

## 補助事業成果報告書

補助事業の名称	視覚障害者・ディスレクシアのための音声を使った読書方法の研究開発
補助事業の概要	以下に記す

### 【研究開発の実施内容と成果】

#### 1. 本年度研究の概要

YourEyes (<https://youreyes.jp>) は、老眼や弱視を含む広義の視覚障害者、さらには識字障害の一種であるディスレクシア症例を持つ人への読書支援サービスである。2021年12月にプレスリリースが株式会社ポニーキャニオンより行われ、2022年4月に一般に公開された。その後、2022年、ポニーキャニオンから株式会社スプリュームへ運営会社の変更があったが、弊社想隆社は引き続き開発チームとして本プロジェクトのコア技術の研究開発に関わっている。

本サービスを採用された図書館・モニター図書館などで視覚障害者に利用してもらい、その使用感についてアンケートをとったところ、スマホで本を1ページずつ画面に捉えて撮影するのが難しいという声があった。本サービスは、スマホ上部にとりつけて、撮影できる『YourEyesボックス』という箱状のアダプタ（図1）を販売しているが1ページごとに撮影することを前提としており、見開きで撮影できたほうがよいという声が少なからず寄せられた。



図1 YourEyesBox

そこで、見開きで撮影することで、たとえばノンプルの無いページがあっても、それが見開きの片方であればもう片方のページからページ番号を推定できるといった副次的な効果も得られる。令和4年度は、昨年得られたこのようなユーザからの問題点を解決するべく、『YourEyesボックス』がなくとも見開きで撮影し、OCRの精度を保ちつつ、レイアウト解析も行えることを研究のテーマにした。

具体的には、次の3つのサブテーマに分けて研究を行った。

- ① 見開きのための台の試作
- ② ①の台に対応する事前画像処理の研究
- ③ ②の画像に適用する機械学習を使ったレイアウト解析

- ① は見開きのためであり、②は次の図のように、見開きを本の外側（小口側）で止めると上部が湾曲するため、精度の高い OCR にするためには、これを平面にするといった処理が必要になるからである。
- ② ③については従来以上に複雑なレイアウトの書籍を正しく読み上げるために必要であるためである。

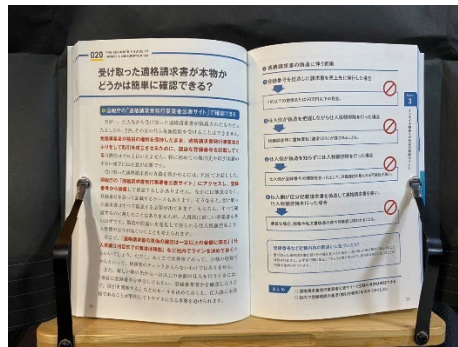


図2 湾曲の補正の必要性

本サービスは書籍を視覚障害者に音声で情報としてその内容を届けることが目的であるが、そのためには、OCRされた書籍のページ（版面データと呼ぶ）をレイアウト解析し、柱とよばれるページ上部のヘッダー部や、ページ番号であるノズルを本文の中に入れてないように分離する技術が必要となる。そうしなければページ番号を本文に取り込んでしまったり、ページとページに分かれる文章が音声にしたときに切れてしまったり、空白の音声が入ってしまうからである。

また、伝統的なボランティアの音訳と言われる作業では、グラフなどの図も説明して音声にするが、そのような本文と挿絵や図であるかといった違いが重要になるからである。

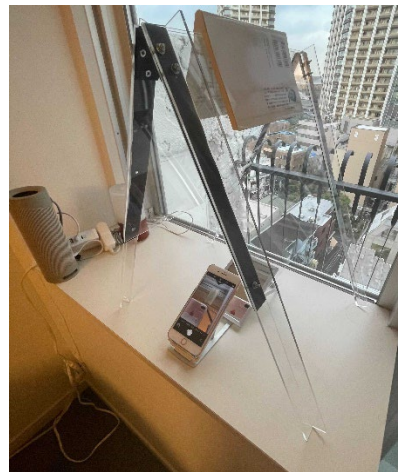
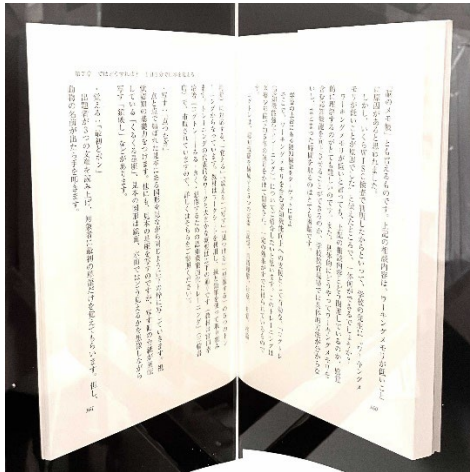
昨年までは、これを新書、文庫といったシリーズになっている書籍では、定型の版面デザインが行われるため、ある ISBN の本が特定の新書・文庫等の提携フォーマットであった場合、その定型情報からレイアウトの概要を得てレイアウト解析に利用するという方法をとった。OCRエンジンの性能だけに頼らず、外部ナレッジを使ったOCR精度の向上を目指したもので、精度は大きく向上した。

しかしながら、レイアウトが複雑なものや外部ナレッジにない書籍を読むときには、このレイアウトの解析の精度が高くないという欠点があった。

そこで本年度は、機械学習を使い、未知のレイアウトに対してもある程度モデルの汎化性を期待してレイアウト解析ができるような手法の研究を行うことを目標とした。

## 2. 各サブテーマの結果

### ① 見開きの台の試作



見開き固定が容易にできる台を複数試作したが、残念ながら実用に耐えるものは完成しなかった。例えば、上の写真は、その1つであるが、アクリルが反射し、逆にこれが精度を低下させる結果となった。透過率の低反射のアクリル板は非常にコストが高く、この方式での台は断念した。このように現YourEyesBoxのような製作コストでBoxなしにするような台に関しては難しいと判断し、本サブテーマはいったん試作品の製作をもって中断した。

## ② 事前画像処理の研究

図2のような湾曲したデータを上部の曲線をスプライン曲線によって近似し、直線に変換する関数（計算ルーチン）を考案した。この関数は勾配降下法を使うため、計算に時間がかかるため実用上はさらに研究を重ねることが必要であるとの結論に達した。

## ③ 機械学習を使ったレイアウト解析の研究

畳み込みニューラルネットワーク（CNN）と勾配ブースティングと呼ばれる機械学習の手法をハイブリッドにする方法を用いた。

レイアウト解析モデル - CNN as an image features extract

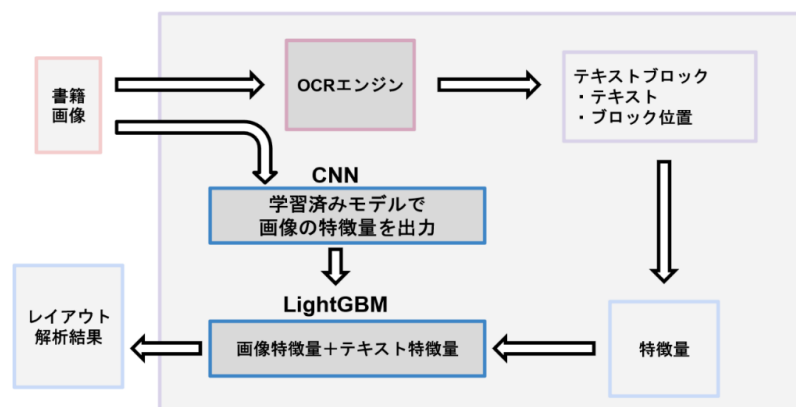


図3 CNN と勾配ブースティングのハイブリッドなモデル

機械学習では、教師データと呼ばれる見本のデータを用意することが必要だが、人間がレ

レイアウトの分類をアノテーションと呼ばれるラベル付けする作業が必要で、時間とコストが大きくかかる。一般的に、アノテーションによる特徴量が大きければ大きいほど学習にはよいが、アノテーションのコストとの兼ね合いになる。

今回は、画像ニューラルネットワークに入力し、その特徴量を得る方法、(パターン1)、OCRから得られたテキストデータを利用するものなど3パターンの方法を試みた。

	テキストから生成した特徴量	ベクトル化テキスト	CNNで生成した画像特徴量
パターン1	○		
パターン2	○	○	
パターン3	○		○

表 1. 特徴量の組み合わせのパターン

この結果、従来の新書や文庫のような文字主体のもの、レイアウトが文庫などよりも複雑で、段組されているようなもの、さらに技術書のようなレイアウトの複雑なものについて、見出しや本文の区別が従来よりもできるようになった。次に示すのは様々なレイアウトの書籍の紙面であるが、色分けされた部分がレイアウトの同一種類のパーツであり、うまく分類できていることがわかる。

### 例 3 : 技術書

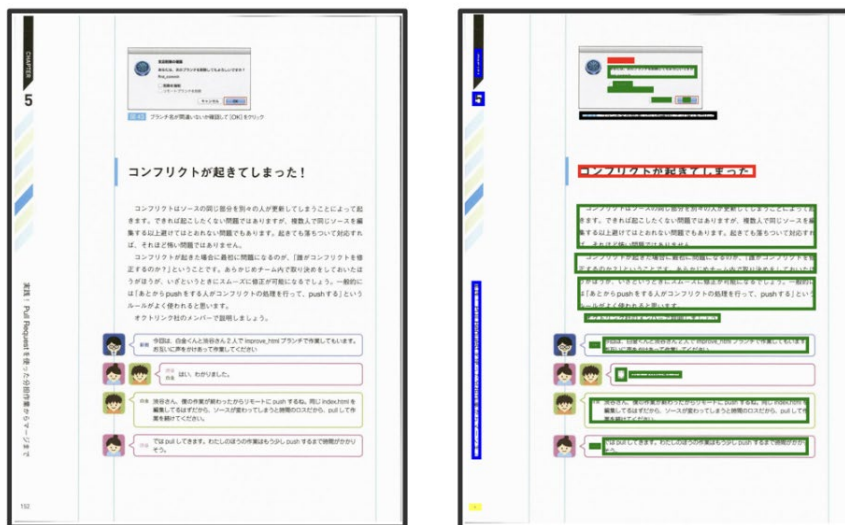


図 4 機械学習によって判断されたレイアウト解析結果

また、これら3つのパターンについて、精度に関する3つの指標 (Accuracy, macro-

f1,wehght-f1) についての調べたところ、以下のようになった。

	accuracy	macro-f1	weighted-f1
パターン1	0.921	0.87	0.92
パターン2	0.934	0.89	0.93
パターン3	0.920	0.85	0.92

表2. 学習時に与える特徴量別パターンに対する精度について

単純に比較できるものではないが、国立国会図書館の資料デジタル化プロジェクトのレイアウト解析で開発されたモデル[1]では Accuracy が 0.92 であるため、これに勝るとも劣らないモデルを構築することができたといえる。

[1] [https://www.jstage.jst.go.jp/article/pjsai/JSAI2020/0/JSAI2020\\_3Rin472/\\_pdf/-char/ja](https://www.jstage.jst.go.jp/article/pjsai/JSAI2020/0/JSAI2020_3Rin472/_pdf/-char/ja)