SpecialTokenを導入することでSequenceの中で、画像内のテキストを単語単位で抽出(例:<ocr>text</ocr>)
- ②Decoder最終層とLM HeadをSequenceHeadとCoordinateHeadに分離することで、テキストだけではなく座標も推論
- ③CoordinateHeadは単語の終わりを意味する</ocr>トークンが出る場合にアクティベーションされるので、単語ごとの位置を獲得

図3:「CREPE」の概要図



<図3の説明>

提案手法はImageEncoderとTextDecoderで構成され、TextDecoderは情報抽出を担当しSequenceHeadで単語ごとのテキスト認識を担当。CoordinateHeadは単語の終わりを表す</ocr>という特殊トークンをトリガーに単語の座標を出力。

LINE WORKS

■LINE WORKS株式会社が提供するAIサービス、研究開発について

LINE WORKS株式会社では、ビジネス現場のコミュニケーションツール「LINE WORKS」に加え、LINE WORKS AI製品として「LINE WORKS AiCall(電話対応AIサービス)」や「LINE WORKS OCR(AI-OCRサービス)」などの高度なAI技術を活用したサービスを提供しています。

さらに、AI技術そのものの研究開発にも注力しており、今後は、今回の論文で提案した手法をさらに発展させると共に、プロダクトへの適用や新たな機能・サービスの創出に努めてまいります。

<国際会議やコンペティションでの主な実績>

【画像・文字認識】計6分野で世界No. 1を獲得

- ・ICDAR(文書解析・認識に関する国際会議)
- ・ECVV(コンピューター分野における国際会議)

【音声認識】権威ある世界最大規模の国際学会で論文採択

- ・ICASSP(音声・音響信号処理における国際学会)
- ・INTERSPEECH(音声・音響信号処理における世界最大規模の国際学会)

<代表的な研究・開発>

- ・テーブル認識の研究(TRACE)
- ・書類偽造検出の研究(Forgery Detection)
- ・AIに読書させる研究
- ・LLMによる文書画像解析(CREAM)
- ・効率的なE2Eモデル(GREPE)
- ・学習していない単語を認識させやすくする研究(InterBiasing)
- ・音声認識モデルにLLMの知識蒸留を行う研究
- ・数式認識
など

■会社概要

社名:LINE WORKS株式会社

本社:東京都渋谷区神宮前1-5-8 神宮前タワービルディング11F

設立:2015年6月

代表者:島岡 岳史

資本金:55億2,000万円

URL: <https://line-works.com/>

※記載の会社名、製品名は、それぞれ会社の商標または登録商標です。

※本プレスリリース記載の情報は発表日現在の情報です。予告なしに変更されることがありますので、予めご了承ください。